

LIST OF DSBDA LAB PRACTICALS

PR EXAM May 23-24

- 1) Design a distributed application using MapReduce(Using Java) which processes a log file of a system.
- 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. Find the average departure delay per day in 2008.
- 3) Write an application using HiveQL for flight information system which will include
 - a. Load table with data, insert new values and field in the table, Join tables with Hive
 - b. Create index on Flight Information Table
 - e. Find the average departure delay per day in 2008.
- 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 Air quality and Heart Diseases data sets
 - a. Data cleaning
 - b. Data integration
 - c. Data transformation

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

Error correcting

Data model building

7) Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs for Air quality and Heart Diseases data sets.

8) Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs for forest fire dataset.

9) 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

d. Temporal Data Visualization

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

Multidimensional Data Visualization

Tree/ Hierarchical Data visualization

Network Data visualization

11) Create a Review Scraper for any e-commerce website to fetch real time comments reviews, ratings, comment tags, customer Name using Python.