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' ;

2) 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.

a) 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

```
hive> select * from stockcompanies;
OK
COTY    Coty; Inc       Consumer Staples        Personal Products       New York; NY
CCI     Crown Castle International Corp.        Real Estate     REITs   Houston; Texas
CSRA    CSRA Inc.       Information Technology  IT Consulting & Other Services  Falls Church; Virginia
CSX     CSX Corp.       Industrials     Railroads       Jacksonville; Florida
CMI     Cummins Inc.    Industrials     Industrial Machinery    Columbus; Indiana
CVS     CVS Health      Consumer Staples        Drug Retail     Woonsocket; Rhode Island
DHI     D. R. Horton    Consumer Discretionary  Homebuilding    Fort Worth; Texas
DHR     Danaher Corp.   Industrials     Industrial Conglomerates        Washington; D.C.
DRI     Darden Restaurants      Consumer Discretionary  Restaurants     Orlando; Florida
DVA     DaVita Inc.     Health Care     Health Care Facilities  Denver; Colorado
DE      Deere & Co.     Industrials     Construction & Farm Machinery & Heavy Trucks    Moline; Illinois
DLPH    Delphi Automotive       Consumer Discretionary  Auto Parts & Equipment  Gillingham; Kent; United Kingdom
DAL     Delta Air Lines Industrials     Airlines        Atlanta; Georgia
XRAY    Dentsply Sirona Health Care     Health Care Supplies    York; Pennsylvania
DVN     Devon Energy Corp.      Energy  Oil & Gas Exploration & Production      Oklahoma City; Oklahoma
DLR     Digital Realty Trust    Real Estate     Specialized REITs       San Francisco; California
DFS     Discover Financial Services     Financials      Consumer Finance        Riverwoods; Illinois
DISCA   Discovery Communications-A      Consumer Discretionary  Cable & Satellite       Silver Spring; Maryland
DISCK   Discovery Communications-C      Consumer Discretionary  Cable & Satellite       Silver Spring; Maryland
DG      Dollar General  Consumer Discretionary  General Merchandise Stores      Goodlettsville; Tennessee
DLTR    Dollar Tree     Consumer Discretionary  General Merchandise Stores      Chesapeake; Virginia
D       Dominion Resources      Utilities       Electric Utilities      Richmond; Virginia
DOV     Dover Corp.     Industrials     Industrial Machinery    Downers Grove; Illinois
DOW     Dow Chemical    Materials       Diversified Chemicals   Midland; Michigan
DPS     Dr Pepper Snapple Group Consumer Staples        Soft Drinks     Plano; Texas
DTE     DTE Energy Co.  Utilities       Multiutilities  Detroit; Michigan
DD      Du Pont (E.I.)  Materials       Diversified Chemicals   Wilmington; Delaware
DUK     Duke Energy     Utilities       Electric Utilities      Charlotte; North Carolina
DNB     Dun & Bradstreet        Industrials     Research & Consulting Services  Short Hills; New Jersey
ETFC    E*Trade Financials      Investment Banking & Brokerage  New York; New York
EMN     Eastman Chemical        Materials       Diversified Chemicals   Kingsport; Tennessee
ETN     Eaton Corporation       Industrials     Industrial Conglomerates        Dublin; Ireland
EBAY    eBay Inc.       Information Technology  Internet Software & Services    San Jose; California
ECL     Ecolab Inc.     Materials       Specialty Chemicals     St. Paul; Minnesota
```

b) 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

```
hive> select * from stockprice limit 5;
OK
05-01-2016      WLTW    123.43  125.839996      122.309998      126.25  2163600
06-01-2016      WLTW    125.239998      119.980003      119.940002      125.540001      2386400
07-01-2016      WLTW    116.379997      114.949997      114.93  119.739998      2489500
08-01-2016      WLTW    115.480003      116.620003      113.5   117.440002      2006300
11-01-2016      WLTW    117.010002      114.970001      114.089996      117.330002      1408600
Time taken: 0.067 seconds, Fetched: 5 row(s)
hive>
```

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 ;

