

Flow of Document

1. Introduction to Flume, Sqoop, Pig
2. Step I to put the data on hdfs using Flume
3. Step II to convert the xml file into csv using Pig Script
4. PIG UDF's code and explanation
5. Screen shots and explanation of commands and flow of script and program for Question 1
6. Screen shots and explanation of commands and flow of script and program for Question 2
7. Shell Scripting

Introduction

1. Flume

What is Flume?

Apache Flume is a tool/service/data ingestion mechanism for collecting aggregating and transporting large amounts of streaming data such as log data, events (etc...) from various webserves to a centralized data store.

It is a highly reliable, distributed, and configurable tool that is principally designed to transfer streaming data from various sources to HDFS.

What Flume Does?

Flume lets Hadoop users ingest high-volume streaming data into HDFS for storage. Specifically, Flume allows users to:

Feature	Description
Stream data	Ingest streaming data from multiple sources into Hadoop for storage and analysis
Insulate systems	Buffer storage platform from transient spikes, when the rate of incoming data exceeds the rate at which data can be written to the destination
Guarantee data delivery	Flume NG uses channel-based transactions to guarantee reliable message delivery. When a message moves from one agent to another, two transactions are started, one on the agent that delivers the event and the other on the agent that receives the event. This ensures guaranteed delivery semantics
Scale horizontally	To ingest new data streams and additional volume as needed

Table that shows the features of the flume

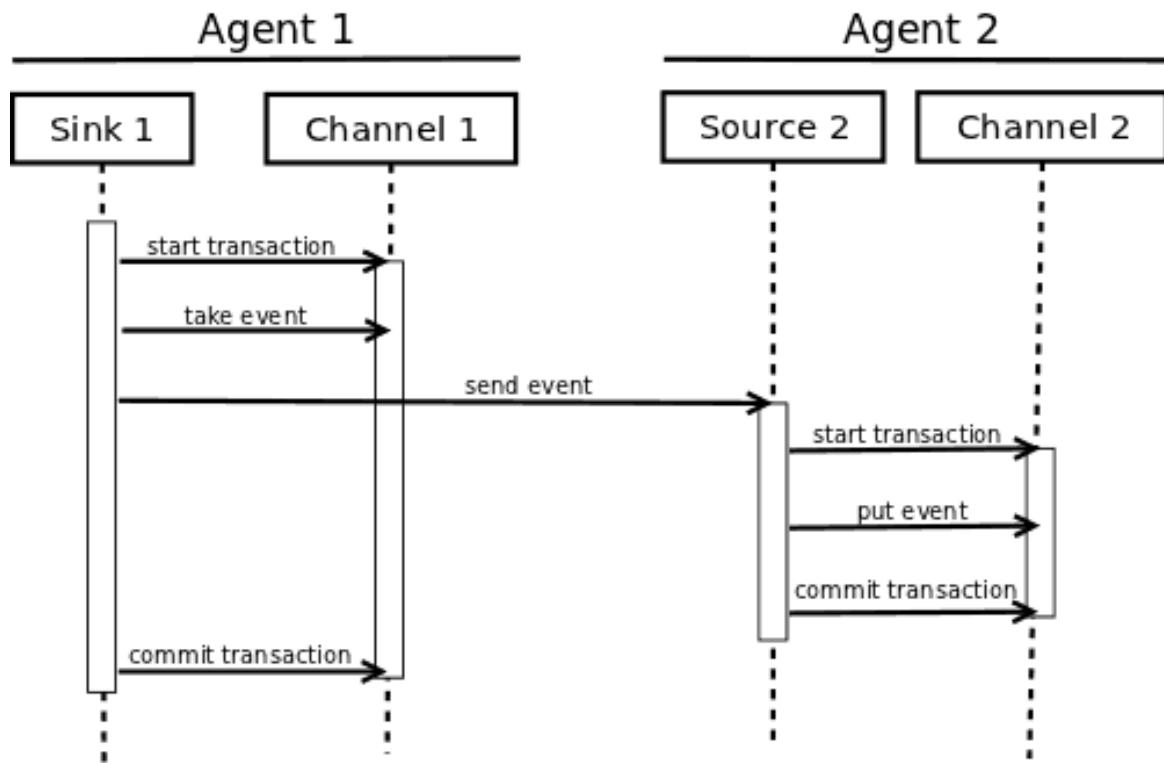


Fig. showing sink and source agents and flow of data

2. SQQOP

What is sqoop?

When Big Data storages and analyzers such as MapReduce, Hive, Hbase, Cassandra, Pig, etc. of the Hadoop ecosystem came into picture, they required a tool to interact with the relational database servers for importing and exporting the Big Data residing in them. Here, Sqoop occupies a place in the Hadoop ecosystem to provide feasible interaction between relational database server and Hadoop's HDFS.

Sqoop: “SQL to Hadoop and Hadoop to SQL”

Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases. It is provided by the Apache Software Foundation.

Where is Sqoop used?

Relational database systems are widely used to interact with the traditional business applications. So, relational database systems has become one of the sources that generate Big Data.

As we are dealing with Big Data, Hadoop stores and processes the Big Data using different processing frameworks like MapReduce, Hive, Hbase, Cassandra, Pig etc and storage frameworks like HDFS to achieve benefit of distributed computing and distributed storage. In order to store and analyze the Big Data from relational databases, Data need to be transferred between database systems and Hadoop Distributed File System (HDFS). Here, Sqoop comes into picture. Sqoop acts like a intermediate layer between Hadoop and relational database systems. You can import data and export data between relational database systems and Hadoop and its eco-systems directly using sqoop.

What Sqoop Does?

Apache Sqoop does the following to integrate bulk data movement between Hadoop and structured datastores:

Function	Benefit
Import sequential datasets from mainframe	Satisfies the growing need to move data from mainframe to HDFS
Import direct to ORCFiles	Improved compression and light-weight indexing for improved query performance
Data imports	Moves certain data from external stores and EDWs into Hadoop to optimize cost-effectiveness of combined data storage and processing
Parallel data transfer	For faster performance and optimal system utilization
Fast data copies	From external systems into Hadoop
Efficient data analysis	Improves efficiency of data analysis by combining structured data with unstructured data in a schema-on-read data lake
Load balancing	Mitigates excessive storage and processing loads to other systems

Table function and benefit of sqoop

3. PIG

What Is Pig?

Pig is a high level scripting language that is used with Apache Hadoop. Pig enables data workers to write complex data transformations without knowing Java. Pig's simple SQL-like scripting language is called Pig Latin, and appeals to developers already familiar with scripting languages and SQL.

Pig is complete, so you can do all required data manipulations in Apache Hadoop with Pig. Through the User Defined Functions(UDF) facility in Pig, Pig can invoke code in many language like Java. You can also embed Pig scripts in other languages. The result is that you can use Pig as a component to build larger and more complex applications that tackle real business problems.

To write data analysis programs, Pig provides a high-level language known as Pig Latin. This language provides various operators using which programmers can develop their own functions for reading, writing, and processing data.

To analyze data using Apache Pig, programmers need to write scripts using Pig Latin language. All these scripts are internally converted to Map and Reduce tasks. Apache Pig has a component known as Pig Engine that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs.

Features of Pig

Apache Pig comes with the following features –

- **Rich set of operators** – It provides many operators to perform operations like join, sort, filer, etc.
- **Ease of programming** – Pig Latin is similar to SQL and it is easy to write a Pig script if you are good at SQL.
- **Optimization opportunities** – The tasks in Apache Pig optimize their execution automatically, so the programmers need to focus only on semantics of the language.
- **Extensibility** – Using the existing operators, users can develop their own functions to read, process, and write data.
- **UDF's** – Pig provides the facility to create **User-defined Functions** in other programming languages such as Java and invoke or embed them in Pig Scripts.
- **Handles all kinds of data** – Apache Pig analyzes all kinds of data, both structured as well as unstructured. It stores the results in HDFS.

Pig UDF's:

In addition to the built-in functions, Apache Pig provides extensive support for User Defined Functions (UDF's). Using these UDF's, we can define our own functions and use them. The UDF support is provided in six programming languages, namely, Java, Jython, Python, JavaScript, Ruby and Groovy.

For writing UDF's, complete support is provided in Java and limited support is provided in all the remaining languages. Using Java, you can write UDF's involving all parts of the processing like data load/store, column transformation, and aggregation. Since Apache Pig has been written in Java, the UDF's written using Java language work efficiently compared to other languages.

In Apache Pig, we also have a Java repository for UDF's named Piggybank. Using Piggybank, we can access Java UDF's written by other users, and contribute our own UDF's

Types of UDF's in Java:

While writing UDF's using Java, we can create and use the following three types of functions –

- **Filter Functions** – The filter functions are used as conditions in filter statements. These functions accept a Pig value as input and return a Boolean value.
- **Eval Functions** – The Eval functions are used in FOREACH-GENERATE statements. These functions accept a Pig value as input and return a Pig result.
- **Algebraic Functions** – The Algebraic functions act on inner bags in a FOREACHGENERATE statement. These functions are used to perform full MapReduce operations on an inner bag.

Step I: Flume

We have used flume to place the data into HDFS. For that first we have created the directory /project using the command

Hadoop fs –mkdir /project.

Then used the following conf file to provide the configuration and the path to store the data into the HDFS.

```
agent1.sources = mysrc
agent1.sinks = hdfsdest
agent1.channels = mychannel

agent1.sources.mysrc.type = exec
agent1.sources.mysrc.command = hadoop dfs -put /home/nitesh/Downloads/statewiseDistrictwisePhysicalProgress.xml /project

agent1.sinks.hdfsdest.type = hdfs
agent1.sinks.hdfsdest.hdfs.path = hdfs://localhost:9000/flume_import

agent1.channels.mychannel.type = memory

agent1.sources.mysrc.channels = mychannel
agent1.sinks.hdfsdest.channel = mychannel
```

The above conf file is used to specify the path of the sink and the source also the channels are also set in this file.

Once the directory is made and conf file is set then the flume command is executed

Flume-ng agent -n agent1 -conf -f <path of conf file>

Execution of the command is shown in below screen shots:

```
nitesh@nitesh-Lenovo-Z50-70: ~
2.7.0/share/hadoop/mapreduce/hadoop-mapreduce-client-shuffle-2.7.0.jar:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/hadoop-mapreduce-e
xamples-2.7.0.jar:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/lib:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/lib/examples
/usr/local/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/sources:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/conf:/home/nitesh/apache-hive-2.1.0-b
in/lib/* -Djava.library.path=/usr/local/hadoop/hadoop-2.7.0/lib/native:/usr/local/hadoop/hadoop-2.7.0/lib/native org.apache.flume.node.Applic
ation -n agent1 -f /home/nitesh/apache-flume-1.6.0-bin/conf/filecopy.conf
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/apache-flume-1.6.0-bin/lib/slf4j-log4j12-1.6.1.jar!/:org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/nitesh/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4.1.jar!/:org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
17/05/23 22:30:04 INFO node.PollingPropertiesfileConfigurationProvider: Configuration provider starting
17/05/23 22:30:04 INFO node.PollingPropertiesfileConfigurationProvider: Reloading configuration file:/home/nitesh/apache-flume-1.6.0-bin/conf/f
ilecopy.conf
17/05/23 22:30:04 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/05/23 22:30:04 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/05/23 22:30:04 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/05/23 22:30:04 INFO conf.FlumeConfiguration: Added sinks: hdfsdest Agent: agent1
17/05/23 22:30:04 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent1]
17/05/23 22:30:04 INFO node.AbstractConfigurationProvider: Creating channels
17/05/23 22:30:04 INFO channel.DefaultChannelFactory: Creating instance of channel mychannel type memory
17/05/23 22:30:04 INFO node.AbstractConfigurationProvider: Created channel mychannel
17/05/23 22:30:04 INFO source.DefaultSourceFactory: Creating instance of source mysrc, type exec
17/05/23 22:30:04 INFO sink.DefaultSinkFactory: Creating instance of sink: hdfsdest, type: hdfs
17/05/23 22:30:04 INFO node.AbstractConfigurationProvider: Channel mychannel connected to [mysrc, hdfsdest]
17/05/23 22:30:04 INFO node.Application: Starting new configuration:{ sourceRunners:{mysrc=EventDrivenSourceRunner: { source:org.apache.flume.s
ource.ExecSource{name:mysrc,state:IDLE} } } sinkRunners:{hdfsdest=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@80d82fc coun
tGroup:{ name:null counters:{} } } } channels:{mychannel=org.apache.flume.channel.MemoryChannel{name: mychannel}} }
17/05/23 22:30:04 INFO node.Application: Starting Channel mychannel
17/05/23 22:30:04 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: CHANNEL, name: mychannel: Successfully register
ed new MBean.
17/05/23 22:30:04 INFO instrumentation.MonitoringCounterGroup: Component type: CHANNEL, name: mychannel started
17/05/23 22:30:04 INFO node.Application: Starting Sink hdfsdest
17/05/23 22:30:04 INFO node.Application: Starting Source mysrc
17/05/23 22:30:04 INFO source.ExecSource: Exec source starting with command:hadoop dfs -put /home/nitesh/Downloads/StatewiseDistrictwisePhysica
lProgress.xml /project/
17/05/23 22:30:04 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: SOURCE, name: mysrc: Successfully registered ne
w MBean.
17/05/23 22:30:04 INFO instrumentation.MonitoringCounterGroup: Component type: SOURCE, name: mysrc started
17/05/23 22:30:04 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: SINK, name: hdfsdest: Successfully registered ne
w MBean.
17/05/23 22:30:04 INFO instrumentation.MonitoringCounterGroup: Component type: SINK, name: hdfsdest started
17/05/23 22:30:06 INFO source.ExecSource: Command [hadoop dfs -put /home/nitesh/Downloads/StatewiseDistrictwisePhysicalProgress.xml /project/] ex
ited with 0
```

In the above screen shot we can see in the first red box the command to put in the hdfs and then we can find the file is written in the specified folder.

Here we can see the file is written in the specified folder as we use the ls command to locate the file in the specified folder then using the cat command for authenticity.

```
nitesh@nitesh-Lenovo-Z50-70:~  
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /project  
FOUND 1 items  
-rw-r--r-- 4 nitesh supergroup 717414 2017-05-23 22:30 /project/StatewiseDistrictwisePhysicalProgress.xml  
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -cat /project/StatewiseDistrictwisePhysicalProgress.xml  
<PhysicalProgress>  
    <row>  
        <State_Name>Andhra Pradesh</State_Name>  
        <District_Name>ADILABAD</District_Name>  
        <Project_Objectives_IHHL_BPL>247475</Project_Objectives_IHHL_BPL>  
        <Project_Objectives_IHHL_APL>148181</Project_Objectives_IHHL_APL>  
        <Project_Objectives_IHHL_TOTAL>395656</Project_Objectives_IHHL_TOTAL>  
        <Project_Objectives_SCW>0</Project_Objectives_SCW>  
        <Project_Objectives_School_Toilets>4462</Project_Objectives_School_Toilets>  
        <Project_Objectives_Anganwadi_Toilets>427</Project_Objectives_Anganwadi_Toilets>  
        <Project_Objectives_RSM>10</Project_Objectives_RSM>  
        <Project_Objectives_PC>0</Project_Objectives_PC>  
        <Project_Performance-IHHL_BPL>176300</Project_Performance-IHHL_BPL>  
        <Project_Performance-IHHL_APL>52431</Project_Performance-IHHL_APL>  
        <Project_Performance-IHHL_TOTAL>228731</Project_Performance-IHHL_TOTAL>  
        <Project_Performance-SCW>0</Project_Performance-SCW>  
        <Project_Performance-School_Toilets>4462</Project_Performance-School_Toilets>  
        <Project_Performance-Anganwadi_Toilets>427</Project_Performance-Anganwadi_Toilets>  
        <Project_Performance-RSM>0</Project_Performance-RSM>  
        <Project_Performance-PC>0</Project_Performance-PC>  
    </row>  
    <row>  
        <State_Name>Andhra Pradesh</State_Name>  
        <District_Name>ANANTAPUR</District_Name>  
        <Project_Objectives_IHHL_BPL>363314</Project_Objectives_IHHL_BPL>  
        <Project_Objectives_IHHL_APL>181335</Project_Objectives_IHHL_APL>  
        <Project_Objectives_IHHL_TOTAL>544649</Project_Objectives_IHHL_TOTAL>  
        <Project_Objectives_SCW>0</Project_Objectives_SCW>  
        <Project_Objectives_School_Toilets>3421</Project_Objectives_School_Toilets>  
        <Project_Objectives_Anganwadi_Toilets>284</Project_Objectives_Anganwadi_Toilets>  
        <Project_Objectives_RSM>10</Project_Objectives_RSM>  
        <Project_Objectives_PC>0</Project_Objectives_PC>  
        <Project_Performance-IHHL_BPL>366557</Project_Performance-IHHL_BPL>  
        <Project_Performance-IHHL_APL>42000</Project_Performance-IHHL_APL>  
        <Project_Performance-IHHL_TOTAL>408557</Project_Performance-IHHL_TOTAL>  
        <Project_Performance-SCW>0</Project_Performance-SCW>  
        <Project_Performance-School_Toilets>34258</Project_Performance-School_Toilets>  
        <Project_Performance-Anganwadi_Toilets>302</Project_Performance-Anganwadi_Toilets>  
  
nitesh@nitesh-Lenovo-Z50-70:~  
    <Project_Performance-School_Toilets>4122</Project_Performance-School_Toilets>  
    <Project_Performance-Anganwadi_Toilets>477</Project_Performance-Anganwadi_Toilets>  
    <Project_Performance-RSM>0</Project_Performance-RSM>  
    <Project_Performance-PC>0</Project_Performance-PC>  
    <row>  
        <State_Name>Andhra Pradesh</State_Name>  
        <District_Name>WEST GODAVARI</District_Name>  
        <Project_Objectives_IHHL_BPL>344272</Project_Objectives_IHHL_BPL>  
        <Project_Objectives_IHHL_APL>294425</Project_Objectives_IHHL_APL>  
        <Project_Objectives_IHHL_TOTAL>638697</Project_Objectives_IHHL_TOTAL>  
        <Project_Objectives_SCW>385</Project_Objectives_SCW>  
        <Project_Objectives_School_Toilets>4551</Project_Objectives_School_Toilets>  
        <Project_Objectives_Anganwadi_Toilets>216</Project_Objectives_Anganwadi_Toilets>  
        <Project_Objectives_RSM>10</Project_Objectives_RSM>  
        <Project_Objectives_PC>0</Project_Objectives_PC>  
        <Project_Performance-IHHL_BPL>319477</Project_Performance-IHHL_BPL>  
        <Project_Performance-IHHL_APL>294425</Project_Performance-IHHL_APL>  
        <Project_Performance-IHHL_TOTAL>613902</Project_Performance-IHHL_TOTAL>  
        <Project_Performance-SCW>837</Project_Performance-SCW>  
        <Project_Performance-School_Toilets>4524</Project_Performance-School_Toilets>  
        <Project_Performance-Anganwadi_Toilets>316</Project_Performance-Anganwadi_Toilets>  
        <Project_Performance-RSM>97</Project_Performance-RSM>  
        <Project_Performance-PC>0</Project_Performance-PC>  
    </row>  
    <row>  
        <State_Name>Arunachal Pradesh</State_Name>  
        <District_Name>ANJAW</District_Name>  
        <Project_Objectives_IHHL_BPL>3232</Project_Objectives_IHHL_BPL>  
        <Project_Objectives_IHHL_APL>0</Project_Objectives_IHHL_APL>  
        <Project_Objectives_IHHL_TOTAL>3232</Project_Objectives_IHHL_TOTAL>  
        <Project_Objectives_SCW>8</Project_Objectives_SCW>  
        <Project_Objectives_School_Toilets>44</Project_Objectives_School_Toilets>  
        <Project_Objectives_Anganwadi_Toilets>46</Project_Objectives_Anganwadi_Toilets>  
        <Project_Objectives_RSM>1</Project_Objectives_RSM>  
        <Project_Objectives_PC>0</Project_Objectives_PC>  
        <Project_Performance-IHHL_BPL>1490</Project_Performance-IHHL_BPL>  
        <Project_Performance-IHHL_APL>0</Project_Performance-IHHL_APL>  
        <Project_Performance-IHHL_TOTAL>1490</Project_Performance-IHHL_TOTAL>  
        <Project_Performance-SCW>2</Project_Performance-SCW>  
        <Project_Performance-School_Toilets>48</Project_Performance-School_Toilets>  
        <Project_Performance-Anganwadi_Toilets>48</Project_Performance-Anganwadi_Toilets>  
        <Project_Performance-RSM>0</Project_Performance-RSM>
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
<Project_Performance-PC>7</Project_Performance-PC>
</row>
<row>
<State_Name>West Bengal</State_Name>
<District_Name>SOUTH 24 PARAGANAS</District_Name>
<Project_Objectives_IHHL_BPL>628712</Project_Objectives_IHHL_BPL>
<Project_Objectives_IHHL_APL>521192</Project_Objectives_IHHL_APL>
<Project_Objectives_IHHL_TOTAL>1149904</Project_Objectives_IHHL_TOTAL>
<Project_Objectives_SCW>50</Project_Objectives_SCW>
<Project_Objectives_School_Toilets>8940</Project_Objectives_School_Toilets>
<Project_Objectives_Anganwadi_Toilets>5448</Project_Objectives_Anganwadi_Toilets>
<Project_Objectives_RSM>30</Project_Objectives_RSM>
<Project_Objectives_PC>0</Project_Objectives_PC>
<Project_Performance-IHHL_BPL>593712</Project_Performance-IHHL_BPL>
<Project_Performance-IHHL_APL>162487</Project_Performance-IHHL_APL>
<Project_Performance-IHHL_TOTAL>756199</Project_Performance-IHHL_TOTAL>
<Project_Performance-SCW>31</Project_Performance-SCW>
<Project_Performance-School_Toilets>7257</Project_Performance-School_Toilets>
<Project_Performance-Anganwadi_Toilets>1631</Project_Performance-Anganwadi_Toilets>
<Project_Performance-RSM>29</Project_Performance-RSM>
<Project_Performance-PC>29</Project_Performance-PC>
</row>
<row>
<State_Name>West Bengal</State_Name>
<District_Name>UTTAR DINAJPUR</District_Name>
<Project_Objectives_IHHL_BPL>257662</Project_Objectives_IHHL_BPL>
<Project_Objectives_IHHL_APL>301645</Project_Objectives_IHHL_APL>
<Project_Objectives_IHHL_TOTAL>559307</Project_Objectives_IHHL_TOTAL>
<Project_Objectives_SCW>50</Project_Objectives_SCW>
<Project_Objectives_School_Toilets>4806</Project_Objectives_School_Toilets>
<Project_Objectives_Anganwadi_Toilets>1556</Project_Objectives_Anganwadi_Toilets>
<Project_Objectives_RSM>30</Project_Objectives_RSM>
<Project_Objectives_PC>0</Project_Objectives_PC>
<Project_Performance-IHHL_BPL>148802</Project_Performance-IHHL_BPL>
<Project_Performance-IHHL_APL>180619</Project_Performance-IHHL_APL>
<Project_Performance-IHHL_TOTAL>329421</Project_Performance-IHHL_TOTAL>
<Project_Performance-SCW>30</Project_Performance-SCW>
<Project_Performance-School_Toilets>2562</Project_Performance-School_Toilets>
<Project_Performance-Anganwadi_Toilets>2041</Project_Performance-Anganwadi_Toilets>
<Project_Performance-RSM>17</Project_Performance-RSM>
<Project_Performance-PC>0</Project_Performance-PC>
</row>
</PhysicalProgress>nitesh@nitesh-Lenovo-Z50-70: ~$
```

The above screen shots show the data is written using flume command in the hdfs. The data is in the xml format.

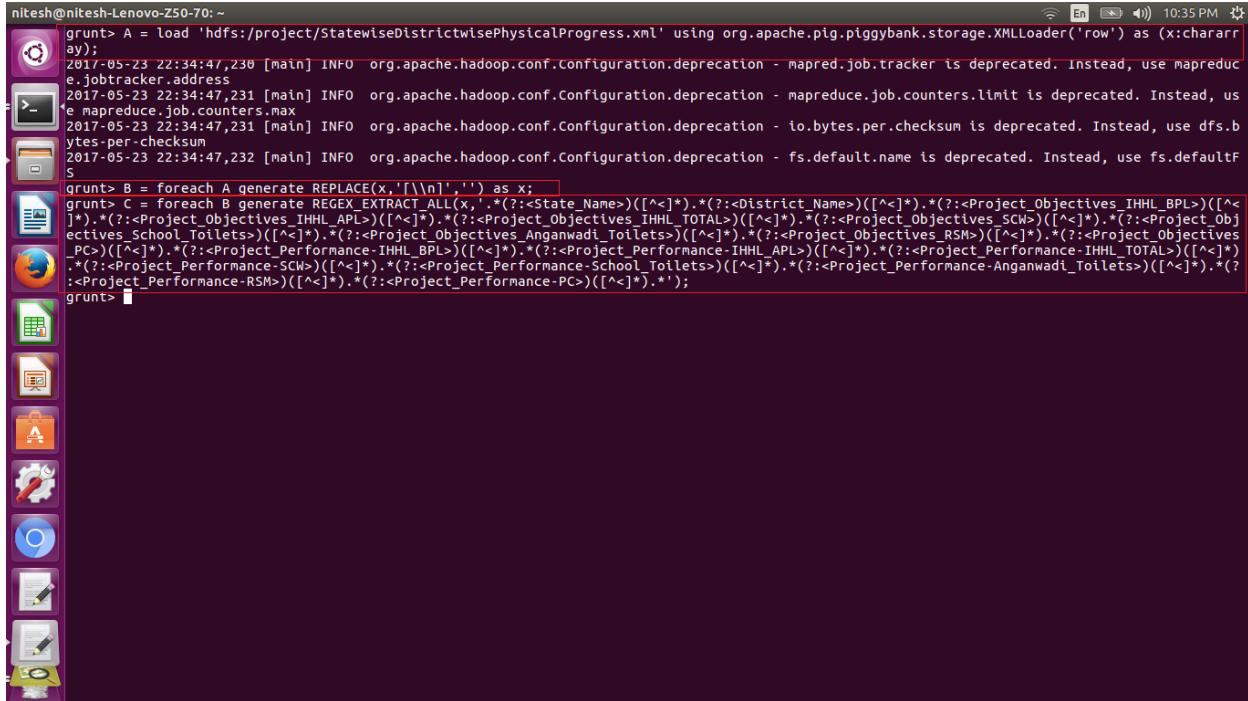
Step II: Conversion of xml file to csv file using the pig script

Here we have the xml file in the hdfs from the first step.

