Project 1.2

Project 1.2 - State-Wise Development Analysis in India

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Contents

1. Project Overview

To develop the System to analyze the log data (In XML format) of government progress of various development activities.

1.1 Purpose and Scope of this Specification

The following requirement will be addressed in phase 1 of Project:

- Developing system to handle the incoming log feed and store the information in HadoopCluster (Flume)
- Analyze the data and understand the progress
- Store the results in Hbase/RDBMS

Out of scope

We can use this data and visualization and get more insights

2. Product/Service Description

2.1 Assumptions

Log will be generated in XML format and stored in a server.

2.2 Constraints

Describe any item that will constrain the design options, including

- This system may not be used for searching for now. But it will be used for analysis and saving the relevant information as of now.
- System will be using mySql as a database

3. Requirements

- The FLUME job which will format the data and place the data to HDFS
- Pig/MapReduce job for parsing the XML data.
- Create Pig scripts/MapReduce jobs to analyze the data
- Create the Sqoop job to store the data in database

Priority Definitions

The following definitions are intended as a guideline to prioritize requirements.

- **Priority 1** – Create FLUME job for fetching log files from spool directory the data
- **Priority 2** – MapReduce/pig job to preprocess

4. Dataset

Download the dataset using the below link:

Link: https://drive.google.com/file/d/0Bxr27gVaXO5sUjd2RWFQS3hQQUE/view?usp=sharing

Refer the below steps to understand the actual steps to create the above project.

Step 1:

Copy dataset from local file system to HDFS using flume.

Note: use the conf file by downloading from below link.

filecopy.conf

Command:

```
flume-ng  agent –n agent1 –c conf –f <path to filecopy.conf>
```

Step 2:

Input file is in the XML format use Map reduce or pig to parse the data and get the results for the below problem statements.

5. Problem statement

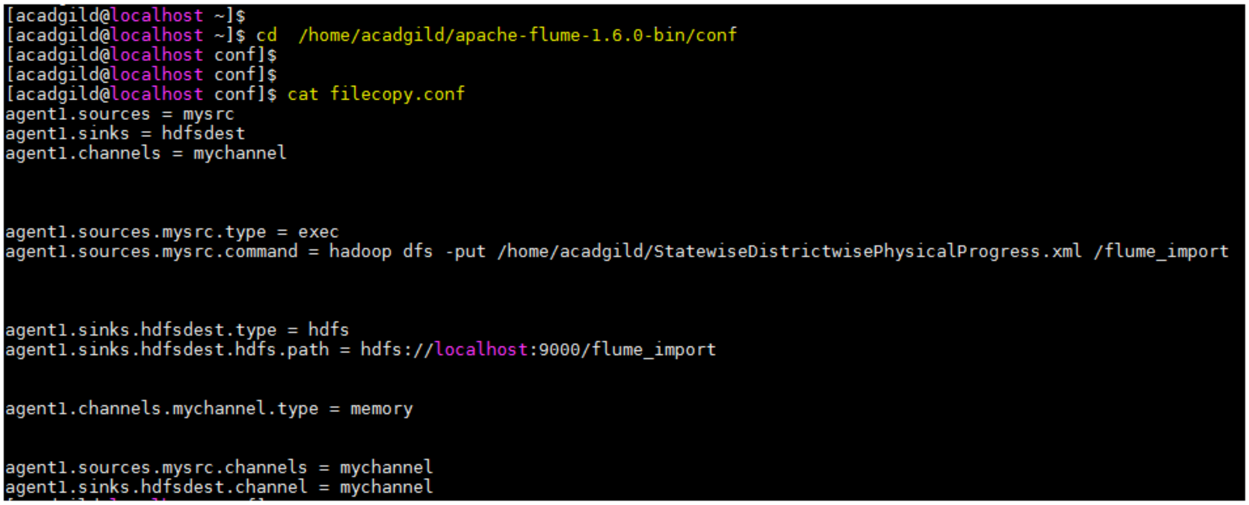
1. Find out the districts who achieved 100 percent objective in BPL cards Export the results to mysql using sqoop
2. Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards. Export the results to MySQL using Sqoop.

Project Execution

Problem Statement1 - Find out the districts who achieved 100 percent objective in BPL cards Export the results to mysql using sqoop

Task 1 – Place Dataset in the target using flume,

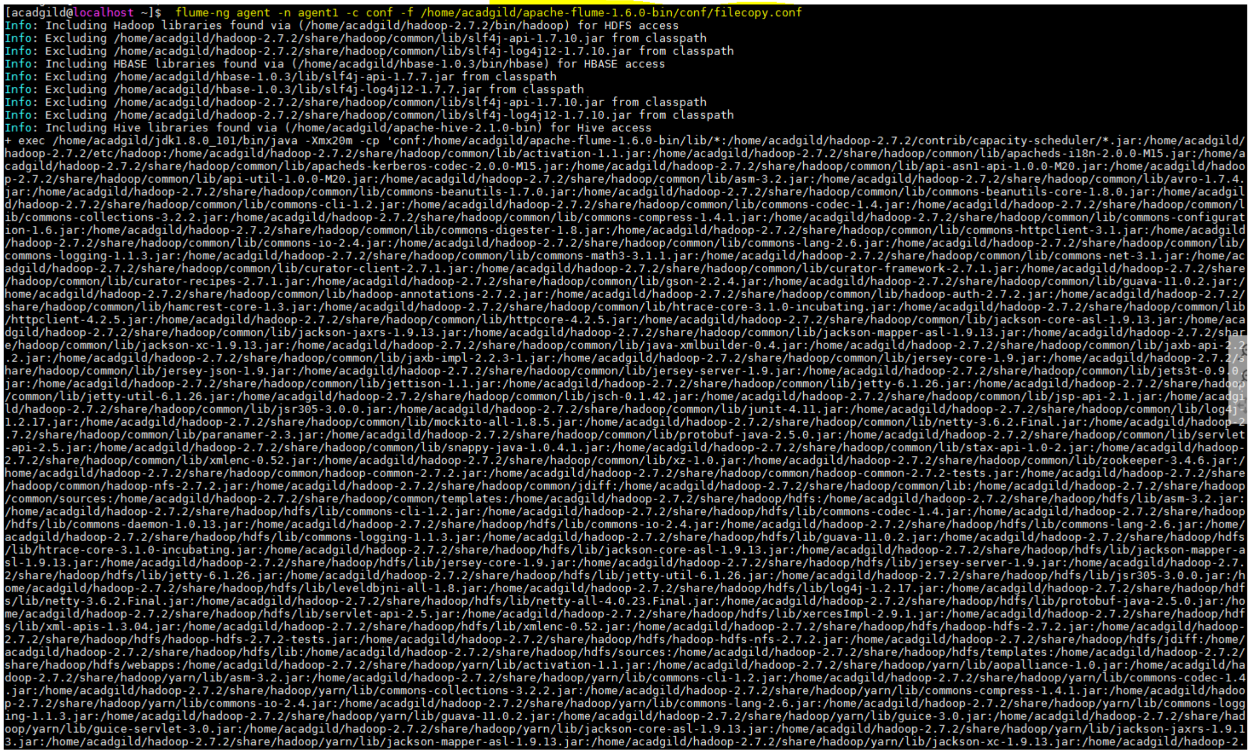
Place the flume config file provided at the location, **/home/acadgild/apache-flume-1.6.0-bin/conf**



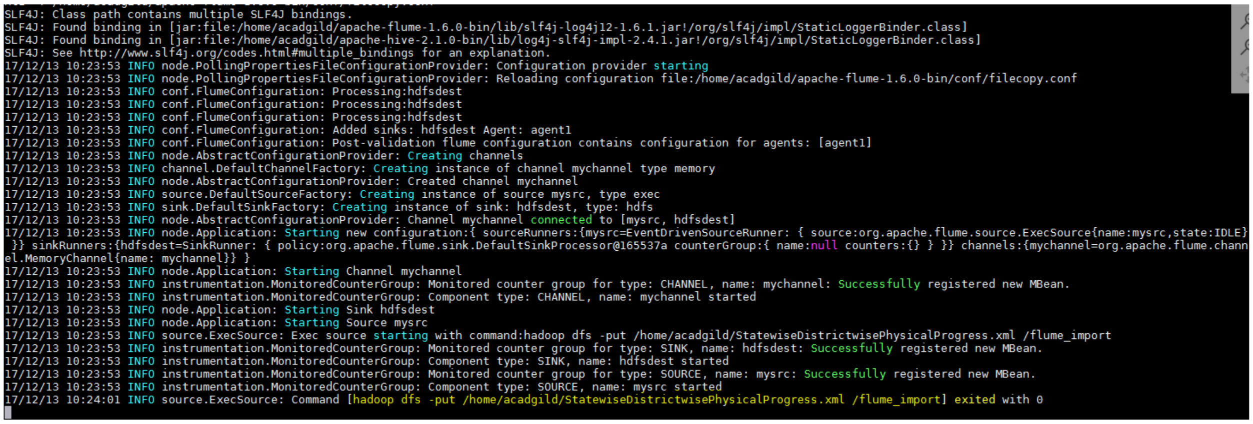
```
[acadgild@localhost ~]$
[acadgild@localhost ~]$ cd  /home/acadgild/apache-flume-1.6.0-bin/conf
[acadgild@localhost conf]$
[acadgild@localhost conf]$
[acadgild@localhost conf]$ cat filecopy.conf
agent1.sources = mysrc
agent1.sinks = hdfsdest
agent1.channels = mychannel



agent1.sources.mysrc.type = exec
agent1.sources.mysrc.command = hadoop dfs -put /home/acadgild/StatewiseDistrictwisePhysicalProgress.xml /flume_import



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
```

Copy the dataset downloaded from the link from local file system to HDFS using flume using the below command,

flume-ng agent -n agent1 -c conf -f /home/acadgild/apache-flume-1.6.0-bin/conf/filecopy.conf



