

PROJECT SOURCE CODE:

STEP 1: Locate the table wfrom where data need to be migrated

MYSQL CODE :

```
USE BDHS_PROJECTS;
```

```
Show tables;
```

```
select * from STOCK_PRICES limit 5;
```

```
select * from STOCK_COMPANIES limit 5;
```

Step 2: Create your own hive database

HIVE :

```
beeline
```

```
!connect jdbc:hive2://localhost:10000
```

```
Create Database 13may2020;
```

```
use 13may2020;
```

Step 3: From terminal impert data from mysql table to hive table below import command create and puts the data in hive database.

Terminal:

```
sqoop import --connect jdbc:mysql://ip-10-0-1-10.ec2.internal/BDHS_PROJECT --username labuser  
--password simplilearn --table=STOCK_COMPANIES --hive-import --hive-database=13may2020 -m 1
```

```
sqoop import --connect jdbc:mysql://ip-10-0-1-10.ec2.internal/BDHS_PROJECT --username labuser  
--password simplilearn --table=STOCK_PRICES --hive-import --hive-database=13may2020 -m 1
```

Step 4: Create Table with data as asked in the project

HIVE:

```
create table stock_data as

select trading_year,
trading_month,
sc.ticker_symbol as symbol,
Security as company_name,
trim(split(headquarter,";")[1]) state,
sector,
Sub_Industry as 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.ticker_symbol = sp.symbol;
```

DATA ANALYSIS USING HIVE:

Question 1 : Find the top five companies that are good for investment

Create table company_horizon_view 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 chv.company_name,
```

```

open from stock_data sd,
company_horizon_view chv
where sd.trading_year = chv.min_year
and sd.trading_month = chv.min_month
and sd.company_name = chv.company_name
) stock_start,
(
select
chv.company_name,
close from stock_data sd,
company_horizon_view chv
where sd.trading_year = chv.max_year
and sd.trading_month = chv.max_month
and sd.company_name = chv.company_name
) stock_end
where stock_start.company_name = stock_end.company_name
sort by growth_percent desc limit 5;

```

Question 2 : Show the best-growing industry by each state, having at least two or more industries mapped.

Create table company_growth 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_data sd,company_horizon_view chv

where sd.trading_year = chv.min_year and sd.trading_month = chv.min_month and sd.company_name
= chv.company_name)

stock_start,

(

select

chv.company_name,

close

from stock_data sd,company_horizon_view chv

where sd.trading_year = chv.max_year

and sd.trading_month = chv.max_month

and sd.company_name = chv.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 industry_growth as
```

```
select
```

```
state,
```

```
sub_industry,
```

```
avg(growth_percent) ind_growth
```

```
from company_growth
```

```
group by state,sub_industry;
```

```
Select
```

```
ig.state,
```

```
sub_industry,
```

```
ind_growth
```

```
from industry_growth ig,
```

```
(
```

```
select
```

```
state,
```

```
max(ind_growth) max_growth
```

```
from industry_growth
```

```
group by state
```

```
having count(sub_industry) >= 2
```

```
) inn_ig
where inn_ig.state = ig.state
and ig.ind_growth = inn_ig.max_growth;
```

Question 3: For each sector find the following.

a. Worst year

b. Best year

c. Stable year

```
create table sector_growth 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;
```

a. Worst year

```
select
x.sector,
x.trading_year,
x.growth
from
sector_growth x,
(
```

```
select sector,  
min(growth) growth  
from sector_growth  
group by sector  
) y  
where x.sector=y.sector  
and x.growth = y.growth;
```

b. Best year

```
select  
x.sector,  
x.trading_year,  
x.growth  
from  
sector_growth x,  
(  
select sector,  
max(growth) growth  
from sector_growth  
group by sector  
) y  
where x.sector=y.sector  
and x.growth = y.growth;
```

c. Stable year

```
select
x.sector,
x.trading_year,
x.growth
from
sector_growth x,
(
select sector,
avg(growth) growth
from sector_growth
group by sector
) y
where x.sector=y.sector
and x.growth = y.growth;
```