The file is loaded in the variable ‘A’

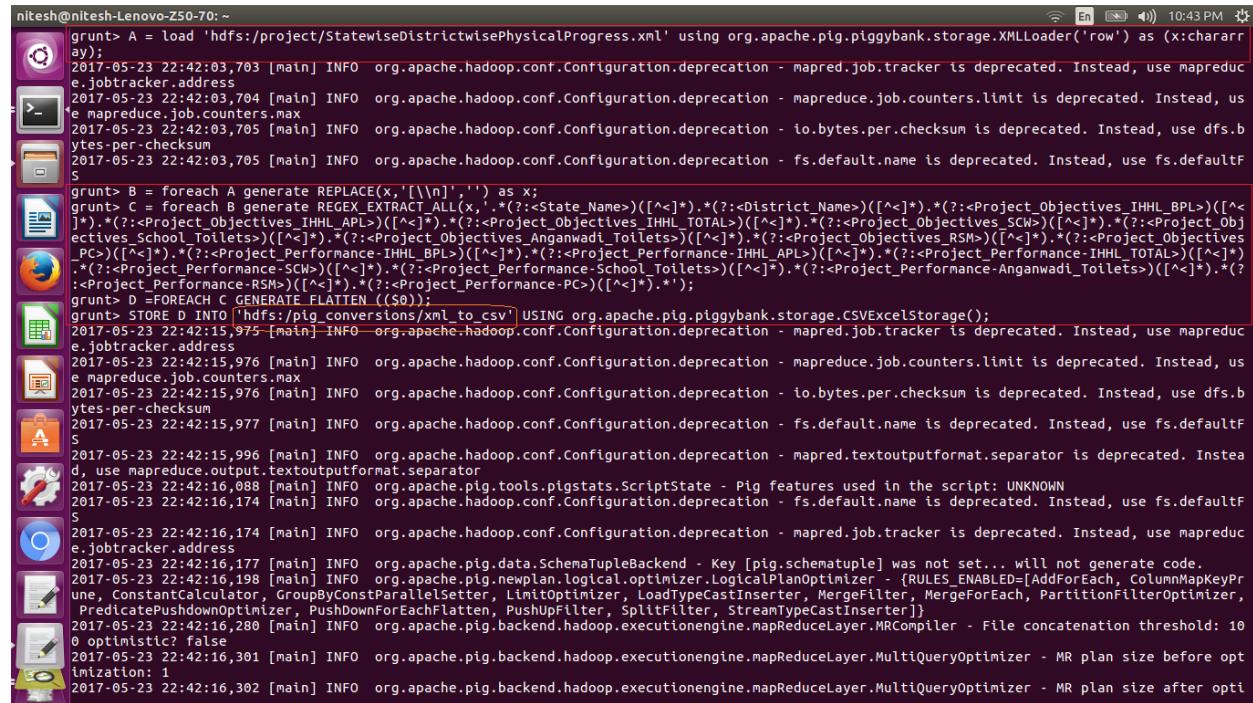
Using variable ‘B’ we replaced the new line by ‘,’

In variable ‘C’ we have removed the tags and the information is fetched such as <State_Name> is removed to get the name of the state only.



```
nitesh@nitesh-Lenovo-Z50-70: ~
grunt> A = load 'hdfs:/project/StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig.piggybank.storage.XMLLoader('row') as (x:chararray)
2017-05-23 22:34:47,230 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-23 22:34:47,231 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-23 22:34:47,231 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-23 22:34:47,232 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = foreach A generate REPLACE(x,'[\n]', '') as x;
grunt> C = foreach B generate REGEX_EXTRACT_ALL(x,.*(?:<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>)({[^<]*}).*';
grunt>
```

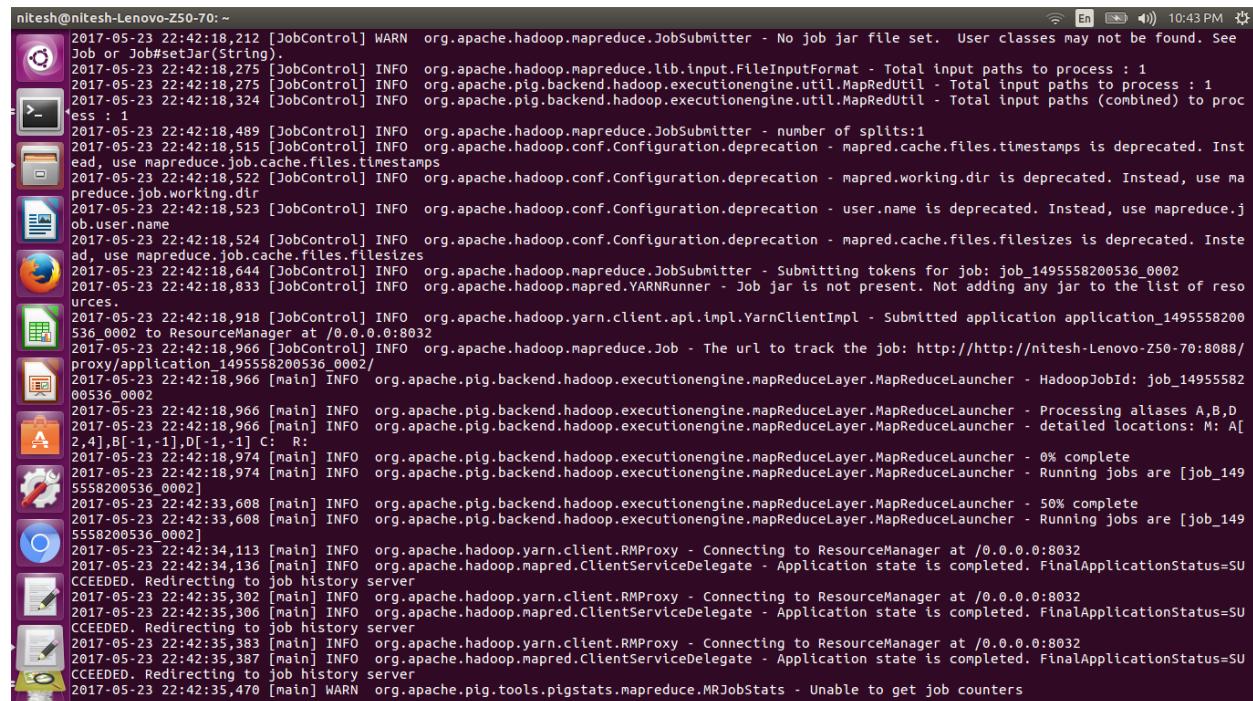
After that we used the store command to store the converted file into the desired location as we have stored it in the '/pig_conversion/xml_to_csv' by specifying hdfs: before the path means the described path is of hdfs.



```

nitesh@nitesh-Lenovo-Z50-70:~$ grunt> A = load 'hdfs:/project/StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig.piggybank.storage.XMLLoader('row') as (x:chararray);
2017-05-23 22:42:03,703 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-23 22:42:03,704 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-23 22:42:03,705 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-23 22:42:03,705 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = foreach A generate REPLACE(x,'[\n]', '') as x;
grunt> C = foreach B generate REGEX_EXTRACT_ALL(x,'*(?:<State_Name>)(&[^<]*).*(?:<District_Name>)(&[^<]*).*(?:<Project_Objectives_IHHL_BPL>)(&[^<]*).*(?:<Project_Objectives_IHHL_APL>)(&[^<]*).*(?:<Project_Objectives_IHHL_TOTAL>)(&[^<]*).*(?:<Project_Objectives_SCW>)(&[^<]*).*(?:<Project_Objectives_RSM>)(&[^<]*).*(?:<Project_Performance-IHHL_BPL>)(&[^<]*).*(?:<Project_Performance-IHHL_APL>)(&[^<]*).*(?:<Project_Performance-IHHL_TOTAL>)(&[^<]*).*(?:<Project_Performance-SCW>)(&[^<]*).*(?:<Project_Performance-School_Toilets>)(&[^<]*).*(?:<Project_Performance-RSM>)(&[^<]*).*(?:<Project_Performance-PC>)(&[^<]*).*');
grunts D =FOREACH C GENERATE FLATTEN ($0);
grunts STORE D INTO [hdfs:/pig_conversions/xml_to_csv] USING org.apache.pig.piggybank.storage.csvExcelStorage();
2017-05-23 22:42:15,975 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-23 22:42:15,976 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-23 22:42:15,976 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-23 22:42:15,977 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-23 22:42:15,996 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2017-05-23 22:42:16,088 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2017-05-23 22:42:16,174 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-23 22:42:16,174 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-23 22:42:16,177 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-05-23 22:42:16,198 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES_ENABLED=[AddForEach, ColumnMapKeyProcessor, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]]
2017-05-23 22:42:16,280 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 10
0 optimistic? false
2017-05-23 22:42:16,301 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2017-05-23 22:42:16,302 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization:

```



```

nitesh@nitesh-Lenovo-Z50-70:~$ 
2017-05-23 22:42:18,212 [JobControl] WARN  org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2017-05-23 22:42:18,275 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-23 22:42:18,275 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-05-23 22:42:18,324 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2017-05-23 22:42:18,489 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-23 22:42:18,515 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-23 22:42:18,522 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-23 22:42:18,523 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-23 22:42:18,524 [JobControl] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-23 22:42:18,644 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495558200536_0002
2017-05-23 22:42:18,833 [JobControl] INFO  org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.
2017-05-23 22:42:18,918 [JobControl] INFO  org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495558200
536_0002 to ResourceManager at /0.0.0.0:8032
2017-05-23 22:42:18,966 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495558200536_0002/
2017-05-23 22:42:18,966 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1495558200536_0002
2017-05-23 22:42:18,966 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,B,D
2017-05-23 22:42:18,966 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[2,4],B[-1,-1],D[-1,-1] C: R
2017-05-23 22:42:18,974 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2017-05-23 22:42:18,974 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495558200536_0002]
2017-05-23 22:42:33,008 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2017-05-23 22:42:33,008 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495558200536_0002]
2017-05-23 22:42:34,113 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-23 22:42:34,136 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-23 22:42:35,302 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-23 22:42:35,306 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-23 22:42:35,383 [main] INFO  org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-23 22:42:35,387 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-23 22:42:35,470 [main] WARN  org.apache.pig.tools.pigstats.mapreduce.MRJobStats - Unable to get job counters

```

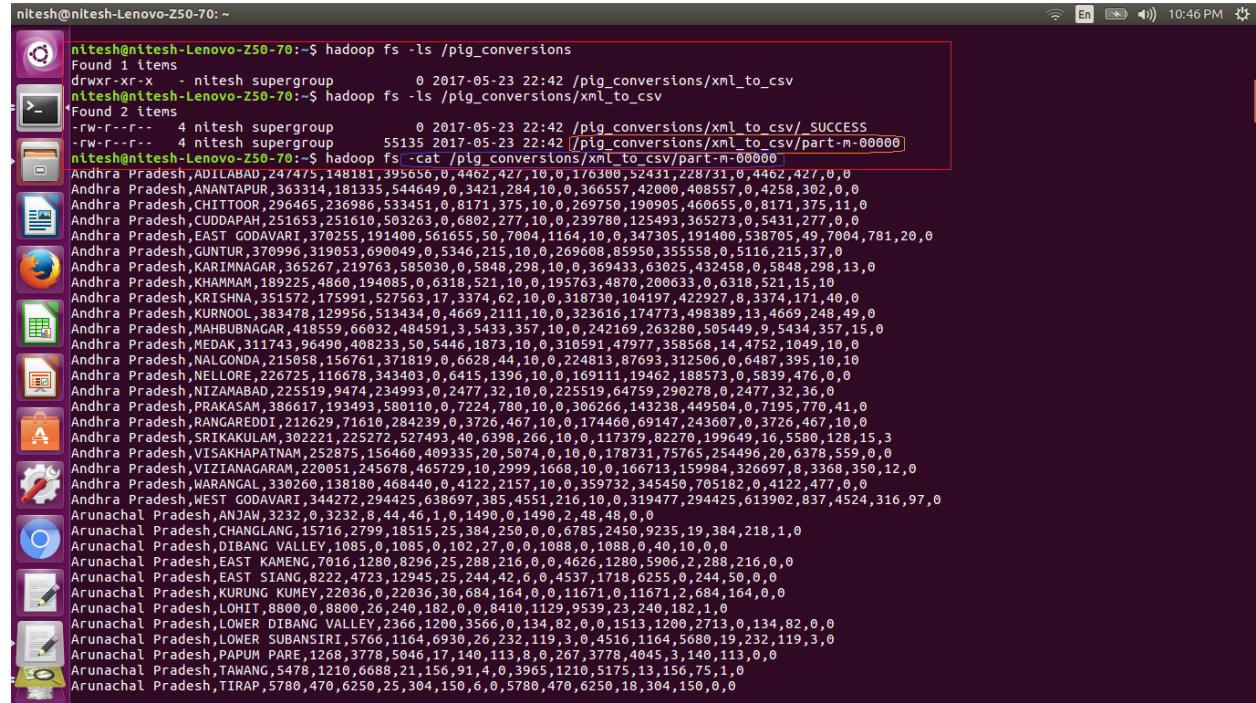
```
nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-23 22:42:35,482 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-23 22:42:16 2017-05-23 22:42:35 UNKNOWN
Success!
Job Stats (time in seconds):
Jobid Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
educetime Alias Feature Outputs
job_1495558200536_0002 1 0 5 5 5 0 0 0 A,B,D MAP_ONLY hdfs:/pig_conve
rsions/xml_to_csv,
Input(s):
Successfully read 0 records from: "hdfs:/project/StatewiseDistrictwisePhysicalProgress.xml"
Output(s):
Successfully stored 0 records in: "hdfs:/pig_conversions/xml_to_csv"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495558200536_0002
2017-05-23 22:42:35,484 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-23 22:42:35,488 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-23 22:42:35,565 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-23 22:42:35,570 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-23 22:42:35,615 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-23 22:42:35,619 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-23 22:42:35,659 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to
compute warning aggregation.
2017-05-23 22:42:35,660 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> ■
```

The above screen shot show that the csv file(converted from the xml) is stored successfully as we can see in the red blocks.

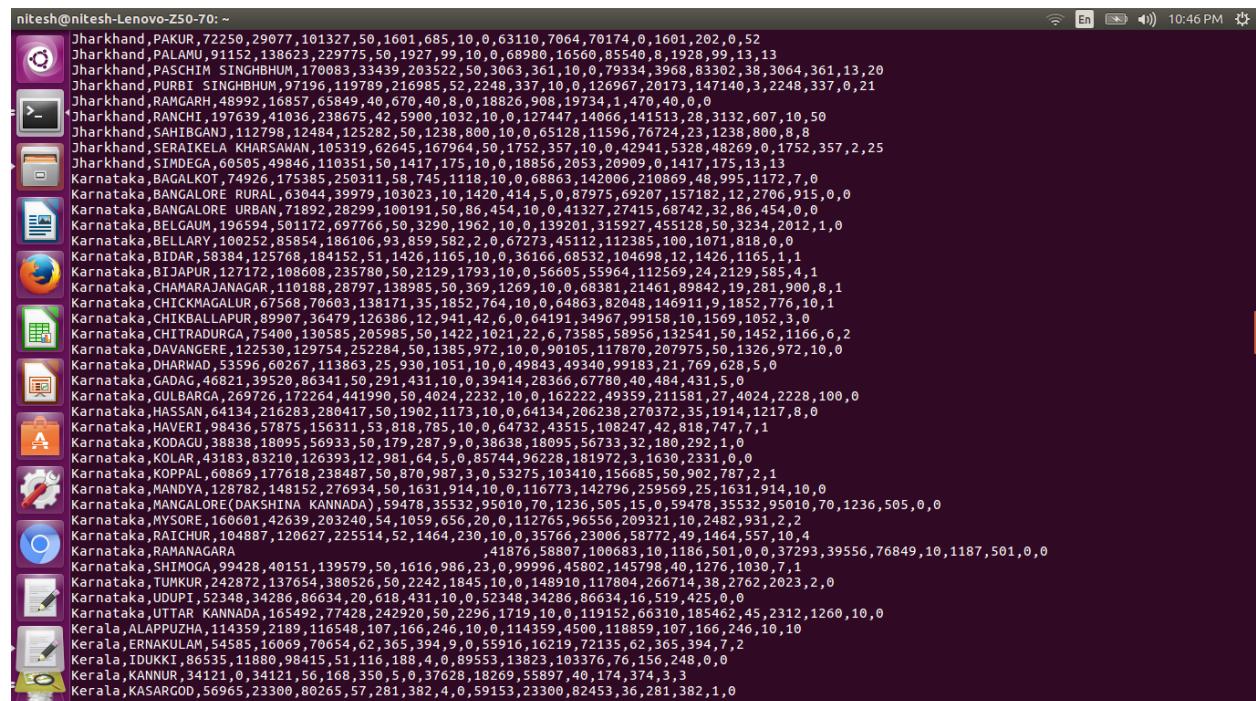
The file is stored as the part-m-00000 in the given path as we can seen in the below screen shots:

```
nitesh@nitesh-Lenovo-Z50-70: ~
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /pig_conversions
Found 1 items
drwxr-xr-x - nitesh supergroup 0 2017-05-23 22:42 /pig_conversions/xml_to_csv
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /pig_conversions/xml_to_csv
Found 2 items
-rw-r--r-- 4 nitesh supergroup 0 2017-05-23 22:42 /pig_conversions/xml_to_csv/_SUCCESS
-rw-r--r-- 4 nitesh supergroup 55135 2017-05-23 22:42 /pig_conversions/xml_to_csv/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ ■
```

Here we have performed cat command on the stored file to check if we got the desired converted file.



```
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /pig_conversions
Found 1 items
drwxr-xr-x - nitesh supergroup          0 2017-05-23 22:42 /pig_conversions/xml_to_csv
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /pig_conversions/xml_to_csv
Found 2 items
-rw-r--r-- 4 nitesh supergroup          0 2017-05-23 22:42 /pig_conversions/xml_to_csv/_SUCCESS
-rw-r--r-- 4 nitesh supergroup      55135 2017-05-23 22:42 /pig_conversions/xml_to_csv/part-n_00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -cat /pig_conversions/xml_to_csv/part-n_00000
Andhra Pradesh,ADILABAD,247475,148181,395656,0,4462,427,10,0,176300,52431,228731,0,4462,427,0,0
Andhra Pradesh,ANANTAPUR,363314,181335,544649,0,3421,284,10,0,366557,42000,408557,0,4258,302,0,0
Andhra Pradesh,CHITTOOR,296465,236986,533451,0,8171,375,10,0,269750,19985,466655,0,8171,375,11,0
Andhra Pradesh,CUDDAPAH,251053,251610,503263,0,6802,277,10,0,239780,125493,365273,0,5431,277,0,0
Andhra Pradesh,EAST GODAVARI,370255,191400,561655,50,7004,1164,10,0,347305,191400,538705,49,7004,781,20,0
Andhra Pradesh,GUNTHU,370996,319053,690049,0,5346,215,10,0,269608,85950,355558,0,5116,215,37,0
Andhra Pradesh,KARIMNAGAR,365267,219763,585030,0,5848,298,10,0,369433,63025,432458,0,5848,298,13,0
Andhra Pradesh,KHAMMAM,189225,4860,194085,0,6318,521,10,0,195763,4870,200633,0,6318,521,15,10
Andhra Pradesh,KRISHNA,351572,175991,527563,17,3374,62,10,0,318730,104197,422927,8,3374,171,40,0
Andhra Pradesh,KRISHNA,383478,129956,513434,0,4669,2111,10,0,323616,174773,498389,13,4669,248,49,0
Andhra Pradesh,MAHBUBNAGAR,418559,66032,484591,3,5433,357,10,0,242169,263280,505449,9,5434,357,15,0
Andhra Pradesh,MEDAK,311743,96490,488233,50,5446,1873,10,0,310591,47977,358568,14,4752,1849,10,0
Andhra Pradesh,NALGONDA,215058,156761,371819,0,6628,44,10,0,224813,87693,312586,0,6487,395,10,10
Andhra Pradesh,NELLORE,226725,116678,343403,0,6415,1396,10,0,169111,19462,188573,0,5839,476,0,0
Andhra Pradesh,NIZAMABAD,225519,9474,234993,0,2477,32,10,0,225519,64759,290278,0,2477,32,36,0
Andhra Pradesh,PAKASAN,386617,193493,580110,0,7224,780,10,0,306266,143238,49504,0,7195,770,41,0
Andhra Pradesh,RANGAREDDI,212629,71618,284239,0,3726,467,10,0,174460,69147,243607,0,3726,467,10,0
Andhra Pradesh,SRIRAKULAM,302221,225272,527493,40,6398,266,10,0,117379,82270,19649,16,5580,128,15,3
Andhra Pradesh,VISAKHAPATNAM,252875,156466,409335,20,5074,0,16,0,178731,75765,254496,20,6378,559,0,0
Andhra Pradesh,VIZIANAGARAM,220951,245678,465729,10,2999,1664,10,0,166713,159984,326697,8,3368,350,12,0
Andhra Pradesh,WARANGAL,330260,138180,468440,0,4122,2157,10,0,359732,345450,705182,0,4122,477,0,0
Andhra Pradesh,WEST GODAVARI,344272,294425,638697,385,4551,216,10,0,319477,294425,613902,837,4524,316,97,0
Arunachal Pradesh,ANJAN,3232,0,3232,8,44,16,10,1490,0,1490,2,48,48,0,0
Arunachal Pradesh,CHANGLANG,15716,2799,18515,25,384,250,0,0,6785,2456,9235,19,384,218,1,0
Arunachal Pradesh,DIBANG VALLEY,1085,0,1085,0,102,27,7,0,0,1088,0,1088,0,40,10,0,0
Arunachal Pradesh,EAST KAMENG,7016,1280,8299,25,288,216,0,0,4626,1280,5906,2,288,216,0,0
Arunachal Pradesh,EST SIANG,8222,4723,12945,25,244,42,6,0,4537,1718,6255,0,244,50,0,0
Arunachal Pradesh,KURUNG KUNEY,22036,22036,30,684,164,0,0,11671,0,11671,2,684,164,0,0
Arunachal Pradesh,LOHIT,8800,0,8800,26,240,182,0,0,8410,1129,9539,23,240,182,1,0
Arunachal Pradesh,LOWER DIBANG VALLEY,2366,1200,3566,0,134,82,0,0,1513,1200,2713,0,134,82,0,0
Arunachal Pradesh,LOWER SUBANSIRI,5766,1164,6930,26,232,119,3,0,4516,1164,5680,19,232,119,3,0
Arunachal Pradesh,PAPUM PARE,1268,3778,5046,17,140,113,8,0,267,3778,4045,3,140,113,0,0
Arunachal Pradesh,TAWANG,5478,1210,6688,21,156,91,4,0,3965,1210,5175,13,156,75,1,0
Arunachal Pradesh,TIRAP,5780,470,6250,25,304,150,6,0,5780,470,6250,18,304,150,0,0
```



```
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /pig_conversions
Found 1 items
drwxr-xr-x - nitesh supergroup          0 2017-05-23 22:42 /pig_conversions/xml_to_csv
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /pig_conversions/xml_to_csv
Found 2 items
-rw-r--r-- 4 nitesh supergroup          0 2017-05-23 22:42 /pig_conversions/xml_to_csv/_SUCCESS
-rw-r--r-- 4 nitesh supergroup      55135 2017-05-23 22:42 /pig_conversions/xml_to_csv/part-n_00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -cat /pig_conversions/xml_to_csv/part-n_00000
Jharkhand,PAKUR,72250,29077,191327,50,1601,685,10,0,63110,7064,70174,0,1601,202,0,52
Jharkhand,PALAMU,91152,138623,229775,56,1927,99,10,0,68980,16560,85540,8,1928,99,13,13
Jharkhand,PASCHIM SINGHBUM,170083,33439,203522,50,3063,361,10,0,79334,3968,83302,38,3064,361,13,20
Jharkhand,PURBI SINGHBUM,97196,119789,216989,52,2248,337,10,0,126967,20173,147140,3,2248,337,0,21
Jharkhand,RAMGARH,48992,16857,65849,46,670,40,8,0,18826,908,19734,1,470,40,0,0
Jharkhand,RANCHI,197639,41036,42,5900,1032,10,0,127447,14066,141513,28,3132,607,10,50
Jharkhand,SAHIBGANJ,112798,12484,125282,50,1238,800,10,0,65128,11596,76724,23,1238,800,8,8
Jharkhand,SEKAIKHARA KHARASWAN,105319,62645,167964,50,1752,357,10,0,42941,5328,48269,0,1752,357,2,25
Jharkhand,SIMDEGA,60505,49846,110851,50,1417,175,10,0,18856,2053,20999,0,1417,175,13,13
Karnataka,BAGALKOT,74926,175385,250311,58,745,1118,10,0,68863,142006,210869,49,995,1172,7,0
Karnataka,BANGALORE RURAL,63044,39979,103023,10,1420,414,5,0,87957,69207,157182,12,2706,915,0,0
Karnataka,BANGALORE URBAN,71892,28299,100191,50,86,454,18,0,41327,27415,68742,32,86,454,0,0
Karnataka,BELGAUM,196594,501172,697766,50,3290,1962,10,0,139201,315927,455128,50,3234,2012,1,0
Karnataka,BELLARY,106252,85854,186106,93,859,582,2,0,67273,45112,112385,100,1671,818,0,0
Karnataka,BIDAR,53834,125768,184152,51,1426,1165,10,0,36160,68532,104698,12,1426,1165,1,1
Karnataka,BIJAPUR,127172,108608,235780,50,2129,1793,10,0,55605,55964,112569,24,2129,585,4,1
Karnataka,CHAMARAJANAGAR,110188,28797,138985,50,369,1269,10,0,68381,21461,89842,19,281,900,8,1
Karnataka,CHICKMAGALUR,67568,70663,138171,35,1852,764,10,0,64863,82048,146911,9,1852,776,16,1
Karnataka,CHIKBALLAPUR,88997,36479,126386,12,941,42,6,0,64191,34967,99158,10,1569,1052,3,0
Karnataka,CHITRADURGA,75400,130585,205985,50,1422,1021,22,6,73585,58596,132541,50,1452,1166,6,2
Karnataka,DAVANGERE,122530,129754,252284,50,1385,972,10,0,90105,117870,207975,50,1326,972,10,0
Karnataka,DHARWAD,53596,60267,113863,25,930,1051,10,0,49843,49340,99183,21,769,628,5,0
Karnataka,DAGAD,48621,39520,86341,50,291,431,10,0,39414,28366,67780,40,484,431,5,0
Karnataka,GULBARGA,269726,172264,441990,50,4624,2232,10,0,162222,49359,211581,27,4024,2228,100,0
Karnataka,HASSAN,64134,216283,286417,50,1902,1173,10,0,64134,206238,270373,35,1914,1217,8,0
Karnataka,HAVERI,98436,57875,156311,53,818,785,10,0,64732,43515,188247,42,818,747,7,1
Karnataka,KODAGU,38838,18095,56933,50,179,287,9,0,38638,18095,56733,32,180,292,1,0
Karnataka,KOLAR,43183,83210,126393,12,981,64,5,0,85744,96228,181972,3,1630,2331,0,0
Karnataka,KOPPAL,60869,177618,238487,50,870,987,3,0,53275,103410,156683,50,902,787,2,1
Karnataka,MANDYA,128782,148152,276934,50,1631,914,10,0,116773,142796,259568,25,1631,914,10,0
Karnataka,MANGALORE(DAKSHINA KANNADA),59478,35532,95010,70,1236,505,0,59478,35532,95010,70,1236,505,0,0
Karnataka,MYSORE,106061,42639,203240,54,1059,656,20,0,112765,96556,209321,10,2482,931,2,2
Karnataka,RAICHUR,104887,120627,225514,52,1464,230,10,0,35766,23086,58772,49,1464,557,10,4
Karnataka,RAMANAGARA,104887,120627,225514,52,1464,230,10,0,35766,23086,58772,49,1464,557,10,4
Karnataka,SHIMOGA,99428,40151,139579,50,1616,986,23,0,99996,45882,145798,40,1276,1030,7,1
Karnataka,TUMKUR,242872,137654,380526,50,2242,1845,16,0,148910,117804,266714,38,2762,2023,2,0
Karnataka,UTTAR KANNADA,165492,77428,242920,50,2296,1719,10,0,119152,66310,185462,45,2312,1260,10,0
Kerala,ALAPPUZHA,114359,2189,116548,167,166,246,10,0,114359,4500,118859,107,166,246,10,10
Kerala,ERNAKULAM,54585,16669,70654,62,365,394,9,0,55916,16219,72135,62,365,394,7,2
Kerala,IDUKKI,86535,11880,98415,51,116,188,4,0,89553,13823,103376,76,156,248,0,0
Kerala,KANNUR,34121,0,34121,56,168,350,5,0,37628,18269,55897,40,174,374,3,3
Kerala,KASARGOD,56965,23300,80265,57,281,382,4,0,59153,23300,82453,36,281,382,1,0
```

```

nitesh@nitesh-Lenovo-Z50-70: ~
Uttar Pradesh, SANT KABIR NAGAR, 121112, 227284, 348396, 35, 2368, 991, 4, 0, 93881, 184716, 198597, 35, 2368, 991, 3, 0
Uttar Pradesh, SANT RAVIDAS NAGAR( BHADOHI), 75119, 103247, 178366, 43, 2083, 1233, 5, 0, 69904, 85676, 155580, 43, 1602, 1233, 1, 0
Uttar Pradesh, SHAHJAHANPUR, 194645, 114246, 308891, 18, 4842, 1347, 6, 0, 194959, 94064, 289023, 10, 4842, 1347, 6, 9
Uttar Pradesh, SHRAVASTI, 104902, 54772, 159674, 10, 1838, 650, 3, 0, 98761, 41818, 140579, 10, 1838, 650, 0, 0
Uttar Pradesh, SIDDHARTHNAGAR, 139597, 133650, 273247, 50, 4128, 1481, 5, 0, 124597, 111651, 236248, 50, 4128, 1481, 2, 0
Uttar Pradesh, SITAPUR, 305299, 255574, 560873, 25, 7397, 2307, 3, 1, 273463, 185138, 458601, 25, 5874, 2307, 1, 1
Uttar Pradesh, SONBHADRA, 138370, 79419, 217789, 113, 3176, 1051, 0, 0, 138370, 57900, 196270, 113, 2978, 755, 10, 0
Uttar Pradesh, SULTANPUR, 168843, 262071, 430914, 40, 4898, 2244, 6, 0, 168843, 215646, 384489, 40, 4898, 2244, 6, 0
Uttar Pradesh, UNNAO, 229599, 141734, 371333, 20, 4700, 1683, 8, 0, 223630, 131996, 355626, 20, 4700, 1683, 2, 2
Uttar Pradesh, VARANASI, 105408, 249056, 354464, 47, 4471, 1532, 0, 0, 102430, 147309, 249739, 47, 2516, 1283, 3, 0
Uttarakhand, ALMORA, 45572, 48151, 93723, 40, 214, 247, 18, 0, 32277, 38661, 70938, 30, 239, 4, 0, 0
Uttarakhand, BAGESWAR, 21447, 16508, 37955, 20, 111, 149, 3, 0, 17859, 18603, 36462, 4, 105, 2, 2, 0
Uttarakhand, CHAMOLI, 27147, 26145, 53292, 18, 265, 115, 7, 2, 26043, 25844, 51887, 6, 222, 7, 3, 0
Uttarakhand, CHAMPAWAT, 22991, 9085, 32076, 20, 375, 78, 2, 1, 18323, 14178, 32501, 3, 160, 6, 1, 0
Uttarakhand, DEHRADUN, 37212, 24463, 61675, 0, 497, 19, 5, 1, 31724, 24651, 56373, 0, 396, 19, 1, 0
Uttarakhand, HARIDWAR, 42500, 90074, 132574, 60, 200, 20, 2, 0, 38188, 77694, 115882, 37, 100, 13, 0, 0
Uttarakhand, NAINITAL, 14314, 26585, 40899, 58, 208, 95, 6, 2, 18039, 29625, 47064, 1, 208, 19, 2, 0
Uttarakhand, PAURI(GARHWAL), 53399, 37146, 90545, 50, 620, 173, 10, 0, 35852, 42518, 78370, 8, 367, 23, 3, 0
Uttarakhand, PITHORAGARH, 41110, 29253, 78363, 50, 670, 215, 6, 2, 34597, 31566, 66163, 10, 439, 102, 3, 0
Uttarakhand, RUDRAPRAYAG, 13150, 22322, 35472, 40, 100, 124, 3, 0, 13810, 17522, 31332, 9, 100, 70, 2, 0
Uttarakhand, TEHRI GARHWAL, 55173, 34580, 89753, 30, 307, 120, 8, 0, 46293, 35076, 81369, 0, 486, 10, 0, 0
Uttarakhand, UDHAM SINGH NAGAR, 39427, 59658, 99085, 50, 118, 123, 5, 2, 37664, 51384, 88988, 0, 119, 75, 4, 1
Uttarakhand, UTTARKASHI, 28189, 20780, 48889, 58, 249, 123, 4, 8, 25523, 21017, 46540, 4, 248, 0, 3, 0
West Bengal, BANKURA, 198152, 333832, 531984, 58, 7544, 4130, 29, 0, 105545, 243191, 348736, 46, 7687, 1340, 26, 0
West Bengal, BARDHAMAN, 700047, 341920, 1641967, 133, 9891, 7988, 10, 0, 601906, 277914, 879820, 140, 9890, 7724, 10, 19
West Bengal, BIRBHAMU, 338989, 299893, 638882, 50, 5617, 3816, 4, 22, 266347, 186599, 452946, 58, 5563, 2233, 19, 0
West Bengal, COOCH BEHAR, 335236, 254422, 589658, 50, 3715, 1718, 15, 0, 262294, 164038, 426332, 144, 5764, 1818, 15, 12
West Bengal, DAKSHIN DINAJPUR, 182621, 194577, 377198, 50, 3712, 2642, 10, 0, 181453, 49448, 233661, 19, 2632, 939, 8, 8
West Bengal, DARJEELING, 66648, 130666, 196714, 58, 1784, 468, 0, 0, 32921, 3035, 35956, 18, 1435, 574, 8, 8
West Bengal, HOOGHLY, 271737, 195510, 467247, 53, 6821, 4168, 19, 0, 269779, 191294, 461073, 49, 6764, 3435, 18, 18
West Bengal, HOWRAH, 231860, 143308, 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, 439290, 25, 6578, 4664, 17, 14
West Bengal, MALDA, 452324, 270288, 722532, 58, 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, 566031, 210, 10149, 2882, 8, 17
West Bengal, MIDNAPUR WEST, 509496, 432096, 941592, 50, 16498, 5825, 10, 0, 596291, 322659, 918956, 73, 13452, 2787, 0, 0
West Bengal, MURSHIDABAD, 702442, 506963, 1209405, 58, 10260, 7012, 18, 0, 498998, 198174, 697172, 47, 7838, 2423, 26, 26
West Bengal, NADIA, 346696, 278335, 625031, 58, 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

```

As we can see from the above screen shots the output is the desired one i.e. comma separated values and tags are deleted.

