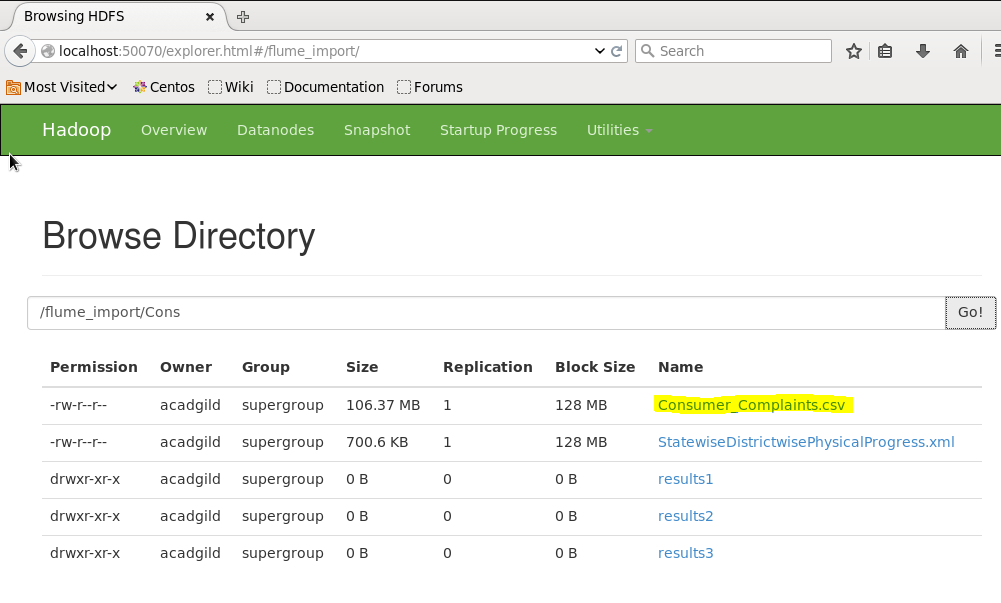
**Note: You need to copy the dataset into HDFS using Flume and the results of the problem statements**

