

Exam Date & Time: 22-Jun-2024 (02:30 PM - 05:30 PM)



## MANIPAL ACADEMY OF HIGHER EDUCATION

SIXTH SEMESTER B.TECH END SEMESTER MAKE UP EXAMINATIONS, JUN 2024  
**BIG DATA MODELLING AND MANAGEMENT [CRA 4055]**

**Marks: 50**

**Duration: 180 mins.**

**A**

**Answer all the questions.**

Section Duration: 180 mins

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

- 1) How do variety, valence, velocity, and veracity acts as an important characteristic of big data? Explain each with a suitable example. (4)
  - A)
  - B) How is data science defined? Discuss on the five distinct components that can be defined as the P's of data science? (3)
  - C) Identify the various steps involved in the data science process and explain the role of each in data analysis process. (3)
- 2) Explain Hadoop Ecosystem in brief with a neat block diagram. (5)
  - A)
  - B) Suppose there is an image file which contains the map of a country, by exploring the array data model of an image, how can we extract the various location from the image? Demonstrate how to identify whether some location is in ocean or not? (3)
  - C) What is the benefit of commodity clusters? Draw a simple commodity cluster architecture. (2)
- 3) List various management and processing challenges for streaming data. Explain with examples for each (4)
  - A)
  - B) Compare and contrast data in motion and data at rest with examples for both. (4)
  - C) Describe the Connectedness property and connected components of a graph with the help of a graph. (2)

- 4) Explain the advantages of DBMS. (5)
- A)
- B) Illustrate how Aerospike handles transaction? (3)
- C) Explain column store in Vertica. (2)
- 5) What is a data lake? Explain. Give the data lake definition as given by Pentaho Corporation's CTO James Dixon. (4)
- A)
- B) Differentiate parallel and distributed database system. (2)
- C) Explain in brief the benefits of big data processing with respect to real-world scenarios. (4)

-----End-----