



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

Ms. B N Nithya (B1)

Signature

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

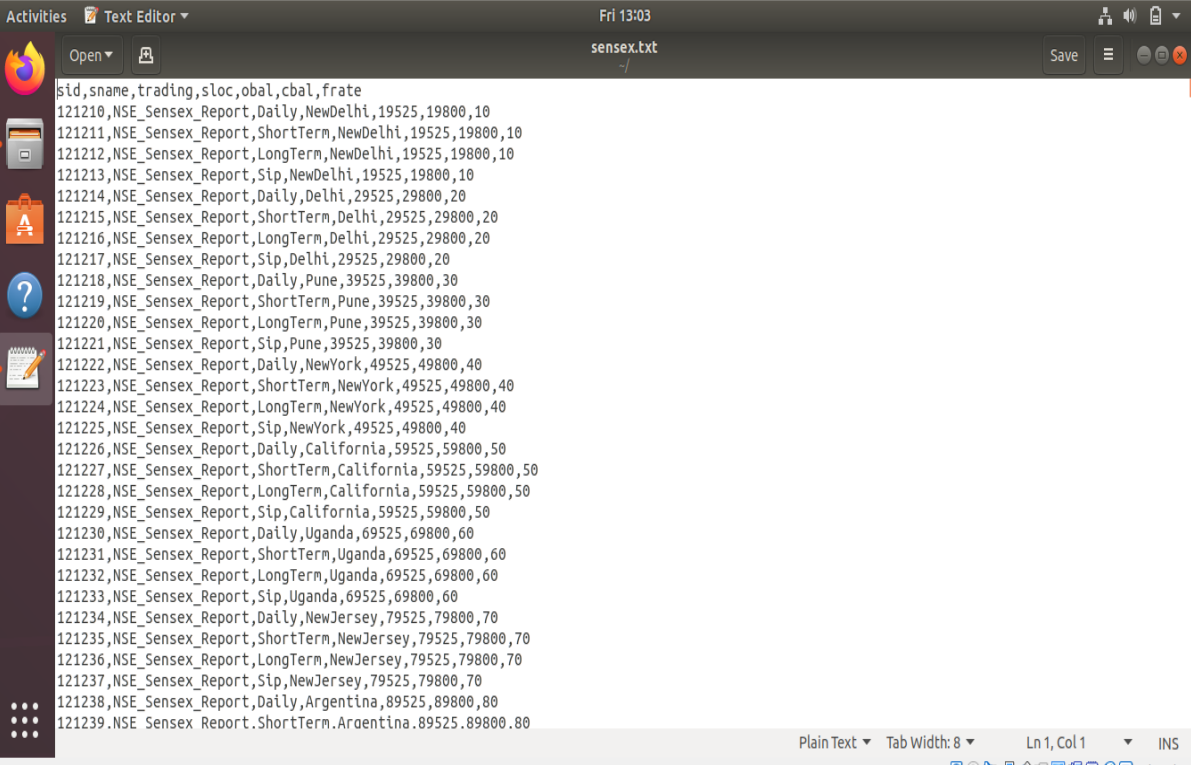
Flume is used to load a original data(sensex.txt file) which is in local memory to HDFS, Hadoop is used to store the data from flume, by creating a directory in it.

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.

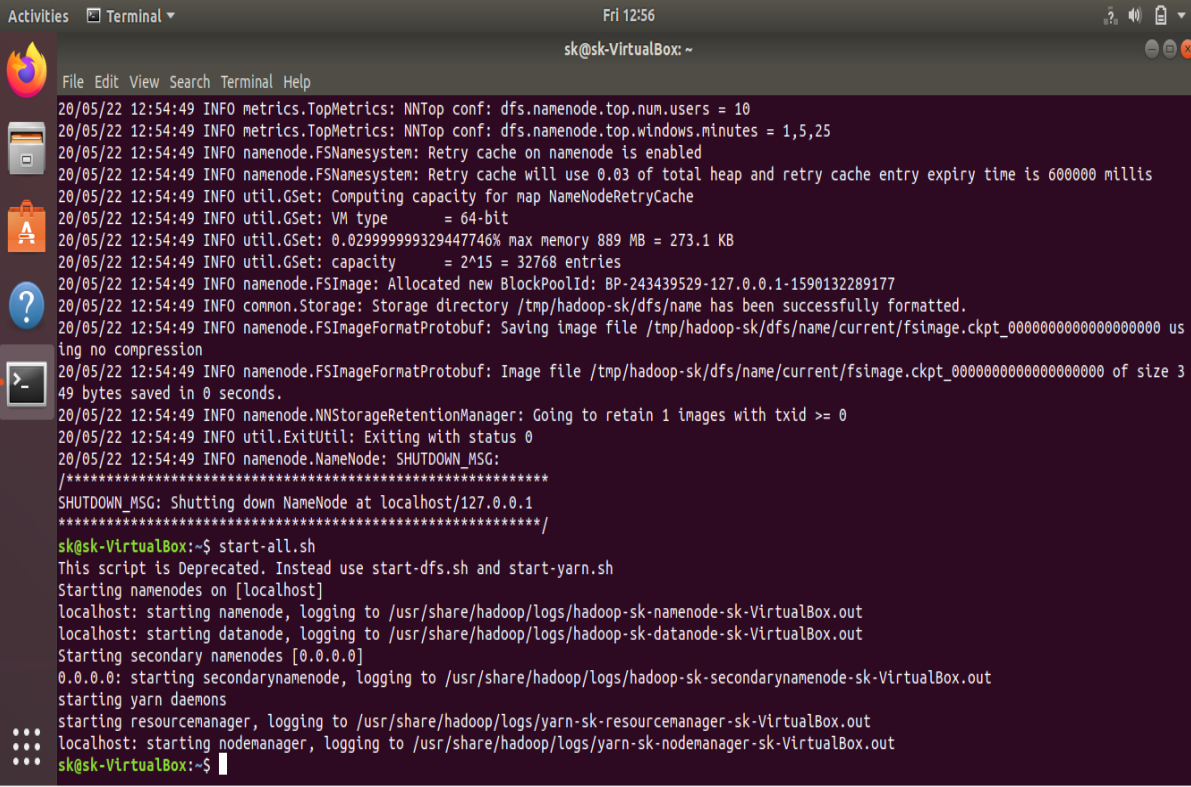
The fields contain **sid, sname, trading, sloc, obal, cbal, frate**



```
sid,sname,trading,sloc,obal,cbal,frate
121210,NSE_Sensex_Report,Daily,NewDelhi,19525,19800,10
121211,NSE_Sensex_Report,ShortTerm,NewDelhi,19525,19800,10
121212,NSE_Sensex_Report,LongTerm,NewDelhi,19525,19800,10
121213,NSE_Sensex_Report,Sip,NewDelhi,19525,19800,10
121214,NSE_Sensex_Report,Daily,Delhi,29525,29800,20
121215,NSE_Sensex_Report,ShortTerm,Delhi,29525,29800,20
121216,NSE_Sensex_Report,LongTerm,Delhi,29525,29800,20
121217,NSE_Sensex_Report,Sip,Delhi,29525,29800,20
121218,NSE_Sensex_Report,Daily,Pune,39525,39800,30
121219,NSE_Sensex_Report,ShortTerm,Pune,39525,39800,30
121220,NSE_Sensex_Report,LongTerm,Pune,39525,39800,30
121221,NSE_Sensex_Report,Sip,Pune,39525,39800,30
121222,NSE_Sensex_Report,Daily,NewYork,49525,49800,40
121223,NSE_Sensex_Report,ShortTerm,NewYork,49525,49800,40
121224,NSE_Sensex_Report,LongTerm,NewYork,49525,49800,40
121225,NSE_Sensex_Report,Sip,NewYork,49525,49800,40
121226,NSE_Sensex_Report,Daily,California,59525,59800,50
121227,NSE_Sensex_Report,ShortTerm,California,59525,59800,50
121228,NSE_Sensex_Report,LongTerm,California,59525,59800,50
121229,NSE_Sensex_Report,Sip,California,59525,59800,50
121230,NSE_Sensex_Report,Daily,Uganda,69525,69800,60
121231,NSE_Sensex_Report,ShortTerm,Uganda,69525,69800,60
121232,NSE_Sensex_Report,LongTerm,Uganda,69525,69800,60
121233,NSE_Sensex_Report,Sip,Uganda,69525,69800,60
121234,NSE_Sensex_Report,Daily,NewJersey,79525,79800,70
121235,NSE_Sensex_Report,ShortTerm,NewJersey,79525,79800,70
121236,NSE_Sensex_Report,LongTerm,NewJersey,79525,79800,70
121237,NSE_Sensex_Report,Sip,NewJersey,79525,79800,70
121238,NSE_Sensex_Report,Daily,Argentina,89525,89800,80
121239,NSE_Sensex_Report,ShortTerm,Argentina,89525,89800,80
```

Command to check services in Hadoop:

