1. Flume conf file Source Code

agent1.sources = mysrc

agent1.sinks = hdfsdest

agent1.channels = mychannel

agent1.sources.mysrc.type = exec

agent1.sources.mysrc.command = hadoop dfs -put /home/acadgild/Project2/Consumer\_Complaints.csv /flume\_import1

agent1.sinks.hdfsdest.type = hdfs

agent1.sinks.hdfsdest.hdfs.path = hdfs://localhost:50070/flume\_import1

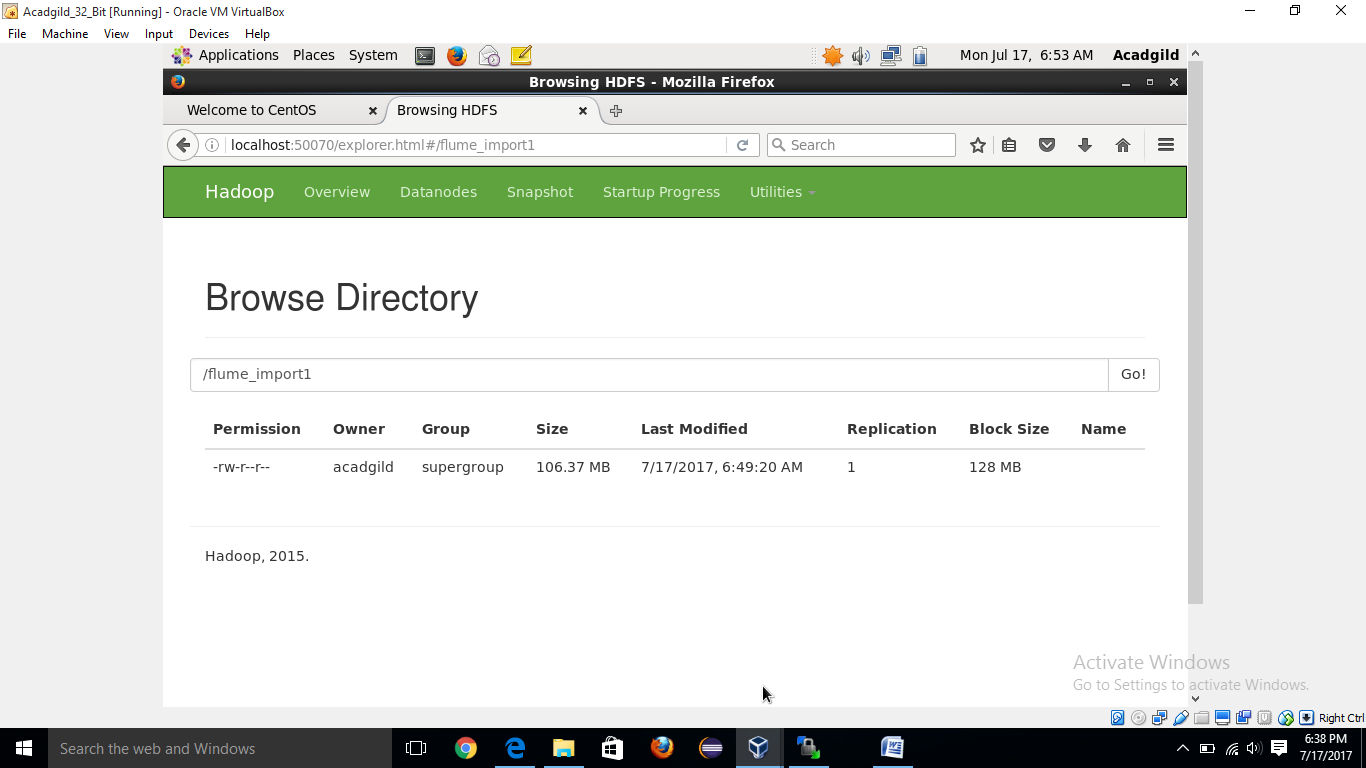
agent1.channels.mychannel.type = memory

agent1.sources.mysrc.channels = mychannel

agent1.sinks.hdfsdest.channel = mychannel

--Source Code

flume-ng agent --conf-file /home/acadgild/Project2/filecopy1.conf --name agent1



**--Loading the file**

complaint = LOAD '/home/acadgild/Project2/Consumer\_Complaints.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',') as (date:chararray, product:chararray, subp:chararray, issue:chararray, subi:chararray, complaint:chararray, response:chararray, company:chararray, state:chararray, zip:chararray, via:chararray, dateSent:chararray, responseToConsumer:chararray, timelyRes:chararray, disputed:chararray, id:int);

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

A = FILTER complaint BY $13 == 'Yes';

B = GROUP A ALL;

C = FOREACH B GENERATE group, COUNT(A.$13);

dump C;

**Ans:** (all,455383)

1. **Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company**

A2 = FILTER complaint BY $0 == $11;

B2 = GROUP A2 ALL;

C2 = FOREACH B2 GENERATE group, COUNT(A.$13);

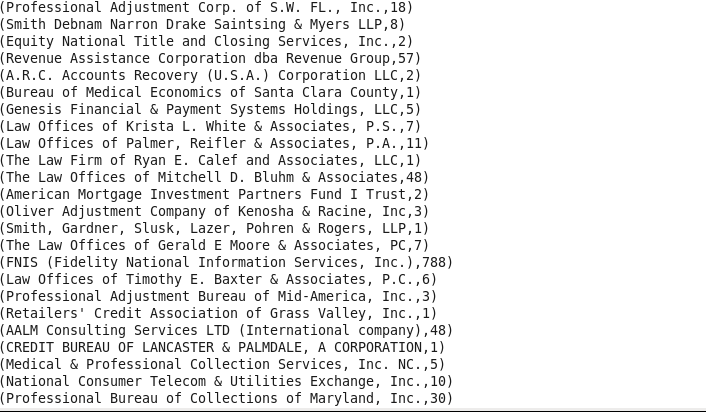
**Ans:** (all,161004)

1. **Write a pig script to find list of companies toping in complaint chart (companies with maximum number of complaints)**

A3 = GROUP complaint BY $7;

B3 = FOREACH A3 GENERATE group, COUNT(complaint.$5);

dump B3;



1. **Write a pig script to find no of complaints filed with product type has “ Debt collection” for the year 2015**

A4 = FILTER complaint BY $1 == 'Debt collection' AND SUBSTRING($0, 7, 4) == '2015';

B4 = GROUP A4 ALL;

C4 = FOREACH B4 GENERATE group, COUNT(A4.$15);

dump C4;

**Ans:** (all,31828)