**Hadoop Stock Exchange Case study**

**Hive Hands-on**

a. Create database NYSE in hive

*create database NYSE;*

****

b. Create a table called nyse\_data file with following colummns

Exchang string

Ticker string

Date Date

Open Decimal (10,2)

High Decimal (10,2)

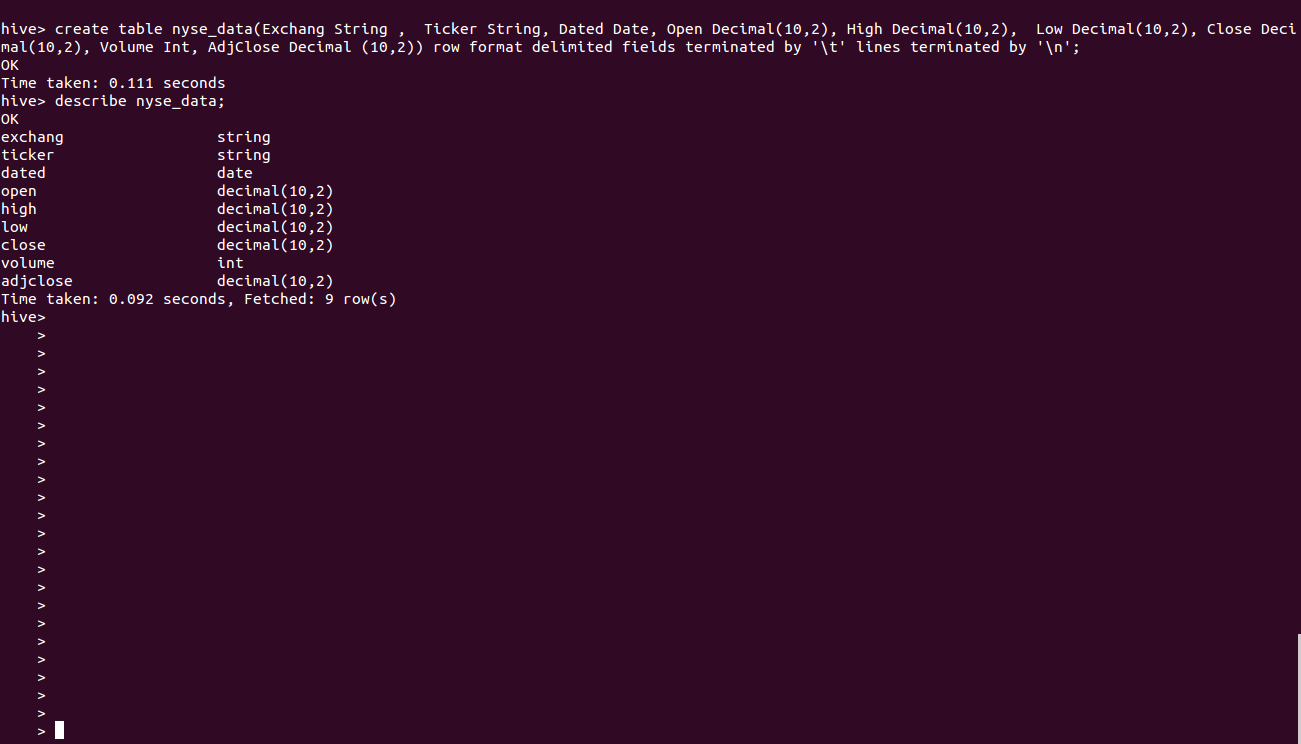
Low Decimal (10,2)

Close Decimal (10,2)