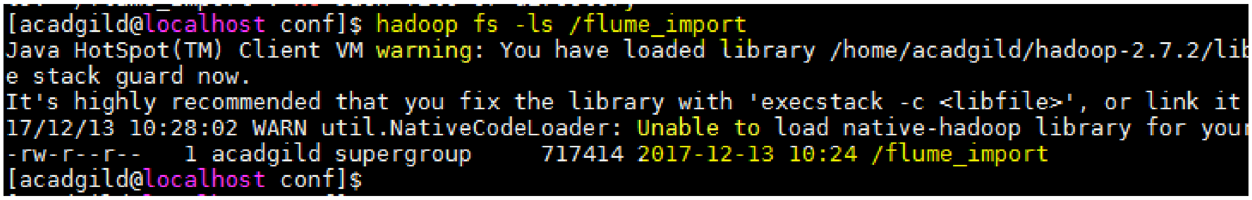
```
[acadgild@localhost ~]$  flume-ng agent -n agent1 -c conf -f /home/acadgild/apache-flume-1.6.0-bin/conf/filecopy.conf
Info: Including Hadoop libraries found via (/home/acadgild/hadoop-2.7.2/bin/hadoop) for HDFS access
Info: Excluding /home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/slf4j-api-1.7.10.jar from classpath
Info: Excluding /home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar from classpath
Info: Including HBASE libraries found via (/home/acadgild/hbase-1.0.3/bin/hbase) for HBASE access
Info: Excluding /home/acadgild/hbase-1.0.3/lib/slf4j-api-1.7.7.jar from classpath
Info: Excluding /home/acadgild/hbase-1.0.3/lib/slf4j-log4j12-1.7.7.jar from classpath
Info: Excluding /home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/slf4j-api-1.7.10.jar from classpath
Info: Excluding /home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar from classpath
Info: Including Hive libraries found via (/home/acadgild/apache-hive-2.1.0-bin) for Hive access
+ exec /home/acadgild/jdk1.8.0_101/bin/java -Xmx20m -cp 'conf:/home/acadgild/apache-flume-1.6.0-bin/lib/*:/home/acadgild/hadoop-2.7.2/contrib/capacity-scheduler/*.jar:/home/acadgild/
hadoop-2.7.2/etc/hadoop:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/activation-1.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/apacheds-i18n-2.0.0-M15.jar:/home/a
cadgild/hadoop-2.7.2/share/hadoop/common/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/api-asn1-api-1.0.0-M20.jar:/home/acadgild/hadoo
p-2.7.2/share/hadoop/common/lib/api-util-1.0.0-M20.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/asm-3.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/avro-1.7.4.
jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-beanutils-core-1.8.0.jar:/home/acadgil
d/hadoop-2.7.2/share/hadoop/common/lib/commons-cli-1.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-codec-1.4.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/l
ib/commons-collections-3.2.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-compress-1.4.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-configurat
ion-1.6.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-digester-1.8.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-httpclient-3.1.jar:/home/acadgild
/hadoop-2.7.2/share/hadoop/common/lib/commons-io-2.4.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-lang-2.6.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/
commons-logging-1.1.3.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-math3-3.1.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/commons-net-3.1.jar:/home/ac
adgild/hadoop-2.7.2/share/hadoop/common/lib/curator-client-2.7.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/curator-framework-2.7.1.jar:/home/acadgild/hadoop-2.7.2/share
/hadoop/common/lib/curator-recipes-2.7.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/gson-2.2.4.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/guava-11.0.2.jar:/
home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/hadoop-annotations-2.7.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/hadoop-auth-2.7.2.jar:/home/acadgild/hadoop-2.7.2/
share/hadoop/common/lib/hamcrest-core-1.3.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/htrace-core-3.1.0-incubating.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib
/httpclient-4.2.5.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/httpcore-4.2.5.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/home/aca
dgild/hadoop-2.7.2/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jackson-mapper-asl-1.9.13.jar:/home/acadgild/hadoop-2.7.2/shar
e/hadoop/common/lib/jackson-xc-1.9.13.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/java-xmlbuilder-0.4.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jaxb-api-2.2
.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jersey-core-1.9.jar:/home/acadgild/hadoop-2.7.2/s
hare/hadoop/common/lib/jersey-json-1.9.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jersey-server-1.9.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jets3t-0.9.0.
jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jettison-1.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jetty-6.1.26.jar:/home/acadgild/hadoop-2.7.2/share/hadoop
/common/lib/jetty-util-6.1.26.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jsch-0.1.42.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/jsp-api-2.1.jar:/home/acadgi
ld/hadoop-2.7.2/share/hadoop/common/lib/jsr305-3.0.0.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/junit-4.11.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/log4j-
1.2.17.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/mockito-all-1.8.5.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/netty-3.6.2.Final.jar:/home/acadgild/hadoop-2
.7.2/share/hadoop/common/lib/paranamer-2.3.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/servlet
-api-2.5.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/snappy-java-1.0.4.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/stax-api-1.0-2.jar:/home/acadgild/hadoop-
2.7.2/share/hadoop/common/lib/xmlenc-0.52.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/xz-1.0.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/zookeeper-3.4.6.jar:/
home/acadgild/hadoop-2.7.2/share/hadoop/common/hadoop-common-2.7.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/hadoop-common-2.7.2-tests.jar:/home/acadgild/hadoop-2.7.2/share
/hadoop/common/hadoop-nfs-2.7.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/common/jdiff:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib:/home/acadgild/hadoop-2.7.2/share/hadoop
/common/sources:/home/acadgild/hadoop-2.7.2/share/hadoop/common/templates:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/asm-3.2.jar:
/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/commons-cli-1.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/commons-codec-1.4.jar:/home/acadgild/hadoop-2.7.2/share/hadoop
/hdfs/lib/commons-daemon-1.0.13.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/commons-io-2.4.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/commons-lang-2.6.jar:/home/
acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/commons-logging-1.1.3.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/guava-11.0.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs
/lib/htrace-core-3.1.0-incubating.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/jackson-core-asl-1.9.13.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/jackson-mapper-a
sl-1.9.13.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/jersey-core-1.9.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/jersey-server-1.9.jar:/home/acadgild/hadoop-2.7.
2/share/hadoop/hdfs/lib/jetty-6.1.26.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/jetty-util-6.1.26.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/jsr305-3.0.0.jar:/h
ome/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/leveldbjni-all-1.8.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/log4j-1.2.17.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdf
s/lib/netty-3.6.2.Final.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/netty-all-4.0.23.Final.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/protobuf-java-2.5.0.jar:/ho
me/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/servlet-api-2.5.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/xercesImpl-2.9.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdf
s/lib/xml-apis-1.3.04.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib/xmlenc-0.52.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/hadoop-hdfs-2.7.2.jar:/home/acadgild/hadoop-
2.7.2/share/hadoop/hdfs/hadoop-hdfs-2.7.2-tests.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/hadoop-hdfs-nfs-2.7.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/jdiff:/home/
acadgild/hadoop-2.7.2/share/hadoop/hdfs/lib:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/sources:/home/acadgild/hadoop-2.7.2/share/hadoop/hdfs/templates:/home/acadgild/hadoop-2.7.2/
share/hadoop/hdfs/webapps:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/activation-1.1.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/aopalliance-1.0.jar:/home/acadgild/ha
doop-2.7.2/share/hadoop/yarn/lib/asm-3.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/commons-cli-1.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/commons-codec-1.4
.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/commons-collections-3.2.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/commons-compress-1.4.1.jar:/home/acadgild/hadoo
p-2.7.2/share/hadoop/yarn/lib/commons-io-2.4.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/commons-lang-2.6.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/commons-logg
ing-1.1.3.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/guava-11.0.2.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/guice-3.0.jar:/home/acadgild/hadoop-2.7.2/share/had
oop/yarn/lib/guice-servlet-3.0.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/jackson-core-asl-1.9.13.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/jackson-jaxrs-1.9.1
3.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/jackson-mapper-asl-1.9.13.jar:/home/acadgild/hadoop-2.7.2/share/hadoop/yarn/lib/jackson-xc-1.9.13.jar:/home/acadgild/hadoop-2.
```



```
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/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/acadgild/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.
17/12/13 10:23:53 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting
17/12/13 10:23:53 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:/home/acadgild/apache-flume-1.6.0-bin/conf/filecopy.conf
17/12/13 10:23:53 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/12/13 10:23:53 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/12/13 10:23:53 INFO conf.FlumeConfiguration: Processing:hdfsdest
17/12/13 10:23:53 INFO conf.FlumeConfiguration: Added sinks: hdfsdest Agent: agent1
17/12/13 10:23:53 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent1]
17/12/13 10:23:53 INFO node.AbstractConfigurationProvider: Creating channels
17/12/13 10:23:53 INFO channel.DefaultChannelFactory: Creating instance of channel mychannel type memory
17/12/13 10:23:53 INFO node.AbstractConfigurationProvider: Created channel mychannel
17/12/13 10:23:53 INFO source.DefaultSourceFactory: Creating instance of source mysrc, type exec
17/12/13 10:23:53 INFO sink.DefaultSinkFactory: Creating instance of sink: hdfsdest, type: hdfs
17/12/13 10:23:53 INFO node.AbstractConfigurationProvider: Channel mychannel connected to [mysrc, hdfsdest]
17/12/13 10:23:53 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@165537a counterGroup:{ name:null counters:{} } }} channels:{mychannel=org.apache.flume.chann
el.MemoryChannel{name: mychannel}} }
17/12/13 10:23:53 INFO node.Application: Starting Channel mychannel
17/12/13 10:23:53 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: mychannel: Successfully registered new MBean.
17/12/13 10:23:53 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: mychannel started
17/12/13 10:23:53 INFO node.Application: Starting Sink hdfsdest
17/12/13 10:23:53 INFO node.Application: Starting Source mysrc
17/12/13 10:23:53 INFO source.ExecSource: Exec source starting with command:hadoop dfs -put /home/acadgild/StatewiseDistrictwisePhysicalProgress.xml /flume_import
17/12/13 10:23:53 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: hdfsdest: Successfully registered new MBean.
17/12/13 10:23:53 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: hdfsdest started
17/12/13 10:23:53 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: mysrc: Successfully registered new MBean.
17/12/13 10:23:53 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: mysrc started
17/12/13 10:24:01 INFO source.ExecSource: Command [hadoop dfs -put /home/acadgild/StatewiseDistrictwisePhysicalProgress.xml /flume_import] exited with 0
```

Verify whether the file is copied in the target,

Hadoop fs –ls /flume_import



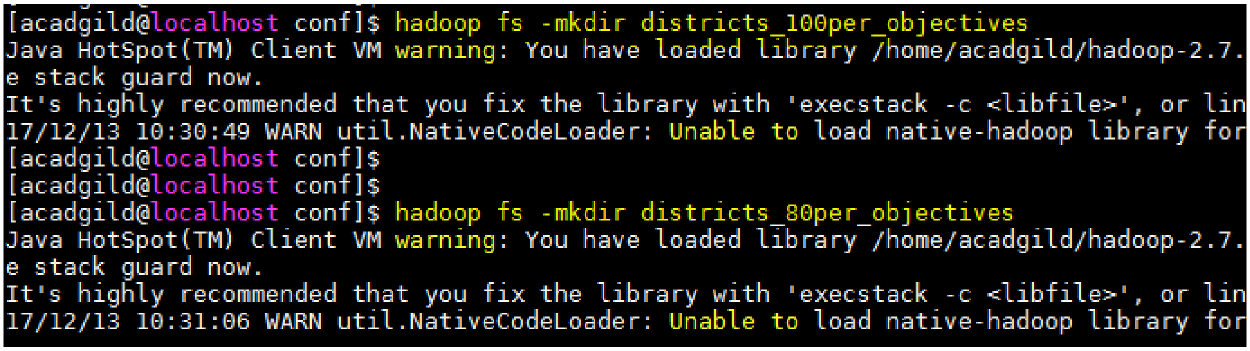
```
[acadgild@localhost conf]$ hadoop fs -ls /flume_import
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it
17/12/13 10:28:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
-rw-r--r--   1 acadgild supergroup     717414 2017-12-13 10:24 /flume_import
[acadgild@localhost conf]$
```

Task2 – Create folders in the HDFS to store the outputs,

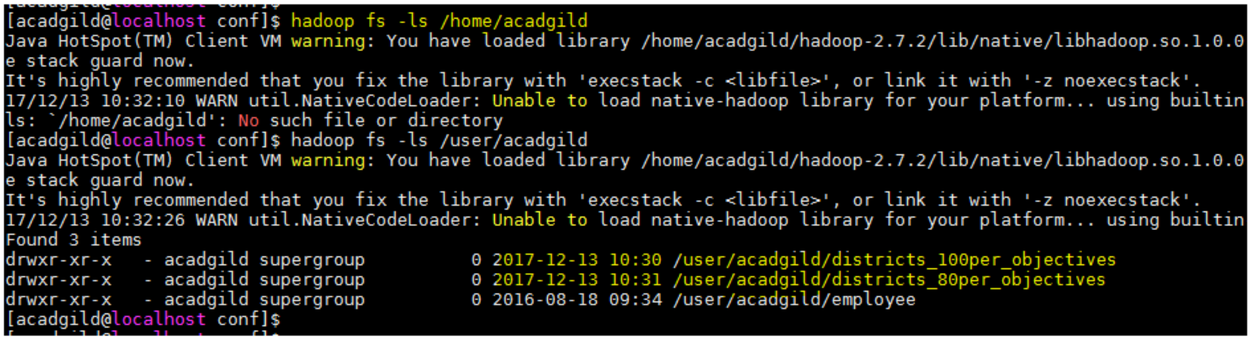
Create 2 folders in the HDFS where we are going to store the output from PIG execution,

hadoop fs -mkdir districts_100per_objectives

hadoop fs -mkdir districts_80per_objectives



```
[acadgild@localhost conf]$ hadoop fs -mkdir districts_100per_objectives
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or lin
17/12/13 10:30:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
[acadgild@localhost conf]$
[acadgild@localhost conf]$
[acadgild@localhost conf]$ hadoop fs -mkdir districts_80per_objectives
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or lin
17/12/13 10:31:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
```



```
[acadgild@localhost conf]$ hadoop fs -ls /home/acadgild
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/12/13 10:32:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
ls: `/home/acadgild': No such file or directory
[acadgild@localhost conf]$ hadoop fs -ls /user/acadgild
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/12/13 10:32:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
Found 3 items
drwxr-xr-x   - acadgild supergroup          0 2017-12-13 10:30 /user/acadgild/districts_100per_objectives
drwxr-xr-x   - acadgild supergroup          0 2017-12-13 10:31 /user/acadgild/districts_80per_objectives
drwxr-xr-x   - acadgild supergroup          0 2016-08-18 09:34 /user/acadgild/employee
[acadgild@localhost conf]$
```

Task3 – Create Database and the Tables in the mysql

Start mysql> sudo service mysqld start

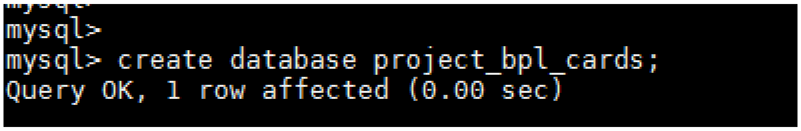
Login as root user,

create database project_bpl_cards;

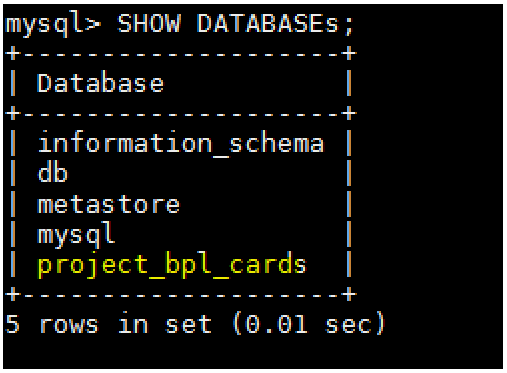
use project_bpl_cards;

create table districts_100percent_objective (district_name varchar(50));

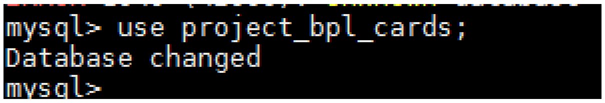
create table districts_80percent_objective (district_name varchar(50));



