

Proof of Concept (POC) on Stock Market Analysis



PoC On Stock Market Analysis

By:-Karan Pratap Singh

B.Tech in Computer Science

Engineering. JK Lakshmipat University

Context:

The dataset includes a list of all the stocks contained therein and associated key financials such as price, market capitalization, earnings, price/earnings ratio, price to book

Input Format:

Comma Separated values (.CSV) format of all the given datasets-

Data Publicly available at website:

<https://datahub.io/core/s-and-p-500-companies-financials#readm>

Field information

Field Name	Order	Type (Format)	Description
Symbol	1	string	
Name	2	string	
Sector	3	string	
Price	4	number	
Price/Earnings	5	number	
Dividend Yield	6	number	
Earnings/Share	7	number	
52 Week Low	8	number	
52 Week High	9	number	
Market Cap	10	number	

Price_to_Earning helps investors determine the market value of a stock as compared to the company's earnings. In short, the P/E shows what the market is willing to pay today for a stock based on its past or future earnings. A high P/E could mean that a stock's price is high relative to earnings and possibly overvalued. Conversely, a low P/E might indicate that the current stock price is low relative to earnings.

P/E ratios are used by investors and analysts to determine the relative value of a company's shares in an apples-to-apples comparison. It can also be used to compare a company against its own historical record or to compare aggregate markets against one another or over time.

Dividend yield is the ratio of a company's annual dividend compared to its share price

Earnings per share (EPS) is calculated as a company's profit divided by the outstanding shares of its common stock. The resulting number serves as an indicator of a company's profitability.

Market cap refers to the total value of all a company's shares of stock. It is calculated by multiplying the price of a stock by its total number of outstanding shares.

52_Week_Low : A 52-week low is the highest or lowest price at which a stock has traded in the previous year.

52_Week_High: A 52-week high is the highest or lowest price at which a stock has traded in the previous year.

Problem Statement:

1. Select sectors with high market cap.
2. Select sectors with low market cap.
3. Select industry and sectors with high P/E ratio.
4. Select industry and sector with low P/E ratios
5. Select five sectors with lowest stock price trader;
6. Select five sectors with highest stock price trader;
7. Select five sectors with market cap with lowest price
at which stock has traded;
8. Select five sectors with market cap with highest price
at which stock has traded

Big data analytics is a form of advanced analytics, which involves complex applications with elements such as predictive models, statistical algorithms and analysis powered by high-performance analytics systems.

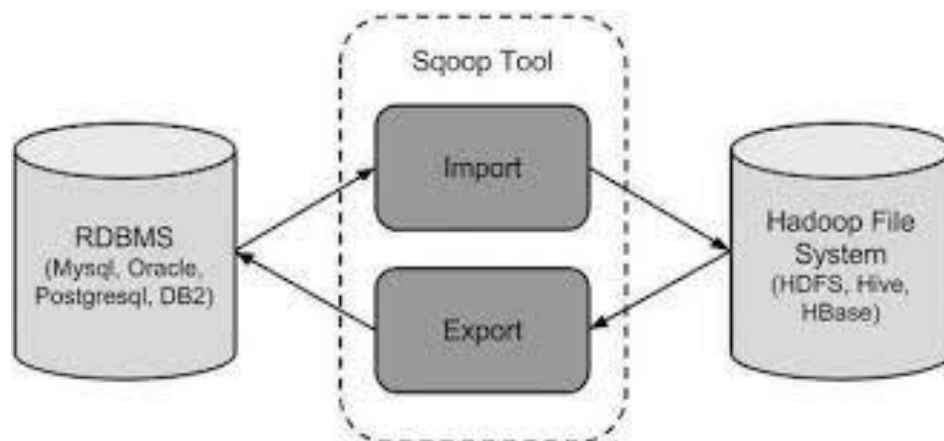
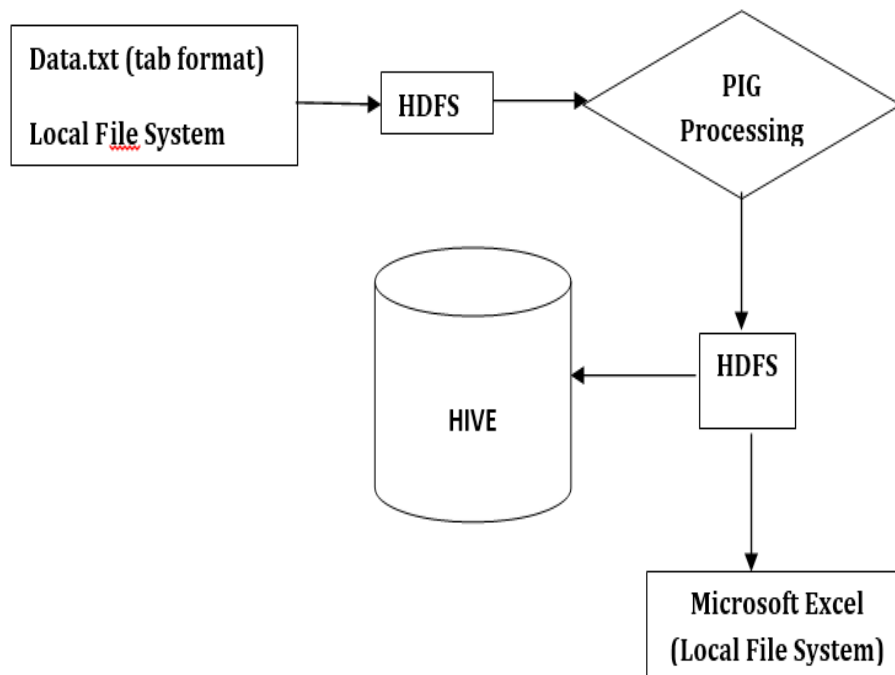
Importance of Bigdata:

Big data analytics applications enable big data analysts, data scientists, predictive modelers, statisticians and other analytics professionals to analyze growing volumes of structured transaction data, plus other forms of data that are often left untapped by conventional business intelligence (BI) and analytics programs. That encompasses a mix of semi-structured and unstructured data -- for example, internet clickstream data, web server logs, social media content, text from customer emails and survey responses, mobile phone records, and machine data captured by sensors connected to the internet of things.

