Hadoop Hive Lab:

# Assignment:

Complete following tasks based on the data set mentioned above:

1. Create NYSE\_Partition table based on the **company symbol** field.
2. This table needs to store the data as ORC file.
3. Load data of only those records where open price of the day is greater than 68 and high price of the day is less than 70.

# Solution:

**Step 1:** *Creating new database in hive*

create database may29c;

**Step 2:** *Using existing database*

use may29c;

**Step 3:** *Creation of table without partition*

CREATE TABLE NYSE\_WP(

stock\_exchange STRING,

company\_symbol STRING,

idate STRING,

open\_price\_of\_the\_day FLOAT,

high\_of\_the\_day FLOAT,

low\_of\_the\_day FLOAT,

close\_of\_the\_day FLOAT,

volume BIGINT,

adjusted\_close\_price FLOAT

)

ROW FORMAT delimited fields terminated BY ','

STORED AS textfile;

**Step 4:** *Creation of Partition table based on company symbol field*

CREATE TABLE NYSE\_PARTITION(

stock\_exchange STRING,

idate STRING,

open\_price\_of\_the\_day FLOAT,

high\_of\_the\_day FLOAT,

low\_of\_the\_day FLOAT,

close\_of\_the\_day FLOAT,

volume BIGINT,

adjusted\_close\_price FLOAT

)

PARTITIONED BY(company\_symbol STRING)

ROW FORMAT delimited fields terminated BY ','

STORED AS orc;

**Step 5:** *Loading data in to table without partition*

load data local inpath '/mnt/home/arun.kpselvam\_gmail/NYSE\_daily\_File.csv' into table NYSE\_WP;

**Step 6:** *Checking if the data got loaded correctly*

select \* from NYSE\_WP where open\_price\_of\_the\_day > 68.0 and high\_of\_the\_day < 70.0 limit 5;

**Step 7:** *Enabling Dynamic Partition since it is disabled by default*

set hive.exec.dynamic.partition=true;

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

set hive.exec.max.dynamic.partitions=100;

set hive.exec.max.dynamic.partitions.pernode=100;

**Step 8:** *Insertion of data in to partition table based on conditions*

insert into table NYSE\_PARTITION partition(company\_symbol) select stock\_exchange,idate,open\_price\_of\_the\_day,high\_of\_the\_day,low\_of\_the\_day,close\_of\_the\_day,volume,adjusted\_close\_price,company\_symbol from NYSE\_WP where open\_price\_of\_the\_day > 68.0 and high\_of\_the\_day < 70.0;

**Step 9:** *Query to test the partition*

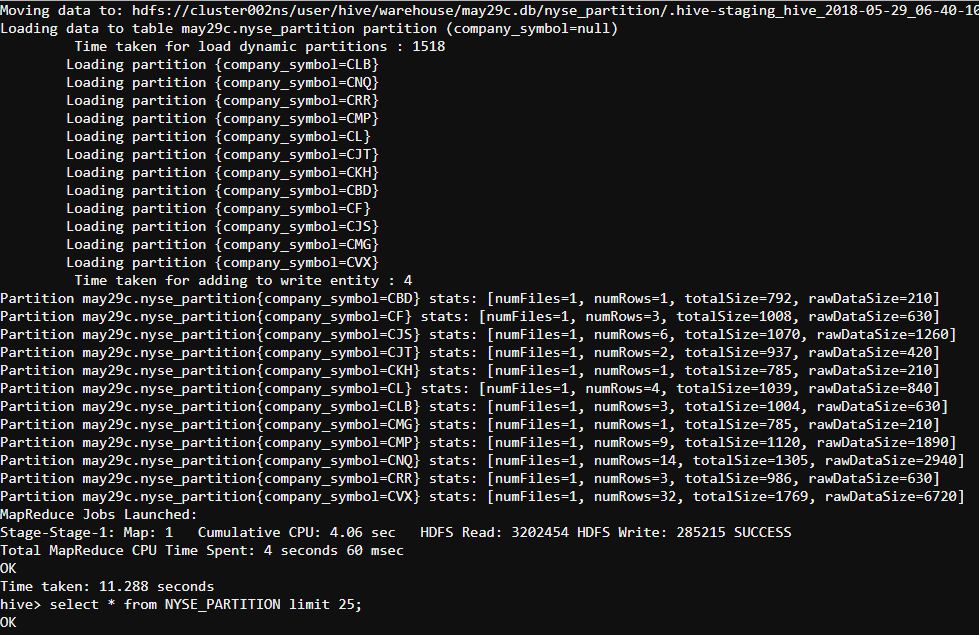
select \* from NYSE\_PARTITION limit 5;

