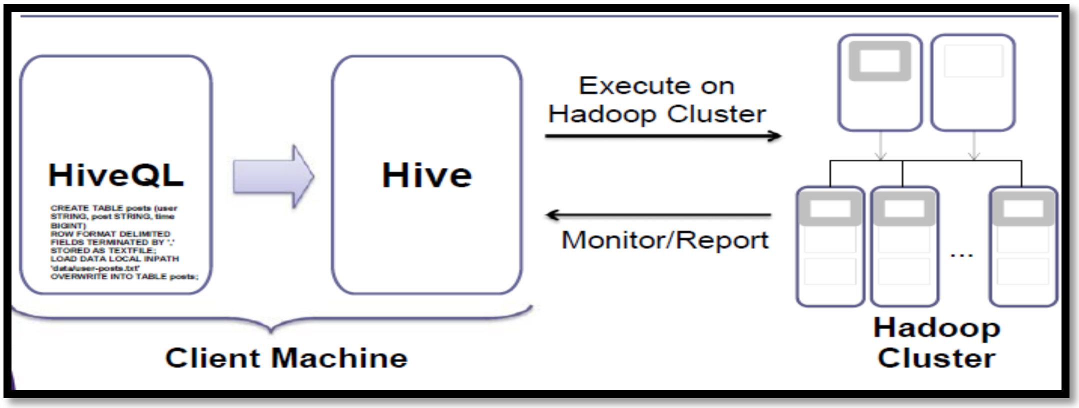
Apache Hive Exercise
(with Enterprise_Retail_Dataset)

BACKGROUND ABOUT THE HIVE

What is Apache Hive?

Apache Hive is a data warehouse infrastructure that facilitates querying and managing large data sets which resides in distributed storage system. It is built on top of Hadoop and developed by Facebook. **Hive** provides a way to query the data using a SQL-like query language called **HiveQL (Hive query Language).**

Internally, a compiler translates **HiveQL** statements into **MapReduce** jobs, which are then submitted to **Hadoop framework** for execution as shown below.



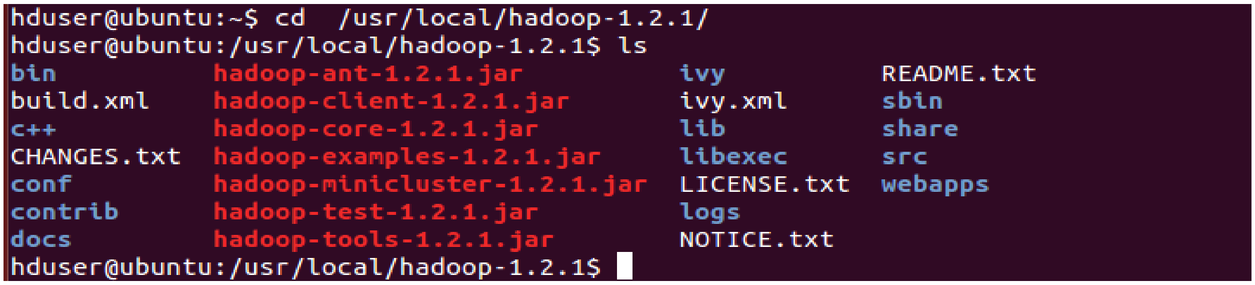
```
CREATE TABLE posts (user
STRING, post STRING, time
BIGINT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
LOAD DATA LOCAL INPATH
'data/user-posts.txt'
OVERWRITE INTO TABLE posts;
```

PREREQUISITES

1. We have already installed Hadoop and Hive in our ubuntu. Now its the time to understand the directory structure of both.

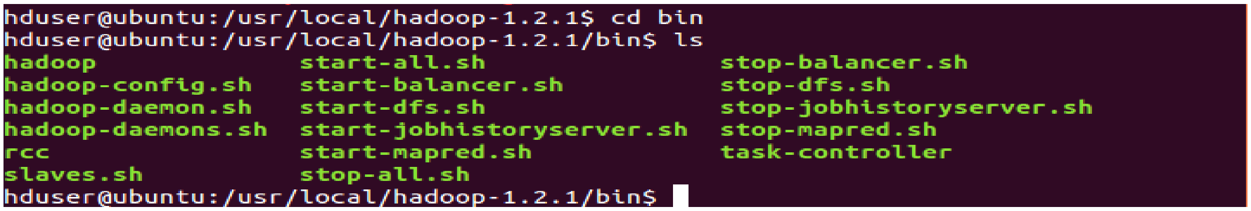
a. Hadoop Directory Structure:

```
$ cd  /usr/local/hadoop-1.2.1/

$ ls
```



```
hduser@ubuntu:~$ cd  /usr/local/hadoop-1.2.1/
hduser@ubuntu:/usr/local/hadoop-1.2.1$ ls
bin            hadoop-ant-1.2.1.jar          ivy          README.txt
build.xml      hadoop-client-1.2.1.jar       ivy.xml      sbin
c++            hadoop-core-1.2.1.jar         lib          share
CHANGES.txt    hadoop-examples-1.2.1.jar     libexec      src
conf           hadoop-minicluster-1.2.1.jar  LICENSE.txt  webapps
contrib        hadoop-test-1.2.1.jar         logs
docs           hadoop-tools-1.2.1.jar        NOTICE.txt
hduser@ubuntu:/usr/local/hadoop-1.2.1$
```

```
$ cd bin

$ ls
```

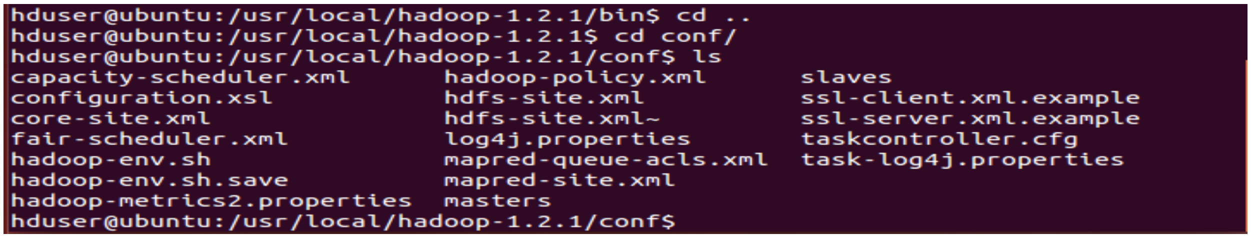


```
hduser@ubuntu:/usr/local/hadoop-1.2.1$ cd bin
hduser@ubuntu:/usr/local/hadoop-1.2.1/bin$ ls
hadoop              start-all.sh                stop-balancer.sh
hadoop-config.sh    start-balancer.sh           stop-dfs.sh
hadoop-daemon.sh    start-dfs.sh                stop-jobhistoryserver.sh
hadoop-daemons.sh   start-jobhistoryserver.sh   stop-mapred.sh
rcc                 start-mapred.sh             task-controller
slaves.sh           stop-all.sh
hduser@ubuntu:/usr/local/hadoop-1.2.1/bin$
```

[**bin**- directory which contains binary executable files has mainly:

start-all.sh- To start Hadoop daemons
stop-all.sh- To stop Hadoop daemons
hadoop- To work with Hadoop Distributed operations]

```
$ cd ..

$ cd conf/

$ ls
```

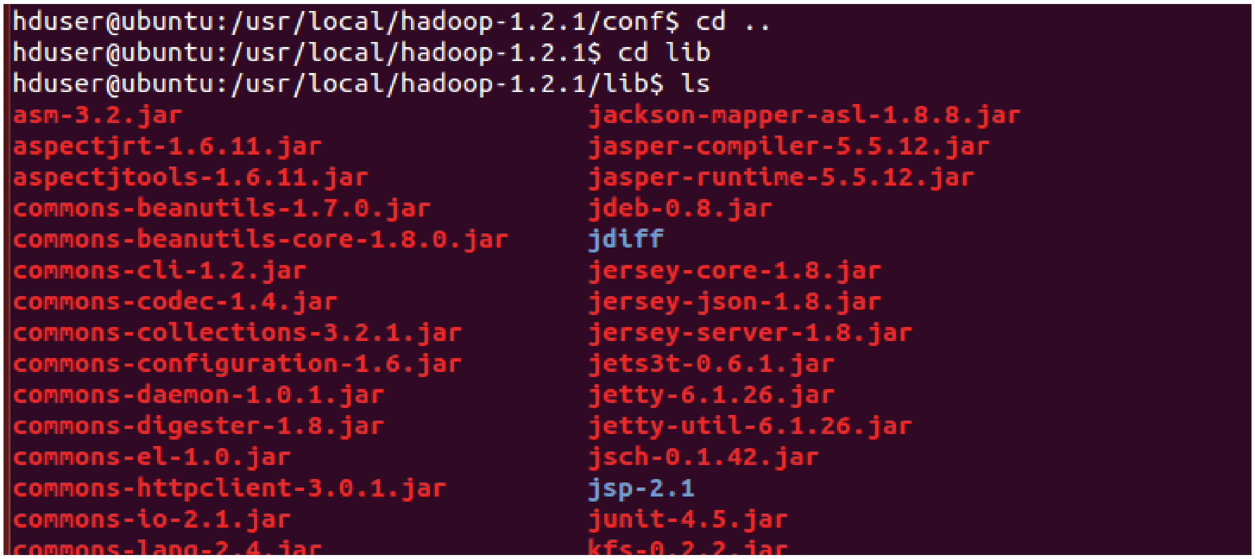


```
hduser@ubuntu:/usr/local/hadoop-1.2.1/bin$ cd ..
hduser@ubuntu:/usr/local/hadoop-1.2.1$ cd conf/
hduser@ubuntu:/usr/local/hadoop-1.2.1/conf$ ls
capacity-scheduler.xml        hadoop-policy.xml        slaves
configuration.xsl             hdfs-site.xml            ssl-client.xml.example
core-site.xml                 hdfs-site.xml~           ssl-server.xml.example
fair-scheduler.xml            log4j.properties         taskcontroller.cfg
hadoop-env.sh                 mapred-queue-acls.xml    task-log4j.properties
hadoop-env.sh.save            mapred-site.xml
hadoop-metrics2.properties    masters
hduser@ubuntu:/usr/local/hadoop-1.2.1/conf$
```

[**conf**- directory which contains all configuration files which contains mainly:

hadoop-env.sh- To configure java integration with Hadoop.
core-site.xml- To configure namenode and datanode
mapred-site.xml- To configure jobtracker and tasktracker
hdfs-site.xml- To configure number replication]

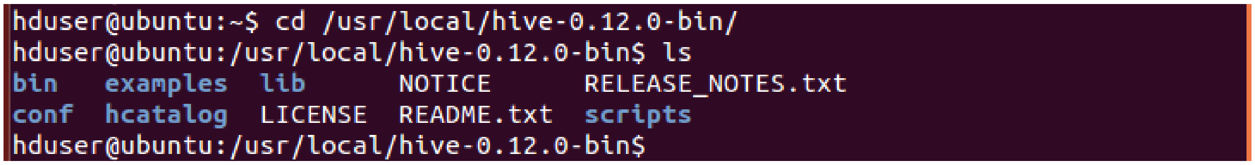
```
$ cd ..

$ cd lib

$ ls
```



```
hduser@ubuntu:/usr/local/hadoop-1.2.1/conf$ cd ..
hduser@ubuntu:/usr/local/hadoop-1.2.1$ cd lib
hduser@ubuntu:/usr/local/hadoop-1.2.1/lib$ ls
asm-3.2.jar                          jackson-mapper-asl-1.8.8.jar
aspectjrt-1.6.11.jar                 jasper-compiler-5.5.12.jar
aspectjtools-1.6.11.jar              jasper-runtime-5.5.12.jar
commons-beanutils-1.7.0.jar          jdeb-0.8.jar
commons-beanutils-core-1.8.0.jar     jdiff
commons-cli-1.2.jar                  jersey-core-1.8.jar
commons-codec-1.4.jar                jersey-json-1.8.jar
commons-collections-3.2.1.jar        jersey-server-1.8.jar
commons-configuration-1.6.jar        jets3t-0.6.1.jar
commons-daemon-1.0.1.jar             jetty-6.1.26.jar
commons-digester-1.8.jar             jetty-util-6.1.26.jar
commons-el-1.0.jar                   jsch-0.1.42.jar
commons-httpclient-3.0.1.jar         jsp-2.1
commons-io-2.1.jar                   junit-4.5.jar
commons-lang-2.4.jar                 kfs-0.2.2.jar
```

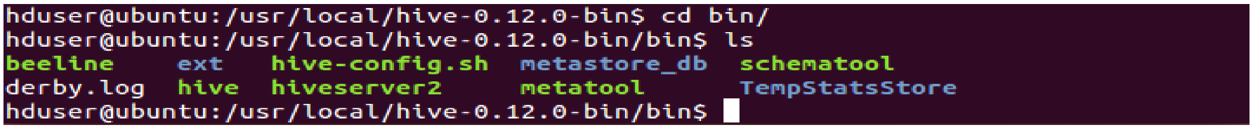
b. Hive Directory Structure:

```
$ cd /usr/local/hive-0.12.0-bin/

$ ls
```



```
hduser@ubuntu:~$ cd /usr/local/hive-0.12.0-bin/
hduser@ubuntu:/usr/local/hive-0.12.0-bin$ ls
bin   examples  lib      NOTICE      RELEASE_NOTES.txt
conf  hcatalog  LICENSE  README.txt  scripts
hduser@ubuntu:/usr/local/hive-0.12.0-bin$
```

```
$ cd bin

$ ls
```

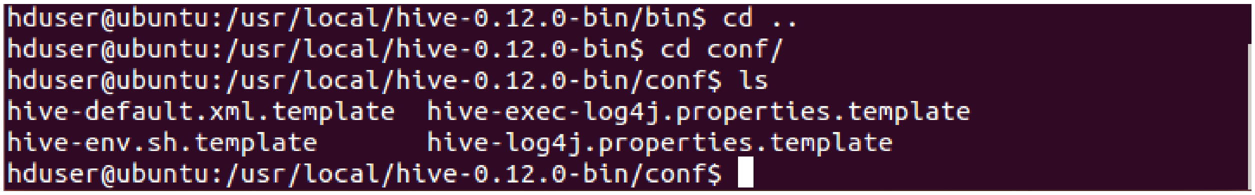


```
hduser@ubuntu:/usr/local/hive-0.12.0-bin$ cd bin/
hduser@ubuntu:/usr/local/hive-0.12.0-bin/bin$ ls
beeline    ext   hive-config.sh  metastore_db  schematool
derby.log  hive  hiveserver2     metatool      TempStatsStore
hduser@ubuntu:/usr/local/hive-0.12.0-bin/bin$
```

[**bin**- directory which contains binary executable files which contains mainly:

hive- To start Hive command line interface

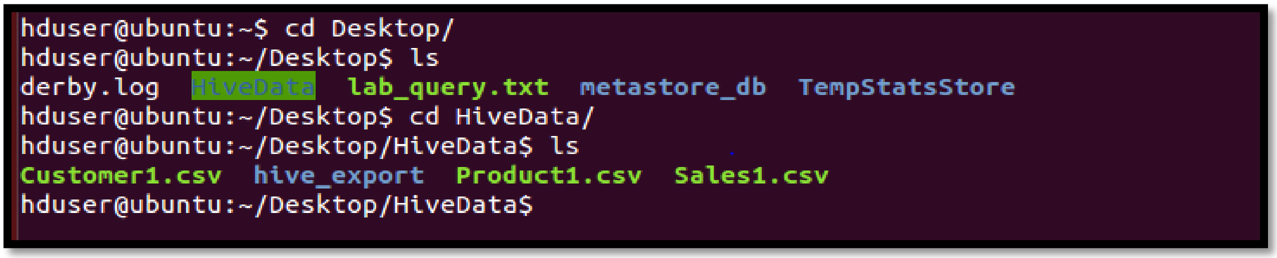
```
$ cd ..

$ cd conf/

$ ls
```



```
hduser@ubuntu:/usr/local/hive-0.12.0-bin/bin$ cd ..
hduser@ubuntu:/usr/local/hive-0.12.0-bin$ cd conf/
hduser@ubuntu:/usr/local/hive-0.12.0-bin/conf$ ls
hive-default.xml.template  hive-exec-log4j.properties.template
hive-env.sh.template       hive-log4j.properties.template
hduser@ubuntu:/usr/local/hive-0.12.0-bin/conf$
```

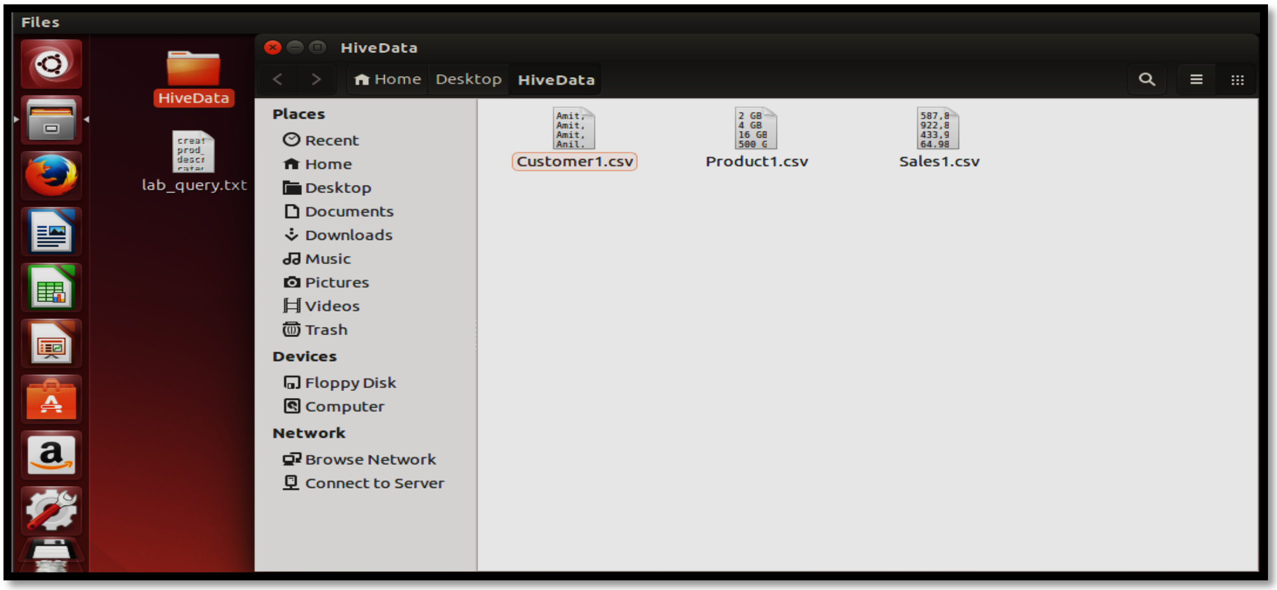
2. The history of retail data should be placed on our ubuntu system Desktop.

```
$ cd /home/hduser/Desktop/HiveData/
```



```
hduser@ubuntu:~$ cd Desktop/
hduser@ubuntu:~/Desktop$ ls
derby.log  HiveData  lab_query.txt  metastore_db  TempStatsStore
hduser@ubuntu:~/Desktop$ cd HiveData/
hduser@ubuntu:~/Desktop/HiveData$ ls
Customer1.csv  hive_export  Product1.csv  Sales1.csv
hduser@ubuntu:~/Desktop/HiveData$
```

