

**II SEMESTER EXAMINATION, 2022 – 23**  
**Year: 1<sup>st</sup>, Programme: M.Tech,**  
**Branch: Computer Science & Engineering**  
**Subject: Data Preparation and Analysis**

Duration: 3:00 hrs

Max Marks: 100

*Note: - Attempt all questions. All Questions carry equal marks. In case of any ambiguity or missing data, the same may be assumed and state the assumption made in the answer.*

|      |   |          |
|------|---|----------|
| Q 1. | Answer any four parts of the following.<br>a) Explain the Scalability issues in data preparation.<br>b) Write the overview of preparing data tables.<br>c) Explain Regression ANOVA.<br>d) Discuss about Big Data and its importance.<br>e) Explain the working of Hadoop.<br>f) Explain predictive analysis with suitable example.               | 5x4=20   |
| Q 2. | Answer any four parts of the following.<br>a) Explain the 4 V's of Big Data.<br>b) Describe about Converting Continuous Data to Categories.<br>c) What are the data cleaning methods?<br>d) Discuss about EDA.<br>e) Explain Data Visualization with suitable example<br>f) Differentiate between correlation and simple linear regression.       | 5x4=20   |
| Q 3. | Answer any two parts of the following.<br>a) Explain creating the components of Hadoop Map reduce jobs<br>b) Discuss the installation of and running Hive QL.<br>c) Explain Oracle Big Data in detail.  | 10x2= 20 |
| Q 4. | Answer any two parts of the following.<br>a) Explain investigating the Hadoop Distributed File System Selecting appropriate execution modes: local, pseudo-distributed, fully distributed.<br>b) Explain Inter- and Trans-Firewall Analytics also explain information management<br>c) Describe with an example of Geolocated data visualization. | 10x2= 20 |
| Q 5. | Answer any two parts of the following.<br>a) Explain how to deal with the missing data in data cleaning process?<br>b) Explain distributing data processing across server farms in detail with example<br>c) Explain how to Visualize similarities between social network groups using multidimensional scaling (MDS)                             | 10x2= 20 |

\*\*\*\*\*