**should be exported into RDBMS(Mysql) using sqoop.**

****

**The aim of this project is to analyze performance of various companies on aspects like:**

1. **Write a pig script to find no of complaints which got timely response**

***Flume script to move the csv data into HDFS***

[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

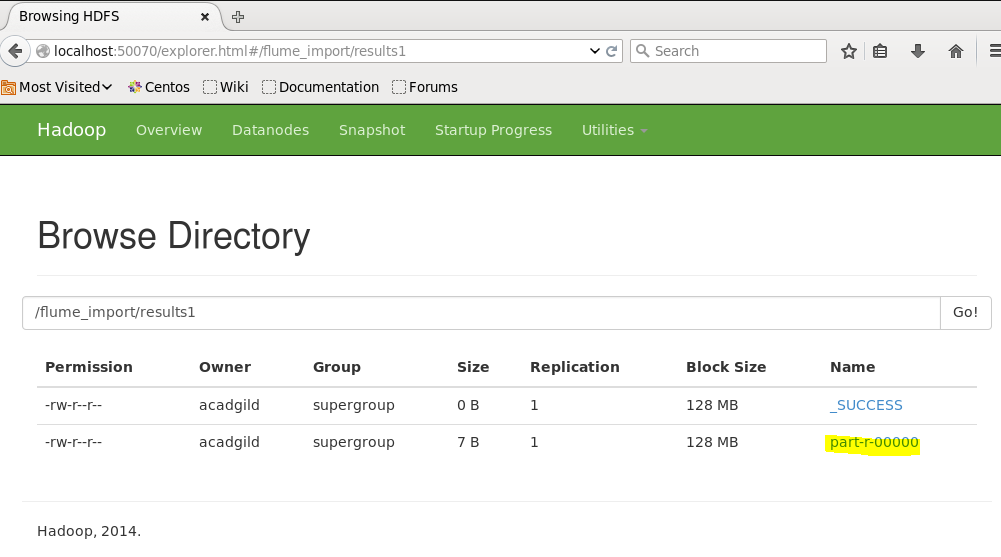
agent1.channels.mychannel.type = memory

agent1.sources.mysrc.channels = mychannel

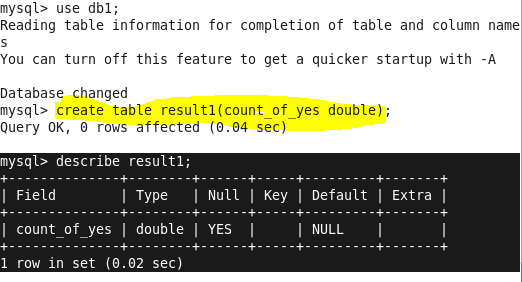
agent1.sinks.hdfsdest.channel = mychannel

***Command to run the flume agent:***

[acadgild@localhost project]$ flume-ng agent -n agent1 -f /usr/local/flume/conf/filecopy.conf

****

***Create a empty table to hold the data in MySQL:***

****

***Execute the below script:***

result1.pig:

============

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

A = LOAD '/flume\_import/Consumer\_Complaints.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','WRITE\_OUTPUT\_HEADER');

B = foreach A generate $13;

C = filter B by ($0=='Yes');

D = group C all;

E =foreach D generate COUNT(C.$0);

store E into '/flume\_import/result1' using PigStorage(',');

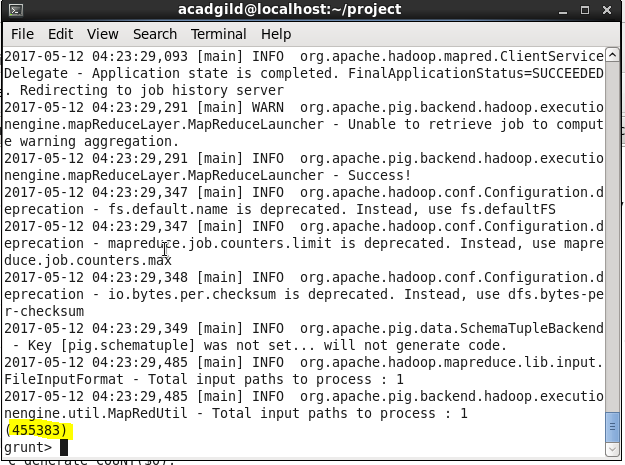
prob\_sol1.sh

============

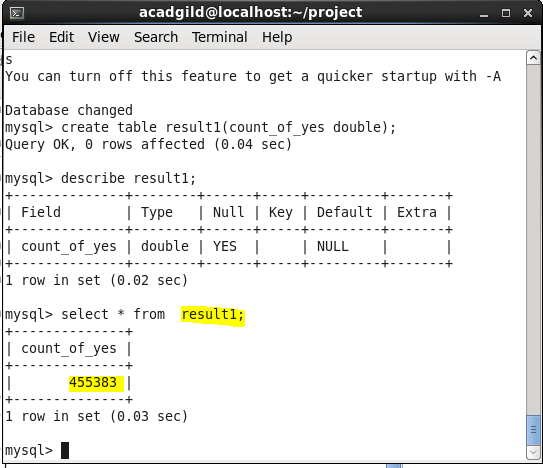
#! /bin/bash

pig result1.pig

sqoop export -m 1 --connect jdbc:mysql://localhost/db1 --username root --table 'result1' --export-dir '/flume\_import/result1/part-r-00000' --input-fields-terminated-by ','

****

***Check the table in MySQL which has been populated with the results:***

****

**2. Write a pig script to find no of complaints where consumer forum forwarded the complaint**

**same day they received to respective company**

**Create a empty table ‘result2’ in MySQL, and run the shell script.**

***prob\_sol2.sh:***

#! /bin/bash

pig result2.pig

sqoop export -m 1 --connect jdbc:mysql://localhost/db1 --username root --table 'result2' --export-dir '/flume\_import/result2/part-r-00000' --input-fields-terminated-by ','