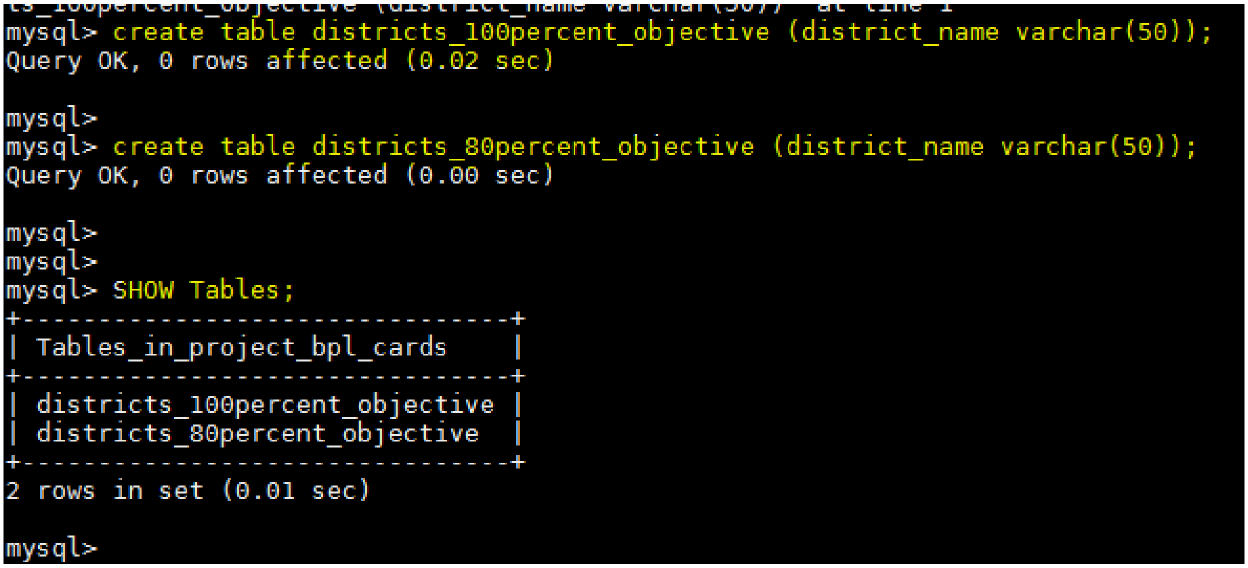
```
mysql>
mysql> create database project_bpl_cards;
Query OK, 1 row affected (0.00 sec)
```



```
mysql> SHOW DATABASEs;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| db                 |
| metastore          |
| mysql              |
| project_bpl_cards  |
+--------------------+
5 rows in set (0.01 sec)
```



```
mysql> use project_bpl_cards;
Database changed
mysql>
```



```
ts_100percent_objective (district_name varchar(50))' at line 1
mysql> create table districts_100percent_objective (district_name varchar(50));
Query OK, 0 rows affected (0.02 sec)

mysql>
mysql> create table districts_80percent_objective (district_name varchar(50));
Query OK, 0 rows affected (0.00 sec)

mysql>
mysql>
mysql> SHOW Tables;
+--------------------------------+
| Tables_in_project_bpl_cards    |
+--------------------------------+
| districts_100percent_objective |
| districts_80percent_objective  |
+--------------------------------+
2 rows in set (0.01 sec)

mysql>
```

Task4 - PIG query to process XML and store into PIG table

In this section we are going to Load data from HDFS to PIG alias ***StatewiseDistrictwisePhysicalProgress*** using below query:

PIG Queries,

DEFINE XPath org.apache.pig.piggybank.evaluation.xml.XPath;

StatewiseDistrictwisePhysicalProgress = LOAD 'hdfs://localhost:9000/flume_import' USING org.apache.pig.piggybank.storage.XMLLoader('row') as (row:chararray);

Next, iterate over each row and load into alias ***StatewiseDistrictwisePhysicalProgress*** which has schema fields same as XML schema hyphen (-) are replaced with underscore (_)

PhysicalProgress = FOREACH StatewiseDistrictwisePhysicalProgress GENERATE XPath(row, 'row/State_Name') AS State_name,

XPath(row, 'row/District_Name') AS District_name,

XPath(row, 'row/Project_Objectives_IHHL_BPL') AS Project_Objectives_IHHL_BPL,

XPath(row, 'row/Project_Objectives_IHHL_APL') AS Project_Objectives_IHHL_APL,

XPath(row, 'row/Project_Objectives_IHHL_TOTAL') AS Project_Objectives_IHHL_TOTAL,

XPath(row, 'row/Project_Objectives_SCW') AS Project_Objectives_SCW,

XPath(row, 'row/Project_Objectives_Anganwadi_Toilets') AS Project_Objectives_Anganwadi_Toilets,

XPath(row, 'row/Project_Objectives_RSM') AS Project_Objectives_RSM,

XPath(row, 'row/Project_Objectives_PC') AS Project_Objectives_PC,

XPath(row, 'row/Project_Performance-IHHL_BPL') AS Project_Performance_IHHL_BPL,

XPath(row, 'row/Project_Performance-IHHL_APL') AS Project_Performance_IHHL_APL,

XPath(row, 'row/Project_Performance-IHHL_TOTAL') AS Project_Performance_IHHL_TOTAL,

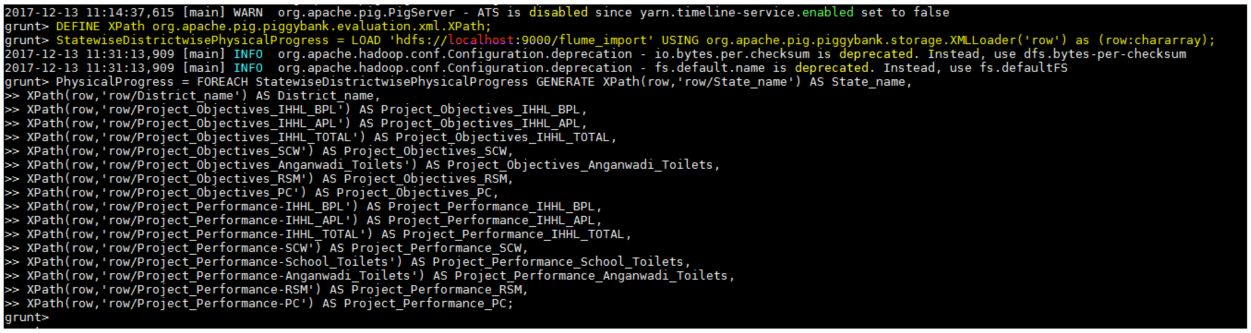
XPath(row, 'row/Project_Performance-SCW') AS Project_Performance_SCW,

XPath(row, 'row/Project_Performance-School_Toilets') AS Project_Performance_School_Toilets,

XPath(row, 'row/Project_Performance-Anganwadi_Toilets') AS Project_Performance_Anganwadi_Toilets,

XPath(row, 'row/Project_Performance-RSM') AS Project_Performance_RSM,

XPath(row, 'row/Project_Performance-PC') AS Project_Performance_PC;



```
2017-12-13 11:14:37,615 [main] WARN  org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt> DEFINE XPath org.apache.pig.piggybank.evaluation.xml.XPath;
grunt> StatewiseDistrictwisePhysicalProgress = LOAD 'hdfs://localhost:9000/flume_import' USING org.apache.pig.piggybank.storage.XMLLoader('row') as (row:chararray);
2017-12-13 11:31:13,909 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-12-13 11:31:13,909 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> PhysicalProgress = FOREACH StatewiseDistrictwisePhysicalProgress GENERATE XPath(row,'row/State_name') AS State_name,
>> XPath(row,'row/District_name') AS District_name,
>> XPath(row,'row/Project_Objectives_IHHL_BPL') AS Project_Objectives_IHHL_BPL,
>> XPath(row,'row/Project_Objectives_IHHL_APL') AS Project_Objectives_IHHL_APL,
>> XPath(row,'row/Project_Objectives_IHHL_TOTAL') AS Project_Objectives_IHHL_TOTAL,
>> XPath(row,'row/Project_Objectives_SCW') AS Project_Objectives_SCW,
>> XPath(row,'row/Project_Objectives_Anganwadi_Toilets') AS Project_Objectives_Anganwadi_Toilets,
>> XPath(row,'row/Project_Objectives_RSM') AS Project_Objectives_RSM,
>> XPath(row,'row/Project_Objectives_PC') AS Project_Objectives_PC,
>> XPath(row,'row/Project_Performance-IHHL_BPL') AS Project_Performance_IHHL_BPL,
>> XPath(row,'row/Project_Performance-IHHL_APL') AS Project_Performance_IHHL_APL,
>> XPath(row,'row/Project_Performance-IHHL_TOTAL') AS Project_Performance_IHHL_TOTAL,
>> XPath(row,'row/Project_Performance-SCW') AS Project_Performance_SCW,
>> XPath(row,'row/Project_Performance-School_Toilets') AS Project_Performance_School_Toilets,
>> XPath(row,'row/Project_Performance-Anganwadi_Toilets') AS Project_Performance_Anganwadi_Toilets,
>> XPath(row,'row/Project_Performance-RSM') AS Project_Performance_RSM,
>> XPath(row,'row/Project_Performance-PC') AS Project_Performance_PC;
grunt>
```

Task5 – Find the districts who achieved 100 percent objective in BPL cards

Filter the records by ***Project_Objectives_IHHL_BPL*** is equal to ***Project_Performance_IHHL_BPL***

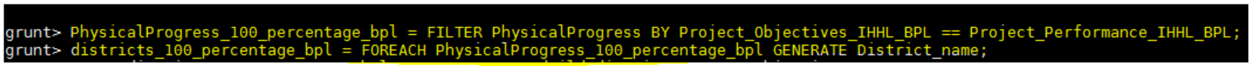
PhysicalProgress_100_percentage_bpl = FILTER PhysicalProgress BY Project_Objectives_IHHL_BPL == Project_Performance_IHHL_BPL;

Select only District_Name column,

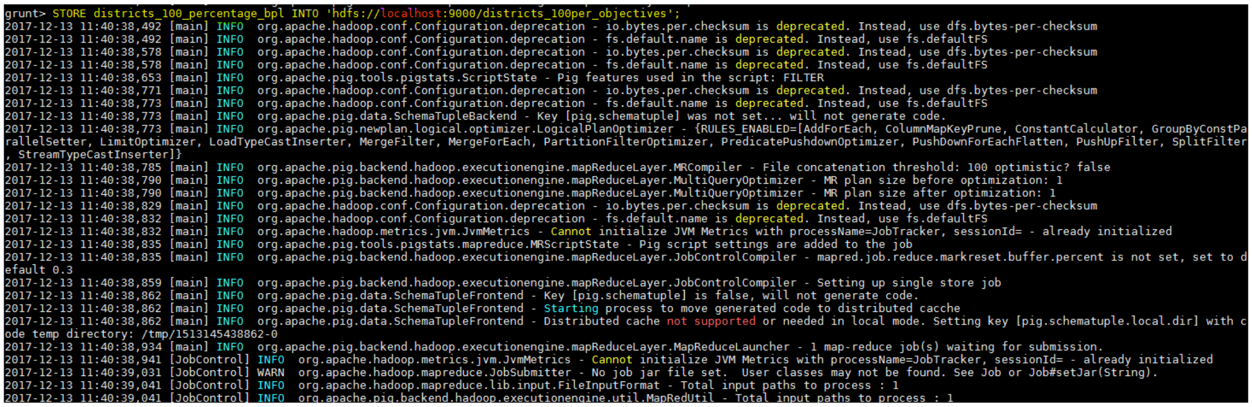
districts_100_percentage_bpl = FOREACH PhysicalProgress_100_percentage_bpl GENERATE District_name;

Now store the data we received from the PIG alias ***districts_100_percentage_bpl*** into the HDFS location where we created at the Task2

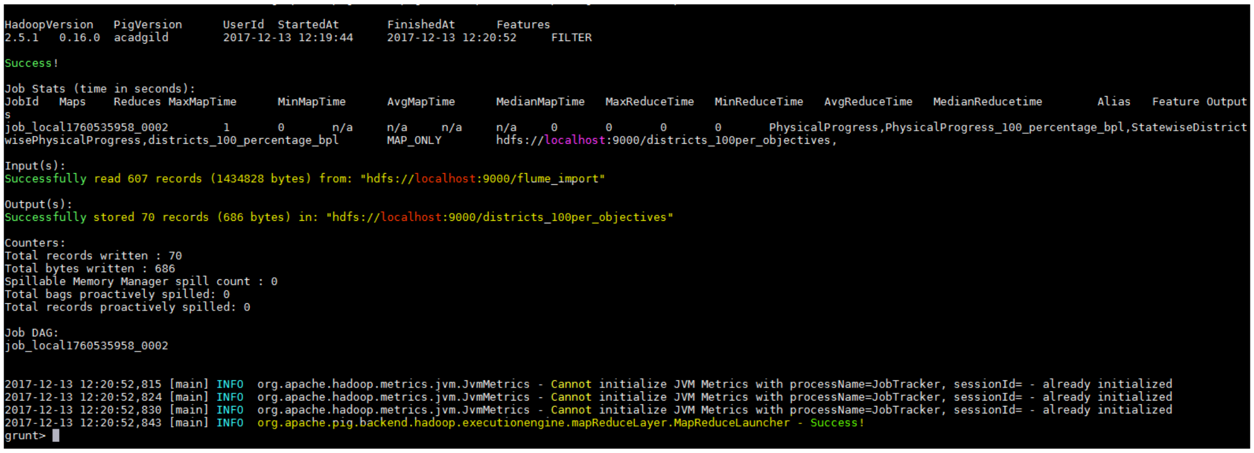
STORE districts_100_percentage_bpl INTO 'hdfs://localhost:9000/districts_100per_objectives';



```
grunt> PhysicalProgress_100_percentage_bpl = FILTER PhysicalProgress BY Project_Objectives_IHHL_BPL == Project_Performance_IHHL_BPL;
grunt> districts_100_percentage_bpl = FOREACH PhysicalProgress_100_percentage_bpl GENERATE District_name;
```