UDF

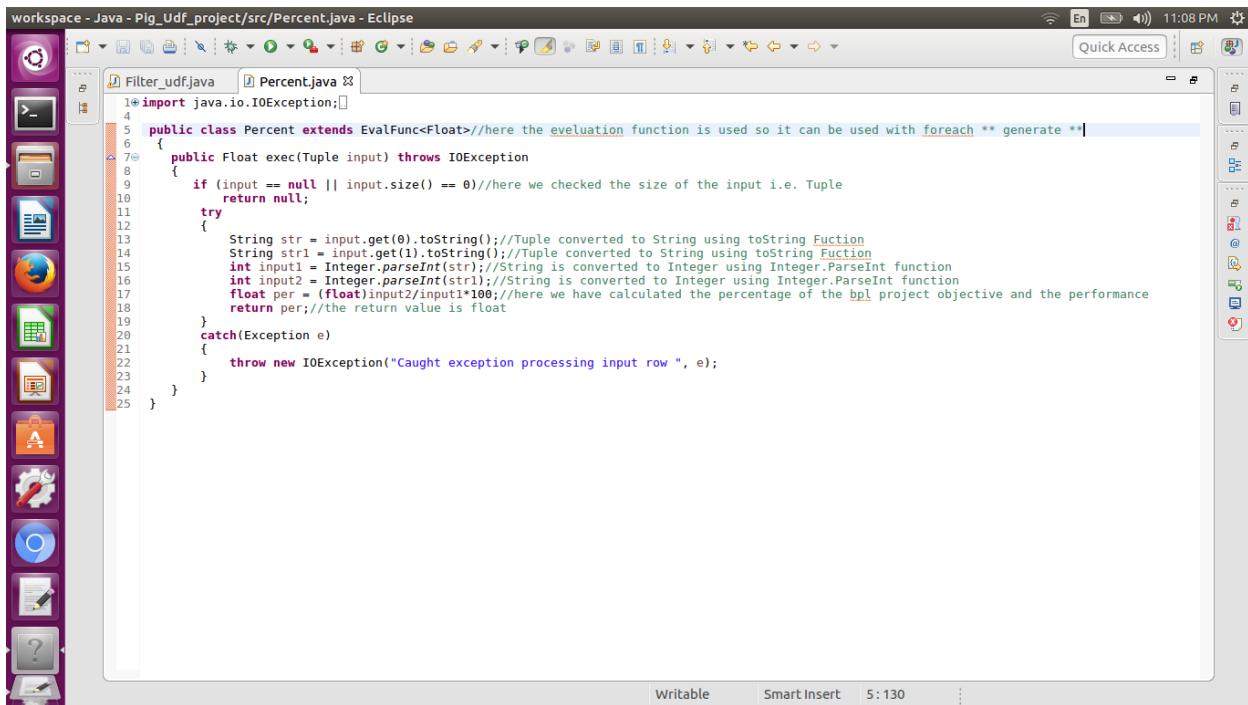
User Defined Function(UDF's) are the functions that programmer define for the purpose to be achieved.

Here we have defined two UDF's one to fetch the percentage and the other one to filter the data on the basis of the achieved performance percentage.

Percent UDF:

Steps:

- a. We have extended class to the EvalFunc which is used to generate the desired output same as the avg,sum,etc.
- b. The two arguments are passed to this udf and the percentage is calculated and returned
- c. The first input is stored in str and converted to string similarly for the second we used the variable str1.
- d. Both the str and str1 variables are now parsed to integer and the percentage is calculated.
- e. The percentage is stored in the float variable and the return type of this udf is float. So we get the percentage in the float datatype.



The screenshot shows the Eclipse IDE interface with the workspace titled "Java - Pig_Udf_project/src/Percent.java - Eclipse". The code editor displays the "Percent.java" file, which contains the following Java code:

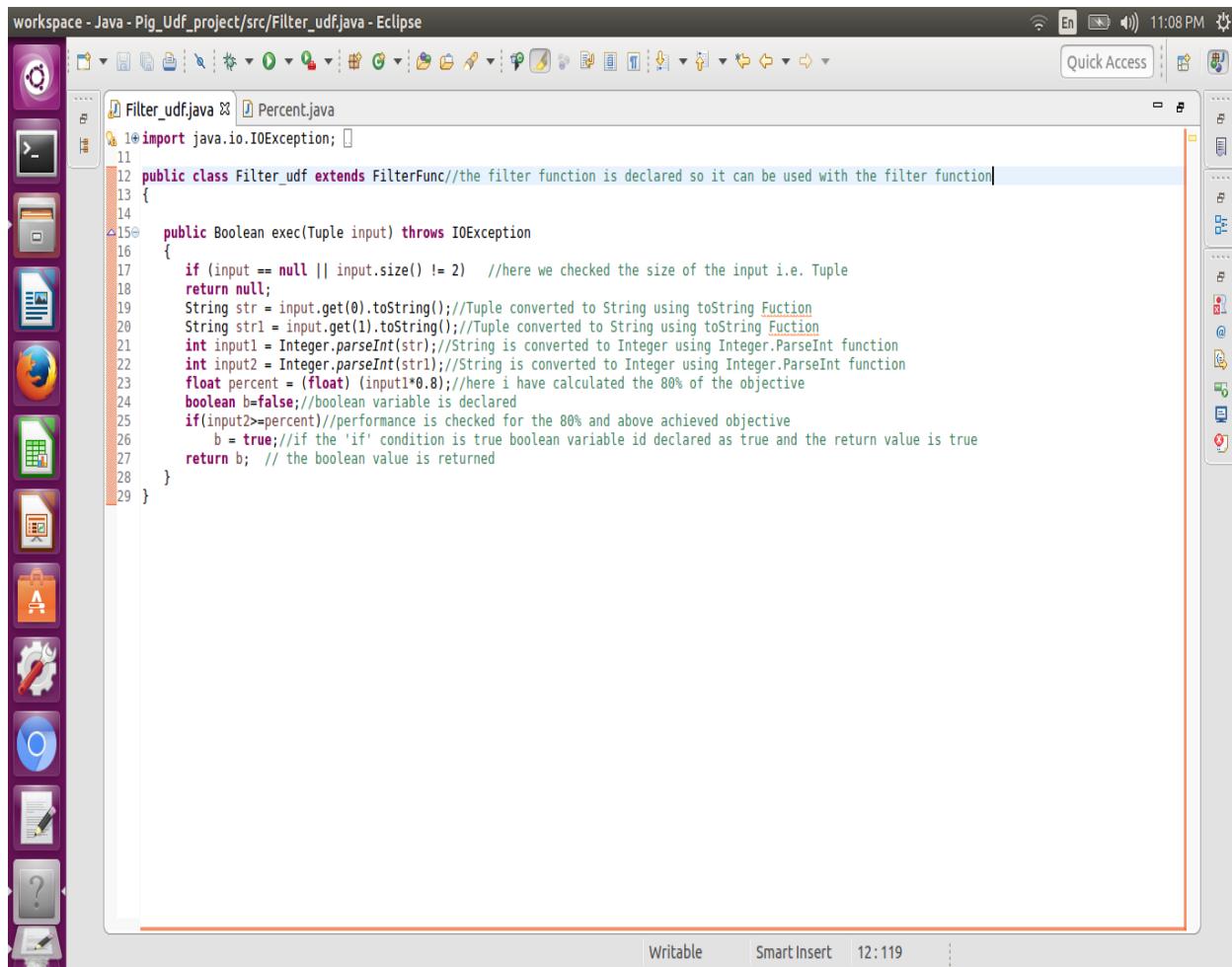
```
workspace - Java - Pig_Udf_project/src/Percent.java - Eclipse
File Edit View Insert Run Project Properties Help
1+ import java.io.IOException;
2+
3+ public class Percent extends EvalFunc<Float> //here the evaluation function is used so it can be used with foreach ** generate **
4+ {
5+     public Float exec(Tuple input) throws IOException
6+     {
7+         if (input == null || input.size() == 0) //here we checked the size of the input i.e. Tuple
8+             return null;
9+         try
10+         {
11+             String str = input.get(0).toString(); //Tuple converted to String using toString Function
12+             String str1 = input.get(1).toString(); //Tuple converted to String using toString Function
13+             int input1 = Integer.parseInt(str); //String is converted to Integer using Integer.parseInt function
14+             int input2 = Integer.parseInt(str1); //String is converted to Integer using Integer.parseInt function
15+             float per = (float)input2/input1*100; //here we have calculated the percentage of the bpl project objective and the performance
16+             return per; //the return value is float
17+         }
18+         catch(Exception e)
19+         {
20+             throw new IOException("Caught exception processing input row ", e);
21+         }
22+     }
23+ }
24+
25+ }
```

The code implements a UDF named "Percent" that extends the "EvalFunc<Float>" class. It takes a single tuple input and returns a float value representing the percentage. The code uses the "toString" method to convert tuple elements to strings, then parses them back into integers using "Integer.parseInt". Finally, it calculates the percentage by dividing the second integer by the first and multiplying by 100. A catch block handles any exceptions that might occur during the conversion or calculation.

Here we have a filter udf which is used to filter the data for the condition we want.

Steps

- a. Here the class is extended to the FilterFunc and the return type of the udf is Boolean.
- b. This udf have two arguments passed and the logic to check whether the performance have achieved 80% or not.
- c. If 80% is achieved then return value is true else false
- d. This udf is implemented in the filter function of the pig script to filter the data on the basis of the percentage achieved in reference to the objective.



The screenshot shows the Eclipse IDE interface with the title bar "workspace - Java - Pig_Udf_project/src/Filter_udf.java - Eclipse". The main editor window displays the Java code for "Filter_udf.java". The code implements a FilterFunc to filter data based on a performance threshold of 80%.

```
workspace - Java - Pig_Udf_project/src/Filter_udf.java - Eclipse
File Edit View Insert Run Project Properties Help
1 import java.io.IOException;
2
3 public class Filter_udf extends FilterFunc//the filter function is declared so it can be used with the filter function
4 {
5
6     public Boolean exec(Tuple input) throws IOException
7     {
8         if (input == null || input.size() != 2) //here we checked the size of the input i.e. Tuple
9             return null;
10        String str = input.get(0).toString(); //Tuple converted to String using toString Function
11        String str1 = input.get(1).toString(); //Tuple converted to String using toString Function
12        int input1 = Integer.parseInt(str); //String is converted to Integer using Integer.ParseInt function
13        int input2 = Integer.parseInt(str1); //String is converted to Integer using Integer.ParseInt function
14        float percent = (float) (input1*0.8); //here i have calculated the 80% of the objective
15        boolean b=false; //boolean variable is declared
16        if(input2>=percent)//performance is checked for the 80% and above achieved objective
17            b = true; //if the 'if' condition is true boolean variable id declared as true and the return value is true
18        return b; // the boolean value is returned
19    }
20}
```

Problem statement 1

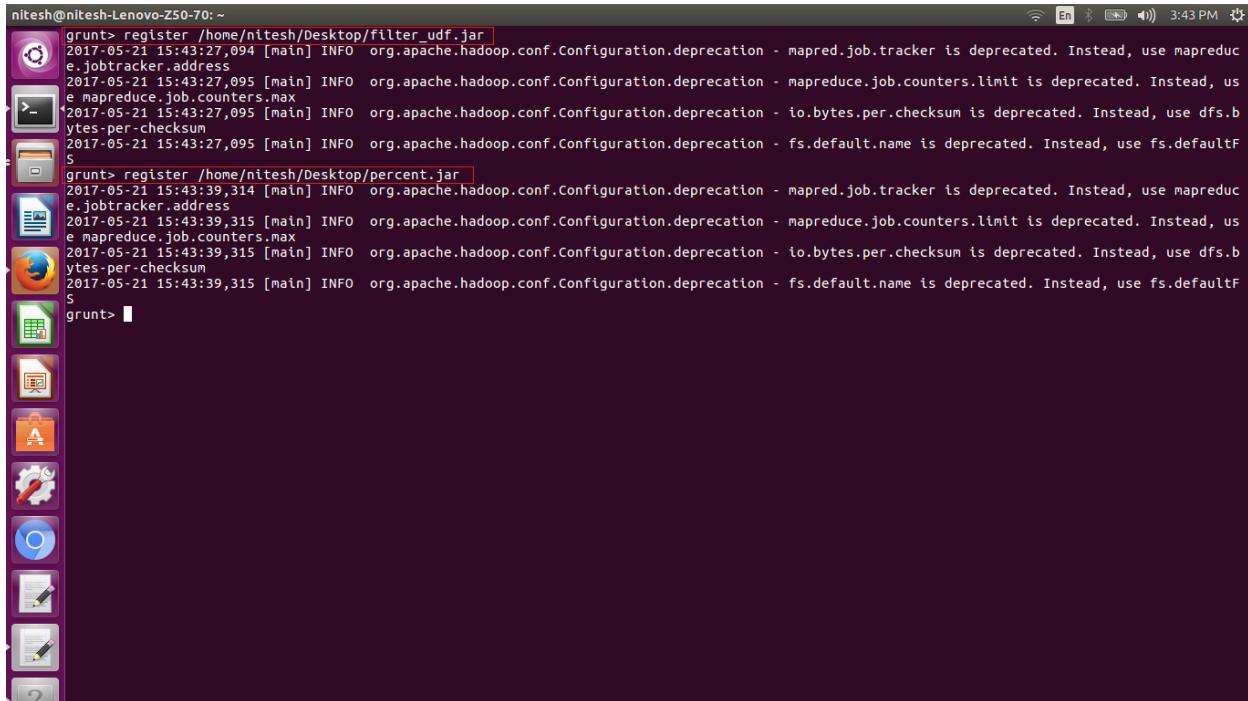
Find out the districts who achieved 100 percent objective in BPL cards Export the results to mysql using sqoop

Answer:

The first problem statement is to fetch the districts who achieved 100% in BPL cards and export the result in the mysql using sqoop.

Here we have implemented the logic to fetch the data for exact 100% achieved and we have stored state, district and percentage.

- I. The udfs are used to create the jar and the created jar are registered using the command register
- II. We have registered both the jars to fetch the percentage and to filter by calculating percentage_percent.jar and filter_udf.jar respectively.



nitesh@nitesh-Lenovo-Z50-70: ~

```
grunt> register /home/nitesh/Desktop/filter_udf.jar
2017-05-21 15:43:27,094 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
S
grunt> register /home/nitesh/Desktop/percent.jar
2017-05-21 15:43:39,314 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
S
grunt>
```

1. Now as we have registered jars we need to load the data set in the variable Input_data.

2. Loading of data is done using the following command

```
load 'hdfs:input path' Using PigStorage('delimiter') as (variable:datatype);
```

4. After the dataset is loaded in Input_data variable we have applied filter function so we can filter the loaded data to fetch the data which is exactly 100% using

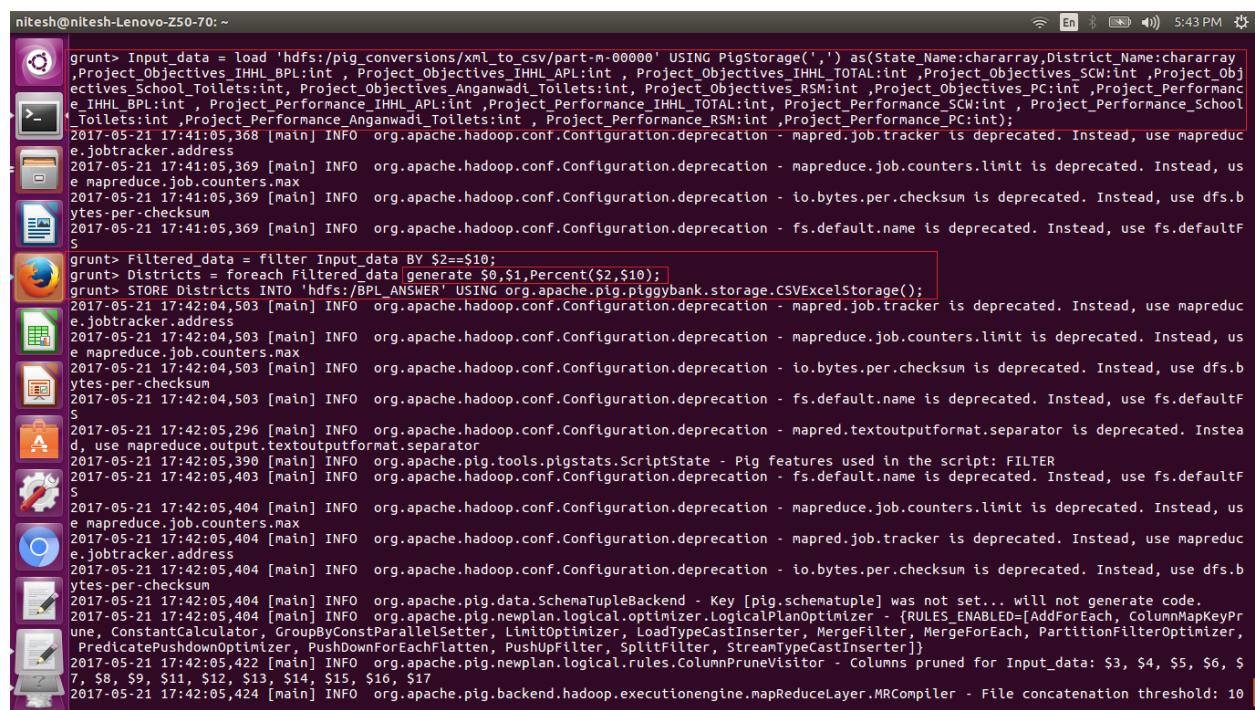
filter 'dataset' BY condition;

5. Once the filter operation is done we need to generate the columns to be stored using following syntax

```
foreach 'filtered data' generate state,district,percent;
```

6. The required columns are generated and now we need to store the data into the hdfs so we use

```
store 'generated data' into 'path' using org.apache.pig.piggybank.storage.CSVExcelStorage();
```



The screenshot shows a terminal window on a Linux desktop environment. The terminal title is "nitesh@nitesh-Lenovo-Z50-70: ~". The window contains the following Pig Latin script:

```
grunt> Input_data = load 'hdfs:/pig_conversions/xml_to_csv/part-m-00000' USING PigStorage(',') as(State_Name:chararray,District_Name:chararray,Project_Objectives_IHML_BPL:Int , Project_Objectives_IHML_AP:Int , Project_Objectives_IHML_TOTAL:Int ,Project_Objectives_SCW:Int ,Project_Objectives_School_Toilets:int , Project_Objectives_Anganwadi_Toilets:int ,Project_Objectives_RSM:Int ,Project_Objectives_PC:Int ,Project_Performance_IHML_BPL:Int , Project_Performance_IHML_AP:Int ,Project_Performance_IHML_TOTAL:Int ,Project_Performance_SCW:Int ,Project_Performance_School_Toilets:int ,Project_Performance_Anganwadi_Toilets:int , Project_Performance_RSM:Int ,Project_Performance_PC:Int);
2017-05-21 17:41:05,368 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 17:41:05,369 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 17:41:05,369 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 17:41:05,369 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> Filtered_data = filter Input_data BY $2==$10;
grunts> Districts = foreach Filtered_data generate $0,$1,Percent($2,$10);
grunts> STORE Districts INTO 'hdfs:/BPL_ANSWER' USING org.apache.pig.piggybank.storage.CSVExcelStorage();
2017-05-21 17:42:04,503 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 17:42:04,503 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 17:42:04,503 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 17:42:04,503 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-21 17:42:05,296 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2017-05-21 17:42:05,390 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2017-05-21 17:42:05,403 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-21 17:42:05,404 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 17:42:05,404 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 17:42:05,404 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 17:42:05,404 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-05-21 17:42:05,404 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypecastInserter]}
2017-05-21 17:42:05,422 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for Input_data: $3, $4, $5, $6, $7, $8, $9, $11, $12, $13, $14, $15, $16, $17
2017-05-21 17:42:05,424 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 10
```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 17:42:09,642 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2017-05-21 17:42:09,642 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2017-05-21 17:42:09,642 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2017-05-21 17:42:09,642 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserializ
e []
2017-05-21 17:42:09,649 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiti
ng for submission.
2017-05-21 17:42:09,651 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:42:10,126 [JobControl] WARN org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set. User classes may not be found. See
Job or JobSetJar(String).
2017-05-21 17:42:10,526 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-21 17:42:10,526 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-05-21 17:42:10,548 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to proc
ess : 1
2017-05-21 17:42:12,426 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-21 17:42:13,052 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495350191822_0049
2017-05-21 17:42:13,089 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of reso
urces.
2017-05-21 17:42:13,906 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495350191
822_0049 to ResourceManager at /0.0.0.0:8032
2017-05-21 17:42:14,259 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/
proxy/application_1495350191822_0049/
2017-05-21 17:42:14,260 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_149
5350191822_0049
2017-05-21 17:42:14,260 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Distri
cts_Filtered_data,Input_data
2017-05-21 17:42:14,260 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: In
put_data[10,13],Input_data[-1,-1],Filtered_data[11,16],Districts[12,12] C: R:
2017-05-21 17:42:14,543 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2017-05-21 17:42:14,543 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149
5350191822_0049]
2017-05-21 17:42:54,423 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2017-05-21 17:42:54,423 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149
5350191822_0049]
2017-05-21 17:42:59,931 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:43:00,021 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCCEEDED. Redirecting to job history server
2017-05-21 17:43:04,950 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:43:05,065 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCCEEDED. Redirecting to job history server
2017-05-21 17:43:05,345 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:43:05,570 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 17:43:05,692 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-21 17:42:05 2017-05-21 17:43:05 FILTER
Success!
Job Stats (time in seconds):
Jobid Maps Reduces MaxMapTime MlnMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
educetime Alias Feature Outputs
job_1495350191822_0049 1 0 22 22 22 0 0 0 0 Districts,Filtered_data,Input_data M
AP_ONLY hdfs:/BPL_ANSWER,
Input(s):
Successfully read 0 records from: "hdfs:/pig_conversions/xml_to_csv/part-m-00000"
Output(s):
Successfully stored 0 records in: "hdfs:/BPL_ANSWER"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495350191822_0049
2017-05-21 17:43:05,692 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:43:05,764 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCCEEDED. Redirecting to job history server
2017-05-21 17:43:05,875 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:43:05,925 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCCEEDED. Redirecting to job history server
2017-05-21 17:43:05,971 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:43:06,064 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCCEEDED. Redirecting to job history server
2017-05-21 17:43:06,135 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to
compute warning aggregation.
2017-05-21 17:43:06,135 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> ■

```

6. When the store command is executed the dataset are written in the specified directory as we can see in the above screen shot.

7. The stored data can be seen in the hdfs so we use

adoop fs -ls '/stored path' here we highlighted using the red box as shown below.

```
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /
Found 23 items
drwxr-xr-x  - nitesh supergroup      0 2017-04-23 00:03 /15_grd
drwxr-xr-x  - nitesh supergroup      0 2017-04-23 01:52 /15_mvN
drwxr-xr-x  - nitesh supergroup      0 2017-04-22 23:20 /15_sbt
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 16:25 /20_1
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 17:32 /20_2
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 18:30 /7_1
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 18:51 /9_1
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 21:56 /Achieved_80_district_percent
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 21:48 /Achieved_80_percent
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 17:42 /BPL_ANSWER
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 18:00 /BPL_ANSWER_District
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 18:47 /BPL_ANSWER_greater_District
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 17:52 /BPL_ANSWER_greater_equal
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 23:50 /pig_conversions
drwxr-xr-x  - nitesh supergroup 391461 2017-05-20 17:29 /piggybank-0.15.0.jar
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 16:56 /prjct
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 17:04 /project
drwxr-xr-x  - nitesh supergroup 717414 2017-05-20 16:34 /project1
drwxrwxrwx  - nitesh supergroup      0 2017-05-21 17:12 /projectinput
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 17:18 /tmp
drwxr-xr-x  - nitesh supergroup      0 2017-05-13 18:34 /user
drwxr-xr-x  - nitesh supergroup      0 2017-04-05 00:04 /usr
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 23:59 /xml_to_csv
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /BPL_ANSWER
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 18:00 /BPL_ANSWER/_SUCCESS
-rw-r--r--  4 nitesh supergroup 690 2017-05-21 18:00 /BPL_ANSWER/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls //BPL_ANSWER
ls: No FileSystem for scheme: null
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_percent
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 17:42 /BPL_ANSWER/_SUCCESS
-rw-r--r--  4 nitesh supergroup 1886 2017-05-21 17:42 /BPL_ANSWER/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_district_percent
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 21:48 /Achieved_80_percent/_SUCCESS
-rw-r--r--  4 nitesh supergroup 10382 2017-05-21 21:48 /Achieved_80_percent/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_district_percent
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 21:56 /Achieved_80_district_percent/_SUCCESS
-rw-r--r--  4 nitesh supergroup 3356 2017-05-21 21:56 /Achieved_80_district_percent/part-m-00000
```

8. Now as we need to export the data into the mysql, we need to create a database, to create table, in which the stored dataset from hdfs can be loaded to the mysql.

9. The below screen shot shows the creation of the database and the table

for that purpose we have used the following commands

create database 'database name'

a. use 'database name'

b. create table 'table name' (columns:data type)

we have also described the table as shown in the below picture

using the command describe table

```

nitesh@nitesh-Lenovo-Z50-70: ~
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| hivedb |
| metastore |
| mysql |
| performance_schema |
| sys |
+-----+
6 rows in set (0.54 sec)

mysql> CREATE DATABASE BPL_ANS;
Query OK, 1 row affected (0.05 sec)

mysql> USE BPL_ANS;
Database changed
mysql> CREATE TABLE BPL_ANS_ALL (STATE VARCHAR(100), DISTRICT VARCHAR(100), PERCENT FLOAT(10));
Query OK, 0 rows affected (1.80 sec)

mysql> DESCRIBE BPL_ANS;
ERROR 1146 (42502): Table 'BPL_ANS.BPL_ANS' doesn't exist
mysql> DESCRIBE BPL_ANS_ALL;
+-----+
| Field | Type   | Null | Key | Default | Extra |
+-----+
| STATE | varchar(100) | YES |     | NULL    |          |
| DISTRICT | varchar(100) | YES |     | NULL    |          |
| PERCENT | float    | YES |     | NULL    |          |
+-----+
3 rows in set (0.02 sec)

mysql>

```

10. Now as we have generated the table in mysql we can export the data from the hdfs to mysql using sqoop. The following is the command to do so

`sqoop export -m 1 --connect jdbc:mysql://localhost:3306/'database name' --username username --password password --table 'table name' --export-dir 'hdfs path of the file to be exported' --driver com.mysql.jdbc.Driver`

```

nitesh@nitesh-Lenovo-Z50-70: ~
nitesh@nitesh-Lenovo-Z50-70:~$ sqoop export -m 1 --connect jdbc:mysql://localhost:3306/BPL_ANS --username root --password nss8055 --table BPL_ANS_ALL
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../hcatalog does not exist! HCatalog jobs will fail.
Please set SHCAT_HOME to the root of your HCatalog installation.
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-05-21 23:25:59,598 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-05-21 23:25:59,611 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-05-21 23:25:59,671 WARN [main] sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate connection manager is not being set (via --connection-manager). Sqoop is going to fall back to org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.
2017-05-21 23:25:59,682 INFO [main] manager.SqlManager: Using default fetchSize of 1000
2017-05-21 23:25:59,683 INFO [main] tool.CodeGenTool: Beginning code generation
2017-05-21 23:26:00,068 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_ALL AS t WHERE 1=0
2017-05-21 23:26:00,152 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_ALL AS t WHERE 1=0
2017-05-21 23:26:00,239 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop/hadoop-2.7.0
Note: /tmp/sqoop-nitesh/compile/1d8ceef3/e8c89a55c8cbfec005ae6cb/BPL_ANS_ALL.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-05-21 23:26:04,874 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-nitesh/compile/1d8ceef3/e8c89a55c8cbfec005ae6cb/BPL_ANS_ALL.jar
2017-05-21 23:26:04,977 INFO [main] mapreduce.ExportJobBase: Beginning export of BPL_ANS_ALL
2017-05-21 23:26:04,977 INFO [main] Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-05-21 23:26:05,536 INFO [main] Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2017-05-21 23:26:07,991 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_ALL AS t WHERE 1=0
2017-05-21 23:26:07,994 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-21 23:26:07,999 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-21 23:26:07,999 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-21 23:26:08,216 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 23:26:36,328 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:26:36,329 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:26:37,317 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-21 23:26:37,434 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-21 23:26:37,434 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.f

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 23:26:07,994 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-21 23:26:07,999 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-21 23:26:07,999 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-21 23:26:08,210 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 23:26:36,328 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:26:36,329 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:26:37,317 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-21 23:26:37,434 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-21 23:26:37,434 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-21 23:26:37,434 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-21 23:26:37,434 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.in
putdir
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.ma
p.speculative
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputform
at.class
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputfo
rmat.class
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.v
alue.class
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key
.class
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.
files
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-05-21 23:26:37,435 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.fi
les.filesizes
2017-05-21 23:26:38,460 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495350191822_0057
2017-05-21 23:26:39,218 INFO [main] impl.YarnClientImpl: Submitted application application_1495350191822_0057 to ResourceManager at /0.0.0.0:8
032
2017-05-21 23:26:39,442 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495350
191822_0057/
2017-05-21 23:26:39,443 INFO [main] mapreduce.Job: Running job: job_1495350191822_0057
2017-05-21 23:27:07,041 INFO [main] mapreduce.Job: Job job_1495350191822_0057 running in uber mode : false
2017-05-21 23:27:07,133 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-05-21 23:27:26,429 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-05-21 23:27:34,113 INFO [main] mapreduce.Job: Job job_1495350191822_0057 completed successfully

```

we can see that the sqoop command executed successfully as shown above in the picture and it is highlighted

11. Now we need to check if the required data is written in mysql, for that we logged into the mysql using command

`mysql -u root -p`

12. Now we need to use the database we created to create the table to store the data.

So we use the same database using

`use 'database'`

`show 'tables'` will show the tables created in the database

to check if the data is loaded in the table we use the select query to verify

`select * from 'table name'`

it will display all the entries in the table and we can see the required data is written from the hdfs to mysql using sqoop.