**result2.pig:**

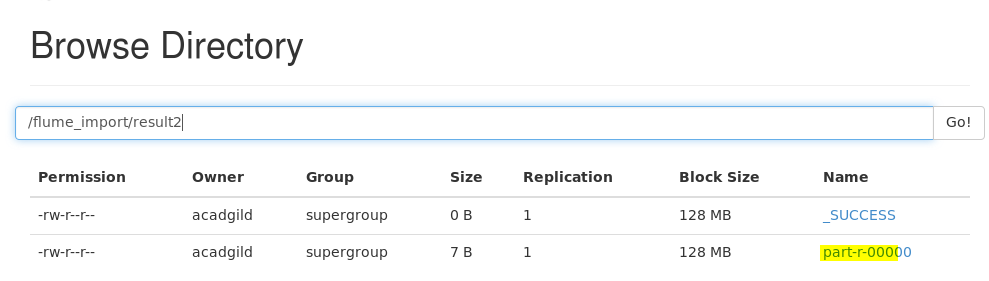
REGISTER /usr/local/pig/lib/piggybank.jar

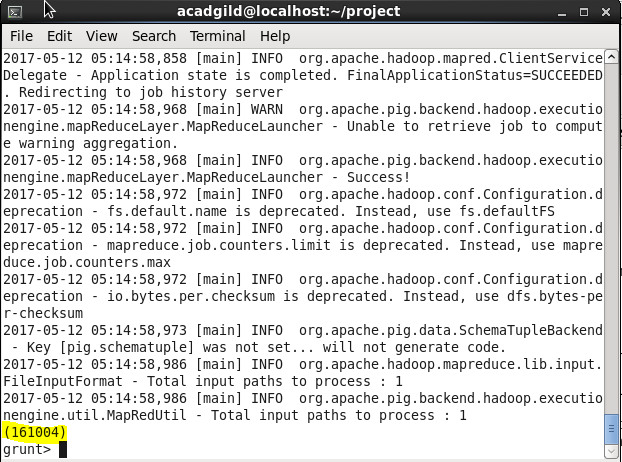
A = LOAD '/flume\_import/Consumer\_Complaints.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','WRITE\_OUTPUT\_HEADER');

B = filter A by ($0==$11);

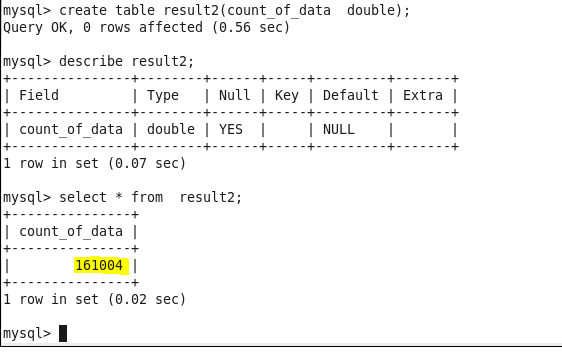
D =foreach C generate COUNT(B);

store D into '/flume\_import/result2' using PigStorage(',');



****

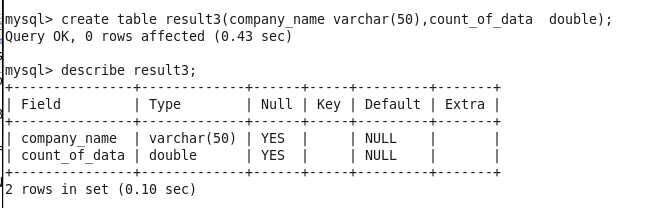
***Check in MySQL if the table has been populated with the result data:***

