

1. Copy the file from source to the local directory

2. Create the database

```
hive (default)> SET hive.execution.engine=mr;  
hive (default)> create database ;
```

3. Create Table

We cant load text data into a ORC table. We can load data from an existing table into ORC table. Or we load data from a ORC file into ORC table.

As we got text file, I created the external table as the staging table.

To create the external table in hive using the location where the file is uploaded to the HDFS

```
create external table nyse_daily  
(stock_exchange string ,company_symbol string, date string, open_price float,  
high_price float, low_price float, close_price float ,  
volume float, adjusted_close_price float)  
row format delimited fields terminated by ',' stored as textfile  
location '/user/test';--here specify the hdfs location to which you've uploaded the data
```

```
select count(*) from nyse_daily
```

4. Create the ORC table with Company Symbol as Partition from the original table and load data into it according to the requirement and get the output.

before creating the ORC dynamic partitioned table and inserting the data, change the partition mode from strict to non-strict for dynamic partition

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

-- to create a ORC partitioned table use below query

```
create table nyse_daily_dynamic_partition  
(stock_exchange string ,date string, open_price float,  
high_price float, low_price float, close_price float ,  
volume float, adjusted_close_price float) partitioned by (company_symbol string)  
row format delimited fields terminated by ',' stored as ORC
```

-- to insert the data into the ORC partitioned table from the staging table and to apply filters while loading the data

***insert into table nyse_daily_dynamic_partition partition(company_symbol) select
Stock_exchange,date,open_price,high_price,low_price,close_price,volume,adjusted_close_price,company_symbol from nyse_daily where open_price >68 and high_price < 70***

-- use this query to see the partitions

show partitions nyse_daily_dynamic_partition

-- this is used to see the data in the partitioned table(79 rows will come as output of this query)

select * from nyse_daily_dynamic_partition