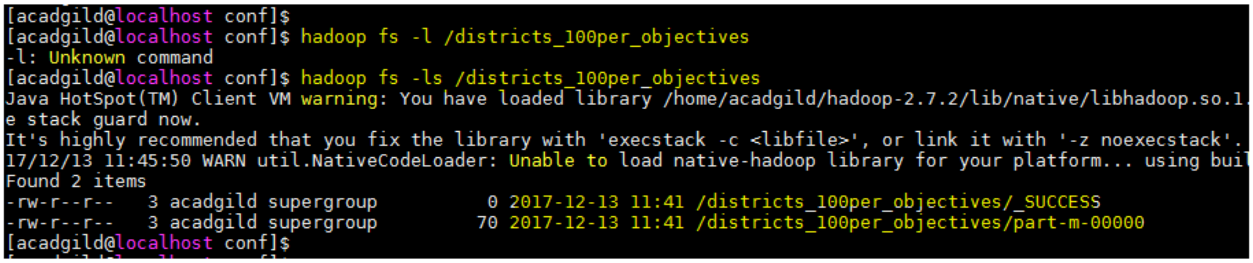
```
grunt> STORE districts_100_percentage_bpl INTO 'hdfs://localhost:9000/districts_100per_objectives';
2017-12-13 11:40:38,492 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-12-13 11:40:38,492 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-12-13 11:40:38,578 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-12-13 11:40:38,578 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-12-13 11:40:38,653 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2017-12-13 11:40:38,771 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-12-13 11:40:38,773 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-12-13 11:40:38,773 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-12-13 11:40:38,773 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstPa
rallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter
, StreamTypeCastInserter]}
2017-12-13 11:40:38,785 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2017-12-13 11:40:38,790 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2017-12-13 11:40:38,790 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2017-12-13 11:40:38,829 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-12-13 11:40:38,832 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-12-13 11:40:38,832 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-12-13 11:40:38,835 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2017-12-13 11:40:38,835 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to d
efault 0.3
2017-12-13 11:40:38,859 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2017-12-13 11:40:38,862 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2017-12-13 11:40:38,862 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2017-12-13 11:40:38,862 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Distributed cache not supported or needed in local mode. Setting key [pig.schematuple.local.dir] with c
ode temp directory: /tmp/1513145438862-0
2017-12-13 11:40:38,934 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2017-12-13 11:40:38,941 [JobControl] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-12-13 11:40:39,031 [JobControl] WARN  org.apache.hadoop.mapreduce.JobSubmitter - No job jar file set.  User classes may not be found. See Job or Job#setJar(String).
2017-12-13 11:40:39,041 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-12-13 11:40:39,041 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
```



```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.5.1   0.16.0  acadgild        2017-12-13 12:19:44     2017-12-13 12:20:52     FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   Feature Output
s
job_local1760535958_0002        1       0       n/a     n/a     n/a     n/a     0       0       0       0       PhysicalProgress,PhysicalProgress_100_percentage_bpl,StatewiseDistrict
wisePhysicalProgress,districts_100_percentage_bpl       MAP_ONLY        hdfs://localhost:9000/districts_100per_objectives,

Input(s):
Successfully read 607 records (1434828 bytes) from: "hdfs://localhost:9000/flume_import"

Output(s):
Successfully stored 70 records (686 bytes) in: "hdfs://localhost:9000/districts_100per_objectives"

Counters:
Total records written : 70
Total bytes written : 686
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1760535958_0002


2017-12-13 12:20:52,815 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-12-13 12:20:52,824 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-12-13 12:20:52,830 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-12-13 12:20:52,843 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> 
```

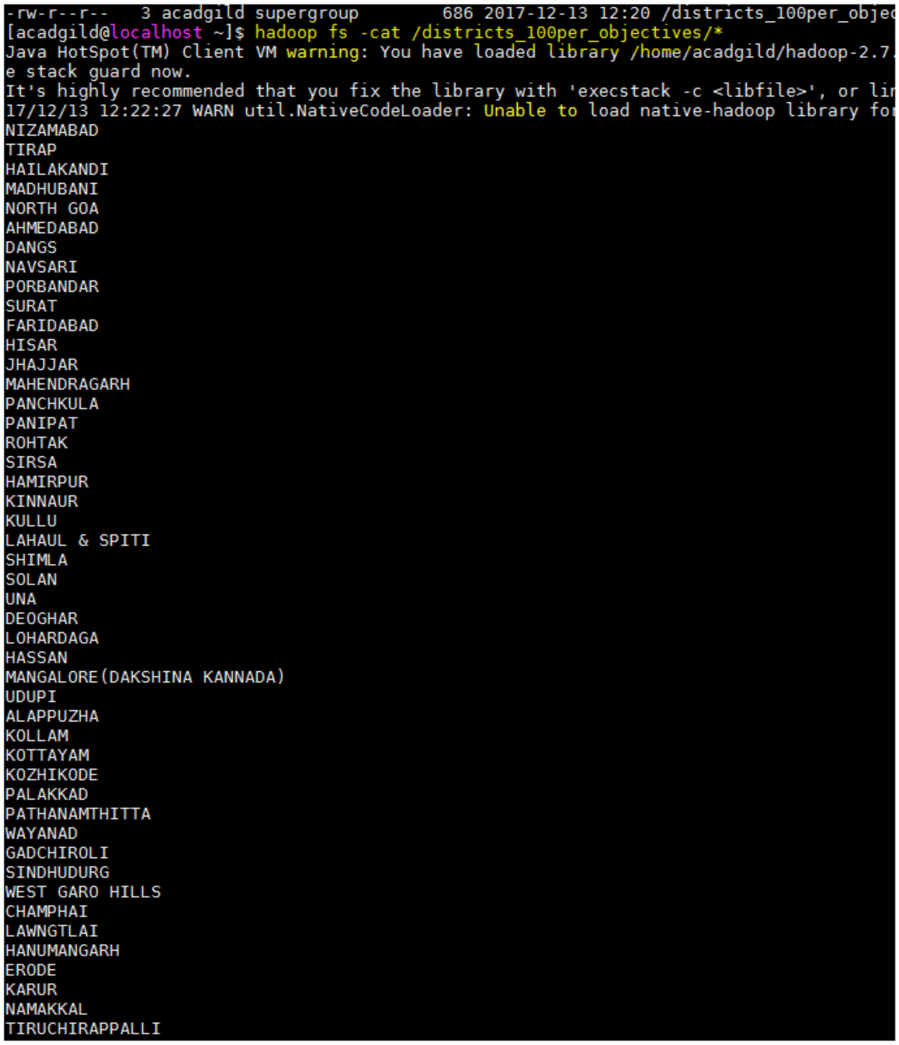
Task6 – Verifying the stored results in the HDFS

hadoop fs -ls /districts_100per_objectives

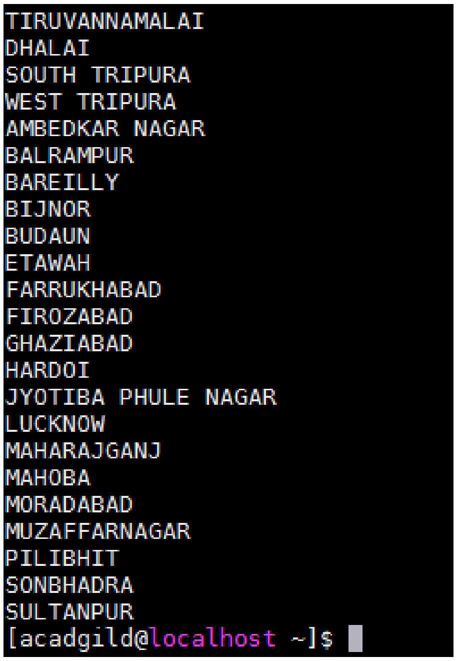


```
[acadgild@localhost conf]$
[acadgild@localhost conf]$ hadoop fs -l /districts_100per_objectives
-l: Unknown command
[acadgild@localhost conf]$ hadoop fs -ls /districts_100per_objectives
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/12/13 11:45:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bui
Found 2 items
-rw-r--r--   3 acadgild supergroup          0 2017-12-13 11:41 /districts_100per_objectives/_SUCCESS
-rw-r--r--   3 acadgild supergroup         70 2017-12-13 11:41 /districts_100per_objectives/part-m-00000
[acadgild@localhost conf]$
```

hadoop fs -cat /districts_100per_objectives/*



```
-rw-r--r--   3 acadgild supergroup        686 2017-12-13 12:20 /districts_100per_objec
[acadgild@localhost ~]$ hadoop fs -cat /districts_100per_objectives/*
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or lin
17/12/13 12:22:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
NIZAMABAD
TIRAP
HAILAKANDI
MADHUBANI
NORTH GOA
AHMEDABAD
DANGS
NAVSARI
PORBANDAR
SURAT
FARIDABAD
HISAR
JHAJJAR
MAHENDRAGARH
PANCHKULA
PANIPAT
ROHTAK
SIRSA
HAMIRPUR
KINNAUR
KULLU
LAHAUL & SPITI
SHIMLA
SOLAN
UNA
DEOGHAR
LOHARDAGA
HASSAN
MANGALORE(DAKSHINA KANNADA)
UDUPI
ALAPPUZHA
KOLLAM
KOTTAYAM
KOZHIKODE
PALAKKAD
PATHANAMTHITTA
WAYANAD
GADCHIROLI
SINDHUDURG
WEST GARO HILLS
CHAMPHAI
LAWNGTLAI
HANUMANGARH
ERODE
KARUR
NAMAKKAL
TIRUCHIRAPPALLI
```

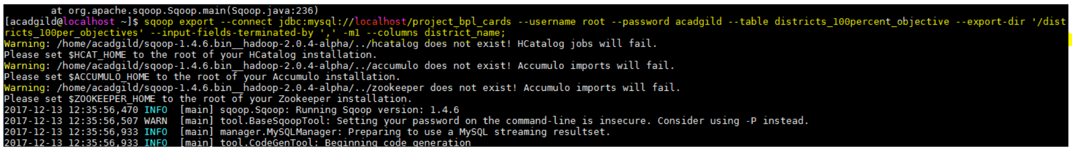


```
TIRUVANNAMALAI
DHALAI
SOUTH TRIPURA
WEST TRIPURA
AMBEDKAR NAGAR
BALRAMPUR
BAREILLY
BIJNOR
BUDAUN
ETAWAH
FARRUKHABAD
FIROZABAD
GHAZIABAD
HARDOI
JYOTIBA PHULE NAGAR
LUCKNOW
MAHARAJGANJ
MAHOBA
MORADABAD
MUZAFFARNAGAR
PILIBHIT
SONBHADRA
SULTANPUR
[acadgild@localhost ~]$
```

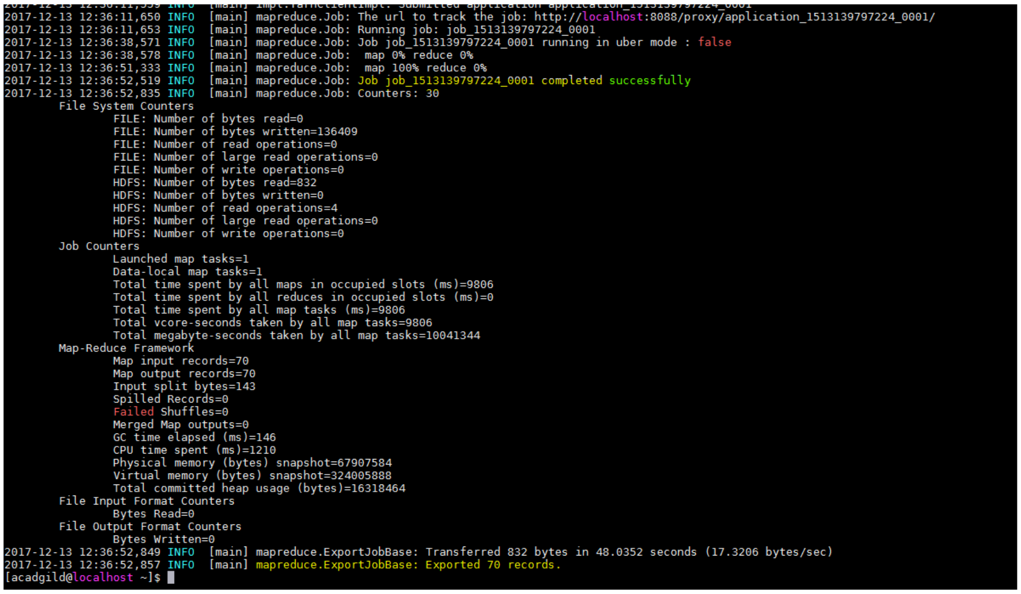
Task7 – Export the results into mysql using sqoop

Sqoop command to export,

sqoop export --connect jdbc:mysql://localhost/project_bpl_cards --username root --password acadgild --table districts_100percent_objective --export-dir '/districts_100per_objectives' --input-fields-terminated-by ',' -m1 --columns district_name



```
        at org.apache.sqoop.Sqoop.main(Sqoop.java:236)
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/project_bpl_cards --username root --password acadgild --table districts_100percent_objective --export-dir '/dist
ricts_100per_objectives' --input-fields-terminated-by ',' -m1 --columns district_name;
Warning: /home/acadgild/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acadgild/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/acadgild/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.
2017-12-13 12:35:56,470 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-12-13 12:35:56,507 WARN  [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-12-13 12:35:56,933 INFO  [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-12-13 12:35:56,933 INFO  [main] tool.CodeGenTool: Beginning code generation
```

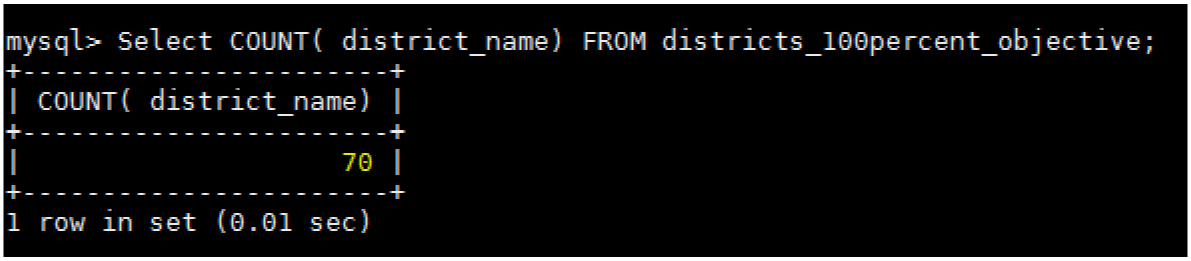


```
2017-12-13 12:36:11,359 INFO  [main] impl.YarnClientImpl: Submitted application application_1513139797224_0001
2017-12-13 12:36:11,650 INFO  [main] mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1513139797224_0001/
2017-12-13 12:36:11,653 INFO  [main] mapreduce.Job: Running job: job_1513139797224_0001
2017-12-13 12:36:38,571 INFO  [main] mapreduce.Job: Job job_1513139797224_0001 running in uber mode : false
2017-12-13 12:36:38,578 INFO  [main] mapreduce.Job:  map 0% reduce 0%
2017-12-13 12:36:51,333 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2017-12-13 12:36:52,519 INFO  [main] mapreduce.Job: Job job_1513139797224_0001 completed successfully
2017-12-13 12:36:52,835 INFO  [main] mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=136409
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=832
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=9806
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=9806
                Total vcore-seconds taken by all map tasks=9806
                Total megabyte-seconds taken by all map tasks=10041344
        Map-Reduce Framework
                Map input records=70
                Map output records=70
                Input split bytes=143
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=146
                CPU time spent (ms)=1210
                Physical memory (bytes) snapshot=67907584
                Virtual memory (bytes) snapshot=324005888
                Total committed heap usage (bytes)=16318464
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
2017-12-13 12:36:52,849 INFO  [main] mapreduce.ExportJobBase: Transferred 832 bytes in 48.0352 seconds (17.3206 bytes/sec)
2017-12-13 12:36:52,857 INFO  [main] mapreduce.ExportJobBase: Exported 70 records.
[acadgild@localhost ~]$
```

