Sensex Logdata Data Analysis

Submitted to the

Department of Master of Computer Applications

in partial fulfillment of the requirements of

Hadoop Ecosystems (MCAE03)

by

Name: Debojit Paul
USN:1MS18MCA12

RAMAIAH INSTITUTE OF TECHNOLOGY
(Autonomous Institute, Affiliated to VTU)
Accredited by National Board of Accreditation & NAAC with 'A'
Grade MSR Nagar, MSRIT Post, Bangalore-560054
www.msrit.edu **2019-2020**

CERTIFICATE

This is to certify that the Mini Project entitled **"Sensex Logdata Data Analysis"** is carried out by **Debojit Paul - 1MS18MCA12,** a bonafide student of Dept. of Master of Computer Applications, Ramaiah Institute of Technology, Bangalore, in partial fulfillment for the **Hadoop Ecosystems (MCAE03)** during the year 2019- 2020. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the mini-project report deposited in the department.

Faculty **Signature**

Ms. B N Nithya (B1)

ABSTRACT:

The project **'Sensex Logdata Data Analysis'** contains the field such as particular Sensex Name, having a Open Balance and Closing Balance, type of Trading and having Location, and frate.

The purpose of this project is to analysis, for identifying sensex of a particular Trading, estimating the Open Balance and Closing Balance, filtering the type of Trading and Location, Ordering the Open Balance in Descending order, grouping the sensex data and finding the average of Closing Balance, finding unique sensex names.

Tools used:

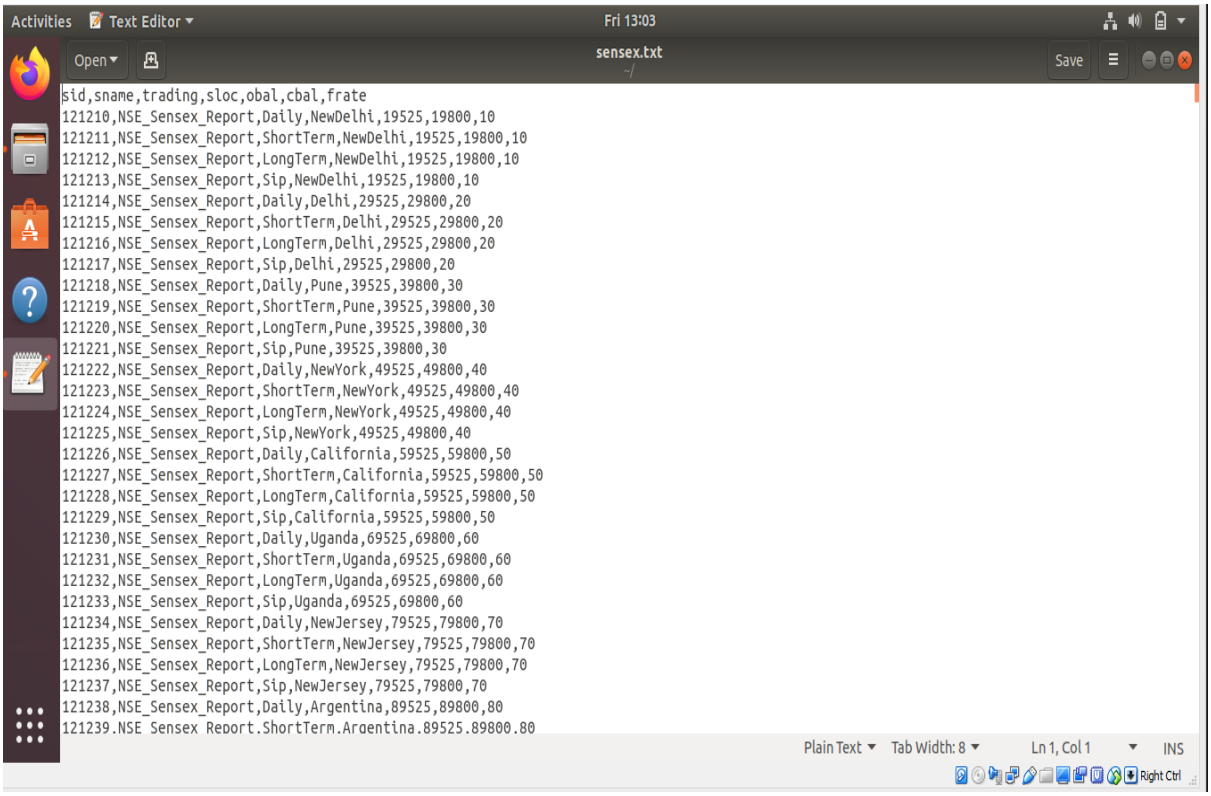
- Flume
- Hadoop
- Pig

Flume is used to load a original data(sensex.txt file) which is in local memory to HDFS, Hadoop is used to store the data from flume, by creating a directory in it.

Pig is used to process the commands, the data loaded in hdfs are processed here to produce a desired output.

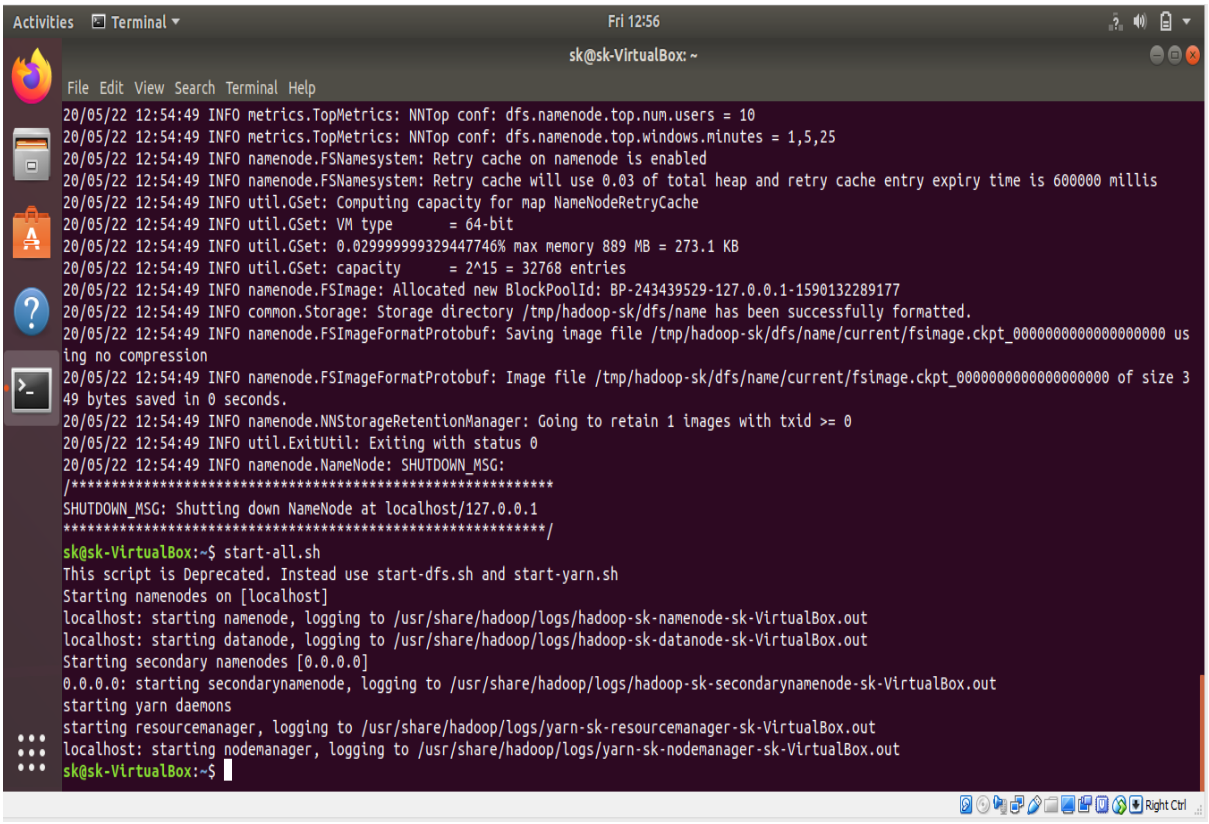
Original Data:

This is the sample data of **SENSEX,** taken to analyse.
The fields contain **sid, sname, trading, sloc, obal, cbal, frate**