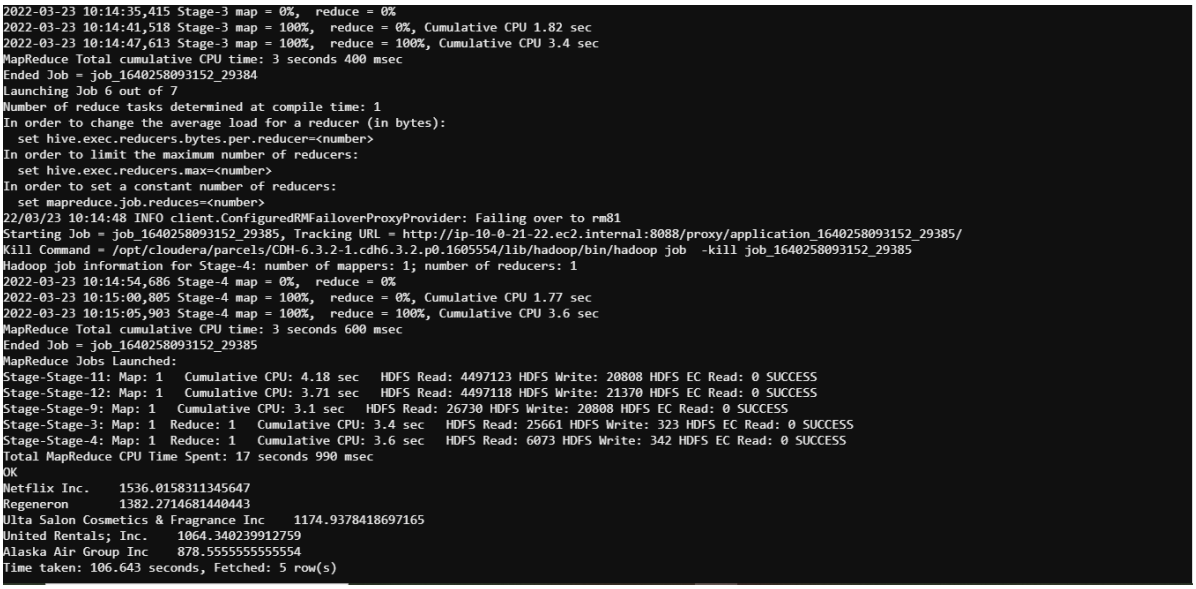
```
Stage-Stage-5: Map: 1   Cumulative CPU: 4.67 sec   HDFS Read: 2797801 HDFS Write: 4489513 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 25 seconds 480 msec
OK
Time taken: 45.884 seconds
hive> select * from stock_data5 limit 5;
OK
2010    01      A       Agilent Technologies Inc        California      Health Care     Health Care Equipment   21.72   21.61   21.4    21.86   4208442.11
2011    01      A       Agilent Technologies Inc        California      Health Care     Health Care Equipment   30.29   30.29   29.96   30.65   4496845.0
2012    01      A       Agilent Technologies Inc        California      Health Care     Health Care Equipment   28.54   28.78   28.22   29.08   5069975.0
2013    01      A       Agilent Technologies Inc        California      Health Care     Health Care Equipment   31.2    31.26   30.97   31.45   4567819.05
2014    01      A       Agilent Technologies Inc        California      Health Care     Health Care Equipment   42.01   42.04   41.66   42.36   3494200.0
Time taken: 0.064 seconds, Fetched: 5 row(s)
hive>
```

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;



4) 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;