HiveData folder contains the data in 3 separate files. Namely : Customer1.csv , Product1.csv , Sales1.csv as shown below

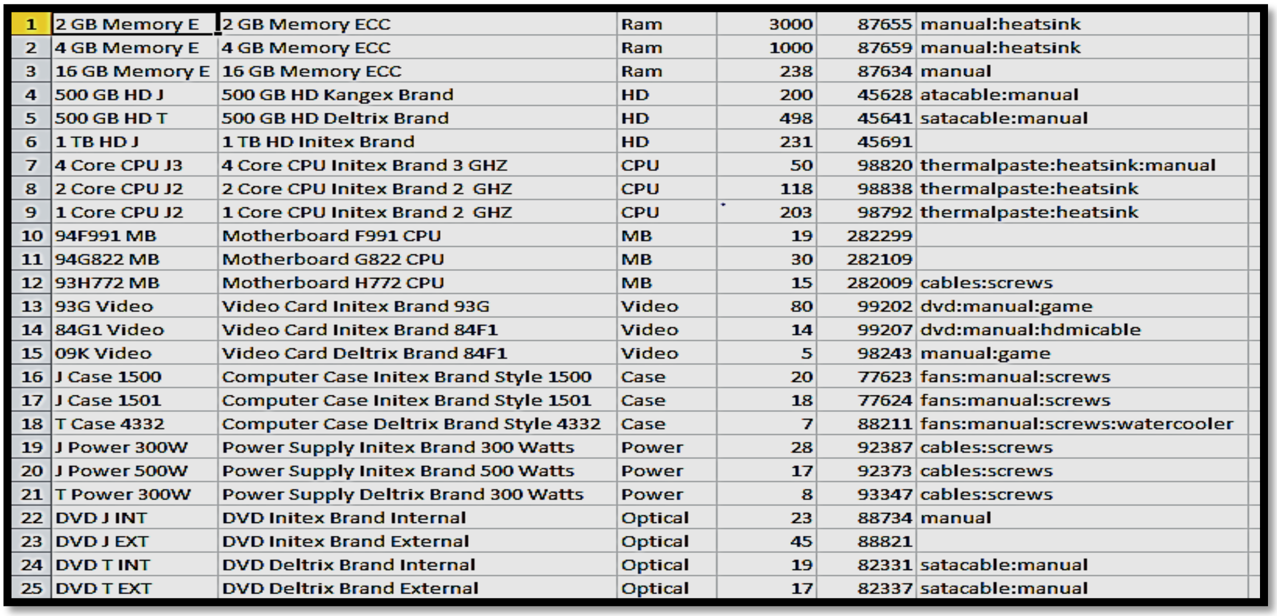


Now we will goto the terminal. **Login to hduser account.**

[Short cut to open ubuntu terminal:**CTRL+ALT+t**]

To open the products1.csv. Double click Products1.csv on Desktop

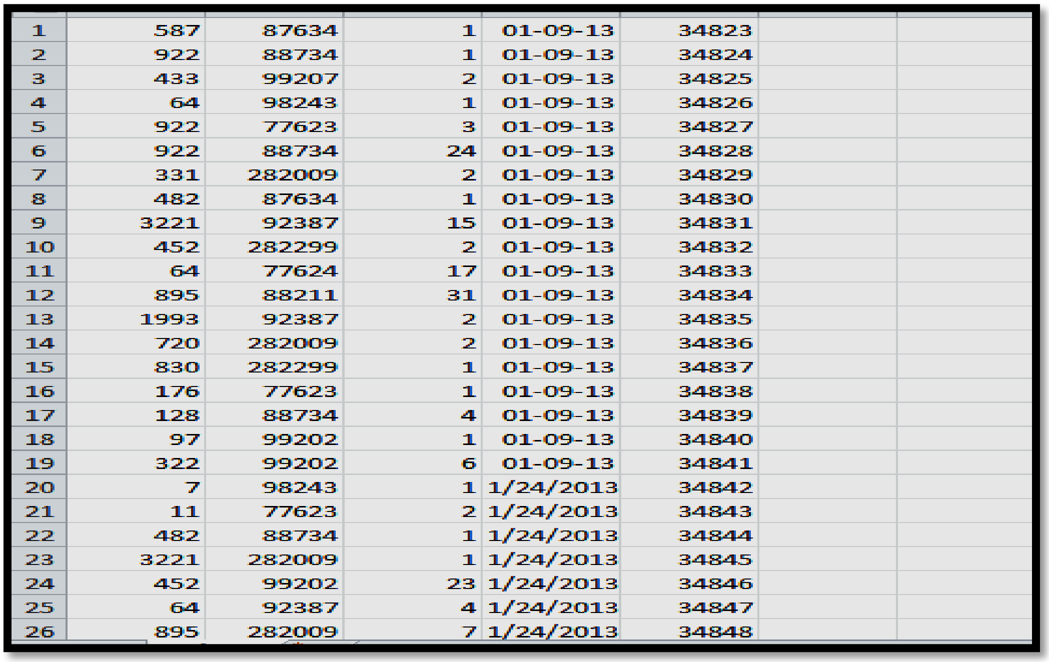
SlNo | prod_name | description | category| qty_on_hand |*prod_num* | packaged_with



1	2 GB Memory E	2 GB Memory ECC	Ram	3000	87655	manual:heatsink
2	4 GB Memory E	4 GB Memory ECC	Ram	1000	87659	manual:heatsink
3	16 GB Memory E	16 GB Memory ECC	Ram	238	87634	manual
4	500 GB HD J	500 GB HD Kangex Brand	HD	200	45628	atacable:manual
5	500 GB HD T	500 GB HD Deltrix Brand	HD	498	45641	satacable:manual
6	1 TB HD J	1 TB HD Initex Brand	HD	231	45691	
7	4 Core CPU J3	4 Core CPU Initex Brand 3 GHZ	CPU	50	98820	thermalpaste:heatsink:manual
8	2 Core CPU J2	2 Core CPU Initex Brand 2 GHZ	CPU	118	98838	thermalpaste:heatsink
9	1 Core CPU J2	1 Core CPU Initex Brand 2 GHZ	CPU	203	98792	thermalpaste:heatsink
10	94F991 MB	Motherboard F991 CPU	MB	19	282299	
11	94G822 MB	Motherboard G822 CPU	MB	30	282109	
12	93H772 MB	Motherboard H772 CPU	MB	15	282009	cables:screws
13	93G Video	Video Card Initex Brand 93G	Video	80	99202	dvd:manual:game
14	84G1 Video	Video Card Initex Brand 84F1	Video	14	99207	dvd:manual:hdmicable
15	09K Video	Video Card Deltrix Brand 84F1	Video	5	98243	manual:game
16	J Case 1500	Computer Case Initex Brand Style 1500	Case	20	77623	fans:manual:screws
17	J Case 1501	Computer Case Initex Brand Style 1501	Case	18	77624	fans:manual:screws
18	T Case 4332	Computer Case Deltrix Brand Style 4332	Case	7	88211	fans:manual:screws:watercooler
19	J Power 300W	Power Supply Initex Brand 300 Watts	Power	28	92387	cables:screws
20	J Power 500W	Power Supply Initex Brand 500 Watts	Power	17	92373	cables:screws
21	T Power 300W	Power Supply Deltrix Brand 300 Watts	Power	8	93347	cables:screws
22	DVD J INT	DVD Initex Brand Internal	Optical	23	88734	manual
23	DVD J EXT	DVD Initex Brand External	Optical	45	88821	
24	DVD T INT	DVD Deltrix Brand Internal	Optical	19	82331	satacable:manual
25	DVD T EXT	DVD Deltrix Brand External	Optical	17	82337	satacable:manual

To open the sales1.csv. Double click on Sales1.csv on Desktop

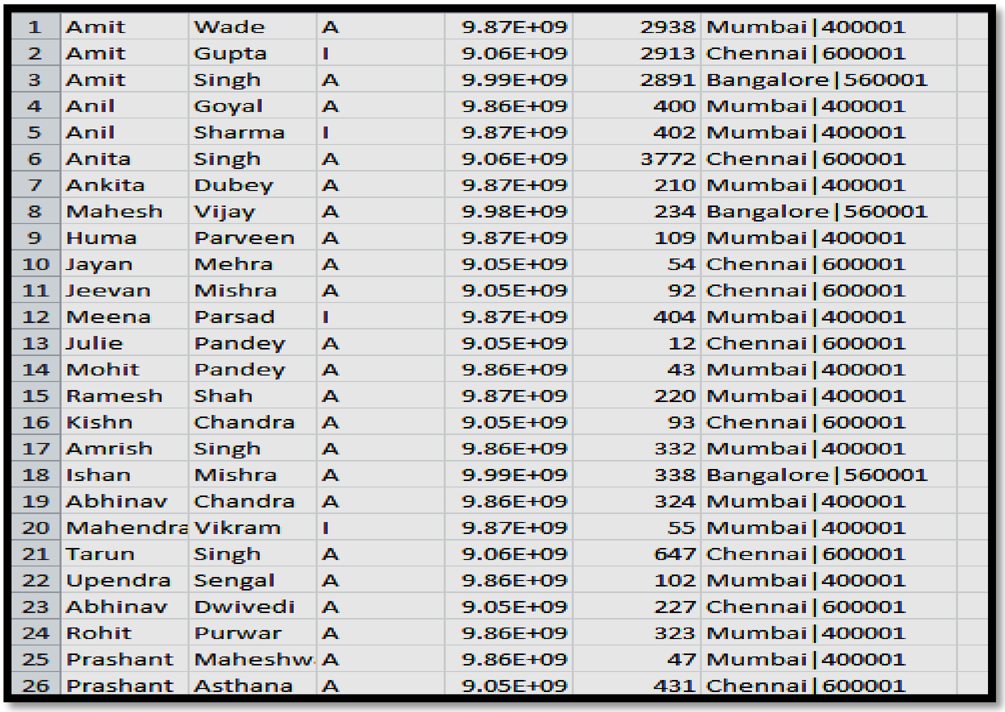
SlNo | *cust_id* | *prod_num* | qty | sales_date | sales_id



SlNo	*cust_id*	*prod_num*	qty	sales_date	sales_id
1	587	87634	1	01-09-13	34823
2	922	88734	1	01-09-13	34824
3	433	99207	2	01-09-13	34825
4	64	98243	1	01-09-13	34826
5	922	77623	3	01-09-13	34827
6	922	88734	24	01-09-13	34828
7	331	282009	2	01-09-13	34829
8	482	87634	1	01-09-13	34830
9	3221	92387	15	01-09-13	34831
10	452	282299	2	01-09-13	34832
11	64	77624	17	01-09-13	34833
12	895	88211	31	01-09-13	34834
13	1993	92387	2	01-09-13	34835
14	720	282009	2	01-09-13	34836
15	830	282299	1	01-09-13	34837
16	176	77623	1	01-09-13	34838
17	128	88734	4	01-09-13	34839
18	97	99202	1	01-09-13	34840
19	322	99202	6	01-09-13	34841
20	7	98243	1	1/24/2013	34842
21	11	77623	2	1/24/2013	34843
22	482	88734	1	1/24/2013	34844
23	3221	282009	1	1/24/2013	34845
24	452	99202	23	1/24/2013	34846
25	64	92387	4	1/24/2013	34847
26	895	282009	7	1/24/2013	34848

To open the products1.csv. Double click on Customer1.csv on Desktop

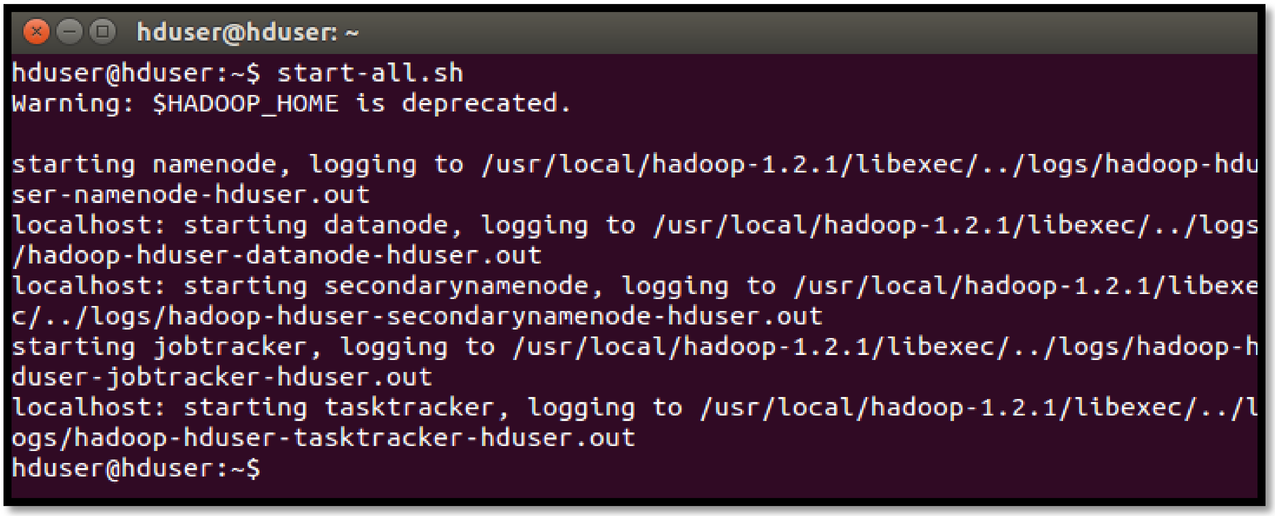
SlNo | fname | lname | status | telno | *cust_id* | city_zip



SlNo	fname	lname	status	telno	*cust_id*	city_zip
1	Amit	Wade	A	9.87E+09	2938	Mumbai|400001
2	Amit	Gupta	I	9.06E+09	2913	Chennai|600001
3	Amit	Singh	A	9.99E+09	2891	Bangalore|560001
4	Anil	Goyal	A	9.86E+09	400	Mumbai|400001
5	Anil	Sharma	I	9.87E+09	402	Mumbai|400001
6	Anita	Singh	A	9.06E+09	3772	Chennai|600001
7	Ankita	Dubey	A	9.87E+09	210	Mumbai|400001
8	Mahesh	Vijay	A	9.98E+09	234	Bangalore|560001
9	Huma	Parveen	A	9.87E+09	109	Mumbai|400001
10	Jayan	Mehra	A	9.05E+09	54	Chennai|600001
11	Jeevan	Mishra	A	9.05E+09	92	Chennai|600001
12	Meena	Parsad	I	9.87E+09	404	Mumbai|400001
13	Julie	Pandey	A	9.05E+09	12	Chennai|600001
14	Mohit	Pandey	A	9.86E+09	43	Mumbai|400001
15	Ramesh	Shah	A	9.87E+09	220	Mumbai|400001
16	Kishn	Chandra	A	9.05E+09	93	Chennai|600001
17	Amrish	Singh	A	9.86E+09	332	Mumbai|400001
18	Ishan	Mishra	A	9.99E+09	338	Bangalore|560001
19	Abhinav	Chandra	A	9.86E+09	324	Mumbai|400001
20	Mahendra	Vikram	I	9.87E+09	55	Mumbai|400001
21	Tarun	Singh	A	9.06E+09	647	Chennai|600001
22	Upendra	Sengal	A	9.86E+09	102	Mumbai|400001
23	Abhinav	Dwivedi	A	9.05E+09	227	Chennai|600001
24	Rohit	Purwar	A	9.86E+09	323	Mumbai|400001
25	Prashant	Maheshw	A	9.86E+09	47	Mumbai|400001
26	Prashant	Asthana	A	9.05E+09	431	Chennai|600001

3. We should start the hadoop cluster before starting the Hive.

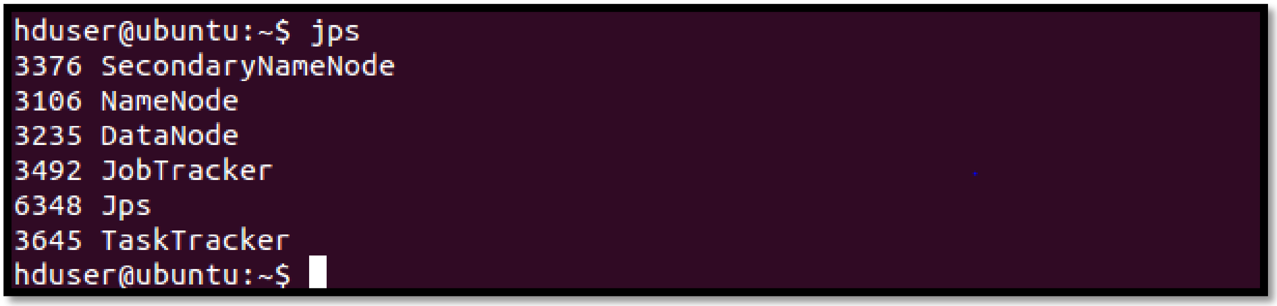
```
$ start-all.sh
            OR
$ /usr/local/hadoop-1.2.1/bin/start-all.sh
```



```
hduser@hduser:~$ start-all.sh
Warning: $HADOOP_HOME is deprecated.

starting namenode, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-hdu
ser-namenode-hduser.out
localhost: starting datanode, logging to /usr/local/hadoop-1.2.1/libexec/../logs
/hadoop-hduser-datanode-hduser.out
localhost: starting secondarynamenode, logging to /usr/local/hadoop-1.2.1/libexe
c/../logs/hadoop-hduser-secondarynamenode-hduser.out
starting jobtracker, logging to /usr/local/hadoop-1.2.1/libexec/../logs/hadoop-h
duser-jobtracker-hduser.out
localhost: starting tasktracker, logging to /usr/local/hadoop-1.2.1/libexec/../l
ogs/hadoop-hduser-tasktracker-hduser.out
hduser@hduser:~$
```

Now check whether the hadoop daemons are started.