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

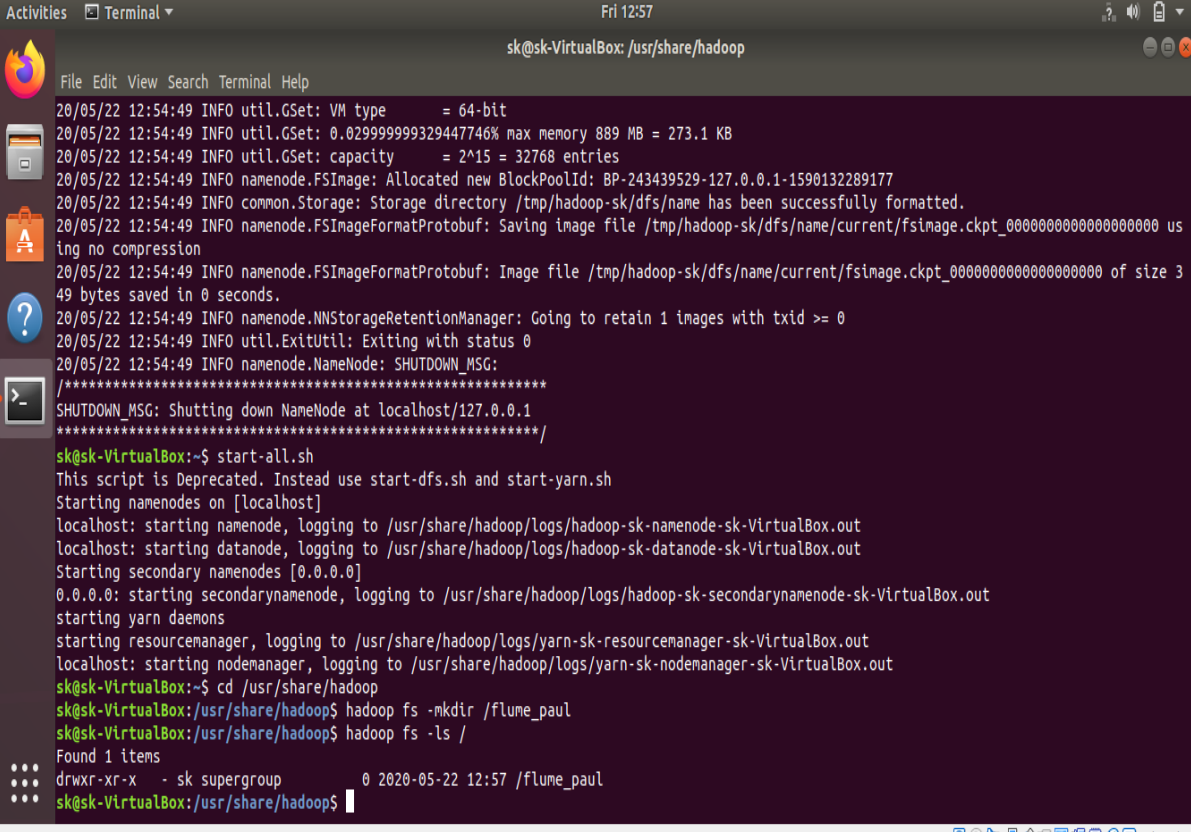


The screenshot shows a terminal window titled "sk@sk-VirtualBox: ~" with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays a series of log messages from Hadoop components, including metrics, namenode configuration, and storage formatting. After the logs, the user enters the command `start-all.sh`. The output indicates that the script is deprecated and suggests using `start-dfs.sh` and `start-yarn.sh` instead. It then shows the process of starting namenodes on localhost, starting datanodes, and starting yarn daemons, with logging paths specified for each.

```
20/05/22 12:54:49 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
20/05/22 12:54:49 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
20/05/22 12:54:49 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
20/05/22 12:54:49 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
20/05/22 12:54:49 INFO util.GSet: Computing capacity for map NameNodeRetryCache
20/05/22 12:54:49 INFO util.GSet: VM type = 64-bit
20/05/22 12:54:49 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
20/05/22 12:54:49 INFO util.GSet: capacity = 2^15 = 32768 entries
20/05/22 12:54:49 INFO namenode.FSImage: Allocated new BlockPoolId: BP-243439529-127.0.0.1-1590132289177
20/05/22 12:54:49 INFO common.Storage: Storage directory /tmp/hadoop-sk/dfs/name has been successfully formatted.
20/05/22 12:54:49 INFO namenode.FSImageFormatProtobuf: Saving image file /tmp/hadoop-sk/dfs/name/current/fsimage.ckpt_00000000000000000000 using no compression
20/05/22 12:54:49 INFO namenode.FSImageFormatProtobuf: Image file /tmp/hadoop-sk/dfs/name/current/fsimage.ckpt_00000000000000000000 of size 349 bytes saved in 0 seconds.
20/05/22 12:54:49 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
20/05/22 12:54:49 INFO util.ExitUtil: Exiting with status 0
20/05/22 12:54:49 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at localhost/127.0.0.1
*****/
sk@sk-VirtualBox:~$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/share/hadoop/logs/hadoop-sk-namenode-sk-VirtualBox.out
localhost: starting datanode, logging to /usr/share/hadoop/logs/hadoop-sk-datanode-sk-VirtualBox.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/share/hadoop/logs/hadoop-sk-secondarynamenode-sk-VirtualBox.out
starting yarn daemons
starting resourcemanager, logging to /usr/share/hadoop/logs/yarn-sk-resourcemanager-sk-VirtualBox.out
localhost: starting nodemanager, logging to /usr/share/hadoop/logs/yarn-sk-nodemanager-sk-VirtualBox.out
sk@sk-VirtualBox:~$
```

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.

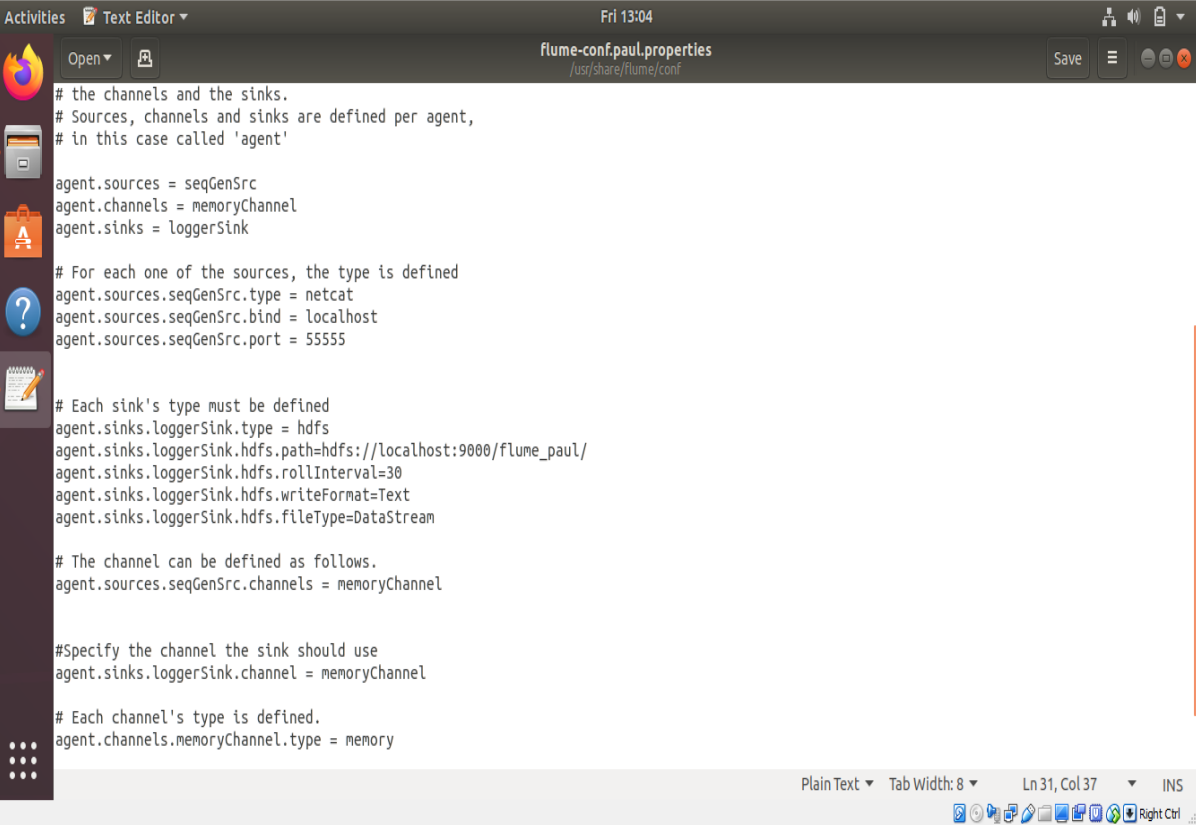


```
Activities  Terminal  Fri 12:57
sk@sk-VirtualBox: /usr/share/hadoop

20/05/22 12:54:49 INFO util.GSet: VM type      = 64-bit
20/05/22 12:54:49 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
20/05/22 12:54:49 INFO util.GSet: capacity    = 2^15 = 32768 entries
20/05/22 12:54:49 INFO namenode.FSImage: Allocated new BlockPoolId: BP-243439529-127.0.0.1-1590132289177
20/05/22 12:54:49 INFO common.Storage: Storage directory /tmp/hadoop-sk/dfs/name has been successfully formatted.
20/05/22 12:54:49 INFO namenode.FSImageFormatProtobuf: Saving image file /tmp/hadoop-sk/dfs/name/current/fsimage.ckpt_000000000000000000 using no compression
20/05/22 12:54:49 INFO namenode.FSImageFormatProtobuf: Image file /tmp/hadoop-sk/dfs/name/current/fsimage.ckpt_000000000000000000 of size 349 bytes saved in 0 seconds.
20/05/22 12:54:49 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
20/05/22 12:54:49 INFO util.ExitUtil: Exiting with status 0
20/05/22 12:54:49 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at localhost/127.0.0.1
*****/
sk@sk-VirtualBox:~$ start-all.sh
This script is deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/share/hadoop/logs/hadoop-sk-namenode-sk-VirtualBox.out
localhost: starting datanode, logging to /usr/share/hadoop/logs/hadoop-sk-datanode-sk-VirtualBox.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/share/hadoop/logs/hadoop-sk-secondarynamenode-sk-VirtualBox.out
starting yarn daemons
starting resourcemanager, logging to /usr/share/hadoop/logs/yarn-sk-resourcemanager-sk-VirtualBox.out
localhost: starting nodemanager, logging to /usr/share/hadoop/logs/yarn-sk-nodemanager-sk-VirtualBox.out
sk@sk-VirtualBox:~$ cd /usr/share/hadoop
sk@sk-VirtualBox:/usr/share/hadoop$ hadoop fs -mkdir /flume_paul
sk@sk-VirtualBox:/usr/share/hadoop$ hadoop fs -ls /
Found 1 items
drwxr-xr-x  - sk supergroup          0 2020-05-22 12:57 /flume_paul
sk@sk-VirtualBox:/usr/share/hadoop$
```

Flume Configuration image:

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



The image shows a text editor window titled "flume-conf.paul.properties" with the file path "/usr/share/flume/conf". The editor contains the following configuration:

```
# the channels and the sinks.
# Sources, channels and sinks are defined per agent,
# in this case called 'agent'

agent.sources = seqGenSrc
agent.channels = memoryChannel
agent.sinks = loggerSink

# For each one of the sources, the type is defined
agent.sources.seqGenSrc.type = netcat
agent.sources.seqGenSrc.bind = localhost
agent.sources.seqGenSrc.port = 55555

# Each sink's type must be defined
agent.sinks.loggerSink.type = hdfs
agent.sinks.loggerSink.hdfs.path=hdfs://localhost:9000/flume_paul/
agent.sinks.loggerSink.hdfs.rollInterval=30
agent.sinks.loggerSink.hdfs.writeFormat=Text
agent.sinks.loggerSink.hdfs.fileType=DataStream

# The channel can be defined as follows.
agent.sources.seqGenSrc.channels = memoryChannel

#Specify the channel the sink should use
agent.sinks.loggerSink.channel = memoryChannel

# Each channel's type is defined.
agent.channels.memoryChannel.type = memory
```

The status bar at the bottom indicates "Plain Text", "Tab Width: 8", "Ln 31, Col 37", and "INS".

