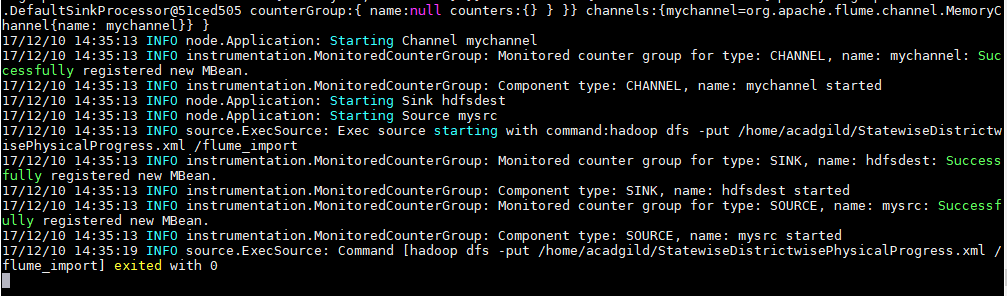
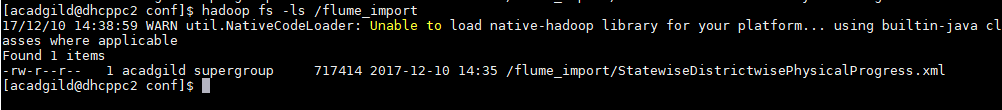
Project

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



Verify the file in HDFS



Read XML in PIG:

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

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

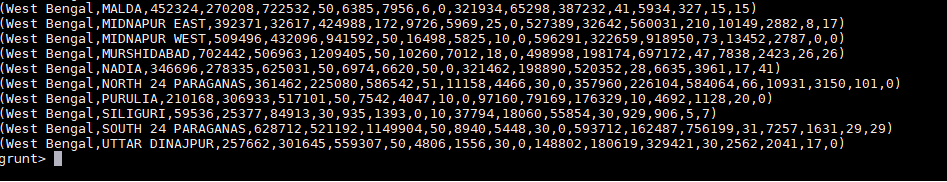
,XPath(x,'row/Project\_Performance-School\_Toilets') as Project\_Performance\_School\_Toilets

,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

UDF function to

**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**;

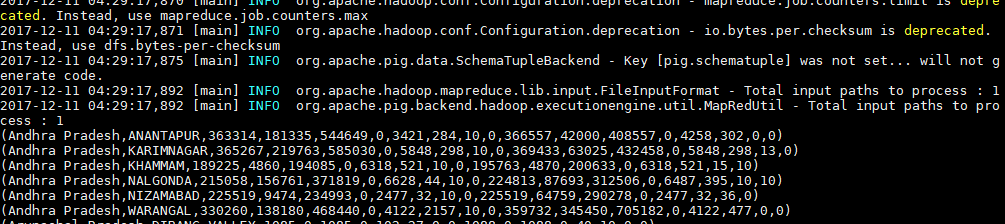
}

}

REGISTER /home/acadgild/nikidir/projectOne.jar

DEFINE isHundred HundredPercent();

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



create table hundred\_percent\_group(

state varchar(30),

district varchar(30),

bpl\_objective int,

bpl\_progress int

);

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.