**Project 1 State-Wise Development Analysis In India**

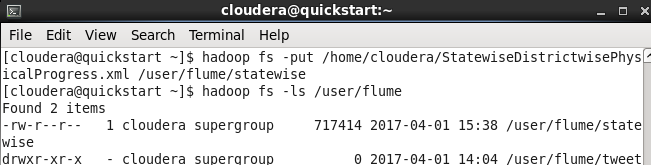
1.Find out the districts who achieved 100 percent objective in BPL

-**Store the data in Hadoop Cluster using Flume**

**-Clean the data using pig script**

**-Store the results in HBASE/RDBMS using Sqoop**

**-Insert datasets in HDFS from local file system**

****

**-Starting FLUME job**

**-Create a conf file in flume-ng/conf folder**

**-Named it as state-wise.conf**

**-Edit the configuration file with below**

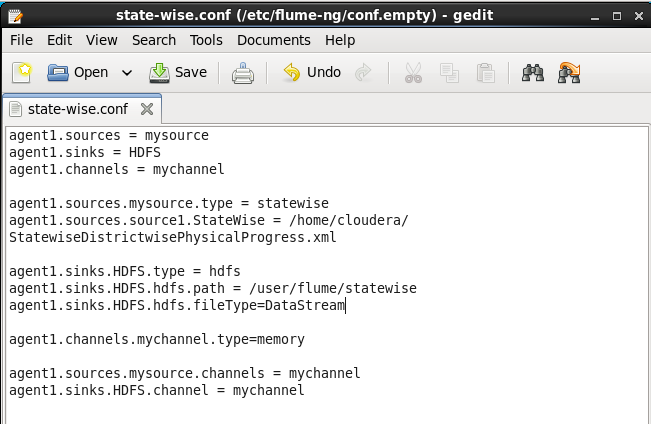
- Name the components of the agent

- Describe the source where should read

- Describe the sink where should write

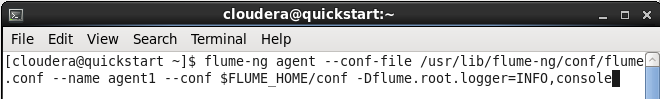
- Use a channel which moves in memory so when moving into HDFS give it as a file

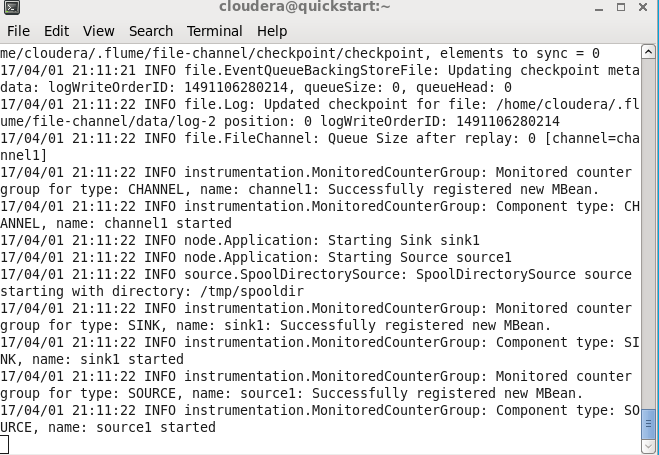
- Define the source directory, bind the sources and sinks to the channel

****

**-Given the configuration file**

**-Start flume-ng agent as follows :**

****

****

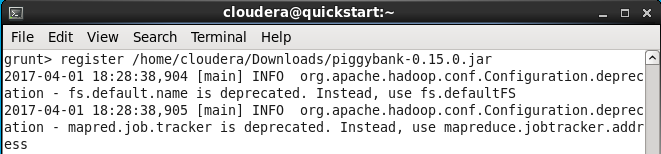
**-PERFORMING PIG**

**-Clean the data using PIG**

**-Start the grunt shell with help of pig command**

**-To read xml format we XMLLoader() in piggybank. So for this we need to register a piggybank-0.15.0.jar**

**-Download the jar and using register command register the jar**

****

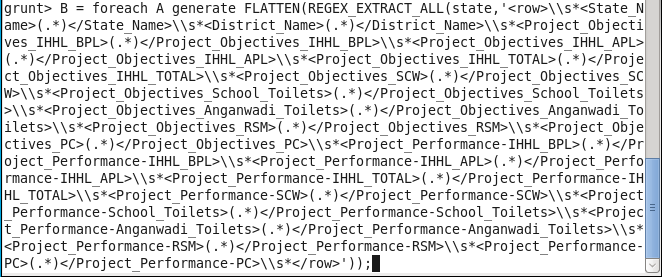
**-Load the xml file using the default XML Loader available in pig, inside the XML loader we are specifying that our root element is property**

