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 process it using a pseudo distribution mode on Hadoop platform.
2. Design a distributed application using MapReduce(Using Java) which processes a temperature file containing temperature per year. Display the year with maximum temperature.
3. Write an application using HiveQL for flight information system which will include ( database 1. FlightDB)
4. Creating Two Tables in FlightDB 1. FlightDetails (Fields 1. FlightID 2. Source 3. Destination 4. FlightRate 2.FlightTime (Fields 1. FlightID 2.DepartureDate 3. DepartureTime 4. DepartureDelay (in Hours))

b. Show Dropping, and altering Database tables.

**c.** Creating an external Hive table. Show how it works?

**d.** Load table with data stored in csv file

**e.** Find the average departure delay per day in year 2020.

1. Write an application using HiveQL for student information system which will include (Create Database StudentDB)

a.Add two tables 1. StudentDetails (Fields : RollNo,Name,Address,Class) 2. StudentPerformnce( Fields 1. RollNo, 2.Sub1\_marks 2.Sub2-Marks 3. Sub3\_Marks 4.Percentage)

b. Show alter table/column command on tables from same database

**c.** Creating an external Hive table and show how it works?

**d.** Load table with data stored in csv file.

e. Display name of students having percentage > 60

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

**a.** Create data subsets ( Minimum 3 ways)

**b.** Merge Data (Apply on FB dataset ,perform concat and merge operation)

**c.** Sort Data

**d.** Transposing Data

**e.** Shape and reshape Data

**6.** Perform the following operations using Python on Heart Diseases dataset

**a.** Data cleaning (Handle null-values, outliers)

**b.** Data integration (Perform data integration using sqlite/mysql use any database of your choice):

1. Perform create table

2. Select records from database.

**7.** Perform the following operations using Python on Heart Diseases dataset

**A.** Data cleaning (Handle null-values, outliers)

B Build any classification model on Heart-disease dataset.

1. Visualize the data using Python libraries matplotlib,seaborn by plotting the following graphs on iris dataset.
2. Bar-chart
3. Histogram
4. Pie-Chart
5. Scatter-Plot
6. Box-Plot for different features in dataset.
7. Compare different groups in given dataset.
8. Visualize the data using Python libraries matplotlib,seaborn by plotting the following graphs on titanic dataset.
9. Bar-chart
10. Histogram
11. Pie-Chart
12. Scatter-Plot
13. Box-Plot for different features in dataset.
14. Compare different groups in given dataset.

**10.** Perform the following data visualization operations using Tableau on Adult dataset considering different variables and relationships among them,

**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

**11.** Perform the following data visualization operations using Tableau on SuperStore dataset considering different variables and relationships among them,

**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

1. **A.** Perform the following operations using Python on Air Quality dataset

Data cleaning (Handle null-values,Categorical values)

B. Build Regression model for Weight-Height dataset

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

**a.** Create data subsets ( Minimum 3 ways)

**b.** Merge Data (Apply on FB dataset ,perform concat and merge operation)

**c.** Sort Data

**d.** Transposing Data

**e.** Shape and reshape Data

14.Visualize the data using Python libraries matplotlib,seaborn by plotting the following graphs on titanic dataset.

1. Bar-chart
2. Histogram
3. Pie-Chart
4. Scatter-Plot
5. Box-Plot for different features in dataset.
6. Compare different groups in given dataset.

**15.** Perform the following operations using Python on Heart Diseases dataset

**A.** Data cleaning (Handle null-values, outliers)

B Build any classification model on Heart-disease dataset.