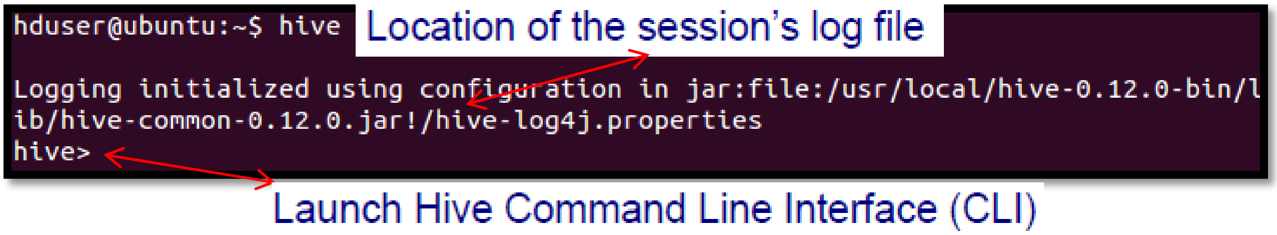
```
$ jps
```



```
hduser@ubuntu:~$ jps
3376 SecondaryNameNode
3106 NameNode
3235 DataNode
3492 JobTracker
6348 Jps
3645 TaskTracker
hduser@ubuntu:~$
```

4. Now start the hive

```
$ hive
        OR
$ /usr/local/hive-0.12.0-bin/bin/hive
```



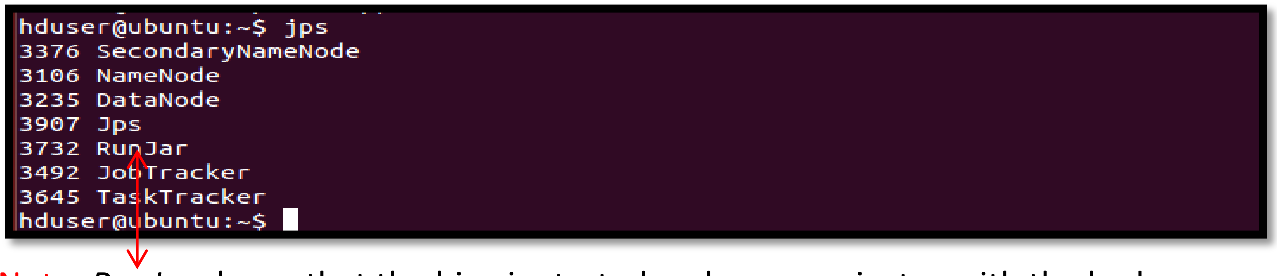
```
hduser@ubuntu:~$ hive

Logging initialized using configuration in jar:file:/usr/local/hive-0.12.0-bin/l
ib/hive-common-0.12.0.jar!/hive-log4j.properties
hive>
```

Location of the session's log file

Launch Hive Command Line Interface (CLI)

Now check whether the hive is started with the hadoop

```
$ jps
```



```
hduser@ubuntu:~$ jps
3376 SecondaryNameNode
3106 NameNode
3235 DataNode
3907 Jps
3732 RunJar
3492 JobTracker
3645 TaskTracker
hduser@ubuntu:~$
```

Note: *RunJar* shows that the hive is started and communicates with the hadoop

PROCEDURE

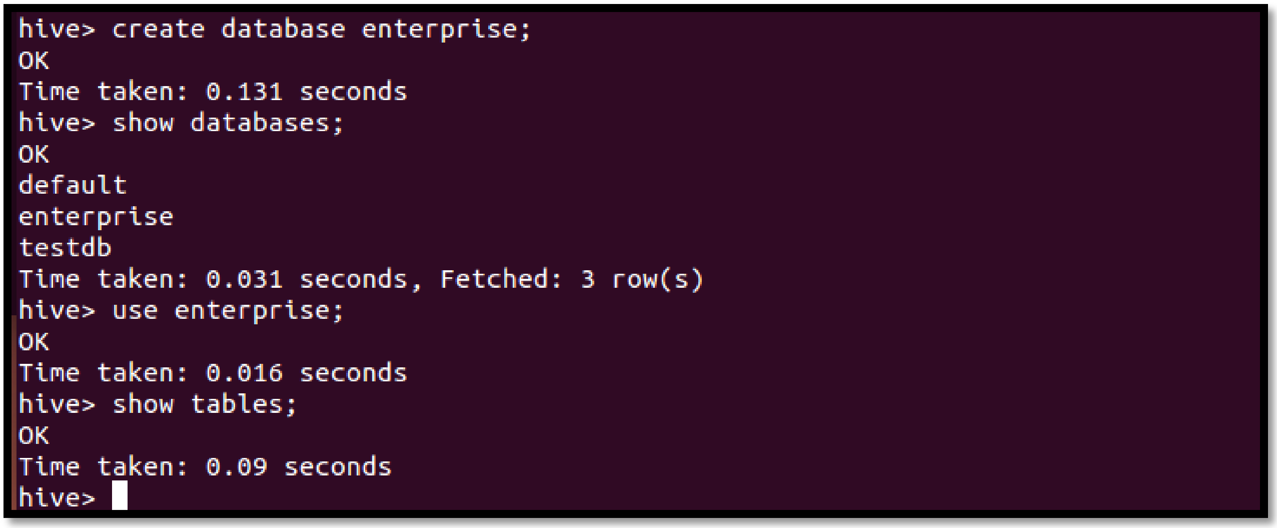
Hive stores its tables on HDFS and those locations needs to be bootstrapped.

```
$ hadoop dfs  -mkdir /temp

$ hadoop dfs  -mkdir /user/hive/warehouse

$ hadoop dfs  -chmod g+w /temp

$ hadoop dfs  -chmod g+w /user/hive/warehouse
```

Lets us start with the Hive queries:

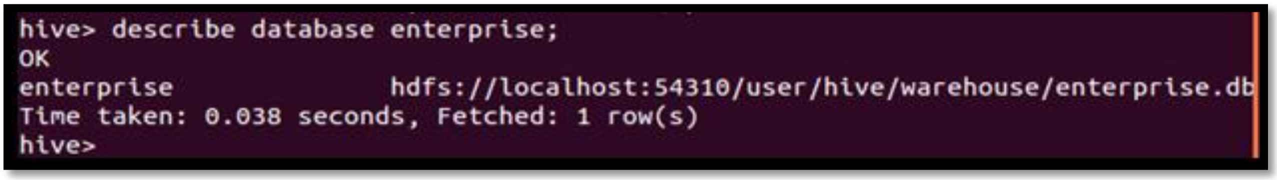
CREATE DATABASE <data base name> to create the new database in the Hive.
USE <data base name> to use existing database
SHOW <table name> to display tables

```
hive> create database enterprise;

hive> use enterprise;

hive> show tables;
```



```
hive> create database enterprise;
OK
Time taken: 0.131 seconds
hive> show databases;
OK
default
enterprise
testdb
Time taken: 0.031 seconds, Fetched: 3 row(s)
hive> use enterprise;
OK
Time taken: 0.016 seconds
hive> show tables;
OK
Time taken: 0.09 seconds
hive>
```

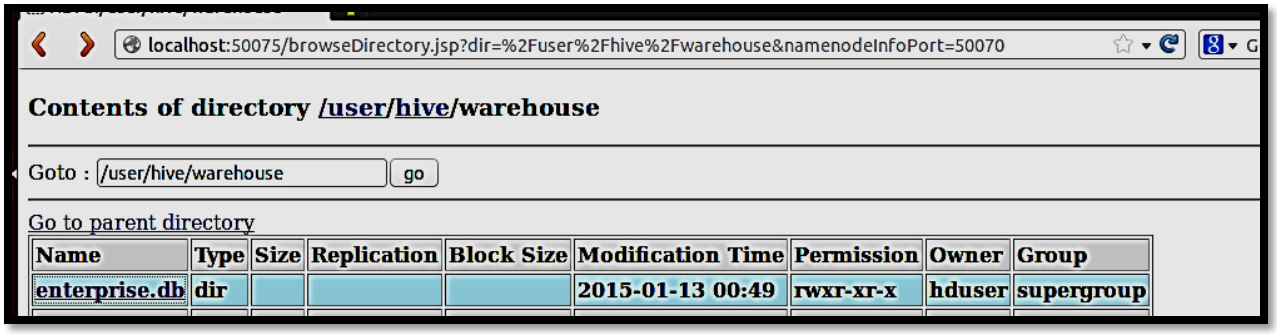
```
hive>describe database enterprise;
```



```
hive> describe database enterprise;
OK
enterprise              hdfs://localhost:54310/user/hive/warehouse/enterprise.db
Time taken: 0.038 seconds, Fetched: 1 row(s)
hive>
```

To see the location where the enterprise database stored in the browser

http://localhost:50075/browseDirectory.jsp?dir=%2Fuser%2Fhive%2Fwarehouse&namenodeInfoPort=50070



localhost:50075/browseDirectory.jsp?dir=%2Fuser%2Fhive%2Fwarehouse&namenodeInfoPort=50070

Contents of directory /user/hive/warehouse

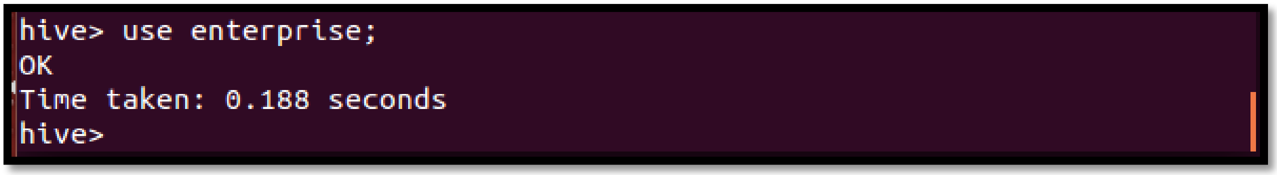
Goto : /user/hive/warehouse go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
enterprise.db	dir				2015-01-13 00:49	rwxr-xr-x	hduser	supergroup

Now let us starts working with the real data set called enterprise retail data

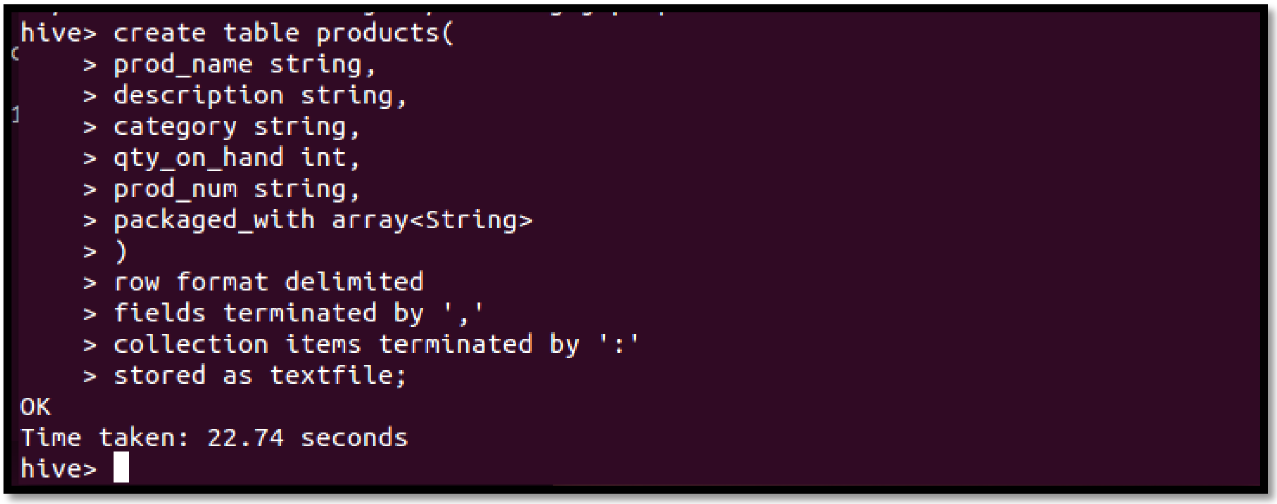
```
hive>use enterprise;
```



```
hive> use enterprise;
OK
Time taken: 0.188 seconds
hive>
```

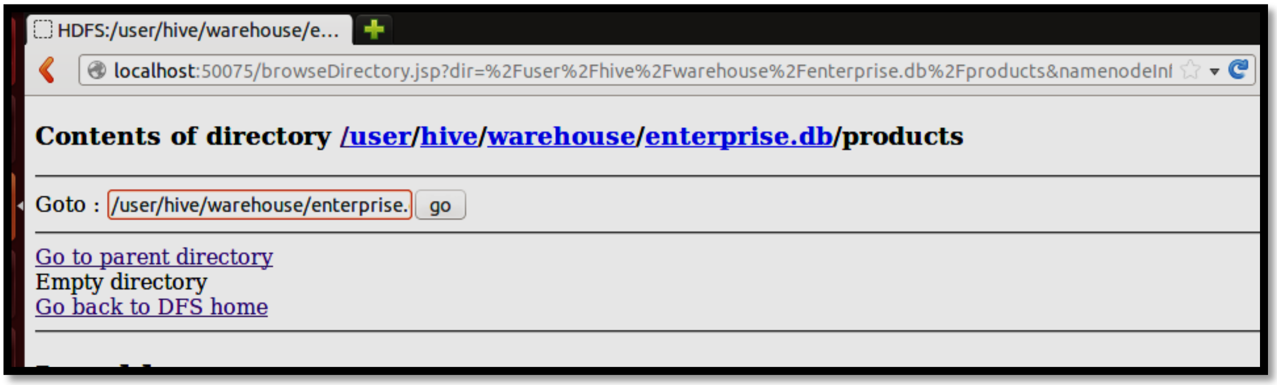
Now **create a products table inside the enterprise database**

```
create table products(
prod_name string,
description string,
category string,
qty_on_hand int,
prod_num string,
packaged_with array<String>
)
row format delimited
fields terminated by ','
collection items terminated by ':'
stored as textfile;
```



```
hive> create table products(
    > prod_name string,
    > description string,
    > category string,
    > qty_on_hand int,
    > prod_num string,
    > packaged_with array<String>
    > )
    > row format delimited
    > fields terminated by ','
    > collection items terminated by ':'
    > stored as textfile;
OK
Time taken: 22.74 seconds
hive>
```

In GUI:



localhost:50075/browseDirectory.jsp?dir=%2Fuser%2Fhive%2Fwarehouse%2Fenterprise.db%2Fproducts&namenodeInf

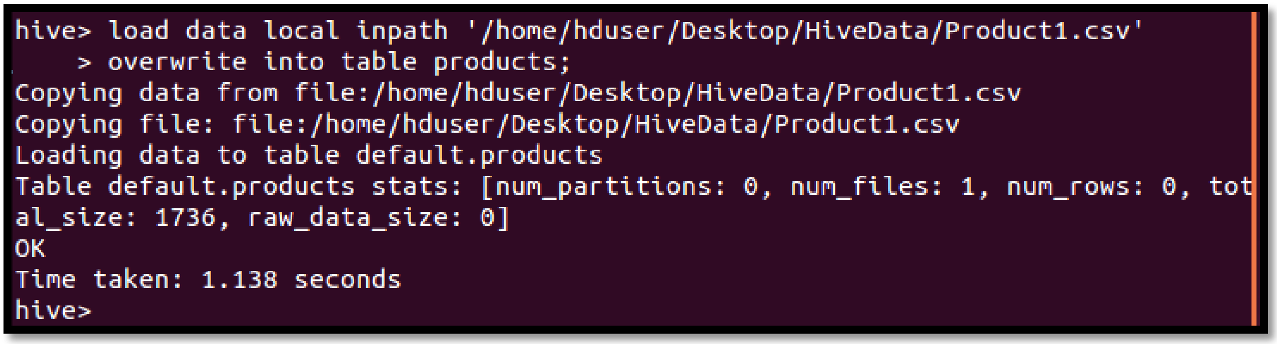
Contents of directory /user/hive/warehouse/enterprise.db/products

Goto : /user/hive/warehouse/enterprise. go

Go to parent directory
Empty directory
Go back to DFS home

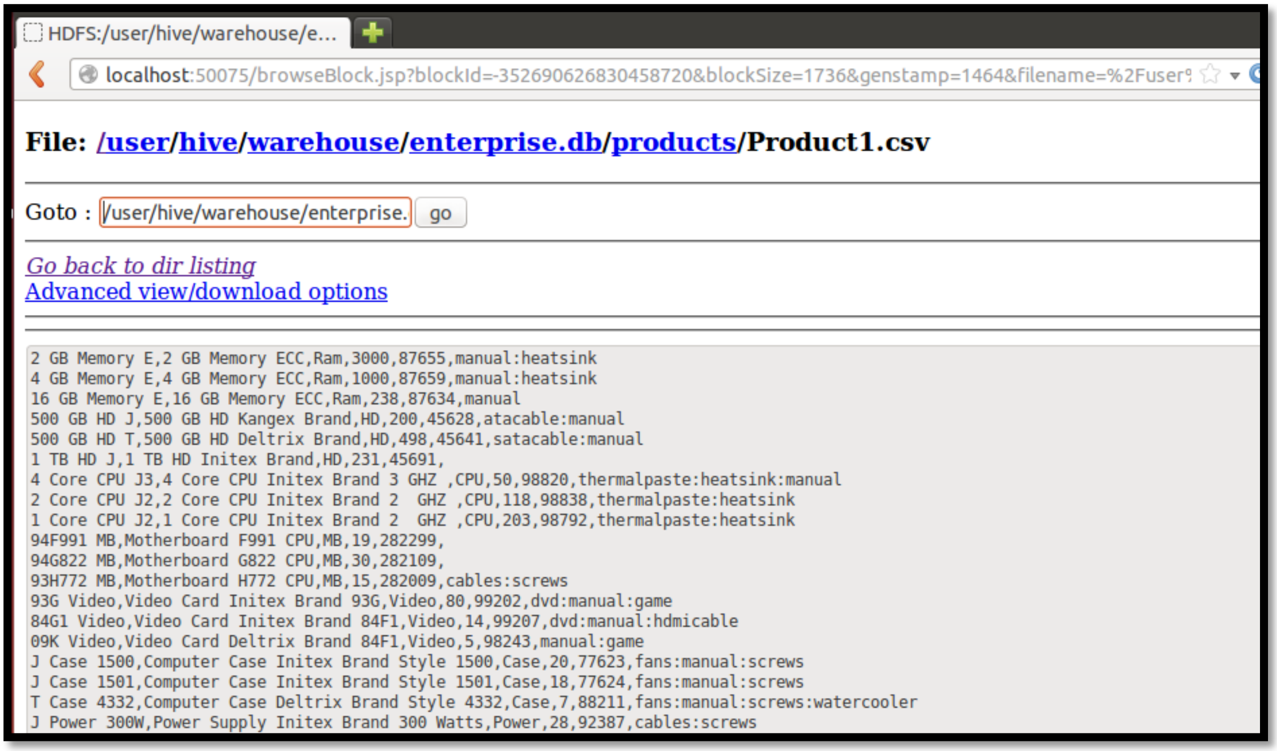
Now copy **Product1.csv file** located at local file system path '/home/hduser/Desktop/HiveData/Product1.csv' to the **products table** created.

```
load data local inpath '/home/hduser/Desktop/HiveData/Product1.csv'
overwrite into table products;
```



```
hive> load data local inpath '/home/hduser/Desktop/HiveData/Product1.csv'
    > overwrite into table products;
Copying data from file:/home/hduser/Desktop/HiveData/Product1.csv
Copying file: file:/home/hduser/Desktop/HiveData/Product1.csv
Loading data to table default.products
Table default.products stats: [num_partitions: 0, num_files: 1, num_rows: 0, tot
al_size: 1736, raw_data_size: 0]
OK
Time taken: 1.138 seconds
hive>
```

In GUI:



HDFS:/user/hive/warehouse/e...
localhost:50075/browseBlock.jsp?blockId=-3526906268304587208&blockSize=1736&genstamp=1464&filename=%2Fuser%