**-we are storing the whole thing with an alias name** state **as** chararray

****

**-Pig job for parsing xml data**

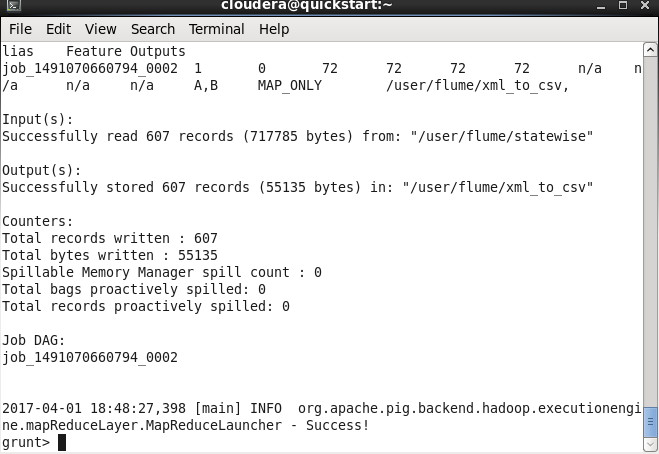
**-Now get the values of each state from the relation A and store it into another relation named B using the foreach operator and separate the record by using flatten function**

****

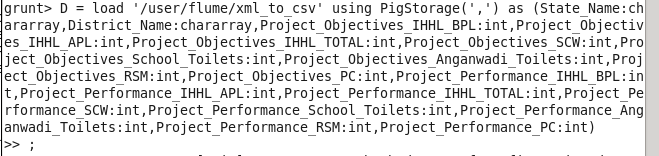
**-Define XPath**

**-Now store the relation in HDFS directory ‘/user/flume/xml\_to\_csv’ inside the CSVExcelStorage**

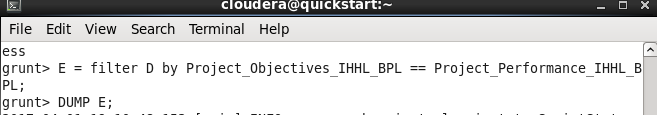
****

****

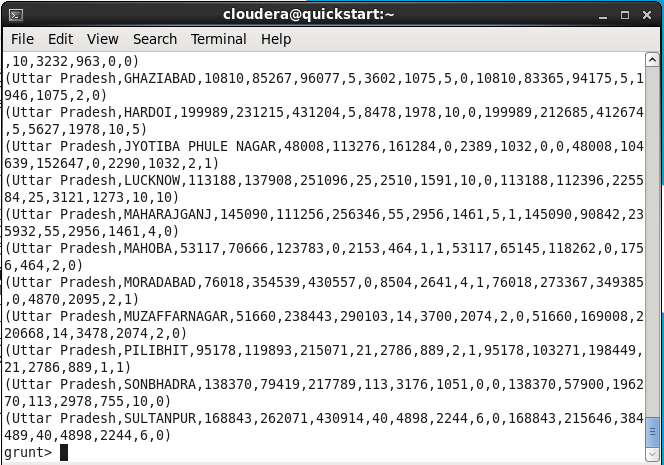
**-Loading the csv file into pig**

****

**-Now use the Filter operator to get the details of the Project\_Objective\_IIHL\_BPL**

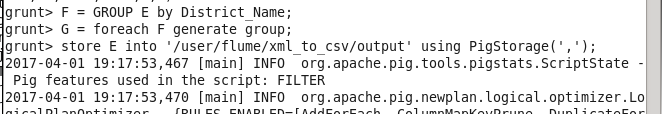