****

**3. Write a pig script to find list of companies toping in complaint chart (companies with**

**maximum number of complaints)**

***Create a empty table to hold the data in MySQL:***

******

***Execute the below script:***

Result3.pig:

============

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

A = LOAD '/flume\_import/Consumer\_Complaints.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','WRITE\_OUTPUT\_HEADER');

B = group A by $7;

C =foreach B generate $0,COUNT(A);

D = order C by $1 desc;

E = LIMIT D 10;

store E into '/flume\_import/result3' using PigStorage(',');

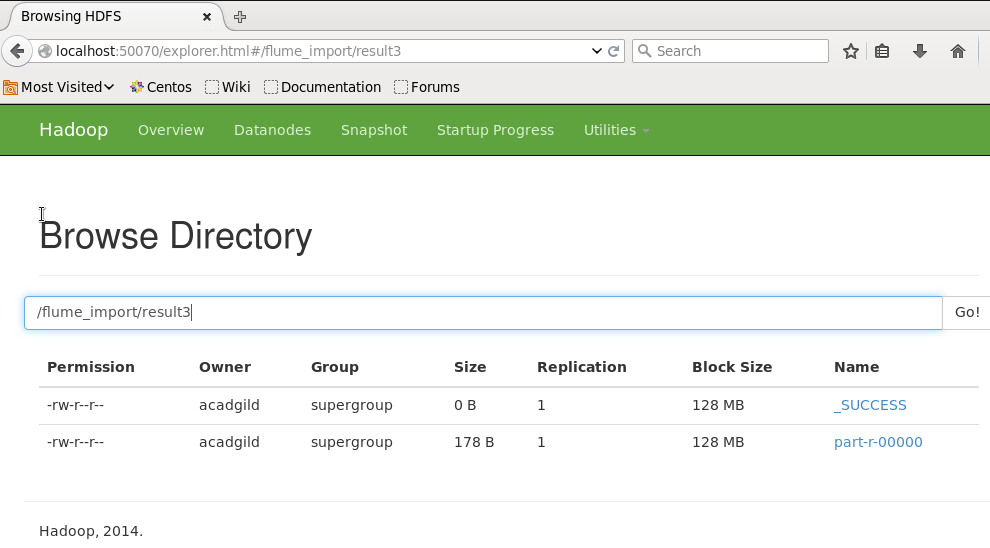
prob\_sol3.sh

============

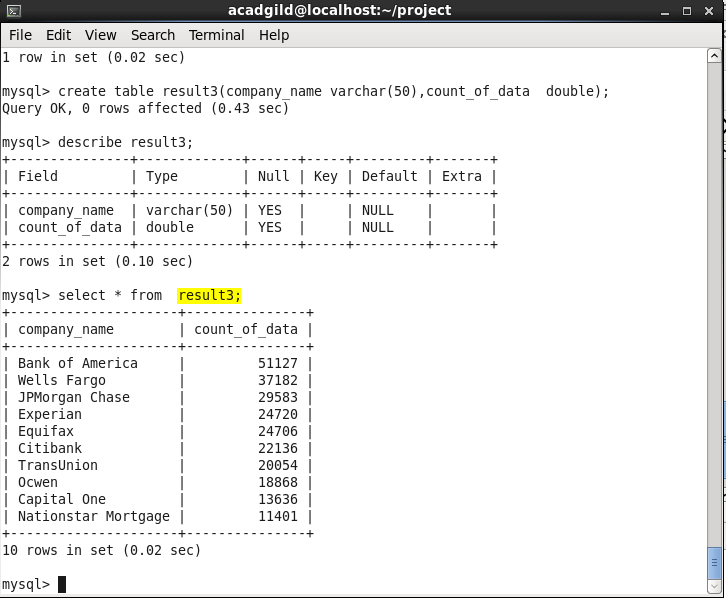
#! /bin/bash

pig result3.pig

sqoop export -m 1 --connect jdbc:mysql://localhost/db1 --username root --table 'result3' --export-dir '/flume\_import/result3/part-r-00000' --input-fields-terminated-by ','