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 ;

```
hive> select * from industry_growth limit 50;
OK
California      Apparel Retail  261.8019767671436
California      Application Software    120.31201428397625
California      Biotechnology   189.855597859821
California      Health Care Equipment   147.29321666303753
California      Home Entertainment Software     298.3868203665769
California      Internet Software & Services    336.56041800779366
California      MultiUtilities  62.895455776620366
California      Networking Equipment    16.60504085306744
California      REITs   131.08399426163498
California      Retail REITs    107.50814027970536
California      Semiconductor Equipment 147.9653267908011
California      Semiconductors  282.4593566657631
Connecticut     Industrial Conglomerates        75.02785859563353
Florida Hotels; Resorts & Cruise Lines  137.37933900437363
Florida Industrial Conglomerates        177.39424996713996
Florida Internet Software & Services    163.10267039273256
Illinois        Construction & Farm Machinery & Heavy Trucks    73.11111617775501
Illinois        Health Care Equipment   41.94627917684228
Illinois        Industrial Machinery    136.03903698890892
Illinois        Packaged Foods & Meats  55.38509384373717
Illinois        REITs   65.02753772568181
Ireland Industrial Conglomerates        136.6124307695158
Ireland Pharmaceuticals 245.14488964557233
Maryland        Cable & Satellite       85.87024737549106
Massachusetts   Biotechnology   260.2045747013016
Massachusetts   Health Care Equipment   169.22809835926253
Massachusetts   REITs   68.24052887161051
Massachusetts   Semiconductors  285.17414566777273
Michigan        MultiUtilities  143.76404421430016
Minnesota       Packaged Foods & Meats  166.69719818630523
Missouri        Industrial Conglomerates        85.58764014977949
New Jersey      Health Care Equipment   120.88893239332754
New Jersey      Research & Consulting Services  115.46492537344713
```

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;

```
22/03/23 11:30:05 INFO client.ConfiguredRMFailoverProxyProvider: Failing over to rm81
Starting Job = job_1640258093152_29554, Tracking URL = http://ip-10-0-21-22.ec2.internal:8088/proxy/application_1640258093152_29554/
Kill Command = /opt/cloudera/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554/lib/hadoop/bin/hadoop job  -kill job_1640258093152_29554
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
2022-03-23 11:30:12,035 Stage-4 map = 0%,  reduce = 0%
2022-03-23 11:30:19,213 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 3.65 sec
MapReduce Total cumulative CPU time: 3 seconds 650 msec
Ended Job = job_1640258093152_29554
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.43 sec   HDFS Read: 10966 HDFS Write: 833 HDFS EC Read: 0 SUCCESS
Stage-Stage-4: Map: 1   Cumulative CPU: 3.65 sec   HDFS Read: 7158 HDFS Write: 1409 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 80 msec
OK
California      Internet Software & Services    336.56041800779366
Connecticut     Industrial Conglomerates        75.02785859563353
Florida Industrial Conglomerates        177.39424996713996
Illinois        Industrial Machinery    136.03903698890892
Ireland Pharmaceuticals 245.14488964557233
Maryland        Cable & Satellite       85.87024737549106
Massachusetts   Semiconductors  285.17414566777273
Michigan        MultiUtilities  143.76404421430016
Minnesota       Packaged Foods & Meats  166.69719818630523
Missouri        Industrial Conglomerates        85.58764014977949
New Jersey      Health Care Equipment   120.88893239332754
New York        Diversified Financial Services  244.27936670090116
North Carolina  Apparel; Accessories & Luxury Goods     234.5365478318533
Ohio    Banks   171.58635235361513
Oklahoma        Oil & Gas Exploration & Production      129.87772667777514
Pennsylvania    Diversified Chemicals   163.20832113323664
Texas   Airlines        571.0256197386555
United Kingdom  Insurance Brokers       101.0504280362911
Virginia        Aerospace & Defense     255.37107674669656
Washington      Internet & Direct Marketing Retail      323.86488316718413
Wisconsin       Electric Utilities      131.94734372014784
Time taken: 43.652 seconds, Fetched: 21 row(s)
hive>
```