Technologies used by Bigdata:

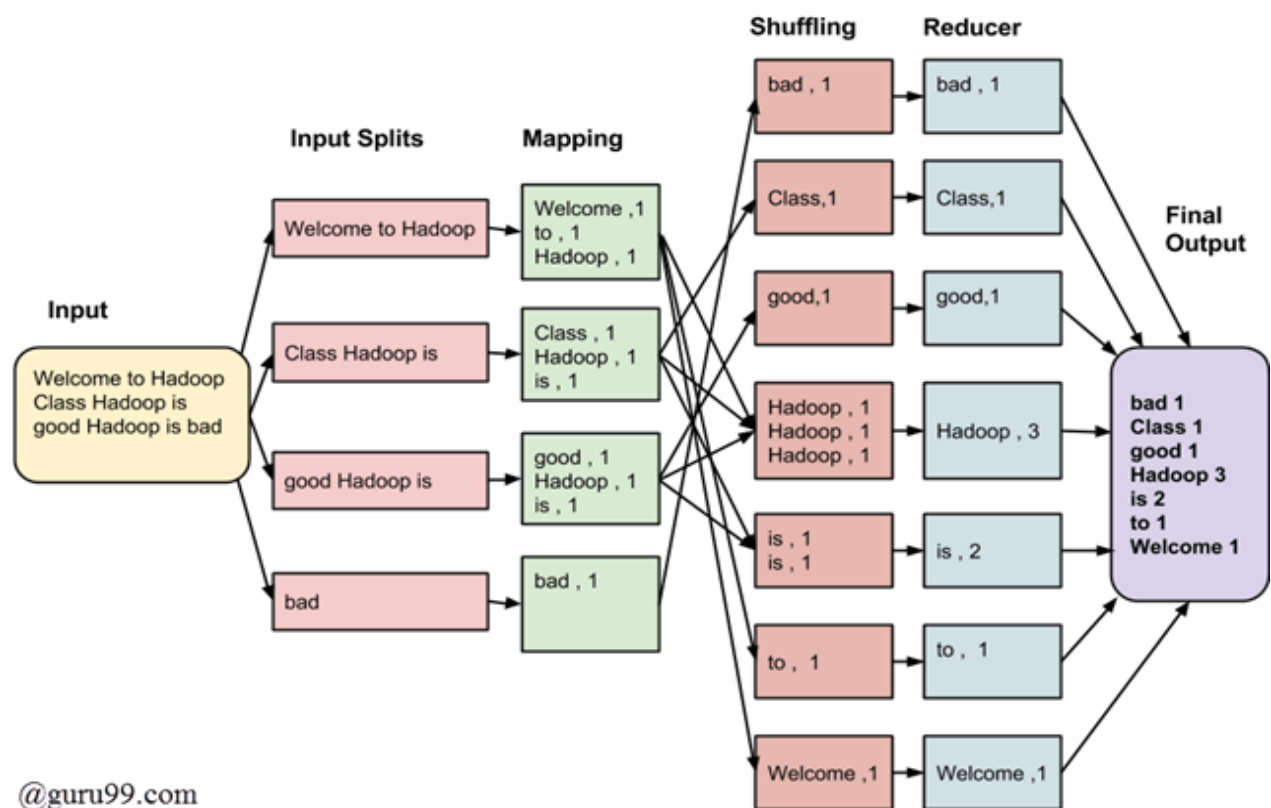
Many of the organizations that collect, process and analyze big data turn to NoSQL databases, as well as Hadoop and its companion tools, including:

- YARN: a cluster management technology and one of the key features in second-generation Hadoop.
- MapReduce: a software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers
- Spark: an open source, parallel processing framework that enables users to run large-scale data analytics applications across clustered systems.
- HBase: a column-oriented key/value data store built to run on top of the Hadoop Distributed File System (HDFS).
- Hive: an open source data warehouse system for querying and analyzing large data sets stored in Hadoop files.
- Kafka: a distributed publish/subscribe messaging system designed to replace traditional message brokers.
- Pig: an open source technology that offers a high-level mechanism for the parallel programming of MapReduce jobs executed on Hadoop clusters.



Hadoop Distributed File System

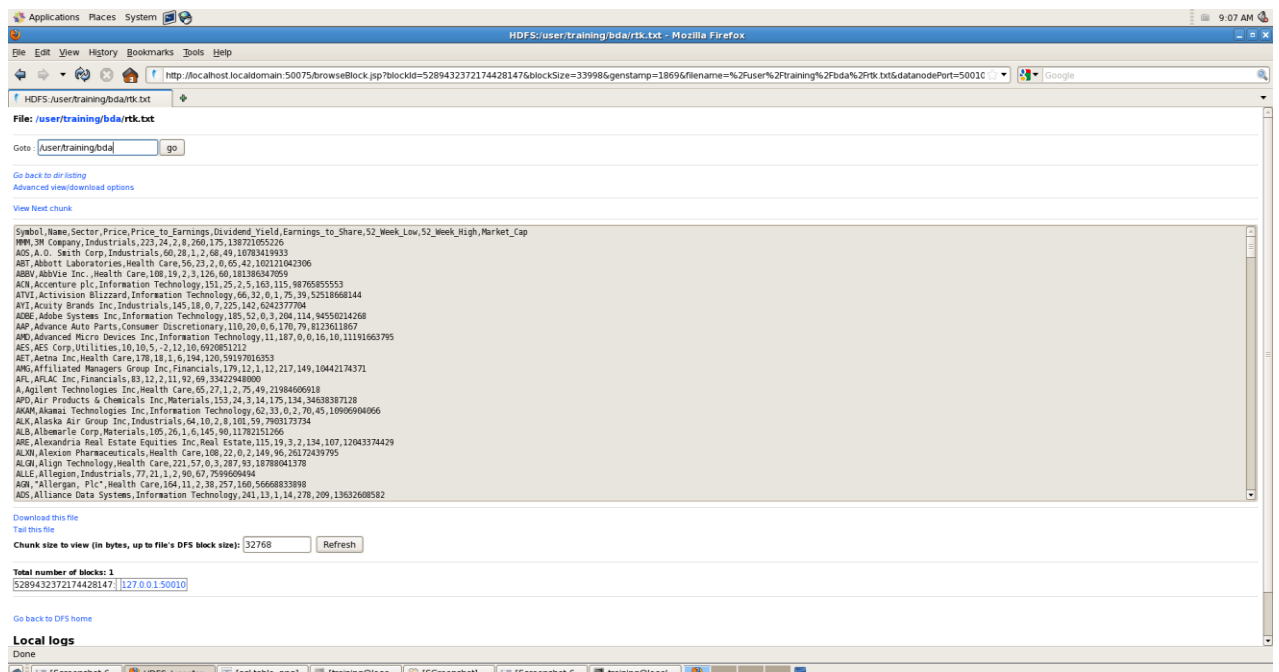
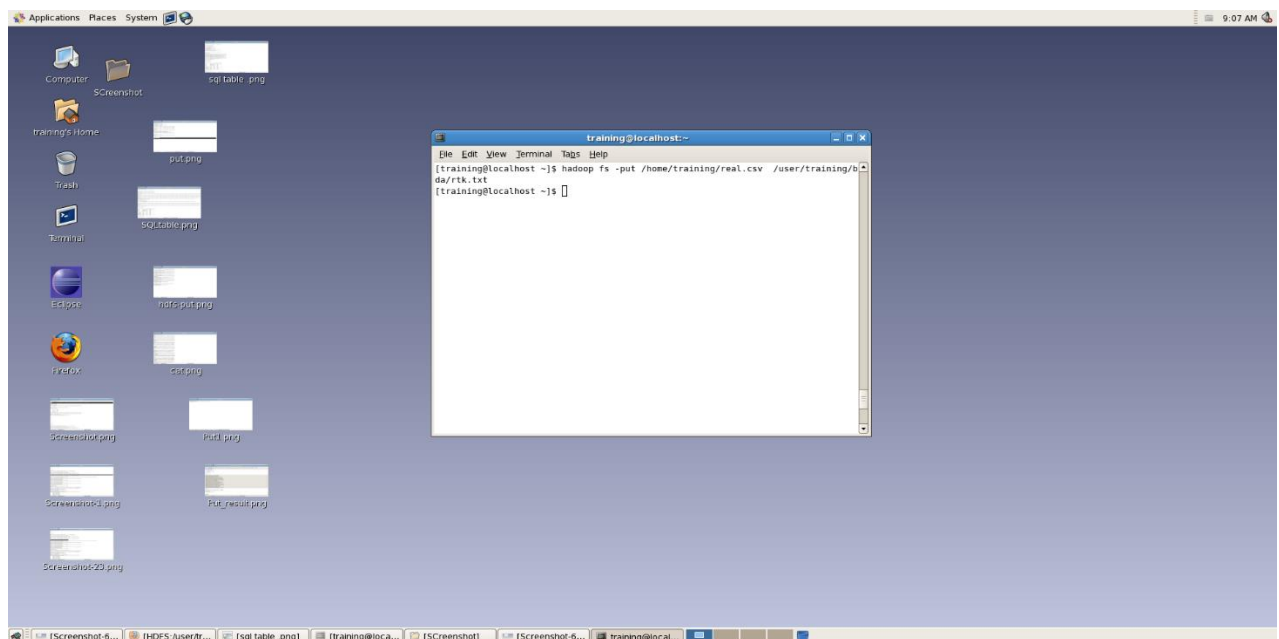
MapReduce is a programming model suitable for processing of huge data. Hadoop is capable of running MapReduce programs written in various languages: Java, Ruby, Python, and C++. MapReduce programs are parallel in nature, thus are very useful for performing large-scale data analysis using multiple machines in the cluster.