File: /user/hive/warehouse/enterprise.db/products/Product1.csv

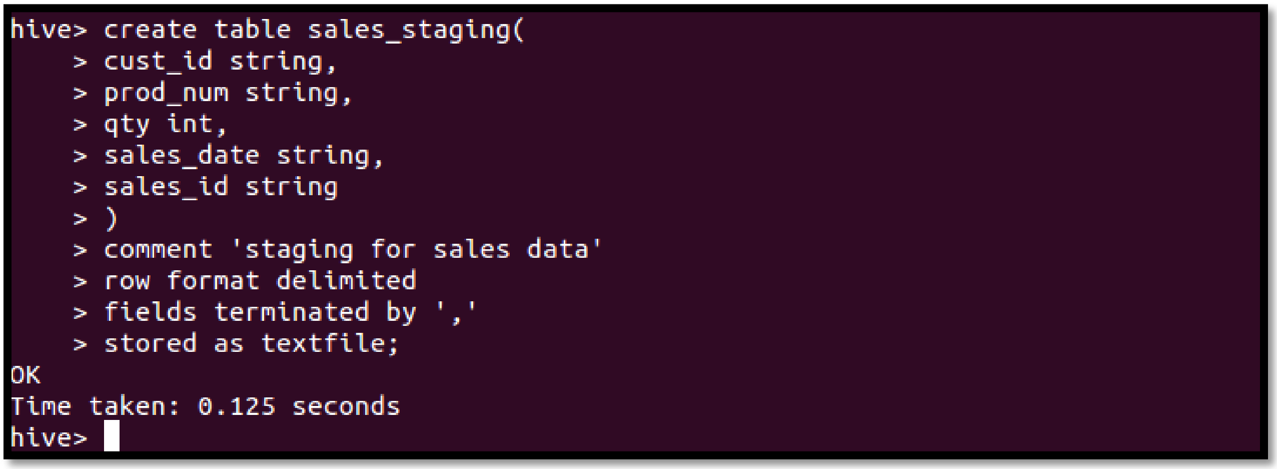
Goto : /user/hive/warehouse/enterprise. go

Go back to dir listing
Advanced view/download options

```
2 GB Memory E,2 GB Memory ECC,Ram,3000,87655,manual:heatsink
4 GB Memory E,4 GB Memory ECC,Ram,1000,87659,manual:heatsink
16 GB Memory E,16 GB Memory ECC,Ram,238,87634,manual
500 GB HD J,500 GB HD Kangex Brand,HD,200,45628,atacable:manual
500 GB HD T,500 GB HD Deltrix Brand,HD,498,45641,satacable:manual
1 TB HD J,1 TB HD Initex Brand,HD,231,45691,
4 Core CPU J3,4 Core CPU Initex Brand 3 GHZ ,CPU,50,98820,thermalpaste:heatsink:manual
2 Core CPU J2,2 Core CPU Initex Brand 2  GHZ ,CPU,118,98838,thermalpaste:heatsink
1 Core CPU J2,1 Core CPU Initex Brand 2  GHZ ,CPU,203,98792,thermalpaste:heatsink
94F991 MB,Motherboard F991 CPU,MB,19,282299,
94G822 MB,Motherboard G822 CPU,MB,30,282109,
93H772 MB,Motherboard H772 CPU,MB,15,282009,cables:screws
93G Video,Video Card Initex Brand 93G,Video,80,99202,dvd:manual:game
84G1 Video,Video Card Initex Brand 84F1,Video,14,99207,dvd:manual:hdmicable
09K Video,Video Card Deltrix Brand 84F1,Video,5,98243,manual:game
J Case 1500,Computer Case Initex Brand Style 1500,Case,20,77623,fans:manual:screws
J Case 1501,Computer Case Initex Brand Style 1501,Case,18,77624,fans:manual:screws
T Case 4332,Computer Case Deltrix Brand Style 4332,Case,7,88211,fans:manual:screws:watercooler
J Power 300W,Power Supply Initex Brand 300 Watts,Power,28,92387,cables:screws
```

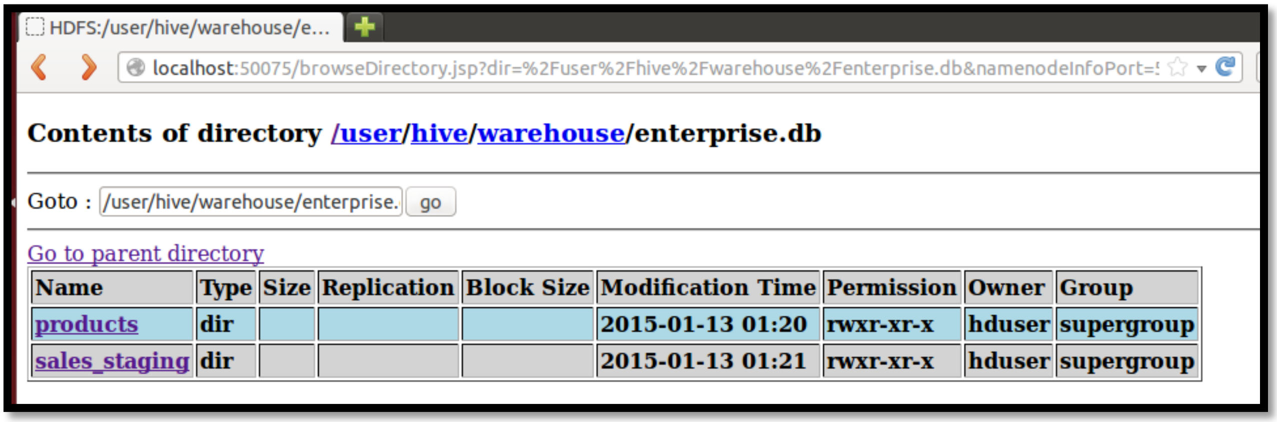
Now create a **sales_staging table** inside the **enterprise database**

```
create table sales_staging(
cust_id string,
prod_num string,
qty int,
sales_date string,
sales_id string
)
comment 'staging for sales data'
row format delimited
fields terminated by ','
stored as textfile;
```



```
hive> create table sales_staging(
    > cust_id string,
    > prod_num string,
    > qty int,
    > sales_date string,
    > sales_id string
    > )
    > comment 'staging for sales data'
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 0.125 seconds
hive>
```

In GUI:



localhost:50075/browseDirectory.jsp?dir=%2Fuser%2Fhive%2Fwarehouse%2Fenterprise.db&namenodeInfoPort=

Contents of directory /user/hive/warehouse/enterprise.db

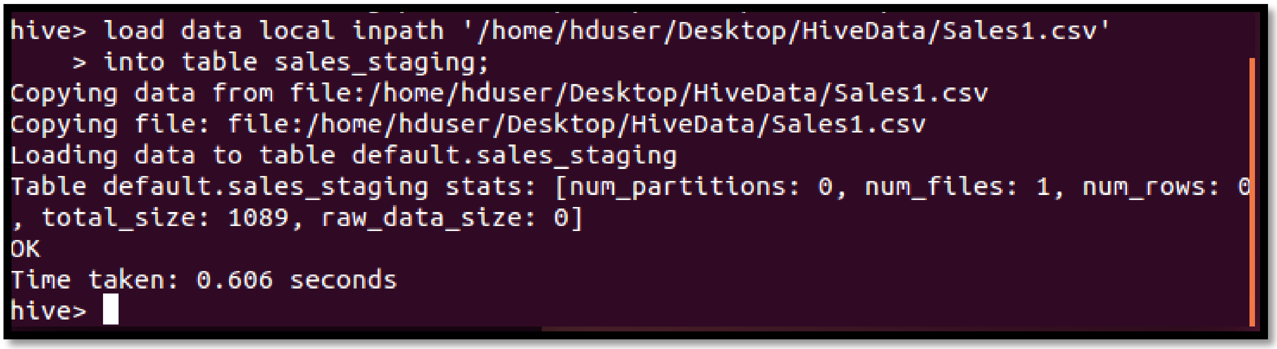
Goto : /user/hive/warehouse/enterprise.

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
products	dir				2015-01-13 01:20	rwxr-xr-x	hduser	supergroup
sales_staging	dir				2015-01-13 01:21	rwxr-xr-x	hduser	supergroup

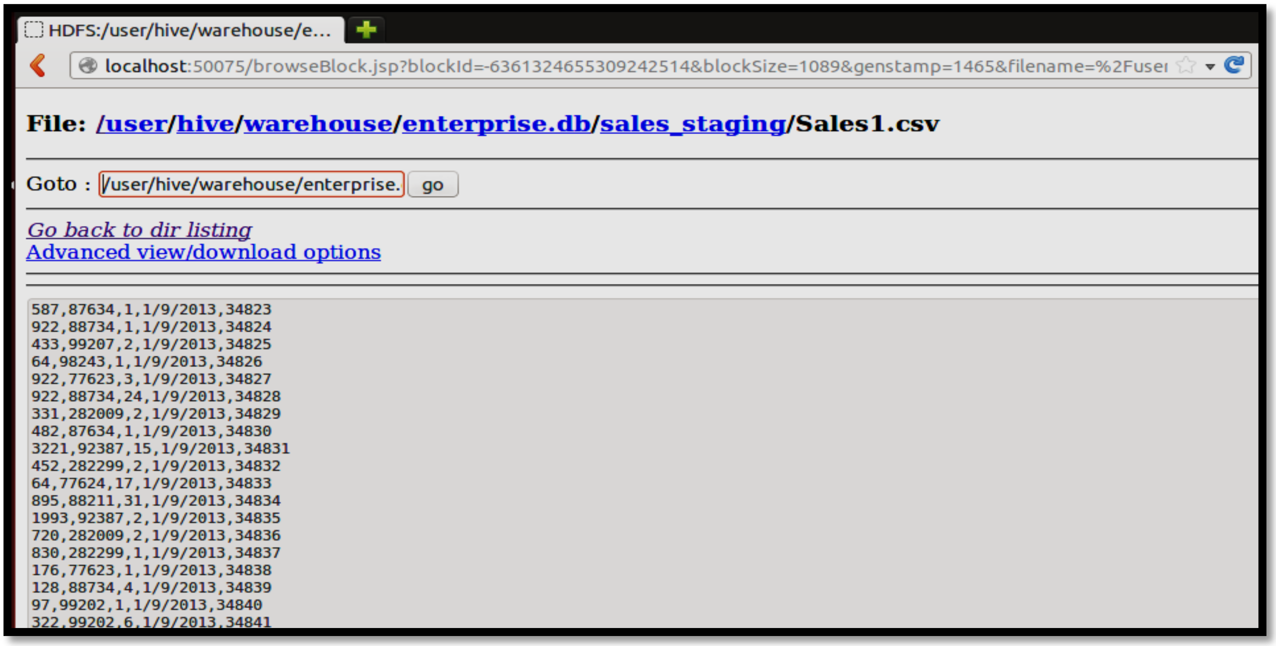
Now copy **Sales1.csv file** located at local file system path '/home/hduser/Desktop/HiveData/Sales1.csv' to the **Sales table** created.

```
 load data local inpath '/home/hduser/Desktop/HiveData/Sales1.csv'
into table sales_staging;
```



```
hive> load data local inpath '/home/hduser/Desktop/HiveData/Sales1.csv'
    > into table sales_staging;
Copying data from file:/home/hduser/Desktop/HiveData/Sales1.csv
Copying file: file:/home/hduser/Desktop/HiveData/Sales1.csv
Loading data to table default.sales_staging
Table default.sales_staging stats: [num_partitions: 0, num_files: 1, num_rows: 0
, total_size: 1089, raw_data_size: 0]
OK
Time taken: 0.606 seconds
hive>
```

In GUI:



HDFS:/user/hive/warehouse/e...
localhost:50075/browseBlock.jsp?blockId=-6361324655309242514&blockSize=1089&genstamp=1465&filename=%2Fuser
go

File: /user/hive/warehouse/enterprise.db/sales_staging/Sales1.csv

Goto : /user/hive/warehouse/enterprise.

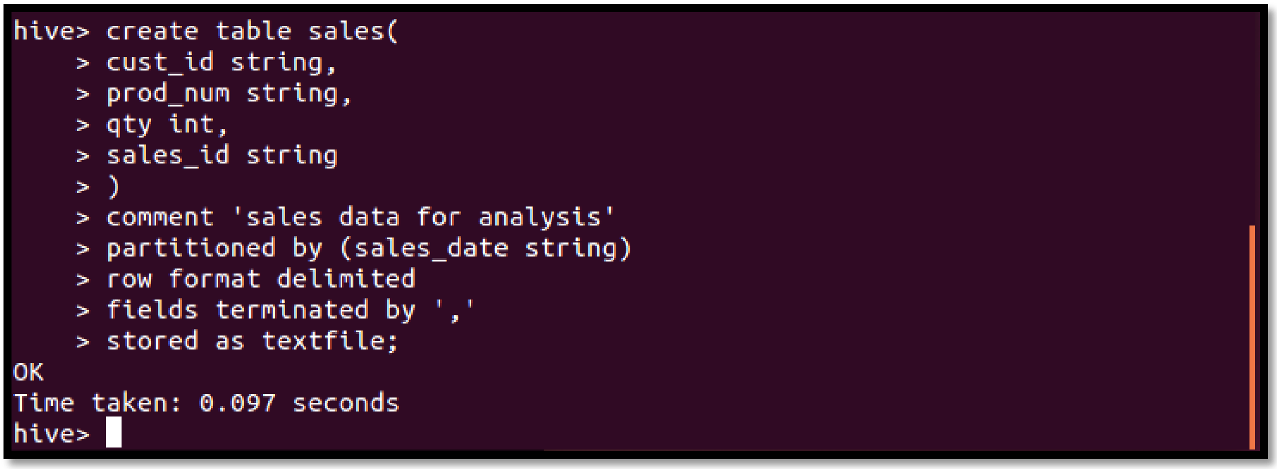
Go back to dir listing
Advanced view/download options

```
587,87634,1,1/9/2013,34823
922,88734,1,1/9/2013,34824
433,99207,2,1/9/2013,34825
64,98243,1,1/9/2013,34826
922,77623,3,1/9/2013,34827
922,88734,24,1/9/2013,34828
331,282009,2,1/9/2013,34829
482,87634,1,1/9/2013,34830
3221,92387,15,1/9/2013,34831
452,282299,2,1/9/2013,34832
64,77624,17,1/9/2013,34833
895,88211,31,1/9/2013,34834
1993,92387,2,1/9/2013,34835
720,282009,2,1/9/2013,34836
830,282299,1,1/9/2013,34837
176,77623,1,1/9/2013,34838
128,88734,4,1/9/2013,34839
97,99202,1,1/9/2013,34840
322,99202,6,1/9/2013,34841
```

Now create **a partitioned sales table** inside the **enterprise database** based on **sales_date**

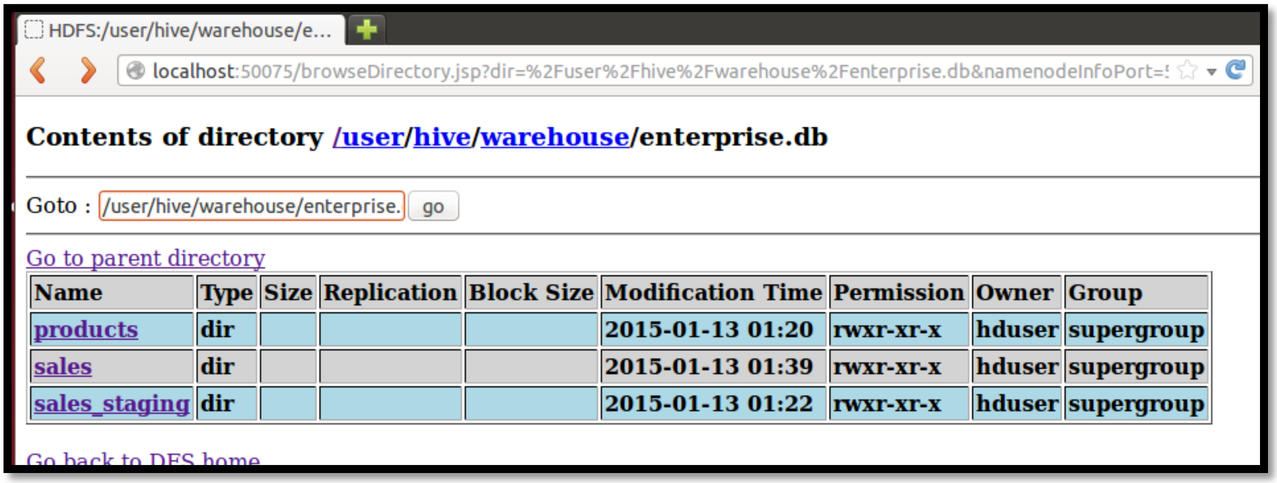
```
create table sales(
cust_id string,
prod_num string,
qty int,
sales_id string
)
comment 'sales data for analysis'
partitioned by (sales_date string)
row format delimited
fields terminated by ','
stored as textfile;
```

- **To increase performance Hive has the capability to partition data**
  - The values of partitioned column divide a table into segments
  - Entire partitions can be ignored at query time
  - Similar to relational databases' indexes but not as granular
- **Partitions have to be properly crated by users**
  - When inserting data must specify a partition
- **At query time, whenever appropriate, Hive will automatically filter out partitions**



```
hive> create table sales(
    > cust_id string,
    > prod_num string,
    > qty int,
    > sales_id string
    > )
    > comment 'sales data for analysis'
    > partitioned by (sales_date string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 0.097 seconds
hive> 
```

In GUI:



localhost:50075/browseDirectory.jsp?dir=%2Fuser%2Fhive%2Fwarehouse%2Fenterprise.db&namenodeInfoPort=!

