BIGDATA HADOOP AND SPARK DEVELOPER

STOCK EXCHANGE DATA ANALYSIS PROJECT - SNAPSHOTS

**RAVEENA PRABAKARAN**

**TIGER ANALYTICS**

**RAVEENA PRABAKARAN**

**TIGER ANALYTICS**

1) Create a data pipeline using sqoop to pull the data from the table below from MYSQL server into Hive.

-> stockcompanies

Text

Description automatically generated

-> stockprices

Graphical user interface, text

Description automatically generated

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

-> stock\_data5

Text

Description automatically generated

**DATA ANALYSIS USING HIVE**

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

Text

Description automatically generated

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

-> companyanalysis1

Text

Description automatically generated

-> industry\_growth

A screenshot of a computer

Description automatically generated

-> best-growing industry by each state

Text

Description automatically generated

5) For each sector find the following.

-> sectorwisegrowth

Text

Description automatically generated

* Worst year

Text

Description automatically generated

* b. Best year

Text

Description automatically generated

* c. Stable year

Text

Description automatically generated