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 ;

```
2022-03-23 11:45:20,195 Stage-5 map = 0%,  reduce = 0%
2022-03-23 11:45:26,312 Stage-5 map = 100%,  reduce = 0%, Cumulative CPU 3.15 sec
MapReduce Total cumulative CPU time: 3 seconds 150 msec
Ended Job = job_1640258093152_29595
Moving data to directory hdfs://nameservice1/user/hive/warehouse/bdhsproject.db/sectorwisegrowth
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.62 sec   HDFS Read: 4500562 HDFS Write: 3561 HDFS EC Read: 0 SUCCESS
Stage-Stage-4: Map: 1  Reduce: 1   Cumulative CPU: 6.84 sec   HDFS Read: 4500570 HDFS Write: 3561 HDFS EC Read: 0 SUCCESS
Stage-Stage-5: Map: 1   Cumulative CPU: 3.15 sec   HDFS Read: 9536 HDFS Write: 3047 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 610 msec
OK
Time taken: 61.978 seconds
hive> select * from sectorwisegrowth limit 20;
OK
Consumer Discretionary  2010    14.404047387211953
Consumer Discretionary  2011    4.860126582278504
Consumer Discretionary  2012    7.730506329113922
Consumer Discretionary  2013    24.392572590011653
Consumer Discretionary  2014    9.603809523809531
Consumer Discretionary  2015    6.202380952380949
Consumer Discretionary  2016    14.681904761904747
Consumer Staples        2010    4.1823529411764468
Consumer Staples        2011    4.158529411764697
Consumer Staples        2012    4.537941176470589
Consumer Staples        2013    8.819092436974778
Consumer Staples        2014    9.322857142857139
Consumer Staples        2015    4.892626984126977
Consumer Staples        2016    3.183055555555569
Energy  2010    7.710909090909105
Energy  2011    -3.5948484848484491
Energy  2012    -1.9490909090909128
Energy  2013    12.584444444444436
Energy  2014    -6.2644444444444443
Energy  2015    -10.09944444444443
Time taken: 0.055 seconds, Fetched: 20 row(s)
hive>
```

STEP2 : SELECT YEAR

a) worst year

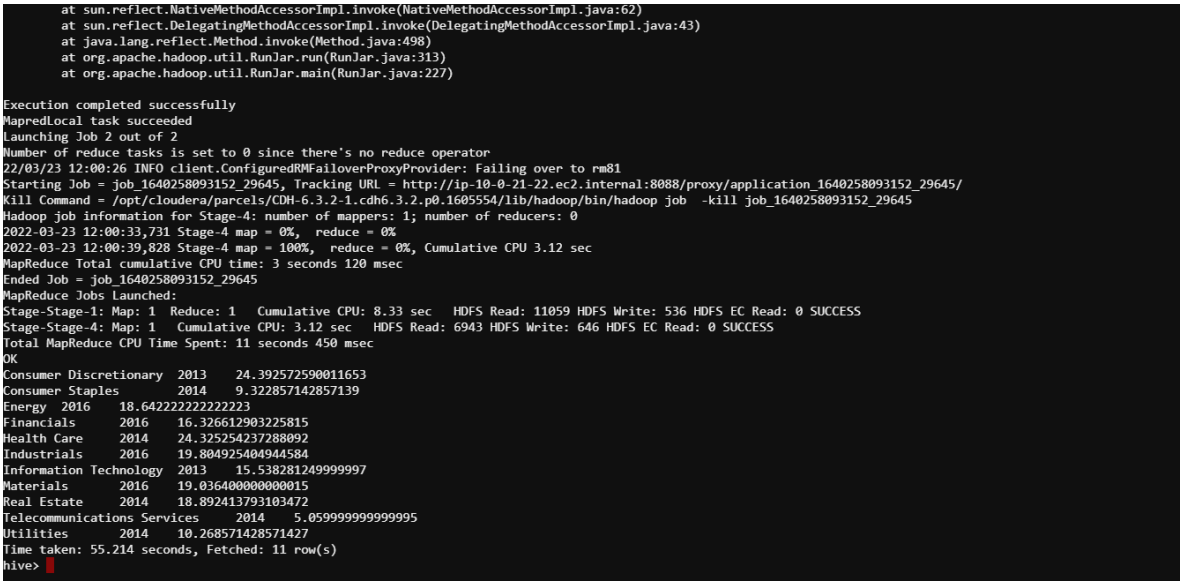
select sg.sector ,sg.trading_year , sg.growth from sectorwisegrowth sg , (select sector, min(growth) growth from sectorwisegrowth group by sector) sg1 where sg.sector = sg1.sector and sg.growth = sg1.growth ;

```
        at org.apache.logging.log4j.core.appender.AbstractManager.getManager(AbstractManager.java:114)
        at org.apache.logging.log4j.core.appender.OutputStreamManager.getManager(OutputStreamManager.java:114)
        at org.apache.logging.log4j.core.appender.RandomAccessFileManager.getFileManager(RandomAccessFileManager.java:74)
        at org.apache.hadoop.hive.ql.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
        ... 35 more

Execution completed successfully
MapredLocal task succeeded
Launching Job 2 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
22/03/23 11:54:50 INFO client.ConfiguredRMFailoverProxyProvider: Failing over to rm81
Starting Job = job_1640258093152_29629, Tracking URL = http://ip-10-0-21-22.ec2.internal:8088/proxy/application_1640258093152_29629/
Kill Command = /opt/cloudera/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554/lib/hadoop/bin/hadoop job  -kill job_1640258093152_29629
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
2022-03-23 11:54:56,172 Stage-4 map = 0%,  reduce = 0%
2022-03-23 11:55:02,272 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 2.88 sec
MapReduce Total cumulative CPU time: 2 seconds 880 msec
Ended Job = job_1640258093152_29629
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.97 sec   HDFS Read: 10935 HDFS Write: 536 HDFS EC Read: 0 SUCCESS
Stage-Stage-4: Map: 1   Cumulative CPU: 2.88 sec   HDFS Read: 6839 HDFS Write: 646 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 850 msec
OK
Consumer Discretionary  2011    4.860126582278504
Consumer Staples        2016    3.183055555555569
Energy  2015    -10.09944444444443
Financials      2011    -6.860689655172415
Health Care     2016    2.080508474576485
Industrials     2015    -2.640000000000029
Information Technology  2013    -2.902539682825396896
Materials       2011    -3.967083333333335
Real Estate     2013    -4.463448275862078
Telecommunications Services     2015    -2.29399999999999
Utilities       2015    -6.473928571428566
Time taken: 42.649 seconds, Fetched: 11 row(s)
hive>
```

