

Stock data Analysis using Hive.

This class assignment is for the NYSE stock Data set which is provided as part of this assignment.

We have taken this NYSE dataset Daily stock data of each company from yahoo finance website. The Yahoo finance website contains stock data for each stock exchange worldwide.

We have taken the NYSE stock exchange data for this study. The data set is composed of: **stock exchange,company symbol, date, open price of the day, high of the day, low of the day, close of the day , volume and adjusted close price.**

Please complete following tasks based on this dataset.

1. Create NYSE\_Partition table based on the date field

hive> create table NYSE(stock STRING,company STRING,`timestamp` DATETIME,open Decimal(4,2),high Decimal(4,2),low Decimal(4,2),close Decimal(4,2),volume INT) partitioned by (`timestamp`) row format delimited fields terminated by '\t' stored as textfile;

1. This table needs to store the data as ORC file
2. hive> create table NYSE(stock STRING,company STRING,`timestamp` DATETIME,open Decimal(4,2),high Decimal(4,2),low Decimal(4,2),close Decimal(4,2),volume INT) partitioned by (`timestamp`) row format delimited fields terminated by '\t' stored as orc;
3. In this partition table, Load data of only those records where open price of the day is greater than 68 and high close price is less than 70.

Select \* from NYSE where open>68 and high<70;

Please note: You need to copy NYSE file “NYSE\_daily\_File” in the EMR cluster using WinSCP as described in 1st session for copying the files in EMR.