Step 1: Moving data into Hadoop Distributed File System (HDFS)

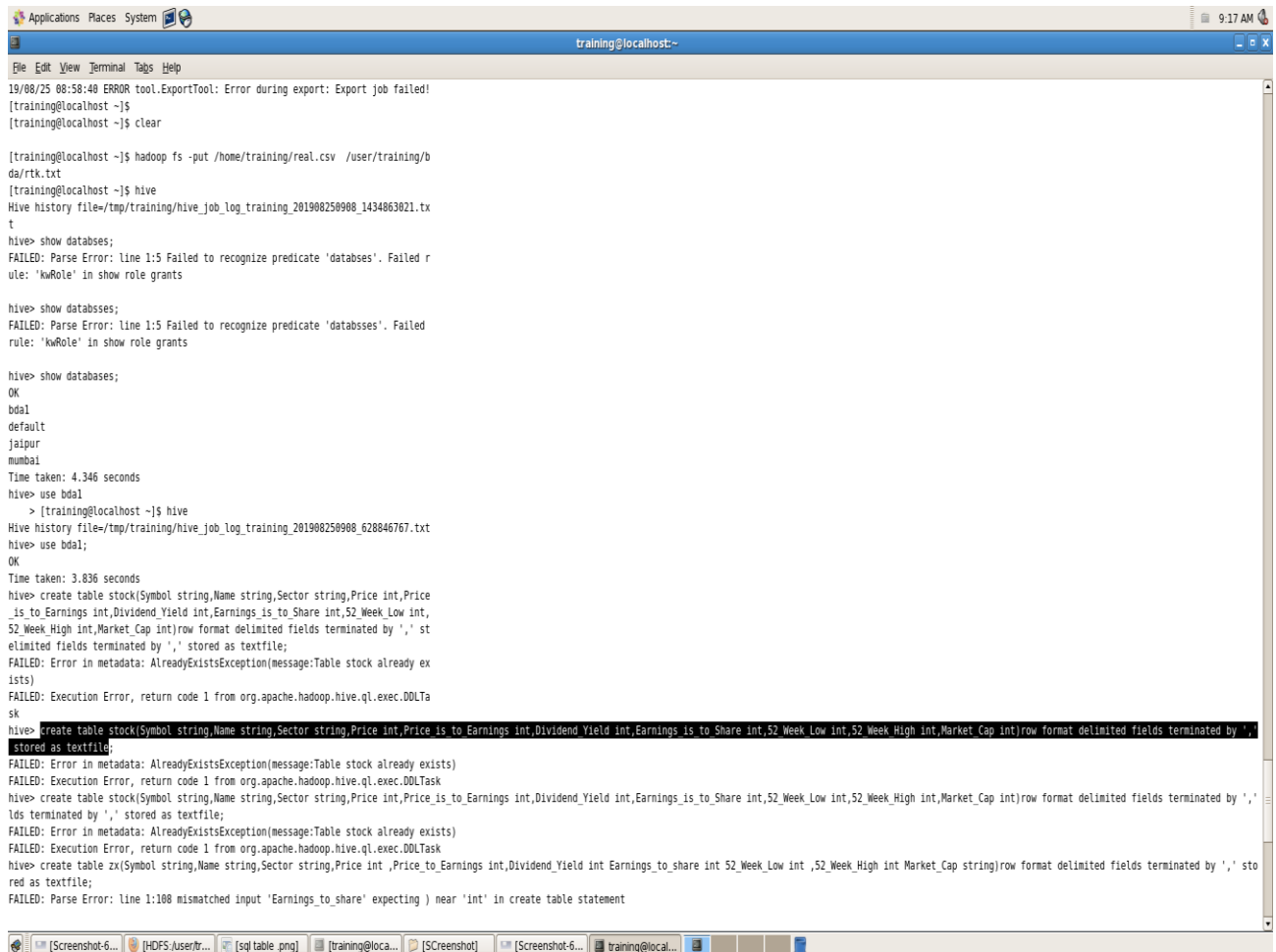
For any queries in Hive we need to load the data in HDFS. For queries in hive we will load the data in HDFS and for queries in Pig we will use the data from local file system.

The snapshot also contains some hadoop fs `-put/get/cat` commands and `mkdir` to create directory in hadoop file system.