# Results:

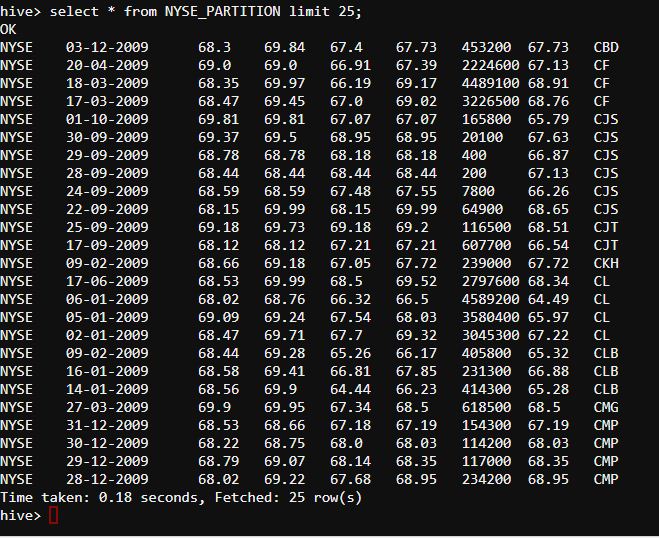
**Hadoop Hive Console Log:**

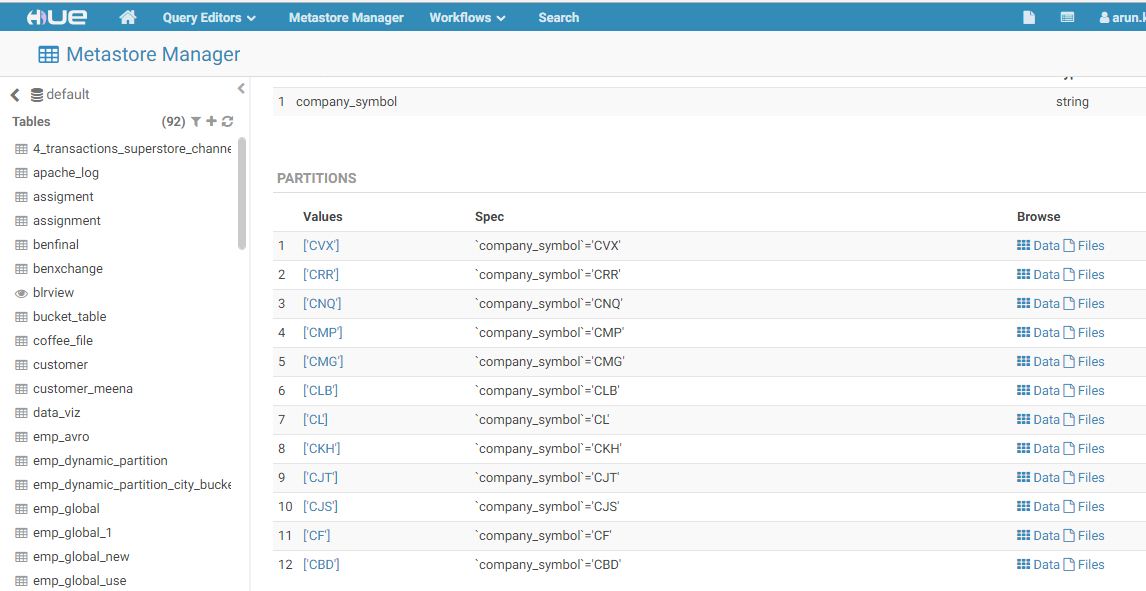
****

**ScreenShots:**

**1.Data loaded to Partition Table:**

**2.Data from Partition table retrieved successfully**

****

**3.Partitions in HIVE**