Task8 – verify the data exported to mysql

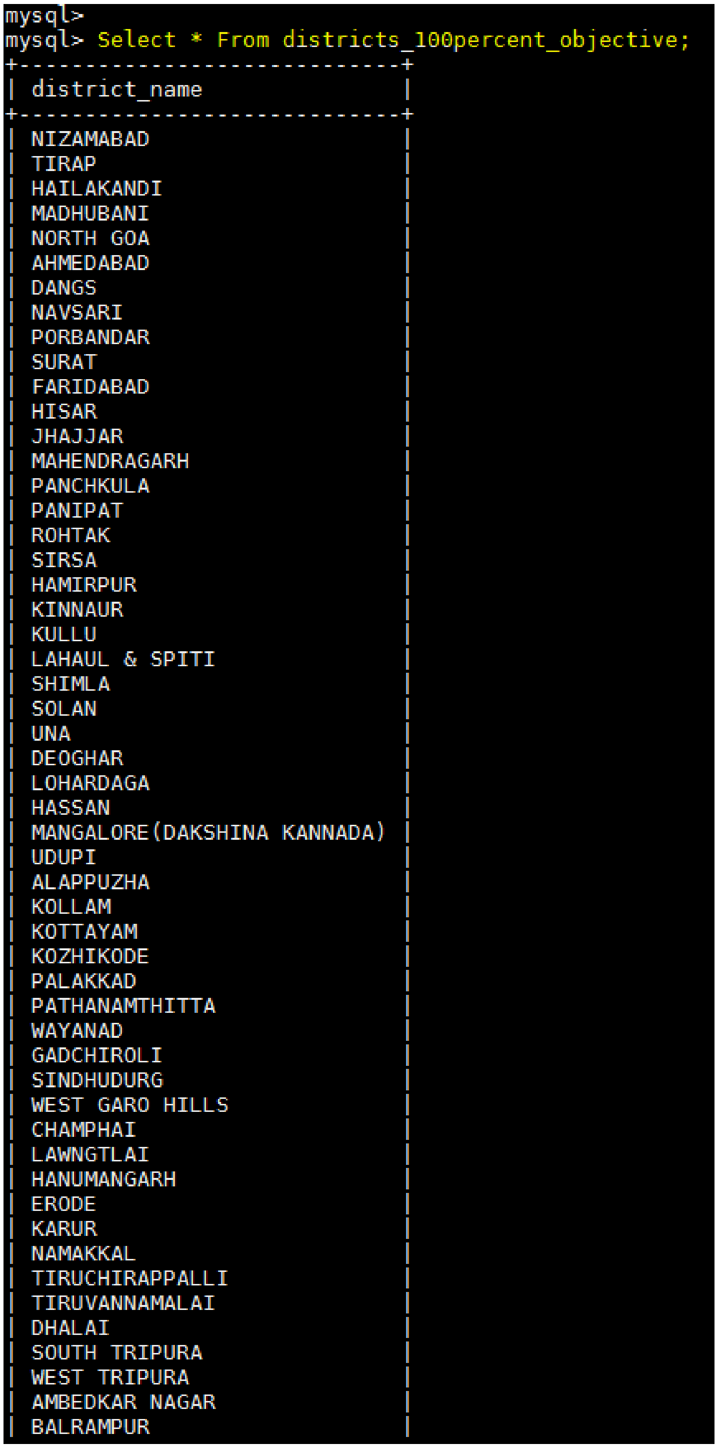
Use the following command in mysql to verify results in mysql

Select COUNT(district_name) FROM districts_100percent_objective;

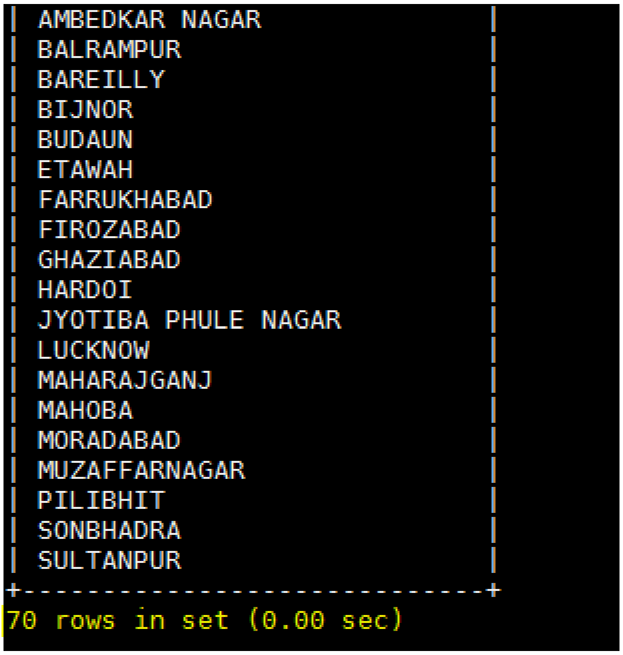


```
mysql> Select COUNT( district_name) FROM districts_100percent_objective;
+------------------------+
| COUNT( district_name) |
+------------------------+
|                     70 |
+------------------------+
1 row in set (0.01 sec)
```

select * from districts_100percent_objective;



```
mysql>
mysql> Select * From districts_100percent_objective;
+-----------------------------+
| district_name               |
+-----------------------------+
| NIZAMABAD                   |
| TIRAP                       |
| HAILAKANDI                  |
| MADHUBANI                   |
| NORTH GOA                   |
| AHMEDABAD                   |
| DANGS                       |
| NAVSARI                     |
| PORBANDAR                   |
| SURAT                       |
| FARIDABAD                   |
| HISAR                       |
| JHAJJAR                     |
| MAHENDRAGARH                |
| PANCHKULA                   |
| PANIPAT                     |
| ROHTAK                      |
| SIRSA                       |
| HAMIRPUR                    |
| KINNAUR                     |
| KULLU                       |
| LAHAUL & SPITI              |
| SHIMLA                      |
| SOLAN                       |
| UNA                         |
| DEOGHAR                     |
| LOHARDAGA                   |
| HASSAN                      |
| MANGALORE(DAKSHINA KANNADA) |
| UDUPI                       |
| ALAPPUZHA                   |
| KOLLAM                      |
| KOTTAYAM                    |
| KOZHIKODE                   |
| PALAKKAD                    |
| PATHANAMTHITTA              |
| WAYANAD                     |
| GADCHIROLI                  |
| SINDHUDURG                  |
| WEST GARO HILLS             |
| CHAMPHAI                    |
| LAWNGTLAI                   |
| HANUMANGARH                 |
| ERODE                       |
| KARUR                       |
| NAMAKKAL                    |
| TIRUCHIRAPPALLI             |
| TIRUVANNAMALAI              |
| DHALAI                      |
| SOUTH TRIPURA               |
| WEST TRIPURA                |
| AMBEDKAR NAGAR              |
| BALRAMPUR                   |
```



```
| AMBEDKAR NAGAR              |
| BALRAMPUR                   |
| BAREILLY                    |
| BIJNOR                      |
| BUDAUN                      |
| ETAWAH                      |
| FARRUKHABAD                 |
| FIROZABAD                   |
| GHAZIABAD                   |
| HARDOI                      |
| JYOTIBA PHULE NAGAR         |
| LUCKNOW                     |
| MAHARAJGANJ                 |
| MAHOBA                      |
| MORADABAD                   |
| MUZAFFARNAGAR               |
| PILIBHIT                    |
| SONBHADRA                   |
| SULTANPUR                   |
+-----------------------------+
70 rows in set (0.00 sec)
```

Thus, as per the problem statement 1, we have successfully exported the result from HDFS to mysql database **project_bpl_cards** and into the table **districts_100percent_objective.**

Problem statemet2 - Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards. Export the results to MySQL using Sqoop.

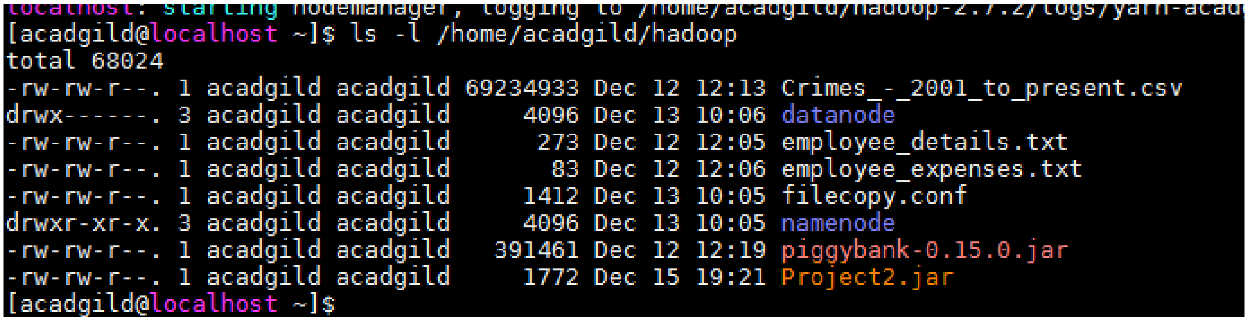
Task1 – Create a PIG UDF using Java

Create a Maven project **StateAnalysis** and Write a Java class **StateAnalysis** in eclipse which will filter those tuples for which 80 percent objective in BPL cards are achieved. The logic put in exec method is value of **Project_Performance_IHHL_BPL** is equal to more than 80% of **Project_Objectives_IHHL_BPL.**

Java code

```
package StateAnalysis;
import java.io.IOException;
import org.apache.pig.FilterFunc;
import org.apache.pig.backend.executionengine.ExecException;
import org.apache.pig.data.Tuple;

public class StateAnalysis extends FilterFunc
{
        @Override
        public Boolean exec(Tuple input) throws IOException
        {
                try
                {
                        if(input == null || input.size() == 0)
                        {
                                return false;
                        }
                        Object valueTuple = input.get(0);
                        if (valueTuple instanceof Tuple)
                        {
                                Object value1 = ((Tuple) valueTuple).get(0);
                                Object value2 = ((Tuple) valueTuple).get(1);
                                long objective_value = Long.valueOf((String) value1);
                                long performance_value = Long.valueOf((String) value2);

                        if(performance_value>objective_value*80/100)
                        {
                                return true;
                        }

                        }
                }
                catch(ExecException ee)
                {
                        throw ee;
                }
                return false;
        }
}
```

Compile this project and Export the project as .jar file to the acadgild local file system. Here we named the jar file as ***Project2.jar.***



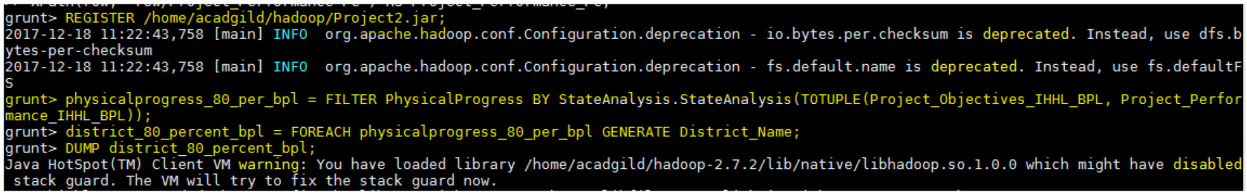
```
localhost: starting nodemanager, logging to /home/acadgild/hadoop-2.7.2/logs/yarn-acad
[acadgild@localhost ~]$ ls -l /home/acadgild/hadoop
total 68024
-rw-rw-r--. 1 acadgild acadgild 69234933 Dec 12 12:13 Crimes_-_2001_to_present.csv
drwx------. 3 acadgild acadgild     4096 Dec 13 10:06 datanode
-rw-rw-r--. 1 acadgild acadgild      273 Dec 12 12:05 employee_details.txt
-rw-rw-r--. 1 acadgild acadgild       83 Dec 12 12:06 employee_expenses.txt
-rw-rw-r--. 1 acadgild acadgild     1412 Dec 13 10:05 filecopy.conf
drwxr-xr-x. 3 acadgild acadgild     4096 Dec 13 10:05 namenode
-rw-rw-r--. 1 acadgild acadgild   391461 Dec 12 12:19 piggybank-0.15.0.jar
-rw-rw-r--. 1 acadgild acadgild     1772 Dec 15 19:21 Project2.jar
[acadgild@localhost ~]$
```

Task2 - Write PIG query to find out the districts who achieved 80 percent objective in BPL cards

REGISTER /home/acadgild/hadoop/Project2.jar;

