BIGDATA HADOOP AND SPARK DEVELOPER

STOCK EXCHANGE DATA ANALYSIS PROJECT

**RAVEENA PRABAKARAN**

**TIGER ANALYTICS**

**PREREQUISITES**

1. Create a stockprices mysql table and load data from StockPrice.csv

create table stockprices(Trading\_date varchar(100),Symbol varchar(20),Open double,Close Double,Low double,High double,Volume int)

load data local infile 'StockPrices.csv' into table stockprices fields terminated by ',' lines terminated by '\n' ;

1. Create a stock\_companies mysql table and load data from Stockcompaines.csv in hdfs using sqoop

create table stock\_companies(Symbol varchar(20),Company\_name varchar(50), Sector varchar(50), Sub\_industry varchar(50), Headquarter varchar(50));

sqoop export --connect jdbc:mysql://sqoopdb.slbdh.cloudlabs.com/raveenaprabakartigeranaly --username raveenaprabakartigeranaly --password raveenaprabakartigeranalyneeix --table stock\_companies --export-dir Stockcompanies.csv --input-fields-terminated-by ','

 1) Create a data pipeline using sqoop to pull the data from the table below from MYSQL server into Hive.

1. sqoop import --connect jdbc:mysql://sqoopdb.slbdh.cloudlabs.com/raveenaprabakartigeranaly --username raveenaprabakartigeranaly --password raveenaprabakartigeranalyneeix --table stock\_companies --hive-import --hive-database bdhsproject --hive-table stockcompanies -m 1

Text

Description automatically generated

1. sqoop import --connect jdbc:mysql://sqoopdb.slbdh.cloudlabs.com/raveenaprabakartigeranaly --username raveenaprabakartigeranaly --password raveenaprabakartigeranalyneeix --table stock\_prices --hive-import --hive-database bdhsproject --hive-table stockprice -m 1

Graphical user interface, text

Description automatically generated

2) Create a new hive table with the following fields by joining the above two hive tables.  
Please use appropriate Hive built-in functions for columns (a,b,e and h to l).

* Trading\_year: Should contain YYYY for each record
* Trading\_month: Should contain MM or MMM for each record
* Symbol: Ticker code
* CompanyName: Legal name of the listed company
* State: State to be extracted from headquarters value.
* Sector: Business vertical of the listed company
* Sub\_Industry: Business domain of the listed company within a sector
* Open: Average of intra-day opening price by month and year for each listed company
* Close: Average of intra-day closing price by month and year for each listed company
* Low: Average of intra-day lowest price by month and year for each listed company
* High: Average of intra-day highest price by month and year for each listed company
* Volume: Average of number of shares traded by month and year for each listed company

create table stock\_data5 as select trading\_year,trading\_month, sc.Symbol, Company\_name ,trim(split(Headquarter,"\;")[1]) state,Sector,Sub\_industry, open, close, low, high, volume from stockcompanies sc, (select Symbol, split(Trading\_date,'-')[2] trading\_year,split(Trading\_date,'-')[1] trading\_month,round(avg(Open),2) open, round(avg(Close),2) close, round(avg(Low),2) low,round(avg(High),2) high,round(avg(Volume),2) volume from stockprice group by Symbol, split(Trading\_date,'-')[1],split(Trading\_date,'-')[2]) sp where sc.Symbol=sp.Symbol ;

Text

Description automatically generated

**DATA ANALYSIS USING HIVE**

  3) Find the top five companies that are good for investment

create table companyanalysis as ( select company\_name, min(trading\_year) min, max(trading\_year) max,min(trading\_month) min\_month, max(trading\_month) max\_month from stock\_data5 group by company\_name; )

select startstock.company\_name,((close-open)/open)\*100 growth\_percent from (select ca.company\_name, open from stock\_data5 sd5, companyanalysis ca where sd5.trading\_year = ca.min and sd5.trading\_month = ca.min\_month and sd5.company\_name = ca.company\_name) startstock,(select ca.company\_name, close from stock\_data5 sd5, companyanalysis ca where sd5.trading\_year = ca.max and sd5.trading\_month = ca.max\_month and sd5.company\_name = ca.company\_name) endstock where startstock.company\_name = endstock.company\_name sort by growth\_percent desc limit 5;

