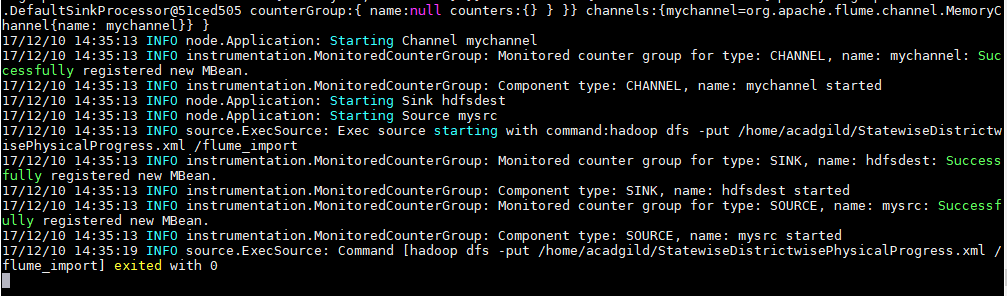
**Project**

* Copy dataset from local file system to HDFS using flume.

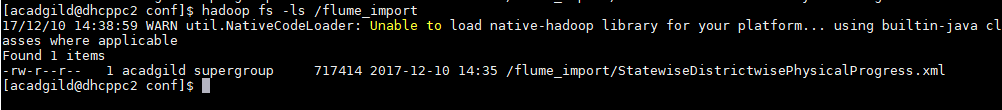
1. Create FLUME job for fetching log files from spool directory the data

flume-ng agent -n agent1 -f /usr/local/flume/conf/filecopy.conf



1. Verify the file in HDFS

hadoop fs –ls /flume\_import



* Input file is in the XML format use Map reduce or pig to parse the data and get the results

Read XML in PIG:

1. Register piggybank jar and Define XPath method

REGISTER /usr/local/pig/lib/piggybank.jar

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

1. Load Data from xml to pig relation

A =LOAD 'hdfs://localhost:9000/flume\_import/StatewiseDistrictwisePhysicalProgress.xml' using org.apache.pig.piggybank.storage.XMLLoader('row') as (x:chararray);

1. Parse the xml file using XPath

B =FOREACH A GENERATE XPath(x,'row/State\_Name') as State\_Name

,XPath(x, 'row/District\_Name') as District\_Name

,XPath(x,'row/Project\_Objectives\_IHHL\_BPL') as Project\_Objectives\_IHHL\_BPL

,XPath(x,'row/Project\_Objectives\_IHHL\_APL') as Project\_Objectives\_IHHL\_APL

,XPath(x,'row/Project\_Objectives\_IHHL\_TOTAL') as Project\_Objectives\_IHHL\_TOTAL

,XPath(x,'row/Project\_Objectives\_SCW') as Project\_Objectives\_SCW

,XPath(x,'row/Project\_Objectives\_School\_Toilets') as Project\_Objectives\_School\_Toilets

,XPath(x,'row/Project\_Objectives\_Anganwadi\_Toilets') as Project\_Objectives\_Anganwadi\_Toilets

,XPath(x,'row/Project\_Objectives\_RSM') as Project\_Objectives\_RSM

,XPath(x,'row/Project\_Objectives\_PC') as Project\_Objectives\_PC

,XPath(x,'row/Project\_Performance-IHHL\_BPL') as Project\_Performance\_IHHL\_BPL

,XPath(x,'row/Project\_Performance-IHHL\_APL') as Project\_Performance\_IHHL\_APL

,XPath(x,'row/Project\_Performance-IHHL\_TOTAL') as Project\_Performance\_IHHL\_TOTAL

,XPath(x,'row/Project\_Performance-SCW') as Project\_Performance\_SCW

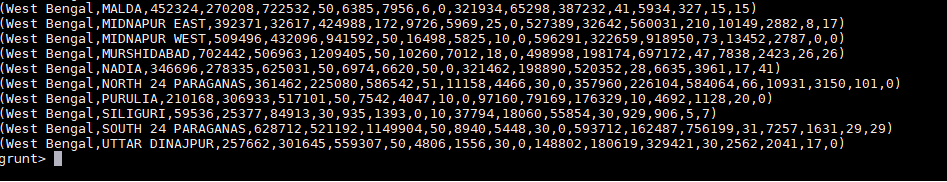
,XPath(x,'row/Project\_Performance-School\_Toilets') as Project\_Performance\_School\_Toilets

,XPath(x,'row/Project\_Performance-Anganwadi\_Toilets') as Project\_Performance\_Anganwadi\_Toilets