Next, using the UDF filter those tuple for which **Project_Performance_IHHL_BPL** is equal to more than 80% of **Project_Objectives_IHHL_BPL,**

physicalprogress_80_per_bpl = FILTER PhysicalProgress BY StateAnalysis.StateAnalysis(TOTUPLE(Project_Objectives_IHHL_BPL, Project_Performance_IHHL_BPL));



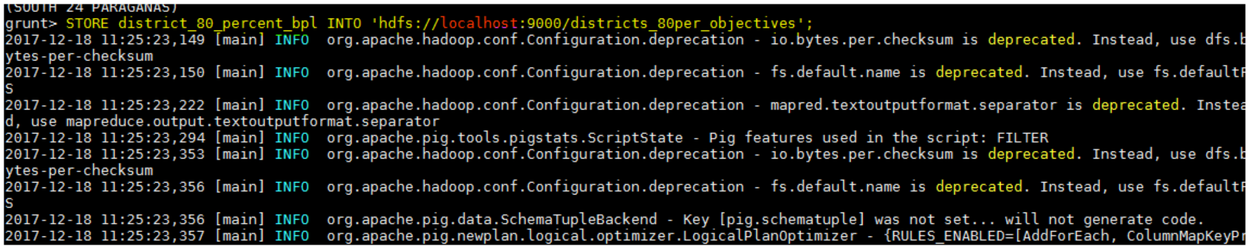
```
grunt> REGISTER /home/acadgild/hadoop/Project2.jar;
2017-12-18 11:22:43,758 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b
ytes-per-checksum
2017-12-18 11:22:43,758 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultF
S
grunt> physicalprogress_80_per_bpl = FILTER PhysicalProgress BY StateAnalysis.StateAnalysis(TOTUPLE(Project_Objectives_IHHL_BPL, Project_Perfor
mance_IHHL_BPL));
grunt> district_80_percent_bpl = FOREACH physicalprogress_80_per_bpl GENERATE District_Name;
grunt> DUMP district_80_percent_bpl;
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled
 stack guard. The VM will try to fix the stack guard now.
```

Next, select only **District_Name** field using command below:

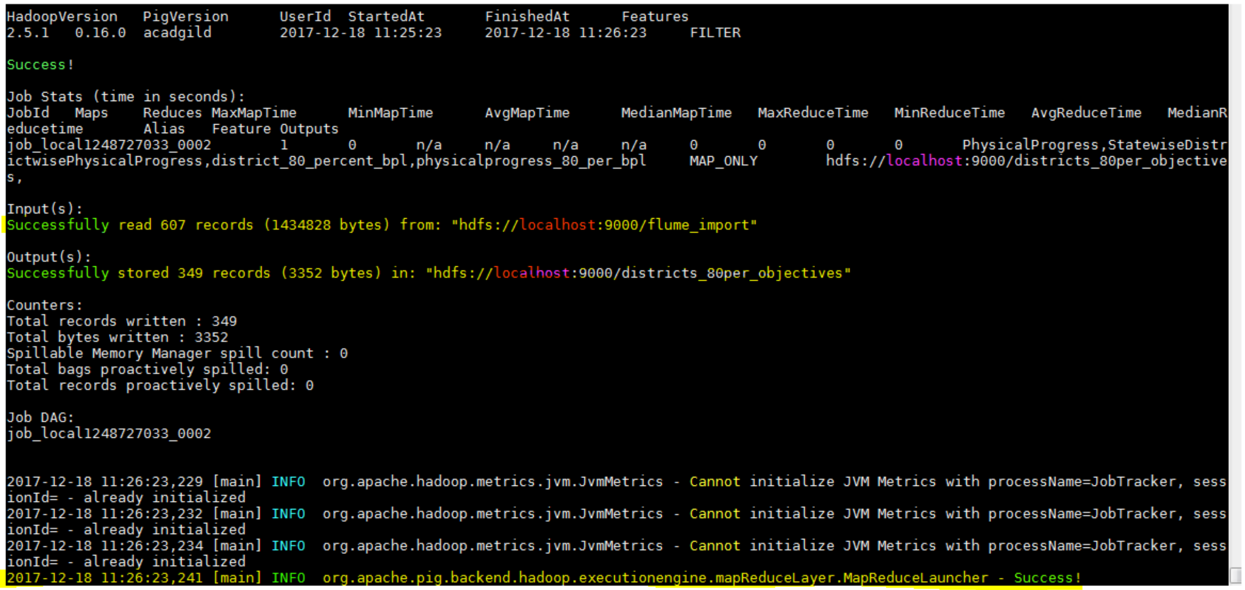
district_80_percent_bpl = FOREACH physicalprogress_80_per_bpl GENERATE District_Name;

Now store the data we received from the PIG alias ***district_80_percent_bpl*** into the HDFS location where we created at the Task2

STORE district_80_percent_bpl INTO 'hdfs://localhost:9000/districts_having_80percent_objectives';



```
(SOUTH 24 PARAGANAS)
grunt> STORE district_80_percent_bpl INTO 'hdfs://localhost:9000/districts_80per_objectives';
2017-12-18 11:25:23,149 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b
ytes-per-checksum
2017-12-18 11:25:23,150 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultF
S
2017-12-18 11:25:23,222 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instea
d, use mapreduce.output.textoutputformat.separator
2017-12-18 11:25:23,294 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2017-12-18 11:25:23,353 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.b
ytes-per-checksum
2017-12-18 11:25:23,356 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultF
S
2017-12-18 11:25:23,356 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-12-18 11:25:23,357 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPr
```

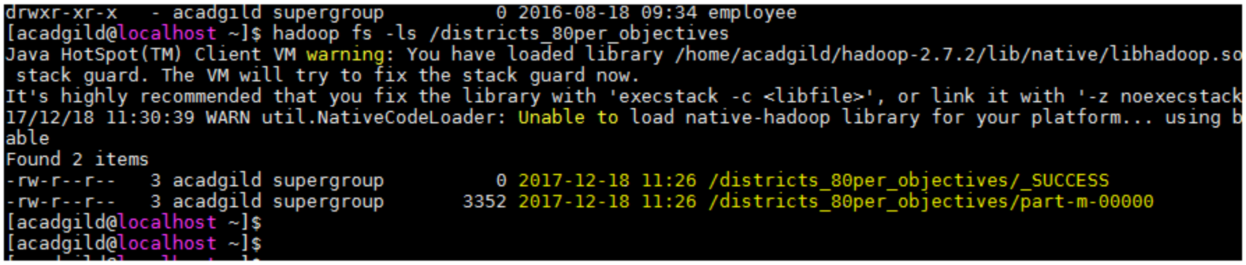


```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.5.1   0.16.0  acadgild        2017-12-18 11:25:23     2017-12-18 11:26:23     FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianR
educetime       Alias   Feature Outputs
job_local1248727033_0002        1       0       n/a     n/a     n/a     n/a     0       0       0       0       PhysicalProgress,StatewiseDistr
ictwisePhysicalProgress,district_80_percent_bpl,physicalprogress_80_per_bpl     MAP_ONLY        hdfs://localhost:9000/districts_80per_objective
s,

Input(s):
Successfully read 607 records (1434828 bytes) from: "hdfs://localhost:9000/flume_import"

Output(s):
Successfully stored 349 records (3352 bytes) in: "hdfs://localhost:9000/districts_80per_objectives"

Counters:
Total records written : 349
Total bytes written : 3352
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1248727033_0002


2017-12-18 11:26:23,229 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sess
ionId= - already initialized
2017-12-18 11:26:23,232 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sess
ionId= - already initialized
2017-12-18 11:26:23,234 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sess
ionId= - already initialized
2017-12-18 11:26:23,241 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

Task2 – verify the result stored in the HDFS

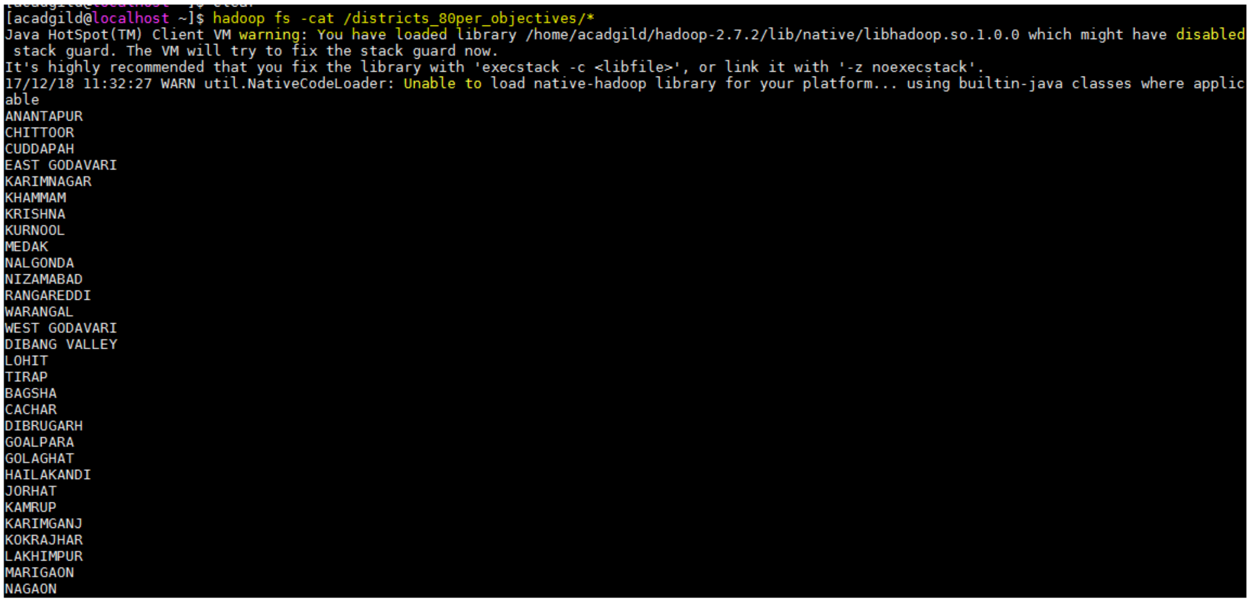
The following command shows that folders are created under districts_having_100percent_objectives,
hadoop fs -ls / districts_80per_objectives
hadoop fs –ls / districts_80per_objectives/part-m-00000

The output file has been generated in the HDFS location,

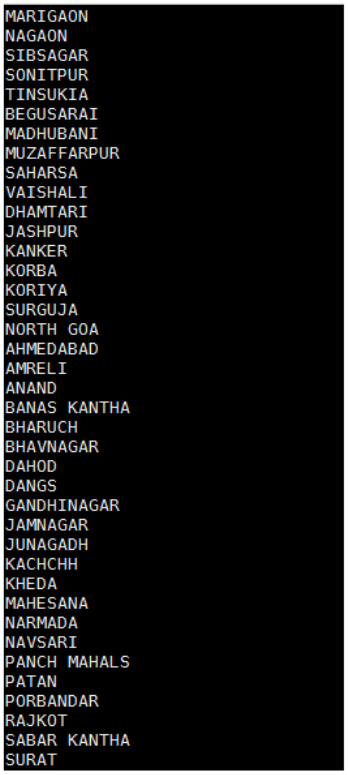


```
drwxr-xr-x   - acadgild supergroup          0 2016-08-18 09:34 employee
[acadgild@localhost ~]$ hadoop fs -ls /districts_80per_objectives
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so
 stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack
17/12/18 11:30:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using b
able
Found 2 items
-rw-r--r--   3 acadgild supergroup          0 2017-12-18 11:26 /districts_80per_objectives/_SUCCESS
-rw-r--r--   3 acadgild supergroup       3352 2017-12-18 11:26 /districts_80per_objectives/part-m-00000
[acadgild@localhost ~]$
[acadgild@localhost ~]$
```

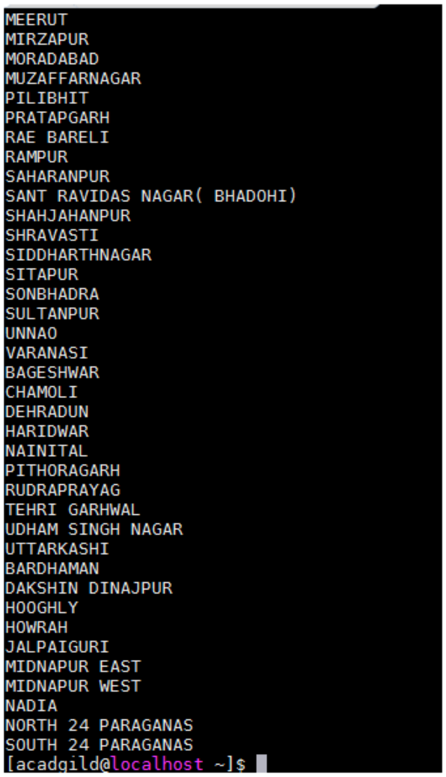
hadoop fs -cat /districts_80per_objectives/*



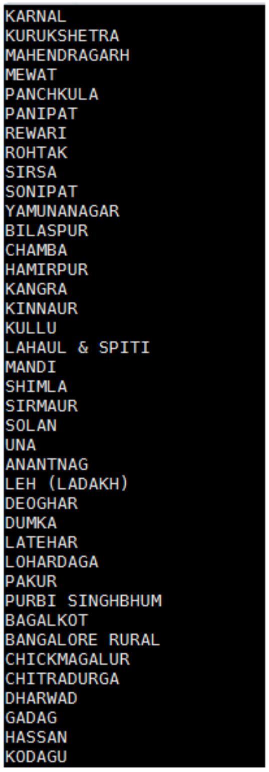
```
[acadgild@localhost ~]$ hadoop fs -cat /districts_80per_objectives/*
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled
 stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/12/18 11:32:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applic
able
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
```