```
nitesh@nitesh-Lenovo-Z50-70: ~ nitesh@nitesh-Lenovo-Z50-70:~$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 24
Server version: 5.7.11-0ubuntu6 (Ubuntu)

Copyright (c) 2000, 2016, 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> use BPL_ANS;
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_BPL_ANS |
+-----+
| BPL_ANS_ALL |
| BPL_ANS_DIST |
+-----+
2 rows in set (0.00 sec)

mysql> SELECT * FROM BPL_ANS_ALL;
+-----+-----+-----+
| STATE | DISTRICT | PERCENT |
+-----+-----+-----+
| Andhra Pradesh | NIZAMABAD | 100 |
| Arunachal Pradesh | TIRAP | 100 |
| Assam | HAILAKANDI | 100 |
| Bihar | MADHUBANI | 100 |
| Goa | NORTH GOA | 100 |
| Gujarat | AHMEDABAD | 100 |
| Gujarat | DANGS | 100 |
| Gujarat | NAVSARI | 100 |
| Gujarat | PORBANDAR | 100 |
| Gujarat | SURAT | 100 |
| Haryana | FARIDABAD | 100 |

```

```
nitesh@nitesh-Lenovo-Z50-70: ~
+-----+-----+-----+
| Gujarat | SURAT | 100 |
| Haryana | FARIDABAD | 100 |
| Haryana | HISAR | 100 |
| Haryana | JHAJJAR | 100 |
| Haryana | MAHENDRAGARH | 100 |
| Haryana | PANCHKULA | 100 |
| Haryana | PANIPAT | 100 |
| Haryana | ROHTAK | 100 |
| Haryana | SIRSIA | 100 |
| Himachal Pradesh | HAMIRPUR | 100 |
| Himachal Pradesh | KINNAUR | 100 |
| Himachal Pradesh | KULLU | 100 |
| Himachal Pradesh | LAHUAL & SPITI | 100 |
| Himachal Pradesh | SHIMLA | 100 |
| Himachal Pradesh | SOLAN | 100 |
| Himachal Pradesh | UNA | 100 |
| Jharkhand | DEOGARH | 100 |
| Jharkhand | LOHARDAGA | 100 |
| Karnataka | HASSAN | 100 |
| Karnataka | MANGALORE(DAKSHINA KANNADA) | 100 |
| Karnataka | UDUPI | 100 |
| Kerala | ALAPPUZHA | 100 |
| Kerala | KOLLAM | 100 |
| Kerala | KOTTAYAM | 100 |
| Kerala | KOZHIKODE | 100 |
| Kerala | PALAKKAD | 100 |
| Kerala | PATHANAMTHITTA | 100 |
| Kerala | WAYANAD | 100 |
| Maharashtra | GADCHIROLI | 100 |
| Maharashtra | SINDHUDURG | 100 |
| Meghalaya | WEST GARO HILLS | 100 |
| Mizoram | CHAMPHAI | 100 |
| Mizoram | LAWNGTLAI | 100 |
| Rajasthan | HANUMANGARH | 100 |
| Tamil Nadu | ERODE | 100 |
| Tamil Nadu | KARUR | 100 |
| Tamil Nadu | NAMAKKAL | 100 |
| Tamil Nadu | TIRUCHIRAPPALLI | 100 |
| Tamil Nadu | TIRUVANNAMALAI | 100 |
| Tripura | DHALAI | 100 |
| Tripura | SOUTH TRIPURA | 100 |
| Tripura | WEST TRIPURA | 100 |
| Uttar Pradesh | AMBEDKAR NAGAR | 100 |
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
+-----+-----+-----+
| Kerala | KOLLAM | 100 |
| Kerala | KOTTAYAM | 100 |
| Kerala | KOZHIKODE | 100 |
| Kerala | PALAKKAD | 100 |
| Kerala | PATHANAMTHITTA | 100 |
| Kerala | WAYANAD | 100 |
| Maharashtra | GADCHIROLI | 100 |
| Maharashtra | SINDHUDURG | 100 |
| Meghalaya | WEST GARO HILLS | 100 |
| Mizoram | CHAMPHAI | 100 |
| Mizoram | LAWNGTALAI | 100 |
| Rajasthan | HANUMANGARH | 100 |
| Tamil Nadu | ERODE | 100 |
| Tamil Nadu | KARUR | 100 |
| Tamil Nadu | NAMAKKAL | 100 |
| Tamil Nadu | TIRUCHIRAPPALLI | 100 |
| Tamil Nadu | TIRUVANNAMALAI | 100 |
| Tripura | DHALAI | 100 |
| Tripura | SOUTH TRIPURA | 100 |
| Tripura | WEST TRIPURA | 100 |
| Uttar Pradesh | AMBEDKAR NAGAR | 100 |
| Uttar Pradesh | BALRAMPUR | 100 |
| Uttar Pradesh | BAREILLY | 100 |
| Uttar Pradesh | BIJNOR | 100 |
| Uttar Pradesh | BUDAUN | 100 |
| Uttar Pradesh | ETAWAH | 100 |
| Uttar Pradesh | FARRUKHABAD | 100 |
| Uttar Pradesh | FIROZABAD | 100 |
| Uttar Pradesh | GHAZIABAD | 100 |
| Uttar Pradesh | HARDOT | 100 |
| Uttar Pradesh | JYOTIBA PHULE NAGAR | 100 |
| Uttar Pradesh | LUCKNOW | 100 |
| Uttar Pradesh | MAHARAJGANJ | 100 |
| Uttar Pradesh | MAHOBIA | 100 |
| Uttar Pradesh | MORADABAD | 100 |
| Uttar Pradesh | MUZAFFARNAGAR | 100 |
| Uttar Pradesh | PILIBHIT | 100 |
| Uttar Pradesh | SONBHADRA | 100 |
| Uttar Pradesh | SULTANPUR | 100 |
+-----+-----+-----+
70 rows in set (0.01 sec)

mysql> 
```

the above are the few screen shots of the answer

there are 70 entries of the exact 100%.

Code is for exact 100% of performance reference to objective and displaying district only.

Here we have done the same processes to get the data, but we fetched only districts to store in hdfs and then into the mysql using sqoop.

```
nitesh@nitesh-Lenovo-Z50-70:~$ grunt> register /home/nitesh/Desktop/filter_udf.jar
2017-05-21 15:43:27,094 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> register /home/nitesh/Desktop/percent.jar
2017-05-21 15:43:39,314 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt>
```

```
nitesh@nitesh-Lenovo-Z50-70:~$ grunt>
grunt> Input_data = load 'hdfs://pig conversions/xml_to_csv/part-m-00000' USING PigStorage(' ') as(State_Name:chararray,District_Name:chararray
,Project_Objectives_IHML_BPL:int , Project_Objectives_IHML_AP:int , Project_Objectives_IHML_TOTAL:int ,Project_Objectives_SCW:int ,Project_Ob
jectives_School_Toilets:int ,Project_Objectives_Anganwadi_Toilets:int ,Project_Objectives_RSM:int ,Project_Objectives_PC:int ,Project_Performanc
e_IHML_BPL:int , Project_Performance_IHML_AP:Int ,Project_Performance_IHML_TOTAL:Int ,Project_Performance_SCW:int , Project_Performance_Schoo
l>Toilets:int ,Project_Performance_Anganwadi_Toilets:int , Project_Performance_RSM:int ,Project_Performance_PC:int);
2017-05-21 17:57:27,753 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-21 17:57:27,753 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us
e mapreduce.job.counters.max
2017-05-21 17:57:27,753 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc
e.jobtracker.address
2017-05-21 17:57:27,753 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b
ytes-per-checksum
2017-05-21 17:57:30,009 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc
e.jobtracker.address
2017-05-21 17:57:30,009 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us
e mapreduce.job.counters.max
2017-05-21 17:57:30,010 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b
ytes-per-checksum
2017-05-21 17:57:30,010 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> Filtered_data = filter Input_data BY $2==$10;
grunts> Districts = foreach Filtered_data generate $1;
grunts> STORE Districts INTO 'hdfs://BPL_ANSWER_District' USING org.apache.pig.piggypbank.storage.CSVExcelStorage();
2017-05-21 17:58:55,455 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc
e.jobtracker.address
2017-05-21 17:58:55,456 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us
e mapreduce.job.counters.max
2017-05-21 17:58:55,456 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b
ytes-per-checksum
2017-05-21 17:58:55,456 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-21 17:58:56,892 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2017-05-21 17:58:56,934 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-21 17:58:56,934 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us
e mapreduce.job.counters.max
2017-05-21 17:58:56,934 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc
e.jobtracker.address
2017-05-21 17:58:56,934 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b
ytes-per-checksum
2017-05-21 17:58:56,935 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
l/pig/lib/joda-time-2.5.jar to DistributedCache through /tmp/temp797228565/tmp-953216838/joda-time-2.5.jar
2017-05-21 17:59:03,946 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store
job
2017-05-21 17:59:03,946 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2017-05-21 17:59:03,946 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
e []
2017-05-21 17:59:03,953 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiti
ng for submission.
2017-05-21 17:59:03,954 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:59:04,422 [JobControl] WARN org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set. User classes may not be found. See
Job or Job#setJar(String).
2017-05-21 17:59:04,997 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-21 17:59:04,997 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
2017-05-21 17:59:05,027 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths (combined) to proc
ess : 1
2017-05-21 17:59:06,872 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-21 17:59:07,882 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495350191822_0053
2017-05-21 17:59:08,156 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of reso
urces.
2017-05-21 17:59:09,521 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495350191
822_0053 to ResourceManager at /0.0.0.0:8032
2017-05-21 17:59:10,099 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/
proxy/application_1495350191822_0053/
2017-05-21 17:59:10,099 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_14953501
91822_0053
2017-05-21 17:59:10,099 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Distri
cts,Filtered_data,Input_data
2017-05-21 17:59:10,277 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: In
put_data[23,13],Input_data[-1,-1],Filtered_data[24,16],Districts[25,12] C: R:
2017-05-21 17:59:10,277 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2017-05-21 17:59:10,277 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149
5350191822_0053]
2017-05-21 18:00:30,809 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2017-05-21 18:00:30,809 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149
5350191822_0053]
2017-05-21 18:00:46,825 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 18:00:46,917 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 18:01:58,580 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 18:01:58,762 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 18:01:58,831 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 18:01:59,156 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-21 17:58:56 2017-05-21 18:01:59 FILTER
Success!
Job Stats (time in seconds):
Jobid Maps Reduces MaxMapTime MlnMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
educetime Alias Feature Outputs
job_1495350191822_0053 1 0 49 49 49 49 0 0 0 0 Districts,Filtered_data,Input_data M
AP_ONLY hdfs:/BPL_ANSWER_District,
Input(s):
Successfully read 0 records from: "hdfs:/pig_conversions/xml_to_csv/part-m-00000"
Output(s):
Successfully stored 0 records in: "hdfs:/BPL_ANSWER_District"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495350191822_0053

2017-05-21 18:01:59,157 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 18:01:59,192 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 18:01:59,433 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 18:01:59,611 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 18:01:59,667 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 18:01:59,712 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 18:01:59,734 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to
compute warning aggregation.
2017-05-21 18:01:59,734 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> ■
```

```
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /
Found 23 items
drwxr-xr-x  - nitesh supergroup      0 2017-04-23 00:03 /15_grd
drwxr-xr-x  - nitesh supergroup      0 2017-04-23 01:52 /15_mvN
drwxr-xr-x  - nitesh supergroup      0 2017-04-22 23:20 /15_sbt
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 16:25 /20_1
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 17:32 /20_2
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 10:30 /7_1
drwxr-xr-x  - nitesh supergroup      0 2017-04-30 10:51 /9_1
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 21:56 /Achieved_80_district_percent
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 21:48 /Achieved_80_percent
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 17:42 /BPL_ANSWER
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 18:00 /BPL_ANSWER_District
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 18:47 /BPL_ANSWER_greater_District
drwxr-xr-x  - nitesh supergroup      0 2017-05-21 17:52 /BPL_ANSWER_greater_equal
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 23:58 /pig_conversions
-rw-r--r--  4 nitesh supergroup 391461 2017-05-20 17:29 /piggybank-0.15.0.jar
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 16:56 /prjct
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 17:04 /project
drwxr-xr-x  - nitesh supergroup 717414 2017-05-20 16:34 /project1
drwxrwxrwx  - nitesh supergroup      0 2017-05-21 17:18 /tmp
drwxr-xr-x  - nitesh supergroup      0 2017-05-13 18:34 /user
drwxr-xr-x  - nitesh supergroup      0 2017-04-05 00:04 /usr
drwxr-xr-x  - nitesh supergroup      0 2017-05-20 23:59 /xml_to_csv
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /BPL_ANSWER_District
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 18:00 /BPL_ANSWER_District/_SUCCESS
-rw-r--r--  4 nitesh supergroup 690 2017-05-21 18:00 //BPL_ANSWER_District/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls //BPL_ANSWER
ls: No FileSystem for scheme: null
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /BPL_ANSWER
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 17:42 /BPL_ANSWER/_SUCCESS
-rw-r--r--  4 nitesh supergroup 1886 2017-05-21 17:42 /BPL_ANSWER/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_percent
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 21:48 /Achieved_80_percent/_SUCCESS
-rw-r--r--  4 nitesh supergroup 10382 2017-05-21 21:48 /Achieved_80_percent/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_district_percent
Found 2 items
-rw-r--r--  4 nitesh supergroup      0 2017-05-21 21:56 /Achieved_80_district_percent/_SUCCESS
-rw-r--r--  4 nitesh supergroup 3356 2017-05-21 21:56 /Achieved_80_district_percent/part-m-00000
```

```
nitesh@nitesh-Lenovo-Z50-70:~$ mysql> CREATE TABLE BPL_ANS_DIST (DISTRICT VARCHAR(100));
Query OK, 0 rows affected (0.83 sec)

mysql> DESCRIBE BPL_ANS_DIST;
+-----+-----+-----+-----+-----+
| Field | Type   | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| DISTRICT | varchar(100) | YES |   | NULL |   |
+-----+-----+-----+-----+-----+
1 row in set (0.03 sec)

mysql>
```

```
nitesh@nitesh-Lenovo-Z50-70:~ nitesh@nitesh-Lenovo-Z50-70:~$ sqoop export -m 1 --connect jdbc:mysql://localhost:3306/BPL_ANS --username root --password nss8055 --table BPL_A  
NS_DIST --export-dir /BPL_ANSWER_District/part-m-00000 --driver com.mysql.jdbc.Driver  
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../catalog does not exist! HCatalog jobs will fail.  
Please set $HCAT_HOME to the root of your HCatalog installation.  
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../accumulo does not exist! Accumulo imports will fail.  
Please set $ACCUMULO_HOME to the root of your Accumulo installation.  
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../zookeeper does not exist! Accumulo imports will fail.  
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.  
2017-05-21 23:39:14,808 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6  
2017-05-21 23:39:14,819 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.  
2017-05-21 23:39:14,872 WARN [main] sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate Connection manager  
is not being set (via --connection-manager). Sqoop is going to fall back to org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.  
2017-05-21 23:39:14,881 INFO [main] manager.SqlManager: Using default fetchSize of 1000  
2017-05-21 23:39:14,881 INFO [main] tool.CodeGenTool: Beginning code generation  
2017-05-21 23:39:15,243 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_DIST AS t WHERE 1=0  
2017-05-21 23:39:15,261 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_DIST AS t WHERE 1=0  
2017-05-21 23:39:15,286 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop/hadoop-2.7.0  
Note: /tmp/sqoop-nitesh/compile/65abdc166ba22d999e6dc7a4c824dec/BPL_ANS_DIST.java uses or overrides a deprecated API.  
Note: Recompile with -Xlint:deprecation for details.  
2017-05-21 23:39:16,431 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-nitesh/compile/65abdc166ba22d999e6dc7a4c824dec/BPL_A  
NS_DIST.jar  
2017-05-21 23:39:16,440 INFO [main] mapreduce.ExportJobBase: Beginning export of BPL_ANS_DIST  
2017-05-21 23:39:16,440 INFO [main] Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/lib/slf4j-jlog4j12-1.6.4.jar!/org/slf4j.impl/StaticLoggerBinder.cl  
ass]  
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-jlog4j12-1.7.10.jar!/org/slf4j.impl/StaticLoggerB  
inder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
2017-05-21 23:39:16,634 INFO [main] Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar  
2017-05-21 23:39:17,468 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_DIST AS t WHERE 1=0  
2017-05-21 23:39:17,471 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce  
.reduce.speculative  
2017-05-21 23:39:17,474 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.ma  
p.speculative  
2017-05-21 23:39:17,474 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps  
2017-05-21 23:39:17,580 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
2017-05-21 23:39:33,369 INFO [main] input.FileInputFormat: Total input paths to process : 1  
2017-05-21 23:39:33,370 INFO [main] input.FileInputFormat: Total input paths to process : 1  
2017-05-21 23:39:34,764 INFO [main] mapreduce.JobSubmitter: number of splits:1  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.f  
iles.timestamps  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.in  
putdir  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use Mapreduce.ma  
p.speculative  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputform  
at.class  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputfo  
rmat.class  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.v  
alue.class  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key  
.class  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.  
files  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.fi  
les.filesizes  
2017-05-21 23:39:35,152 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495350191822_0059  
2017-05-21 23:39:37,714 INFO [main] impl.YarnClientImpl: Submitted application application_1495350191822_0059 to ResourceManager at /0.0.0.0:  
032  
2017-05-21 23:39:37,889 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495350  
191822_0059/  
2017-05-21 23:39:37,889 INFO [main] mapreduce.Job: Running job: job_1495350191822_0059  
2017-05-21 23:39:57,122 INFO [main] mapreduce.Job: Job job_1495350191822_0059 running in uber mode : false  
2017-05-21 23:39:57,158 INFO [main] mapreduce.Job: map 0% reduce 0%  
2017-05-21 23:40:17,378 INFO [main] mapreduce.Job: map 100% reduce 0%  
2017-05-21 23:40:23,115 INFO [main] mapreduce.Job: Job job_1495350191822_0059 completed successfully
```

```
nitesh@nitesh-Lenovo-Z50-70:~ nitesh@nitesh-Lenovo-Z50-70:~$ Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce  
.reduce.speculative  
2017-05-21 23:39:17,474 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.ma  
p.speculative  
2017-05-21 23:39:17,474 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps  
2017-05-21 23:39:17,580 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
2017-05-21 23:39:33,369 INFO [main] input.FileInputFormat: Total input paths to process : 1  
2017-05-21 23:39:33,370 INFO [main] input.FileInputFormat: Total input paths to process : 1  
2017-05-21 23:39:34,764 INFO [main] mapreduce.JobSubmitter: number of splits:1  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.f  
iles.timestamps  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.in  
putdir  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use Mapreduce.ma  
p.speculative  
2017-05-21 23:39:34,922 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputform  
at.class  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputfo  
rmat.class  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.v  
alue.class  
2017-05-21 23:39:34,923 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key  
.class  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.  
files  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces  
2017-05-21 23:39:34,924 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.fi  
les.filesizes  
2017-05-21 23:39:35,152 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495350191822_0059  
2017-05-21 23:39:37,714 INFO [main] impl.YarnClientImpl: Submitted application application_1495350191822_0059 to ResourceManager at /0.0.0.0:  
032  
2017-05-21 23:39:37,889 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495350  
191822_0059/  
2017-05-21 23:39:37,889 INFO [main] mapreduce.Job: Running job: job_1495350191822_0059  
2017-05-21 23:39:57,122 INFO [main] mapreduce.Job: Job job_1495350191822_0059 running in uber mode : false  
2017-05-21 23:39:57,158 INFO [main] mapreduce.Job: map 0% reduce 0%  
2017-05-21 23:40:17,378 INFO [main] mapreduce.Job: map 100% reduce 0%  
2017-05-21 23:40:23,115 INFO [main] mapreduce.Job: Job job_1495350191822_0059 completed successfully
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
mysql> use BPL_ANS;
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_BPL_ANS |
+-----+
| BPL_ANS_ALL |
| BPL_ANS_DIST |
+-----+
2 rows in set (0.00 sec)

mysql> SELECT * FROM BPL_ANS_DIST;
+-----+
| DISTRICT |
+-----+
| NIZAMABAD |
| TIRAP |
| HAILAKANDI |
| MADHUBANI |
| NORTH GOA |
| AHMEDABAD |
| DANGS |
| NAVSARI |
| PORBANDAR |
| SURAT |
| FARIDABAD |
| HTSAR |
| JHAJJAR |
| MAHENDRAGARH |
| PANCHKULA |
| PANIPAT |
| ROHTAK |
| SIRSA |
| HAMIRPUR |
| KINNAUR |
| KULLU |
| LAHAUL & SPITI |
| SHIMLA |
| SOLAN |
| UNA |
+-----+
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
+-----+
| UNA |
| DEOGHAR |
| LOHARDAGA |
| HASSAN |
| MANGALORE(DAKSHINA KANNADA) |
| UDUPI |
| ALAPPUZHA |
| KOLLAM |
| KOTTAYAM |
| KOZHIKODE |
| PALAKKAD |
| PATHANAMTHITTA |
| WAYANAD |
| GADCHIROLI |
| SINDHUDURG |
| WEST GARO HILLS |
| CHAMPHAI |
| LAWNGTIAI |
| HANUMANGARH |
| ERODE |
| KARUR |
| NAMAKKAL |
| TIRUCHIRAPPALLI |
| TIRUVANNAMALAI |
| DHALAI |
| SOUTH TRIPURA |
| WEST TRIPURA |
| AMBEDKAR NAGAR |
| BALRAMPUR |
| BAREILLY |
| BTJNOR |
| BUDAUN |
| ETAWAH |
| FARRUKHABAD |
| FIROZABAD |
| GHAZIABAD |
| HARDOI |
| JVOTIBA PHULE NAGAR |
| LUCKNOW |
| MAHARAJGANJ |
| MAHOBIA |
| MORADABAD |
| MUZAFFARNAGAR |
+-----+
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
+----+
| KOLLAM
| KOTTAYAM
| KOZHIKODE
| PALAKKAD
| PATHANAMTHITTA
| WAYANAD
| GADCHIROLI
| SINDHUDURG
| WEST GARO HILLS
| CHAMPHAI
| LAWNGTIAI
| HANUMANGARH
| ERODE
| KARUR
| NAMAKKAL
| TIRUCHIRAPPALLI
| TIRUVANNAMALAI
| DHALAI
| SOUTH TRIPURA
| WEST TRIPURA
| AMBEDKAR NAGAR
| BALRAMPUR
| BAREILLY
| BIJNOR
| BUDAUN
| ETAWAH
| FARRUKHABAD
| FIROZABAD
| GHAZIABAD
| HARDOI
| JYOTIBA PHULE NAGAR
| LUCKNOW
| MAHARAJGANJ
| MAHOBIA
| MORADABAD
| MUZAFFARNAGAR
| PILIBHIT
| SONBHADRA
| SULTANPUR
+----+
70 rows in set (0.00 sec)

mysql> |
```

Here we have a code for greater than or equal to 100% of performance in reference to objective and we have displayed all state, district and percentage.

Here we have done the same processes to get the data, and the data is stored in hdfs and then into the mysql using sqoop.

```
nitesh@nitesh-Lenovo-Z50-70: ~
grunt> register /home/nitesh/Desktop/filter_udf.jar
2017-05-21 15:43:27,094 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> register /home/nitesh/Desktop/percent.jar
2017-05-21 15:43:39,314 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt>
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
grunt> Input_data = load 'hdfs:/pig_conversions/xml_to_csv/part-m-00000' USING PigStorage(',') as(State_Name:chararray,District_Name:chararray ,Project_Objectives_IHML_BPL:int ,Project_Objectives_IHML_AP:int ,Project_Objectives_IHML_TOTAL:int ,Project_Objectives_SCW:int ,Project_Objectives_School_Toilets:int ,Project_Objectives_Anganwadi_Toilets:int ,Project_Objectives_RSM:int ,Project_Objectives_PC:int ,Project_Performance_IHML_BPL:int ,Project_Performance_IHML_AP:int ,Project_Performance_IHML_TOTAL:int ,Project_Performance_SCW:int ,Project_Performance_School_Toilets:int ,Project_Performance_Anganwadi_Toilets:int ,Project_Performance_RSM:int ,Project_Performance_PC:int);
2017-05-21 17:50:01,370 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 17:50:01,371 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 17:50:01,371 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 17:50:01,371 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> Filtered_data = filter Input_data BY $2<=$10;
grunt> Districts = foreach Filtered_data generate $0,$1,Percent($2,$10);
grunt> STORE Districts INTO 'hdfs:/BPL_ANSWER_greater_equal' USING org.apache.pig.piggypack.storage.CSVExcelStorage();
2017-05-21 17:51:37,582 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 17:51:37,583 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 17:51:37,583 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 17:51:37,583 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-21 17:51:38,209 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2017-05-21 17:51:38,224 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-21 17:51:38,224 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 17:51:38,224 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 17:51:38,224 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 17:51:38,225 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-05-21 17:51:38,225 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-05-21 17:51:38,226 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for Input_data: $3, $4, $5, $6, $7, $8, $9, $11, $12, $13, $14, $15, $16, $17
2017-05-21 17:51:38,227 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 10
0 optimistic? false
2017-05-21 17:51:38,228 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before opt
```

```

nitesh@nitesh-Lenovo-Z50-70: ~
job
2017-05-21 17:51:49,831 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2017-05-21 17:51:49,831 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2017-05-21 17:51:49,831 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserizalize []
2017-05-21 17:51:49,876 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2017-05-21 17:51:49,880 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:51:50,309 [JobControl] WARN org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2017-05-21 17:51:51,268 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-21 17:51:51,268 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapredUtil - Total input paths to process : 1
2017-05-21 17:51:51,374 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapredUtil - Total input paths (combined) to process : 1
2017-05-21 17:51:53,003 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-21 17:51:54,415 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495350191822_0052
2017-05-21 17:51:54,598 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.
2017-05-21 17:51:55,698 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495350191822_0052 to ResourceManager at /0.0.0.0:8032
2017-05-21 17:51:55,699 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495350191822_0052/
2017-05-21 17:51:55,699 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1495350191822_0052
2017-05-21 17:51:55,699 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Districts_Filtered_data_Input_data
2017-05-21 17:51:55,699 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: Input_data[19,13],Input_data[-1,-1],Filtered_data[20,16],Districts[21,12] C: R:
2017-05-21 17:51:56,712 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2017-05-21 17:51:56,712 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495350191822_0052]
2017-05-21 17:52:20,950 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2017-05-21 17:52:20,950 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495350191822_0052]
2017-05-21 17:52:26,457 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:52:26,493 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-21 17:52:28,937 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:52:28,956 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-21 17:52:29,006 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:52:29,044 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 17:52:29,075 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-21 17:51:38 2017-05-21 17:52:29 FILTER
Success!
Job Stats (time in seconds):
Jobid Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
educetime Alias Feature Outputs
job_1495350191822_0052 1 0 11 11 11 11 0 0 0 Districts_Filtered_data_Input_data M
AP_ONLY hdfs:/BPL_ANSWER_greater_equal,
Input(s):
Successfully read 0 records from: "hdfs:/pig_conversions/xml_to_csv/part-m-00000"
Output(s):
Successfully stored 0 records in: "hdfs:/BPL_ANSWER_greater_equal"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495350191822_0052
2017-05-21 17:52:29,076 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:52:29,107 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-21 17:52:29,173 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:52:29,214 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-21 17:52:29,296 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 17:52:29,433 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESSED. Redirecting to job history server
2017-05-21 17:52:29,562 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-05-21 17:52:29,562 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> 

```

```

nitesh@nitesh-Lenovo-Z50-70:~ nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /
Found 23 items
drwxr-xr-x  - nitesh supergroup          0 2017-04-23 00:03 /15_grd
drwxr-xr-x  - nitesh supergroup          0 2017-04-23 01:52 /15_mvN
drwxr-xr-x  - nitesh supergroup          0 2017-04-22 23:20 /15_sbt
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 16:25 /20_1
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 17:32 /20_2
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 10:30 /7_1
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 10:51 /9_1
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 21:56 /Achieved_80_district_percent
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 21:48 /Achieved_80_percent
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 17:42 /BPL_ANSWER
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 18:00 /BPL_ANSWER_District
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 18:47 /BPL_ANSWER_greater_District
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 17:52 /BPL_ANSWER_greater_equal
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 23:58 /piggybank
drwxr-xr-x  - nitesh supergroup          391461 2017-05-20 17:29 /piggybank-0.15.0.jar
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 16:56 /prjct
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 17:04 /project
drwxr-xr-x  - nitesh supergroup          717414 2017-05-20 16:34 /project1
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 17:12 /projectininput
drwxrwxrwx  - nitesh supergroup          0 2017-05-21 17:18 /tmp
drwxr-xr-x  - nitesh supergroup          0 2017-05-13 18:34 /user
drwxr-xr-x  - nitesh supergroup          0 2017-04-05 00:04 /usr
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 23:59 /xml_to_csv
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /BPL_ANSWER_District
Found 2 items
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 18:00 /BPL_ANSWER_District/_SUCCESS
-rw-r--r--  4 nitesh supergroup          690 2017-05-21 18:00 /BPL_ANSWER_District/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls //BPL_ANSWER
ls: No FileSystem for scheme: null
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /BPL_ANSWER
Found 2 items
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 17:42 /BPL_ANSWER/_SUCCESS
-rw-r--r--  4 nitesh supergroup          1886 2017-05-21 17:42 /BPL_ANSWER/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_percent
Found 2 items
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 21:48 /Achieved_80_percent/_SUCCESS
-rw-r--r--  4 nitesh supergroup          10382 2017-05-21 21:48 /Achieved_80_percent/part-m-00000
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_district_percent
Found 2 items
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 21:56 /Achieved_80_district_percent/_SUCCESS
-rw-r--r--  4 nitesh supergroup          3356 2017-05-21 21:56 /Achieved_80_district_percent/part-m-00000

```