Volume Int

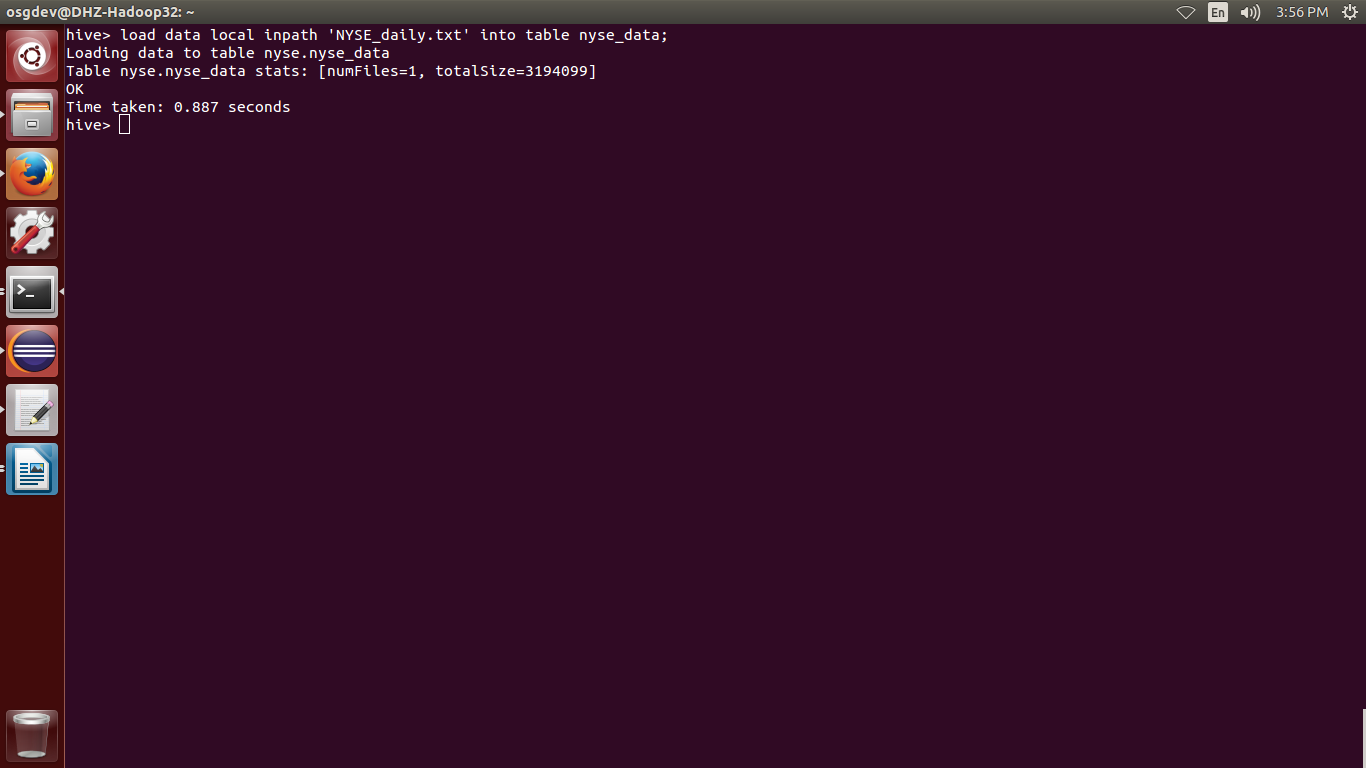
AdjClose Decimal (10,2)

*create table nyse\_data(Exchang String , Ticker String, Dated Date, Open Decimal(10,2), High Decimal(10,2), Low Decimal(10,2), Close Decimal(10,2), Volume Int, AdjClose Decimal (10,2)) row format delimited fields terminated by '\t' lines terminated by '\n';*



c. Load data from NYSE\_daily.txt into table

*load data local inpath 'NYSE\_daily.txt' into table nyse\_data;*



d. Create a table called nyse\_by\_month partitioned by year, month having following columns

Ticker string

Date Date

Open Decimal (10,2)

High Decimal (10,2)

