**Question 2:** You have transformed the Ford company share data in assignment 1. Use the table(s) saved in Hive and import the data into Spark. In Spark, using RDD transformations, create new column ‘Quarter’ in the file and list the top 3 prices in each quarter. Perform this activity in Cloudera? (You may do it has an application or command line)

**Bonus 2:** Compute Average Price of every quarter and rate of change to next quarter.

1. Ford.txt is loaded in spark on databricks and records are showed.

**Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated**

1. Ford\_Table is created via createTempView command for querying purpose. The date format is changed according to unix timestamp as shown.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

1. The quarter is extracted from the date using quarter function and temporary table is created as “ford\_edited”. The ford\_edited table is displayed.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

1. Rank is given to each stock price in its quarter with help of rank and partition by command.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application

Description automatically generated

1. The temporary table is created as “rank\_ford\_table” from which rank with 1,2,3 are taken for every quarter.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

1. The average on stock price is calculated for every quarter and shown in order.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

1. The next table is created using lead value, over function and rate of change is calculated

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generated

1. Here I took lag value for rate of change.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Link[:](:%20https:/databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/3166984684234423/2287749801263735/8643682393510715/latest.html) https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/3166984684234423/2287749801263735/8643682393510715/latest.html