Contents of directory /user/hive/warehouse/enterprise.db

Goto : /user/hive/warehouse/enterprise. go

Go to parent directory

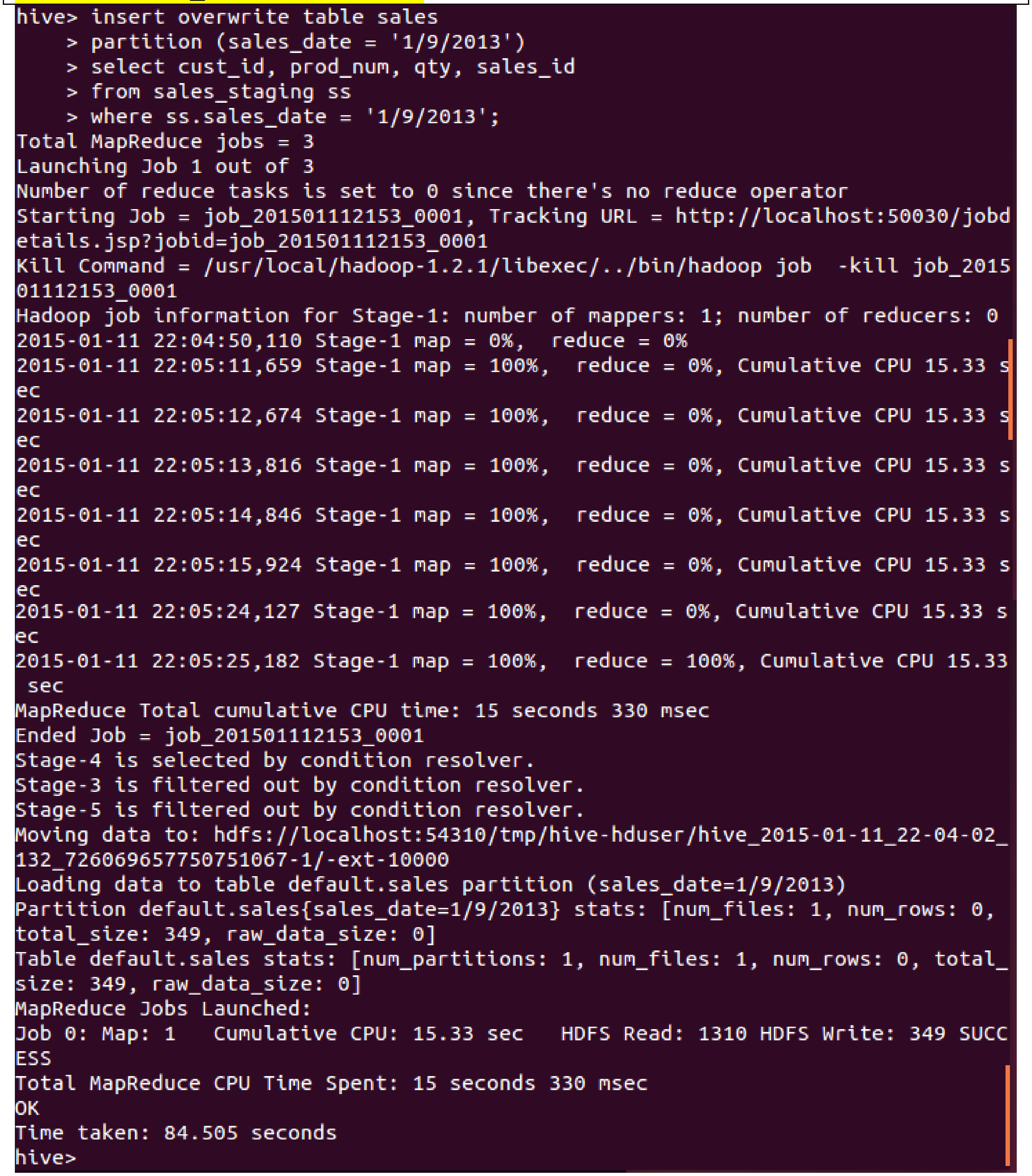
Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
products	dir				2015-01-13 01:20	rwxr-xr-x	hduser	supergroup
sales	dir				2015-01-13 01:39	rwxr-xr-x	hduser	supergroup
sales_staging	dir				2015-01-13 01:22	rwxr-xr-x	hduser	supergroup

Go back to DFS home

Now insert the data into **sales table** from **sales_staging table** based on **sales_date='1/9/2013'**

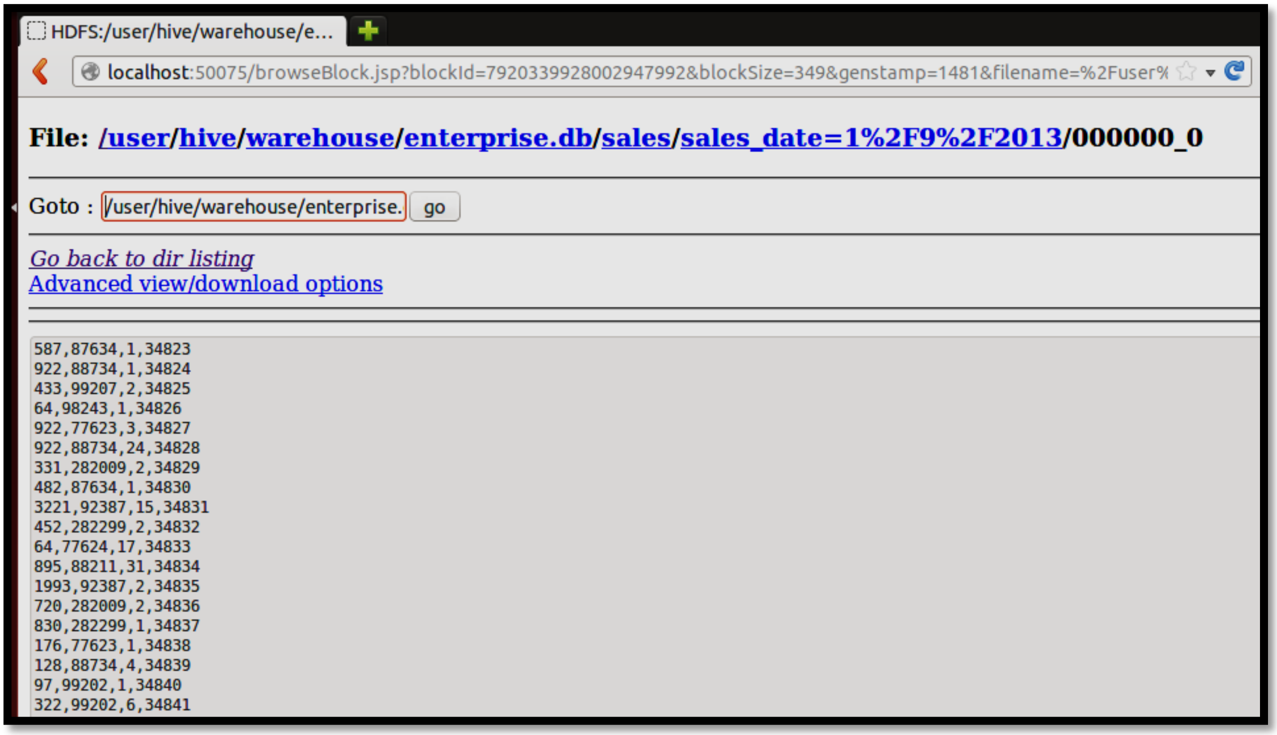
```
insert overwrite table sales
partition (sales_date = '1/9/2013')
select cust_id, prod_num, qty, sales_id
from sales_staging ss
```

```
where ss.sales_date = '1/9/2013';
```



```
hive> insert overwrite table sales
    > partition (sales_date = '1/9/2013')
    > select cust_id, prod_num, qty, sales_id
    > from sales_staging ss
    > where ss.sales_date = '1/9/2013';
Total MapReduce jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201501112153_0001, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501112153_0001
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01112153_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-01-11 22:04:50,110 Stage-1 map = 0%,  reduce = 0%
2015-01-11 22:05:11,659 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 15.33 s
ec
2015-01-11 22:05:12,674 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 15.33 s
ec
2015-01-11 22:05:13,816 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 15.33 s
ec
2015-01-11 22:05:14,846 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 15.33 s
ec
2015-01-11 22:05:15,924 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 15.33 s
ec
2015-01-11 22:05:24,127 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 15.33 s
ec
2015-01-11 22:05:25,182 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.33
 sec
MapReduce Total cumulative CPU time: 15 seconds 330 msec
Ended Job = job_201501112153_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://localhost:54310/tmp/hive-hduser/hive_2015-01-11_22-04-02_
132_726069657750751067-1/-ext-10000
Loading data to table default.sales partition (sales_date=1/9/2013)
Partition default.sales{sales_date=1/9/2013} stats: [num_files: 1, num_rows: 0,
total_size: 349, raw_data_size: 0]
Table default.sales stats: [num_partitions: 1, num_files: 1, num_rows: 0, total_
size: 349, raw_data_size: 0]
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 15.33 sec   HDFS Read: 1310 HDFS Write: 349 SUCC
ESS
Total MapReduce CPU Time Spent: 15 seconds 330 msec
OK
Time taken: 84.505 seconds
hive>
```

In GUI:



HDFS:/user/hive/warehouse/e...
localhost:50075/browseBlock.jsp?blockId=7920339928002947992&blockSize=349&genstamp=1481&filename=%2Fuser%

File: /user/hive/warehouse/enterprise.db/sales/sales_date=1%2F9%2F2013/000000_0

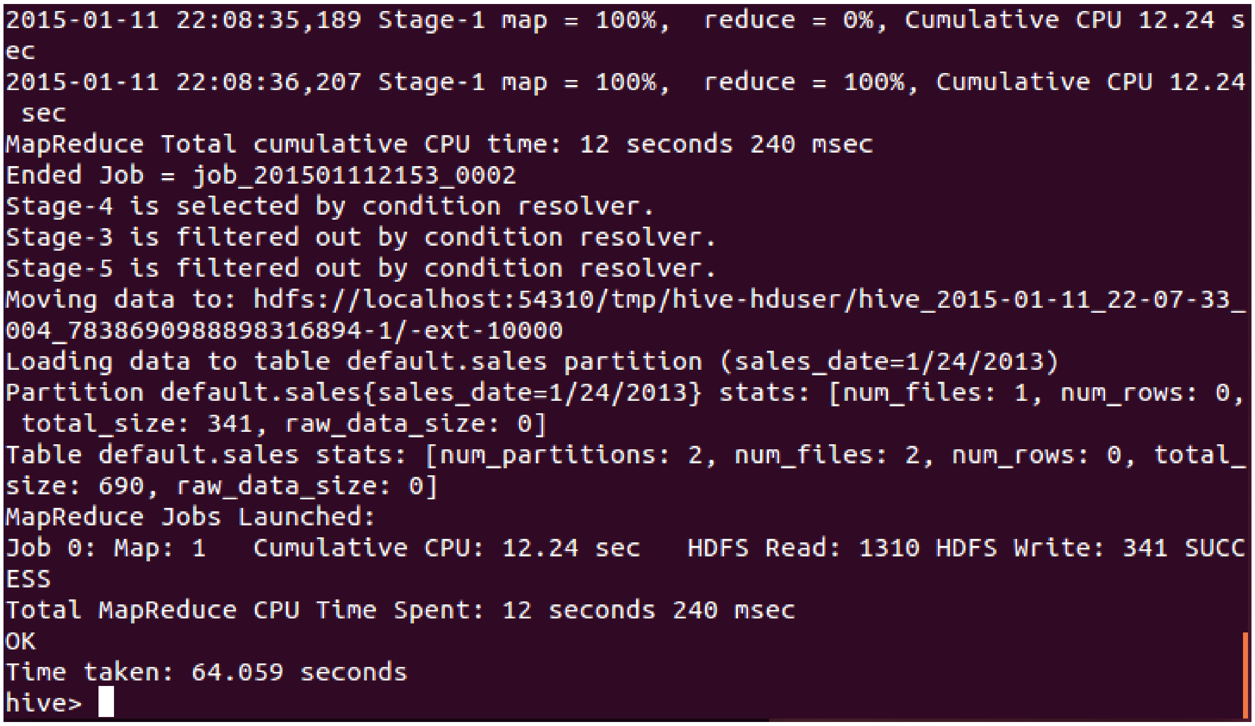
Goto : /user/hive/warehouse/enterprise. go

Go back to dir listing
Advanced view/download options

```
587,87634,1,34823
922,88734,1,34824
433,99207,2,34825
64,98243,1,34826
922,77623,3,34827
922,88734,24,34828
331,282009,2,34829
482,87634,1,34830
3221,92387,15,34831
452,282299,2,34832
64,77624,17,34833
895,88211,31,34834
1993,92387,2,34835
720,282009,2,34836
830,282299,1,34837
176,77623,1,34838
128,88734,4,34839
97,99202,1,34840
322,99202,6,34841
```

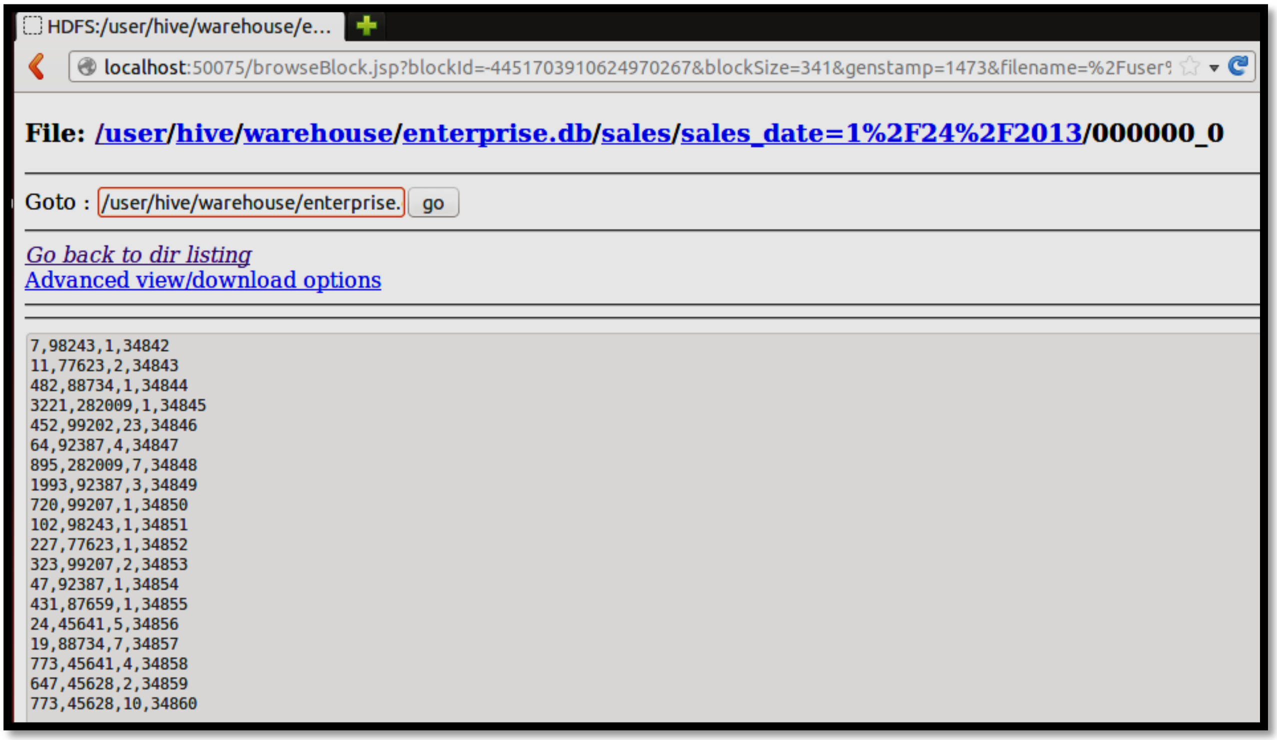
Now insert the data into **sales table** from **sales_staging table** based on matched **sales_date='1/24/2013'**

```
insert overwrite table sales
partition (sales_date = '1/24/2013')
select cust_id, prod_num, qty, sales_id
from sales_staging ss
where ss.sales_date = '1/24/2013';
```



```
2015-01-11 22:08:35,189 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 12.24 s
ec
2015-01-11 22:08:36,207 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 12.24
 sec
MapReduce Total cumulative CPU time: 12 seconds 240 msec
Ended Job = job_201501112153_0002
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://localhost:54310/tmp/hive-hduser/hive_2015-01-11_22-07-33_
004_7838690988898316894-1/-ext-10000
Loading data to table default.sales partition (sales_date=1/24/2013)
Partition default.sales{sales_date=1/24/2013} stats: [num_files: 1, num_rows: 0,
 total_size: 341, raw_data_size: 0]
Table default.sales stats: [num_partitions: 2, num_files: 2, num_rows: 0, total_
size: 690, raw_data_size: 0]
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 12.24 sec   HDFS Read: 1310 HDFS Write: 341 SUCC
ESS
Total MapReduce CPU Time Spent: 12 seconds 240 msec
OK
Time taken: 64.059 seconds
hive>
```

In GUI:



HDFS:/user/hive/warehouse/e...
localhost:50075/browseBlock.jsp?blockId=-4451703910624970267&blockSize=341&genstamp=1473&filename=%2Fuser%

File: /user/hive/warehouse/enterprise.db/sales/sales_date=1%2F24%2F2013/000000_0

Goto : /user/hive/warehouse/enterprise. go

Go back to dir listing
Advanced view/download options

```
7,98243,1,34842
11,77623,2,34843
482,88734,1,34844
3221,282009,1,34845
452,99202,23,34846
64,92387,4,34847
895,282009,7,34848
1993,92387,3,34849
720,99207,1,34850
102,98243,1,34851
227,77623,1,34852
323,99207,2,34853
47,92387,1,34854
431,87659,1,34855
24,45641,5,34856
19,88734,7,34857
773,45641,4,34858
647,45628,2,34859
773,45628,10,34860
```

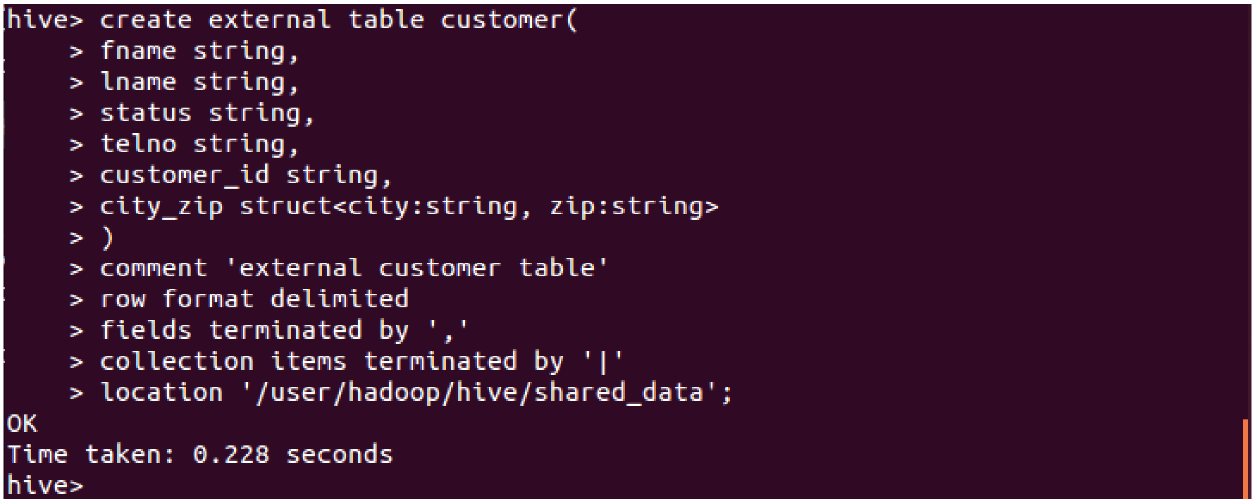
create **/user/hadoop/hive/shared_data** at HDSF and put Customer1.csv inside it.

