STOCKS ANALYSIS-CASE STUDY

The data set is a NYSE_daily_prices_q.csv that contains the stock information such as daily quotes etc at New York stock exchange.

Problem Statement: Using HiveQL to analyse the stock exchange dataset and calculate covariance between the stocks for each month. This will help a stock broker in recommending the stocks to his customer.

Covariance: This finance term represents the degree or amount that two stocks or financial instruments move together or apart from each other. With covariance, investors have the opportunity to seek out different investment options based upon their respective risk profile. It is a statistical measure of how one investment moves in relation to another.

- A positive covariance means that asset returns moved together. If investment instruments or stocks tend to be up or down during the same time periods, they have positive covariance.
- A negative covariance means returns move inversely. If one investment tends to be up while the other is down, they have negative covariance.

SOLUTION:

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 ",";
```

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

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.

```
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
```