**Objective:** To Analyse New York Stock Exchange data

nyse.7z present along with this document contains 2 kind of file

1. NASDAQ\_daily\_prices\*.csv [ A to Z ]
2. NASDAQ\_dividends\*.csv [ A to Z ]

There is also 2 specific file NASDAQ\_daily\_prices0.csv,NASDAQ\_dividends0.csv,that just has column name stored in it with no data.

**NASDAQ\_daily\_prices table**

|  |  |
| --- | --- |
| **Column Name** | **Datatype** |
| Exchange | String |
| stock\_symbol | String |
| Tdate | String |
| stock\_price\_open | Float |
| stock\_price\_high | Float |
| stock\_price\_low | Float |
| stock\_price\_close | Float |
| stock\_volume | Int |
| stock\_price\_adj\_close | Float |

1. **Lets create a database called financials & Use it**

hive>create database financials;

hive>use financials;

1. **Create Table to store data**

**CREATE** **TABLE** NYSE\_DAILY\_PRICES\_HIVE

(

 STOCK\_EXCHANGE STRING,

 STOCK\_SYMBOL STRING,

 TDATE **STRING**,

 STOCK\_PRICE\_OPEN **DOUBLE**,

 STOCK\_PRICE\_HIGH **DOUBLE**,

 STOCK\_PRICE\_LOW **DOUBLE**,

 STOCK\_PRICE\_CLOSE **DOUBLE**,

 STOCK\_VOLUME **BIGINT**,

STOCK\_PRICE\_ADJ\_CLOSE **DOUBLE**

)ROW FORMAT DELIMITED FIELDS TERMINATED **BY** ',';

1. **Load data from local to hive table**

hive>**load** data **local** inpath '/home/hduser/nyse/NASDAQ\_daily\_prices\_\*.csv'  overwrite **into** **table** NYSE\_DAILY\_PRICES\_HIVE;

1. **Create Another table called dividends**

**CREATE**  **TABLE** dividends

(

STOCK\_EXCHANGE STRING,

STOCK\_SYMBOL STRING,

ymd STRING,

dividend **FLOAT**

)

ROW FORMAT DELIMITED FIELDS TERMINATED **BY** ',';

1. **Load data from local to hive table**

**load** data **local** inpath '/home/hduser/nyse/NASDAQ\_dividends\_\*.csv'

overwrite **into** **table** DIVIDENDS;

**Lets Perform some Operation**

1. **Find out all the Unique stock symbol**

hive>select distinct(stock\_symbol) from NYSE\_DAILY\_PRICES\_HIVE;

1. **Find out Top 5 stock\_symbol with highest stock volume**

hive>select stock\_volume from NYSE\_DAILY\_PRICES\_HIVE order by stock\_volume desc limit 5;

stock\_symbol stock\_volume

MSFT 1031788800

ORCL 1030988400

DELL 842560000

ORCL 838823800

CSCO 806732800

1. **Find out the date in which any stock symbol has got highest value and lowest value**

hive> select distinct(b.tdate) as date from (select tdate from nyse\_daily\_prices\_hive where stock\_price\_high in (select max(stock\_price\_high) from nyse\_daily\_prices\_hive)) a join (select tdate from nyse\_daily\_prices\_hive where stock\_price\_low in(select min(stock\_price\_low) from nyse\_daily\_prices\_hive)) b on a.tdate = b.tdate;

**date**

**14-07-2000**

**17-07-2000**

**18-07-2000**

1. **Find out the highest stock volume for each stock symbol**

hive> select stock\_symbol,max(stock\_volume) from nyse\_daily\_prices\_hive group by stock\_symbol

1. **Find out the highest and lowest volumn of 'DELL' till now.**

hive> select stock\_symbol,max(stock\_volume) as Highestvolume,min(stock\_volume) as LowestVolume from nyse\_daily\_prices\_hive where stock\_symbol='DELL' group by stock\_symbol;

**stock\_symbol highestvolume lowestvolume**

**DELL 842560000 556800**