****

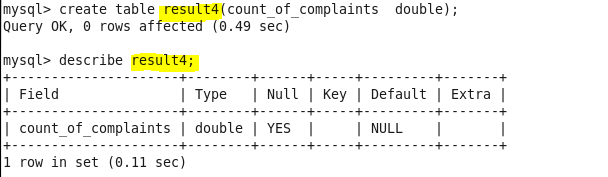
***Check the table in MySQL which has been populated with the results:***

****

**4. Write a pig script to find no of complaints filed with product type has &quot;Debt**

**collection&quot; for the year 2015**

**Create an empty MySql table:**

****

***Execute the below script:***

Result4.pig:

============

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

A = LOAD '/flume\_import/Consumer\_Complaints.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','WRITE\_OUTPUT\_HEADER');

B = filter A by ($1=='Debt collection');

C = foreach B generate $0;

D = foreach C generate FLATTEN(REGEX\_EXTRACT\_ALL($0,'([0-3][0-9])[-/]([0-1][0-9])[-/](15|2015)')) as (data:chararray);

E = group D all;

F =foreach E generate COUNT(D.data);

store F into '/flume\_import/result4' using PigStorage(',');

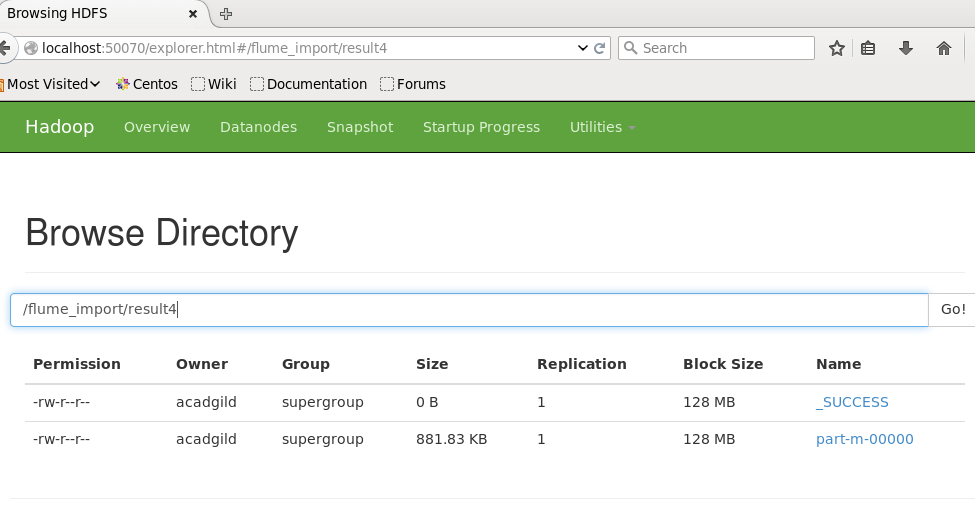
prob\_sol4.sh

============

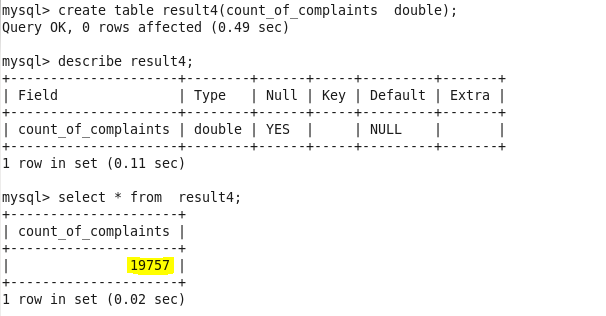
#! /bin/bash

pig result4.pig

sqoop export -m 1 --connect jdbc:mysql://localhost/db1 --username root --table 'result4' --export-dir '/flume\_import/result4/part-r-00000' --input-fields-terminated-by ','

****

***Check the table in MySQL which has been populated with the results:***

****