Note:
Open a new terminal and create **shared_data** folder:

```
$ /usr/local/hadoop-1.2.1/bin/hadoop dfs –mkdir /user/hadoop/
hive/shared_data
```

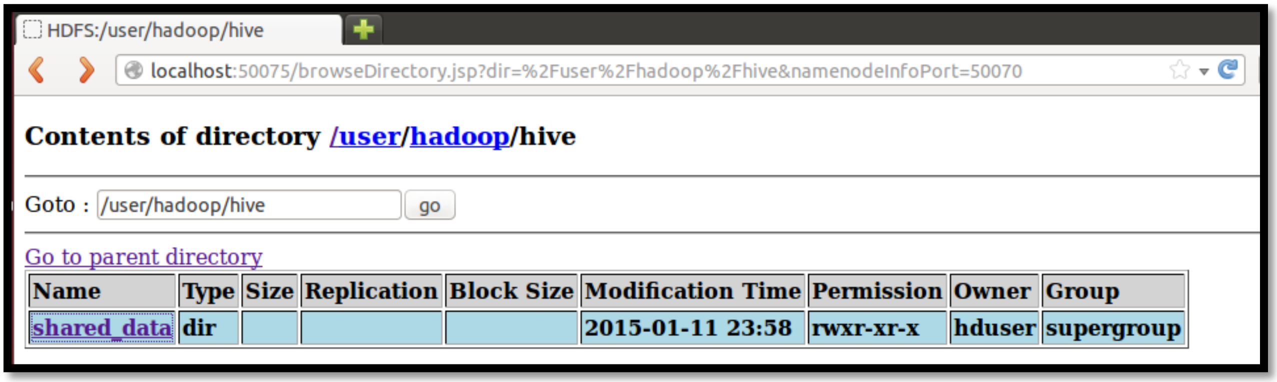
Now go to hive command line interface and create **customer** table outside the Hive Warehouse as an **external table**.

```
create external table customer(
fname string,
lname string,
status string,
telno string,
customer_id string,
city_zip struct<city:string, zip:string>
)
comment 'external customer table'
row format delimited
fields terminated by ','
collection items terminated by '|'
location '/user/hadoop/hive/shared_data';
```



```
hive> create external table customer(
    > fname string,
    > lname string,
    > status string,
    > telno string,
    > customer_id string,
    > city_zip struct<city:string, zip:string>
    > )
    > comment 'external customer table'
    > row format delimited
    > fields terminated by ','
    > collection items terminated by '|'
    > location '/user/hadoop/hive/shared_data';
OK
Time taken: 0.228 seconds
hive>
```

In GUI:


HDFS:/user/hadoop/hive
localhost:50075/browseDirectory.jsp?dir=%2Fuser%2Fhadoop%2Fhive&namenodeInfoPort=50070
go

Contents of directory /user/hadoop/hive

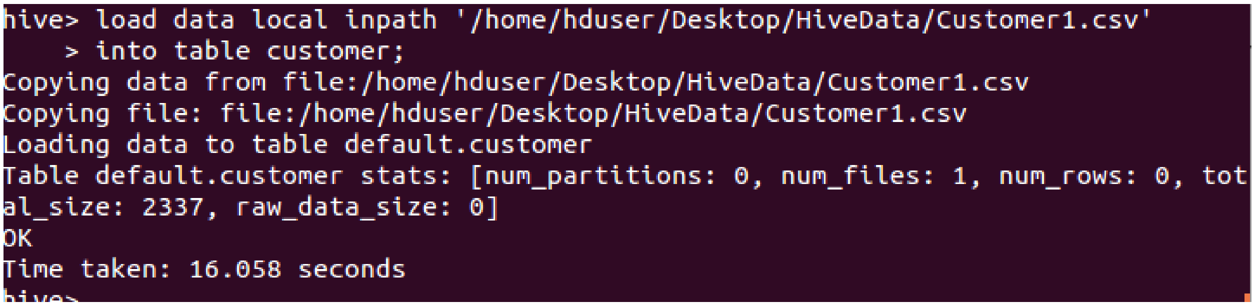
Goto : /user/hadoop/hive

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
shared_data	dir				2015-01-11 23:58	rwxr-xr-x	hduser	supergroup

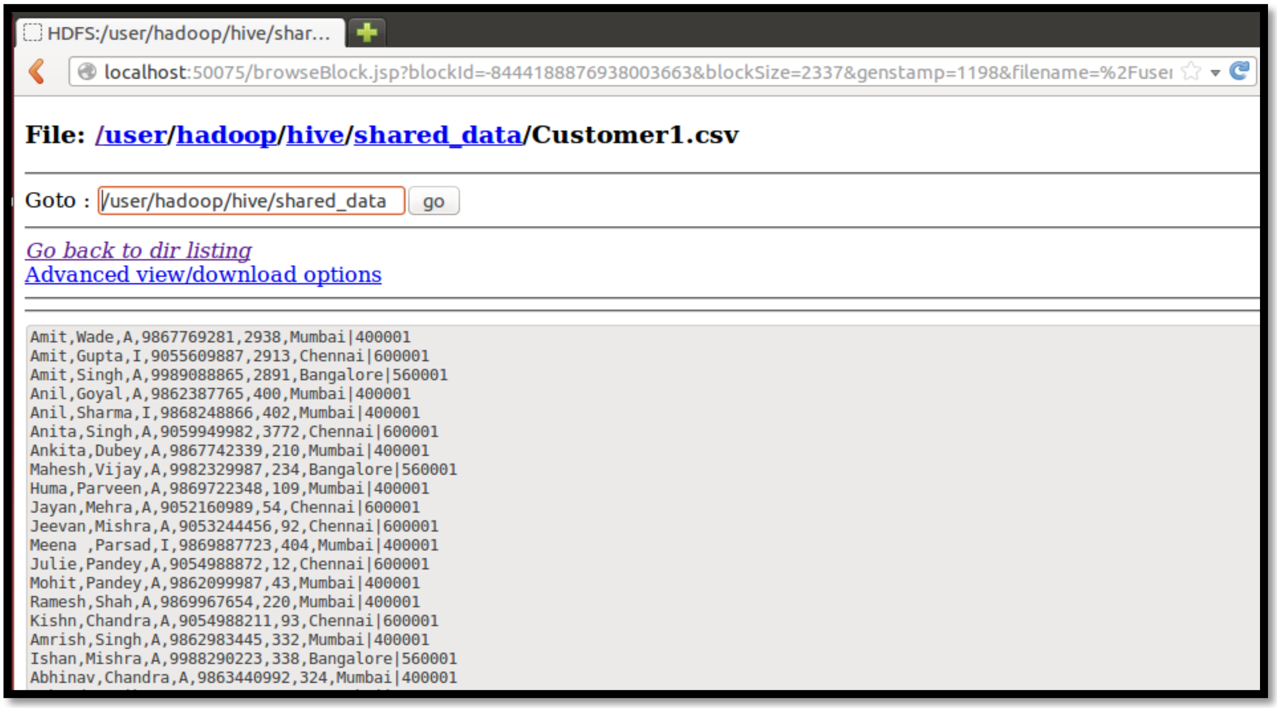
Now copy **Customer1.csv file** located at local file system path '/home/hduser/Desktop/HiveData/Product1.csv' to the **customer table** created.

```
load data local inpath '/home/hduser/Desktop/HiveData/Customer1.csv'
into table customer;
```



```
hive> load data local inpath '/home/hduser/Desktop/HiveData/Customer1.csv'
    > into table customer;
Copying data from file:/home/hduser/Desktop/HiveData/Customer1.csv
Copying file: file:/home/hduser/Desktop/HiveData/Customer1.csv
Loading data to table default.customer
Table default.customer stats: [num_partitions: 0, num_files: 1, num_rows: 0, tot
al_size: 2337, raw_data_size: 0]
OK
Time taken: 16.058 seconds
hive>
```

In GUI:



HDFS:/user/hadoop/hive/shar...
localhost:50075/browseBlock.jsp?blockId=-8444188876938003663&blockSize=2337&genstamp=1198&filename=%2Fuser

File: /user/hadoop/hive/shared_data/Customer1.csv

Goto : /user/hadoop/hive/shared_data go

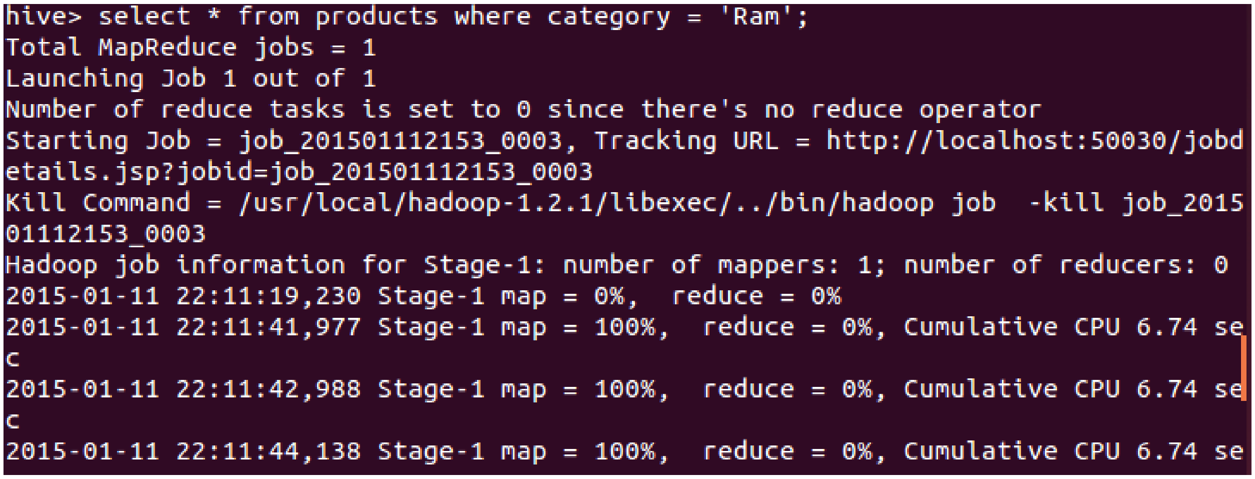
Go back to dir listing
Advanced view/download options

```
Amit,Wade,A,9867769281,2938,Mumbai|400001
Amit,Gupta,I,9055609887,2913,Chennai|600001
Amit,Singh,A,9989088865,2891,Bangalore|560001
Anil,Goyal,A,9862387765,400,Mumbai|400001
Anil,Sharma,I,9868248866,402,Mumbai|400001
Anita,Singh,A,9059949982,3772,Chennai|600001
Ankita,Dubey,A,9867742339,210,Mumbai|400001
Mahesh,Vijay,A,9982329987,234,Bangalore|560001
Huma,Parveen,A,9869722348,109,Mumbai|400001
Jayan,Mehra,A,9052160989,54,Chennai|600001
Jeevan,Mishra,A,9053244456,92,Chennai|600001
Meena ,Parsad,I,9869887723,404,Mumbai|400001
Julie,Pandey,A,9054988872,12,Chennai|600001
Mohit,Pandey,A,9862099987,43,Mumbai|400001
Ramesh,Shah,A,9869967654,220,Mumbai|400001
Kishn,Chandra,A,9054988211,93,Chennai|600001
Amrish,Singh,A,9862983445,332,Mumbai|400001
Ishan,Mishra,A,9988290223,338,Bangalore|560001
Abhinav,Chandra,A,9863440992,324,Mumbai|400001
```

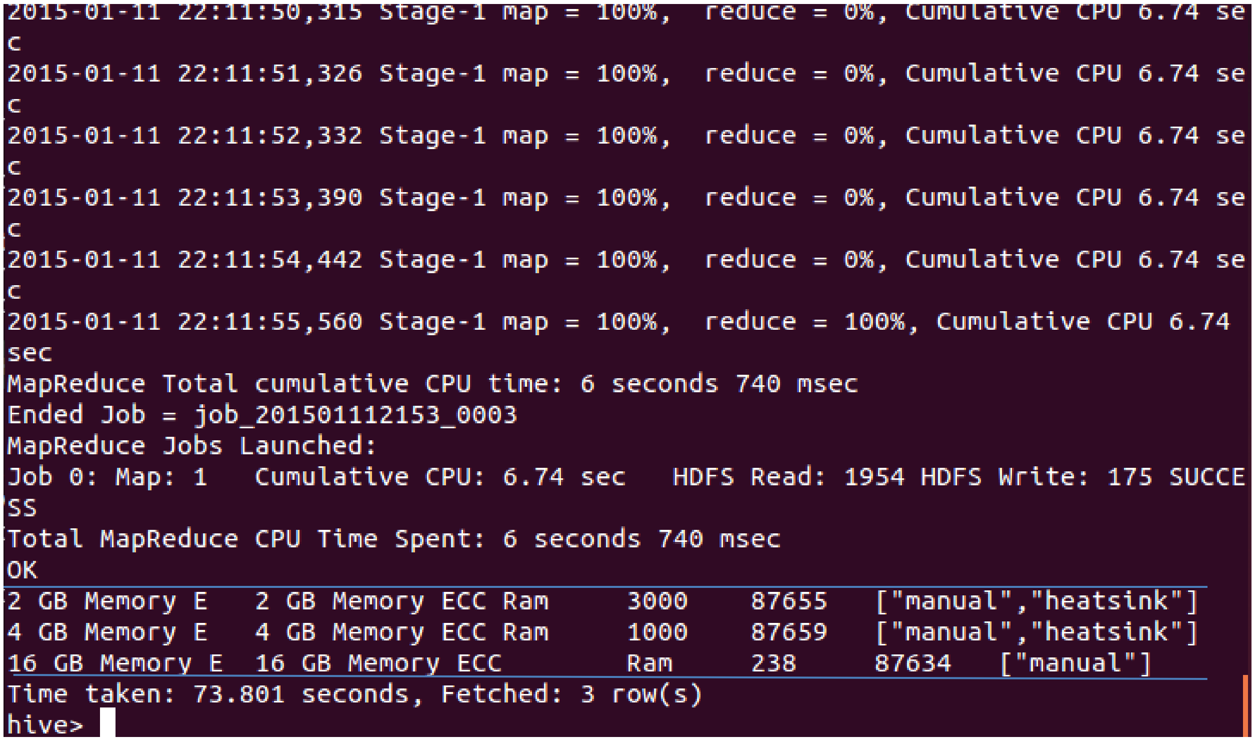
we are done with loading data from local system into the HDFS.

Now let us start with our HiveQL queries.

```
hive> select * from products where category = 'Ram';
```

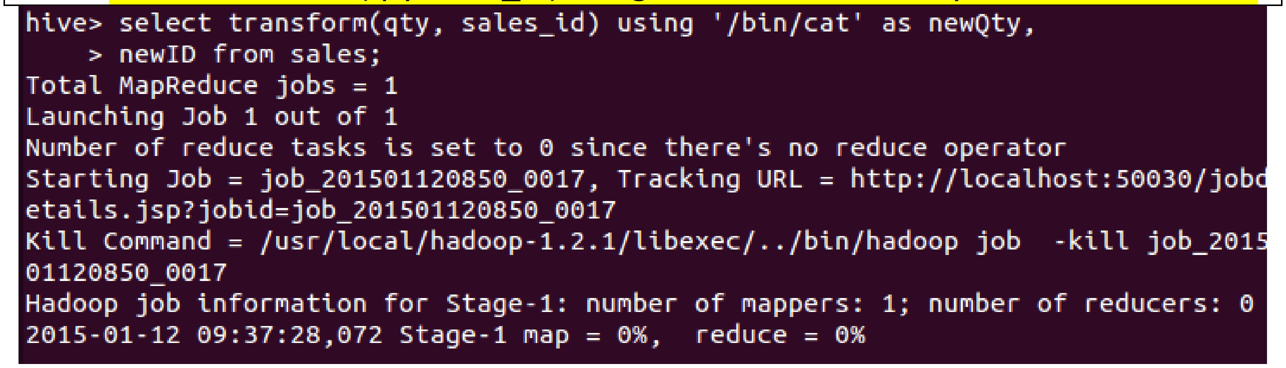


```
hive> select * from products where category = 'Ram';
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201501112153_0003, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501112153_0003
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01112153_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-01-11 22:11:19,230 Stage-1 map = 0%,  reduce = 0%
2015-01-11 22:11:41,977 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
c
2015-01-11 22:11:42,988 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
c
2015-01-11 22:11:44,138 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
```

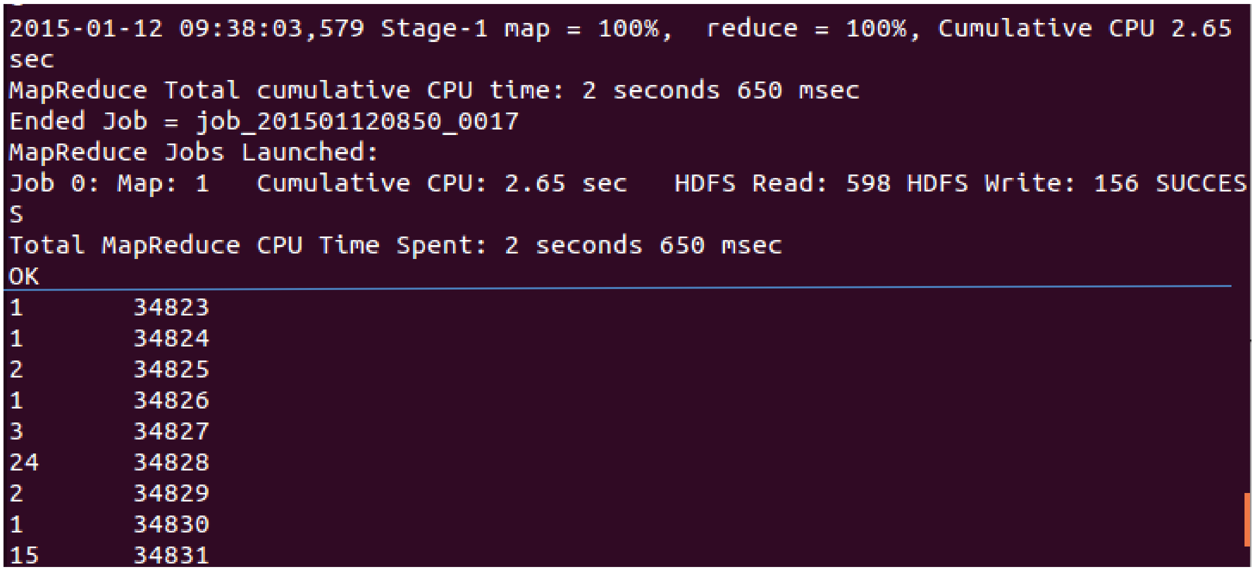


```
2015-01-11 22:11:50,315 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
c
2015-01-11 22:11:51,326 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
c
2015-01-11 22:11:52,332 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
c
2015-01-11 22:11:53,390 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
c
2015-01-11 22:11:54,442 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.74 se
c
2015-01-11 22:11:55,560 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.74
sec
MapReduce Total cumulative CPU time: 6 seconds 740 msec
Ended Job = job_201501112153_0003
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 6.74 sec   HDFS Read: 1954 HDFS Write: 175 SUCCE
SS
Total MapReduce CPU Time Spent: 6 seconds 740 msec
OK
2 GB Memory E   2 GB Memory ECC Ram     3000    87655   ["manual","heatsink"]
4 GB Memory E   4 GB Memory ECC Ram     1000    87659   ["manual","heatsink"]
16 GB Memory E  16 GB Memory ECC        Ram     238     87634   ["manual"]
Time taken: 73.801 seconds, Fetched: 3 row(s)
hive>
```