Step 2: Load Data into Hive

1. Open the terminal.
2. Write the Hive Command as hive
3. Now you need to create a table and load data into it from local file system.



```
Applications Places System
training@localhost:~$
19/08/25 08:58:48 ERROR tool.ExportTool: Error during export: Export job failed!
[training@localhost ~]$
[training@localhost ~]$ clear

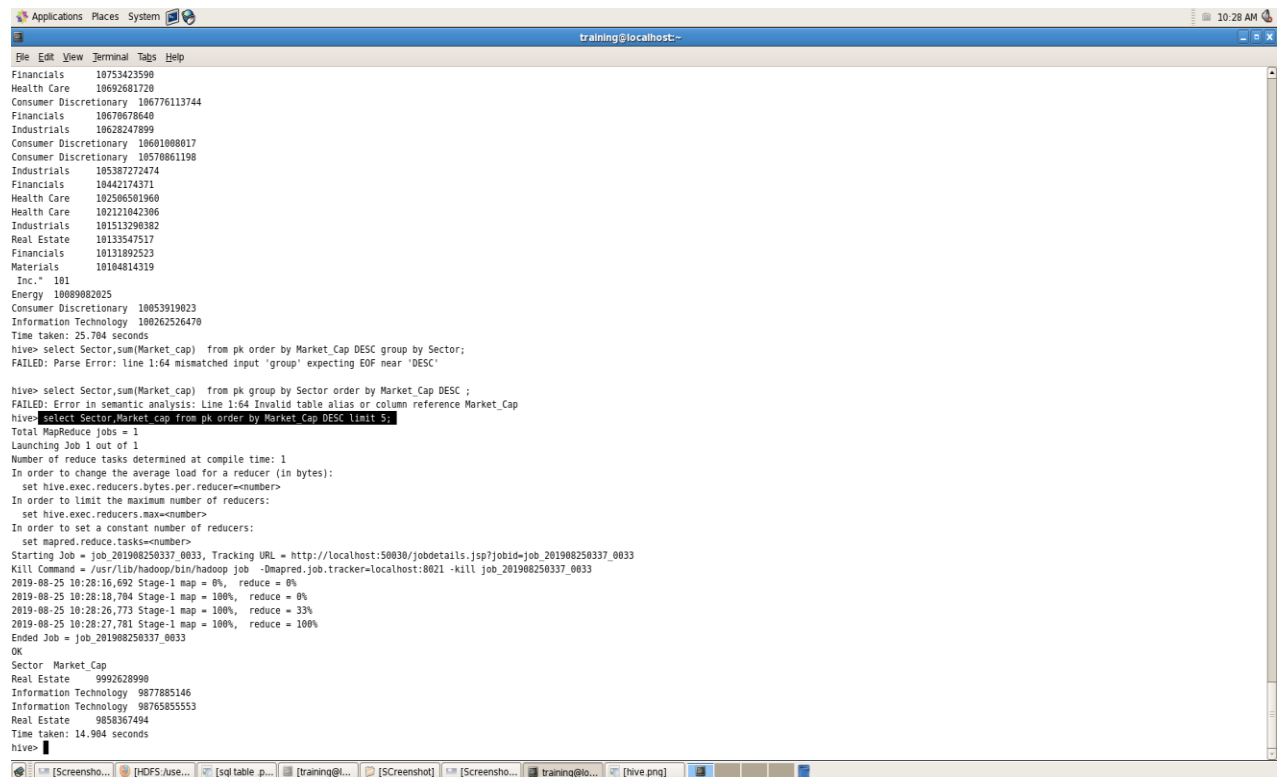
[training@localhost ~]$ hadoop fs -put /home/training/real.csv /user/training/b
da/rtk.txt
[training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_201908250908_1434863021.tx
t
hive> show databases;
FAILED: Parse Error: line 1:5 Failed to recognize predicate 'databases'. Failed r
ule: 'kwRole' in show role grants

hive> show databases;
FAILED: Parse Error: line 1:5 Failed to recognize predicate 'databases'. Failed
rule: 'kwRole' in show role grants

hive> show databases;
OK
bdal
default
jaipur
mumbai
Time taken: 4.346 seconds
hive> use bdal
> [training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_201908250908_628846767.txt
hive> use bdal;
OK
Time taken: 3.836 seconds
hive> create table stock(Symbol string,Name string,Sector string,Price int,Price
_is_to_Earnings int,Dividend_Yield int,Earnings_is_to_Share int,52_Week_Low int,
52_Week_High int,Market_Cap int)row format delimited fields terminated by ',' st
e limited fields terminated by ',' stored as textfile;
FAILED: Error in metadata: AlreadyExistsException(message:Table stock already ex
ists)
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive ql.exec.DDLTa
sk
hive> create table stock(Symbol string,Name string,Sector string,Price int,Price_is_to_Earnings int,Dividend_Yield int,Earnings_is_to_Share int,52_Week_Low int,52_Week_High int,Market_Cap int)row format delimited fields terminated by ','
stored as textfile;
FAILED: Error in metadata: AlreadyExistsException(message:Table stock already exists)
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive ql.exec.DDLTask
hive> create table stock(Symbol string,Name string,Sector string,Price int,Price_is_to_Earnings int,Dividend_Yield int,Earnings_is_to_Share int,52_Week_Low int,52_Week_High int,Market_Cap int)row format delimited fields terminated by ','
lds terminated by ',' stored as textfile;
FAILED: Error in metadata: AlreadyExistsException(message:Table stock already exists)
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive ql.exec.DDLTask
hive> create table zx(Symbol string,Name string,Sector string,Price int ,Price_to_Earnings int,Dividend_Yield int Earnings_to_share int 52_Week_Low int ,52_Week_High int Market_Cap string)row format delimited fields terminated by ',' sto
red as textfile;
FAILED: Parse Error: line 1:108 mismatched input 'Earnings_to_share' expecting ) near 'int' in create table statement
```

Solving Problem Statement:

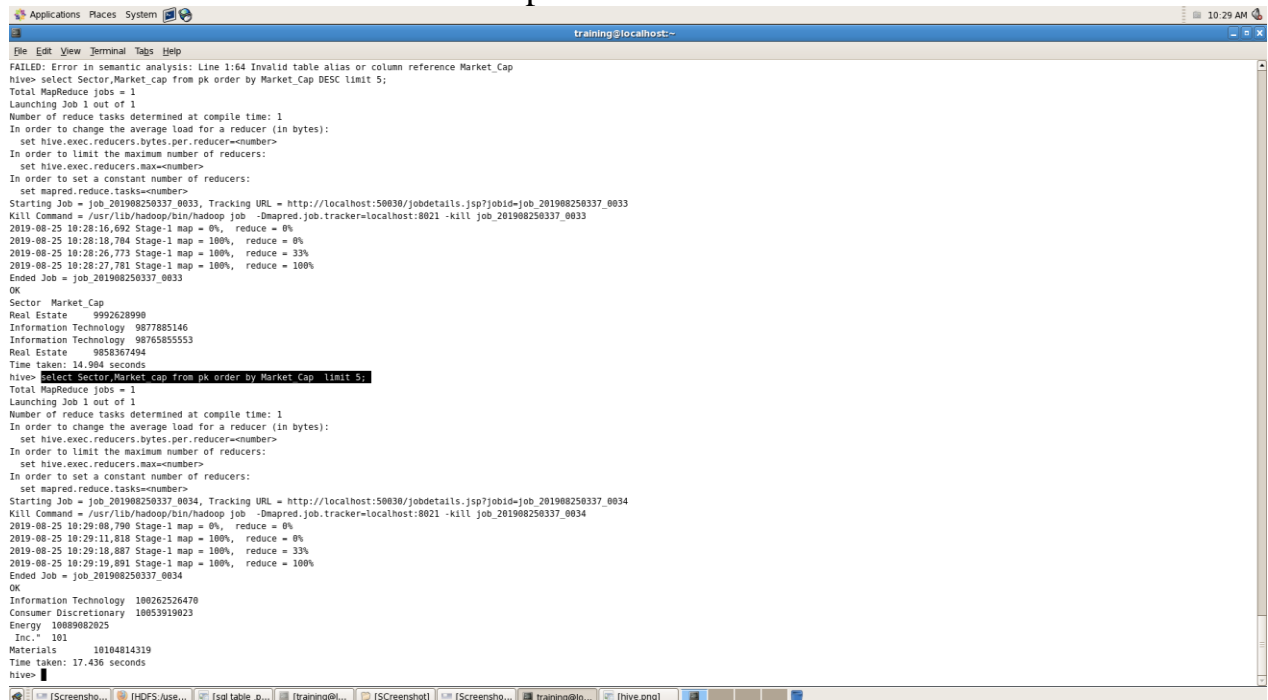
1. Select sectors with high market cap.



```
File Edit View Terminal Tabs Help
training@localhost:~$
Financials 10753423590
Health Care 10692681720
Consumer Discretionary 106776113744
Financials 10670678640
Industrials 10628247899
Consumer Discretionary 10601008017
Consumer Discretionary 10570861198
Industrials 105387272474
Financials 10442174371
Health Care 102506501960
Health Care 102121042306
Industrials 101513290382
Real Estate 10133547517
Financials 10131092523
Materials 10104814319
Inc." 101
Energy 10009082025
Consumer Discretionary 10053919023
Information Technology 100262526470
Time taken: 25.704 seconds
hive> select Sector,sum(Market_cap) from pk order by Market_Cap DESC group by Sector;
FAILED: Parse Error: line 1:64 mismatched input 'group' expecting EOF near 'DESC'

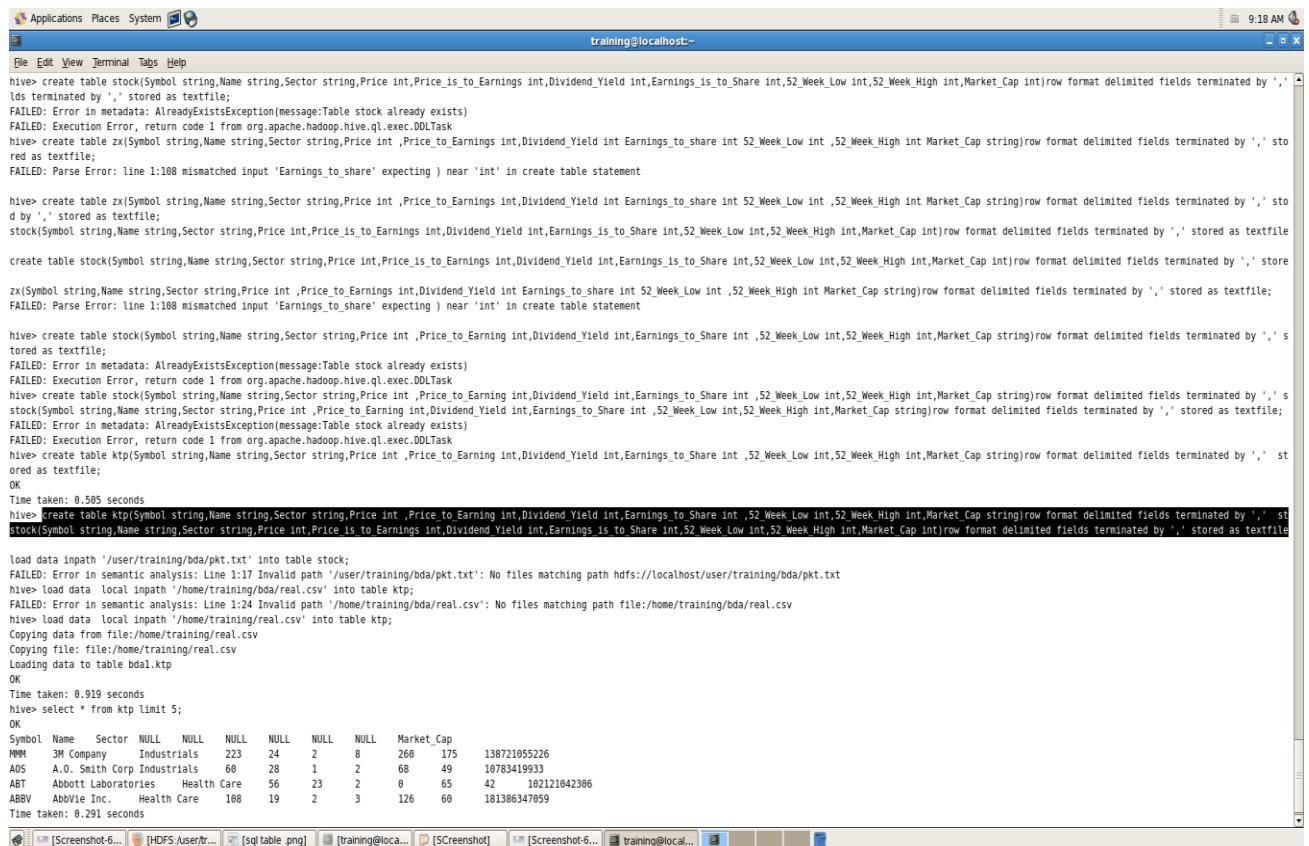
hive> select Sector,sum(Market_cap) from pk group by Sector order by Market_Cap DESC ;
FAILED: Error in semantic analysis: Line 1:64 Invalid table alias or column reference Market_Cap
hive> select Sector,Market_cap from pk order by Market_Cap DESC limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908250337_0033, Tracking URL = http://localhost:50036/jobdetails.jsp?jobid=job_201908250337_0033
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908250337_0033
2019-08-25 10:28:16,692 Stage-1 map = 0%, reduce = 0%
2019-08-25 10:28:18,704 Stage-1 map = 100%, reduce = 0%
2019-08-25 10:28:26,773 Stage-1 map = 100%, reduce = 33%
2019-08-25 10:28:27,781 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908250337_0033
OK
Sector Market_Cap
Real Estate 9992628990
Information Technology 9877885146
Information Technology 9876585553
Real Estate 9858367494
Time taken: 14.904 seconds
hive>
```

2. Select sectors with low market cap.



```
training@localhost:~$ hiveshell
FAILED: Error in semantic analysis: Line 1:64 Invalid table alias or column reference Market_Cap
hive> select Sector,Market_cap from pk order by Market_Cap DESC limit 5;
Total MapReduce Jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908250337_0033, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908250337_0033
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908250337_0033
2019-08-25 10:28:16,692 Stage-1 map = 0%, reduce = 0%
2019-08-25 10:28:18,704 Stage-1 map = 100%, reduce = 0%
2019-08-25 10:28:26,773 Stage-1 map = 100%, reduce = 33%
2019-08-25 10:28:27,781 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908250337_0033
OK
Sector Market_Cap
Real Estate 9992628990
Information Technology 9877885146
Information Technology 9876585553
Real Estate 9858367494
Time taken: 14.904 seconds
hive> select Sector,Market_cap from pk order by Market_Cap limit 5;
Total MapReduce Jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908250337_0034, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908250337_0034
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908250337_0034
2019-08-25 10:29:08,790 Stage-1 map = 0%, reduce = 0%
2019-08-25 10:29:11,818 Stage-1 map = 100%, reduce = 0%
2019-08-25 10:29:18,887 Stage-1 map = 100%, reduce = 33%
2019-08-25 10:29:19,891 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908250337_0034
OK
Information Technology 100262526470
Consumer Discretionary 10053919023
Energy 10089082025
Inc." 101
Materials 10104814319
Time taken: 17.436 seconds
hive>
```

