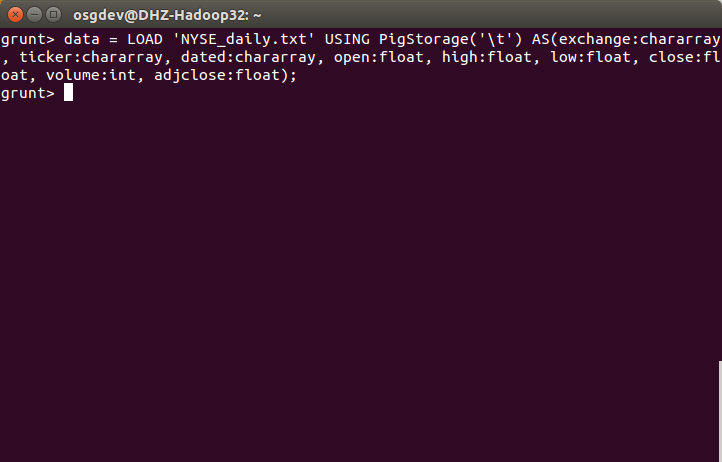
**Hadoop Stock Exchange Case study**

**Pig Hands-on**

1] Create a pig script to do the following:

a. Load the NYSE\_daily.txt file

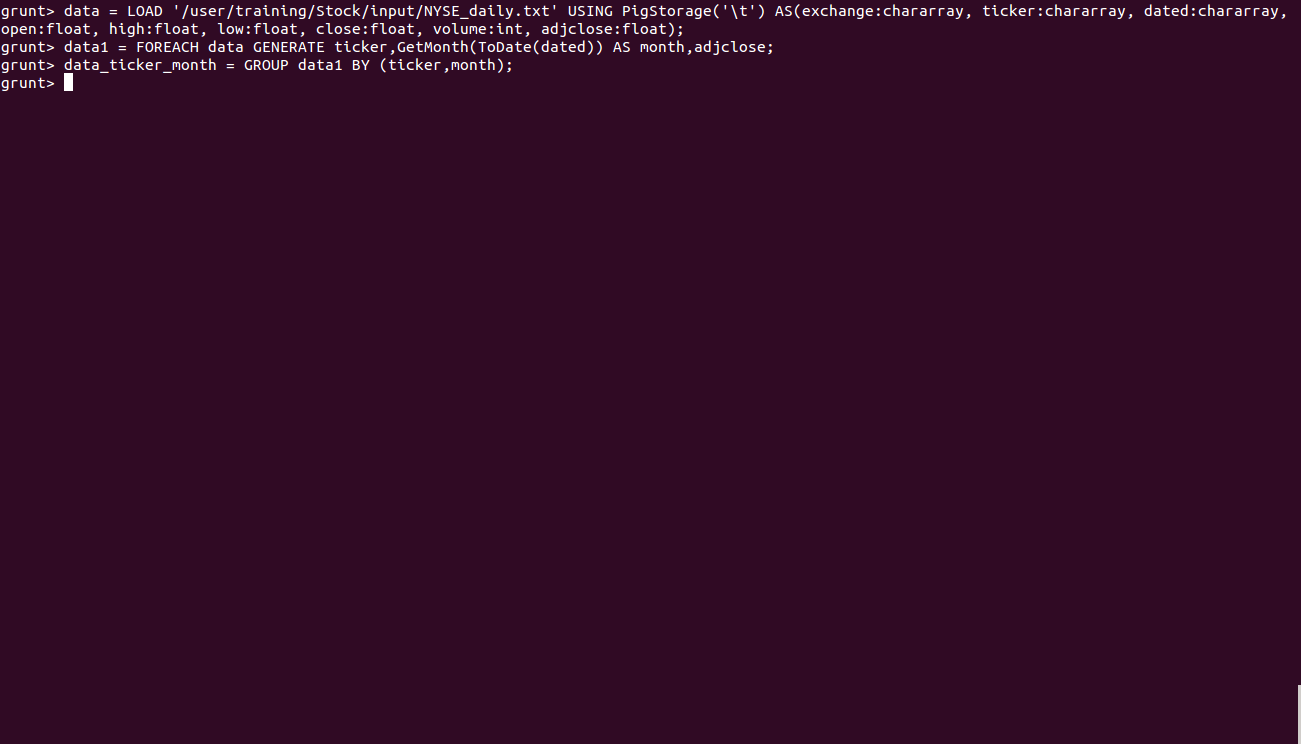
*data = LOAD '/user/training/Stock/input/NYSE\_daily.txt' USING PigStorage('\t') AS(exchange:chararray, ticker:chararray, dated:chararray, open:float, high:float, low:float, close:float, volume:int, adjclose:float);*



b. Group records by Ticker, Month

*data1 = FOREACH data GENERATE ticker,GetMonth(ToDate(dated)) AS month,adjclose;*