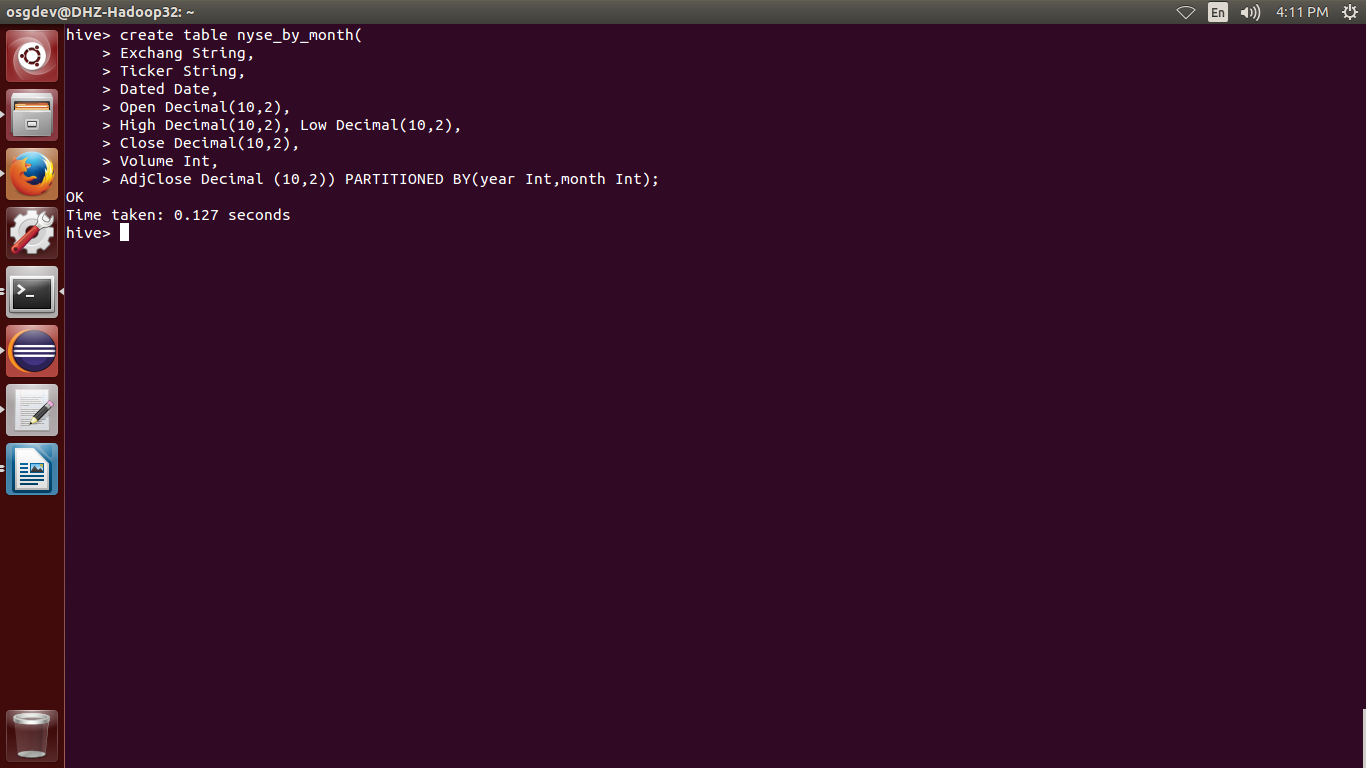
Low Decimal (10,2)

Close Decimal (10,2)

Volume Int

AdjClose Decimal (10,2)