3. Select industry and sectors with high P/E ratio



```
training@localhost:~$ hiveshell
hive> create table stock(Symbol string,Name string,Sector string,Price int,Price_is_to_Earnings int,Dividend_Yield int,Earnings_is_to_Share int,52_Week_Low int,52_Week_High int,Market_Cap int)row format delimited fields terminated by ','
loaded as textfile;
FAILED: Error in metadata: AlreadyExistsException(message:Table stock already exists)
hive> create table zx(Symbol string,Name string,Sector string,Price int ,Price_to_Earnings int,Dividend_Yield int,Earnings_to_Share int ,52_Week_Low int ,52_Week_High int,Market_Cap string)row format delimited fields terminated by ',' stored as textfile;
FAILED: Parse Error: line 1:108 mismatched input 'Earnings_to_Share' expecting ')' near 'int' in create table statement

hive> create table zx(Symbol string,Name string,Sector string,Price int ,Price_to_Earnings int,Dividend_Yield int,Earnings_to_Share int ,52_Week_Low int ,52_Week_High int,Market_Cap string)row format delimited fields terminated by ',' stored as textfile;
FAILED: Error in metadata: AlreadyExistsException(message:Table stock already exists)
hive> create table stock(Symbol string,Name string,Sector string,Price int ,Price_to_Earnings int,Dividend_Yield int,Earnings_to_Share int ,52_Week_Low int,52_Week_High int,Market_Cap string)row format delimited fields terminated by ',' stored as textfile;
FAILED: Error in metadata: AlreadyExistsException(message:Table stock already exists)
hive> create table ktp(Symbol string,Name string,Sector string,Price int ,Price_to_Earnings int,Dividend_Yield int,Earnings_to_Share int ,52_Week_Low int,52_Week_High int,Market_Cap string)row format delimited fields terminated by ',' stored as textfile;
Time taken: 0.505 seconds
hive> create table ktp(Symbol string,Name string,Sector string,Price int ,Price_to_Earnings int,Dividend_Yield int,Earnings_to_Share int ,52_Week_Low int,52_Week_High int,Market_Cap string)row format delimited fields terminated by ',' stored as textfile;
Stock(Symbol string,Name string,Sector string,Price int,Price_is_to_Earnings int,Dividend_Yield int,Earnings_is_to_Share int,52_Week_Low int,52_Week_High int,Market_Cap int)row format delimited fields terminated by ',' stored as textfile

load data inpath '/user/training/bda/pkt.txt' into table stock;
FAILED: Error in semantic analysis: Line 1:17 Invalid path '/user/training/bda/pkt.txt': No files matching path hdfs://localhost/user/training/bda/pkt.txt
hive> load data local inpath '/home/training/bda/real.csv' into table ktp;
FAILED: Error in semantic analysis: Line 1:24 Invalid path '/home/training/bda/real.csv': No files matching path file:/home/training/bda/real.csv
hive> load data local inpath '/home/training/bda/real.csv' into table ktp;
Copying data from file:/home/training/real.csv
Copying file: file:/home/training/real.csv
Loading data to table bda1.ktp
OK
Time taken: 0.919 seconds
hive> select * from ktp limit 5;
OK
Symbol Name Sector NULL NULL NULL NULL NULL Market_Cap
MMM 3M Company Industrials 223 24 2 8 260 175 138721055226
AOS A.O. Smith Corp Industrials 60 20 1 2 68 49 10783419933
ABT Abbott Laboratories Health Care 56 23 2 0 65 42 102121042306
ABBV AbbVie Inc. Health Care 108 19 2 3 126 60 101386347059
Time taken: 0.291 seconds
```

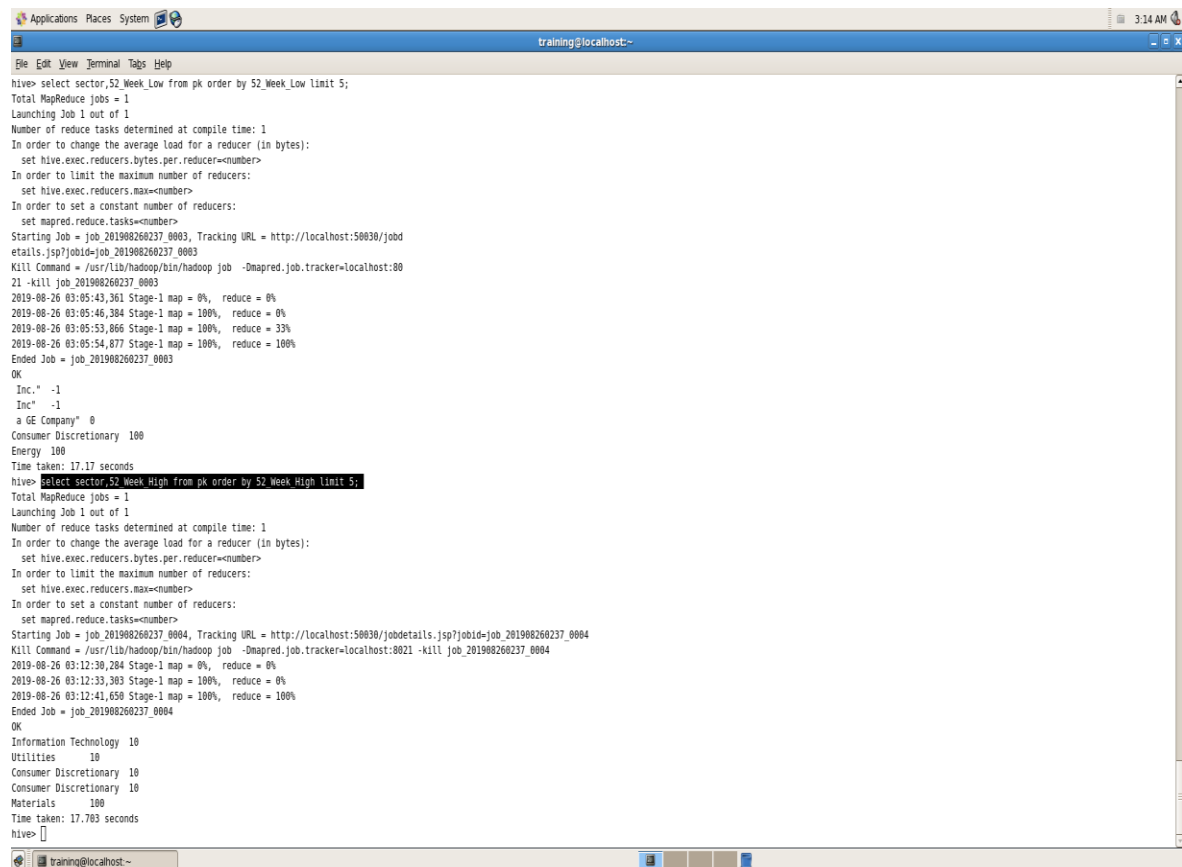
4. Select industry and sectors with low P/E ratio

```
Applications Places System 10:40 AM
training@localhost:~
File Edit View Terminal Tabs Help
FAILED: Error in semantic analysis: Line 1:14 Invalid table alias or column reference Price_to_Earnings
hive> select Sector,Price from pk order by Price limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908250337_0036, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908250337_0036
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908250337_0036
2019-08-25 10:39:40,964 Stage-1 map = 0%, reduce = 0%
2019-08-25 10:39:45,189 Stage-1 map = 100%, reduce = 0%
2019-08-25 10:39:52,265 Stage-1 map = 100%, reduce = 33%
2019-08-25 10:39:53,272 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908250337_0036
OK
Consumer Discretionary 10
Utilities 10
Health Care 100
Consumer Staples 100
Consumer Staples 100
Time taken: 16.956 seconds
hive> select Sector,Price from pk order by Price DESC limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908250337_0037, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908250337_0037
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908250337_0037
2019-08-25 10:40:23,375 Stage-1 map = 0%, reduce = 0%
2019-08-25 10:40:25,384 Stage-1 map = 100%, reduce = 0%
2019-08-25 10:40:32,431 Stage-1 map = 100%, reduce = 33%
2019-08-25 10:40:33,437 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908250337_0037
OK
Inc.* Real Estate
Inc.* Real Estate
Sector Price
Inc.* Information Technology
Inc.* Information Technology
Time taken: 15.012 seconds
hive>
```