*data\_ticker\_month = GROUP data1 BY (ticker,month);*



c. Find Minimum and Maximum AdjClose for each stock ticker in each month

*ans = FOREACH data\_ticker\_month GENERATE FLATTEN(group), MIN($1.$2),MAX($1.$2);*

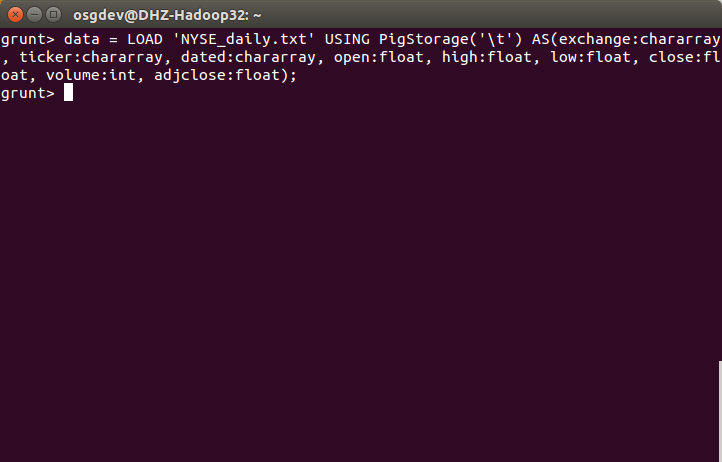
*DUMP ans;*



2] Create a pig script to do the following:

a. Load the NYSE\_daily.txt file

*data = LOAD '/user/training/Stock/input/NYSE\_daily.txt' USING PigStorage('\t') AS(exchange:chararray, ticker:chararray, dated:chararray, open:float, high:float, low:float, close:float, volume:int, adjclose:float);*