Text

Description automatically generated

1. Show the best-growing industry by each state, having at least two or more industries mapped.

STEP 1 : CREATE TABLE COMPANYANALYSIS1

create table companyanalysis1 as select state, sub\_industry, stock\_start.company\_name, ((stock\_end.close-stock\_start.open)/stock\_start.open)\*100 growth\_percent from

(select chv.company\_name,open from stock\_data5 sd, companyanalysis chv where sd.trading\_year=chv.min and sd.trading\_month=chv.min\_month and sd.company\_name=chv.company\_name) stock\_start,

(select chv.company\_name, close from stock\_data5 sd, companyanalysis chv where sd.trading\_year=chv.max and sd.trading\_month=chv.max\_month and sd.company\_name=chv.company\_name) stock\_end,

(select company\_name, state, sub\_industry from stock\_data5 group by company\_name,state,sub\_industry) sd

where (stock\_end.close-stock\_start.open)>0 and stock\_start.company\_name=stock\_end.company\_name and sd.company\_name=stock\_start.company\_name;

Text

Description automatically generated

STEP 2 : CREATE TABLE INDUSTRY\_GROWTH

create table industry\_growth as select state,sub\_industry, avg(growth\_percent) ind\_growth from companyanalysis1 group by state, sub\_industry having count(sub\_industry) >=2 ;

A screenshot of a computer

Description automatically generated

STEP 3 : SELECT BEST GROWING INDUSTRY FOR EACH STATE

select indgwt.state, sub\_industry, ind\_growth from industry\_growth indgwt ,

(select state, max(ind\_growth) maxgrowth from industry\_growth group by state) ig

where ig.state = indgwt.state and indgwt.ind\_growth = ig.maxgrowth;

Text

Description automatically generated

5) For each sector find the following.

* Worst year
* b. Best year
* c. Stable year

STEP 1: CREATE TABLE SECTORWISEGROWTH

create table sectorwisegrowth as select open.sector, open.trading\_year,(close-open) growth from

(select sector,trading\_year,avg(open) open from stock\_data5 where trading\_month = 1 group by sector,trading\_year) open,

(select sector,trading\_year,avg(close) close from stock\_data5 where trading\_month=12 group by sector,trading\_year) close

where open.sector = close.sector and open.trading\_year = close.trading\_year ;

Text

Description automatically generated

STEP2 : SELECT YEAR

a) worst year

select sg.sector ,sg.trading\_year , sg.growth from sectorwisegrowth sg , ( select sector, min(growth) growth from sectorwisegrowth group by sector ) sg1 where sg.sector = sg1.sector and sg.growth = sg1.growth ;

Text

Description automatically generated

B ) best year

select sg.sector ,sg.trading\_year , sg.growth from sectorwisegrowth sg , ( select sector, max(growth) growth from sectorwisegrowth group by sector ) sg1 where sg.sector = sg1.sector and sg.growth = sg1.growth ;

Text

Description automatically generated

c) stable year

select sg.sector ,sg.trading\_year , round(sg.growth,0)from sectorwisegrowth sg , ( select sector,round(avg(growth),0) growth from sectorwisegrowth group by sector ) sg1 where sg.sector = sg1.sector and sg.growth = sg1.growth ;

Text

Description automatically generated

A screenshot of a computer

Description automatically generated

create table stock\_data as select trading\_year,

trading\_month, sc.symbol, company\_name, trim(split(headquarter,”\;”)

[1]) state,

sector, sub\_industry, open, close, low, high, volume

from stock\_companies sc,

(select symbol, year(trading\_date) trading\_year, month(trading\_date)

trading\_month,

round(avg(open),2) open, round(avg(close),2) close, round(avg(low),2)

low,

round(avg(high),2) high, round(avg(volume),2) volume

from stock\_prices

group by symbol, month(trading\_date),year(trading\_date)) sp

where sc.symbol=sp