5. Select 5 sectors with lowest stock price trader;

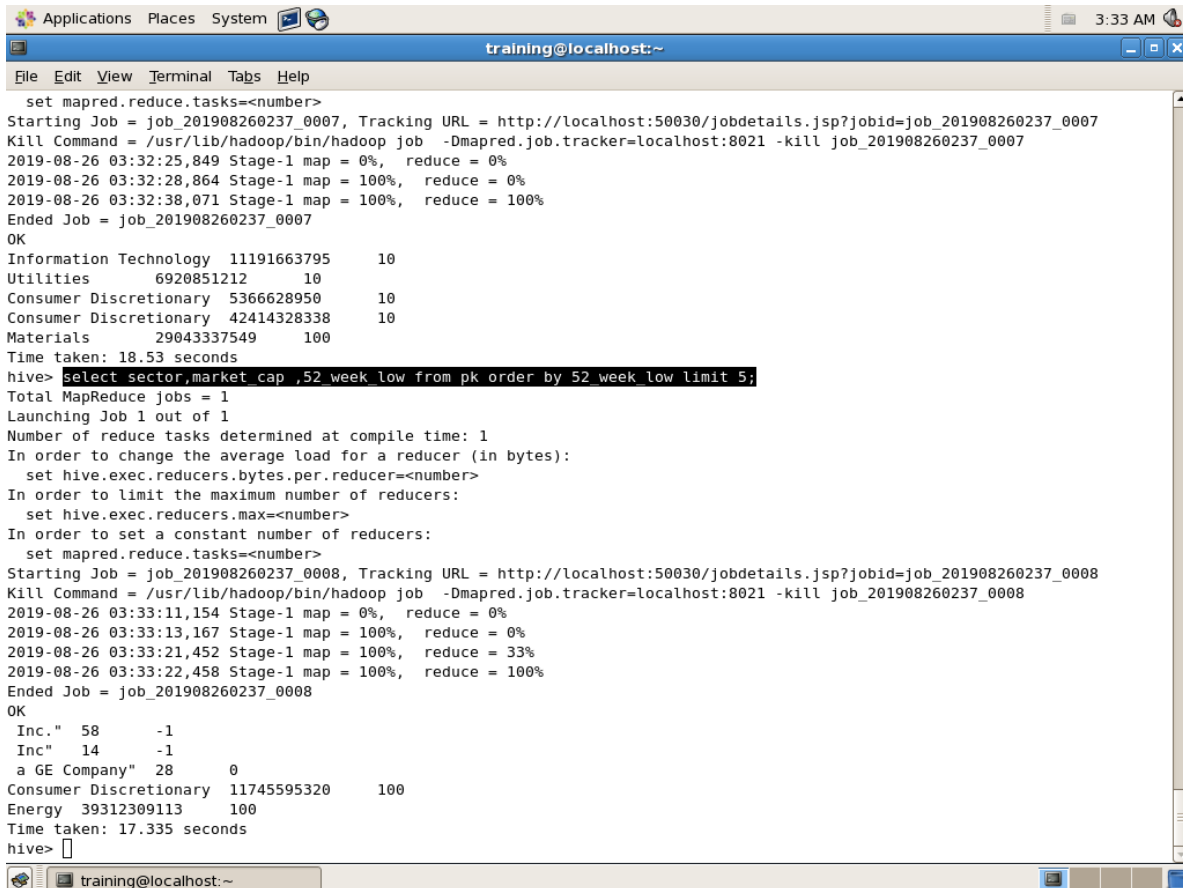
```
Applications Places System 3:08 AM
training@localhost:~
File Edit View Terminal Tabs Help
Inc.* -1
Inc.* -1
a GE Company* 0
Inc.* 2
Time taken: 19.521 seconds
describe pk;
OK
symbol string
name string
sector string
price string
price_to_earning string
dividend_yield string
earnings_to_share string
52_week_low string
52_week_high string
market_cap string
Time taken: 0.109 seconds
select Sector,Name,Price_to_Earnings from pk order by Price_to_Earnings DESC 1
>
>
>;
hive> select sector,52 Week Low from pk order by 52 Week Low limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908260237_0003, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908260237_0003
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908260237_0003
2019-08-26 03:05:43,361 Stage-1 map = 0%, reduce = 0%
2019-08-26 03:05:46,384 Stage-1 map = 100%, reduce = 0%
2019-08-26 03:05:53,866 Stage-1 map = 100%, reduce = 33%
2019-08-26 03:05:54,877 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908260237_0003
OK
Inc.* -1
Inc.* -1
a GE Company* 0
Consumer Discretionary 100
Energy 100
Time taken: 17.17 seconds
hive>
```