```

nitesh@nitesh-Lenovo-Z50-70:~ nitesh@nitesh-Lenovo-Z50-70:~$ sqoop export -m 1 --connect jdbc:mysql://localhost:3306/BPL_ANS --username root --password nss8055 --table BPL_AN
NS_GREATER_ALL --export-dir /BPL_ANS_greater_equal/part-m-00000 --driver com.mysql.jdbc.Driver
Warning: /home/nitesh/sqoop-1.4.6.bln_hadoop-2.0.4-alpha/..//hcatalog does not exist! HCatalog jobs will fail.
Please set SHCAT_HOME to the root of your HCatalog installation.
Warning: /home/nitesh/sqoop-1.4.6.bln_hadoop-2.0.4-alpha/..//accumulo does not exist! Accumulo imports will fail.
Please set ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/nitesh/sqoop-1.4.6.bln_hadoop-2.0.4-alpha/..//zookeeper does not exist! Zookeeper imports will fail.
Please set ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-05-21 23:51:33,009 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-05-21 23:51:33,032 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-05-21 23:51:33,090 WARN [main] sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate connection manager is not being set (via --connection-manager). Sqoop is going to fall back to org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.
2017-05-21 23:51:33,112 INFO [main] manager.SqlManager: Using default fetchSize of 1000
2017-05-21 23:51:33,112 INFO [main] tool.CodeGenTool: Beginning code generation
2017-05-21 23:51:33,498 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_GREATER_ALL AS t WHERE 1=0
2017-05-21 23:51:33,583 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_GREATER_ALL AS t WHERE 1=0
2017-05-21 23:51:33,647 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop/hadoop-2.7.0
Note: /tmp/sqoop-nitesh/compile/d18a9d635c5dec87cd07fa6dde938e7b/BPL_ANS_GREATER_ALL.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-05-21 23:51:35,146 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-nitesh/compile/d18a9d635c5dec87cd07fa6dde938e7b/BPL_AN
S_GREATER_ALL.jar
2017-05-21 23:51:35,164 INFO [main] mapreduce.ExportJobBase: Beginning export of BPL_ANS_GREATER_ALL
2017-05-21 23:51:35,165 INFO [main] Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j.impl/StaticLoggerBinder.cl
ass]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerB
inder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-05-21 23:51:35,439 INFO [main] Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2017-05-21 23:51:36,585 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_ANS_GREATER_ALL AS t WHERE 1=0
2017-05-21 23:51:36,589 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce
.reduce.speculative
2017-05-21 23:51:36,592 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.ma
p.speculative
2017-05-21 23:51:36,613 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-21 23:51:36,775 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 23:52:02,355 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:52:02,370 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:52:03,580 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.f

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 23:51:36,589 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-21 23:51:36,592 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-21 23:51:36,613 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-21 23:51:36,775 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 23:52:02,355 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:52:02,376 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-21 23:52:03,580 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.in
putdir
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.ma
p.speculative
2017-05-21 23:52:03,734 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputform
at.class
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputfo
rmat.class
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.v
alue.class
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key
.class
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.
files
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-05-21 23:52:03,735 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.fi
les.filesizes
2017-05-21 23:52:04,523 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495350191822_0061
2017-05-21 23:52:08,725 INFO [main] impl.YarnClientImpl: Submitted application application_1495350191822_0061 to ResourceManager at /0.0.0.0:8
032
2017-05-21 23:52:08,777 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495350
191822_0061/
2017-05-21 23:52:08,777 INFO [main] mapreduce.Job: Running job: job_1495350191822_0061
2017-05-21 23:52:18,211 INFO [main] mapreduce.Job: Job job_1495350191822_0061 running in uber mode : false
2017-05-21 23:52:18,212 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-05-21 23:52:29,324 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-05-21 23:52:33,549 INFO [main] mapreduce.Job: Job job_1495350191822_0061 completed successfully

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
mysql> SELECT * FROM BPL_ANS_GREATER_ALL;
+-----+-----+-----+
| STATE | DISTRICT | PERCENT |
+-----+-----+-----+
| Andhra Pradesh | ANANTAPUR | 100.893 |
| Andhra Pradesh | KARIMNAGAR | 101.141 |
| Andhra Pradesh | KHAMMAM | 103.455 |
| Andhra Pradesh | NALGONDA | 104.536 |
| Andhra Pradesh | NIZAMABAD | 100 |
| Andhra Pradesh | WARANGAL | 108.924 |
| Arunachal Pradesh | DIBANG VALLEY | 100.276 |
| Arunachal Pradesh | TIRAP | 100 |
| Assam | HAILAKANDI | 100 |
| Bihar | MADHUBANI | 100 |
| Bihar | VAISHALI | 103.094 |
| Chhattisgarh | KORBA | 126.222 |
| Goa | NORTH GOA | 100 |
| Gujarat | AHMEDABAD | 100 |
| Gujarat | BHAVNAGAR | 100.824 |
| Gujarat | DANGS | 100 |
| Gujarat | JAMNAGAR | 105.154 |
| Gujarat | MAHESANA | 100.714 |
| Gujarat | NAVSARI | 100 |
| Gujarat | PATAN | 102.147 |
| Gujarat | PORBANDAR | 100 |
| Gujarat | RAJKOT | 100.867 |
| Gujarat | SURAT | 100 |
| Gujarat | VALSAD | 127.909 |
| Haryana | BHIMANI | 100.613 |
| Haryana | FARIDABAD | 100 |
| Haryana | GURGAON | 140.867 |
| Haryana | HISAR | 100 |
| Haryana | JHAJJAR | 100 |
| Haryana | KARNAL | 100.091 |
| Haryana | KURUKSHETRA | 100.271 |
| Haryana | MAHENDRAGARH | 100 |
| Haryana | PANCHKULA | 100 |
| Haryana | PANIPAT | 100 |
| Haryana | ROHTAK | 100 |
| Haryana | SIRSNA | 100 |
| Haryana | SONIPAT | 101.651 |
| Himachal Pradesh | BILASPUR | 109.614 |

```

nitesh@nitesh-Lenovo-Z50-70: ~

	Himachal Pradesh	BILASPUR	189.614
	Himachal Pradesh	CHAMBA	131.495
	Himachal Pradesh	HAMIRPUR	100
	Himachal Pradesh	KANGRA	137.378
	Himachal Pradesh	KINNAUR	100
	Himachal Pradesh	KULLU	100
	Himachal Pradesh	LAHAUL & SPITI	100
	Himachal Pradesh	MANDI	100.003
	Himachal Pradesh	SHIMLA	100
	Himachal Pradesh	SOLAN	100
	Himachal Pradesh	UNA	100
	Jharkhand	DEOCHAR	100
	Jharkhand	LOHARDAGA	100
	Jharkhand	PURBI SINGHBHUM	130.63
	Karnataka	BANGALORE RURAL	139.545
	Karnataka	HASSAN	100
	Karnataka	KOLAR	198.56
	Karnataka	MANGALORE(DAKSHINA KANNADA)	100
	Karnataka	SHIMOGA	100.571
	Kerala	UDUPI	100
	Kerala	ALAPPUZHA	100
	Kerala	ERNAKULAM	102.438
	Kerala	IDUKKI	103.488
	Kerala	KANNUR	110.278
	Kerala	KASARGOD	103.841
	Kerala	KOLLAM	100
	Kerala	KOTTAYAM	100
	Kerala	KOZHIKODE	100
	Kerala	MALAPPURAM	102.167
	Kerala	PALAKKAD	100
	Kerala	PATHANAMTHitta	100
	Kerala	THIRUVANANTHAPURAM	106.205
	Kerala	THRISSUR	104.824
	Kerala	WAYANAD	100
	Madhya Pradesh	ALIRAJPUR	138.577
	Madhya Pradesh	ANUPPUR	139.684
	Madhya Pradesh	BARWANI	111.848
	Madhya Pradesh	BETUL	100.601
	Madhya Pradesh	BHOPAL	156.822
	Madhya Pradesh	BURHANPUR	105.586
	Madhya Pradesh	DATIA	153.255
	Madhya Pradesh	DEWAS	134.584
	Madhya Pradesh	DHAR	119.608

nitesh@nitesh-Lenovo-Z50-70: ~

	Madhya Pradesh	DHAR	119.608
	Madhya Pradesh	GWALIOR	121.642
	Madhya Pradesh	HARDA	133.728
	Madhya Pradesh	HOSHANGABAD	164.708
	Madhya Pradesh	INDORE	102.355
	Madhya Pradesh	JABALPUR	110.833
	Madhya Pradesh	JHABUA	115.65
	Madhya Pradesh	KATNI	128.858
	Madhya Pradesh	KHANDWA(EAST NIMAR)	137.382
	Madhya Pradesh	KHARGONE	100.412
	Madhya Pradesh	MANDSAUR	145.391
	Madhya Pradesh	MORENA	122.32
	Madhya Pradesh	NEEMUCH	154.167
	Madhya Pradesh	RAISEN	115.4
	Madhya Pradesh	RAJGARH	122.283
	Madhya Pradesh	RATLAM	125.814
	Madhya Pradesh	REWA	121.395
	Madhya Pradesh	SEHORE	224.157
	Madhya Pradesh	SHAHDOOL	102.104
	Madhya Pradesh	SHAJAPUR	114.96
	Madhya Pradesh	SHEOPUR	159.111
	Madhya Pradesh	UJJAIN	113.473
	Madhya Pradesh	UMARIA	134.171
	Maharashtra	GADCHIROLI	100
	Maharashtra	RATNAGIRI	106.744
	Maharashtra	SINDHUDURG	100
	Meghalaya	WEST GARO HILLS	100
	Mizoram	CHAMPHALI	100
	Mizoram	LAWNGT LAI	100
	Punjab	BARNALA	103.757
	Punjab	FATEHgarh SAHIB	105.835
	Punjab	HOSHIARPUR	171.832
	Punjab	LUDHIANA	103.298
	Punjab	NAWANSHAHR	159.233
	Punjab	S.A.S Nagar	100.424
	Rajasthan	HANUMANGARH	100
	Sikkim	EAST SIKKIM	111.377
	Sikkim	NORTH SIKKIM	109.703
	Sikkim	SOUTH SIKKIM	138.751
	Sikkim	WEST SIKKIM	100.755
	Tamil Nadu	DHARMAPURI	100.032
	Tamil Nadu	ERODE	100
	Tamil Nadu	KANCHIPURAM	104.839

Tamil Nadu	KANCHIPURAM	184.839	
Tamil Nadu	KANYAKUMARI(NAGERCOIL)	111.267	
Tamil Nadu	KARUR	100	
Tamil Nadu	NAMAKKAL	100	
Tamil Nadu	PUDUKKOTTAI	112.103	
Tamil Nadu	RAMANATHAPURAM	149.246	
Tamil Nadu	TIRUCHIRAPPALLI	100	
Tamil Nadu	TIRUNELVELI	110.782	
Tamil Nadu	TIRUVANNAMALAI	100	
Tamil Nadu	TIRUVARUR	108.648	
Tripura	DHALAI	100	
Tripura	SOUTH TRIPURA	100	
Tripura	WEST TRIPURA	100	
Uttar Pradesh	AGRA	101.109	
Uttar Pradesh	ALLAHABAD	107.593	
Uttar Pradesh	AMBEDKAR NAGAR	100	
Uttar Pradesh	BALRAMPUR	100	
Uttar Pradesh	BARABANKI	101.396	
Uttar Pradesh	BAREILLY	100	
Uttar Pradesh	BIJNOR	100	
Uttar Pradesh	BUDAUN	100	
Uttar Pradesh	BULANDSHAHAR	110.77	
Uttar Pradesh	CHANDAULI	104.668	
Uttar Pradesh	CHITRAKOOT	112	
Uttar Pradesh	ETAH	111.513	
Uttar Pradesh	ETAWAH	100	
Uttar Pradesh	FARRUKHABAD	100	
Uttar Pradesh	FIROZABAD	100	
Uttar Pradesh	GHAZIABAD	100	
Uttar Pradesh	HARDOI	100	
Uttar Pradesh	JAUNPUR	100.37	
Uttar Pradesh	JHANSI	100.498	
Uttar Pradesh	JYOTIBA PHULE NAGAR	100	
Uttar Pradesh	KANNAUJ	100.117	
Uttar Pradesh	LAKHIMPUR KHERI	131.153	
Uttar Pradesh	LUCKNOW	100	
Uttar Pradesh	MAHAMAYA NAGAR(HATHRAS)	100.493	
Uttar Pradesh	MAHARAJGANJ	100	
Uttar Pradesh	MAHOBIA	100	
Uttar Pradesh	MATHURA	100.42	
Uttar Pradesh	MIRZAPUR	100.862	
Uttar Pradesh	MORADABAD	100	
Uttar Pradesh	MUZAFFARNAGAR	100	

Uttar Pradesh	BALRAMPUR	100	
Uttar Pradesh	BARABANKI	101.396	
Uttar Pradesh	BAREILLY	100	
Uttar Pradesh	BIJNOR	100	
Uttar Pradesh	BUDAUN	100	
Uttar Pradesh	BULANDSHAHAR	110.77	
Uttar Pradesh	CHANDAULI	104.668	
Uttar Pradesh	CHITRAKOOT	112	
Uttar Pradesh	ETAH	111.513	
Uttar Pradesh	ETAWAH	100	
Uttar Pradesh	FARRUKHABAD	100	
Uttar Pradesh	FIROZABAD	100	
Uttar Pradesh	GHAZIABAD	100	
Uttar Pradesh	HARDOI	100	
Uttar Pradesh	JAUNPUR	100.37	
Uttar Pradesh	JHANSI	100.498	
Uttar Pradesh	JYOTIBA PHULE NAGAR	100	
Uttar Pradesh	KANNAUJ	100.117	
Uttar Pradesh	LAKHIMPUR KHERI	131.153	
Uttar Pradesh	LUCKNOW	100	
Uttar Pradesh	MAHAMAYA NAGAR(HATHRAS)	100.493	
Uttar Pradesh	MAHARAJGANJ	100	
Uttar Pradesh	MAHOBIA	100	
Uttar Pradesh	MATHURA	100.42	
Uttar Pradesh	MIRZAPUR	100.862	
Uttar Pradesh	MORADABAD	100	
Uttar Pradesh	MUZAFFARNAGAR	100	
Uttar Pradesh	PILIBHIT	100	
Uttar Pradesh	PRATAPGARH	101.52	
Uttar Pradesh	RAE BARELI	100.065	
Uttar Pradesh	SAHARANPUR	100.259	
Uttar Pradesh	SHAHJAHANPUR	100.161	
Uttar Pradesh	SONBHADRA	100	
Uttar Pradesh	SULTANPUR	100	
Uttarakhand	NAINITAL	126.023	
Uttarakhand	RUDRAPRAYAG	105.019	
West Bengal	DAKSHIN DINAJPUR	100.839	
West Bengal	MIDNAPUR EAST	134.411	
West Bengal	MIDNAPUR WEST	117.035	

Here when we have executed the code for greater than or equal to 100% we get 176 rows .i.e. the 176 entries having performance greater than or equal to 100%

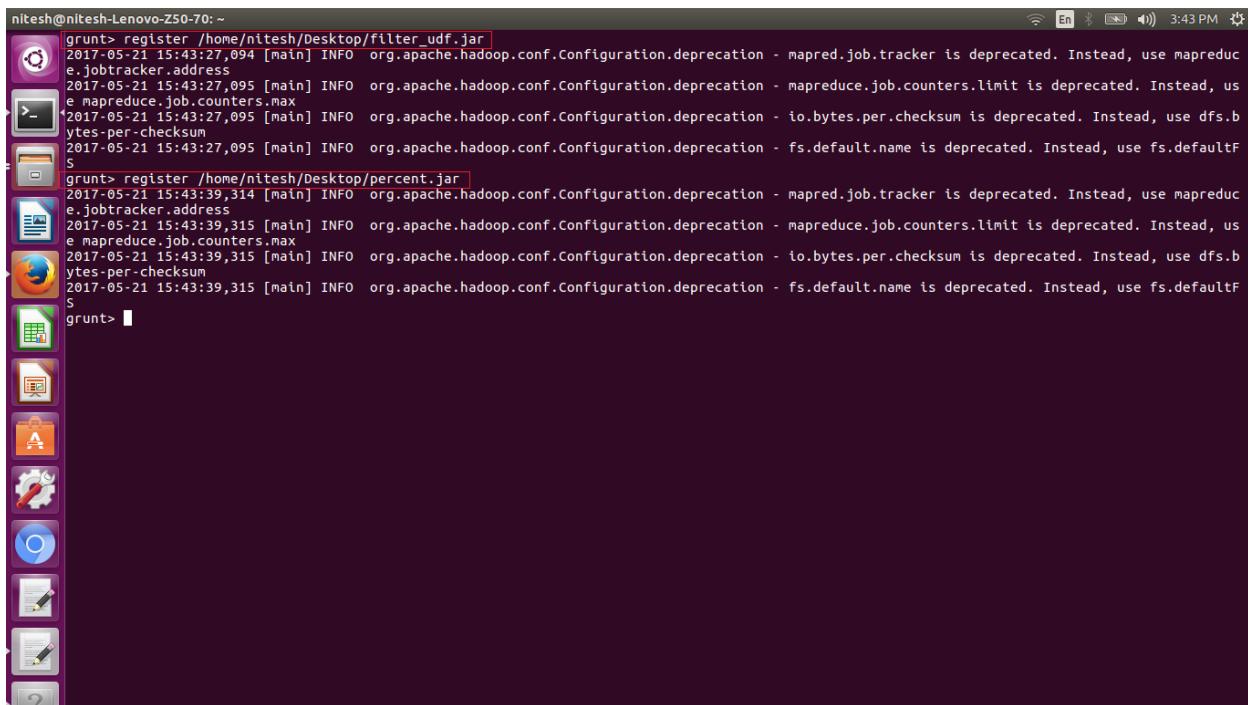
Problem statement 2

Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards.
Export the results to MySQL using Sqoop

Answer:

Here we are supposed to write the pig udf to filter the dataset to fetch the data who reached 80% of objectives. So we have written the udfs as shown in the above udf section.

The created udf's are now registered to use it in the script as shown below using the commands register for both the filter_udf and the percent udf.



```
gnutel@nitesh-Lenovo-Z50-70: ~
grunt> register /home/nitesh/Desktop/filter_udf.jar
2017-05-21 15:43:27,094 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
gnutel@nitesh-Lenovo-Z50-70: ~
grunt> register /home/nitesh/Desktop/percent.jar
2017-05-21 15:43:39,314 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
gnutel@nitesh-Lenovo-Z50-70: ~
gnutel@nitesh-Lenovo-Z50-70: ~
```

1. Now as we have registered jars we need to load the data set in the variable Input_data.

2. Loading of data is done using the following command

```
load 'hdfs:input path' Using PigStorage('delimiter') as (variable:datatype);
```

8. After the dataset is loaded in Input_data variable we have applied filter fuction so we can filter the loaded data to fetch the data who has achieved 80% of objective using

filter 'dataset' BY condition;

9. Once the filter operation is done we need to generate the columns to be stored using following syntax

```
foreach 'filtered data' generate state,district,percent;
```

10. The required columns are generated and now we need to store the data into the hdfs so we use

store ‘generated data’ into ‘path’ using org.apache.pig.piggybank.storage.CSVExcelStorage();

```
nitesh@nitesh-Lenovo-Z50-70:~  
grunt> Input_data = load 'hdfs:/pig_conversions/xml_to_csv/part-m-00000' USING PigStorage(',') as(State_Name:chararray,District_Name:chararray  
,Project_Objectives_IHHL_BPL:int , Project_Objectives_IHHL_AP:int , Project_Objectives_IHHL_TOTAL:int ,Project_Objectives_SCW:int ,Project_Obj  
e_IHHL_BPL:int , Project_Performance_IHHL_AP:Int ,Project_Performance_IHHL_TOTAL:int ,Project_Performance_SCW:int , Project_Performanc  
e_Totlets:int ,Project_Performance_Anganwadi_Totlets:int , Project_Performance_RSM:int ,Project_Performance_PC:int ,Project_Performanc  
e_IHHL_BPL:int , Project_Performance_Anganwadi_Totlets:int , Project_Performance_RSM:int ,Project_Performance_PC:int);  
2017-05-24 08:19:22,164 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc  
e.jobtracker.address  
2017-05-24 08:19:22,164 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us  
e mapreduce.job.counters.max  
2017-05-24 08:19:22,164 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b  
ytes-per-checksum  
2017-05-24 08:19:22,164 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultF  
S  
grunt> Filtered_data = filter Input_data BY Filter_udf($2,$10);  
grunt> Districts = foreach Filtered_data generate $0,$1.Percent($2,$10);  
grunt> STORE Districts INTO 'hdfs://Achieved_80_percent' USING org.apache.pig.piggybank.storage.CSVExcelStorage();  
2017-05-24 08:29:54,901 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc  
e.jobtracker.address  
2017-05-24 08:29:54,901 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us  
e mapreduce.job.counters.max  
2017-05-24 08:29:54,902 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b  
ytes-per-checksum  
2017-05-24 08:29:54,902 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultF  
S  
2017-05-24 08:29:54,948 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instea  
d, use mapreduce.output.textoutputformat.separator  
2017-05-24 08:29:54,980 [main] INFO org.apache.tools.pigstats.ScriptState - Pig features used in the script: FILTER  
2017-05-24 08:29:55,075 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultF  
S  
2017-05-24 08:29:55,075 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc  
e.jobtracker.address  
2017-05-24 08:29:55,077 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2017-05-24 08:29:55,104 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES_ENABLED=AddForEach, ColumnMapKeyPr  
une, ConstantCalculator, GroupByConcurrentSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer,  
PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]  
2017-05-24 08:29:55,145 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for Input_data: $3, $4, $5, $6, $  
7, $8, $9, $11, $12, $13, $14, $15, $16, $17  
2017-05-24 08:29:55,305 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 10  
0 optimistic? false  
2017-05-24 08:29:55,324 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before opt  
imization: 1  
2017-05-24 08:29:55,324 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after opti  
mization:  
  
nitesh@nitesh-Lenovo-Z50-70:~  
2017-05-24 08:29:59,192 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2017-05-24 08:29:59,192 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
2017-05-24 08:29:59,290 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to proc  
ess : 1  
2017-05-24 08:29:59,425 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2017-05-24 08:29:59,472 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.timestamps is deprecated. Inst  
ead, use mapreduce.job.cache.files.timestamps  
2017-05-24 08:29:59,473 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.working.dir is deprecated. Instead, use ma  
preduce.job.working_dir  
2017-05-24 08:29:59,473 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.j  
ob.user.name  
2017-05-24 08:29:59,473 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.filesizes is deprecated. Inste  
ad, use mapreduce.job.cache.files.filesizes  
2017-05-24 08:30:00,047 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495593828721_0001  
2017-05-24 08:30:00,346 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of reso  
urces.  
2017-05-24 08:30:00,648 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - submitted application application_1495593828  
721_0001 to ResourceManager at /0.0.0.0:8032  
2017-05-24 08:30:00,808 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/  
proxy/application_1495593828721_0001/  
2017-05-24 08:30:00,814 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_14955938  
28721_0001  
2017-05-24 08:30:00,814 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Distri  
cts_Filtered_data_Input_data  
2017-05-24 08:30:00,814 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: In  
put_data[4,13],Input_data[-1,-1],Filtered_data[5,16],Districts[6,12] C: R  
2017-05-24 08:30:00,859 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete  
2017-05-24 08:30:00,859 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149  
5593828721_0001]  
2017-05-24 08:30:29,876 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete  
2017-05-24 08:30:29,876 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149  
5593828721_0001]  
2017-05-24 08:30:35,891 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-05-24 08:30:35,980 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=S  
UCCEEDED. Redirecting to job history server  
2017-05-24 08:30:37,133 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-05-24 08:30:37,139 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=S  
UCCEEDED. Redirecting to job history server  
2017-05-24 08:30:37,243 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-05-24 08:30:37,247 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=S  
UCCEEDED. Redirecting to job history server  
2017-05-24 08:30:37,359 [main] WARN org.apache.tools.pigstats.mapreduce.MRJobStats - Unable to get job counters  
java.io.IOException: java.lang.IllegalArgumentException: No enum constant org.apache.hadoop.mapreduce.JobCounter.MB_MILLIS_MAPS
```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-24 08:30:37,372 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-24 08:29:55 2017-05-24 08:30:37 FILTER
Success!
Job Stats (time in seconds):
Jobid Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
educetime Alias Feature Outputs job_1495593828721_0001 1 0 9 9 9 0 0 0 0 Districts,Filtered_data,Input_data M
AP_ONLY hdfs:/Achieved_80_percent,
Input(s):
Successfully read 0 records from: "hdfs:/pig_conversions/xml_to_csv/part-m-00000"
Output(s):
Successfully stored 0 records in: "hdfs:/Achieved_80_percent"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495593828721_0001
2017-05-24 08:30:37,374 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-24 08:30:37,380 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-24 08:30:37,427 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-24 08:30:37,432 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-24 08:30:37,493 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-24 08:30:37,553 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-24 08:30:37,591 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-05-24 08:30:37,591 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> 

```

store command was successfully executed and the filtered data is written in the hdfs as we can see in the following image. Where we have used

adoop fs –ls /

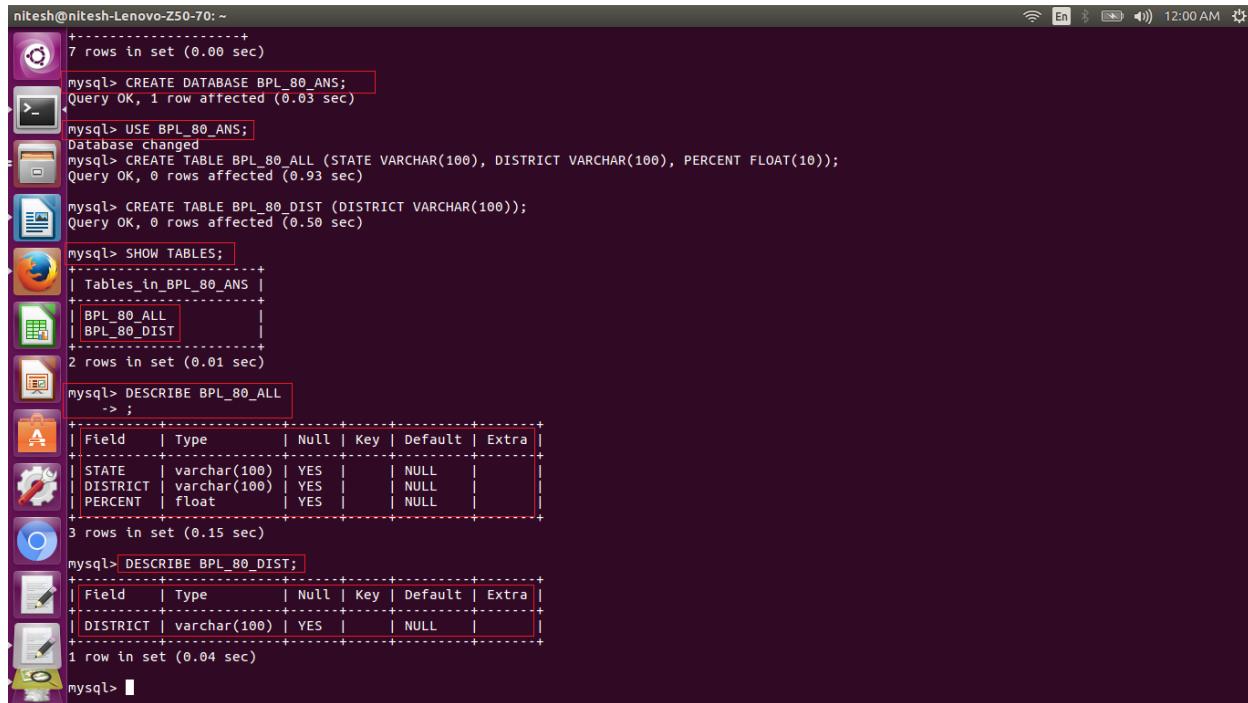
to get the folder where we have stored and it is marked with red box in the image

```

nitesh@nitesh-Lenovo-Z50-70: ~
nitesh@nitesh-Lenovo-Z50-70: ~$ hadoop fs -ls /
Found 23 items
drwxr-xr-x - nitesh supergroup 0 2017-04-23 00:03 /15_grd
drwxr-xr-x - nitesh supergroup 0 2017-04-23 01:52 /15_mvn
drwxr-xr-x - nitesh supergroup 0 2017-04-22 23:20 /15_sbt
drwxr-xr-x - nitesh supergroup 0 2017-04-30 16:25 /20_1
drwxr-xr-x - nitesh supergroup 0 2017-04-30 17:32 /20_2
drwxr-xr-x - nitesh supergroup 0 2017-04-30 10:30 /7_1
drwxr-xr-x - nitesh supergroup 0 2017-04-30 10:51 /9_1
drwxr-xr-x - nitesh supergroup 0 2017-05-21 21:56 /Achieved_80_district_percent
drwxr-xr-x - nitesh supergroup 0 2017-05-21 21:48 /Achieved_80_percent
drwxr-xr-x - nitesh supergroup 0 2017-05-21 17:42 /BPL_ANSWER
drwxr-xr-x - nitesh supergroup 0 2017-05-21 18:00 /BPL_ANSWER_District
drwxr-xr-x - nitesh supergroup 0 2017-05-21 18:47 /BPL_ANSWER_greater_District
drwxr-xr-x - nitesh supergroup 0 2017-05-21 17:52 /BPL_ANSWER_greater_equal
drwxr-xr-x - nitesh supergroup 0 2017-05-20 23:50 /pig_conversions
-rw-r--r-- 4 nitesh supergroup 391461 2017-05-20 17:29 /piggybank-0.15.0.jar
drwxr-xr-x - nitesh supergroup 0 2017-05-20 16:56 /pjct
drwxr-xr-x - nitesh supergroup 0 2017-05-20 17:04 /project
-rw-r--r-- 4 nitesh supergroup 717414 2017-05-20 16:34 /project1
drwxr-xr-x - nitesh supergroup 0 2017-05-20 17:12 /projectinput
drwxrwxrwx - nitesh supergroup 0 2017-05-21 17:18 /tmp
drwxr-xr-x - nitesh supergroup 0 2017-05-13 18:34 /user
drwxr-xr-x - nitesh supergroup 0 2017-04-05 00:04 /usr
drwxr-xr-x - nitesh supergroup 0 2017-05-20 23:59 /xml_to_csv
nitesh@nitesh-Lenovo-Z50-70: ~$ hadoop fs -ls /BPL_ANSWER_District
Found 2 items
-rw-r--r-- 4 nitesh supergroup 0 2017-05-21 18:00 /BPL_ANSWER_District/_SUCCESS
-rw-r--r-- 4 nitesh supergroup 690 2017-05-21 18:00 /BPL_ANSWER_District/part-m-00000
nitesh@nitesh-Lenovo-Z50-70: ~$ hadoop fs -ls //BPL_ANSWER
ls: No FileSystem for scheme: null
nitesh@nitesh-Lenovo-Z50-70: ~$ hadoop fs -ls /BPL_ANSWER
Found 2 items
-rw-r--r-- 4 nitesh supergroup 0 2017-05-21 17:42 /BPL_ANSWER/_SUCCESS
-rw-r--r-- 4 nitesh supergroup 1886 2017-05-21 17:42 /BPL_ANSWER/part-m-00000
nitesh@nitesh-Lenovo-Z50-70: ~$ hadoop fs -ls /Achieved_80_percent
Found 2 items
-rw-r--r-- 4 nitesh supergroup 0 2017-05-21 21:48 /Achieved_80_percent/_SUCCESS
-rw-r--r-- 4 nitesh supergroup 10382 2017-05-21 21:48 /Achieved_80_percent/part-m-00000
nitesh@nitesh-Lenovo-Z50-70: ~$ hadoop fs -ls /Achieved_80_district_percent
Found 2 items
-rw-r--r-- 4 nitesh supergroup 0 2017-05-21 21:56 /Achieved_80_district_percent/_SUCCESS
-rw-r--r-- 4 nitesh supergroup 3356 2017-05-21 21:56 /Achieved_80_district_percent/part-m-00000

