1. **Create a data base in Hive (DATABASE NAME: BDHS\_PROJECT)**
2. **Stock\_prices**
3. **Stock\_companies**

CREATE DATABASE BDHS\_Project1;

CREATE EXTERNAL TABLE Stock\_Prices

(

Trading\_date DATE,

Symbol STRING,

Open DOUBLE,

Close DOUBLE,

Low DOUBLE,

High DOUBLE,

Volume INT

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

LOAD DATA INPATH '/user/anabig114231/StockPrices.csv' INTO TABLE stock\_prices;

ALTER TABLE stock\_prices

SET TBLPROPERTIES ("skip.header.line.count"="1");

select \* from stock\_prices;

CREATE EXTERNAL TABLE Stock\_Companies

(

Symbol STRING,

Company\_name STRING,

Sector STRING,

Sub\_industry STRING,

Headquarter STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

LOAD DATA INPATH '/user/anabig114231/Stockcompanies.csv' INTO TABLE stock\_companies;

ALTER TABLE stock\_companies

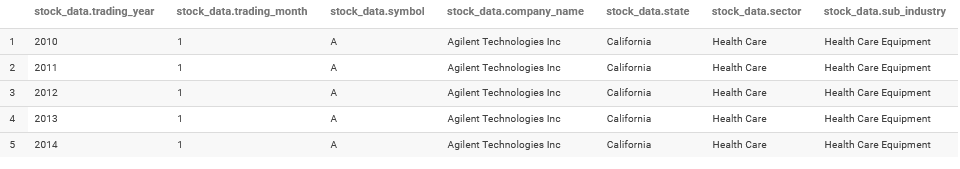
SET TBLPROPERTIES ("skip.header.line.count"="1");

select \* from stock\_companies;

1. **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\_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;



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

CREATE table stock\_table\_2 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\_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, (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\_table\_2 limit 5;



create table stock\_tbl\_3 as select state, sub\_industry, avg(growth\_percent) ind\_growth from stock\_table\_2 group by state, sub\_industry having count(sub\_industry>=2); select \* from stock\_tbl3 limit 10;

select t3.state, sub\_industry, ind\_growth from stock\_tbl\_3 t3, (select state, max(ind\_growth) max\_growth from stock\_tbl\_3 group by state) max\_ind where max\_ind.state = t3.state and t3.ind\_growth = max\_ind.max\_growth;

1. **For each sector find the following.**

**• Worst year**

**• Best 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 \* from stock\_table4 limit 5;



**WORST 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;



**BEST YEAR**

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