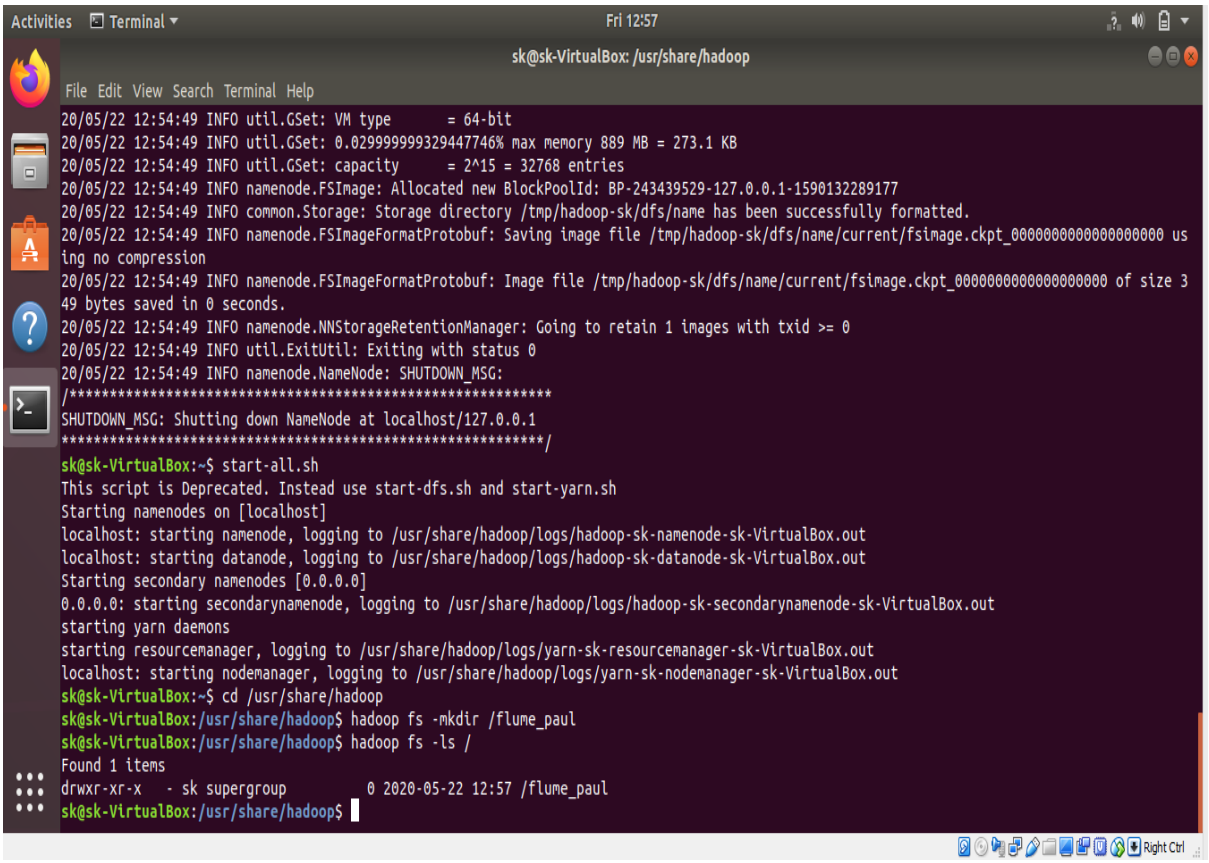
Command to check services in Hadoop:

Initially we have to start the Hadoop using **start-all.sh.**



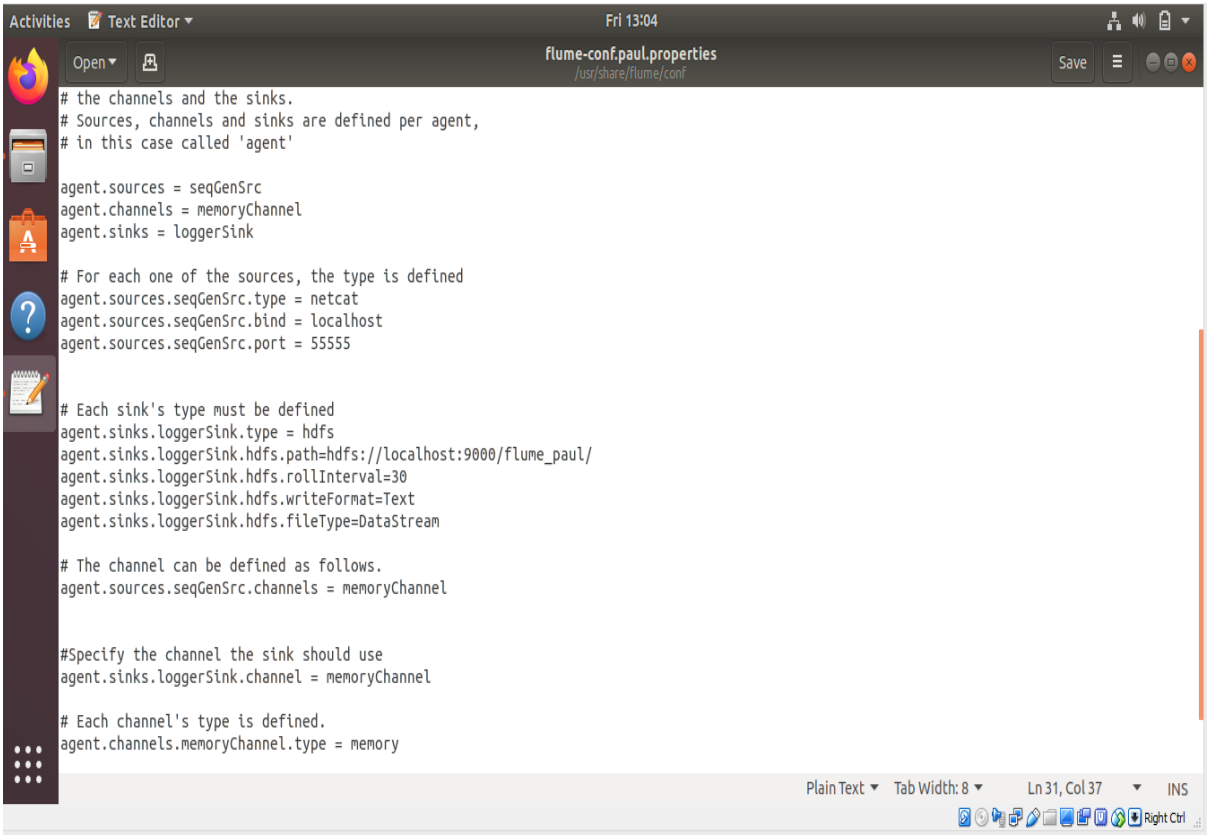
Creating a directory:

Creating a directory in Hadoop named flume_paul, the purpose of creating this directory is to store the flume data in hdfs.



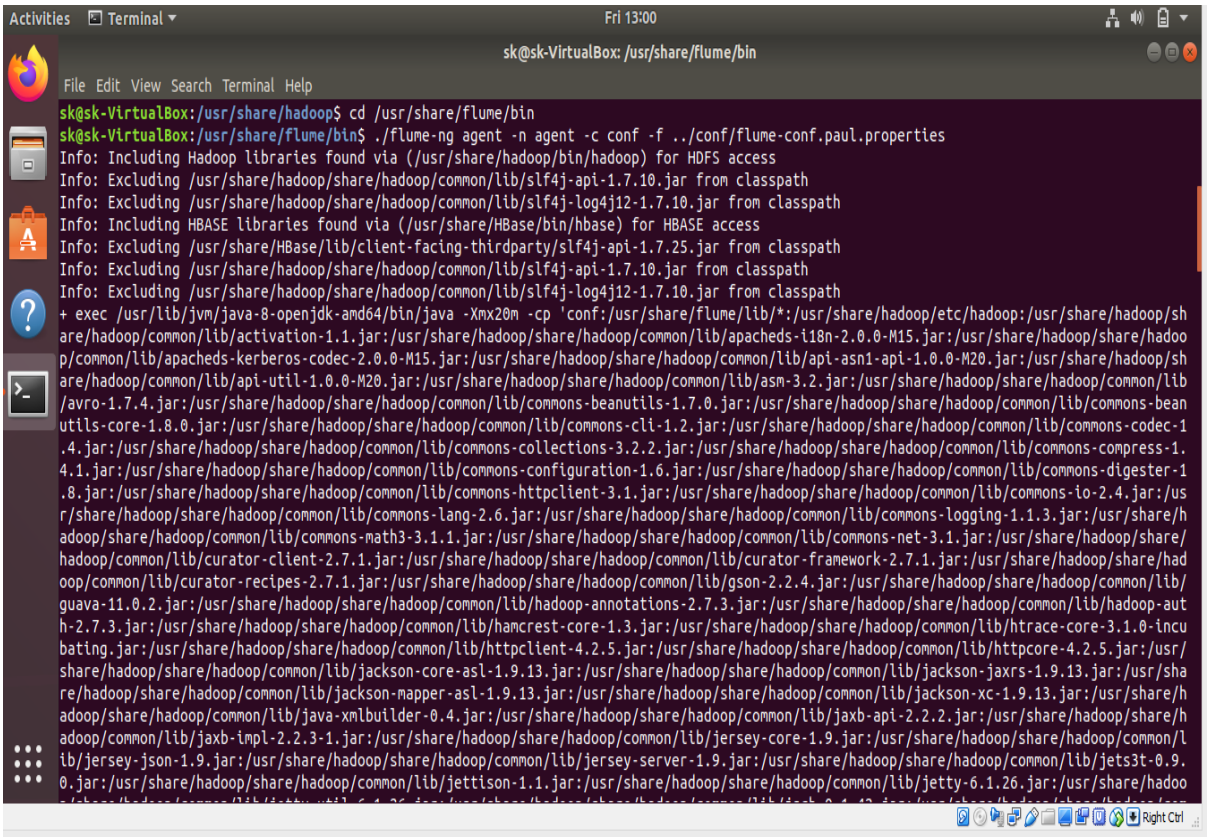
Flume Configuration image:

Configuration has to setup such as type, bind , port number, type of file ,write format, and rollInterval.