```

As we now want to export the stored filtered data from hdfs to mysql db we need to create the database to create the table with the required number of columns to export the data from the hdfs to mysql using sqoop as per our requirement.



```
nitesh@nitesh-Lenovo-Z50-70:~  
+-----+  
7 rows in set (0.00 sec)  
  
mysql> CREATE DATABASE BPL_80_ANS;  
Query OK, 1 row affected (0.03 sec)  
  
mysql> USE BPL_80_ANS;  
Database changed  
mysql> CREATE TABLE BPL_80_ALL (STATE VARCHAR(100), DISTRICT VARCHAR(100), PERCENT FLOAT(10));  
Query OK, 0 rows affected (0.93 sec)  
  
mysql> CREATE TABLE BPL_80_DIST (DISTRICT VARCHAR(100));  
Query OK, 0 rows affected (0.50 sec)  
  
mysql> SHOW TABLES;  
+-----+  
| Tables_in_BPL_80_ANS |  
+-----+  
| BPL_80_ALL |  
| BPL_80_DIST |  
+-----+  
2 rows in set (0.01 sec)  
  
mysql> DESCRIBE BPL_80_ALL;  
-> ;  
+-----+-----+-----+-----+-----+  
| Field | Type   | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+  
| STATE | varchar(100) | YES |   | NULL    |       |  
| DISTRICT | varchar(100) | YES |   | NULL    |       |  
| PERCENT | float      | YES |   | NULL    |       |  
+-----+-----+-----+-----+-----+  
3 rows in set (0.15 sec)  
  
mysql> DESCRIBE BPL_80_DIST;  
+-----+-----+-----+-----+-----+  
| Field | Type   | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+  
| DISTRICT | varchar(100) | YES |   | NULL    |       |  
+-----+-----+-----+-----+-----+  
1 row in set (0.04 sec)  
  
mysql>
```

In the below screen shot we can see that the sqoop command is used to export the data from the hdfs to mysql. The command can be used as per below

```
sqoop export -m 1 --connect jdbc:mysql://localhost:3306/'database name' --username username  
--password password --table 'table name' --export-dir 'hdfs path of the file to be exported' --driver  
com.mysql.jdbc.Driver
```

```

nitesh@nitesh-Lenovo-Z50-70: ~
nitesh@nitesh-Lenovo-Z50-70:~$ sqoop export -m 1 --connect jdbc:mysql://localhost:3306/BPL_80_ANS --username root --password nss8055 --table BPL_80_ALL --export-dir /Achieved_80_percent/part-m-00000 --driver com.mysql.jdbc.Driver
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../catalog does not exist! HCatalog jobs will fail.
Please set $CAT_HOME to the root of your HCatalog installation.
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop-2.0.4-alpha/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-05-22 00:06:20,720 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-05-22 00:06:20,731 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-05-22 00:06:20,791 WARN [main] sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate Connection manager is not being set (via --connection-manager). Sqoop is going to fall back to org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.
2017-05-22 00:06:20,803 INFO [main] manager.SqlManager: Using default fetchSize of 1000
2017-05-22 00:06:20,803 INFO [main] tool.CodeGenTool: Beginning code generation
2017-05-22 00:06:21,183 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_80_ALL AS t WHERE 1=0
2017-05-22 00:06:21,209 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_80_ALL AS t WHERE 1=0
2017-05-22 00:06:21,253 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop/hadoop-2.7.0
Note: /tmp/sqoop-nitesh/compile/41ff98647e6452fde0d47b8d14020776/BPL_80_ALL.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-05-22 00:06:23,223 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-nitesh/compile/41ff98647e6452fde0d47b8d14020776/BPL_80_ALL.jar
2017-05-22 00:06:23,237 INFO [main] mapreduce.ExportJobBase: Beginning export of BPL_80_ALL
2017-05-22 00:06:23,238 INFO [main] Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/lib/slf4j-jlog4j12-1.6.4.jar!/org/slf4j.impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-jlog4j12-1.7.10.jar!/org/slf4j.impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-05-22 00:06:23,466 INFO [main] Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2017-05-22 00:06:24,323 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_80_ALL AS t WHERE 1=0
2017-05-22 00:06:24,328 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-22 00:06:24,333 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-22 00:06:24,334 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-22 00:06:24,439 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-22 00:06:43,380 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:06:43,382 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:06:43,726 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-22 00:06:43,795 INFO [main] configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-22 00:06:43,795 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.f

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-22 00:06:24,328 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-22 00:06:24,333 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-22 00:06:24,334 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-22 00:06:24,439 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-22 00:06:43,380 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:06:43,382 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:06:43,726 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-22 00:06:43,795 INFO [main] configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-22 00:06:43,795 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-22 00:06:43,795 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-22 00:06:43,795 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2017-05-22 00:06:43,795 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-22 00:06:43,795 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputformat.class
2017-05-22 00:06:43,795 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputformat.class
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.value.class
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key.class
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.files
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-05-22 00:06:43,796 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-22 00:06:44,459 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495350191822_0063
2017-05-22 00:06:45,990 INFO [main] impl.YarnClientImpl: Submitted application application_1495350191822_0063 to ResourceManager at /0.0.0.0:8032
2017-05-22 00:06:46,138 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495350191822_0063/
2017-05-22 00:06:46,138 INFO [main] mapreduce.Job: Running job: job_1495350191822_0063
2017-05-22 00:06:55,009 INFO [main] mapreduce.Job: Job job_1495350191822_0063 running in uber mode : false
2017-05-22 00:06:55,012 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-05-22 00:07:06,773 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-05-22 00:07:10,805 INFO [main] mapreduce.Job: Job job_1495350191822_0063 completed successfully

```

Here in the above image we can see in the red box that the job is completed successfully. So the data is being exported to mysql database.

To verify we open the mysql using the command

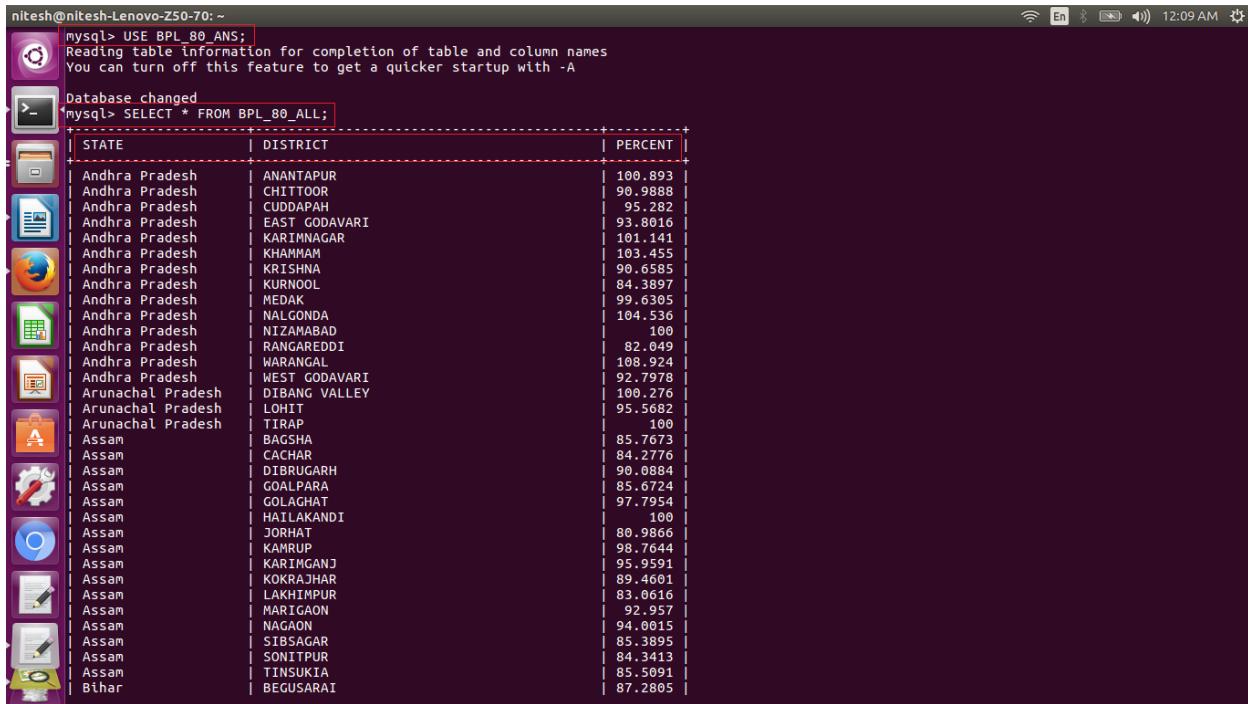
mysql -u root -p

After the mysql shell is open we need to use the created database using the command use ‘database name’

then use use the table to query the data in the table using the select query as

```
select * from ‘table name’;
```

and we fetch all the data from the table that is exported using sqoop from hdfs.



```
nitesh@nitesh-Lenovo-Z50-70: ~
mysql> USE BPL_80_ANS;
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> SELECT * FROM BPL_80_ALL;
+-----+-----+-----+
| STATE | DISTRICT | PERCENT |
+-----+-----+-----+
| Andhra Pradesh | ANANTAPUR | 100.893 |
| Andhra Pradesh | CHITTOOR | 90.9888 |
| Andhra Pradesh | CUDDAPAH | 95.282 |
| Andhra Pradesh | EAST GODAVARI | 93.8016 |
| Andhra Pradesh | KARIMNAGAR | 101.141 |
| Andhra Pradesh | KHAMMAM | 103.455 |
| Andhra Pradesh | KRISHNA | 90.6585 |
| Andhra Pradesh | KURNOOL | 84.3897 |
| Andhra Pradesh | MEDAK | 99.6305 |
| Andhra Pradesh | NALGONDA | 104.536 |
| Andhra Pradesh | NIZAMABAD | 100 |
| Andhra Pradesh | RANGAREDDI | 82.049 |
| Andhra Pradesh | WARANGAL | 108.924 |
| Andhra Pradesh | WEST GODAVARI | 92.7978 |
| Arunachal Pradesh | DIBANG VALLEY | 100.276 |
| Arunachal Pradesh | LOHIT | 95.5682 |
| Arunachal Pradesh | TIRAP | 100 |
| Assam | BAGSHA | 85.7673 |
| Assam | CACHAR | 84.2776 |
| Assam | DIBRUGARH | 90.0884 |
| Assam | GOALPARA | 85.6724 |
| Assam | GOLAGHAT | 97.7954 |
| Assam | HAILAKANDI | 100 |
| Assam | JORHAT | 80.9866 |
| Assam | KAMRUP | 98.7644 |
| Assam | KARIMGANJ | 95.9591 |
| Assam | KOKRAJHAR | 89.4601 |
| Assam | LAKHIMPUR | 83.0616 |
| Assam | MARIGAON | 92.957 |
| Assam | NAGAON | 94.0015 |
| Assam | STBSAGAR | 85.3895 |
| Assam | SONITPUR | 84.3413 |
| Assam | TINSUKIA | 85.5091 |
| Bihar | BEGUSARAI | 87.2805 |
+-----+-----+-----+
```

Haryana	REWARI	88.4933		
Haryana	ROHTAK	100		
Haryana	SIRSA	100		
Haryana	SONIPAT	101.651		
Haryana	YAMUNANAGAR	98.8123		
Himachal Pradesh	BILASPUR	109.614		
Himachal Pradesh	CHAMBA	131.495		
Himachal Pradesh	HAMIRPUR	100		
Himachal Pradesh	KANGRA	137.378		
Himachal Pradesh	KINNAUR	100		
Himachal Pradesh	KULLU	100		
Himachal Pradesh	LAHAUL & SPITI	100		
Himachal Pradesh	MANDI	100.003		
Himachal Pradesh	SHIMLA	100		
Himachal Pradesh	SIRMAUR	99.4038		
Himachal Pradesh	SOLAN	100		
Himachal Pradesh	UNA	100		
Jammu & Kashmir	ANANTNAG	82.0632		
Jammu & Kashmir	LEH (LADAKH)	86.9622		
Jharkhand	DEOGHAR	100		
Jharkhand	DUMKA	85.328		
Jharkhand	LATEHAR	83.4495		
Jharkhand	LOHARDAGA	100		
Jharkhand	PAKUR	87.3495		
Jharkhand	PURBI SINGHBHUM	130.63		
Karnataka	BAGALKOT	91.908		
Karnataka	BANGALORE RURAL	139.545		
Karnataka	CHICKMAGALUR	95.9966		
Karnataka	CHITRADURGA	97.5928		
Karnataka	DHARWAD	92.9976		
Karnataka	GADAG	84.1802		
Karnataka	HASSAN	100		
Karnataka	KODAGU	99.485		
Karnataka	KOLAR	198.56		
Karnataka	KOPPAL	87.524		
Karnataka	MANDYA	96.6749		
Karnataka	MANGALORE(DAKSHINA KANNADA)	100		
Karnataka	RAMANAGARA	89.0558		
Karnataka	SHIMOGA	100.571		
Karnataka	UDUPI	100		
Kerala	ALAPPUZHA	100		
Kerala	ERNAKULAM	102.438		
Kerala	IDUKKI	103.488		

Maharashtra	HINGOLI	91.2322		
Maharashtra	JALNA	97.2189		
Maharashtra	KOLHAPUR	88.2574		
Maharashtra	NAGPUR	88.3415		
Maharashtra	OSMANABAD	99.9779		
Maharashtra	PARBHANI	87.1717		
Maharashtra	PUNE	97.7041		
Maharashtra	RATNAGIRI	106.744		
Maharashtra	SANGLI	82.9894		
Maharashtra	SATARA	85.8256		
Maharashtra	SINDHUDURG	100		
Maharashtra	THANE	95.5074		
Maharashtra	WARDHA	92.7833		
Manipur	BISHNUPUR	87.0639		
Manipur	IMPHAL EAST	95.1551		
Manipur	TAMENGLONG	88.5351		
Meghalaya	RI BHOI	84.4249		
Meghalaya	SOUTH GARO HILLS	90.5283		
Meghalaya	WEST GARO HILLS	100		
Mizoram	CHAMPHALI	100		
Mizoram	KOLASIB	98.0867		
Mizoram	LAWNGTIAI	100		
Mizoram	LUNGLEI	89.362		
Mizoram	MAMIT	97.7102		
Mizoram	SAIHA	88.6571		
Mizoram	SERCHHIP	88.9024		
Nagaland	KOHIMA	93.6392		
Nagaland	MOKOKCHUNG	96.9273		
Nagaland	PHEK	86.9651		
Orissa	BALESWAR	96.7402		
Orissa	JAGATSINGHAPUR	95.7636		
Punjab	BARNALA	103.757		
Punjab	FATEHGARH SAHIB	105.835		
Punjab	HOSHIARPUR	171.832		
Punjab	JALANDHAR	98.8333		
Punjab	KAPURTHALA	95.8627		
Punjab	LUDHIANA	103.298		
Punjab	MANSA	93.592		
Punjab	NAWANSHAHR	159.233		
Punjab	S.A.S Nagar	100.424		
Rajasthan	AJMER	89.486		
Rajasthan	CHURU	94.6775		
Rajasthan	DUNGARPUR	82.7233		

nitesh@nitesh-Lenovo-Z50-70: ~

Tamil Nadu	TIRUVANNAMALAI	100	
Tamil Nadu	TIRUVARUR	108.648	
Tamil Nadu	VELLORE	96.2776	
Tamil Nadu	VIRUDHUNAGAR	87.6451	
Tripura	DHALAI	100	
Tripura	NORTH TRIPURA	94.0234	
Tripura	SOUTH TRIPURA	100	
Tripura	WEST TRIPURA	100	
Uttar Pradesh	AGRA	101.109	
Uttar Pradesh	ALIGARH	92.1	
Uttar Pradesh	ALLAHABAD	107.593	
Uttar Pradesh	AMBEDKAR NAGAR	100	
Uttar Pradesh	AZAMGARH	86.3251	
Uttar Pradesh	BAGPAT	85.086	
Uttar Pradesh	BALLIA	97.2767	
Uttar Pradesh	BALRAMPUR	100	
Uttar Pradesh	BANDA	92.4382	
Uttar Pradesh	BARABANKI	101.396	
Uttar Pradesh	BAREILLY	100	
Uttar Pradesh	BASTI	81.1466	
Uttar Pradesh	BITJNOR	100	
Uttar Pradesh	BUDAUN	100	
Uttar Pradesh	BULANDSHahr	110.77	
Uttar Pradesh	CHANDAULI	104.668	
Uttar Pradesh	CHITRAKOOT	112	
Uttar Pradesh	DEORIA	95.0321	
Uttar Pradesh	ETAH	111.513	
Uttar Pradesh	ETAWAH	100	
Uttar Pradesh	FAIZABAD	98.3188	
Uttar Pradesh	FARRUKHABAD	100	
Uttar Pradesh	FATEHPUR	95.7352	
Uttar Pradesh	FIROZABAD	100	
Uttar Pradesh	GAUTAM BUDDHA NAGAR	98.1337	
Uttar Pradesh	GHAZIABAD	100	
Uttar Pradesh	GHAZIPUR	99.1362	
Uttar Pradesh	GONDA	86.8813	
Uttar Pradesh	GORAKHPUR	99.9891	
Uttar Pradesh	HAMIRPUR	84.0198	
Uttar Pradesh	HARDOI	100	
Uttar Pradesh	JALAN	89.106	
Uttar Pradesh	JAUNPUR	100.37	
Uttar Pradesh	JHANSI	100.498	
Uttar Pradesh	JYOTIBA PHULE NAGAR	100	

nitesh@nitesh-Lenovo-Z50-70: ~

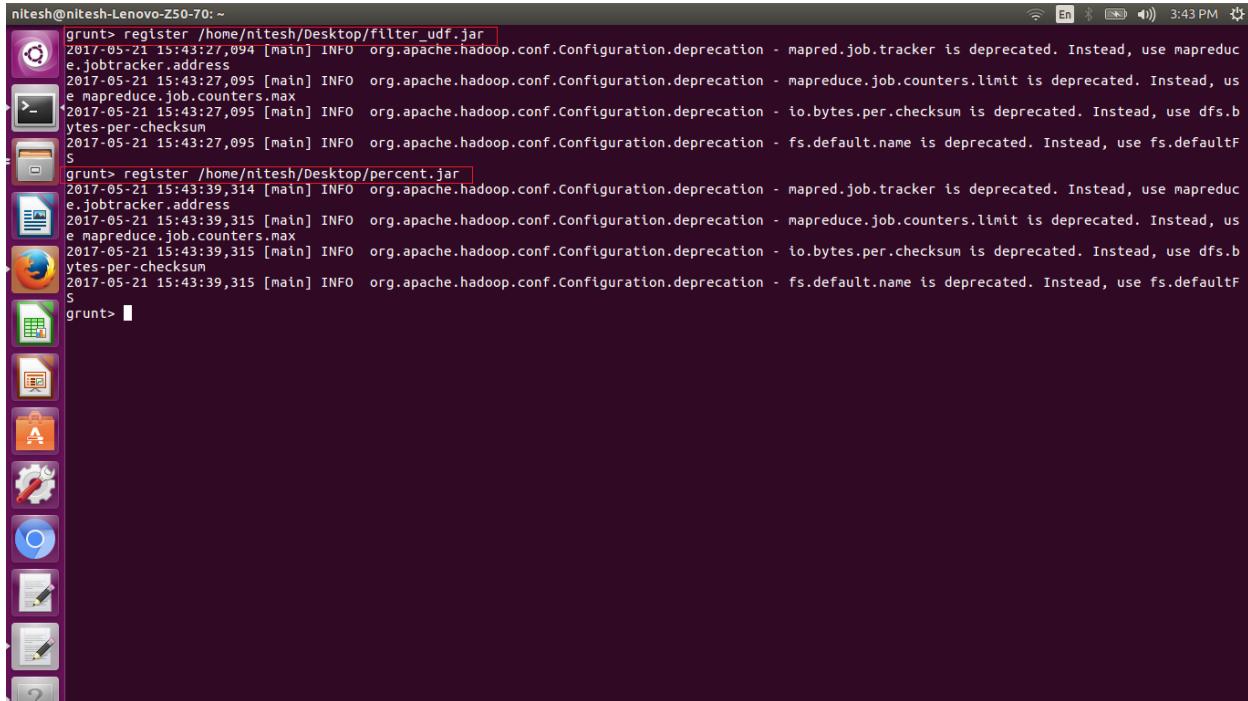
Uttar Pradesh	MAU	81.1694	
Uttar Pradesh	MEERUT	99.3847	
Uttar Pradesh	MIRzapur	100.862	
Uttar Pradesh	MORADABAD	100	
Uttar Pradesh	MUZAFFARNAGAR	100	
Uttar Pradesh	Pilibhit	100	
Uttar Pradesh	PRATAPGARH	101.52	
Uttar Pradesh	RAE BARELI	100.065	
Uttar Pradesh	RAMPUR	91.2306	
Uttar Pradesh	SAHARANPUR	100.259	
Uttar Pradesh	SANT RAVIDAS NAGAR(BHADOHI)	93.0577	
Uttar Pradesh	SHAHJAHANPUR	100.161	
Uttar Pradesh	SHRAVASTI	94.146	
Uttar Pradesh	SIDDHARTHNAGAR	89.2548	
Uttar Pradesh	SITAPUR	89.5722	
Uttar Pradesh	SONBHADRA	100	
Uttar Pradesh	SULTANPUR	100	
Uttar Pradesh	UNNAO	97.4602	
Uttar Pradesh	VARANASI	97.1748	
Uttarakhand	BAGESHWAR	83.2704	
Uttarakhand	CHAMOLI	95.9333	
Uttarakhand	DEHRADUN	85.2521	
Uttarakhand	HARIDWAR	89.8541	
Uttarakhand	NAINITAL	126.023	
Uttarakhand	PITHORAGARH	84.1571	
Uttarakhand	RUDRAPRAYAG	105.019	
Uttarakhand	TEHRI GARHWAL	83.9052	
Uttarakhand	UDHAM SINGH NAGAR	95.3763	
Uttarakhand	UTTARKASHI	98.5424	
West Bengal	BARDHAMAN	85.9808	
West Bengal	DAKSHIN DINAJPUR	100.839	
West Bengal	HOOGLY	99.2794	
West Bengal	HOWRAH	99.2797	
West Bengal	JALPAIGURI	96.5472	
West Bengal	MIDNAPUR EAST	134.411	
West Bengal	MIDNAPUR WEST	117.035	
West Bengal	NADIA	92.7216	
West Bengal	NORTH 24 PARAGANAS	99.0312	
West Bengal	SOUTH 24 PARAGANAS	94.4331	

349 rows in set (0.00 sec)

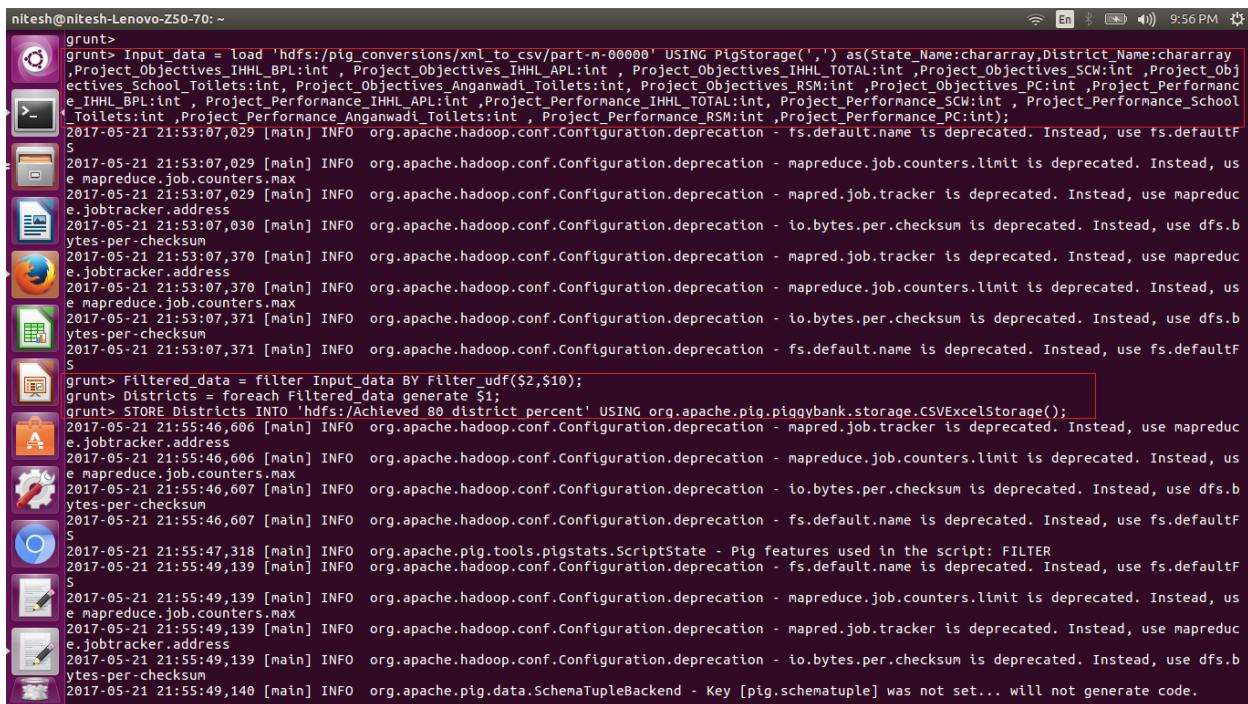
mysql> |

here we have 349 rows in the table that is the number of entries who achieved 80% of the objective

Here we have screen shots of the processes we followed as of above fo the condition greater than or equal to 80% using pig udf to generate all districts only.



```
nitesh@nitesh-Lenovo-Z50-70:~  
grunt> register /home/nitesh/Desktop/filter_udf.jar  
2017-05-21 15:43:27,094 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max  
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2017-05-21 15:43:27,095 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> register /home/nitesh/Desktop/percent.jar  
2017-05-21 15:43:39,314 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max  
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2017-05-21 15:43:39,315 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt>
```



```
nitesh@nitesh-Lenovo-Z50-70:~  
grunt>  
grunt> Input_data = load 'hdfs://pig/conversions/xml_to_csv/part-m-00000' USING PigStorage('') as(State_Name:chararray,District_Name:chararray  
,Project_Objectives_IHML_BPL:int , Project_Objectives_IHML_AP:int , Project_Objectives_IHML_TOTAL:int ,Project_Objectives_SCW:int ,Project_Ob  
jectives_School_Toilets:int , Project_Objectives_Anganwadi_Toilets:int , Project_Objectives_RSM:int ,Project_Objectives_PC:int ,Project_Performanc  
e_IHML_BPL:int , Project_Performance_IHML_AP:Int ,Project_Performance_IHML_TOTAL:Int ,Project_Performance_SCW:int , Project_Performance_Schoo  
l_Toilets:int ,Project_Performance_Anganwadi_Toilets:int , Project_Performance_RSM:int ,Project_Performance_PC:int);  
2017-05-21 21:53:07,029 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-05-21 21:53:07,029 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us  
e mapreduce.job.counters.max  
2017-05-21 21:53:07,029 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc  
e.jobtracker.address  
2017-05-21 21:53:07,030 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b  
ytes-per-checksum  
2017-05-21 21:53:07,370 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc  
e.jobtracker.address  
2017-05-21 21:53:07,370 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us  
e mapreduce.job.counters.max  
2017-05-21 21:53:07,371 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b  
ytes-per-checksum  
2017-05-21 21:53:07,371 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> Filtered_data = filter Input_data BY Filter_udf($2,$10);  
grunts: Districts = foreach Filtered_data generate $1;  
grunts: STORE Districts INTO 'hdfs://Achieved_80_district_percent' USING org.apache.pig.piggypbank.storage.CSVExcelStorage();  
2017-05-21 21:55:46,606 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc  
e.jobtracker.address  
2017-05-21 21:55:46,606 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us  
e mapreduce.job.counters.max  
2017-05-21 21:55:46,607 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b  
ytes-per-checksum  
2017-05-21 21:55:46,607 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-05-21 21:55:47,318 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER  
2017-05-21 21:55:49,139 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-05-21 21:55:49,139 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, us  
e mapreduce.job.counters.max  
2017-05-21 21:55:49,139 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduc  
e.jobtracker.address  
2017-05-21 21:55:49,139 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b  
ytes-per-checksum  
2017-05-21 21:55:49,140 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 21:55:51,842 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2017-05-21 21:55:51,842 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2017-05-21 21:55:51,843 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserializ
e []
2017-05-21 21:55:51,852 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiti
ng for submission.
2017-05-21 21:55:51,854 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 21:55:52,085 [JobControl] WARN org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set. User classes may not be found. See
Job or Job#setJar(String).
2017-05-21 21:55:52,110 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-21 21:55:52,110 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-05-21 21:55:52,110 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to proc
ess : 1
2017-05-21 21:55:52,452 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-21 21:55:53,051 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495350191822_0056
2017-05-21 21:55:53,084 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of reso
urces.
2017-05-21 21:55:53,361 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495350191
822_0056 to ResourceManager at /0.0.0.0:8032
2017-05-21 21:55:53,468 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/
proxy/application_1495350191822_0056/
2017-05-21 21:55:53,468 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_149
5350191822_0056
2017-05-21 21:55:53,469 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Distri
cts,Filtered_data,Input_data
2017-05-21 21:55:53,469 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: In
put_data[35,13],Input_data[-1,1],Filtered_data[36,16],Districts[37,12] C: R:
2017-05-21 21:55:53,484 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2017-05-21 21:55:53,484 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149
5350191822_0056]
2017-05-21 21:56:15,939 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2017-05-21 21:56:15,939 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_149
5350191822_0056]
2017-05-21 21:56:23,950 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 21:56:24,171 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 21:56:26,196 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 21:56:26,226 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 21:56:26,401 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 21:56:26,427 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 21:56:26,610 [main] WARN org.apache.pig.tools.pigstats.mapreduce.MRJobStats - Unable to get job counters

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-21 21:56:26,612 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-21 21:55:49 2017-05-21 21:56:26 FILTER
Success!
Job Stats (time in seconds):
Jobid Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
educetime Alias Feature Outputs
job_1495350191822_0056 1 0 11 11 11 11 0 0 0 Districts,Filtered_data,Input_data M
AP_ONLY hdfsf:/Achieved_80_district_percent,
Input(s):
Successfully read 0 records from: "hdfs:/pig_conversions/xml_to_csv/part-m-00000"
Output(s):
Successfully stored 0 records in: "hdfs:/Achieved_80_district_percent"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495350191822_0056

2017-05-21 21:56:26,614 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 21:56:26,675 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 21:56:26,848 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 21:56:26,917 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 21:56:27,035 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-21 21:56:27,111 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-21 21:56:27,218 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to
compute warning aggregation.
2017-05-21 21:56:27,218 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> ■