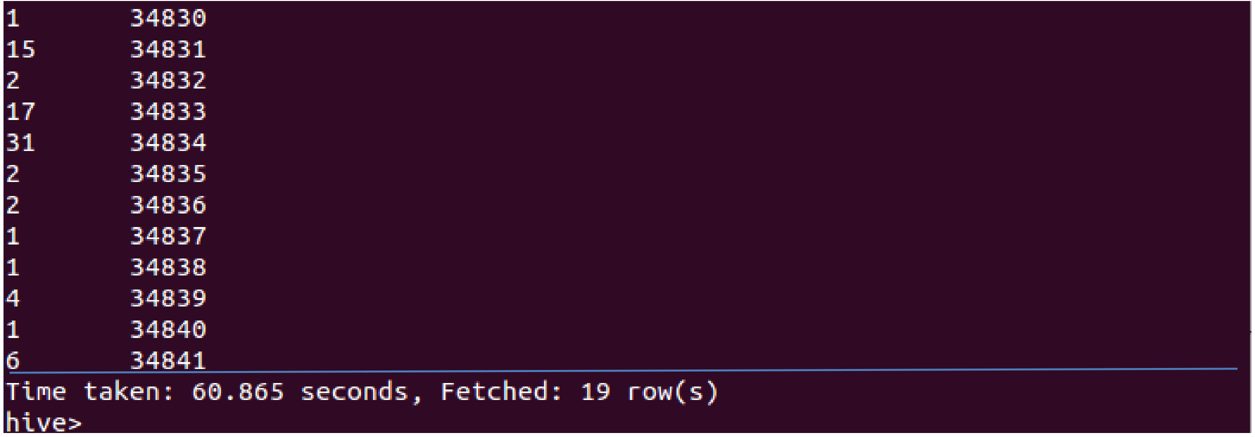
hive> select transform(qty,sales_id) using '/bin/cat' as newQty, newID from sales;



```
hive> select transform(qty, sales_id) using '/bin/cat' as newQty,
    > newID from sales;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201501120850_0017, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501120850_0017
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01120850_0017
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-01-12 09:37:28,072 Stage-1 map = 0%,  reduce = 0%
```

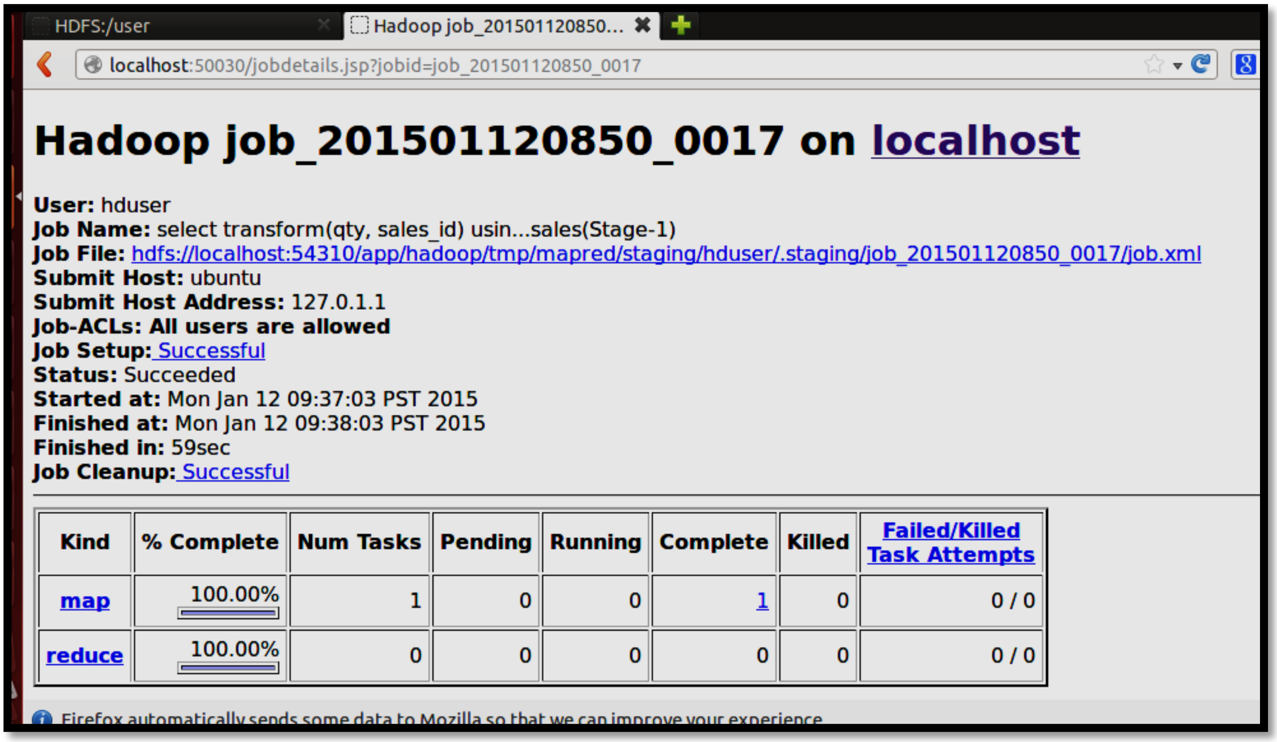


```
2015-01-12 09:38:03,579 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.65
sec
MapReduce Total cumulative CPU time: 2 seconds 650 msec
Ended Job = job_201501120850_0017
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 2.65 sec   HDFS Read: 598 HDFS Write: 156 SUCCES
S
Total MapReduce CPU Time Spent: 2 seconds 650 msec
OK
1       34823
1       34824
2       34825
1       34826
3       34827
24      34828
2       34829
1       34830
15      34831
```



```
1       34830
15      34831
2       34832
17      34833
31      34834
2       34835
2       34836
1       34837
1       34838
4       34839
1       34840
6       34841
Time taken: 60.865 seconds, Fetched: 19 row(s)
hive>
```

TO see the map-reduce jobs in GUI:


HDFS:/user
Hadoop job_201501120850...
localhost:50030/jobdetails.jsp?jobid=job_201501120850_0017

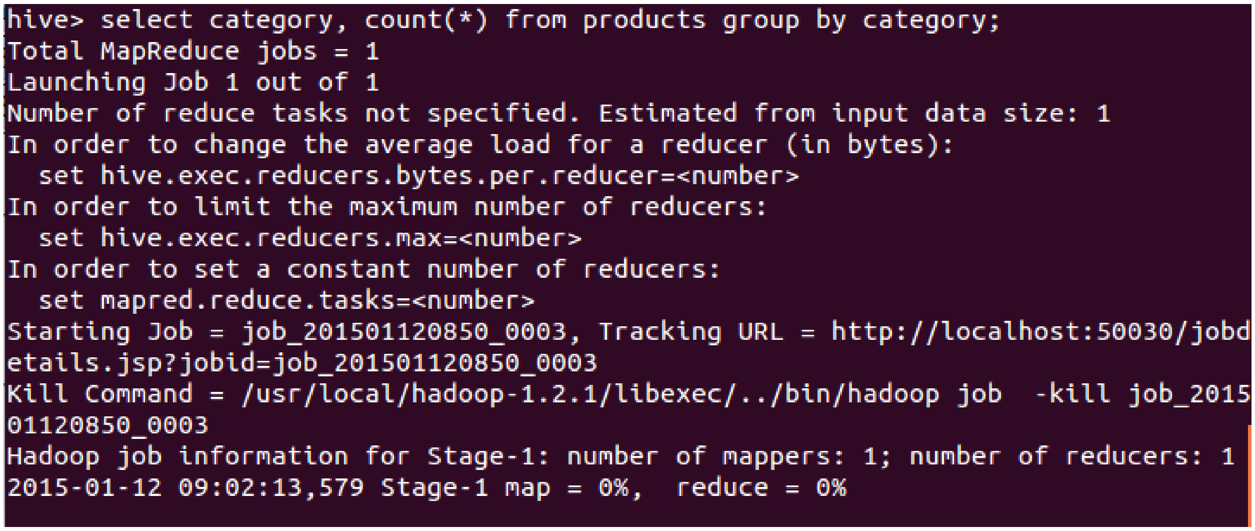
Hadoop job_201501120850_0017 on localhost

User: hduser
Job Name: select transform(qty, sales_id) usin...sales(Stage-1)
Job File: hdfs://localhost:54310/app/hadoop/tmp/mapred/staging/hduser/.staging/job_201501120850_0017/job.xml
Submit Host: ubuntu
Submit Host Address: 127.0.1.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started at: Mon Jan 12 09:37:03 PST 2015
Finished at: Mon Jan 12 09:38:03 PST 2015
Finished in: 59sec
Job Cleanup: Successful

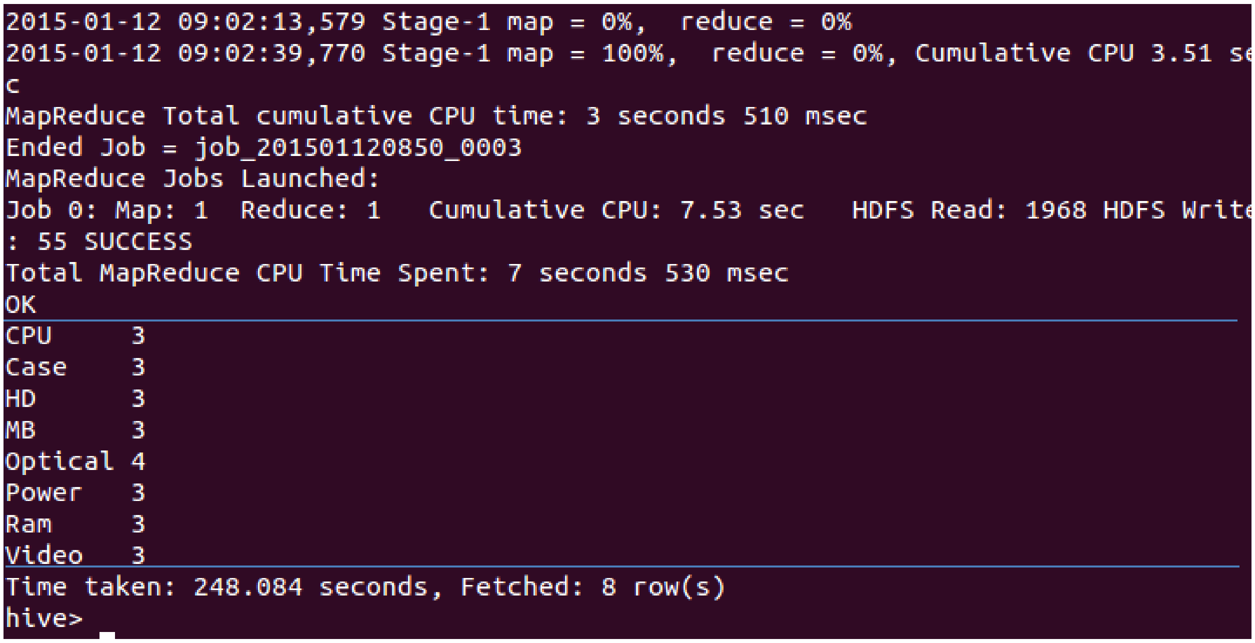
Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	1	0	0	1	0	0 / 0
reduce	100.00%	0	0	0	0	0	0 / 0

Firefox automatically sends some data to Mozilla so that we can improve your experience

>select category, count(*) from products group by category;



```
hive> select category, count(*) from products group by category;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201501120850_0003, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501120850_0003
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01120850_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2015-01-12 09:02:13,579 Stage-1 map = 0%,  reduce = 0%
```

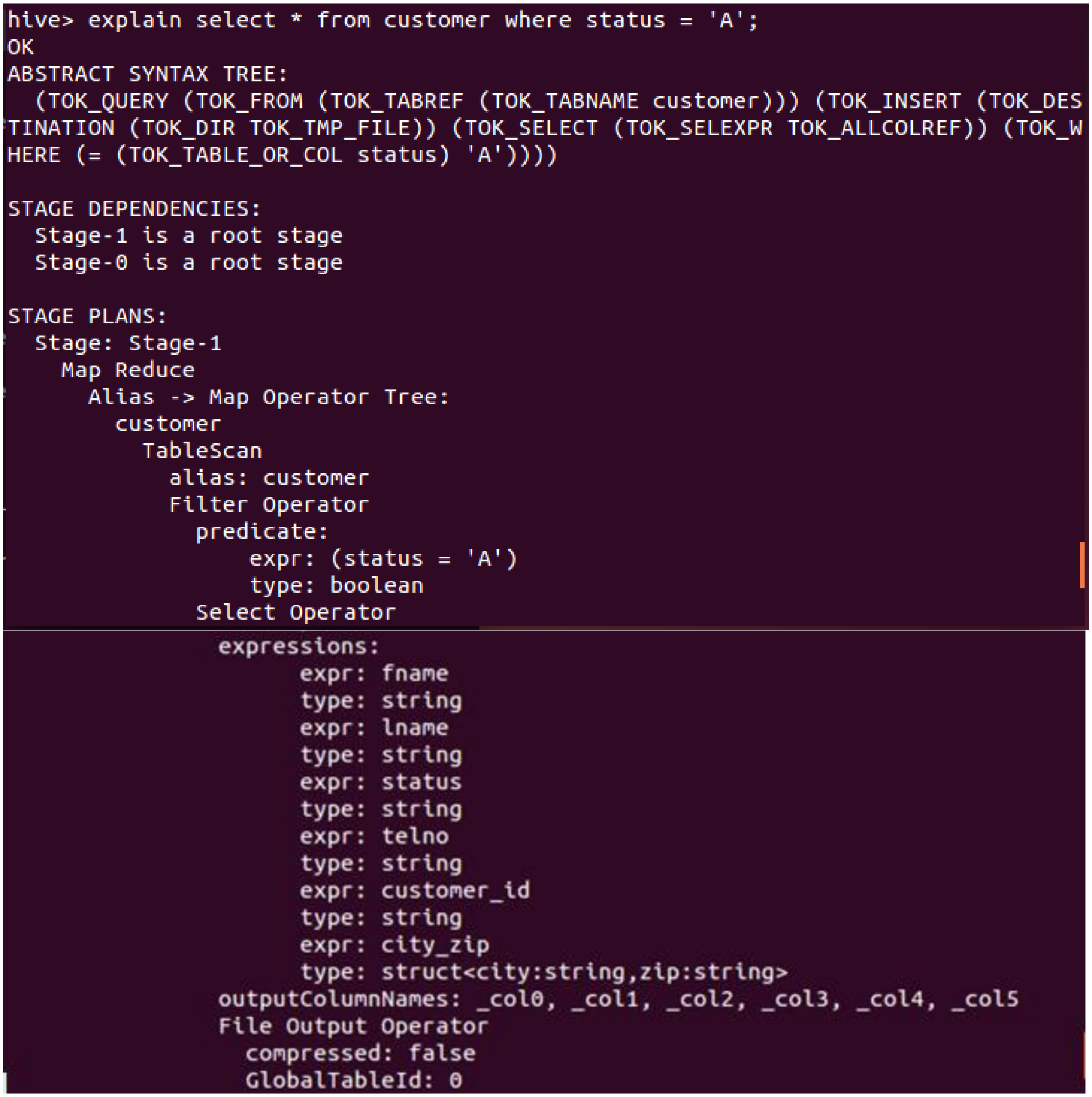


```
2015-01-12 09:02:13,579 Stage-1 map = 0%,  reduce = 0%
2015-01-12 09:02:39,770 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.51 se
c
MapReduce Total cumulative CPU time: 3 seconds 510 msec
Ended Job = job_201501120850_0003
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 7.53 sec   HDFS Read: 1968 HDFS Write
: 55 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 530 msec
OK
CPU     3
Case    3
HD      3
MB      3
Optical 4
Power   3
Ram     3
Video   3
Time taken: 248.084 seconds, Fetched: 8 row(s)
hive> 
```

To display the execution plan of the query based on the condition.

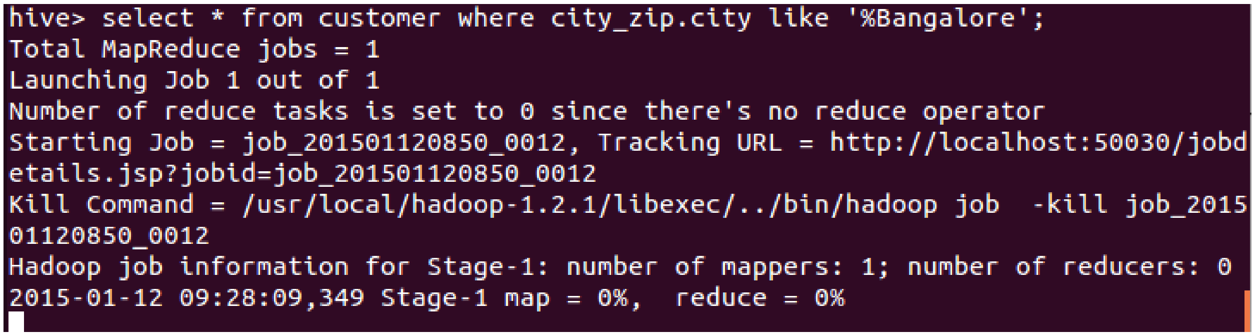
Ex: condition: **status='A'**

```
hive>explain select * from customer where status = 'A';
```

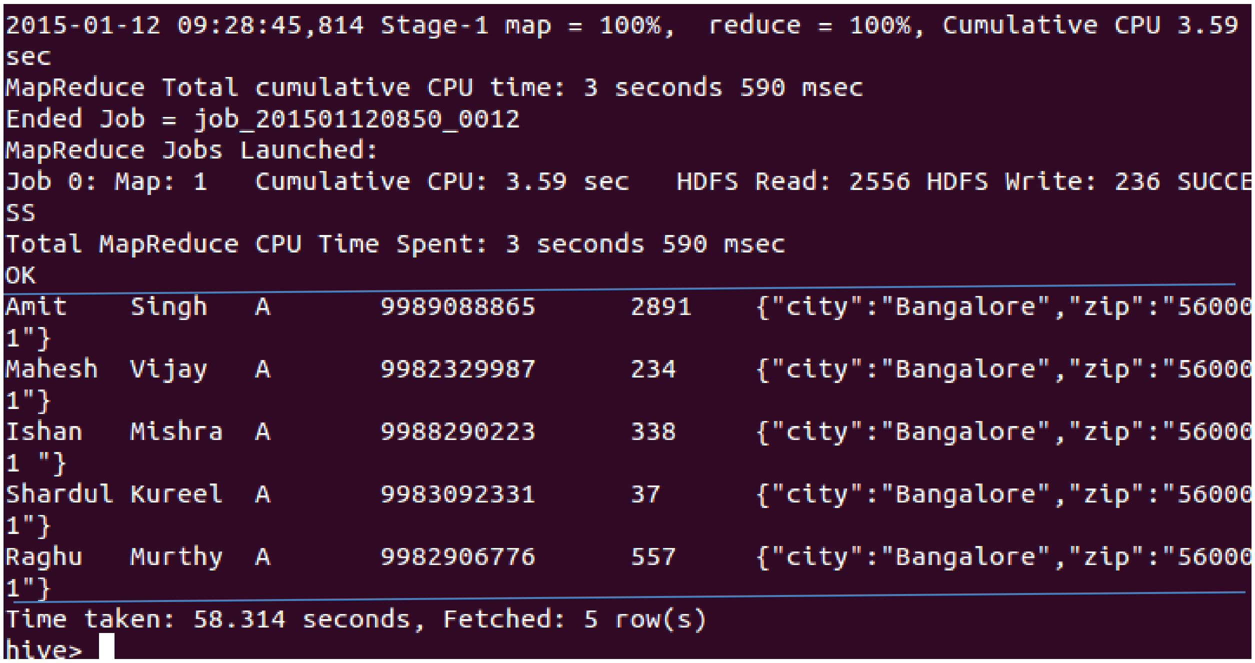


```
hive> explain select * from customer where status = 'A';
OK
ABSTRACT SYNTAX TREE:
  (TOK_QUERY (TOK_FROM (TOK_TABREF (TOK_TABNAME customer))) (TOK_INSERT (TOK_DES
TINATION (TOK_DIR TOK_TMP_FILE)) (TOK_SELECT (TOK_SELEXPR TOK_ALLCOLREF)) (TOK_W
HERE (= (TOK_TABLE_OR_COL status) 'A'))))

STAGE DEPENDENCIES:
  Stage-1 is a root stage
  Stage-0 is a root stage

STAGE PLANS:
  Stage: Stage-1
    Map Reduce
      Alias -> Map Operator Tree:
        customer
          TableScan
            alias: customer
            Filter Operator
              predicate:
                  expr: (status = 'A')
                  type: boolean
              Select Operator
                expressions:
                      expr: fname
                      type: string
                      expr: lname
                      type: string
                      expr: status
                      type: string
                      expr: telno
                      type: string
                      expr: customer_id
                      type: string
                      expr: city_zip
                      type: struct<city:string,zip:string>
                outputColumnNames: _col0, _col1, _col2, _col3, _col4, _col5
                File Output Operator
                  compressed: false
                  GlobalTableId: 0
```

hive> select * from customer where city_zip.city like '%Bangalore';