Transferring 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`

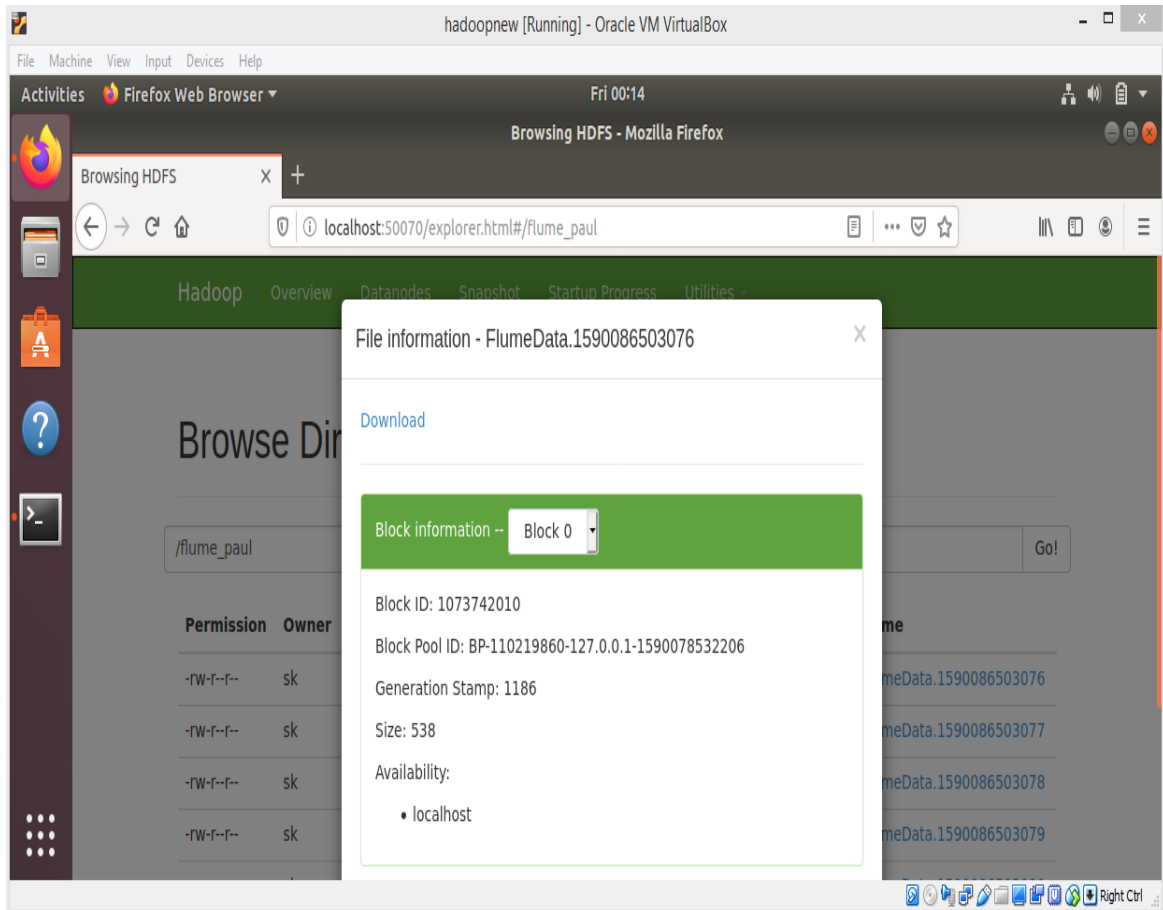
A screenshot of a Linux terminal window. The title bar at the top shows "Activities", "Terminal", and the time "Fri 13:00". The terminal's address bar displays "sk@sk-VirtualBox: /usr/share/flume/bin". The prompt is "sk@sk-VirtualBox:/usr/share/hadoop\$". The user has entered "cd /usr/share/flume/bin" and "ls -l". The output lists various Hadoop-related JAR files, including slf4j-api-1.7.10.jar, log4j-1.7.10.jar, hbase-client-facing-thirdparty-slf4j-api-1.7.25.jar, activation-1.1.jar, apacheds-l18n-2.0.0-M15.jar, api-asn1-api-1.0.0-M20.jar, asm-3.2.jar, commons-beanutils-1.7.0.jar, commons-cli-1.2.jar, commons-codec-1.4.jar, commons-collections-3.2.2.jar, commons-compress-1.4.1.jar, commons-configuration-1.6.jar, commons-digester-1.8.jar, commons-httpclient-3.1.jar, commons-io-2.4.jar, commons-lang-2.6.jar, commons-logging-1.1.3.jar, commons-math3-3.1.1.jar, commons-net-3.1.jar, curator-client-2.7.1.jar, curator-framework-2.7.1.jar, curator-recipes-2.7.1.jar, gson-2.2.4.jar, annotations-2.7.3.jar, aut-h-2.7.3.jar, hamcrest-core-1.3.jar, htrace-core-3.1.0-incubating.jar, httpclient-4.2.5.jar, httpcore-4.2.5.jar, jackson-core-asl-1.9.13.jar, jackson-jaxrs-1.9.13.jar, mapper-asl-1.9.13.jar, xmlbuilder-0.4.jar, jaxb-api-2.2.2.jar, jersey-core-1.9.jar, jersey-server-1.9.jar, jets3t-0.9.0.jar, and jettison-1.1.jar. The terminal interface includes standard icons on the left (Firefox, Files, Applications, Help) and system status icons on the right.

Successfully Transferring of data into hdf5.

[illegible]

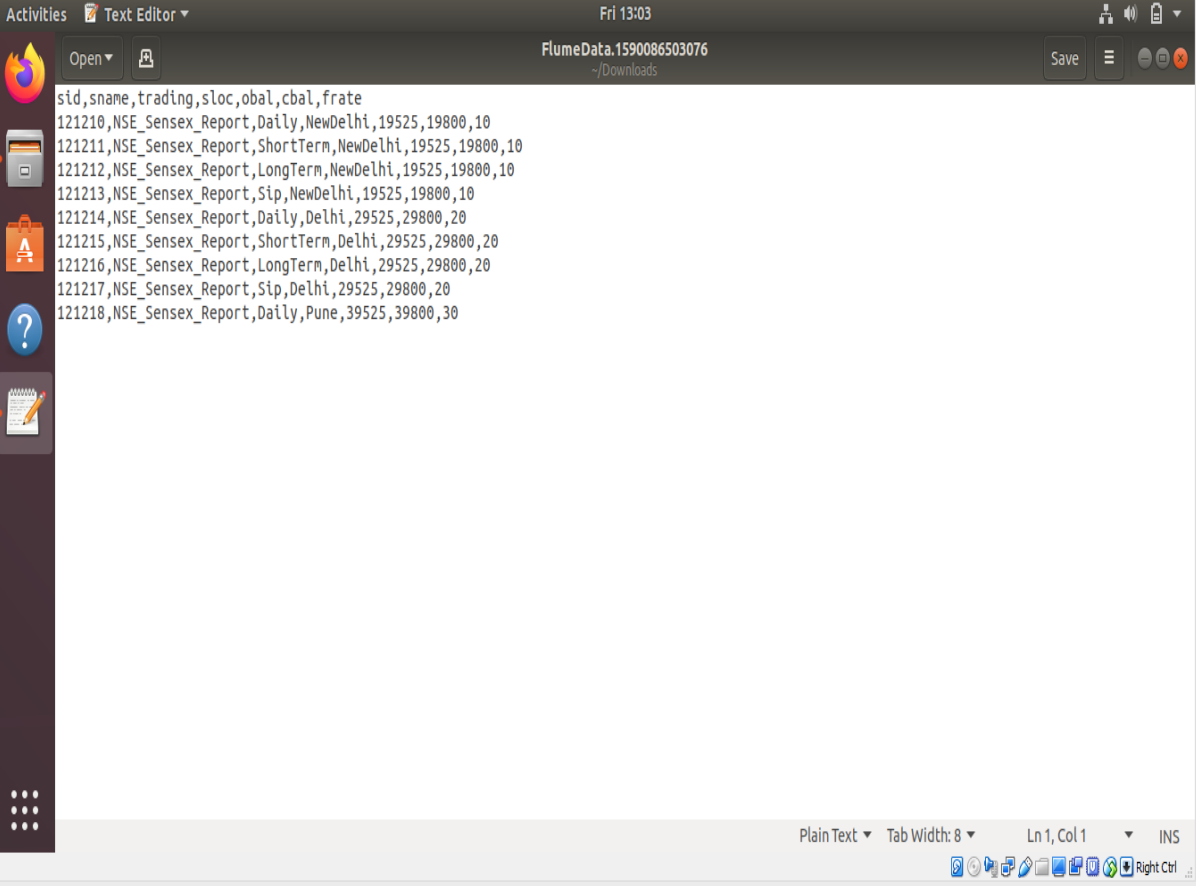
Browser view of Flume data files:

View of flume files loaded in hdfs in browser.



Sample view of flume data file:

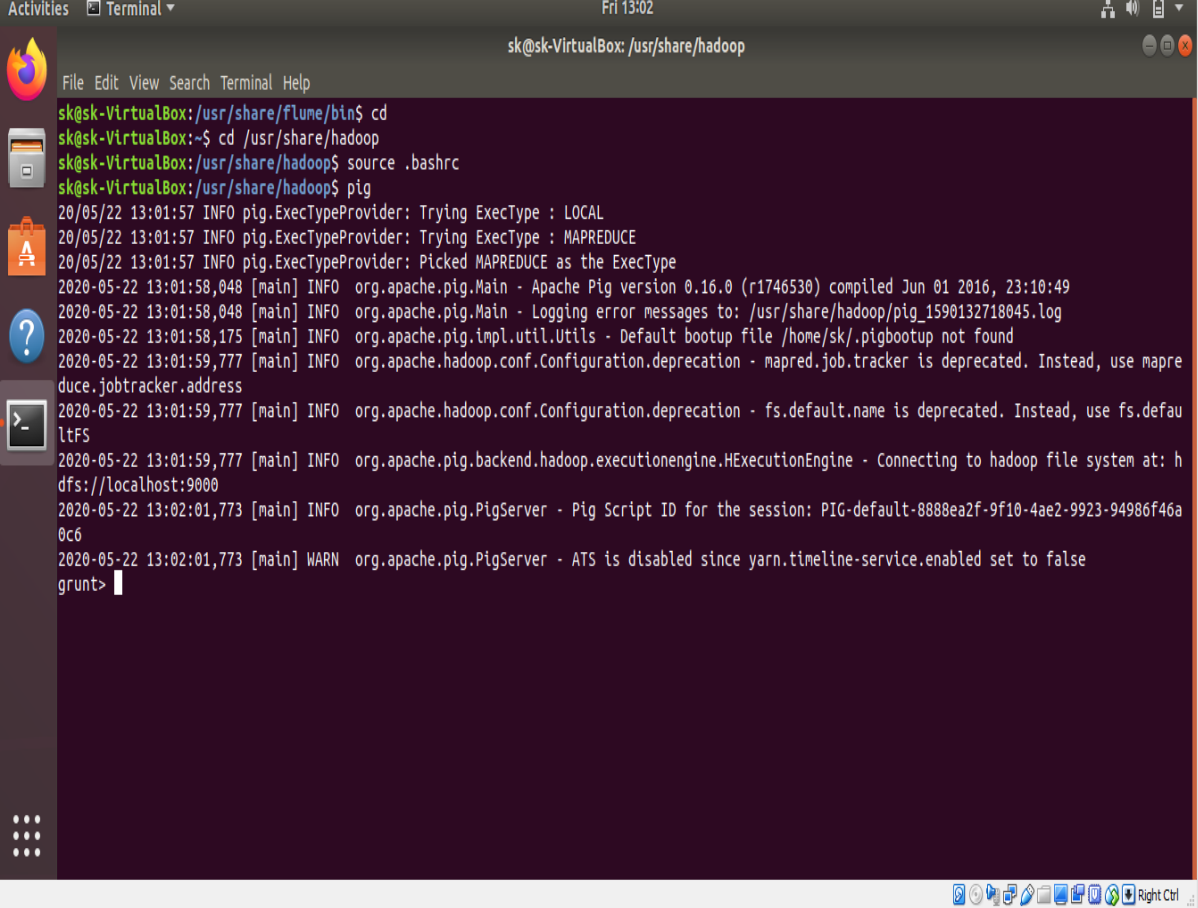
Downloaded flume files from browser where data is transferred in the form of chunks.



The screenshot shows a Linux desktop environment. A text editor window titled 'FlumeData.1590086503076' is open, displaying a list of data records. The records are comma-separated values (CSV) representing various stock market data points. The desktop background is dark, and the text editor has a light background. The window title bar includes a 'Save' button and a close button. The bottom status bar shows 'Plain Text', 'Tab Width: 8', 'Ln 1, Col 1', and 'INS'.

```
sid,sname,trading,sloc,obal,cbal,frate
121210,NSE_Sensex_Report,Daily,NewDelhi,19525,19800,10
121211,NSE_Sensex_Report,ShortTerm,NewDelhi,19525,19800,10
121212,NSE_Sensex_Report,LongTerm,NewDelhi,19525,19800,10
121213,NSE_Sensex_Report,Sip,NewDelhi,19525,19800,10
121214,NSE_Sensex_Report,Daily,Delhi,29525,29800,20
121215,NSE_Sensex_Report,ShortTerm,Delhi,29525,29800,20
121216,NSE_Sensex_Report,LongTerm,Delhi,29525,29800,20
121217,NSE_Sensex_Report,Sip,Delhi,29525,29800,20
121218,NSE_Sensex_Report,Daily,Pune,39525,39800,30
```

Starting a pig:



```
sk@sk-VirtualBox: /usr/share/hadoop
File Edit View Search Terminal Help
sk@sk-VirtualBox: /usr/share/flume/bin$ cd
sk@sk-VirtualBox: ~$ cd /usr/share/hadoop
sk@sk-VirtualBox: /usr/share/hadoop$ source .bashrc
sk@sk-VirtualBox: /usr/share/hadoop$ pig
2020-05-22 13:01:57 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2020-05-22 13:01:57 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2020-05-22 13:01:57 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2020-05-22 13:01:58,048 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2020-05-22 13:01:58,048 [main] INFO org.apache.pig.Main - Logging error messages to: /usr/share/hadoop/pig_1590132718045.log
2020-05-22 13:01:58,175 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/sk/.pigbootup not found
2020-05-22 13:01:59,777 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2020-05-22 13:01:59,777 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2020-05-22 13:01:59,777 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2020-05-22 13:02:01,773 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-8888ea2f-9f10-4ae2-9923-94986f46a0c6
2020-05-22 13:02:01,773 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>
```

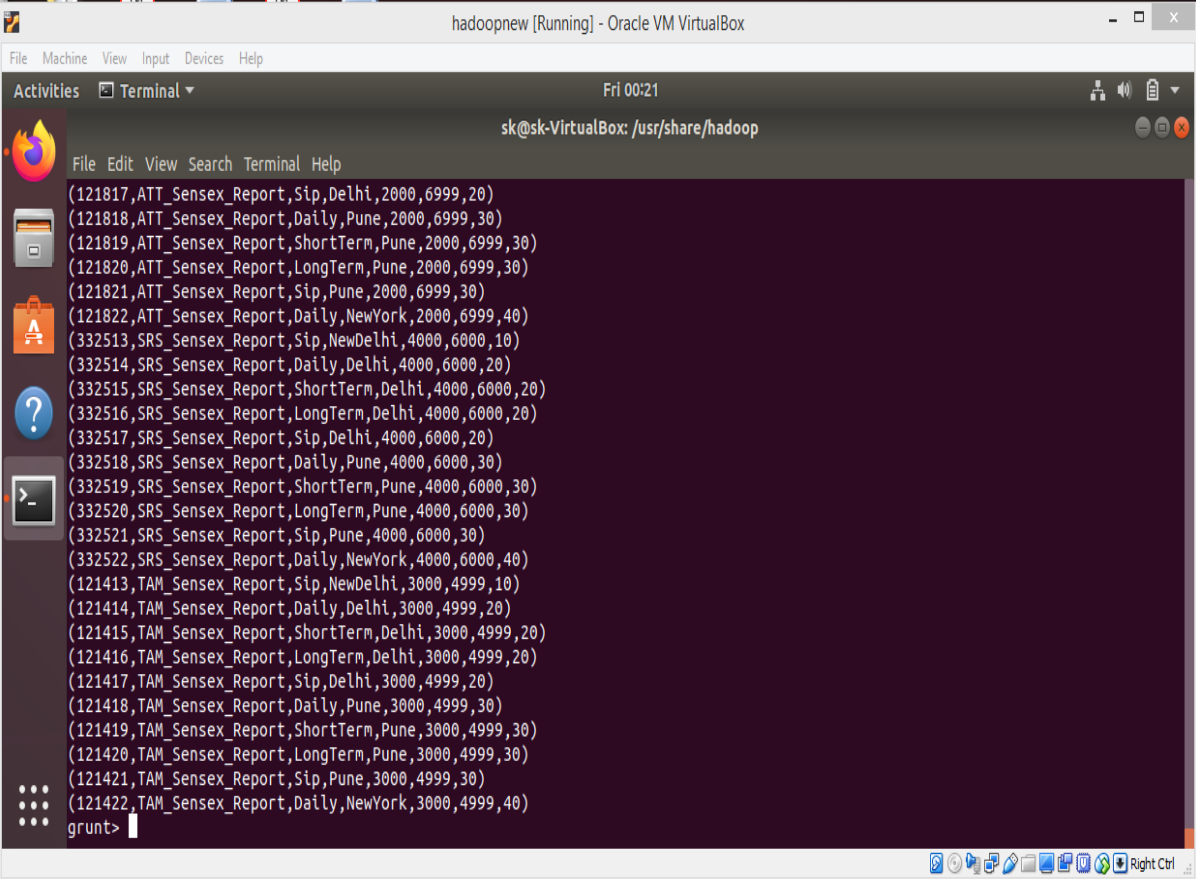
PIG Commands:

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:



```
hadoopnew [Running] - Oracle VM VirtualBox  
Fri 00:21  
sk@sk-VirtualBox: /usr/share/hadoop  
File Edit View Search Terminal Help  
(121817,ATT_Sensex_Report,Sip,Delhi,2000,6999,20)  
(121818,ATT_Sensex_Report,Daily,Pune,2000,6999,30)  
(121819,ATT_Sensex_Report,ShortTerm,Pune,2000,6999,30)  
(121820,ATT_Sensex_Report,LongTerm,Pune,2000,6999,30)  
(121821,ATT_Sensex_Report,Sip,Pune,2000,6999,30)  
(121822,ATT_Sensex_Report,Daily,NewYork,2000,6999,40)  
(332513,SRS_Sensex_Report,Sip,NewDelhi,4000,6000,10)  
(332514,SRS_Sensex_Report,Daily,Delhi,4000,6000,20)  
(332515,SRS_Sensex_Report,ShortTerm,Delhi,4000,6000,20)  
(332516,SRS_Sensex_Report,LongTerm,Delhi,4000,6000,20)  
(332517,SRS_Sensex_Report,Sip,Delhi,4000,6000,20)  
(332518,SRS_Sensex_Report,Daily,Pune,4000,6000,30)  
(332519,SRS_Sensex_Report,ShortTerm,Pune,4000,6000,30)  
(332520,SRS_Sensex_Report,LongTerm,Pune,4000,6000,30)  
(332521,SRS_Sensex_Report,Sip,Pune,4000,6000,30)  
(332522,SRS_Sensex_Report,Daily,NewYork,4000,6000,40)  
(121413,TAM_Sensex_Report,Sip,NewDelhi,3000,4999,10)  
(121414,TAM_Sensex_Report,Daily,Delhi,3000,4999,20)  
(121415,TAM_Sensex_Report,ShortTerm,Delhi,3000,4999,20)  
(121416,TAM_Sensex_Report,LongTerm,Delhi,3000,4999,20)  
(121417,TAM_Sensex_Report,Sip,Delhi,3000,4999,20)  
(121418,TAM_Sensex_Report,Daily,Pune,3000,4999,30)  
(121419,TAM_Sensex_Report,ShortTerm,Pune,3000,4999,30)  
(121420,TAM_Sensex_Report,LongTerm,Pune,3000,4999,30)  
(121421,TAM_Sensex_Report,Sip,Pune,3000,4999,30)  
(121422,TAM_Sensex_Report,Daily,NewYork,3000,4999,40)  
grunt>
```

