

Question Paper

Exam Date & Time: 04-Jun-2024 (10:00 AM - 01:00 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Second Semester Master of Engineering - ME (Cloud Computing) Degree Examination (Makeup Exam) - June 2024

Data Streaming and Visualization [CDC 5203]

Marks: 100

Duration: 180 mins.

Tuesday, 04 June 2024

Answer all the questions.

- 1) Explain how Volume, Velocity, Verity and Veracity influence Big Data. (CO1) (BL 2) (10 marks) (10)
- 2) List the properties required for Batch layer and Speed layer databases. Explain the responsibilities of Batch and Speed layer in lambda architecture. (CO1) (BL 2) (5 + 5 marks) (10)
- 3) With neat diagram, explain the architecture of HDFS. (CO2) (BL 3) (10 marks) (10)
- 4) Define Data Stream Processing. State the goals for stream processing. List the characteristics of stream processing. (CO2) (BL 3) (2 + 4 + 4 marks) (10)
- 5) Text file contains student name, mobile number, email-id, Institute name and Program name. One record is stored in each row. Data is stored in the same sequence. Mobile number and email-id are optional. Write Spark program to compute number of students belonging to each Program and count of students in each institute. (CO2) (BL 4) (10 marks) (10)
- 6) Design python application to scrape movie data from website <https://www.imdb.com/search/title>. Data need to be scraped are for Indian Hindi movies. It should include Movie Title, year of release, IMDB rating and metascore. All movie details are contained in side **div** tag with **id = "movie Container"**. Title is in **h2** tag, Year in **h3** tag, rating in strong tag and metascore if present, inside **span** tag with **class_ = "metascore favorable"**. Store the scraped data in csv file. (CO3) (BL 4) (10 marks) (10)
- 7) Design python applicaton to scrape book data from http://books.toscrape.com/catalogue/category/books/fantasy_19/index.html. Data may be present is multiple pages. "Next" button is provided to move to next page. Tag name of "Next" is li with attribute name 'next'. Data need to be scraped are "Book Title", "Author" and "Price". Tags for Title, Author and Price are **h2**, **h3** and **h4**. (CO3) (BL 4) (10 marks) (10)
- 8) Differentiate between
a) Explorative and Explanatory Visualization
b) Infographics and Data Visualization. (CO4) (BL 4) (5+5 marks) (10)
- 9) Briefly explain any five simple rules for better visualizations. (CO4) (BL 4) (10 marks) (10)
- 10) Design python application to generate do-nut chart showing the popularity of programing languages like Java, Python, C and C#. Use suitable values for each languages used. (CO4) (BL 4) (10 marks) (10)

-----End-----