Transfering a data into flume:

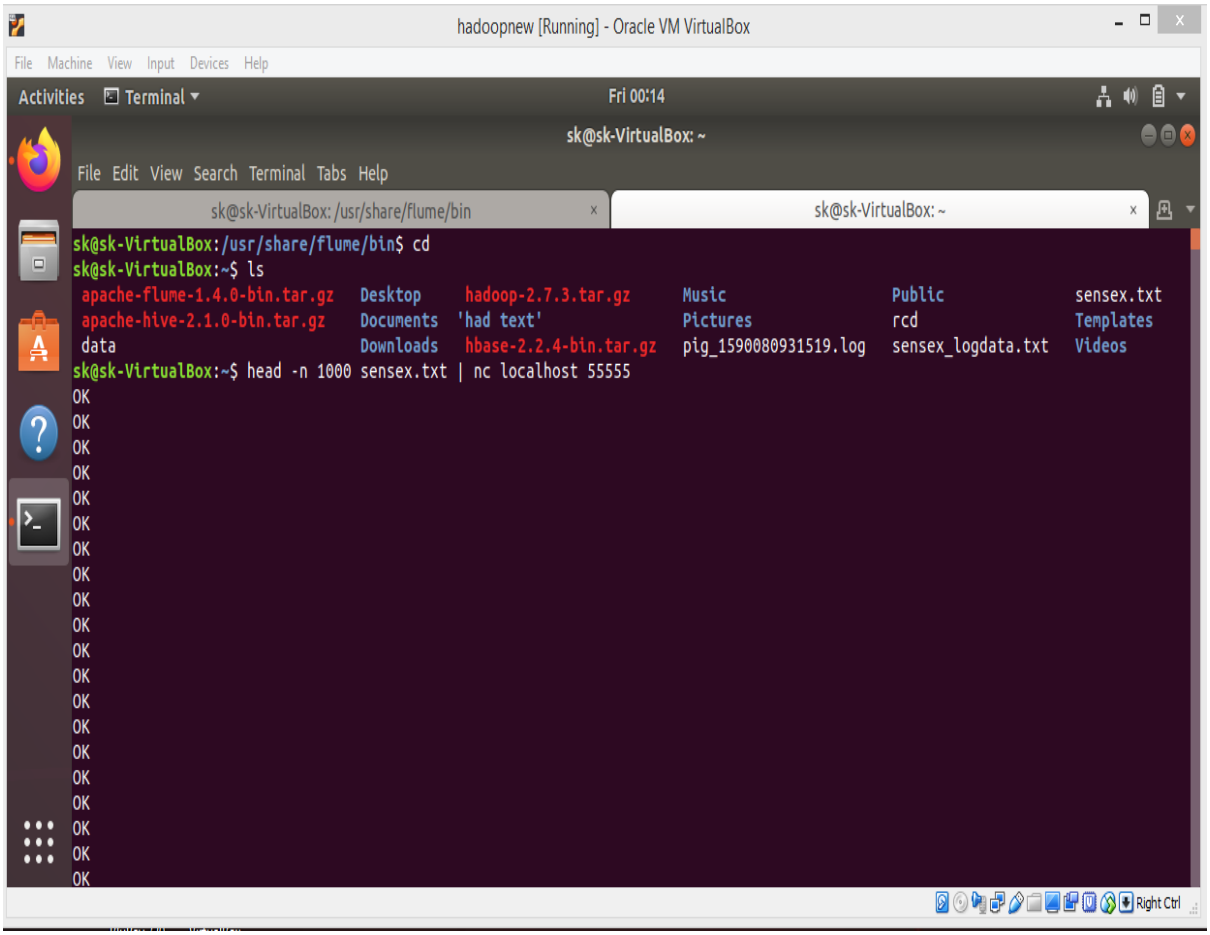
Transferring the data file from local memory to hdfs via flume using a command ./flume- ng agent –n agent –c conf –f ../conf/flume-conf.paul.properties



Loading a data into hdfs:

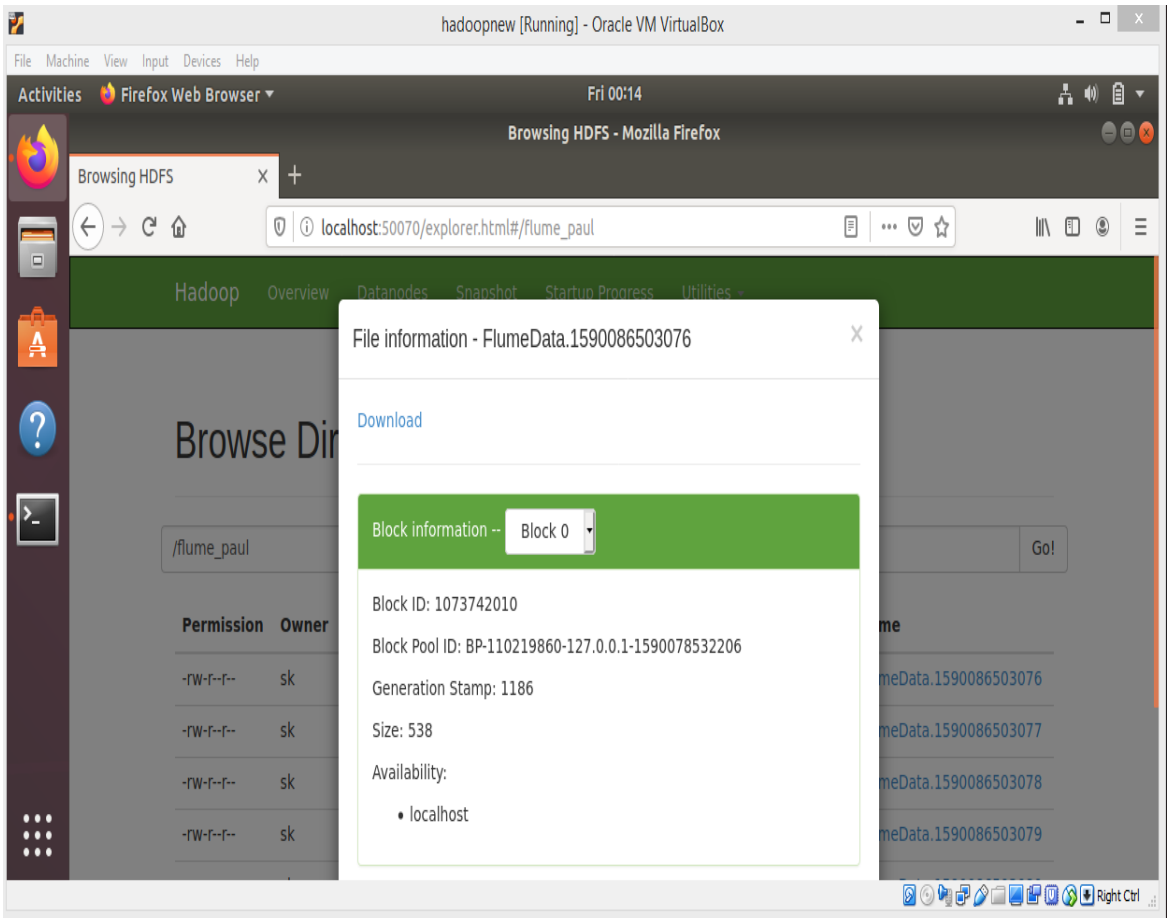
Successfully Transferring of data into hdfs.

