1. Design a distributed application using MapReduce (Using Java) which processes a log file of a system. List out the users who have logged for maximum period on the system. Use simple log file and process it using a pseudo distribution mode on Hadoop platform.

2. Write an application using HiveQL for flight information system which will include
   a. Creating, Dropping, and altering Database tables.
   b. Creating an external Hive table.
   c. Load table with data, insert new values and field in the table, Join tables with Hive
   d. Create index on Flight Information Table
   e. Find the average departure delay per day in 2008.

3. Perform the following operations using Python on the Facebook metrics data sets
   a. Create data subsets
   b. Merge Data
   c. Sort Data
   d. Transposing Data
   e. Shape and reshape Data

4. Perform the following operations using Python on the Heart Diseases data set
   a. Data cleaning
   b. Data integration
   c. Data transformation
   d. Error correcting
   e. Data model building

5. Perform the following operations using Python on the Air quality data set
   a. Data cleaning
   b. Data integration
   c. Data transformation
   d. Error correcting

6. Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs for Heart disease dataset.(Charts : Line chart, Barplot, Heatmap, Scatterplot, histogram, boxplot, violin, timeseries chart)

7. Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs for tips dataset.(Charts : Line chart, Barplot, Heatmap, Scatterplot, histogram, boxplot, violin, timeseries chart)

8. Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs for airquality dataset.(Use Air_quality_forvisualization.csv)(Charts : Line chart, Barplot, Heatmap, Scatterplot, histogram, boxplot, violin, timeseries chart)

9. Perform the following data visualization operations using Tableau on (Superstore dataset)
    - **a.** 1D (Linear) Data visualization
    - **b.** 2D (Planar) Data Visualization
    - **c.** 3D (Volumetric) Data Visualization
    - **d.** Temporal Data Visualization
    - **e.** Multidimensional Data Visualization
    - **f.** Tree/ Hierarchical Data visualization

10. Perform the following data visualization operations using Tableau on (Adult Dataset/Iris Dataset)
    - **a.** 1D (Linear) Data visualization
    - **b.** 2D (Planar) Data Visualization
    - **c.** 3D (Volumetric) Data Visualization
    - **d.** Multidimensional Data Visualization
    - **e.** Tree/ Hierarchical Data visualization
    - **f.** Network Data visualization

11. Create a review scrapper for any ecommerce website to fetch real time comments, reviews, ratings, comment tags, customer name using Python.