**Introduction**

To solve the given problem, you need to follow the following steps:

- Start Hive services and ensure that Hive daemons are running in your Hadoop Cluster

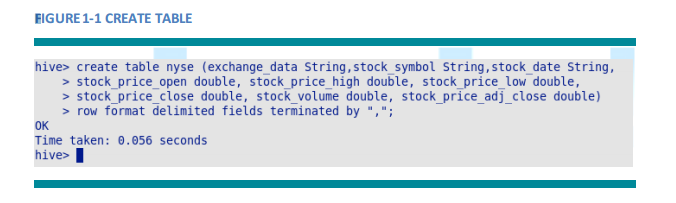
- Use HIveQL to create a table with same column names as given in csv file

- Load the csv file in Hive table - Execute the HiveQL query to get the desired results

**1. Problem Solution**

1.1 Create Hive Table Use ‘create table’ hive command to create the Hive table for your dataset:

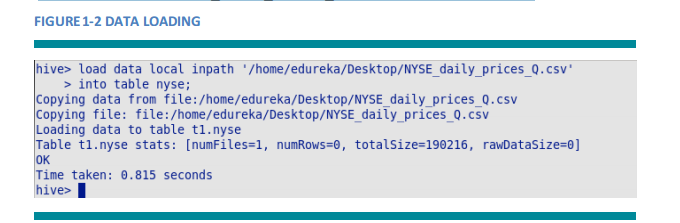
hive> create table nyse (exchange\_data String,stock\_symbol String,stock\_date String,stock\_price\_open double, stock\_price\_high double, stock\_price\_low double, stock\_price\_close double, stock\_volume double, stock\_price\_adj\_close double) row format delimited fields terminated by ",";



**1.2 Load Data to Hive Table**

Use the following Hive command to load data into Hive table:

hive> load data local inpath '/home/cloudera/NYSE\_daily\_prices\_Q.csv' into table nyse;



**1.3 Calculate the Covariance**

Use the following query to calculate the covariance between stocks.

hive> select a.STOCK\_SYMBOL, b.STOCK\_SYMBOL, month(a.STOCK\_DATE), (AVG(a.STOCK\_PRICE\_HIGH\*b.STOCK\_PRICE\_HIGH) - (AVG(a.STOCK\_PRICE\_HIGH)\*AVG(b.STOCK\_PRICE\_HIGH))) from nyse a join nyse b on a.STOCK\_DATE=b.STOCK\_DATE where a.STOCK\_SYMBOL, month(a.STOCK\_DATE);