$ head –n 1000 sensex.txt | nc localhost 55555



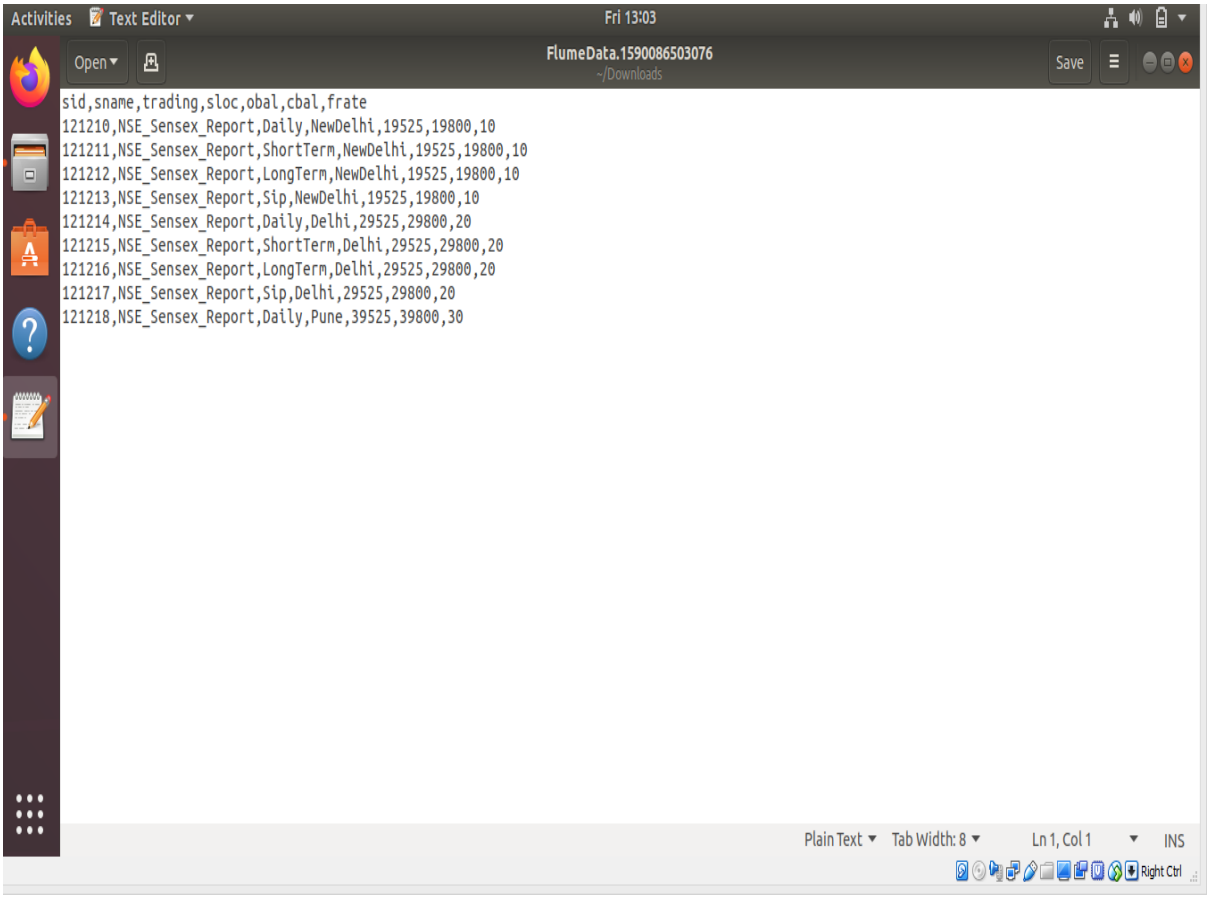
Browser view of Flume data files:

View of flume files loaded in hdfs in browser.

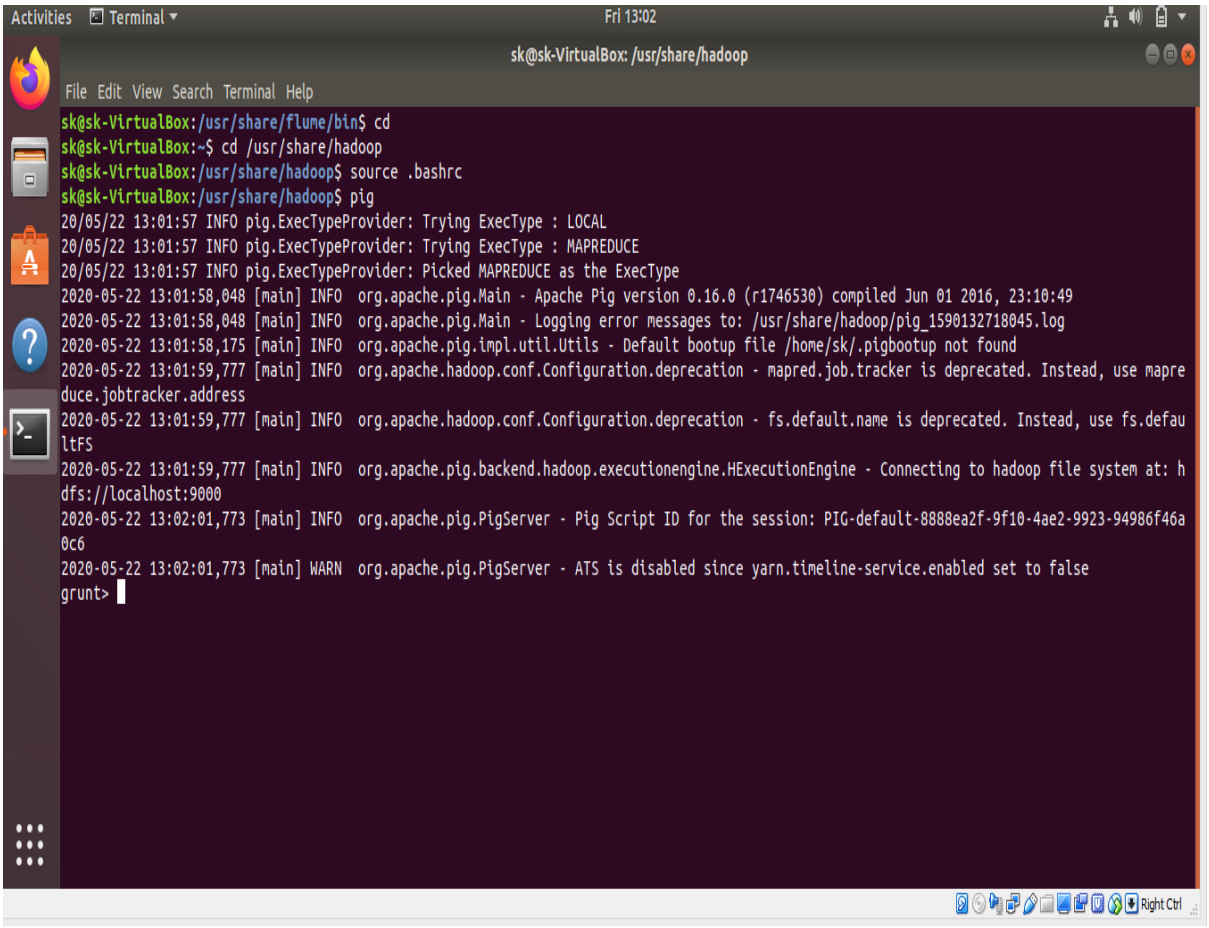


Sample view of flume data file:

Downloaded flume files from browser where data is transferred in the form of chunks.



Starting a pig:



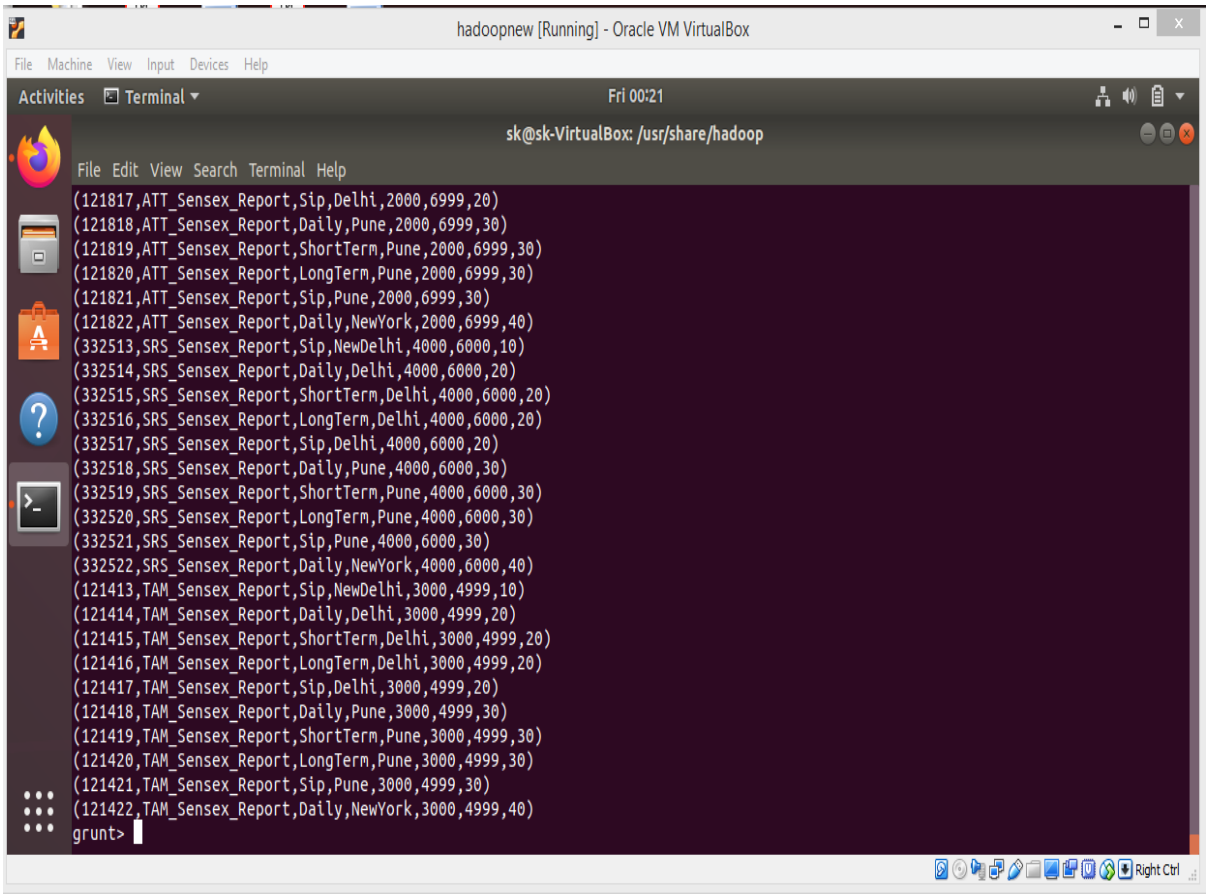
PIG Commands:

1. Display the logical, physical, and MapReduce execution plans of a relation.

A = LOAD '/flume_paul' USING PigStorage(',') as(sid:int,sname:chararray, trading:chararray,sloc:chararray,obal:int,cbal:int,frate:int);

```
grunt> A = load '/flume_paul' using PigStorage(',') as (sid:int,sname:chararray,trading:chararray,sloc:chararray,obal:int,cbal:
int,frate:int);
2020-05-22 00:20:13,110 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead,
 use fs.defaultFS
grunt> dump A;
```

Output:

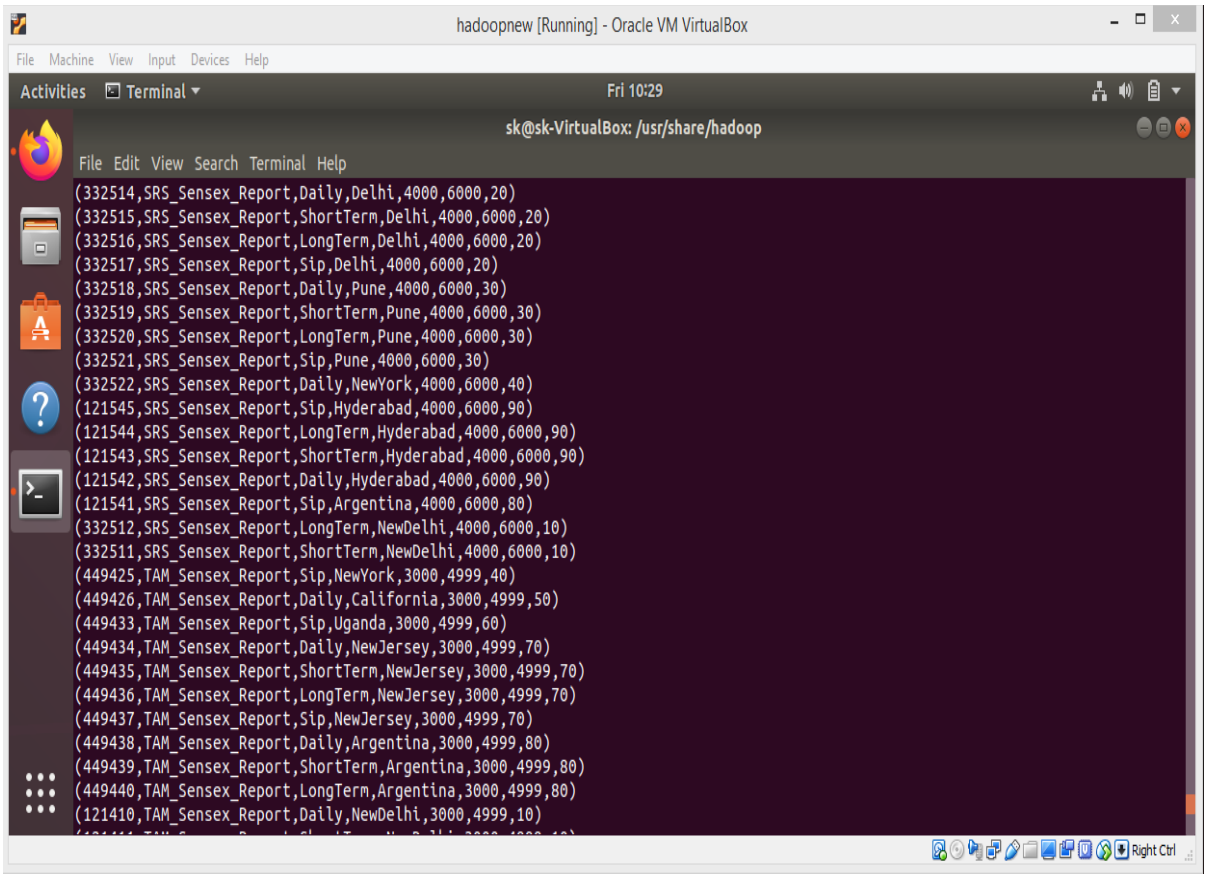


2.Excute Pig Command to List employee records in descending order of their Open Balance

Grunt> q2 = order A by $4 Desc
Grunt> dump q2

```
grunt> q2 = order A by $4 Desc;
grunt> dump q2;
```