2. Execute 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

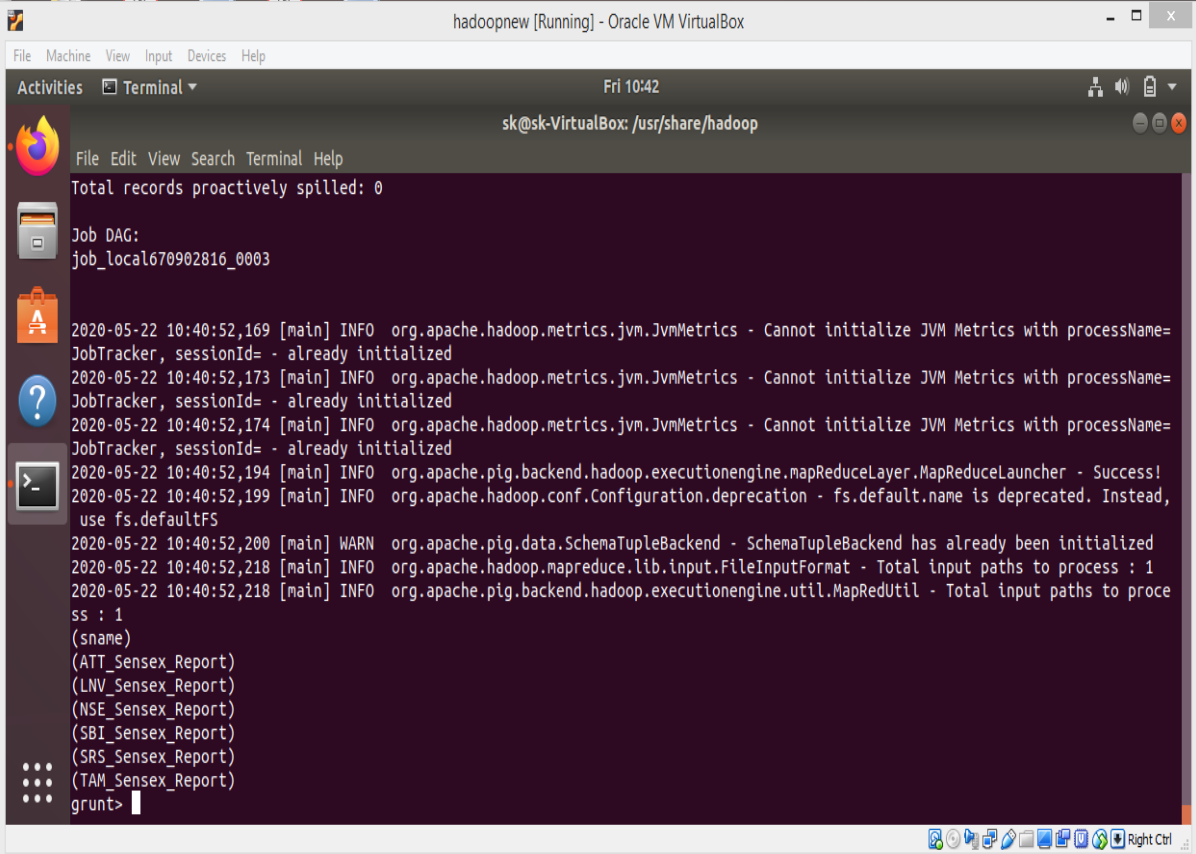
```
hadoopnew [Running] - Oracle VM VirtualBox  
File Machine View Input Devices Help  
Activities Terminal Fri 10:29  
sk@sk-VirtualBox: /usr/share/hadoop  
File Edit View Search Terminal Help  
(332514,SRS_Sensex_Report,Daily,Delhi,4000,6000,20)  
(332515,SRS_Sensex_Report,ShortTerm,Delhi,4000,6000,20)  
(332516,SRS_Sensex_Report,LongTerm,Delhi,4000,6000,20)  
(332517,SRS_Sensex_Report,Sip,Delhi,4000,6000,20)  
(332518,SRS_Sensex_Report,Daily,Pune,4000,6000,30)  
(332519,SRS_Sensex_Report,ShortTerm,Pune,4000,6000,30)  
(332520,SRS_Sensex_Report,LongTerm,Pune,4000,6000,30)  
(332521,SRS_Sensex_Report,Sip,Pune,4000,6000,30)  
(332522,SRS_Sensex_Report,Daily,NewYork,4000,6000,40)  
(121545,SRS_Sensex_Report,Sip,Hyderabad,4000,6000,90)  
(121544,SRS_Sensex_Report,LongTerm,Hyderabad,4000,6000,90)  
(121543,SRS_Sensex_Report,ShortTerm,Hyderabad,4000,6000,90)  
(121542,SRS_Sensex_Report,Daily,Hyderabad,4000,6000,90)  
(121541,SRS_Sensex_Report,Sip,Argentina,4000,6000,80)  
(332512,SRS_Sensex_Report,LongTerm,NewDelhi,4000,6000,10)  
(332511,SRS_Sensex_Report,ShortTerm,NewDelhi,4000,6000,10)  
(449425,TAM_Sensex_Report,Sip,NewYork,3000,4999,40)  
(449426,TAM_Sensex_Report,Daily,California,3000,4999,50)  
(449433,TAM_Sensex_Report,Sip,Uganda,3000,4999,60)  
(449434,TAM_Sensex_Report,Daily,NewJersey,3000,4999,70)  
(449435,TAM_Sensex_Report,ShortTerm,NewJersey,3000,4999,70)  
(449436,TAM_Sensex_Report,LongTerm,NewJersey,3000,4999,70)  
(449437,TAM_Sensex_Report,Sip,NewJersey,3000,4999,70)  
(449438,TAM_Sensex_Report,Daily,Argentina,3000,4999,80)  
(449439,TAM_Sensex_Report,ShortTerm,Argentina,3000,4999,80)  
(449440,TAM_Sensex_Report,LongTerm,Argentina,3000,4999,80)  
(121410,TAM_Sensex_Report,Daily,NewDelhi,3000,4999,10)  
(121411,TAM_Sensex_Report,ShortTerm,NewDelhi,3000,4999,10)
```

3. Execute 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

A screenshot of a terminal window titled 'hadoopnew [Running] - Oracle VM VirtualBox'. The terminal shows the output of a Pig command. It starts with 'Total records proactively spilled: 0' and 'Job DAG: job_local670902816_0003'. Then, it displays several log messages from the Hadoop ecosystem, including warnings and info messages about JVM metrics, configuration deprecation, and SchemaTupleBackend. Finally, it shows the output of the 'dump' command, which lists seven unique sensex names: (ATT_Sensex_Report), (LNV_Sensex_Report), (NSE_Sensex_Report), (SBI_Sensex_Report), (SRS_Sensex_Report), and (TAM_Sensex_Report). The prompt 'grunt>' is visible at the bottom.

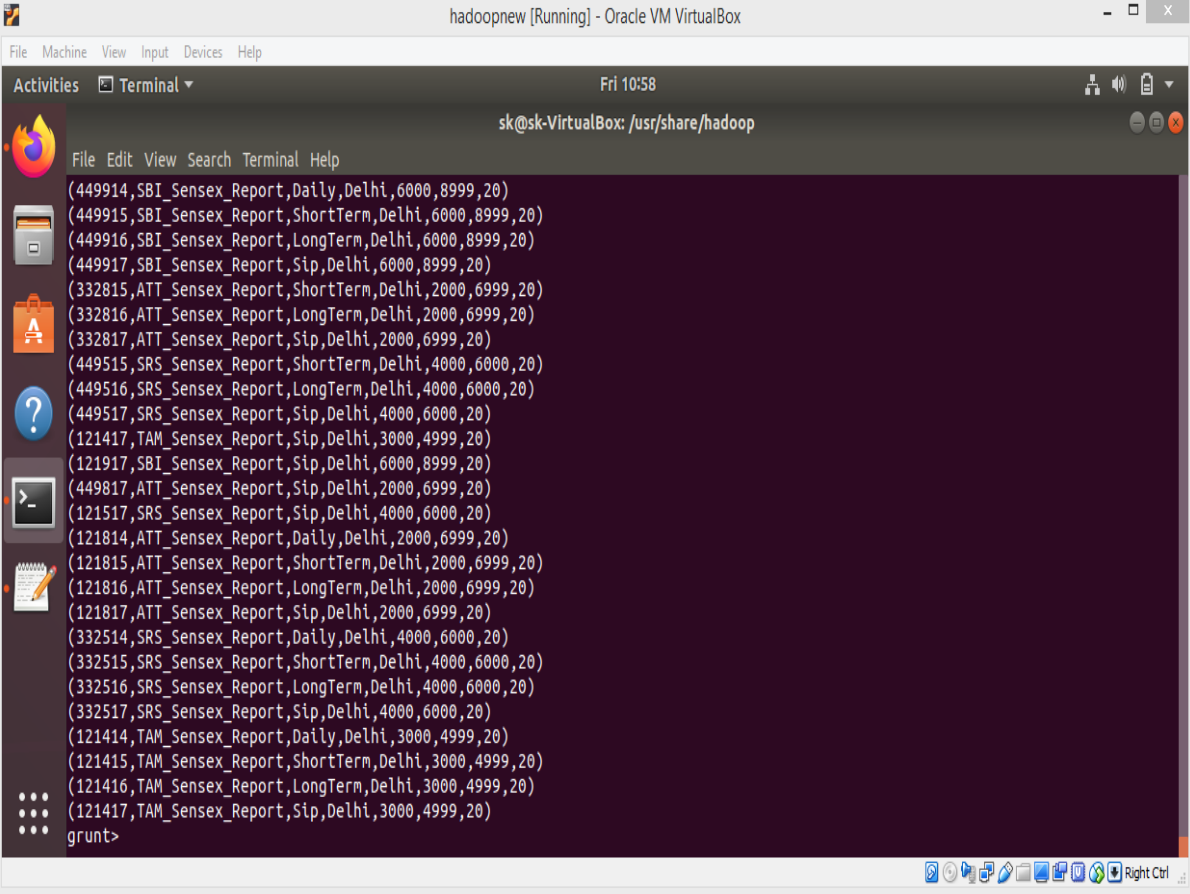
```
hadoopnew [Running] - Oracle VM VirtualBox  
File Machine View Input Devices Help  
Activities Terminal Fri 10:42  
sk@sk-VirtualBox: /usr/share/hadoop  
File Edit View Search Terminal Help  
Total records proactively spilled: 0  
Job DAG:  
job_local670902816_0003  
2020-05-22 10:40:52,169 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2020-05-22 10:40:52,173 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2020-05-22 10:40:52,174 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2020-05-22 10:40:52,194 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2020-05-22 10:40:52,199 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2020-05-22 10:40:52,200 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2020-05-22 10:40:52,218 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2020-05-22 10:40:52,218 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(sname)  
(ATT_Sensex_Report)  
(LNV_Sensex_Report)  
(NSE_Sensex_Report)  
(SBI_Sensex_Report)  
(SRS_Sensex_Report)  
(TAM_Sensex_Report)  
grunt>
```