****

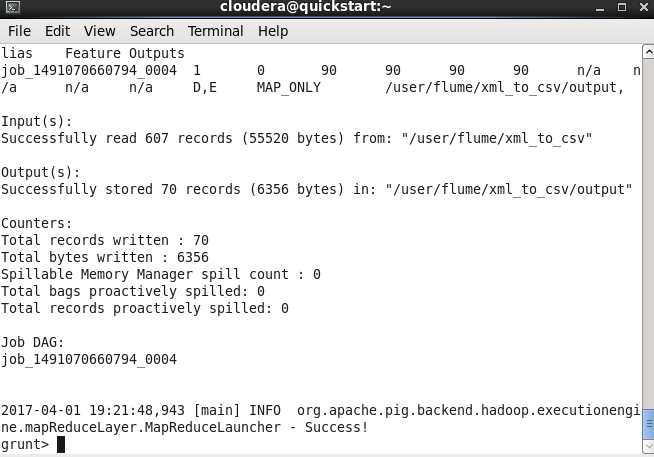
**-dump E;**

****

**-now group the records/tuples in the relation by District\_Name**

**- Store the filter details in ‘/user/flume/xml\_to\_csv/output**

****

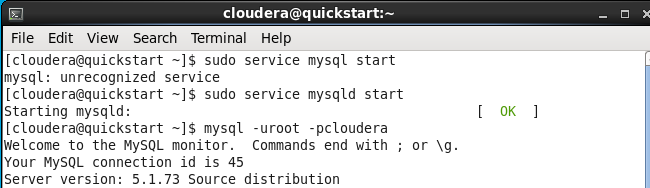


**-PERFORMING SQOOP**

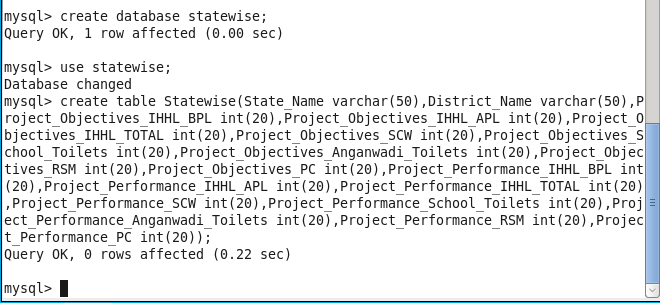
**-Store the result in hbase/RDBMS**

**-Before exporting we need to create a empty table in mysql**

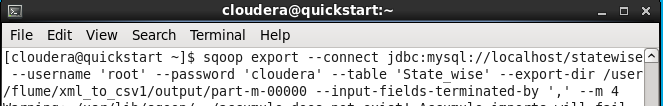
**-Start mysqld**

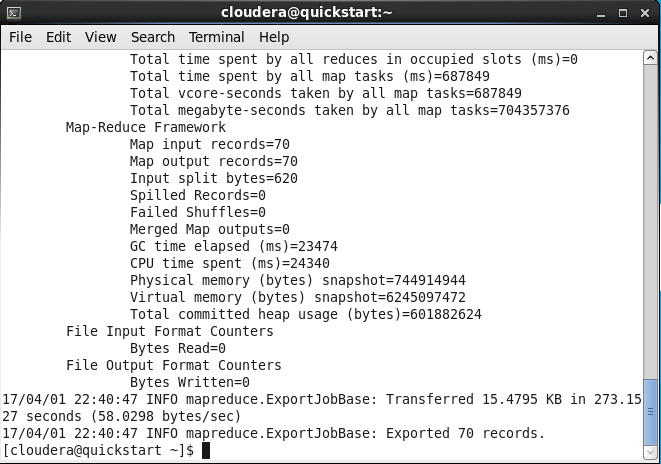
****

**-create table in mysql**

****

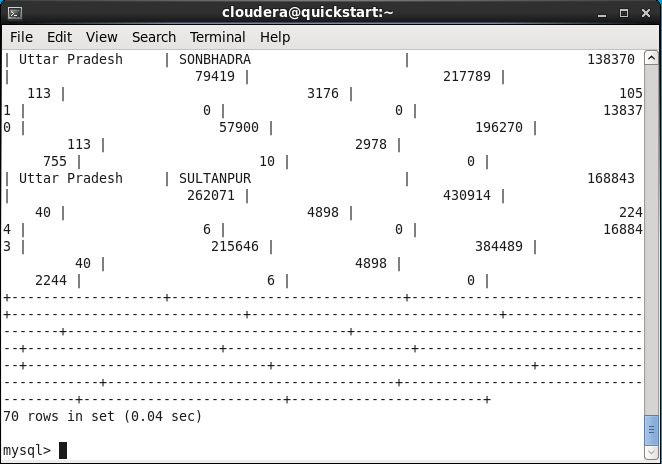
**-Exporting to sql using sqoop**

****

****

**-now we can check the result in mysql by select command**

****

****

**-So the result is**

**-70 districts has achieved 100% objective in BPL cards.**