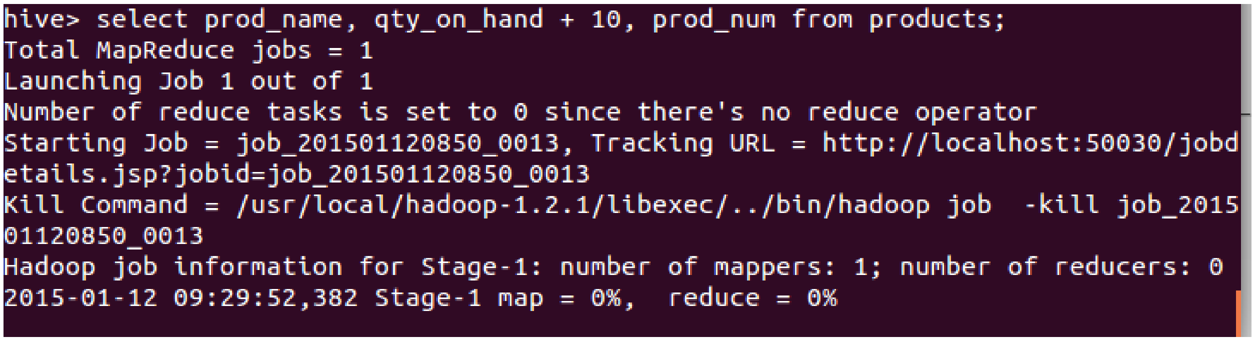


```
hive> select * from customer where city_zip.city like '%Bangalore';
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201501120850_0012, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501120850_0012
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01120850_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-01-12 09:28:09,349 Stage-1 map = 0%,  reduce = 0%
```

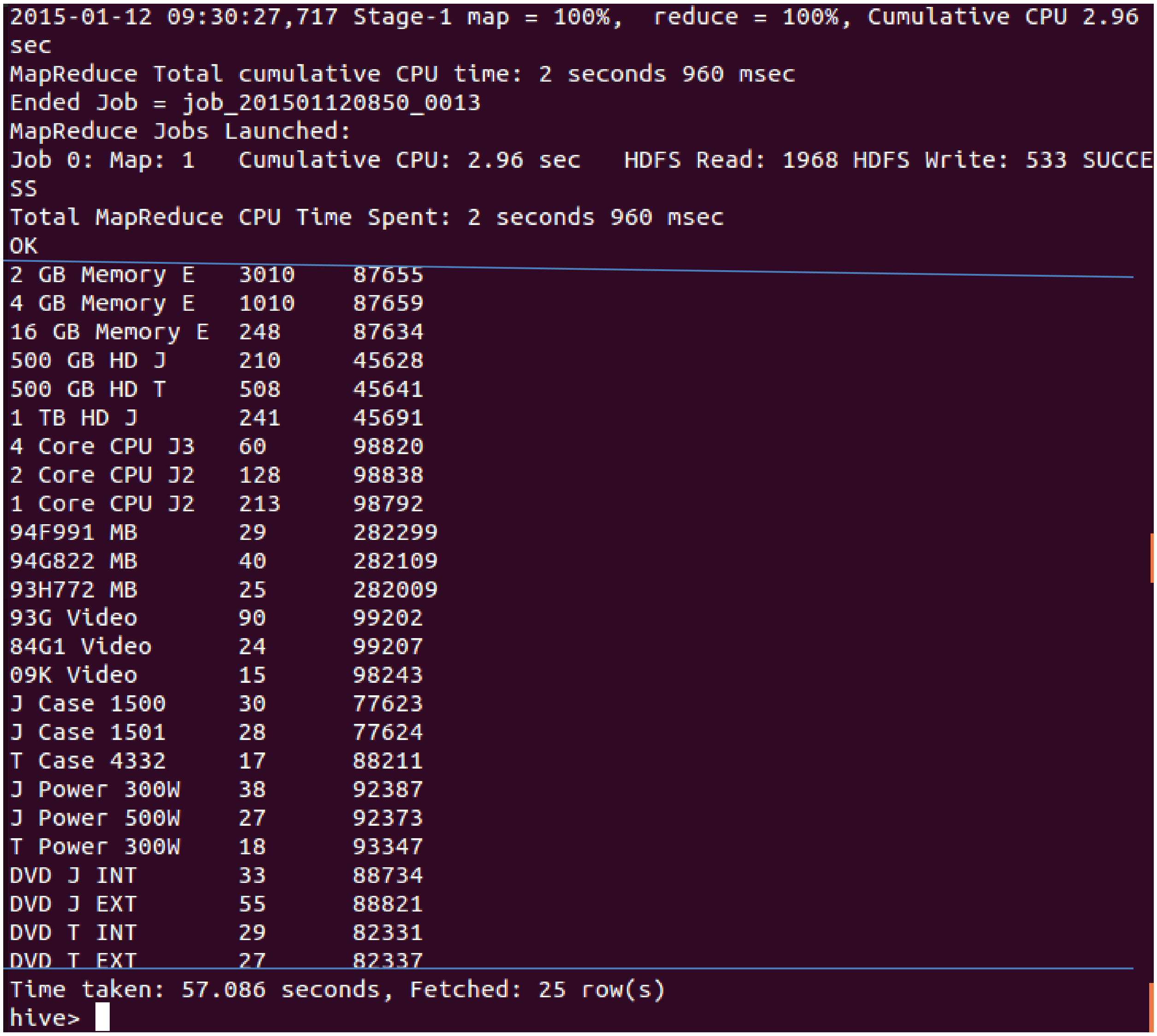


```
2015-01-12 09:28:45,814 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.59
sec
MapReduce Total cumulative CPU time: 3 seconds 590 msec
Ended Job = job_201501120850_0012
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 3.59 sec   HDFS Read: 2556 HDFS Write: 236 SUCCE
SS
Total MapReduce CPU Time Spent: 3 seconds 590 msec
OK
Amit    Singh   A       9989088865      2891    {"city":"Bangalore","zip":"56000
1"}
Mahesh  Vijay   A       9982329987      234     {"city":"Bangalore","zip":"56000
1"}
Ishan   Mishra  A       9988290223      338     {"city":"Bangalore","zip":"56000
1 "}
Shardul Kureel  A       9983092331      37      {"city":"Bangalore","zip":"56000
1"}
Raghu   Murthy  A       9982906776      557     {"city":"Bangalore","zip":"56000
1"}
Time taken: 58.314 seconds, Fetched: 5 row(s)
hive>
```

hive> select prod_name, qty_on_hand + 10, prod_num from products;

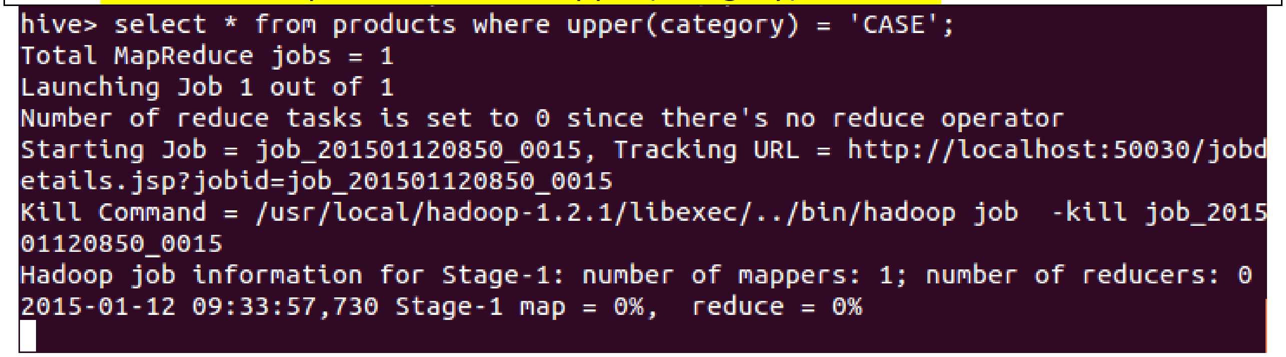


```
hive> select prod_name, qty_on_hand + 10, prod_num from products;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201501120850_0013, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501120850_0013
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01120850_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-01-12 09:29:52,382 Stage-1 map = 0%,  reduce = 0%
```

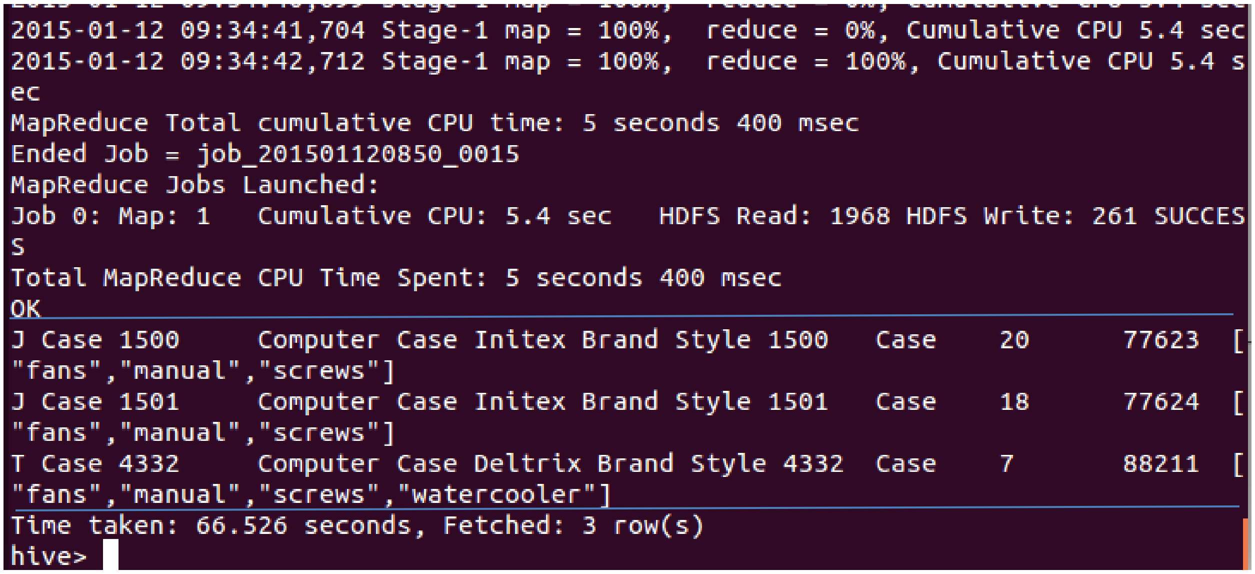


```
2015-01-12 09:30:27,717 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.96
sec
MapReduce Total cumulative CPU time: 2 seconds 960 msec
Ended Job = job_201501120850_0013
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 2.96 sec   HDFS Read: 1968 HDFS Write: 533 SUCCE
SS
Total MapReduce CPU Time Spent: 2 seconds 960 msec
OK
2 GB Memory E   3010    87655
4 GB Memory E   1010    87659
16 GB Memory E  248     87634
500 GB HD J     210     45628
500 GB HD T     508     45641
1 TB HD J       241     45691
4 Core CPU J3   60      98820
2 Core CPU J2   128     98838
1 Core CPU J2   213     98792
94F991 MB       29      282299
94G822 MB       40      282109
93H772 MB       25      282009
93G Video       90      99202
84G1 Video      24      99207
09K Video       15      98243
J Case 1500     30      77623
J Case 1501     28      77624
T Case 4332     17      88211
J Power 300W    38      92387
J Power 500W    27      92373
T Power 300W    18      93347
DVD J INT       33      88734
DVD J EXT       55      88821
DVD T INT       29      82331
DVD T EXT       27      82337
Time taken: 57.086 seconds, Fetched: 25 row(s)
hive>
```

hive>select * from products where upper(category) = ‘CASE’;

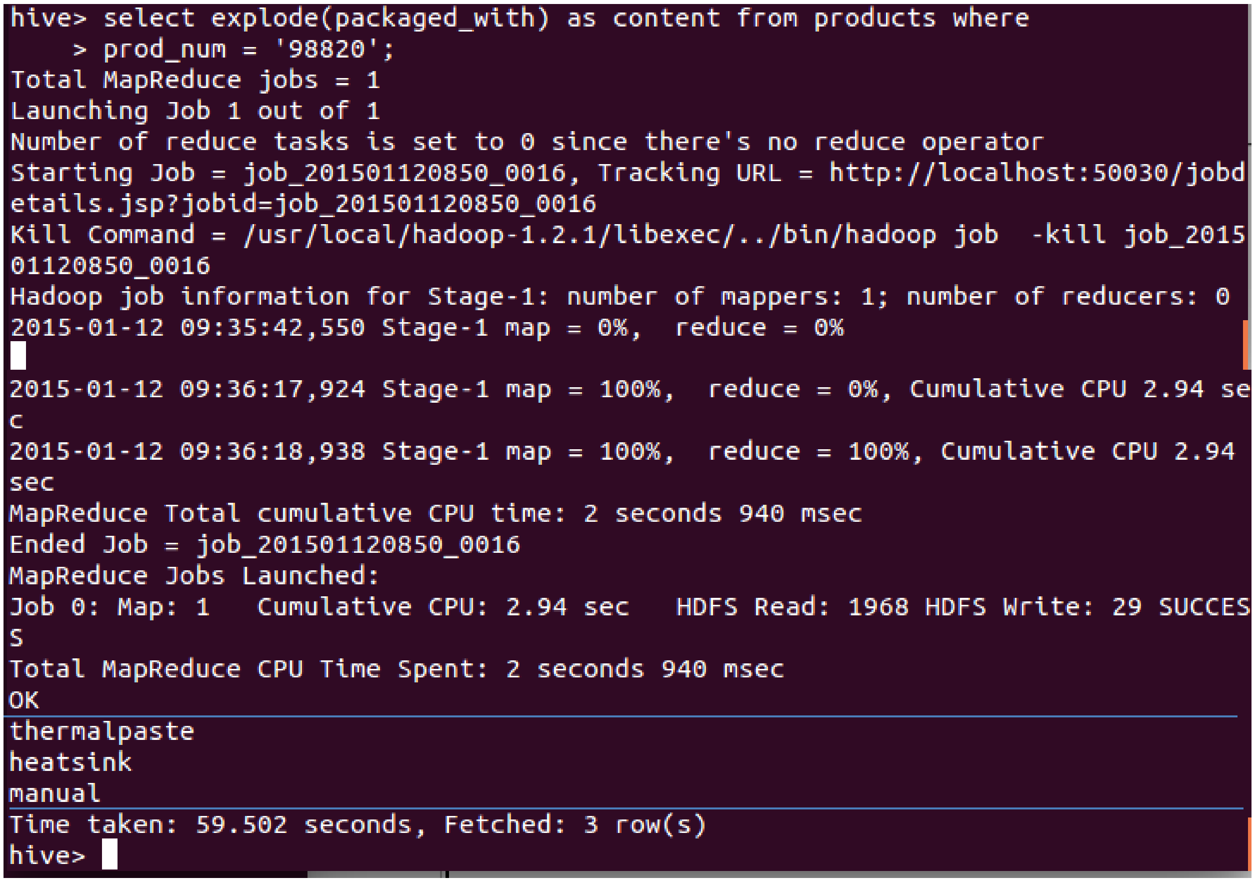


```
hive> select * from products where upper(category) = 'CASE';
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201501120850_0015, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501120850_0015
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01120850_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-01-12 09:33:57,730 Stage-1 map = 0%,  reduce = 0%
```



```
2015-01-12 09:34:41,704 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.4 sec
2015-01-12 09:34:42,712 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.4 s
ec
MapReduce Total cumulative CPU time: 5 seconds 400 msec
Ended Job = job_201501120850_0015
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 5.4 sec   HDFS Read: 1968 HDFS Write: 261 SUCCES
S
Total MapReduce CPU Time Spent: 5 seconds 400 msec
OK
J Case 1500     Computer Case Initex Brand Style 1500   Case    20      77623   [
"fans","manual","screws"]
J Case 1501     Computer Case Initex Brand Style 1501   Case    18      77624   [
"fans","manual","screws"]
T Case 4332     Computer Case Deltrix Brand Style 4332  Case    7       88211   [
"fans","manual","screws","watercooler"]
Time taken: 66.526 seconds, Fetched: 3 row(s)
hive>
```

```
hive>select explode(packaged_with) as content from products where
prod_num='98820';
```



```
hive> select explode(packaged_with) as content from products where
    > prod_num = '98820';
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201501120850_0016, Tracking URL = http://localhost:50030/jobd
etails.jsp?jobid=job_201501120850_0016
Kill Command = /usr/local/hadoop-1.2.1/libexec/../bin/hadoop job  -kill job_2015
01120850_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2015-01-12 09:35:42,550 Stage-1 map = 0%,  reduce = 0%

2015-01-12 09:36:17,924 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.94 se
c
2015-01-12 09:36:18,938 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.94
sec
MapReduce Total cumulative CPU time: 2 seconds 940 msec
Ended Job = job_201501120850_0016
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 2.94 sec   HDFS Read: 1968 HDFS Write: 29 SUCCES
S
Total MapReduce CPU Time Spent: 2 seconds 940 msec
OK
thermalpaste
heatsink
manual
Time taken: 59.502 seconds, Fetched: 3 row(s)
hive>
```

You can also define the different types of queries, if you are familier with the SQL.

For Hive more queries, please refer “**lab_query by Nagarjuna”** notepad.

Now we are done ☺

For further queries, mail us:

1. Mr. Nagarjuna D N
   nagarjunadn.arjun@gmail.com
2. world.of.bigdata.community@gmail.com