Output

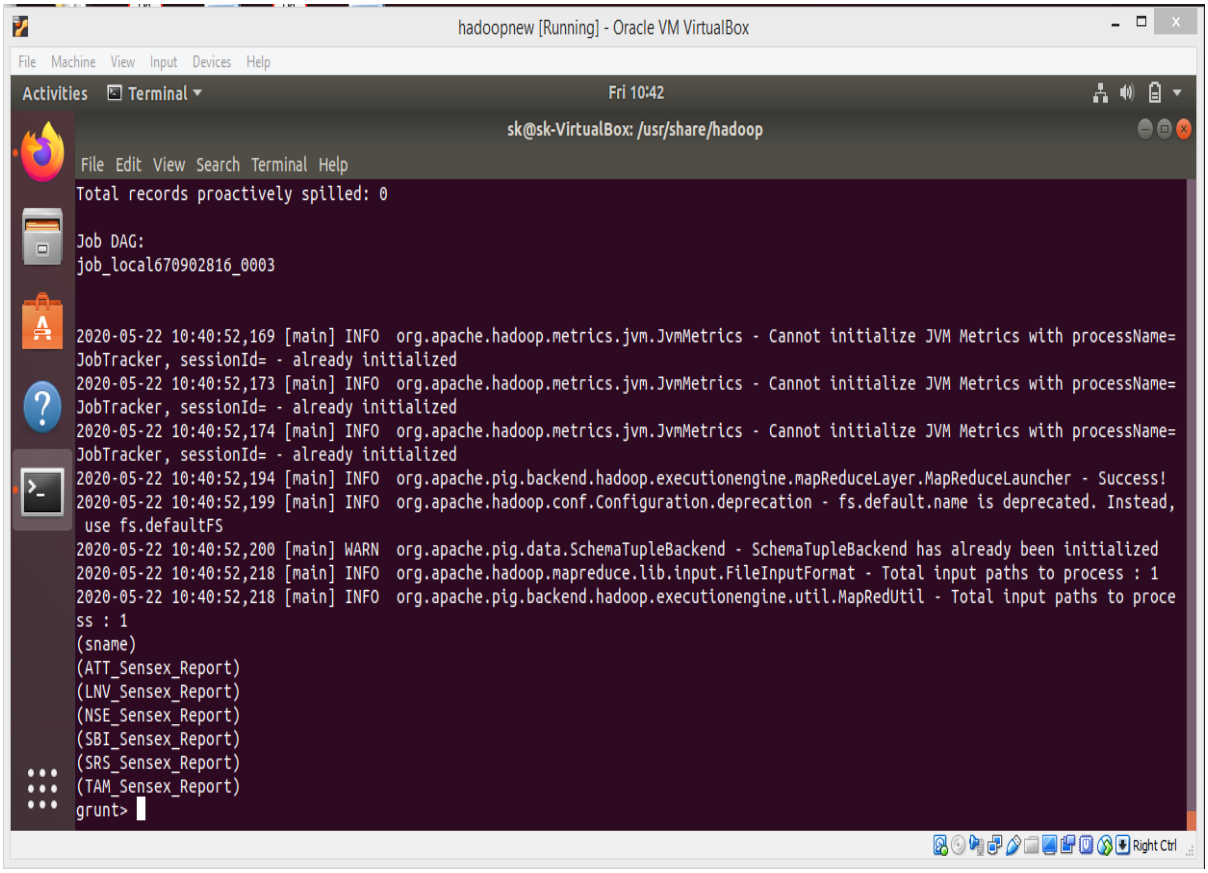


3.Excute Pig Command to List all the unique Sensex Name

Grunt> q3 = foreach A generate sname
Grunt> q31 = distinct q3;
Grunt> dump q31;

```
grunt> q3 = foreach A generate sname;
grunt> q31 = distinct q3;
grunt> dump q31;
```

Output

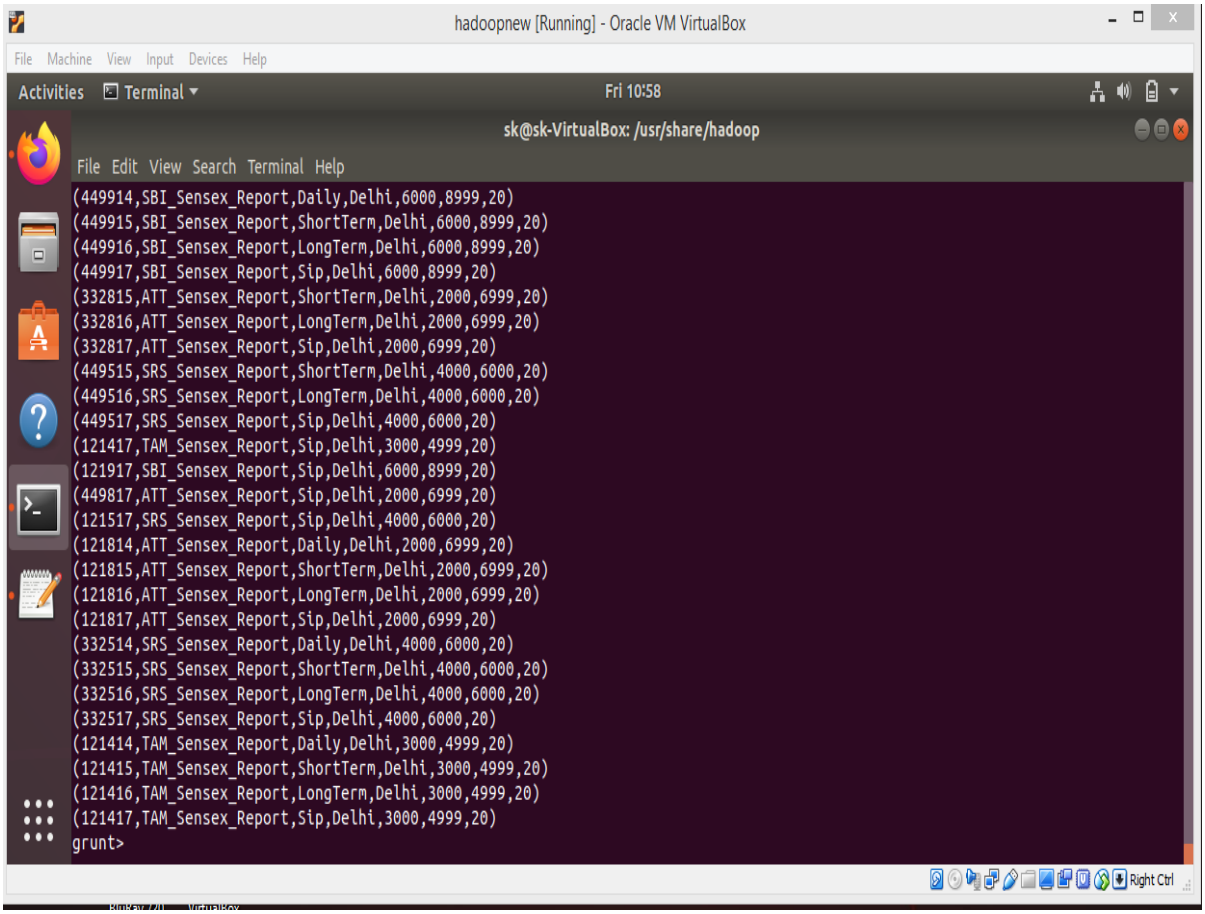


4.Excute Pig Command to List sensex location names starting with ‘D’ and having five characters

Grunt> q4 = filter A by (SIZE($3)==5 AND STARTSWITH($3,’D’));
Grunt> dump q4;

```
grunt> q4 = filter A by (SIZE($3)==5 AND STARTSWITH($3,'D'));
2020-05-22 10:54:34,015 [main] WARN  org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 3 time
(s).
grunt> dump q4;
```

Output

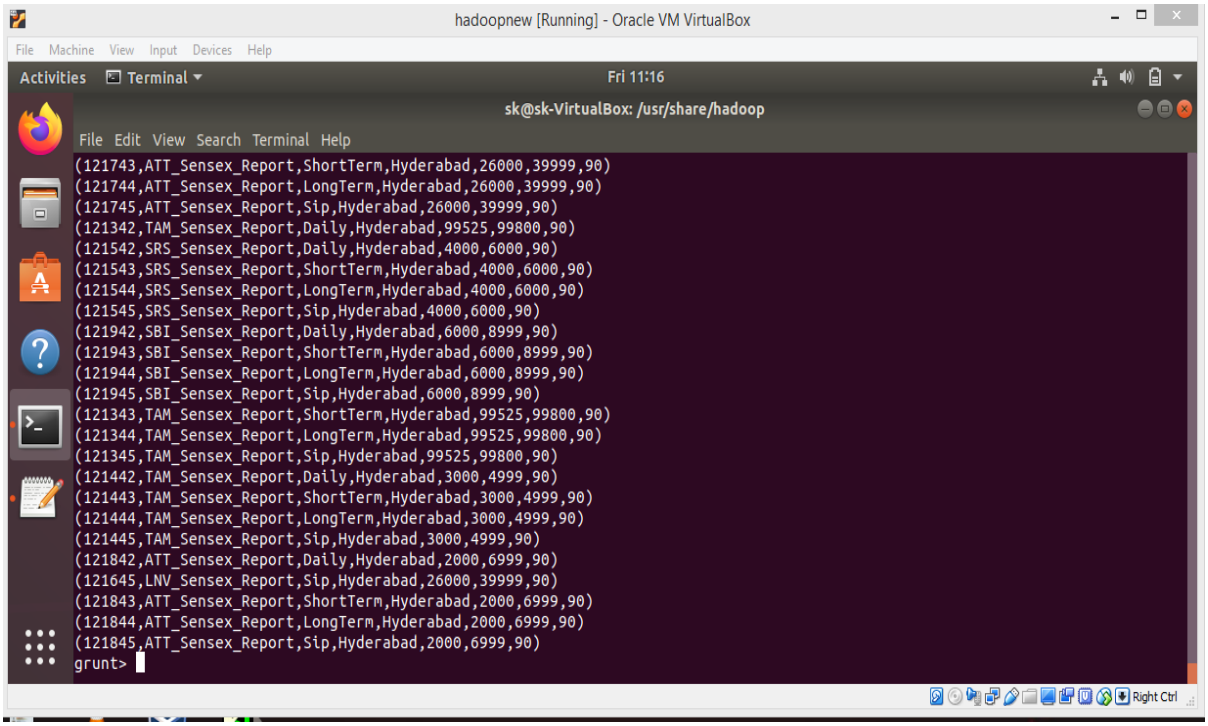


5.Excute Pig Command to List Sensex Location names starting with 'H' and ending with 'd' with 9-character length

