Stock Exchange Data Analysis

**1) Create a database in Hive (DATABASE NAME: BDHS\_PROJECT)**

**i. Stock\_prices ii. Stock\_companies**

Create DATABASE BDHS\_PROJECT;

use BDHS\_PROJECT;

CREATE TABLE stock\_prices (

Trading\_date date,

Symbol varchar(20),

Open double,

Close double,

Low double,

High double,

Volume int

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES

(

"separatorChar" = ",",

"quoteChar" = "\""

) STORED AS TEXTFILE

tblproperties('serialization.null.format'='','skip.header.line.count'='1');

);

CREATE TABLE stock\_companies (

Symbol varchar(25),

Company\_name varchar(120),

Sector varchar(80),

Sub\_industry varchar(80),

Headquarter varchar(120)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES

(

"separatorChar" = ",",

"quoteChar" = "\""

) STORED AS TEXTFILE

tblproperties('serialization.null.format'='','skip.header.line.count'='1');

);

);

LOAD DATA INPATH 'Stockcompanies.csv'

INTO TABLE stock\_companies

;

LOAD DATA inpath 'StockPrices.csv'

INTO TABLE stock\_prices

;

**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).**

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.symbol;

select \* from stock\_data limit 5;

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

create table stock\_table1 as select company\_name, min(trading\_year) min\_year, max(trading\_year) max\_year, min(trading\_month) min\_month, max(trading\_month) max\_month from stock\_data group by company\_name;

select stock\_start.company\_name, ((close-open)/open)\*100 growth\_percent

from(select t1.company\_name, open from stock\_data sd, stock\_table1 t1

where sd.trading\_year = t1.min\_year and sd.trading\_month = t1.min\_month and sd.company\_name = t1.company\_name) stock\_start, (select t1.company\_name, close from stock\_data sd, stock\_table1 t1 where sd.trading\_year = t1.max\_year and sd.trading\_month = t1.max\_month and sd.company\_name = t1.company\_name) stock\_end where stock\_start.company\_name = stock\_end.company\_name sort by growth\_percent desc limit 5;

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

create table stock\_table2 as select

state, sub\_industry, stock\_start.company\_name, ((stock\_end.close-stock\_start.open)/stock\_start.open)\*100 growth\_percent

from (select t1.company\_name,open

from stock\_data sd, stock\_table1 t1

where sd.trading\_year=t1.min and

sd.trading\_month=t1.min\_month and

sd.company\_name=t1.company\_name)stock\_start,

(select t1.company\_name, close

from stock\_data sd, stock\_table1 t1

where sd.trading\_year=t1.max and

sd.trading\_month=t1.max\_month and

sd.company\_name=t1.company\_name)stock\_end,

(select company\_name, state, sub\_industry

from stock\_data

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;

create table stock\_table3 as select state,sub\_industry, avg(growth\_percent)ind\_growth from stock\_table2 group by state, sub\_industry having count(sub\_industry>=2);

select t3.state, sub\_industry, ind\_growth from stock\_table3 t3,

(select state,max(ind\_growth) max\_growth from stock\_table3 group by state) max\_ind where max\_ind.state = t3.state and

t3.ind\_growth = max\_ind.max\_growth;

**5) For each sector find the following.**

**a) Worst Year b) Best Year c)Stabel Year**

create table stock\_table4 as select open.sector, open.trading\_year,

(close-open) growth from (select sector,trading\_year,avg(open) open from stock\_data where trading\_month = 1 group by sector,trading\_year) open, (select sector,trading\_year,avg(close) close from stock\_data

where trading\_month=12 group by sector,trading\_year) close

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

select x.sector,x.trading\_year,x.growth from stock\_table4 x,

(select sector,min(growth) growth from stock\_table4 group by sector) y where x.sector=y.sector and x.growth=y.growth;

select a.sector,a.trading\_year,a.growth from stock\_table4 a,

(select b.sector,max(growth) growth from stock\_table4 group by sector) b where a.sector=b.sector and

a.growth=b.growth;

close

create table stock\_table2 as select

state, sub\_industry, stock\_start.company\_name, ((stock\_end.close-stock\_start.open)/stock\_start.open)\*100 growth\_percent

from (select t1.company\_name,open

from stock\_data sd, stock\_table1 t1

where sd.trading\_year=t1.min and

sd.trading\_month=t1.min\_month and

sd.company\_name=t1.company\_name)stock\_start,

(select t1.company\_name, close

from stock\_data sd, stock\_table1 t1

where sd.trading\_year=t1.max and

sd.trading\_month=t1.max\_month and

sd.company\_name=t1.company\_name)stock\_end,

(select company\_name, state, sub\_industry

from stock\_data

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;

select \* from stock\_table2 limit 5;

create table stock\_table3 as select state,sub\_industry, avg(growth\_percent)ind\_growth from stock\_table2 group by state, sub\_industry having count(sub\_industry>=2);

select t3.state, sub\_industry, ind\_growth from stock\_table3 t3,

(select state,max(ind\_growth) max\_growth from stock\_table3 group by state) max\_ind where max\_ind.state = t3.state and

t3.ind\_growth = max\_ind.max\_growth;

create table stock\_table4 as select open.sector, open.trading\_year,

(close-open) growth from (select sector,trading\_year,avg(open) open from stock\_data where trading\_month = 1 group by sector,trading\_year) open, (select sector,trading\_year,avg(close) close from stock\_data

where trading\_month=12 group by sector,trading\_year) close

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

select x.sector,x.trading\_year,x.growth from stock\_table4 x,

(select sector,min(growth) growth from stock\_table4 group by sector) y where x.sector=y.sector and x.growth=y.growth;

select a.sector,a.trading\_year,a.growth from stock\_table4 a,

(select b.sector,max(growth) growth from stock\_table4 group by sector) b where a.sector=b.sector and

a.growth=b.growth;