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 ;

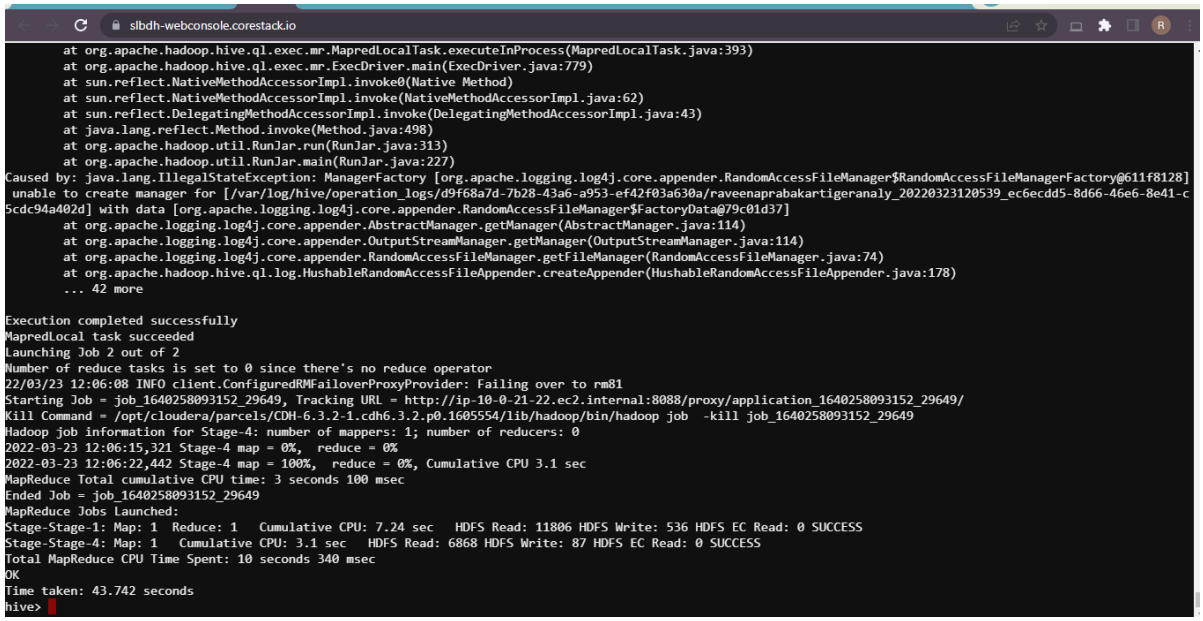


```
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:313)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:227)

Execution completed successfully
MapredLocal task succeeded
Launching Job 2 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
22/03/23 12:00:26 INFO client.ConfiguredRMFailoverProxyProvider: Failing over to rm81
Starting Job = job_1640258093152_29645, Tracking URL = http://ip-10-0-21-22.ec2.internal:8088/proxy/application_1640258093152_29645/
Kill Command = /opt/cloudera/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554/lib/hadoop/bin/hadoop job  -kill job_1640258093152_29645
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
2022-03-23 12:00:33,731 Stage-4 map = 0%,  reduce = 0%
2022-03-23 12:00:39,828 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 3.12 sec
MapReduce Total cumulative CPU time: 3 seconds 120 msec
Ended Job = job_1640258093152_29645
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.33 sec   HDFS Read: 11059 HDFS Write: 536 HDFS EC Read: 0 SUCCESS
Stage-Stage-4: Map: 1   Cumulative CPU: 3.12 sec   HDFS Read: 6943 HDFS Write: 646 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 450 msec
OK
Consumer Discretionary  2013    24.392572590011653
Consumer Staples        2014    9.322857142857139
Energy  2016    18.642222222222223
Financials      2016    16.326612903225815
Health Care     2014    24.325254237288092
Industrials     2016    19.804925404944584
Information Technology  2013    15.538281249999997
Materials       2016    19.036400000000015
Real Estate     2014    18.892413793103472
Telecommunications Services     2014    5.059999999999995
Utilities       2014    10.268571428571427
Time taken: 55.214 seconds, Fetched: 11 row(s)
hive>
```