4. Execute 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



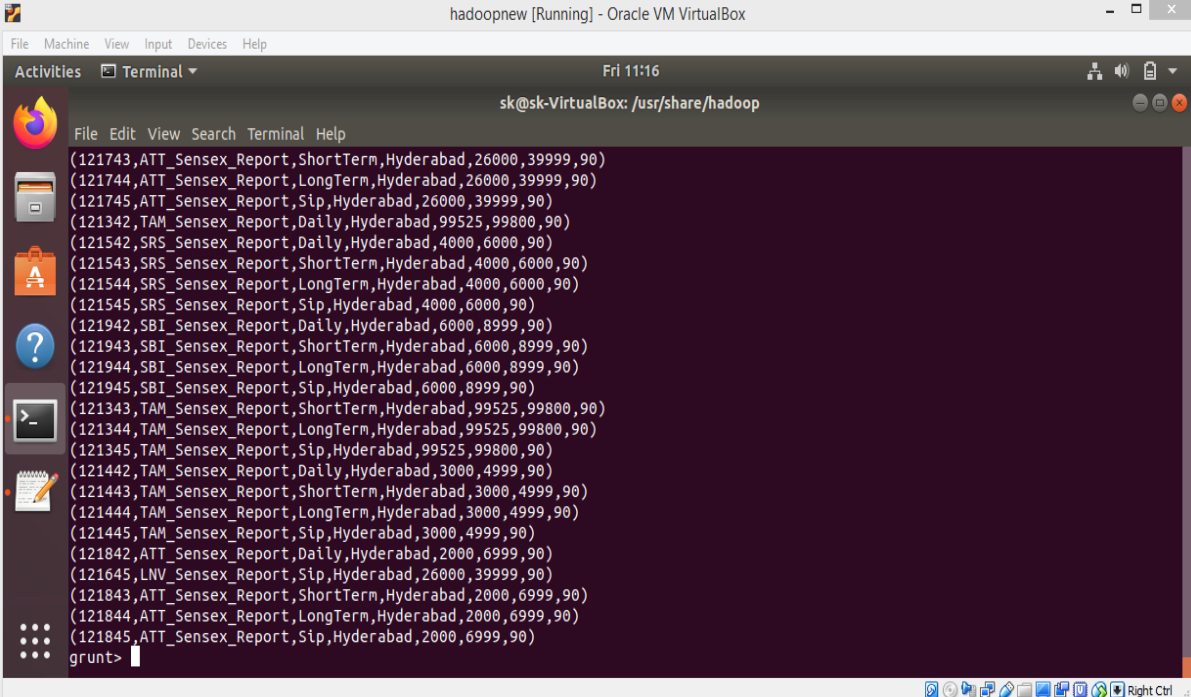
```
hadoopnew [Running] - Oracle VM VirtualBox  
File Machine View Input Devices Help  
Activities Terminal Fri 10:58  
sk@sk-VirtualBox: /usr/share/hadoop  
File Edit View Search Terminal Help  
(449914,SBI_Sensex_Report,Daily,Delhi,6000,8999,20)  
(449915,SBI_Sensex_Report,ShortTerm,Delhi,6000,8999,20)  
(449916,SBI_Sensex_Report,LongTerm,Delhi,6000,8999,20)  
(449917,SBI_Sensex_Report,Sip,Delhi,6000,8999,20)  
(332815,ATT_Sensex_Report,ShortTerm,Delhi,2000,6999,20)  
(332816,ATT_Sensex_Report,LongTerm,Delhi,2000,6999,20)  
(332817,ATT_Sensex_Report,Sip,Delhi,2000,6999,20)  
(449515,SRS_Sensex_Report,ShortTerm,Delhi,4000,6000,20)  
(449516,SRS_Sensex_Report,LongTerm,Delhi,4000,6000,20)  
(449517,SRS_Sensex_Report,Sip,Delhi,4000,6000,20)  
(121417,TAM_Sensex_Report,Sip,Delhi,3000,4999,20)  
(121917,SBI_Sensex_Report,Sip,Delhi,6000,8999,20)  
(449817,ATT_Sensex_Report,Sip,Delhi,2000,6999,20)  
(121517,SRS_Sensex_Report,Sip,Delhi,4000,6000,20)  
(121814,ATT_Sensex_Report,Daily,Delhi,2000,6999,20)  
(121815,ATT_Sensex_Report,ShortTerm,Delhi,2000,6999,20)  
(121816,ATT_Sensex_Report,LongTerm,Delhi,2000,6999,20)  
(121817,ATT_Sensex_Report,Sip,Delhi,2000,6999,20)  
(332514,SRS_Sensex_Report,Daily,Delhi,4000,6000,20)  
(332515,SRS_Sensex_Report,ShortTerm,Delhi,4000,6000,20)  
(332516,SRS_Sensex_Report,LongTerm,Delhi,4000,6000,20)  
(332517,SRS_Sensex_Report,Sip,Delhi,4000,6000,20)  
(121414,TAM_Sensex_Report,Daily,Delhi,3000,4999,20)  
(121415,TAM_Sensex_Report,ShortTerm,Delhi,3000,4999,20)  
(121416,TAM_Sensex_Report,LongTerm,Delhi,3000,4999,20)  
(121417,TAM_Sensex_Report,Sip,Delhi,3000,4999,20)  
grunt>
```


5. Execute 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 time(s).  
grunt> dump q5;
```

output



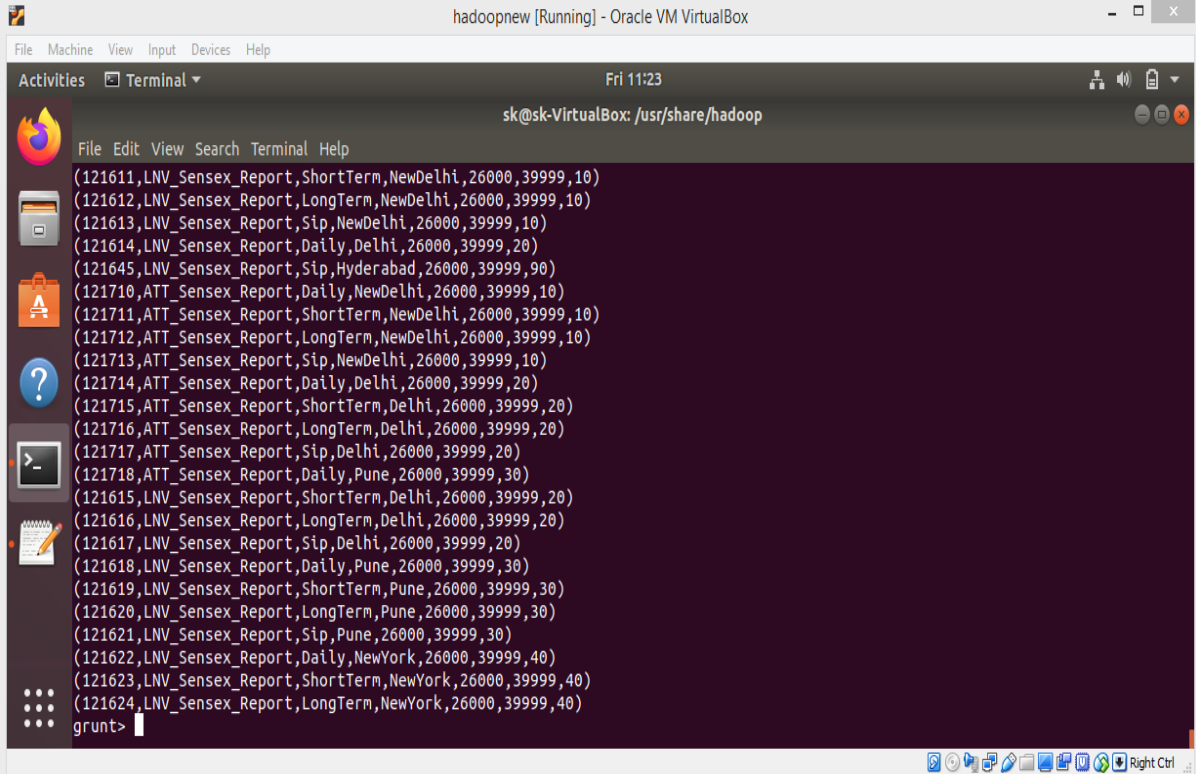
```
hadoopnew [Running] - Oracle VM VirtualBox  
File Machine View Input Devices Help  
Activities Terminal Fri 11:16  
sk@sk-VirtualBox: /usr/share/hadoop  
File Edit View Search Terminal Help  
(121743,ATT_Sensex_Report,ShortTerm,Hyderabad,26000,39999,90)  
(121744,ATT_Sensex_Report,LongTerm,Hyderabad,26000,39999,90)  
(121745,ATT_Sensex_Report,Sip,Hyderabad,26000,39999,90)  
(121342,TAM_Sensex_Report,Daily,Hyderabad,99525,99800,90)  
(121542,SRS_Sensex_Report,Daily,Hyderabad,4000,6000,90)  
(121543,SRS_Sensex_Report,ShortTerm,Hyderabad,4000,6000,90)  
(121544,SRS_Sensex_Report,LongTerm,Hyderabad,4000,6000,90)  
(121545,SRS_Sensex_Report,Sip,Hyderabad,4000,6000,90)  
(121942,SBI_Sensex_Report,Daily,Hyderabad,6000,8999,90)  
(121943,SBI_Sensex_Report,ShortTerm,Hyderabad,6000,8999,90)  
(121944,SBI_Sensex_Report,LongTerm,Hyderabad,6000,8999,90)  
(121945,SBI_Sensex_Report,Sip,Hyderabad,6000,8999,90)  
(121343,TAM_Sensex_Report,ShortTerm,Hyderabad,99525,99800,90)  
(121344,TAM_Sensex_Report,LongTerm,Hyderabad,99525,99800,90)  
(121345,TAM_Sensex_Report,Sip,Hyderabad,99525,99800,90)  
(121442,TAM_Sensex_Report,Daily,Hyderabad,3000,4999,90)  
(121443,TAM_Sensex_Report,ShortTerm,Hyderabad,3000,4999,90)  
(121444,TAM_Sensex_Report,LongTerm,Hyderabad,3000,4999,90)  
(121445,TAM_Sensex_Report,Sip,Hyderabad,3000,4999,90)  
(121842,ATT_Sensex_Report,Daily,Hyderabad,2000,6999,90)  
(121645,LNV_Sensex_Report,Sip,Hyderabad,26000,39999,90)  
(121843,ATT_Sensex_Report,ShortTerm,Hyderabad,2000,6999,90)  
(121844,ATT_Sensex_Report,LongTerm,Hyderabad,2000,6999,90)  
(121845,ATT_Sensex_Report,Sip,Hyderabad,2000,6999,90)  
grunt>
```

6. Execute 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 time(s).  
grunt> dump q6;
```

