

Exam Date & Time: 25-Jun-2024 (02:30 PM - 05:30 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

VI SEMESTER B.TECH MAKE UP EXAMINATION, JUNE 2024

BIG DATA INTEGRATION AND PROCESSING [CRA 4056]

Marks: 50

Duration: 180 mins.

A

Answer all the questions.

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

- 1) Illustrate four different data models popularly used in the field, by specifying an example database application for each of the model in detail. (4)
 - A)
 - B) Differentiate between SQL Select-From-Where query primitive and MongoDB find() primitive with suitable examples. (4)
 - C) Explain the significance of Data Fusion and its uses in BDMS. (2)
- 2) Explain the structure of JSON data. Describe query operations in MongoDB to find documents matching a regular expression pattern and to find documents with values greater than or equal to a specified value. (4)
 - A)
 - B) Describe aggregation functions in MongoDB and assess the significance of the \$match and \$group stages in the aggregation pipeline. (4)
 - C) Elaborate on the implementation in pandas to facilitate the combination of two Data Frames. (2)
- 3) Describe the concept of probabilistic mediated schema with an example. (3)
 - A)
 - B) Demonstrate data exchange problem with a suitable example. (5)
 - C) Outline the queries one would like to ask for integrating multichannel customer analytics. (2)
- 4) Explain in detail about Spark stack significance, layers and components involved. (4)
 - A)

- B) Define data parallel. Explain “split->do->merge” bigdata pipelines with an example. (3)
- C) Illustrate hybrid data processing architecture. Compare and contrast bigdata processing systems. (3)
- 5) Outline the difference between the take action and the collect action in Spark. When would you prioritize using take over collect?. (3)
- A)
- B) Compare Narrow and Wide Transformations. (3)
- C) Summarize the creation and process execution in reading streammed data. (4)

-----End-----