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 ;



```
slbdh-webconsole.corestack.io
        at org.apache.hadoop.hive.ql.exec.mr.MapredLocalTask.executeInProcess(MapredLocalTask.java:393)
        at org.apache.hadoop.hive.ql.exec.mr.ExecDriver.main(ExecDriver.java:779)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:313)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:227)
Caused by: java.lang.IllegalStateException: ManagerFactory [org.apache.logging.log4j.core.appender.RandomAccessFileManager$RandomAccessFileManagerFactory@611f8128]
 unable to create manager for [/var/log/hive/operation_logs/d9f68a7d-7b28-43a6-a953-ef42f03a630a/raveenaprabakartigeranaly_20220323120539_ec6ecdd5-8d66-46e6-8e41-c
5cdc94a402d] with data [org.apache.logging.log4j.core.appender.RandomAccessFileManager$FactoryData@79c01d37]
        at org.apache.logging.log4j.core.appender.AbstractManager.getManager(AbstractManager.java:114)
        at org.apache.logging.log4j.core.appender.OutputStreamManager.getManager(OutputStreamManager.java:114)
        at org.apache.logging.log4j.core.appender.RandomAccessFileManager.getFileManager(RandomAccessFileManager.java:74)
        at org.apache.hadoop.hive.ql.log.HushableRandomAccessFileAppender.createAppender(HushableRandomAccessFileAppender.java:178)
        ... 42 more

Execution completed successfully
MapredLocal task succeeded
Launching Job 2 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
22/03/23 12:06:08 INFO client.ConfiguredRMFailoverProxyProvider: Failing over to rm81
Starting Job = job_1640258093152_29649, Tracking URL = http://ip-10-0-21-22.ec2.internal:8088/proxy/application_1640258093152_29649/
Kill Command = /opt/cloudera/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554/lib/hadoop/bin/hadoop job  -kill job_1640258093152_29649
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
2022-03-23 12:06:15,321 Stage-4 map = 0%,  reduce = 0%
2022-03-23 12:06:22,442 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 3.1 sec
MapReduce Total cumulative CPU time: 3 seconds 100 msec
Ended Job = job_1640258093152_29649
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.24 sec   HDFS Read: 11806 HDFS Write: 536 HDFS EC Read: 0 SUCCESS
Stage-Stage-4: Map: 1   Cumulative CPU: 3.1 sec   HDFS Read: 6868 HDFS Write: 87 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 340 msec
OK
Time taken: 43.742 seconds
hive>
```