

B. Tech. (Computer Science & Information Technology)

8th Semester, (G-Scheme)

Examination, December-2023

**BIGDATA ANALYSIS**

**Paper– PCC-CSE-404G/PCC-IT-403-G**

Time allowed : 3 hours]

[Maximum marks : 75

*Note:- Attempt five questions in total. Question no. 1 is compulsory and select one question from each unit. All questions carry equal marks.*

1. Write short note on the following: 6×2.5=15
- (i) Data Sciences
  - (ii) Data Lakes
  - (iii) RDMS vs. Data warehouse
  - (iv) Data ingestion
  - (v) CRUD operation
  - (vi) Data query and retrieval

**Unit-I**

2. (a) What is Big Data? Explain various characteristics, challenges and applications of Big Data. 8.
- (b) Explain six V's of Big Data with example. 7
3. (a) What is Hadoop? Explain its architecture and components. 8
- (b) What are the various steps involved in data science process? 7

**Unit-II**

4. (a) Define Data Pipeline. Explain different types of Data Pipelines with example. Also advantages and disadvantages of Data Pipelines. 8
- (b) Explain different types of NoSQL databases with example. Differentiate SQL and NoSQL with example. 7
5. Explain various type of data repositories used for big data in detail. 15

**Unit-III**

6. (a) Explain various aspect of big data management system. 8
- (b) Write short note on data operations and real life application of big data. 7
7. What is Modeling? Explain various type of data models with example. 15

**Unit-IV**

8. What is Hive? Explain Hive features, Hive integration and workflow, architecture and Hive query language (HQL) in detail. 15
9. Write short note the following:
  - (a) Big data analytical operations. 8
  - (b) Pig and HDFS Pig commands to process big data. 7