```
MARIGAON
NAGAON
SIBSAGAR
SONITPUR
TINSUKIA
BEGUSARAI
MADHUBANI
MUZAFFARPUR
SAHARSA
VAISHALI
DHAMTARI
JASHPUR
KANKER
KORBA
KORIYA
SURGUJA
NORTH GOA
AHMEDABAD
AMRELI
ANAND
BANAS KANTHA
BHARUCH
BHAVNAGAR
DAHOD
DANGS
GANDHINAGAR
JAMNAGAR
JUNAGADH
KACHCHH
KHEDA
MAHESANA
NARMADA
NAVSARI
PANCH MAHALS
PATAN
PORBANDAR
RAJKOT
SABAR KANTHA
SURAT
```



```
MEERUT
MIRZAPUR
MORADABAD
MUZAFFARNAGAR
PILIBHIT
PRATAPGARH
RAE BARELI
RAMPUR
SAHARANPUR
SANT RAVIDAS NAGAR( BHADOHI)
SHAHJAHANPUR
SHRAVASTI
SIDDHARTHNAGAR
SITAPUR
SONBHADRA
SULTANPUR
UNNAO
VARANASI
BAGESHWAR
CHAMOLI
DEHRADUN
HARIDWAR
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
[acadgild@localhost ~]$
```



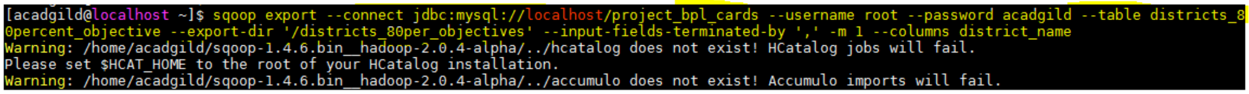
```
KARNAL
KURUKSHETRA
MAHENDRAGARH
MEWAT
PANCHKULA
PANIPAT
REWARI
ROHTAK
SIRSA
SONIPAT
YAMUNANAGAR
BILASPUR
CHAMBA
HAMIRPUR
KANGRA
KINNAUR
KULLU
LAHAUL & SPITI
MANDI
SHIMLA
SIRMAUR
SOLAN
UNA
ANANTNAG
LEH (LADAKH)
DEOGHAR
DUMKA
LATEHAR
LOHARDAGA
PAKUR
PURBI SINGHBHUM
BAGALKOT
BANGALORE RURAL
CHICKMAGALUR
CHITRADURGA
DHARWAD
GADAG
HASSAN
KODAGU
```

Task4 – Export the results into mysql table using sqoop command,

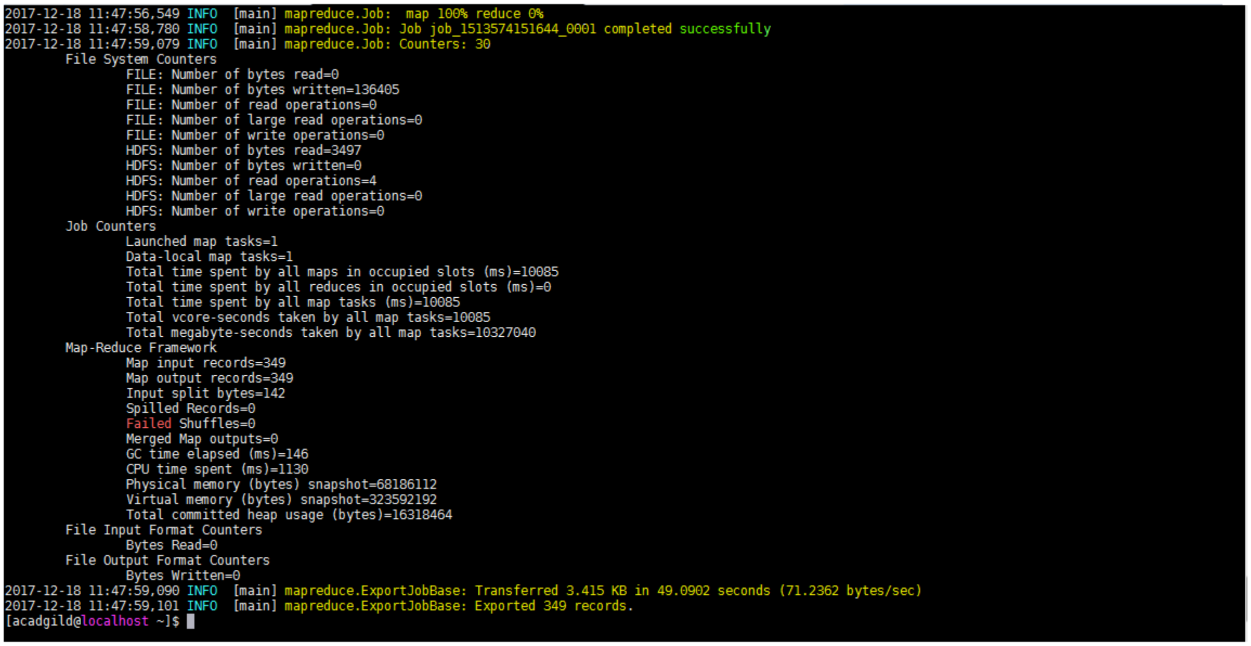
In this task we are going use the sqoop to export the desired output stored in the HDFS location **hdfs://localhost:9000/districts_having_80percent_objectives** to the mysql table **districts_having_80percent_objectives** we created in the database **project_bpl_cards**

Sqoop command,

sqoop export --connect jdbc:mysql://localhost/project_bpl_cards --username root --password acadgild --table districts_80percent_objective --export-dir '/districts_80per_objectives' --input-fields-terminated-by ',' -m 1 --columns district_name



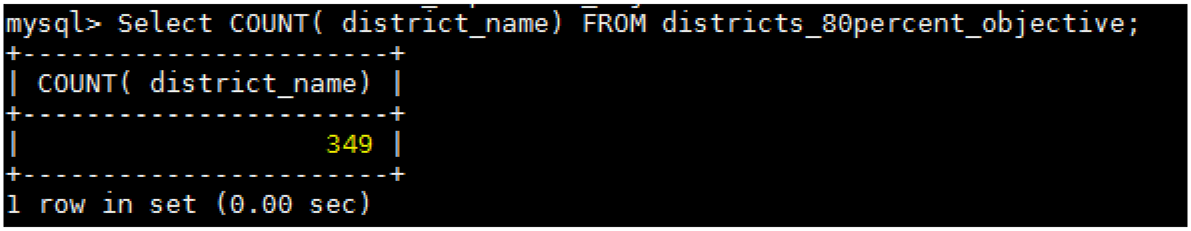
```
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/project_bpl_cards --username root --password acadgild --table districts_8
0percent_objective --export-dir '/districts_80per_objectives' --input-fields-terminated-by ',' -m 1 --columns district_name
Warning: /home/acadgild/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acadgild/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/../accumulo does not exist! Accumulo imports will fail.
```



