Module-5: Hive and HiveQL

Assignment Solution

edureka!

Version 1.0

Module-6: Hive and HiveQL

Assignment Solution – Calculating a Stock's Covariance

Table of Contents

Introduction 2

1. Problem solution 2

    1.1 Create Hive Table 2

    1.2 Load Data to Hive Table 2

    1.3 Calculate the Covariance 3

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 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 ",";
```

FIGURE 1-1 CREATE TABLE

```
hive> create table nyse (exchange 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
    > ",";
OK
Time taken: 0.081 seconds
```

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;
```

FIGURE 1-2 DATA LOADING

```
hive> load data local inpath '/home/user/hadoop/Hive/Assignment_5/NYSE_daily_prices_Q.csv' into table nyse;
Copying data from file:/home/user/hadoop/Hive/Assignment_5/NYSE_daily_prices_Q.csv
Copying file: file:/home/user/hadoop/Hive/Assignment_5/NYSE_daily_prices_Q.csv
Loading data to table default.nyse
Table default.nyse stats: [num_partitions: 0, num_files: 1, num_rows: 0, total_size: 190216, raw_data_size: 0]
OK
```

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<b.STOCK_SYMBOL and
year(a.STOCK_DATE)=2008
group by a.STOCK_SYMBOL, b. STOCK_SYMBOL,
month(a.STOCK_DATE);
```

FIGURE 1-3 CALCULATE COVARIANCE

```
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<b.STOCK_SYMBOL and
    > year(a.STOCK_DATE)=2008
    > group by
    > a.STOCK_SYMBOL, b. STOCK_SYMBOL,
    > month(a.STOCK_DATE);
```

```
QRR     QTM     1       -0.13994965986395158
QRR     QTM     2       2.06000000021489E-4
QRR     QTM     3       0.002930000000027637
QRR     QXM     1       -0.015941496598614435
QRR     QXM     2       0.00512499999992497
QRR     QXM     3       -0.013358000000010861
QTM     QXM     1       -0.003653287981865816
QTM     QXM     2       -0.026352500000005108
QTM     QXM     3       0.00605699999994872
QTM     QXM     4       0.02727107438016851 4
QTM     QXM     5       0.02668866213152121 2
QTM     QXM     6       0.0528705215419442 7
QTM     QXM     7       0.0231260330578519 9
QTM     QXM     8       0.0220612244897981 92
QTM     QXM     9       0.05976031746031651 4
QTM     QXM     10      0.003507939508506030 5
QTM     QXM     11      0.01837174515235462 4
QTM     QXM     12      -0.0038603305785122055
Time taken: 36.907 seconds, Fetched: 18 row(s)
```

You can also create a Hive script (say **'script.sql')** and execute it from the shell rather than writing each statement individually in Hive shell.

FIGURE 1-4 HIVE SCRIPT

```
user@ubuntu:~$ hive -f script.sql
14/04/20 03:54:42 INFO Configuration.deprecation: mapred.input.dir.recursive is deprecated. Instead, use mapred
uce.input.fileinputformat.input.dir.recursive
14/04/20 03:54:42 INFO Configuration.deprecation: mapred.max.split.size is deprecated. Instead, use mapreduce.i
nput.fileinputformat.split.maxsize
```