Grunt> q5 = filter A by (SIZE($3)==9 AND ENDSWITH($3,'d') AND STARTSWITH($3,'H'));
Grunt>dump q5;

```
grunt> q5 = filter A by (SIZE($3)==9 AND ENDSWITH($3,'d') AND STARTSWITH($3,'H'));
2020-05-22 11:14:29,033 [main] WARN  org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 1 tim
e(s).
grunt> dump q5;
```

output

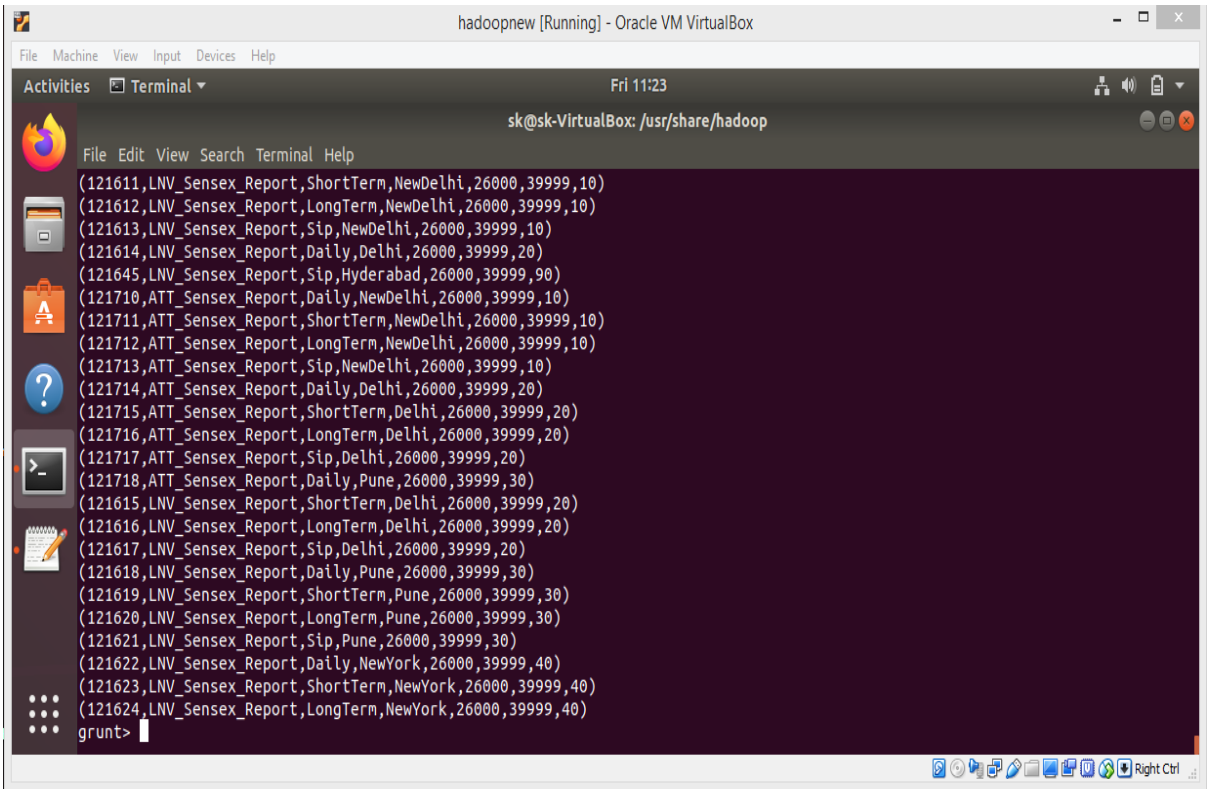


6.Excute Pig Command to List the closing balance is 5-digit number ending with nine

Grunt> q6 = filter A by (Chararray)$5 matches '....9';
Grunt> dump q6;

```
grunt> q6 = filter A by (Chararray)$5 matches '....9';
2020-05-22 11:19:57,886 [main] WARN  org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 1 tim
e(s).
grunt> dump q6;
```

Output

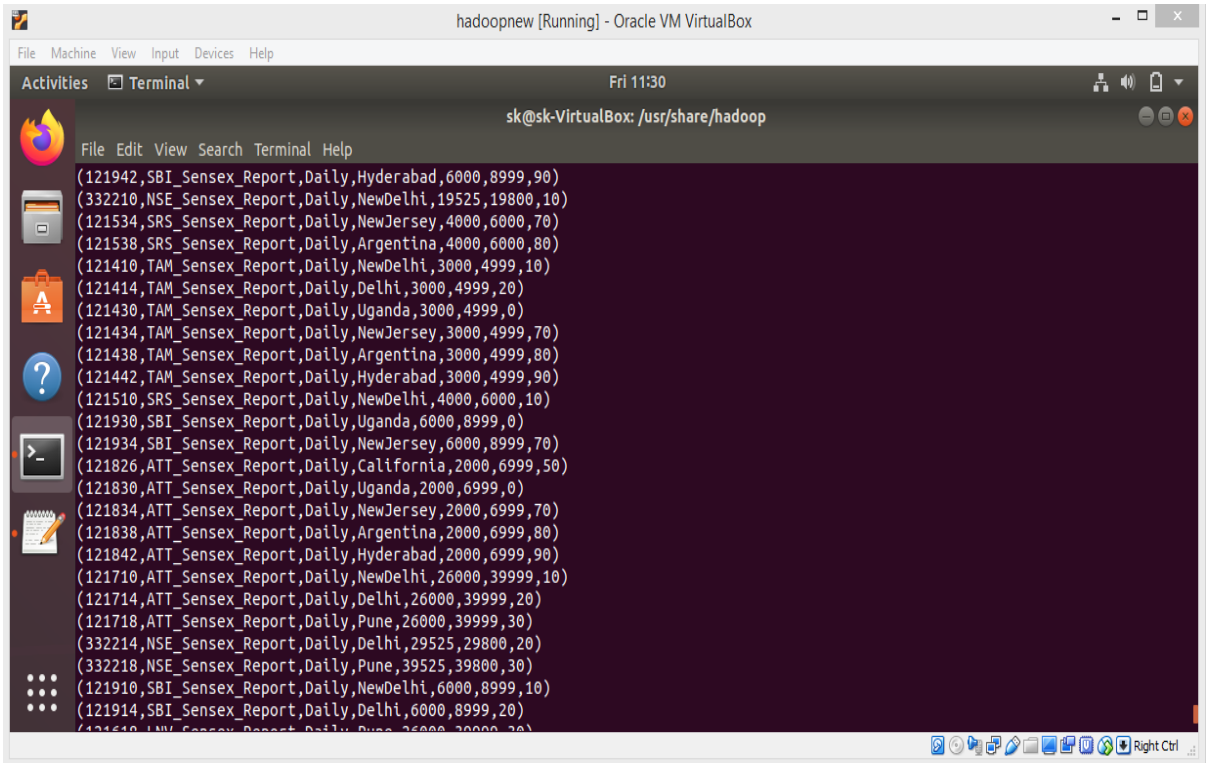


7.Excute Pig Command to List all trading except 'ShortTerm' & 'LongTerm'&'Sip' ascending order of salaries

Grunt> q7 = filter A by trading !='ShortTerm' AND trading !='LongTerm' AND trading !='Sip';
Grunt> dump q7;

```
grunt> q7 = filter A by trading !='ShortTerm' AND trading !='LongTerm' AND trading !='Sip';
2020-05-22 11:27:43,584 [main] WARN  org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 1 tim
e(s).
grunt> dump q7;
```