```
2017-12-18 11:47:56,549 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2017-12-18 11:47:58,780 INFO  [main] mapreduce.Job: Job job_1513574151644_0001 completed successfully
2017-12-18 11:47:59,079 INFO  [main] mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=136405
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=3497
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=10085
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=10085
                Total vcore-seconds taken by all map tasks=10085
                Total megabyte-seconds taken by all map tasks=10327040
        Map-Reduce Framework
                Map input records=349
                Map output records=349
                Input split bytes=142
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=146
                CPU time spent (ms)=1130
                Physical memory (bytes) snapshot=68186112
                Virtual memory (bytes) snapshot=323592192
                Total committed heap usage (bytes)=16318464
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
2017-12-18 11:47:59,090 INFO  [main] mapreduce.ExportJobBase: Transferred 3.415 KB in 49.0902 seconds (71.2362 bytes/sec)
2017-12-18 11:47:59,101 INFO  [main] mapreduce.ExportJobBase: Exported 349 records.
[acadgild@localhost ~]$ 
```

Task5 – Verify the result in the mysql

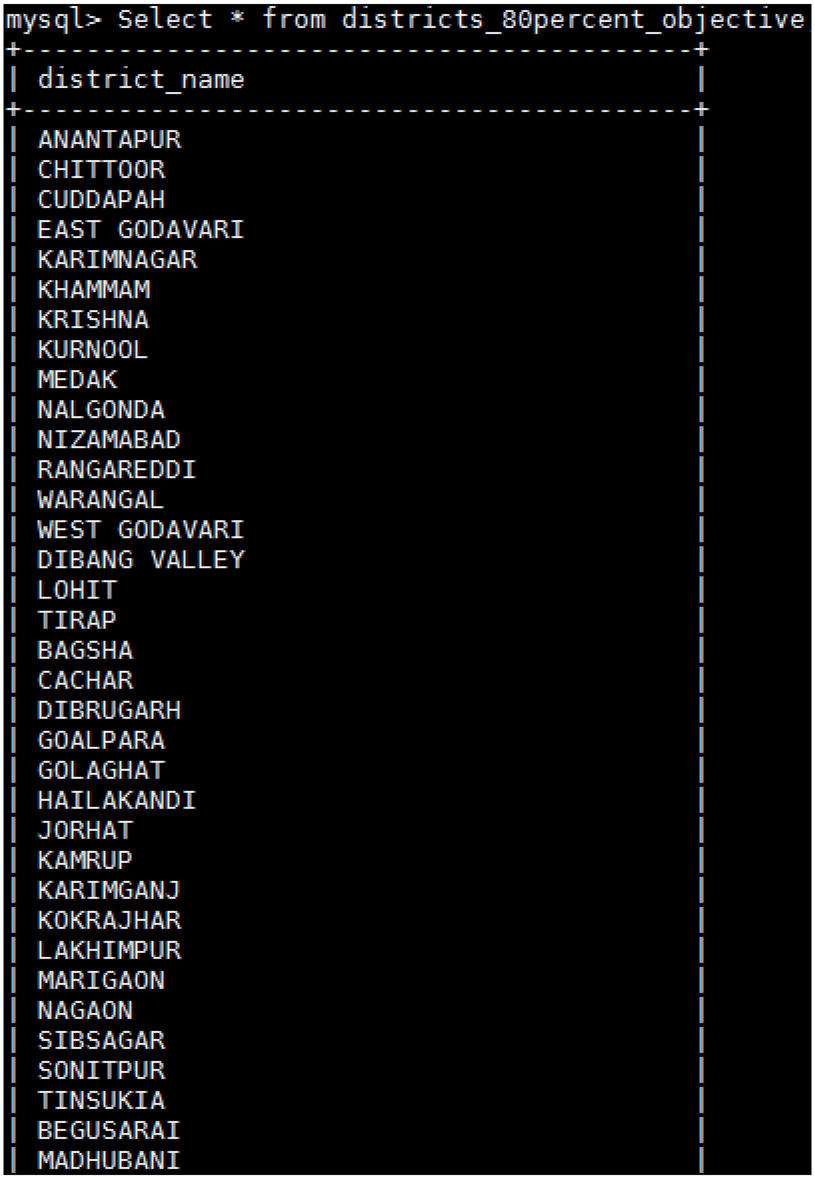
Select COUNT(district_name) FROM districts_80percent_objective;



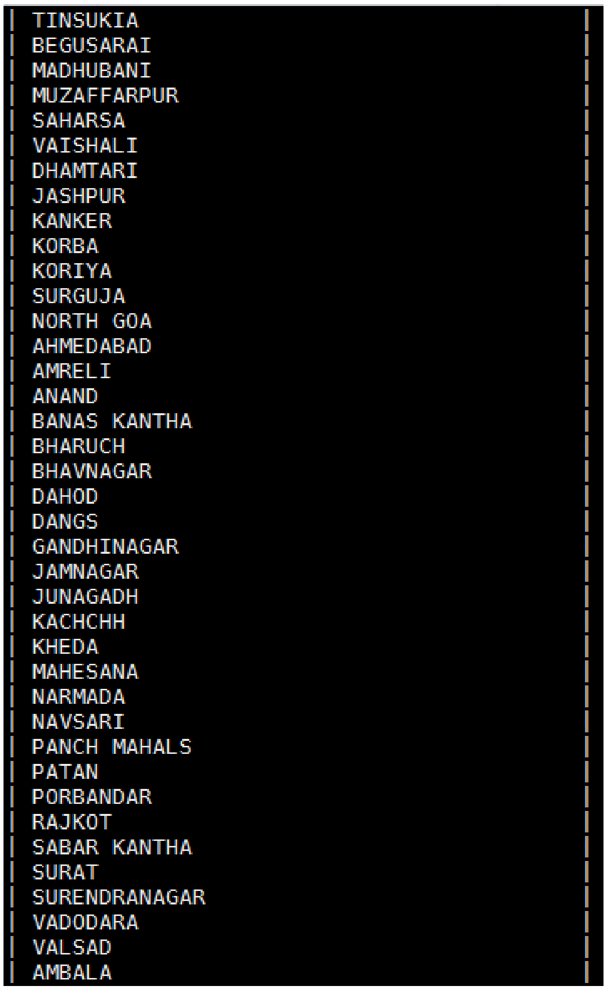
```
mysql> Select COUNT( district_name) FROM districts_80percent_objective;
+-----------------------+
| COUNT( district_name) |
+-----------------------+
|                   349 |
+-----------------------+
1 row in set (0.00 sec)
```

Now, verify the data present in the table,

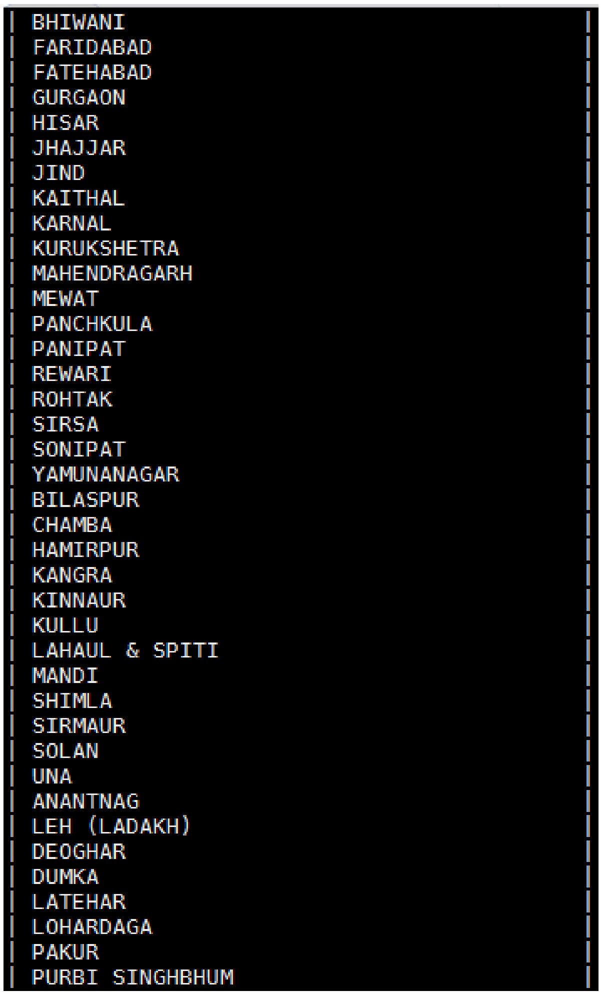
Select * from districts_80percent_objective;



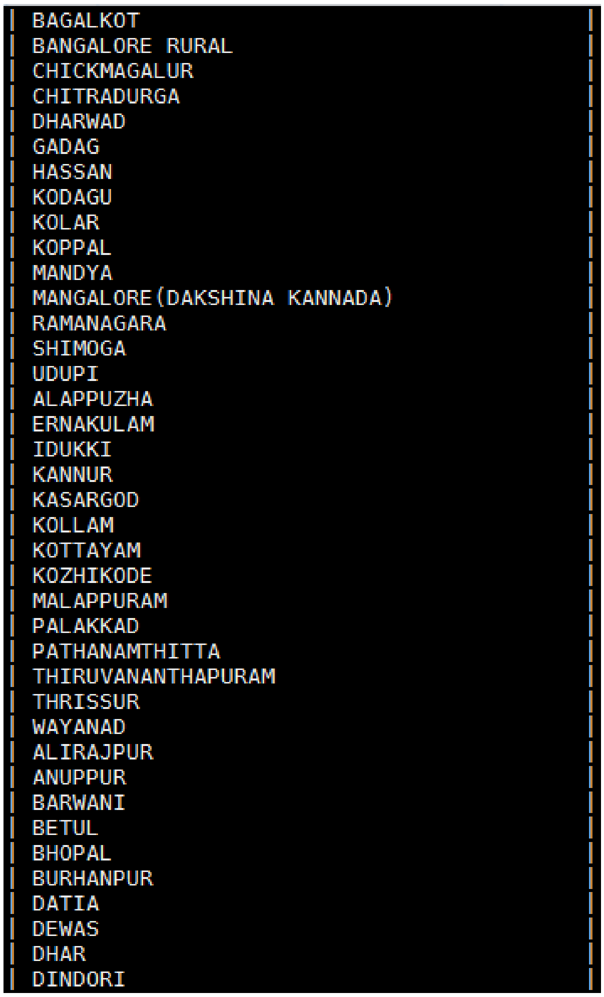
```
mysql> Select * from districts_80percent_objective
+-------------------------------------------+
| district_name                             |
+-------------------------------------------+
| ANANTAPUR                                 |
| CHITTOOR                                  |
| CUDDAPAH                                  |
| 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                                 |
```



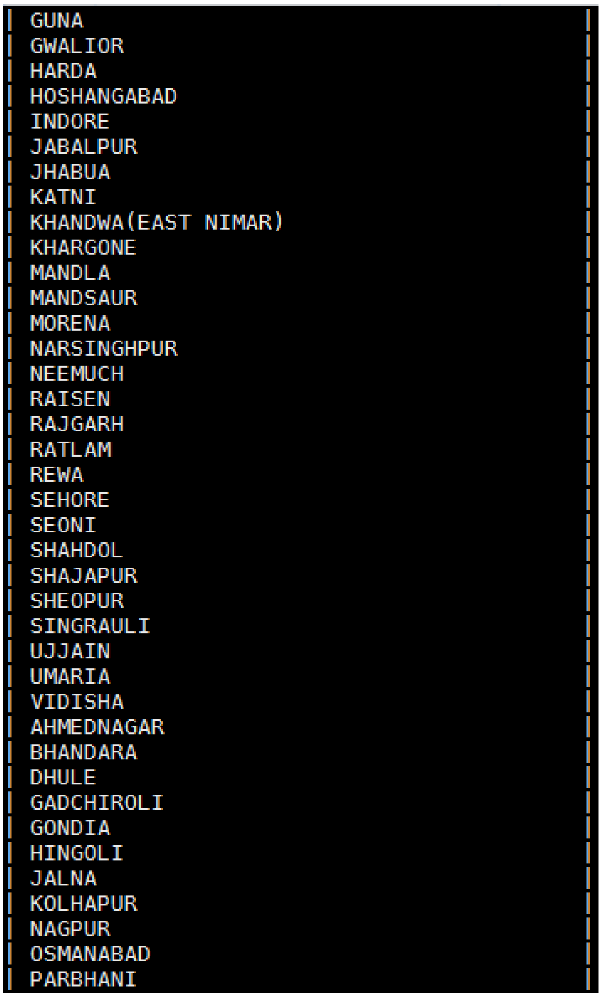
```
| TINSUKIA                                   |
| BEGUSARAI                                  |
| MADHUBANI                                  |
| MUZAFFARPUR                                |
| SAHARSA                                    |
| VAISHALI                                   |
| DHAMTARI                                   |
| JASHPUR                                    |
| KANKER                                     |
| KORBA                                      |
| KORIYA                                     |
| SURGUJA                                    |
| NORTH GOA                                  |
| AHMEDABAD                                  |
| AMRELI                                     |
| ANAND                                      |
| BANAS KANTHA                               |
| BHARUCH                                    |
| BHAVNAGAR                                  |
| DAHOD                                      |
| DANGS                                      |
| GANDHINAGAR                                |
| JAMNAGAR                                   |
| JUNAGADH                                   |
| KACHCHH                                    |
| KHEDA                                      |
| MAHESANA                                   |
| NARMADA                                    |
| NAVSARI                                    |
| PANCH MAHALS                               |
| PATAN                                      |
| PORBANDAR                                  |
| RAJKOT                                     |
| SABAR KANTHA                               |
| SURAT                                      |
| SURENDRANAGAR                              |
| VADODARA                                   |
| VALSAD                                     |
| AMBALA                                     |
```



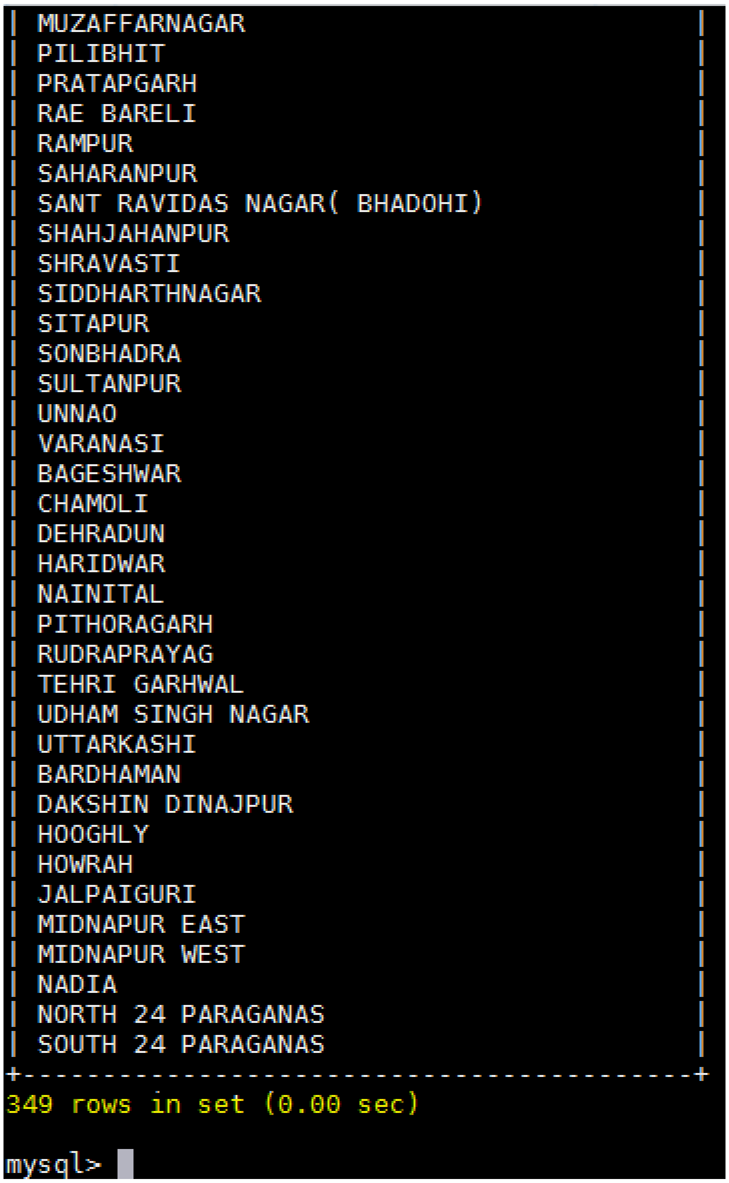
```
| BHIWANI                                    |
| FARIDABAD                                  |
| FATEHABAD                                  |
| GURGAON                                    |
| HISAR                                      |
| JHAJJAR                                    |
| JIND                                       |
| KAITHAL                                    |
| KARNAL                                     |
| KURUKSHETRA                                |
| MAHENDRAGARH                               |
| MEWAT                                      |
| PANCHKULA                                  |
| PANIPAT                                    |
| REWARI                                     |
| ROHTAK                                     |
| SIRSA                                      |
| SONIPAT                                    |
| YAMUNANAGAR                                |
| BILASPUR                                   |
| CHAMBA                                     |
| HAMIRPUR                                   |
| KANGRA                                     |
| KINNAUR                                    |
| KULLU                                      |
| LAHAUL & SPITI                             |
| MANDI                                      |
| SHIMLA                                     |
| SIRMAUR                                    |
| SOLAN                                      |
| UNA                                        |
| ANANTNAG                                   |
| LEH (LADAKH)                               |
| DEOGHAR                                    |
| DUMKA                                      |
| LATEHAR                                    |
| LOHARDAGA                                  |
| PAKUR                                      |
| PURBI SINGHBHUM                            |
```



```
| BAGALKOT                                 |
| BANGALORE RURAL                          |
| CHICKMAGALUR                             |
| CHITRADURGA                              |
| DHARWAD                                  |
| GADAG                                    |
| HASSAN                                   |
| KODAGU                                   |
| KOLAR                                    |
| KOPPAL                                   |
| MANDYA                                   |
| MANGALORE(DAKSHINA KANNADA)              |
| RAMANAGARA                               |
| SHIMOGA                                  |
| UDUPI                                    |
| ALAPPUZHA                                |
| ERNAKULAM                                |
| IDUKKI                                   |
| KANNUR                                   |
| KASARGOD                                 |
| KOLLAM                                   |
| KOTTAYAM                                 |
| KOZHIKODE                                |
| MALAPPURAM                               |
| PALAKKAD                                 |
| PATHANAMTHITTA                           |
| THIRUVANANTHAPURAM                       |
| THRISSUR                                 |
| WAYANAD                                  |
| ALIRAJPUR                                |
| ANUPPUR                                  |
| BARWANI                                  |
| BETUL                                    |
| BHOPAL                                   |
| BURHANPUR                                |
| DATIA                                    |
| DEWAS                                    |
| DHAR                                     |
| DINDORI                                  |
```



```
| GUNA                                     |
| GWALIOR                                  |
| HARDA                                    |
| HOSHANGABAD                              |
| INDORE                                   |
| JABALPUR                                 |
| JHABUA                                   |
| KATNI                                    |
| KHANDWA(EAST NIMAR)                      |
| KHARGONE                                 |
| MANDLA                                   |
| MANDSAUR                                 |
| MORENA                                   |
| NARSINGHPUR                              |
| NEEMUCH                                  |
| RAISEN                                   |
| RAJGARH                                  |
| RATLAM                                   |
| REWA                                     |
| SEHORE                                   |
| SEONI                                    |
| SHAHDOL                                  |
| SHAJAPUR                                 |
| SHEOPUR                                  |
| SINGRAULI                                |
| UJJAIN                                   |
| UMARIA                                   |
| VIDISHA                                  |
| AHMEDNAGAR                               |
| BHANDARA                                 |
| DHULE                                    |
| GADCHIROLI                               |
| GONDIA                                   |
| HINGOLI                                  |
| JALNA                                    |
| KOLHAPUR                                 |
| NAGPUR                                   |
| OSMANABAD                                |
| PARBHANI                                 |
```



```
| MUZAFFARNAGAR                          |
| PILIBHIT                               |
| PRATAPGARH                             |
| RAE BARELI                             |
| RAMPUR                                 |
| SAHARANPUR                             |
| SANT RAVIDAS NAGAR( BHADOHI)           |
| SHAHJAHANPUR                           |
| SHRAVASTI                              |
| SIDDHARTHNAGAR                         |
| SITAPUR                                |
| SONBHADRA                              |
| SULTANPUR                              |
| UNNAO                                  |
| VARANASI                               |
| BAGESHWAR                              |
| CHAMOLI                                |
| DEHRADUN                               |
| HARIDWAR                               |
| 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>
```

Hence, using PIG UDF we have got the required result and stored into the **mysql** table using **sqoop** commands.

End of Document