```

```
nitesh@nitesh-Lenovo-Z50-70:~ nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /  
Found 23 items  
drwxr-xr-x  - nitesh supergroup          0 2017-04-23 00:03 /15_grd  
drwxr-xr-x  - nitesh supergroup          0 2017-04-23 01:52 /15_mv  
drwxr-xr-x  - nitesh supergroup          0 2017-04-22 23:20 /15_sbt  
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 16:25 /20_1  
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 17:32 /20_2  
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 10:30 /7_1  
drwxr-xr-x  - nitesh supergroup          0 2017-04-30 10:51 /9_1  
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 21:56 /Achieved_80_district_percent  
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 21:48 /Achieved_80_percent  
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 17:42 /BPL_ANSWER  
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 18:00 /BPL_ANSWER_District  
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 18:47 /BPL_ANSWER_greater_District  
drwxr-xr-x  - nitesh supergroup          0 2017-05-21 17:52 /BPL_ANSWER_greater_equal  
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 23:58 /pig_conversions  
drwxr-xr-x  4 nitesh supergroup          391461 2017-05-20 17:29 /piggybank-0.15.0.jar  
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 16:56 /prjct  
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 17:04 /project  
drwxr-xr-x  - nitesh supergroup          717414 2017-05-20 16:34 /project1  
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 17:12 /projectininput  
drwxrwxrwx  - nitesh supergroup          0 2017-05-13 18:34 /user  
drwxr-xr-x  - nitesh supergroup          0 2017-04-05 00:04 /usr  
drwxr-xr-x  - nitesh supergroup          0 2017-05-20 23:59 /xml_to_csv  
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls //BPL_ANSWER_District  
Found 2 items  
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 18:00 /BPL_ANSWER_District/_SUCCESS  
-rw-r--r--  4 nitesh supergroup          690 2017-05-21 18:00 /BPL_ANSWER_District/part-m-00000  
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls //BPL_ANSWER  
ls: No FileSystem for scheme: null  
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /BPL_ANSWER  
Found 2 items  
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 17:42 /BPL_ANSWER/_SUCCESS  
-rw-r--r--  4 nitesh supergroup          1886 2017-05-21 17:42 /BPL_ANSWER/part-m-00000  
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_percent  
Found 2 items  
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 21:48 /Achieved_80_percent/_SUCCESS  
-rw-r--r--  4 nitesh supergroup          10382 2017-05-21 21:48 /Achieved_80_percent/part-m-00000  
nitesh@nitesh-Lenovo-Z50-70:~$ hadoop fs -ls /Achieved_80_district_percent  
Found 2 items  
-rw-r--r--  4 nitesh supergroup          0 2017-05-21 21:56 /Achieved_80_district_percent/_SUCCESS  
-rw-r--r--  4 nitesh supergroup          3356 2017-05-21 21:56 /Achieved_80_district_percent/part-m-00000
```

```
nitesh@nitesh-Lenovo-Z50-70:~ mysql> show databases;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| BPL_ANS |  
| hivedb |  
| metastore |  
| mysql |  
| performance_schema |  
| sys |  
+-----+  
7 rows in set (0.00 sec)  
  
mysql> CREATE DATABASE BPL_80_ANS;  
Query OK, 1 row affected (0.03 sec)  
  
mysql> USE BPL_80_ANS;  
Database changed  
mysql> CREATE TABLE BPL_80_ALL (STATE VARCHAR(100), DISTRICT VARCHAR(100), PERCENT FLOAT(10));  
Query OK, 0 rows affected (0.93 sec)  
  
mysql> CREATE TABLE BPL_80_DIST (DISTRICT VARCHAR(100));  
Query OK, 0 rows affected (0.50 sec)  
  
mysql> SHOW TABLES;  
+-----+  
| Tables_in_BPL_80_ANS |  
+-----+  
| BPL_80_ALL |  
| BPL_80_DIST |  
+-----+  
2 rows in set (0.01 sec)  
  
mysql> 
```

```

nitesh@nitesh-Lenovo-Z50-70: ~
+-----+
7 rows in set (0.00 sec)

mysql> CREATE DATABASE BPL_80_ANS;
Query OK, 1 row affected (0.03 sec)

mysql> USE BPL_80_ANS;
Database changed
mysql> CREATE TABLE BPL_80_ALL (STATE VARCHAR(100), DISTRICT VARCHAR(100), PERCENT FLOAT(10));
Query OK, 0 rows affected (0.93 sec)

mysql> CREATE TABLE BPL_80_DIST (DISTRICT VARCHAR(100));
Query OK, 0 rows affected (0.50 sec)

mysql> SHOW TABLES;
+-----+
| Tables_in_BPL_80_ANS |
+-----+
| BPL_80_ALL |
| BPL_80_DIST |
+-----+
2 rows in set (0.01 sec)

mysql> DESCRIBE BPL_80_ALL
-> ;
+-----+-----+-----+-----+-----+
| Field | Type   | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| STATE | varchar(100) | YES |   | NULL    |       |
| DISTRICT | varchar(100) | YES |   | NULL    |       |
| PERCENT | float      | YES |   | NULL    |       |
+-----+-----+-----+-----+-----+
3 rows in set (0.15 sec)

mysql> DESCRIBE BPL_80_DIST;
+-----+-----+-----+-----+-----+
| Field | Type   | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| DISTRICT | varchar(100) | YES |   | NULL    |       |
+-----+-----+-----+-----+-----+
1 row in set (0.04 sec)

mysql> 

```

```

nitesh@nitesh-Lenovo-Z50-70: ~
nitesh@nitesh-Lenovo-Z50-70:~$ sqoop export -m 1 --connect jdbc:mysql://localhost:3306/BPL_80_ANS --username root --password nss8055 --table BPL_80_DIST --export-dir /Achieved_80_district_percent/part-m-000000 -driver com.mysql.jdbc.Driver
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop2-0.4-alpha/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop2-0.4-alpha/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/nitesh/sqoop-1.4.6-bin_hadoop2-0.4-alpha/../zookeeper does not exist! Zookeeper imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-05-22 00:03:22,048 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-05-22 00:03:22,060 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-05-22 00:03:22,127 WARN [main] sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate connection manager is not being set (via --connection-manager). Sqoop is going to fall back to org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.
2017-05-22 00:03:22,138 INFO [main] manager.SqlManager: Using default fetchsize of 1000
2017-05-22 00:03:22,139 INFO [main] tool.CodeGenTool: Beginning code generation
2017-05-22 00:03:22,530 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_80_DIST AS t WHERE 1=0
2017-05-22 00:03:22,553 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_80_DIST AS t WHERE 1=0
2017-05-22 00:03:22,651 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop/hadoop-2.7.0
Note: /tmp/sqoop-nitesh/compile/2ebd8e9dcf761b31619b3b253d70203/BPL_80_DIST.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-05-22 00:03:24,914 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-nitesh/compile/2ebd8e9dcf761b31619b3b253d70203/BPL_80_DIST.jar
2017-05-22 00:03:24,934 INFO [main] mapreduce.ExportJobBase: Beginning export of BPL_80_DIST
2017-05-22 00:03:24,935 INFO [main] Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-05-22 00:03:25,172 INFO [main] Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2017-05-22 00:03:25,898 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM BPL_80_DIST AS t WHERE 1=0
2017-05-22 00:03:25,903 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-22 00:03:25,907 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-22 00:03:25,931 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-22 00:03:26,036 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-22 00:03:43,198 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:03:43,199 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:03:44,172 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.f

```

```
nitesh@nitesh-Lenovo-Z50-70: ~
2017-05-22 00:03:25,903 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-22 00:03:25,907 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-22 00:03:25,931 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-22 00:03:26,036 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-22 00:03:43,198 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:03:43,199 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-22 00:03:44,172 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.in
putdir
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.ma
p.speculative
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputform
at.class
2017-05-22 00:03:44,485 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputfo
rmat.class
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.v
alue.class
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key
.class
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.
files
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-05-22 00:03:44,486 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.fi
les.filesizes
2017-05-22 00:03:44,716 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495350191822_0062
2017-05-22 00:03:47,079 INFO [main] impl.YarnClientImpl: Submitted application application_1495350191822_0062 to ResourceManager at /0.0.0.0:8
032
2017-05-22 00:03:47,350 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495350
191822_0062/
2017-05-22 00:03:47,351 INFO [main] mapreduce.Job: Running job: job_1495350191822_0062
2017-05-22 00:03:57,700 INFO [main] mapreduce.Job: Job job_1495350191822_0062 running in uber mode : false
2017-05-22 00:03:57,725 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-05-22 00:04:13,064 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-05-22 00:04:22,127 INFO [main] mapreduce.Job: Job job_1495350191822_0062 completed successfully
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
mysql> SELECT * FROM BPL_80_DIST;
+-----+
| DISTRICT |
+-----+
| ANANTAPUR |
| CHITTOOR |
| CUDDAPAH |
| EAST GODAVARI |
| KARIMNAGAR |
| KHAMMAM |
| KRISHNA |
| KURNool |
| MEDAK |
| NALGONDA |
| NIZAMABAD |
| RANGAREDDI |
| WARANGAL |
| WEST GODAVARI |
| DIBANG VALLEY |
| LOHIT |
| TIRAP |
| BAGSHA |
| CACHAR |
| DIBRUGARH |
| GOALPARA |
| GOLAghAT |
| HAILAKANDI |
| JORHAT |
| KAMrup |
| KARIMGANj |
| KOKRAJHAR |
| LAKHIMPUR |
| MARIGAON |
| NAGAON |
| SIBSAGAR |
| SONiTPUR |
| TINSUKIA |
| BEGUSARAI |
| MADHUBANI |
| MUZAFFARPUR |
| SAHARSA |
| VAISHALI |
| DHAMTARI |
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
| KURUKSHETRA
| MAHENDRAGARH
| MEWAT
| PANCHKULA
| PANIPAT
| REWARI
| ROHTAK
| SIRSA
| SONIPAT
| YAMUNANAGAR
| BTLASPUR
| CHAMBA
| HAMIRPUR
| KANGRA
| KINNAUR
| KULLU
| LAHAUL & SPITI
| MANDI
| SHIMLA
| SIRMAUR
| SOLAN
| UNA
| ANANTNAG
| LEH (LADAKH)
| DEOGHAR
| DUMKA
| LATEHAR
| LOHARDAGA
| PAKUR
| PURBT SINGHBHUM
| BAGALKOT
| BANGALORE RURAL
| CHICKMAGALUR
| CHITRADURGA
| DHARMAD
| GADAG
| HASSAN
| KODAGU
| KOLAR
| KOPPAL
| MANDYA
| MANGALORE(DAKSHINA KANNADA)
| RAMANAGARA
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
| RAISEN
| RAJGARH
| RATLAM
| REWA
| SEHORE
| SEONI
| SHAHDOL
| SHAJAPUR
| SHEOPUR
| SINGRAULI
| UJJAIN
| UMARIA
| VIDISHA
| AHMEDNAGAR
| BHANDARA
| DHULE
| GADCHIROLI
| GONDIA
| HINGOLI
| JALNA
| KOLHAPUR
| NAGPUR
| OSMANABAD
| PARBHANI
| PUNE
| RATNAGIRI
| SANGLI
| SATARA
| SINDHUDURG
| THANE
| WARDHA
| BISHNUPUR
| IMPHAL EAST
| TAMENGLONG
| RI BHOI
| SOUTH GARO HILLS
| WEST GARO HILLS
| CHAMPHAI
| KOLASIB
| LAWNGTIAI
| LUNGLEI
| MAMIT
| SAIHA
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
+----+
| FAIZABAD
| FARRUKHABAD
| FATEHPUR
| FIROZABAD
| GAUTAM BUDDHA NAGAR
| GHAZIABAD
| GHAZIPUR
| GONDA
| GORAKHPUR
| HAMIRPUR
| HARDOI
| JALAUN
| JAUNPUR
| JHANSI
| JYOTIBA PHULE NAGAR
| KANNAUJ
| KANPUR DEHAT
| KANPUR NAGAR
| KAUSHAMBHI
| KUSHINAGAR
| LAKHIMPUR KHERI
| LALITPUR
| LUCKNOW
| MAHAMAYA NAGAR(HATHRAS)
| MAHARAJGANJ
| MAHOBA
| MAINPURI
| MATHURA
| MAU
| MEERUT
| MIRZAPUR
| MORADABAD
| MUZAFFARNAGAR
| PILIBHIT
| PRATAPGARH
| RAE BARELI
| RAMPUR
| SAHARANPUR
| SANT RAVIDAS NAGAR( BHADOHI)
| SHAHZAHANPUR
| SHRAVASTI
| SIDDHARTHNAGAR
| SITAPUR
+----+
```

```
nitesh@nitesh-Lenovo-Z50-70: ~
+----+
| MAU
| MEERUT
| MIRZAPUR
| MORADABAD
| MUZAFFARNAGAR
| PILIBHIT
| PRATAPGARH
| RAE BARELI
| RAMPUR
| SAHARANPUR
| SANT RAVIDAS NAGAR( BHADOHI)
| SHAHZAHANPUR
| SHRAVASTI
| SIDDHARTHNAGAR
| SITAPUR
| SONGHADRA
| SULTANPUR
| UNNAO
| VARANASI
| BAGESHWAR
| CHAMOLI
| DEHRADUN
| HARIWAR
| NAINITAL
| PITHORAGARH
| RUDRAPRAYAG
| TEHRI GARHWAL
| UDHAM SINGH NAGAR
| UTTARKASHI
| BARDHAMAN
| DAKSHIN DINAJPUR
| HOOGHLY
| HOWRAH
| JALPAIGURI
| MIDNAPUR EAST
| MIDNAPUR WEST
| NADIA
| NORTH 24 PARAGANAS
| SOUTH 24 PARAGANAS
+----+
349 rows in set (0.00 sec)

mysql> |
```

Shell Scripting

What is "the shell"?

Simply put, the shell is a program that takes your commands from the keyboard and gives them to the operating system to perform. In the old days, it was the only user interface available on a Unix computer. Nowadays, we have graphical user interfaces (GUIs) in addition to command line interfaces (CLIs) such as the shell.

On most Linux systems a program called bash (which stands for Bourne Again SHell, an enhanced version of the original Bourne shell program, sh, written by Steve Bourne) acts as the shell program. There are several additional shell programs available on a typical Linux system. These include: ksh, tcsh and zsh.

How to write shell script?

Following steps are required to write shell script:

(1) Use any editor like vi or mcedit to write shell script.

(2) After writing shell script set execute permission for your script as follows

syntax:

chmod permission your-script-name

Examples:

\$ chmod +x your-script-name

\$ chmod 755 your-script-name

This will set read write execute(7) permission for owner, for group and other permission is read and execute only(5).

(3) Execute your script as

syntax:

bash your-script-name

sh your-script-name

./your-script-name

Here we have written the shell script for both the questions in the single script

The flow of the script is as follows:

- First we have started all the daemons
- Secondly we have deleted the directories if they are available and created it again so the previous data is cleared or deleted.
- Then the timeout function is used to kill the flume command to stop it after 30s of execution.
- Then the pig script is used to convert the xml_to_csv. This script is stored in the file with extension .pig
- The pig scripts are passed as pig 'script _name'
- Then we used mysql to create the table, for that we used the syntax

```
Mysql -u root -ppassword <<EOF
```

```
Commands
```

```
EOF
```

This will execute the commands in the mysql. The <<EOF tells that it should execute the commands in EOF till EOF appears again.

- Then we used the sqoop command to export the generated output from hdfs to mysql
- Lastly we displayed the exported result by querying the tables.

A. *The pig script to convert the xml_to_csv is as follows:*

```
A = load 'hdfs:/shell/project/StatewiseDistrictwisePhysicalProgress.xml' using  
org.apache.pig.piggybank.storage.XMLLoader('row') as (x:chararray);
```

```
B = foreach A generate REPLACE(x,'[\n]', '') as x;
```

```
C = foreach B generate  
REGEX_EXTRACT_ALL(x,.*(?:<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>)([^<]*).*)';
```

```
D = FOREACH C GENERATE FLATTEN ((\$0));
```

```
STORE D INTO 'hdfs://localhost:9000/shell/pig_conversions/xml_to_csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage();
```

B. The pig script for question 1 is as follows:

```
register /home/nitesh/Desktop/filter_udf.jar
```

```
register /home/nitesh/Desktop/percent.jar
```

```
Input_data = load 'hdfs://shell/pig_conversions/xml_to_csv/part-m-00000' USING PigStorage(',')  
as(State_Name:chararray,District_Name:chararray ,Project_Objectives_IHHL_BPL:int ,  
Project_Objectives_IHHL_APL:int , Project_Objectives_IHHL_TOTAL:int  
,Project_Objectives_SCW:int ,Project_Objectives_School_Toilets:int,  
Project_Objectives_Anganwadi_Toilets:int, Project_Objectives_RSM:int  
,Project_Objectives_PC:int ,Project_Performance_IHHL_BPL:int ,  
Project_Performance_IHHL_APL:int ,Project_Performance_IHHL_TOTAL:int,  
Project_Performance_SCW:int , Project_Performance_School_Toilets:int  
,Project_Performance_Anganwadi_Toilets:int , Project_Performance_RSM:int  
,Project_Performance_PC:int);
```

```
Filtered_data = filter Input_data BY $2==$10;
```

```
Districts = foreach Filtered_data generate $0,$1,Percent($2,$10);
```

```
STORE Districts INTO 'hdfs://shell/BPL_ANSWER' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage();
```

The flow and explanation of commands is same as the Problem statement 1

C. The pig script for question 2 is as follows:

```
register /home/nitesh/Desktop/filter_udf.jar
```

```
register /home/nitesh/Desktop/percent.jar
```

```
Input_data = load 'hdfs://shell/pig_conversions/xml_to_csv/part-m-00000' USING PigStorage(',')  
as(State_Name:chararray,District_Name:chararray ,Project_Objectives_IHHL_BPL:int ,  
Project_Objectives_IHHL_APL:int , Project_Objectives_IHHL_TOTAL:int  
,Project_Objectives_SCW:int ,Project_Objectives_School_Toilets:int,  
Project_Objectives_Anganwadi_Toilets:int, Project_Objectives_RSM:int  
,Project_Objectives_PC:int ,Project_Performance_IHHL_BPL:int ,  
Project_Performance_IHHL_APL:int ,Project_Performance_IHHL_TOTAL:int,  
Project_Performance_SCW:int , Project_Performance_School_Toilets:int
```

```
,Project_Performance_Anganwadi_Toilets:int , Project_Performance_RSM:int  
,Project_Performance_PC:int);
```

```
Filtered_data = filter Input_data BY Filter_udf($2,$10);
```

```
Districts = foreach Filtered_data generate $0,$1,Percent($2,$10);
```

```
STORE Districts INTO 'hdfs:/shell/BPL_ANSWER_80' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage();
```

D. The shell script for major project is as follows:

```
#####
#####
```

```
#!/bin/bash
```

```
start-all.sh
```

```
hadoop fs -rm -r -skipTrash /shell
```

```
#hadoop fs -rm -r -skipTrash /shell/project
```

```
hadoop fs -mkdir /shell
```

```
hadoop fs -mkdir /shell/pig_conversions/xml_to_csv/
```

```
hadoop fs -mkdir /shell/project/
```

```
timeout 30s flume-ng agent -n agent1 -c conf -f /home/nitesh/apache-flume-1.6.0-  
bin/conf/filecopy.conf
```

```
mr-jobhistory-daemon.sh start historyserver
```

```
pig xml_to_csv.pig
```

```
pig que1.pig
```

```
pig que2.pig
```

```
mysql -u root -p nsss8055 <<EOF
```

```
use shell;
```

```
DROP TABLE IF EXISTS shell_que1;
```

```
CREATE TABLE shell_que1 (state VARCHAR(100),district VARCHAR(100),percent  
FLOAT(10));
```

```
DROP TABLE IF EXISTS shell_que2;
```

```
CREATE TABLE shell_que2 (state VARCHAR(100),district VARCHAR(100),percent  
FLOAT(10));
```

```
EOF
```

```
sqoop export -m 1 --connect jdbc:mysql://localhost:3306/shell --username root --password
nss8055 --table shell_que1 --export-dir /shell/BPL_ANSWER --driver com.mysql.jdbc.Driver

sqoop export -m 1 --connect jdbc:mysql://localhost:3306/shell --username root --password
nss8055 --table shell_que2 --export-dir /shell/BPL_ANSWER_80 --driver
com.mysql.jdbc.Driver

mysql -u root -pnss8055 <<EOF

#here the shell database is already created

use shell;

SELECT * FROM shell_que1;

EOF

echo "OUTPUT OF THE SECOND SCRIPT"

mysql -u root -pnss8055 <<EOF

use shell;

SELECT * FROM shell_que2;

EOF

#####
#####
```

The above script is stored as project2.sh, the extension .sh is not important but we use in general practice.

Below are the screen shots of the execution of the shell script

The command we used for the execution is

./project.sh(shell script name)

```
nitesh@nitesh-Lenovo-Z50-70: ~$ ./Music
[SLF4J]: Class path contains multiple SLF4J bindings.
[SLF4J]: Found binding in [jar:file:/home/nitesh/apache-flume-1.6.0-bin/lib/slf4j-log4j12-1.6.1.jar!/:org/slf4j/impl/StaticLoggerBinder.class]
[SLF4J]: Found binding in [jar:file:/home/nitesh/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4.1.jar!/:org/slf4j/impl/StaticLoggerBinder.class]
[SLF4J]: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
17/05/26 08:37:03 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting
17/05/26 08:37:03 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:/home/nitesh/apache-flume-1.6.0-bin/conf/filecopy.conf
17/05/26 08:37:03 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/05/26 08:37:03 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/05/26 08:37:03 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/05/26 08:37:03 INFO conf.FlumeConfiguration: Added sinks: hdfsdest Agent: agent1
17/05/26 08:37:03 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent1]
17/05/26 08:37:03 INFO node.AbstractConfigurationProvider: Creating channels
17/05/26 08:37:03 INFO channel.DefaultChannelFactory: Creating instance of channel mychannel type memory
17/05/26 08:37:03 INFO node.AbstractConfigurationProvider: Created channel mychannel
17/05/26 08:37:03 INFO source.DefaultSourceFactory: Creating instance of source mysrc, type exec
17/05/26 08:37:03 INFO sink.DefaultSinkFactory: Creating instance of sink: hdfsdest, type: hdfs
17/05/26 08:37:03 INFO node.AbstractConfigurationProvider: Channel mychannel connected to [mysrc, hdfsdest]
17/05/26 08:37:03 INFO node.Application: Starting new configuration:{ sourceRunners:{mysrc=EventDrivenSourceRunner: { source:org.apache.flume.source.ExecSource{name:mysrc,state:IDLE} } } sinkRunners:{hdfsdest=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@2c91f6a count:rGroup:{name:null counters:{} } } } channels:{mychannel=org.apache.flume.channel.MemoryChannel{name: mychannel}} }
17/05/26 08:37:03 INFO node.Application: Starting Channel mychannel
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: CHANNEL, name: mychannel: Successfully registered new MBean.
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Component type: CHANNEL, name: mychannel started
17/05/26 08:37:03 INFO node.Application: Starting Sink hdfsdest
17/05/26 08:37:03 INFO node.Application: Starting Source mysrc
17/05/26 08:37:03 INFO source.ExecSource: Exec source starting with command:hadoop dfs -put /home/nitesh/Downloads/StatewiseDistrictwisePhysicalProgress.xml /shell/project/
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: SINK, name: hdfsdest: Successfully registered new MBean.
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Component type: SINK, name: hdfsdest started
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: SOURCE, name: mysrc: Successfully registered new MBean.
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Component type: SOURCE, name: mysrc started
17/05/26 08:37:06 INFO source.ExecSource: Command [hadoop dfs -put /home/nitesh/Downloads/StatewiseDistrictwisePhysicalProgress.xml /shell/project/] exited with 0
17/05/26 08:37:28 INFO lifecycle.LifecycleSupervisor: Stopping lifecycle supervisor 11
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Component type: SINK, name: hdfsdest started
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: SOURCE, name: mysrc: Successfully registered new MBean.
17/05/26 08:37:03 INFO instrumentation.MonitoringCounterGroup: Component type: SOURCE, name: mysrc started
17/05/26 08:37:06 INFO source.ExecSource: Command [hadoop dfs -put /home/nitesh/downloads/StatewiseDistrictwisePhysicalProgress.xml /shell/project/] exited with 0
17/05/26 08:37:28 INFO lifecycle.LifecycleSupervisor: Stopping lifecycle supervisor 11
17/05/26 08:37:28 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider stopping
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Component type: SINK, name: hdfsdest stopped
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.start.time == 1495768023763
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.stop.time == 1495768048195
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.batch.complete == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.batch.empty == 5
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.batch.underflow == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.connection.closed.count == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.connection.creation.count == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.connection.failed.count == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.event.drain.attempt == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SINK, name: hdfsdest. sink.event.drain.success == 0
17/05/26 08:37:28 INFO source.ExecSource: Stopping exec source with command:hadoop dfs -put /home/nitesh/downloads/StatewiseDistrictwisePhysicalProgress.xml /shell/project/
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Component type: SOURCE, name: mysrc stopped
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. source.start.time == 1495768023767
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. source.stop.time == 1495768048198
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. src.append-batch.accepted == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. src.append-batch.received == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. src.append.accepted == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. src.append.received == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. src.events.received == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: SOURCE, name: mysrc. src.open-connection.count == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Component type: CHANNEL, name: mychannel stopped
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.start.time == 1495768023760
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.stop.time == 1495768048200
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.capacity == 100
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.current.size == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.event.put.attempt == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.event.put.success == 0
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.event.put.success == 0
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.event.take.attempt == 5
17/05/26 08:37:28 INFO instrumentation.MonitoringCounterGroup: Shutdown Metric for type: CHANNEL, name: mychannel. channel.event.take.success == 0
historyserver running as process 4458. Stop it first.
2017-05-26 08:37:30,058 INFO [main] pig.ExecTypeProvider: Trying ExecType : LOCAL
2017-05-26 08:37:30,058 INFO [main] pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2017-05-26 08:37:30,059 INFO [main] pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2017-05-26 08:37:30,132 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2015, 11:44:35
2017-05-26 08:37:30,133 [main] INFO org.apache.pig.Main - Logging error messages to: /home/nitesh/Music/pig_1495768050131.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j.impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j.impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-05-26 08:37:30,539 [main] INFO org.apache.pig.impl.Utils - Default bootup file /home/nitesh/.pigbootup not found
2017-05-26 08:37:30,665 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-26 08:37:30,665 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:37:30,665 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2017-05-26 08:37:30,669 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.used.genericoptionsparser is deprecated. Instead, use mapreduce.client.genericoptionsparser.used
2017-05-26 08:37:31,018 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:9001
2017-05-26 08:37:31,554 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.persist.jobstatus.hours is deprecated. Instead, use mapreduce.jobtracker.persist.jobstatus.hours
2017-05-26 08:37:31,555 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.heartbeats.in.second is deprecated. Instead, use mapreduce.jobtracker.heartbeats.in.second
2017-05-26 08:37:31,555 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - jobclient.completion.poll.interval is deprecated. Instead, use mapreduce.client.completion.pollinterval
2017-05-26 08:37:31,555 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.tasktracker.tasks.sleeptime-before-sigkill is deprecated. Instead, use mapreduce.tasktracker.tasks.sleeptimebeforesigkill
2017-05-26 08:37:31,555 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address
2017-05-26 08:37:31,555 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.replication.considerLoad is deprecated. Instead, use dfs.namenode.replication.considerLoad
2017-05-26 08:37:31,555 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.block.size is deprecated. Instead, use dfs.blocksize
```

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
2017-05-26 08:37:33,329 [JobControl] WARN org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2017-05-26 08:37:33,402 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-26 08:37:33,402 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-05-26 08:37:33,435 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2017-05-26 08:37:33,617 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-26 08:37:33,643 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-26 08:37:33,645 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-26 08:37:33,646 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-26 08:37:33,646 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-26 08:37:33,773 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495765181217_0019
2017-05-26 08:37:33,948 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.
2017-05-26 08:37:34,062 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495765181217_0019 to ResourceManager at /0.0.0.0:8032
2017-05-26 08:37:34,119 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495765181217_0019/
2017-05-26 08:37:34,120 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1495765181217_0019
2017-05-26 08:37:34,120 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases A,B,D
2017-05-26 08:37:34,120 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: A[1,4],B[-1,-1],D[-1,-1] C: R
2017-05-26 08:37:34,134 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2017-05-26 08:37:34,135 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495765181217_0019]
2017-05-26 08:37:46,459 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2017-05-26 08:37:46,459 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495765181217_0019]
2017-05-26 08:37:49,470 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:37:49,526 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2017-05-26 08:37:49,526 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Redirecting to job history server
2017-05-26 08:37:49,813 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:37:49,818 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2017-05-26 08:37:49,826 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Redirecting to job history server
2017-05-26 08:37:49,928 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2017-05-26 08:37:49,972 [main] WARN org.apache.pig.tools.pigstats.mapreduce.MRJobStats - Unable to get job counters

```

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
at org.apache.pig.backend.hadoop.executionengine.shims.HadoopShims.getCounters(HadoopShims.java:130)
... 21 more
2017-05-26 08:37:49,976 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2017-05-26 08:37:49,979 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-26 08:37:32 2017-05-26 08:37:49 UNKNOWN
Success!
Job Stats (time in seconds):
Jobid Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
educeTime Alias Feature Outputs
job_1495765181217_0019 1 0 4 4 4 0 0 0 0 A,B,D MAP_ONLY hdfs://localhost:9000/shell/pig_conversions/xml_to_csv,
Input(s):
Successfully read 0 records from: "hdfs://shell/project/StatewiseDistrictwisePhysicalProgress.xml"
Output(s):
Successfully stored 0 records in: "hdfs://localhost:9000/shell/pig_conversions/xml_to_csv"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495765181217_0019

2017-05-26 08:37:49,981 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:37:49,984 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2017-05-26 08:37:49,984 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Redirecting to job history server
2017-05-26 08:37:50,073 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:37:50,077 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2017-05-26 08:37:50,116 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Redirecting to job history server
2017-05-26 08:37:50,119 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2017-05-26 08:37:50,168 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to