6 Select 5 Sectors with highest stock price trader;



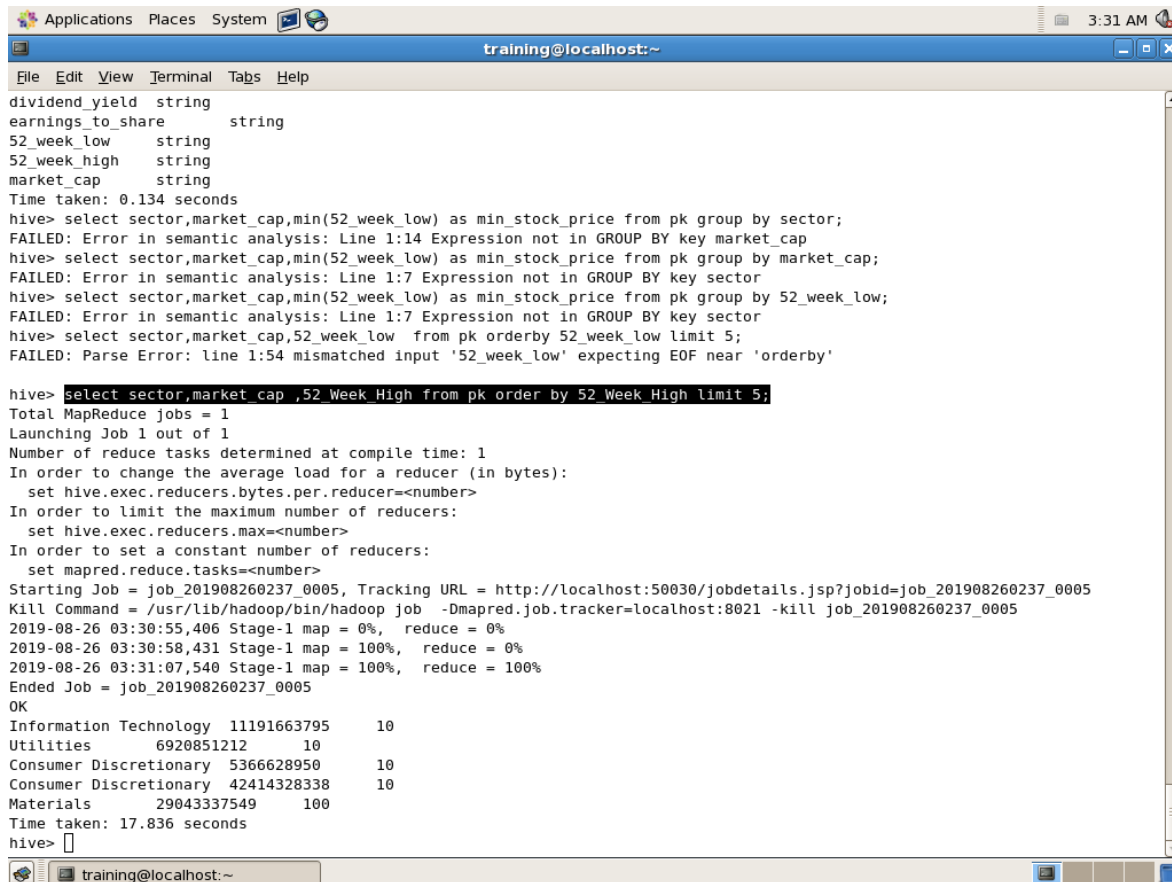
```
Applications Places System 3:14 AM
training@localhost:~
File Edit View Terminal Tags Help
hive> select sector,52_Week_Low from pk order by 52_Week_Low limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908260237_0003, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908260237_0003
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908260237_0003
21 -kill job_201908260237_0003
2019-08-26 03:05:43,361 Stage-1 map = 0%, reduce = 0%
2019-08-26 03:05:46,384 Stage-1 map = 100%, reduce = 0%
2019-08-26 03:05:53,866 Stage-1 map = 100%, reduce = 33%
2019-08-26 03:05:54,877 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908260237_0003
OK
Inc." -1
Inc" -1
a GE Company" 0
Consumer Discretionary 100
Energy 100
Time taken: 17.17 seconds
hive> select sector,52_Week_High from pk order by 52_Week_High limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908260237_0004, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908260237_0004
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908260237_0004
2019-08-26 03:12:30,284 Stage-1 map = 0%, reduce = 0%
2019-08-26 03:12:33,383 Stage-1 map = 100%, reduce = 0%
2019-08-26 03:12:41,650 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908260237_0004
OK
Information Technology 10
Utilities 10
Consumer Discretionary 10
Consumer Discretionary 10
Materials 100
Time taken: 17.763 seconds
hive>
```

7. Select five sectors with market cap with lowest price at which stock has traded;



```
Applications Places System 3:33 AM
training@localhost:~
File Edit View Terminal Tabs Help
set mapred.reduce.tasks=<number>
Starting Job = job_201908260237_0007, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908260237_0007
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908260237_0007
2019-08-26 03:32:25,849 Stage-1 map = 0%, reduce = 0%
2019-08-26 03:32:28,864 Stage-1 map = 100%, reduce = 0%
2019-08-26 03:32:38,071 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908260237_0007
OK
Information Technology 11191663795 10
Utilities 6920851212 10
Consumer Discretionary 5366628950 10
Consumer Discretionary 42414328338 10
Materials 29043337549 100
Time taken: 18.53 seconds
hive> select sector,market_cap ,52_week_low from pk order by 52_week_low limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapred.reduce.tasks=<number>
Starting Job = job_201908260237_0008, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908260237_0008
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908260237_0008
2019-08-26 03:33:11,154 Stage-1 map = 0%, reduce = 0%
2019-08-26 03:33:13,167 Stage-1 map = 100%, reduce = 0%
2019-08-26 03:33:21,452 Stage-1 map = 100%, reduce = 33%
2019-08-26 03:33:22,458 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908260237_0008
OK
Inc." 58 -1
Inc" 14 -1
a GE Company" 28 0
Consumer Discretionary 11745595320 100
Energy 39312309113 100
Time taken: 17.335 seconds
hive>
```

8. Select five sectors with market cap with highest price at which stock has traded;



```
dividend_yield string
earnings_to_share string
52_week_low string
52_week_high string
market_cap string
Time taken: 0.134 seconds
hive> select sector,market_cap,min(52_week_low) as min_stock_price from pk group by sector;
FAILED: Error in semantic analysis: Line 1:14 Expression not in GROUP BY key market_cap
hive> select sector,market_cap,min(52_week_low) as min_stock_price from pk group by market_cap;
FAILED: Error in semantic analysis: Line 1:7 Expression not in GROUP BY key sector
hive> select sector,market_cap,min(52_week_low) as min_stock_price from pk group by 52_week_low;
FAILED: Error in semantic analysis: Line 1:7 Expression not in GROUP BY key sector
hive> select sector,market_cap,52_week_low from pk order by 52_week_low limit 5;
FAILED: Parse Error: line 1:54 mismatched input '52_week_low' expecting EOF near 'orderby'

hive> select sector,market_cap,52 Week High from pk order by 52 Week High limit 5;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201908260237_0005, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201908260237_0005
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201908260237_0005
2019-08-26 03:30:55,406 Stage-1 map = 0%, reduce = 0%
2019-08-26 03:30:58,431 Stage-1 map = 100%, reduce = 0%
2019-08-26 03:31:07,540 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201908260237_0005
OK
Information Technology 11191663795 10
Utilities 6920851212 10
Consumer Discretionary 5366628950 10
Consumer Discretionary 42414328338 10
Materials 29043337549 100
Time taken: 17.836 seconds
hive>
```