Manipal Institute of Technology



Time: 3 Hours

MANIPAL

Constituent Institute of Manipal Academy of Higher Education

VI SEMESTER B. TECH. (COMPUTER SCIENCE AND ENGINEERING) END SEMESTER QUESTION PAPER, MAY 2022

SUBJECT: ELECTIVE I - BIG DATA MODELLING AND MANAGEMENT [CRA 4055] REVISED CREDIT SYSTEM

(19/05/2022)

Max. Marks: 50

(4)

Instructions to Candidates:

• Answer **ALL** questions & missing data may be suitably assumed.

Q1. Explain the importance of exploring data. Briefly explain the methods to perform preliminary analysis of the data. (4)

Q2. How Big Data can be used in Precision Medicine? Explain (3)

Q3. Discuss the two new opportunities that have contributed to the launch of the big data era. 4(3)

Q4. What are the goals of data modelling and management? Also, explain the three components of data model. (5)

Q5. Illustrate how arrays can serve as a data model. (3)

Q6. Write a short note on Graph data model. (2)

Q7. Show with the necessary steps the creation of Term Frequency Matrix, Inverse Document Frequency, tf-idf matrix for the 3 documents shown below.

D1: "new york times" D2: "new york post" D3: "Ios angeles times"

Q8. What is a Distributed File System? How does it enhance overall system performance? (3)

Q9. Write a short note on Commodity Clusters (3)

Q10. Differentiate between "schema on write" and "schema on read". (2)

Q11. Explain data streaming. List down the required characteristics of data streaming systems and challenges associated with it. Elaborate on why real time dynamic streaming data processing is an important part of today's big data applications. (5)

- Q12. What are the two components of Hadoop Distributed File System? Explain in detail. (3)
- Q13. Explain the desirable properties of big data management systems? (4)
- Q14. Explain CAP theorem and its relevance with respect to big data (3)
- Q15. Explain ACID properties and BASE properties. (3)