Output

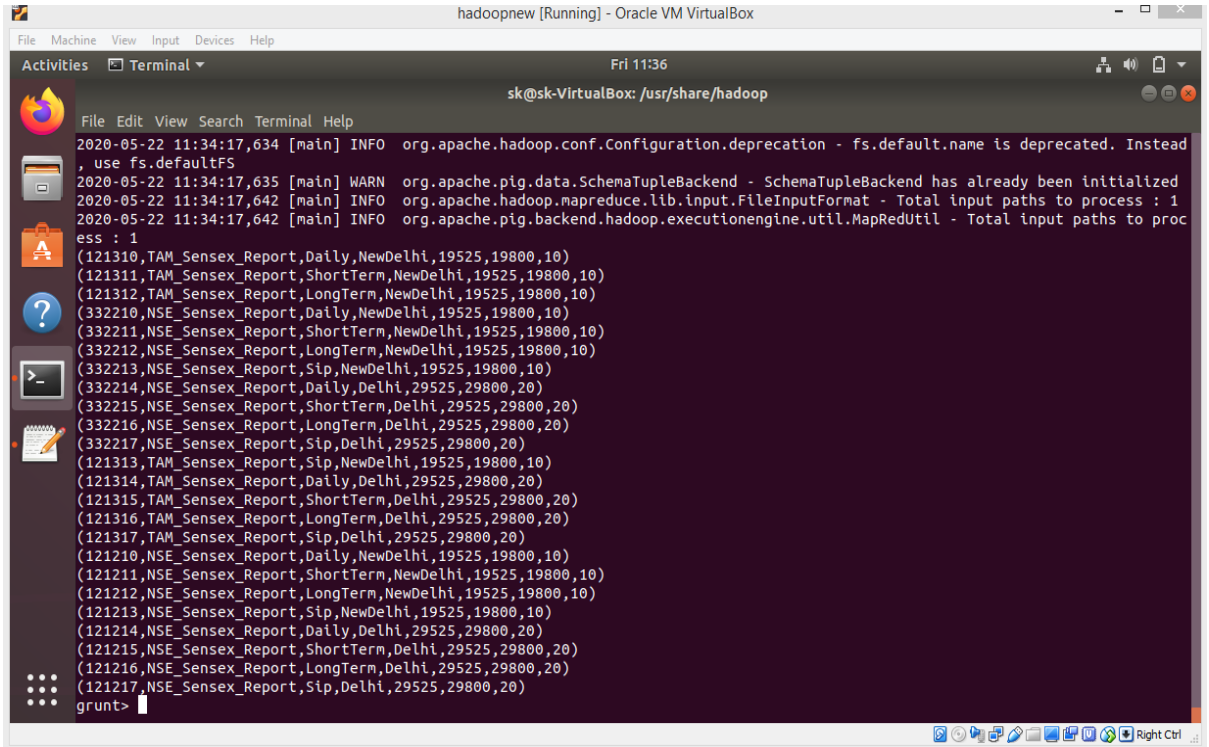


8.Excute Pig Command to List sensex data whose closing balance is ranging from 10000 and 30000

Grunt> q8 = filter A by (cbal > 10000) and (cbal < 30000);
Grunt> dump q8;

```
grunt> q8 = filter A by (cbal > 10000) and (cbal < 30000);
2020-05-22 11:34:03,643 [main] WARN  org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 1 tim
e(s).
grunt> dump q8;
```

Output

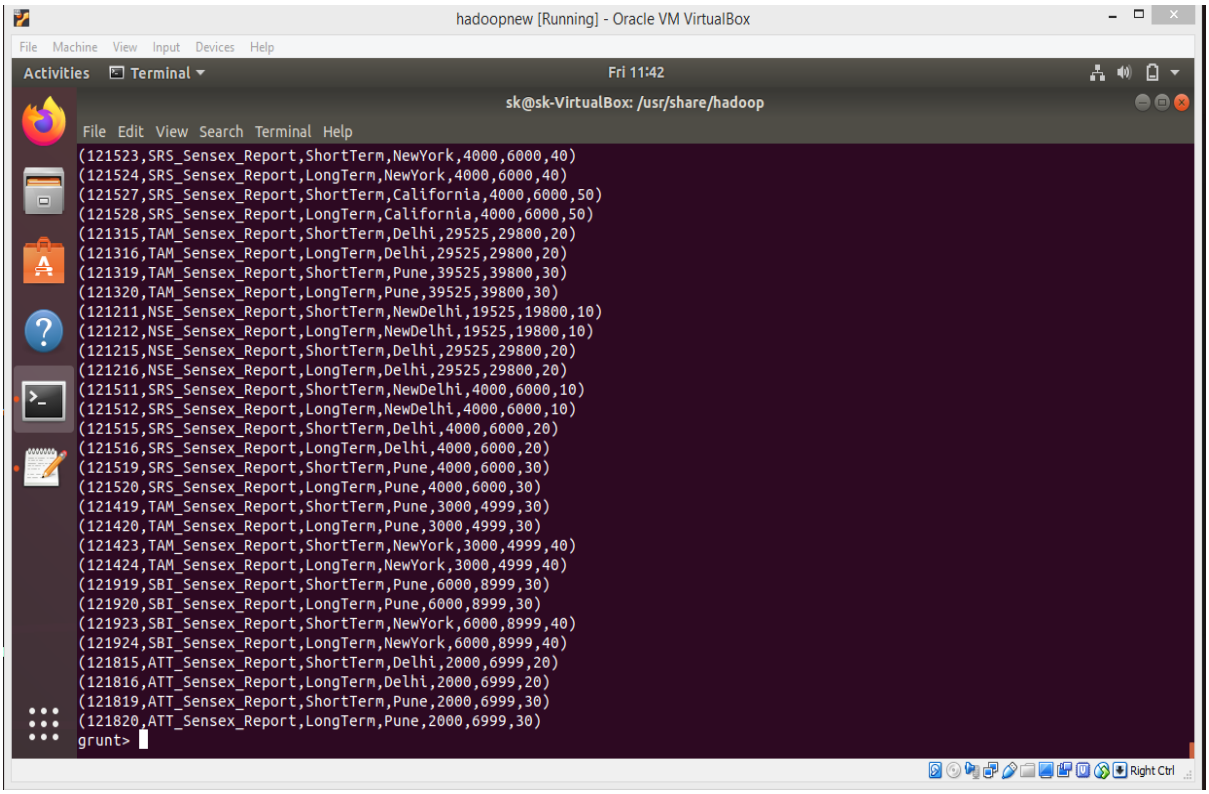


9.Excute Pig Command to List trading data are either ‘ShortTerm’ or ‘LongTerm’.

Grunt> q9 = filter A by (trading==’ShortTerm’) or (trading==’LongTerm’);
Grunt> dump q9

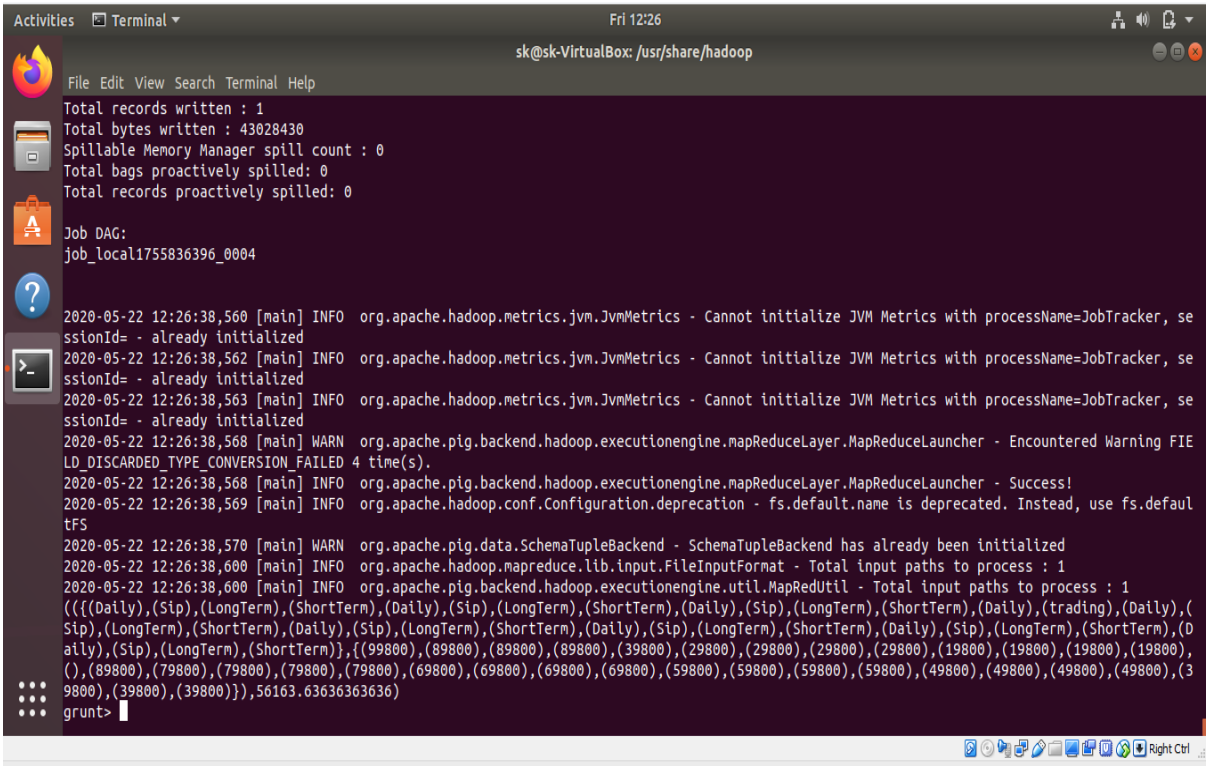
```
grunt> q9 = filter A by (trading=='ShortTerm') or (trading=='LongTerm');
2020-05-22 11:40:24,244 [main] WARN  org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_LONG 1 tim
e(s).
grunt> dump q9;
```

Output



10.Excute Pig Command and to List the trading and average of closing balance.

Grunt> q10 = group A all;
Grunt> q101 = foreach q10 generate (A.trading,A.cbal), AVG(A.cbal);
Grunt> dump q101;

```
grunt> q10 = group A all;
grunt> q101 = foreach q10 generate (A.trading, A.cbal), AVG(A.cbal);
grunt> dump q101;
```

Output