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 from the Internet and process it on Hadoop platform.

2.Write HiveQL for Flight Reservation

1. Creating, Dropping, and altering tables.
2. Insert values in tables/ Load the data from flight dataset
3. Display data from table
4. Create index.
5. Join

**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 Facebook metrics data sets

**a.** Create data subsets

**b.** Merge Data

**c.** Sort Data

**d.** Transposing Data

**e.** Shape and reshape Data

**5.** 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

**6.** Perform the following operations using Python on the Air quality and Heart Diseases data sets

**a.** Data cleaning

**b.** Data integration

**c.** Data transformation

**d.** Error correcting

**e.** Data model building

**7.** Perform the following operations using Python on the Air quality and Heart Diseases data sets

**a.** Data cleaning

**b.** Data integration

**c.** Data transformation

**d.** Error correcting

**e.** Data model building

**8.** Perform the following operations using Python on the Air quality and Heart Diseases data sets

**a.** Data cleaning

**b.** Data integration

**c.** Data transformation

**d.** Error correcting

**e.** Data model building

9. Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs

Use dataset : Air quality and Heart Diseases

10. Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs

Use dataset : Air quality and Heart Diseases

11. Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs

Use dataset : Air quality and Heart Diseases

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

13. Perform the following data visualization operations using Tableau on Adult and Iris datasets.

**a.** 1D (Linear) Data visualization

**b.** 2D (Planar) Data Visualization

**c.** 3D (Volumetric) Data Visualization

&

Perform the following operations using Python on the Air quality data sets

**a.** Data cleaning

**b.** Data integration

14. Perform the following data visualization operations using Tableau on Adult and Iris datasets.

**a.** 1D (Linear) Data visualization

**b.** 2D (Planar) Data Visualization

**c.** 3D (Volumetric) Data Visualization

&

Perform the following operations using Python on the Facebook metrics data sets

**a.** Create data subsets

**b.** Merge Data

15.Write HiveQL for Flight Reservation

1. Creating, Dropping, and altering tables.
2. Insert values in tables/ Load the data from flight dataset
3. Display data from table
4. Create index.
5. Join

16.Write HiveQL for Flight Reservation

1. Creating, Dropping, and altering tables.
2. Insert values in tables/ Load the data from flight dataset
3. Display data from table
4. Create index.
5. Join

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 from the Internet and process it on Hadoop platform.

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