,XPath(x,'row/Project\_Performance-RSM') as Project\_Performance\_RSM

,XPath(x,'row/Project\_Performance-PC') as Project\_Performance\_PC;

dump B;



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

Export the results to mysql using sqoop

1. UDF function to get 100% and above:

**public** **class** HundredPercent **extends** EvalFunc<Boolean>{

@Override

**public** Boolean exec(Tuple tuple) **throws** IOException {

**if** (tuple==**null**)

**return** **null**;

**float** percentage=0;

**float** objective=Float.*valueOf*(((String)tuple.get(0)).trim());

**float** performance=Float.*valueOf*(((String)tuple.get(1)).trim());;

percentage= performance/objective \*100;

**if**(percentage>=100f)

{

**return** **true**;

}

**return** **false**;

}

}

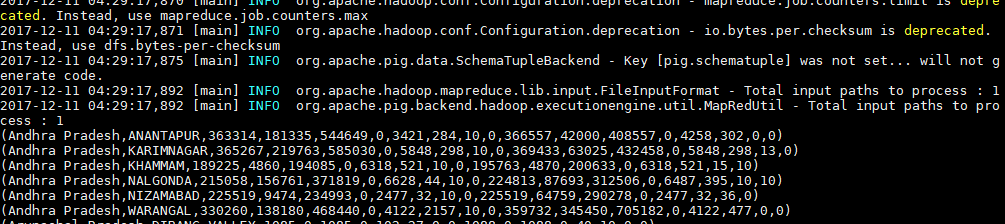
1. Register custom jar and function

REGISTER /home/acadgild/nikidir/projectOne.jar

DEFINE isHundred HundredPercent();

1. Filter the result for 100%

C =Filter B by isHundred(Project\_Objectives\_IHHL\_BPL,Project\_Performance\_IHHL\_BPL);



1. Retrieve only required columns

D =Foreach C generate State\_Name,District\_Name,Project\_Objectives\_IHHL\_BPL,Project\_Performance\_IHHL\_BPL;

1. Store result in hdfs file.

store D into 'output100' using PigStorage(',');

--location: hdfs://localhost:9000/user/acadgild/output100

1. Create table in mysql.

create table hundred\_percent\_group(

state varchar(30),

district varchar(30),

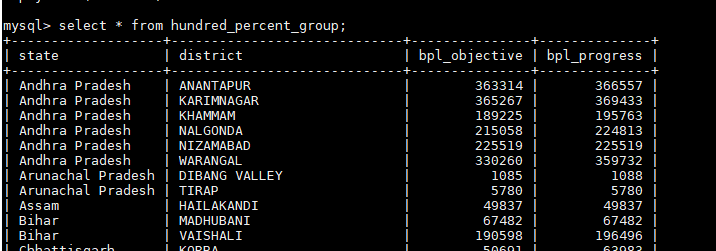
bpl\_objective int,

bpl\_progress int

);

1. Using Sqoop export the hdfs file content to mysql table

sqoop export --connect jdbc:mysql://localhost/mydb --username root -P --table hundred\_percent\_group --export-dir hdfs://localhost:9000/user/acadgild/output100/part-m-00000 --input-fields-terminated-by ','



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

Export the results to MySQL using Sqoop.

UDF function to calculate specified percentage for given values.

**public** **class** CalculatePercent **extends** EvalFunc<Boolean>{

@Override

**public** Boolean exec(Tuple tuple) **throws** IOException {

**if** (tuple==**null**)

**return** **null**;

**float** percentage=0;

**float** objective=Float.*valueOf*(((String)tuple.get(0)).trim());

**float** performance=Float.*valueOf*(((String)tuple.get(1)).trim());

**float** percent =(Integer) tuple.get(2);

percentage= performance/objective \* 100f;

**if**(percentage>=percent)

{

**return** **true**;

}

**return** **false**;

}

}

1. Filter the result for 100%

C =Filter B by isEight(Project\_Objectives\_IHHL\_BPL,Project\_Performance\_IHHL\_BPL,80);

1. Retrieve only required columns

D =Foreach C generate State\_Name,District\_Name,Project\_Objectives\_IHHL\_BPL,Project\_Performance\_IHHL\_BPL;

1. Store result in hdfs file.

store D into 'output80' using PigStorage(',');

--location: hdfs://localhost:9000/user/acadgild/output80

1. Using Sqoop export the hdfs file content to mysql table

sqoop export --connect jdbc:mysql://localhost/mydb --username root -P --table eighty\_percent\_group --export-dir hdfs://localhost:9000/user/acadgild/output80/part-m-00000 --input-fields-terminated-by ','