Output



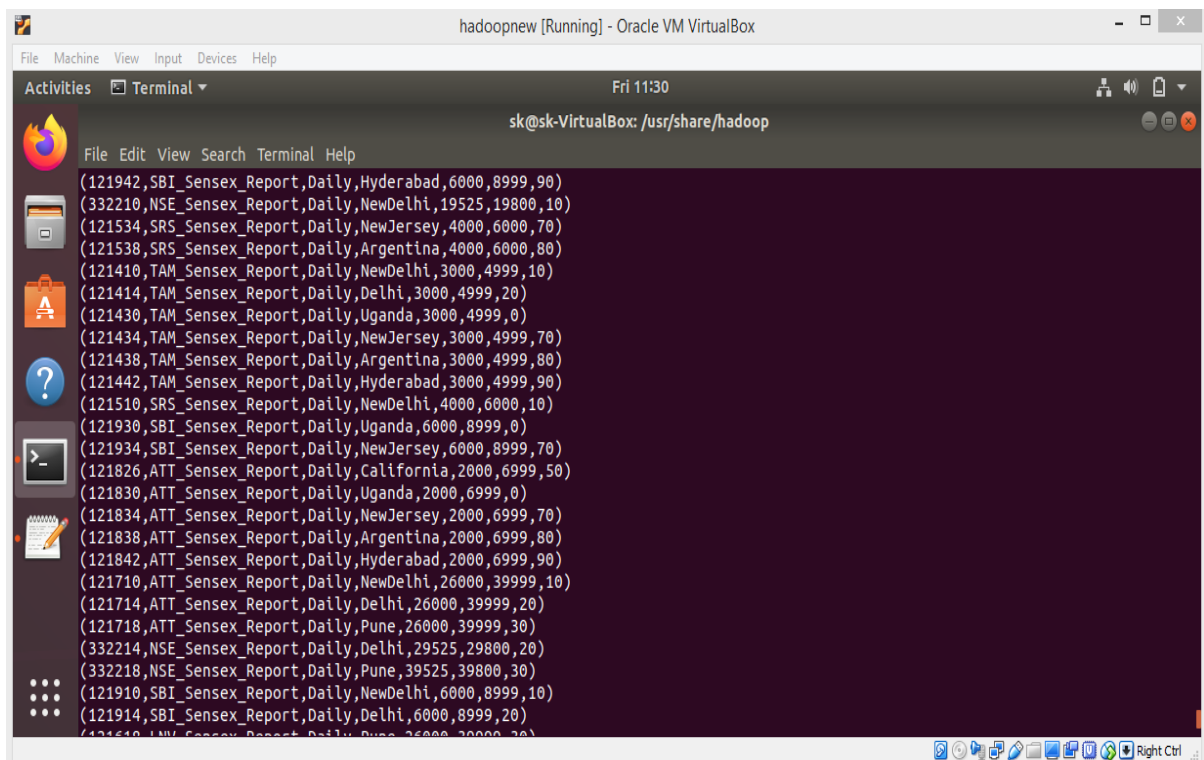
```
hadoopnew [Running] - Oracle VM VirtualBox  
Fri 11:23  
sk@sk-VirtualBox: /usr/share/hadoop  
File Edit View Search Terminal Help  
(121611,LNV_Sensex_Report,ShortTerm,NewDelhi,26000,39999,10)  
(121612,LNV_Sensex_Report,LongTerm,NewDelhi,26000,39999,10)  
(121613,LNV_Sensex_Report,Sip,NewDelhi,26000,39999,10)  
(121614,LNV_Sensex_Report,Daily,Delhi,26000,39999,20)  
(121645,LNV_Sensex_Report,Sip,Hyderabad,26000,39999,90)  
(121710,ATT_Sensex_Report,Daily,NewDelhi,26000,39999,10)  
(121711,ATT_Sensex_Report,ShortTerm,NewDelhi,26000,39999,10)  
(121712,ATT_Sensex_Report,LongTerm,NewDelhi,26000,39999,10)  
(121713,ATT_Sensex_Report,Sip,NewDelhi,26000,39999,10)  
(121714,ATT_Sensex_Report,Daily,Delhi,26000,39999,20)  
(121715,ATT_Sensex_Report,ShortTerm,Delhi,26000,39999,20)  
(121716,ATT_Sensex_Report,LongTerm,Delhi,26000,39999,20)  
(121717,ATT_Sensex_Report,Sip,Delhi,26000,39999,20)  
(121718,ATT_Sensex_Report,Daily,Pune,26000,39999,30)  
(121615,LNV_Sensex_Report,ShortTerm,Delhi,26000,39999,20)  
(121616,LNV_Sensex_Report,LongTerm,Delhi,26000,39999,20)  
(121617,LNV_Sensex_Report,Sip,Delhi,26000,39999,20)  
(121618,LNV_Sensex_Report,Daily,Pune,26000,39999,30)  
(121619,LNV_Sensex_Report,ShortTerm,Pune,26000,39999,30)  
(121620,LNV_Sensex_Report,LongTerm,Pune,26000,39999,30)  
(121621,LNV_Sensex_Report,Sip,Pune,26000,39999,30)  
(121622,LNV_Sensex_Report,Daily,NewYork,26000,39999,40)  
(121623,LNV_Sensex_Report,ShortTerm,NewYork,26000,39999,40)  
(121624,LNV_Sensex_Report,LongTerm,NewYork,26000,39999,40)  
grunt>
```

7. Execute 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 time(s).
grunt> dump q7;
```

Output



```
hadoopnew [Running] - Oracle VM VirtualBox
Fri 11:30
sk@sk-VirtualBox: /usr/share/hadoop

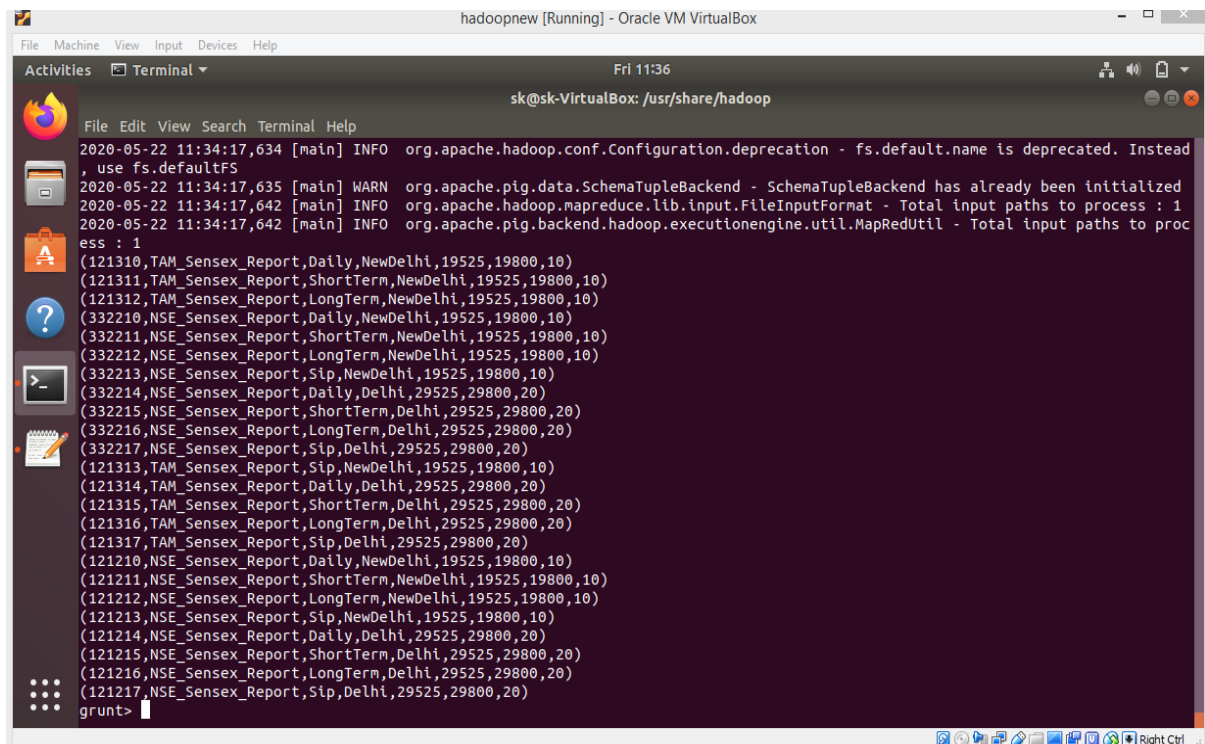
(121942,SBI_Sensex_Report,Daily,Hyderabad,6000,8999,90)
(332210,NSE_Sensex_Report,Daily,NewDelhi,19525,19800,10)
(121534,SRS_Sensex_Report,Daily,NewJersey,4000,6000,70)
(121538,SRS_Sensex_Report,Daily,Argentina,4000,6000,80)
(121410,TAM_Sensex_Report,Daily,NewDelhi,3000,4999,10)
(121414,TAM_Sensex_Report,Daily,Delhi,3000,4999,20)
(121430,TAM_Sensex_Report,Daily,Uganda,3000,4999,0)
(121434,TAM_Sensex_Report,Daily,NewJersey,3000,4999,70)
(121438,TAM_Sensex_Report,Daily,Argentina,3000,4999,80)
(121442,TAM_Sensex_Report,Daily,Hyderabad,3000,4999,90)
(121510,SRS_Sensex_Report,Daily,NewDelhi,4000,6000,10)
(121930,SBI_Sensex_Report,Daily,Uganda,6000,8999,0)
(121934,SBI_Sensex_Report,Daily,NewJersey,6000,8999,70)
(121826,ATT_Sensex_Report,Daily,California,2000,6999,50)
(121830,ATT_Sensex_Report,Daily,Uganda,2000,6999,0)
(121834,ATT_Sensex_Report,Daily,NewJersey,2000,6999,70)
(121838,ATT_Sensex_Report,Daily,Argentina,2000,6999,80)
(121842,ATT_Sensex_Report,Daily,Hyderabad,2000,6999,90)
(121710,ATT_Sensex_Report,Daily,NewDelhi,26000,39999,10)
(121714,ATT_Sensex_Report,Daily,Delhi,26000,39999,20)
(121718,ATT_Sensex_Report,Daily,Pune,26000,39999,30)
(332214,NSE_Sensex_Report,Daily,Delhi,29525,29800,20)
(332218,NSE_Sensex_Report,Daily,Pune,39525,39800,30)
(121910,SBI_Sensex_Report,Daily,NewDelhi,6000,8999,10)
(121914,SBI_Sensex_Report,Daily,Delhi,6000,8999,20)
(121510,SRS_Sensex_Report,Daily,NewDelhi,4000,6000,10)
```

8. Execute 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 time(s).  
grunt> dump q8;
```