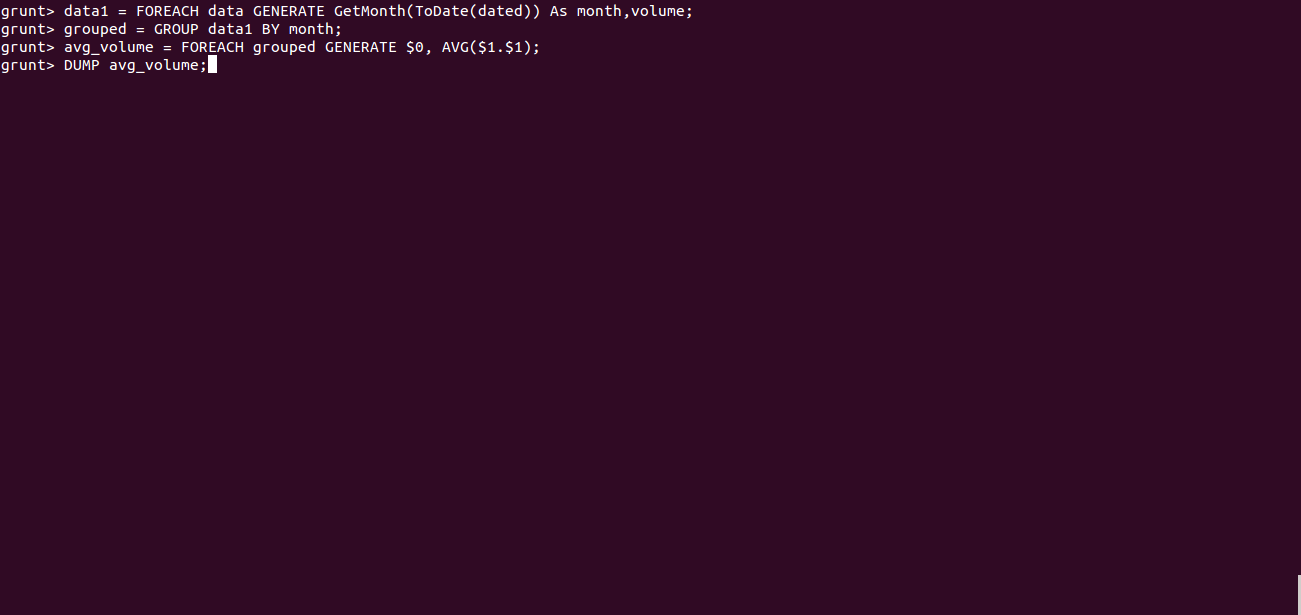


b. Find the average volume for each month

*data1 = FOREACH data GENERATE GetMonth(ToDate(dated)) As month,volume;*

*grouped = GROUP data1 BY month;*

*avg\_volume = FOREACH grouped GENERATE FLATTEN(group),AVG($1.$1);*

*DUMP avg\_volume;*

*c. Extract the following columns*

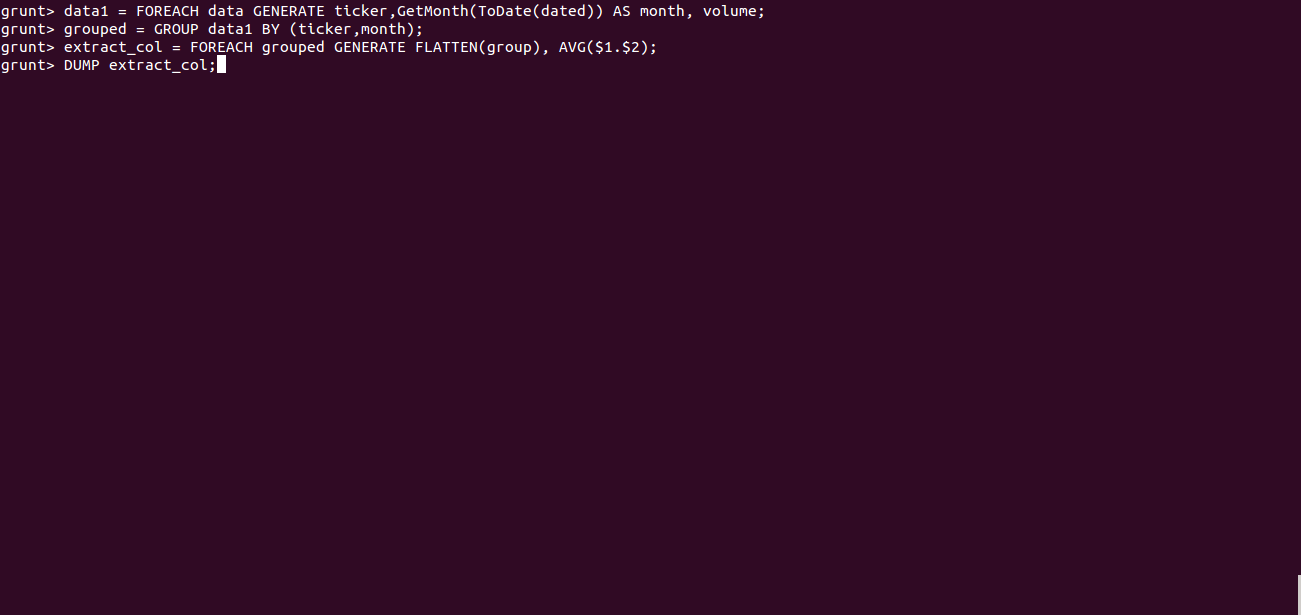
*Ticker, Month, AvgVolume*

*data1 = FOREACH data GENERATE ticker,GetMonth(ToDate(dated)) AS month, volume;*

*grouped = GROUP data1 BY (ticker,month);*

*extract\_col = FOREACH grouped GENERATE FLATTEN(group), AVG($1.$2);*

*DUMP extract\_col;*

**

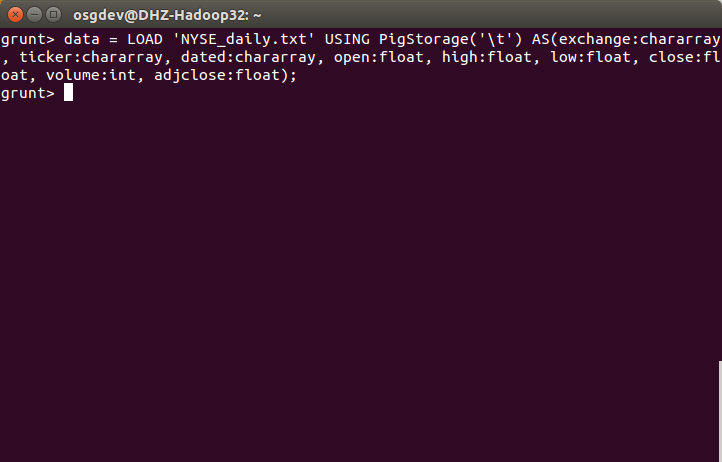
d. Save the output to /user/training/stock/AvgVolume

*STORE extract\_col INTO '/user/training/stock/AvgVolume'*

3] Create a pig script to do the following:

a. Load NYSE\_daily.txt file

*data = LOAD '/user/training/Stock/input/NYSE\_daily.txt' USING PigStorage('\t') AS(exchange:chararray, ticker:chararray, dated:chararray, open:float, high:float, low:float, close:float, volume:int, adjclose:float);*

**

*b. Load AvgVolume data*

*avgvol = GROUP data ALL;*

*avg\_volume = FOREACH avgvol GENERATE AVG(data.volume) as avgvolume;*

*c. Filter out records from NYSE\_daily.txt where Volume >= AvgVolume*

*avg = FILTER data BY (volume >=avg\_volume.avgvolume);*

*d. Save the records for the above condition to* */user/training/stock/HighVolume*

*STORE avg INTO '/user/training/stock/HighVolume';*