

## Step-by-step commands:

**Created new folder "NYSE\_Assmt" in FTP.**

**Uploaded "NYSE\_daily\_File.csv" under folder "NYSE\_Assmt" in FTP.**

```
hdfs dfs -put NYSE_Assmt
```

```
hdfs dfs -ls
```

```
cd NYSE_Assmt
```

```
ls
```

```
cat NYSE_daily_File.csv
```

**Now the NYSE\_daily\_File.csv is in HDFS under NYSE\_Assmt folder. Go back to home directory.**

```
cd ..
```

**To view first 10 lines of the file in HDFS:**

```
hdfs dfs -cat NYSE_Assmt/NYSE_daily_File.csv | head -10
```

**To view last 10 lines of the file in HDFS:**

```
hdfs dfs -cat NYSE_Assmt/NYSE_daily_File.csv | tail -10
```

**Entering HIVE**

```
hive
```

```
create database roopa;
```

```
use roopa;
```

```
create table NYSE_table(stock_exchange string, company_symbol string, date string, open_price float, high_price float, low_price float, close_price float, volume int, adj_close_price float) row format delimited fields terminated by ',' stored as textfile;
```

```
show tables;
```

```
describe NYSE_table;
```

```
load data inpath '/user/roopa.sondur_outlook/NYSE_Assmt' into table NYSE_table;
```

**Question 1: Create NYSE Partition table based on the date field.**

**Question 2: This table needs to store the data as ORC file.**

create table NYSE\_table\_partition(stock\_exchange string, company\_symbol string, open\_price float, high\_price float, low\_price float, close\_price float, volume int, adj\_close\_price float) partitioned by(date string) row format delimited fields terminated by ',' stored as orcfile;

**Question 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.**

from NYSE\_table insert into table NYSE\_table\_partition partition(date) select stock\_exchange, company\_symbol, open\_price, high\_price, low\_price, close\_price, volume, adj\_close\_price, date where open\_price>68 AND high\_price<70;

**\*\*\*WARNING MESSAGE\*\*\***

**Did some configuration settings**

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

from NYSE\_table insert into table NYSE\_table\_partition partition(date) select stock\_exchange, company\_symbol, open\_price, high\_price, low\_price, close\_price, volume, adj\_close\_price, date where open\_price>68 AND high\_price<70;

**\*\*\*WARNING MESSAGE\*\*\***

**Did some configuration settings**

*set hive.exec.max.dynamic.partitions=10000;*

*set hive.exec.max.dynamic.partitions.pernode=500;*

from NYSE\_table insert into table NYSE\_table\_partition partition(date) select stock\_exchange, company\_symbol, open\_price, high\_price, low\_price, close\_price, volume, adj\_close\_price, date where open\_price>68 AND high\_price<70;

**The output which I got can be seen in Hue in the below path:**

**/user/hive/warehouse/roopa.db/nyse\_table\_partition/**