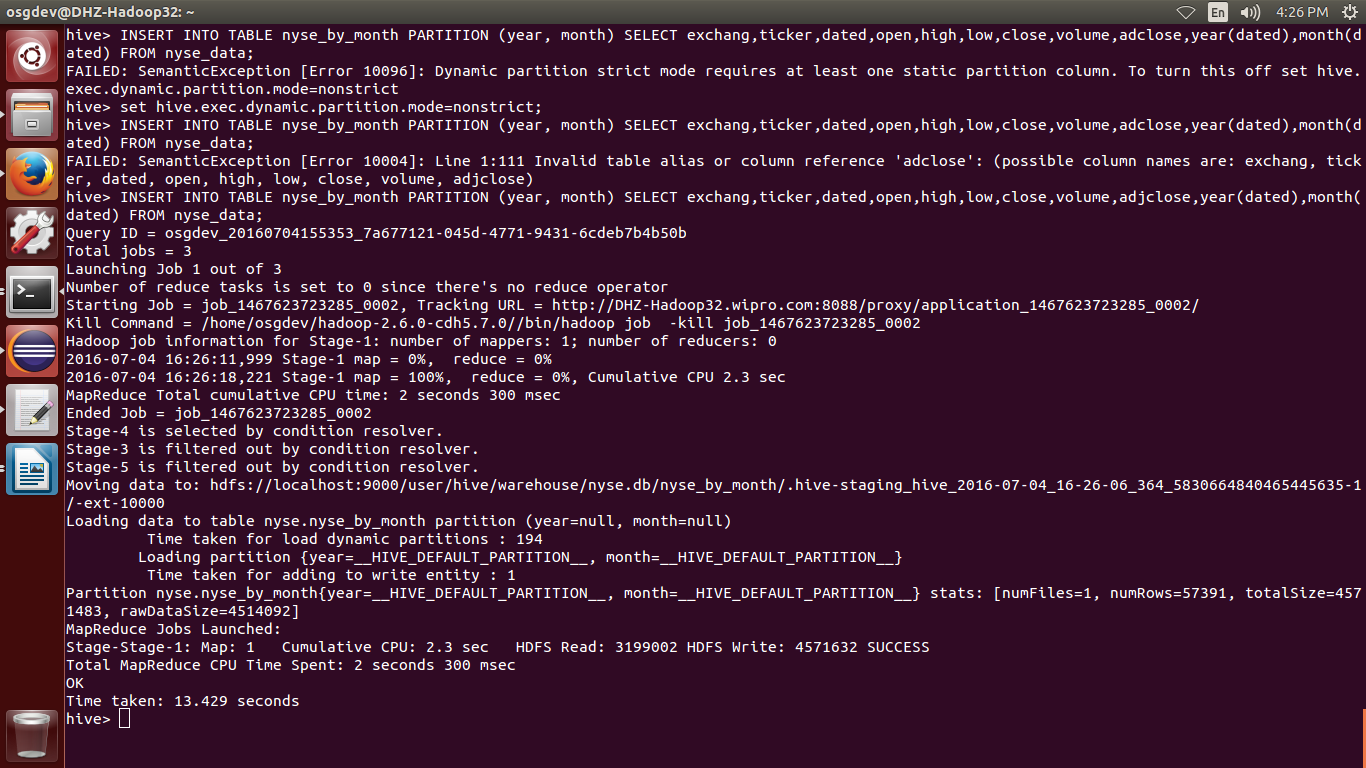
*create table* nyse\_by\_month*(Exchang String,Ticker String,Dated Date,Open Decimal(10,2),High Decimal(10,2), Low Decimal(10,2),Close Decimal(10,2),Volume Int,AdjClose Decimal (10,2))* PARTITIONED BY(year Int,month Int)*;*



e. Load data from NYSE\_daily into nyse\_by\_month table

*set hive.exec.dynamic.partition.mode=nonstrict;*

*INSERT INTO TABLE nyse\_by\_month PARTITION (year, month) SELECT exchang,ticker,dated,open,high,low,close,volume,adjclose,year(dated),month(dated) FROM nyse\_data;*



f. Write a hive sql query to find Minimum and Maximum AdjClose for each stock ticker in each month

*select ticker, month(dated),min(adjclose),max(adjclose) from nyse\_data group by ticker, month(dated);*