Output



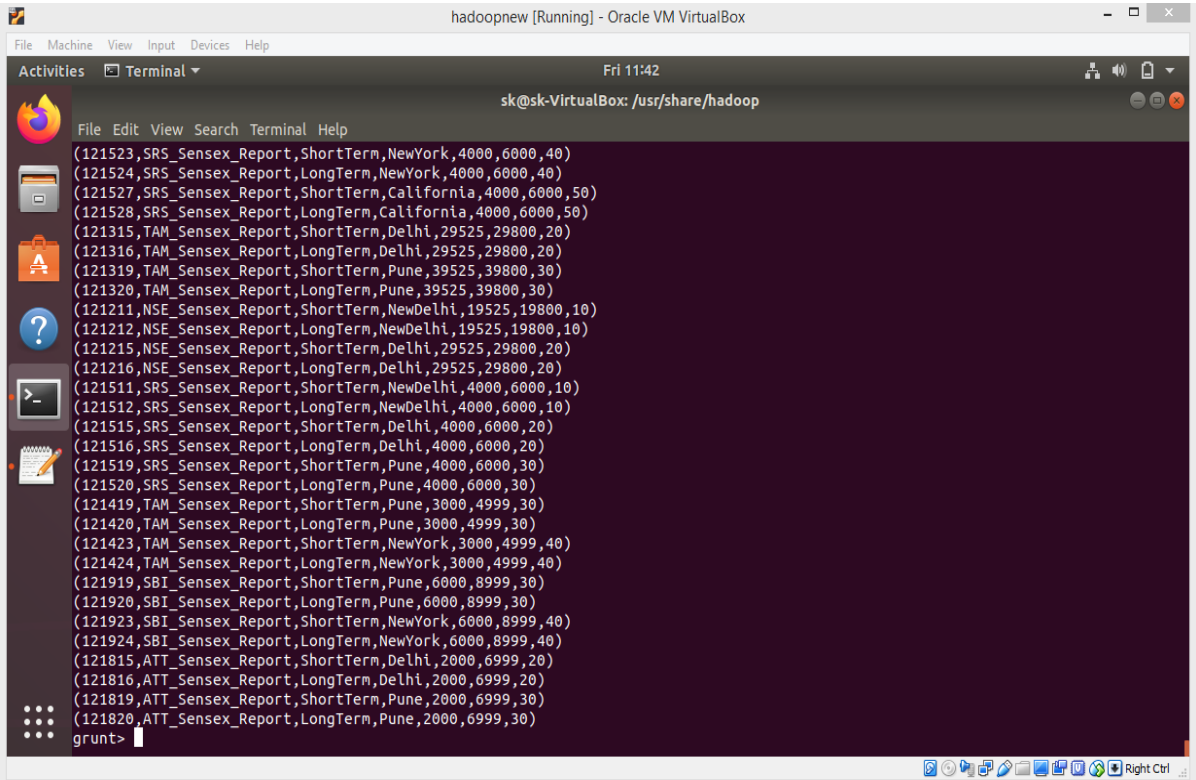
```
hadoopnew [Running] - Oracle VM VirtualBox  
Fri 11:36  
sk@sk-VirtualBox: /usr/share/hadoop  
File Edit View Search Terminal Help  
2020-05-22 11:34:17,634 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead  
, use fs.defaultFS  
2020-05-22 11:34:17,635 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2020-05-22 11:34:17,642 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2020-05-22 11:34:17,642 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to proc  
ess : 1  
(121310,TAM_Sensex_Report,Daily,NewDelhi,19525,19800,10)  
(121311,TAM_Sensex_Report,ShortTerm,NewDelhi,19525,19800,10)  
(121312,TAM_Sensex_Report,LongTerm,NewDelhi,19525,19800,10)  
(332210,NSE_Sensex_Report,Daily,NewDelhi,19525,19800,10)  
(332211,NSE_Sensex_Report,ShortTerm,NewDelhi,19525,19800,10)  
(332212,NSE_Sensex_Report,LongTerm,NewDelhi,19525,19800,10)  
(332213,NSE_Sensex_Report,Stp,NewDelhi,19525,19800,10)  
(332214,NSE_Sensex_Report,Daily,Delhi,29525,29800,20)  
(332215,NSE_Sensex_Report,ShortTerm,Delhi,29525,29800,20)  
(332216,NSE_Sensex_Report,LongTerm,Delhi,29525,29800,20)  
(332217,NSE_Sensex_Report,Stp,Delhi,29525,29800,20)  
(121313,TAM_Sensex_Report,Stp,NewDelhi,19525,19800,10)  
(121314,TAM_Sensex_Report,Daily,Delhi,29525,29800,20)  
(121315,TAM_Sensex_Report,ShortTerm,Delhi,29525,29800,20)  
(121316,TAM_Sensex_Report,LongTerm,Delhi,29525,29800,20)  
(121317,TAM_Sensex_Report,Stp,Delhi,29525,29800,20)  
(121210,NSE_Sensex_Report,Daily,NewDelhi,19525,19800,10)  
(121211,NSE_Sensex_Report,ShortTerm,NewDelhi,19525,19800,10)  
(121212,NSE_Sensex_Report,LongTerm,NewDelhi,19525,19800,10)  
(121213,NSE_Sensex_Report,Stp,NewDelhi,19525,19800,10)  
(121214,NSE_Sensex_Report,Daily,Delhi,29525,29800,20)  
(121215,NSE_Sensex_Report,ShortTerm,Delhi,29525,29800,20)  
(121216,NSE_Sensex_Report,LongTerm,Delhi,29525,29800,20)  
(121217,NSE_Sensex_Report,Stp,Delhi,29525,29800,20)  
grunt>
```

9. Execute 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 time(s).  
grunt> dump q9;
```

Output



```
File Edit View Search Terminal Help  
Fri 11:42  
sk@sk-VirtualBox: /usr/share/hadoop  
(121523,SRS_Sensex_Report,ShortTerm,NewYork,4000,6000,40)  
(121524,SRS_Sensex_Report,LongTerm,NewYork,4000,6000,40)  
(121527,SRS_Sensex_Report,ShortTerm,California,4000,6000,50)  
(121528,SRS_Sensex_Report,LongTerm,California,4000,6000,50)  
(121315,TAM_Sensex_Report,ShortTerm,Delhi,29525,29800,20)  
(121316,TAM_Sensex_Report,LongTerm,Delhi,29525,29800,20)  
(121319,TAM_Sensex_Report,ShortTerm,Pune,39525,39800,30)  
(121320,TAM_Sensex_Report,LongTerm,Pune,39525,39800,30)  
(121211,NSE_Sensex_Report,ShortTerm,NewDelhi,19525,19800,10)  
(121212,NSE_Sensex_Report,LongTerm,NewDelhi,19525,19800,10)  
(121215,NSE_Sensex_Report,ShortTerm,Delhi,29525,29800,20)  
(121216,NSE_Sensex_Report,LongTerm,Delhi,29525,29800,20)  
(121511,SRS_Sensex_Report,ShortTerm,NewDelhi,4000,6000,10)  
(121512,SRS_Sensex_Report,LongTerm,NewDelhi,4000,6000,10)  
(121515,SRS_Sensex_Report,ShortTerm,Delhi,4000,6000,20)  
(121516,SRS_Sensex_Report,LongTerm,Delhi,4000,6000,20)  
(121519,SRS_Sensex_Report,ShortTerm,Pune,4000,6000,30)  
(121520,SRS_Sensex_Report,LongTerm,Pune,4000,6000,30)  
(121419,TAM_Sensex_Report,ShortTerm,Pune,3000,4999,30)  
(121420,TAM_Sensex_Report,LongTerm,Pune,3000,4999,30)  
(121423,TAM_Sensex_Report,ShortTerm,NewYork,3000,4999,40)  
(121424,TAM_Sensex_Report,LongTerm,NewYork,3000,4999,40)  
(121919,SBI_Sensex_Report,ShortTerm,Pune,6000,8999,30)  
(121920,SBI_Sensex_Report,LongTerm,Pune,6000,8999,30)  
(121923,SBI_Sensex_Report,ShortTerm,NewYork,6000,8999,40)  
(121924,SBI_Sensex_Report,LongTerm,NewYork,6000,8999,40)  
(121815,ATT_Sensex_Report,ShortTerm,Delhi,2000,6999,20)  
(121816,ATT_Sensex_Report,LongTerm,Delhi,2000,6999,20)  
(121819,ATT_Sensex_Report,ShortTerm,Pune,2000,6999,30)  
(121820,ATT_Sensex_Report,LongTerm,Pune,2000,6999,30)  
grunt>
```

```
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;
```

```

Fri 12:26
sk@sk-VirtualBox: /usr/share/hadoop

File Edit View Search Terminal Help

Total records written : 1
Total bytes written : 43028430
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1755836396_0004

2020-05-22 12:26:38,560 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, se
ssionId= - already initialized
2020-05-22 12:26:38,562 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, se
ssionId= - already initialized
2020-05-22 12:26:38,563 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, se
ssionId= - already initialized
2020-05-22 12:26:38,568 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIE
LD DISCARDED TYPE_CONVERSION FAILED 4 time(s).
2020-05-22 12:26:38,568 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2020-05-22 12:26:38,569 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
fs
2020-05-22 12:26:38,570 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2020-05-22 12:26:38,600 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2020-05-22 12:26:38,600 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(((Daily),(Sip),(LongTerm),(ShortTerm),(Daily),(Sip),(LongTerm),(ShortTerm),(Daily),(Sip),(LongTerm),(ShortTerm),(Daily),(trading),(Daily),(S
ip),(LongTerm),(ShortTerm),(Daily),(Sip),(LongTerm),(ShortTerm),(Daily),(Sip),(LongTerm),(ShortTerm),(Daily),(Sip),(LongTerm),(ShortTerm),(D
aily),(Sip),(LongTerm),(ShortTerm)),{(99800),(89800),(89800),(89800),(39800),(29800),(29800),(29800),(29800),(19800),(19800),(19800),(19800),
(,),(89800),(79800),(79800),(79800),(79800),(69800),(69800),(69800),(69800),(59800),(59800),(59800),(59800),(49800),(49800),(49800),(49800),(3
9800),(39800)),56163.63636363636)
grunt>

```