```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
e mapreduce.job.counters.max
2017-05-26 08:37:53,076 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-26 08:37:53,076 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:37:53,249 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-26 08:37:53,250 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-26 08:37:53,250 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-26 08:37:53,250 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:37:53,292 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2017-05-26 08:37:53,333 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2017-05-26 08:37:53,425 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:37:53,425 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-26 08:37:53,429 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-05-26 08:37:53,467 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES_ENABLED=AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstrParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]
2017-05-26 08:37:53,511 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for Input_data: $3, $4, $5, $6, $7, $8, $9, $11, $12, $13, $14, $15, $16, $17
2017-05-26 08:37:53,612 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2017-05-26 08:37:53,643 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2017-05-26 08:37:53,643 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2017-05-26 08:37:53,680 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:37:53,789 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:37:53,866 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2017-05-26 08:37:53,871 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markрейset.buffer.percent is not set, set to default 0.3
2017-05-26 08:37:53,873 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
2017-05-26 08:37:54,156 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/nitesh/Desktop/filter_udf.jar to DistributedCache through /tmp/temp301689976/tmp-1468337386/filter_udf.jar
2017-05-26 08:37:54,242 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/nite
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
duce.job.name
2017-05-26 08:37:54,837 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-26 08:37:54,837 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2017-05-26 08:37:54,837 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputformat.class
2017-05-26 08:37:54,837 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputformat.class
2017-05-26 08:37:54,837 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-26 08:37:54,838 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir
2017-05-26 08:37:54,838 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-26 08:37:54,838 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.files
2017-05-26 08:37:54,838 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-26 08:37:54,838 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:37:55,107 [JobControl] WARN org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set. User classes may not be found. See Job or Job#setJar(String).
2017-05-26 08:37:55,173 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-26 08:37:55,173 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-05-26 08:37:55,196 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2017-05-26 08:37:55,462 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-26 08:37:55,487 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-26 08:37:55,488 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-26 08:37:55,488 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-26 08:37:55,488 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-26 08:37:56,072 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495765181217_0020
2017-05-26 08:37:56,213 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.
2017-05-26 08:37:56,371 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495765181217_0020 to ResourceManager at /0.0.0.0:8032
2017-05-26 08:37:56,415 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495765181217_0020/
```

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
2017-05-26 08:38:12,265 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2017-05-26 08:38:12,268 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-26 08:37:53 2017-05-26 08:38:12 FILTER
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
job_1495765181217_0020 1 0 3 3 3 0 0 0 0 Districts,Filtered_data,Input_data M
AP_ONLY hdfs://shell/BPL_ANSWER,
Input(s):
Successfully read 0 records from: "hdfs://shell/pig_conversions/xml_to_csv/part-m-00000"
Output(s):
Successfully stored 0 records in: "hdfs://shell/BPL_ANSWER"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495765181217_0020
2017-05-26 08:38:12,269 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:12,272 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-26 08:38:12,320 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:12,324 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-26 08:38:12,367 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:12,371 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-26 08:38:12,411 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-05-26 08:38:12,411 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

```

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
job_1495765181217_0020
2017-05-26 08:38:12,269 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:12,272 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-26 08:38:12,320 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:12,324 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-26 08:38:12,367 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:12,371 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEEDED. Redirecting to job history server
2017-05-26 08:38:12,411 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-05-26 08:38:12,411 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-26 08:38:12,431 [main] INFO org.apache.pig.Main - Pig script completed in 20 seconds and 929 milliseconds (20929 ms)
2017-05-26 08:38:13,557 [INFO] [main] pig.ExecTypeProvider: Trying ExecType : LOCAL
2017-05-26 08:38:13,558 [INFO] [main] pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2017-05-26 08:38:13,558 [INFO] [main] pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2017-05-26 08:38:13,596 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2015, 11:44:35
2017-05-26 08:38:13,597 [main] INFO org.apache.pig.Main - Logging error messages to: /home/nitesh/Music/pig_1495768093595.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dir/hbase-0.98.4-hadoop2/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j.impl.StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-05-26 08:38:13,971 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/nitesh/.pigbootup not found
2017-05-26 08:38:14,088 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-26 08:38:14,088 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:38:14,088 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2017-05-26 08:38:14,092 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.used.genericoptionsparser is deprecated. Instead, use mapreduce.client.genericoptionsparser.used
2017-05-26 08:38:14,427 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:9001
2017-05-26 08:38:14,501 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.persist.jobstatus.hours is deprecated. Instead, use mapreduce.jobtracker.persist.jobstatus.hours
2017-05-26 08:38:14,501 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.heartbeats.in.second is deprecated. Instead, use mapreduce.jobtracker.heartbeats.in.second
2017-05-26 08:38:14,501 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - jobclient.completion.poll.interval is deprecated. Inste

```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
2017-05-26 08:38:15,283 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2017-05-26 08:38:15,318 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2017-05-26 08:38:15,403 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:38:15,403 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2017-05-26 08:38:15,406 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-05-26 08:38:15,440 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]]
2017-05-26 08:38:15,484 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for Input_data: $3, $4, $5, $6, $7, $8, $9, $11, $12, $13, $14, $15, $16, $17
2017-05-26 08:38:15,602 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 10
0 optimistic? false
2017-05-26 08:38:15,636 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2017-05-26 08:38:15,636 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2017-05-26 08:38:15,671 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-26 08:38:15,747 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:15,810 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptEngine - Pig script settings are added to the job
2017-05-26 08:38:15,814 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.marktreduce.buffer.percent is not set, set to default 0.3
2017-05-26 08:38:15,816 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
2017-05-26 08:38:16,033 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/nitesh/Desktop/filter_udf.jar to DistributedCache through /tmp/temp493358294/tmp171267469/filter_udf.jar
2017-05-26 08:38:16,489 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/home/nitesh/Desktop/percent.jar to DistributedCache through /tmp/temp493358294/tmp-508058949/percent.jar
2017-05-26 08:38:16,664 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig/lib/piggybank.jar to DistributedCache through /tmp/temp493358294/tmp653975393/piggybank.jar
2017-05-26 08:38:16,753 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig/pig-0.15.0-core-h2.jar to DistributedCache through /tmp/temp493358294/tmp-1092567902/pig-0.15.0-core-h2.jar
2017-05-26 08:38:16,819 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp493358294/tmp-670505659/automaton-1.11-8.jar
2017-05-26 08:38:16,854 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp493358294/tmp-1711151938/antlr-runtime-3.4.jar
2017-05-26 08:38:16,897 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig/lib/joda-time-2.5.jar to DistributedCache through /tmp/temp493358294/tmp959770890/joda-time-2.5.jar
2017-05-26 08:38:16,936 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
Job or Job#setJar(String).
2017-05-26 08:38:17,346 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-26 08:38:17,346 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-05-26 08:38:17,386 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2017-05-26 08:38:17,596 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2017-05-26 08:38:17,623 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-26 08:38:17,623 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-26 08:38:17,624 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-26 08:38:17,624 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-26 08:38:17,829 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1495765181217_0021
2017-05-26 08:38:17,988 [JobControl] INFO org.apache.hadoop.mapred.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.
2017-05-26 08:38:18,086 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1495765181217_0021 to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:18,132 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495765181217_0021/
2017-05-26 08:38:18,132 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1495765181217_0021
2017-05-26 08:38:18,132 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Districts_Filtered_data_input_data
2017-05-26 08:38:18,133 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: Input data[6,13],Input data[-1,-1],Filtered data[8,16],Districts[10,12] C: R:
2017-05-26 08:38:18,143 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2017-05-26 08:38:18,143 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495765181217_0021]
2017-05-26 08:38:27,458 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2017-05-26 08:38:27,458 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_1495765181217_0021]
2017-05-26 08:38:33,476 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,487 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-05-26 08:38:33,638 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,641 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-05-26 08:38:33,704 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,707 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-05-26 08:38:33,750 [main] WARN org.apache.pig.tools.pigstats.mapreduce.MRJobStats - Unable to get job counters
```

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
2017-05-26 08:38:33,754 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2017-05-26 08:38:33,757 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.2.0 0.15.0 nitesh 2017-05-26 08:38:15 2017-05-26 08:38:33 FILTER
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianR
job_1495765181217_0021 1 0 3 3 3 0 0 0 0 Districts,Filtered_data,Input_data M
AP_ONLY hdfs:/shell/BPL_ANSWER_80,
Input(s):
Successfully read 0 records from: "hdfs:/shell/pig_conversions/xml_to_csv/part-m-00000"
Output(s):
Successfully stored 0 records in: "hdfs:/shell/BPL_ANSWER_80"
Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1495765181217_0021

2017-05-26 08:38:33,759 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,763 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEDED. Redirecting to job history server
2017-05-26 08:38:33,825 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,830 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEDED. Redirecting to job history server
2017-05-26 08:38:33,873 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,878 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEDED. Redirecting to job history server
2017-05-26 08:38:33,927 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-05-26 08:38:33,927 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

```

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
2017-05-26 08:38:33,759 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,763 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEDED. Redirecting to job history server
2017-05-26 08:38:33,825 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,830 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEDED. Redirecting to job history server
2017-05-26 08:38:33,873 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:33,878 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SU
CCEDED. Redirecting to job history server
2017-05-26 08:38:33,927 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-05-26 08:38:33,927 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-26 08:38:33,954 [main] INFO org.apache.pig.Main - Pig script completed in 20 seconds and 415 milliseconds (20415 ms)
mysql: [Warning] Using a password on the command line interface can be insecure.
Warning: /home/nitesh/sqoop-1.4.6.bn_hadoop-2.0.4-alpha/..//hcatalog does not exist! HCatalog jobs will fail.
Please set SHCAT_HOME to the root of your HCatalog installation.
Warning: /home/nitesh/sqoop-1.4.6.bn_hadoop-2.0.4-alpha/..//accumulo does not exist! Accumulo imports will fail.
Please set SACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/nitesh/sqoop-1.4.6.bn_hadoop-2.0.4-alpha/..//zookeeper does not exist! Accumulo imports will fail.
Please set ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-05-26 08:38:36,975 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-05-26 08:38:36,990 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-05-26 08:38:37,055 WARN [main] sqoop.ConnFactory: Parameter --driver is set to an explicit driver however appropriate connection manager is not being set (via --connection-manager). Sqoop is going to fall back to org.apache.sqoop.manager.GenericJdbcManager. Please specify explicitly which connection manager should be used next time.
2017-05-26 08:38:37,067 INFO [main] manager.SqlManager: Using default fetchSize of 1000
2017-05-26 08:38:37,068 INFO [main] tool.CodeGenTool: Beginning code generation
2017-05-26 08:38:37,396 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM shell_que1 AS t WHERE 1=0
2017-05-26 08:38:37,400 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM shell_que1 AS t WHERE 1=0
2017-05-26 08:38:37,432 INFO [main] orn.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop/hadoop-2.7.0
Note: /tmp/sqoop-nitesh/compile/2c9c65bdb50014b105f54623546931fa/shell_que1.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-05-26 08:38:38,871 INFO [main] orn.CompilationManager: Writing jar file: /tmp/sqoop-nitesh/compile/2c9c65bdb50014b105f54623546931fa/shell_que1.jar
2017-05-26 08:38:38,880 INFO [main] mapreduce.ExportJobBase: Beginning export of shell_que1
2017-05-26 08:38:38,880 INFO [main] Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/nitesh/hbase_dlr/hbase-0.98.4-hadoop2/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]

```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
[main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.tasks.speculative.execution
2017-05-26 08:38:39,079 INFO [main] Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2017-05-26 08:38:39,037 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM shell_quel AS t WHERE 1=0
2017-05-26 08:38:39,641 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.tasks.speculative.execution
2017-05-26 08:38:39,645 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.tasks.speculative.execution
2017-05-26 08:38:39,645 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-26 08:38:39,755 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:43,943 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-26 08:38:43,946 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-26 08:38:44,067 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.tasks.speculative.execution
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputformat.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputformat.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.value.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.files
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-26 08:38:44,211 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495765181217_0022
2017-05-26 08:38:44,459 INFO [main] impl.YarnClientImpl: Submitted application application_1495765181217_0022 to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:44,489 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495765181217_0022/
2017-05-26 08:38:44,490 INFO [main] mapreduce.Job: Running job: job_1495765181217_0022
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
[main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.tasks.speculative.execution
2017-05-26 08:38:39,645 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-26 08:38:39,755 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:43,943 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-26 08:38:43,946 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-26 08:38:44,067 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2017-05-26 08:38:44,103 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.tasks.speculative.execution
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputformat.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputformat.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.value.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key.class
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.files
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-05-26 08:38:44,104 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-26 08:38:44,211 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495765181217_0022
2017-05-26 08:38:44,459 INFO [main] impl.YarnClientImpl: Submitted application application_1495765181217_0022 to ResourceManager at /0.0.0.0:8032
2017-05-26 08:38:44,489 INFO [main] mapreduce.Job: The url to track the job: http://http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495765181217_0022/
2017-05-26 08:38:44,490 INFO [main] mapreduce.Job: Running job: job_1495765181217_0022
2017-05-26 08:38:49,707 INFO [main] mapreduce.Job: Job job_1495765181217_0022 running in uber mode : false
2017-05-26 08:38:49,708 INFO [main] mapreduce.Job: map % reduce %
2017-05-26 08:38:55,935 INFO [main] mapreduce.Job: map 100% reduce %
2017-05-26 08:38:55,952 INFO [main] mapreduce.Job: Job job_1495765181217_0022 completed successfully
2017-05-26 08:38:56,067 ERROR [main] sqoop.Sqoop: Got exception running Sqoop: java.lang.IllegalArgumentException: No enum constant org.apache.hadoop.mapreduce.JobCounter.MB_MILLIS_MAPS
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
2017-05-26 08:38:59,562 INFO [main] Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
2017-05-26 08:38:59,567 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-26 08:38:59,568 INFO [main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-05-26 08:38:59,670 INFO [main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-05-26 08:39:03,158 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-26 08:39:03,162 INFO [main] input.FileInputFormat: Total input paths to process : 1
2017-05-26 08:39:03,294 INFO [main] mapreduce.JobSubmitter: number of splits:1
2017-05-26 08:39:03,318 INFO [main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-05-26 08:39:03,318 INFO [main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job.cache.files.timestamps
2017-05-26 08:39:03,318 INFO [main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-05-26 08:39:03,318 INFO [main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2017-05-26 08:39:03,318 INFO [main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
2017-05-26 08:39:03,318 INFO [main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.inputformat.class
2017-05-26 08:39:03,318 INFO [main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.outputformat.class
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.files
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.output.value.class
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.output.key.class
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.classpath.files
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-05-26 08:39:03,319 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.cache.files.filesizes
2017-05-26 08:39:03,427 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1495765181217_0023
2017-05-26 08:39:03,674 INFO [main] impl.YarnClientImpl: Submitted application application_1495765181217_0023 to ResourceManager at /0.0.0.0:8032
2017-05-26 08:39:03,709 INFO [main] mapreduce.Job: The url to track the job: http://nitesh-Lenovo-Z50-70:8088/proxy/application_1495765181217_0023
2017-05-26 08:39:03,709 INFO [main] mapreduce.Job: Running job: job_1495765181217_0023
2017-05-26 08:39:09,840 INFO [main] mapreduce.Job: Job job_1495765181217_0023 running in uber mode : false
2017-05-26 08:39:09,842 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-05-26 08:39:15,910 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-05-26 08:39:15,925 INFO [main] mapreduce.Job: Job job_1495765181217_0023 completed successfully
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
mysql: [Warning] Using a password on the command line interface can be insecure.
+-----+-----+-----+
| state | district | percent |
+-----+-----+-----+
Andhra Pradesh NIZAMABAD 100
Arunachal Pradesh TIRAP 100
Assam HAILAKANDI 100
Bihar MADHUBANI 100
Goa NORTH GOA 100
Gujarat AHMEDABAD 100
Gujarat DANGS 100
Gujarat NAVSARI 100
Gujarat PORBANDAR 100
Gujarat SURAT 100
Haryana FARIDABAD 100
Haryana HISAR 100
Haryana JHAJJAR 100
Haryana MAHENDRAGARH 100
Haryana PANIPAT 100
Haryana ROHTAK 100
Haryana SIRSA 100
Himachal Pradesh HAMIRPUR 100
Himachal Pradesh KINNAUR 100
Himachal Pradesh KULLU 100
Himachal Pradesh LAHUL & SPITI 100
Himachal Pradesh SHIMLA 100
Himachal Pradesh SOLAN 100
Himachal Pradesh UNA 100
Jharkhand DEOGHAR 100
Jharkhand LOHARDAGA 100
Karnataka HASSAN 100
Karnataka MANGALORE(DAKSHINA KANNADA) 100
Karnataka UDUPI 100
Kerala ALAPPUZHA 100
Kerala KOLLAM 100
Kerala KOTTAYAM 100
Kerala KOZHIKODE 100
Kerala PALAKKAD 100
Kerala PATHANAMTHitta 100
Kerala WAYANAD 100
Maharashtra GADCHIROLI 100
Maharashtra SINDHUDURG 100
Meghalaya WEST GARO HILLS 100
Mizoram CHAMPHAI 100
```

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
Uttar Pradesh AMBEDKAR NAGAR 100
Uttar Pradesh BALRAMPUR 100
Uttar Pradesh BAREILLY 100
Uttar Pradesh BIJNOR 100
Uttar Pradesh BUDAUN 100
Uttar Pradesh ETAWAH 100
Uttar Pradesh FARRUKHABAD 100
Uttar Pradesh FIROZABAD 100
Uttar Pradesh GHAZIABAD 100
Uttar Pradesh HARDI 100
Uttar Pradesh JYOTIBA PHULE NAGAR 100
Uttar Pradesh LUCKNOW 100
Uttar Pradesh MAHARAJGANJ 100
Uttar Pradesh MAHOBIA 100
Uttar Pradesh MORADABAD 100
Uttar Pradesh MUZAFFARNAGAR 100
Uttar Pradesh PILIBHIT 100
Uttar Pradesh SONBHADRA 100
Uttar Pradesh SULTANPUR 100
OUTPUT OF THE SECOND SCRIPT
mysql: [Warning] Using a password on the command line interface can be insecure.
state district percent
Andhra Pradesh ANANTAPUR 100.893
Andhra Pradesh CHITTOR 90.9888
Andhra Pradesh CUDDAPAH 95.282
Andhra Pradesh EAST GODAVARI 93.8016
Andhra Pradesh KARIMNAGAR 101.141
Andhra Pradesh KHAMMAM 103.455
Andhra Pradesh KRISHNA 90.6585
Andhra Pradesh KURNool 84.3897
Andhra Pradesh MEDAK 99.6305
Andhra Pradesh NALGONDA 104.536
Andhra Pradesh NIZAMABAD 100
Andhra Pradesh RANGAREDDI 82.049
Andhra Pradesh WARANGAL 108.924
Andhra Pradesh WEST GODAVARI 92.7978
Arunachal Pradesh DIBANG VALLEY 100.276
Arunachal Pradesh LOHIT 95.5682
Arunachal Pradesh TIRAP 100
Assam BAGSHA 85.7673
Assam CACHAR 84.2776
Assam DIBRUGARH 90.0884
Assam GOALPARA 85.6724

```

From the above screen shot we can see the eco command is executed and we can separate the results of query1 and query2.

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music
Haryana GURGAON 140.867
Haryana HISAR 100
Haryana JHAJJAR 100
Haryana JIND 98.3431
Haryana KATHAL 98.7643
Haryana KARNAL 100.091
Haryana KURUKSHETRA 100.271
Haryana MAHENDRAGARH 100
Haryana MEWAT 90.688
Haryana PANCHKULA 100
Haryana PANIPAT 100
Haryana REWAR 88.4933
Haryana ROHTAK 100
Haryana SIRSA 100
Haryana SONIPAT 101.651
Haryana YAMUNANAGAR 98.8123
Himachal Pradesh BILASPUR 109.614
Himachal Pradesh CHAMBA 131.495
Himachal Pradesh HAMIRPUR 100
Himachal Pradesh KANGRA 137.378
Himachal Pradesh KINNAUR 100
Himachal Pradesh KULLU 100
Himachal Pradesh LAH AUL & SPITI 100
Himachal Pradesh MANDI 100.003
Himachal Pradesh SHIMLA 100
Himachal Pradesh SIRMAUR 99.4038
Himachal Pradesh SOLAN 100
Himachal Pradesh UNA 100
Jammu & Kashmir ANANTNAG 82.0632
Jammu & Kashmir LEH (LADAKH) 86.9622
Jharkhand DEOGHAR 100
Jharkhand DUMKA 85.328
Jharkhand LATEHAR 83.4495
Jharkhand LOHARDAGA 100
Jharkhand PAKUR 87.3495
Jharkhand PURBTI SINGHBHUM 130.63
Karnataka BAGALKOT 91.908
Karnataka BANGALORE RURAL 139.545
Karnataka CHICKMAGALUR 95.9966
Karnataka CHITRADURGA 97.5928
Karnataka DHARWAD 92.9976
Karnataka GADAG 84.1802
Karnataka HASSAN 100

```

nitesh@nitesh-Lenovo-Z50-70: ~/Music		
Madhya Pradesh	DEWAS	134.584
Madhya Pradesh	DHAR	119.608
Madhya Pradesh	DINDORI	94.0017
Madhya Pradesh	GUNA	80.0468
Madhya Pradesh	GWALIOR	121.642
Madhya Pradesh	HARDA	133.728
Madhya Pradesh	HOSHANGABAD	164.708
Madhya Pradesh	INDORE	102.355
Madhya Pradesh	JABALPUR	110.833
Madhya Pradesh	JHABUA	115.65
Madhya Pradesh	KATNI	128.858
Madhya Pradesh	KHANDWA(EAST NIMAR)	137.382
Madhya Pradesh	KHARGONE	100.412
Madhya Pradesh	MANDLA	82.3984
Madhya Pradesh	MANDSAUR	145.391
Madhya Pradesh	MORENA	122.32
Madhya Pradesh	NARSINGHPUR	98.3156
Madhya Pradesh	NEEMUCH	154.167
Madhya Pradesh	RAISEN	115.4
Madhya Pradesh	RAJGARH	122.283
Madhya Pradesh	RATLAM	125.814
Madhya Pradesh	REMA	121.395
Madhya Pradesh	SEHORE	224.157
Madhya Pradesh	SEONI	88.635
Madhya Pradesh	SHAHDOL	102.104
Madhya Pradesh	SHAJAPUR	114.96
Madhya Pradesh	SHEOPUR	159.111
Madhya Pradesh	SINGRAULI	99.315
Madhya Pradesh	UJJAIN	113.473
Madhya Pradesh	UMARTA	134.171
Madhya Pradesh	VIDISHA	85.6183
Maharashtra	AHMEDNAGAR	98.4342
Maharashtra	BHANDARA	87.5362
Maharashtra	DHULE	80.5325
Maharashtra	GADCHIROLI	100
Maharashtra	GONDIA	94.4182
Maharashtra	HINGOLI	91.2322
Maharashtra	JALNA	97.2189
Maharashtra	KOLHAPUR	80.2574
Maharashtra	NAGPUR	88.3415
Maharashtra	OSMANABAD	99.9779
Maharashtra	PARBHANI	87.1717
Maharashtra	PUNE	97.7041

nitesh@nitesh-Lenovo-Z50-70: ~/Music		
Punjab	HOSHIARPUR	171.832
Punjab	JALANDHAR	98.8333
Punjab	KAPURTHALA	95.8627
Punjab	LUDHIANA	103.298
Punjab	MANSA	93.592
Punjab	NAWANSIHR	159.233
Punjab	S.A.S Nagar	100.424
Rajasthan	AJMER	89.486
Rajasthan	CHURU	94.6775
Rajasthan	DUNGARPUR	82.7233
Rajasthan	GANGANAGAR	99.0238
Rajasthan	HANUMANGARH	100
Rajasthan	JAISALMER	92.8388
Rajasthan	NAGAUR	98.6157
Rajasthan	SIKAR	83.4434
Sikkim	EAST SIKKIM	111.377
Sikkim	NORTH SIKKIM	109.703
Sikkim	SOUTH SIKKIM	138.751
Sikkim	WEST SIKKIM	100.755
Tamil Nadu	COIMBATORE	91.9013
Tamil Nadu	CUDDALORE	98.4481
Tamil Nadu	DHARMAPURI	100.032
Tamil Nadu	DINDIGUL	86.3203
Tamil Nadu	ERODE	100
Tamil Nadu	KANCHIPURAM	104.839
Tamil Nadu	KANYAKUMARI(NAGERCOIL)	111.267
Tamil Nadu	KARUR	100
Tamil Nadu	MADURAI	83.0213
Tamil Nadu	NAMAKKAL	100
Tamil Nadu	NILGIRIS(UDHAGAMANDALAM)	80.3326
Tamil Nadu	PERAMBALUR	96.4005
Tamil Nadu	PUDUKKOTTAI	112.183
Tamil Nadu	RAMANATHAPURAM	149.246
Tamil Nadu	SALEM	84.025
Tamil Nadu	SIVAGANGA	94.3665
Tamil Nadu	THENI	97.5968
Tamil Nadu	THOOTHUKUDI	91.5078
Tamil Nadu	TIRUCHIRAPPALLI	100
Tamil Nadu	TIRUNELVELI	110.782
Tamil Nadu	TIRUVANNAMALAI	100
Tamil Nadu	TIRUVARUR	108.648
Tamil Nadu	VELLORE	96.2776
Tamil Nadu	VIRUDHUNAGAR	87.6451

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
Uttar Pradesh BASTI 81.1466
Uttar Pradesh BIJNOR 100
Uttar Pradesh BUDAUN 100
Uttar Pradesh BULANDSHAHAR 110.77
Uttar Pradesh CHANDAULI 104.668
Uttar Pradesh CHITRAKOOT 112
Uttar Pradesh DEORIA 95.0321
Uttar Pradesh ETAH 111.513
Uttar Pradesh ETAWAH 100
Uttar Pradesh FAIZABAD 98.3188
Uttar Pradesh FARRUKHABAD 100
Uttar Pradesh FATEHPUR 95.7352
Uttar Pradesh FIROZABAD 100
Uttar Pradesh GAUTAM BUDDHA NAGAR 98.1337
Uttar Pradesh GHAZIABAD 100
Uttar Pradesh GHAZIPUR 99.1362
Uttar Pradesh GONDA 86.8813
Uttar Pradesh GORAKHPUR 99.9891
Uttar Pradesh HAMIRPUR 84.0198
Uttar Pradesh HARDOI 100
Uttar Pradesh JALAUN 89.106
Uttar Pradesh JAUNPUR 100.37
Uttar Pradesh JHANSI 100.498
Uttar Pradesh JYOTIBA PHULE NAGAR 100
Uttar Pradesh KANNAUJ 100.117
Uttar Pradesh KANPUR DEHAT 95.8658
Uttar Pradesh KANPUR NAGAR 93.7124
Uttar Pradesh KAUSHambi 96.8289
Uttar Pradesh KUSHINAGAR 82.2261
Uttar Pradesh LAKHTIMPUR KHERI 131.153
Uttar Pradesh LALITPUR 98.8225
Uttar Pradesh LUCKNOW 100
Uttar Pradesh MAHAMAYA NAGAR(HATHRAS) 100.493
Uttar Pradesh MAHARAJGANJ 100
Uttar Pradesh MAHOBIA 100
Uttar Pradesh MAINPURI 91.8595
Uttar Pradesh MATHURA 100.42
Uttar Pradesh MAU 81.1694
Uttar Pradesh MEERUT 99.3847
Uttar Pradesh MIRzapur 100.862
Uttar Pradesh MORADABAD 100
Uttar Pradesh MUZAFFARNAGAR 100
Uttar Pradesh PILIBHIT 100
```

```
nitesh@nitesh-Lenovo-Z50-70: ~/Music
Uttar Pradesh MAHOBIA 100
Uttar Pradesh MAINPURI 91.8595
Uttar Pradesh MATHURA 100.42
Uttar Pradesh MAU 81.1694
Uttar Pradesh MEERUT 99.3847
Uttar Pradesh MIRzapur 100.862
Uttar Pradesh MORADABAD 100
Uttar Pradesh MUZAFFARNAGAR 100
Uttar Pradesh PILIBHIT 100
Uttar Pradesh PRATAPGARH 101.52
Uttar Pradesh RAE BARELI 100.065
Uttar Pradesh RAMPUR 91.2306
Uttar Pradesh SAHARANPUR 100.259
Uttar Pradesh SANT RAVIDAS NAGAR( BHADOHI) 93.0577
Uttar Pradesh SHAHZAHANPUR 100.161
Uttar Pradesh SHRAVasti 94.146
Uttar Pradesh SIDDHARTHNAGAR 89.2548
Uttar Pradesh SITAPUR 89.5722
Uttar Pradesh SONBHADRA 100
Uttar Pradesh SULTANPUR 100
Uttar Pradesh UNNAO 97.4002
Uttar Pradesh VARANASI 97.1748
Uttarakhand BAGESHWAR 83.2704
Uttarakhand CHAMOLI 95.9333
Uttarakhand DEHRADUN 85.2521
Uttarakhand HARIDWAR 89.8541
Uttarakhand NAINITAL 126.023
Uttarakhand PITHORAGARH 84.1571
Uttarakhand RUDRAPRAYAG 105.019
Uttarakhand TEHRI GARHWAL 83.9052
Uttarakhand UDHAM SINGH NAGAR 95.3763
Uttarakhand UTTARKASHI 90.5424
West Bengal BARDHAMAN 85.9808
West Bengal DAKSHIN DINAJPUR 100.839
West Bengal HOOGHLY 99.2794
West Bengal HOWRAH 99.2797
West Bengal JALPAIGURI 90.5472
West Bengal MIDNAPUR EAST 134.411
West Bengal MIDNAPUR WEST 117.035
West Bengal NADIA 92.7216
West Bengal NORTH 24 PARAGANAS 99.0312
West Bengal SOUTH 24 PARAGANAS 94.4331
nitesh@nitesh-Lenovo-Z50-70:~/Music$
```

The above screen shots show the results of the second query or problem statement 2