Question Paper

Exam Date & Time: 24-Apr-2024 (02:00 PM - 05:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Second Semester Master of Engineering - ME (Cloud Computing) Degree Examination - April / May 2024

Data Streaming and Visualization [CDC 5203]

Marks: 100 **Duration: 180 mins.**

Wednesday, April 24, 2024

Answer all the questions.

1)	Discuss any four important characteristics of Big Data. (CO1) (BL 2) (10 marks)	(10)
2)	List the different layers of lambda architecture. Discuss the purpose and responsibilities of each layer in lambda architecture. (CO1) (BL 2) (1 + 9 marks)	(10)
3)	Define Check Points. Explain with sketch how edit logs along with Check Points helps to maintain high availability in HDFS. (CO2) (BL 3) (2+ 8 marks)	(10)
4)	Discuss why Spark is called lazy evaluation. With diagram explain the different phases of Spark program execution from dataset to code generation. (CO2) (BL 3) (2 + 8 marks)	(10)
5)	Develop Hadoop map reduce application (required to write map and reduce class codes in Java) to find number of boys and girls students from given Institute. Institute name is passed as command line argument. Dataset contains fields like Institute name, Student name, Gender, Program and CGPA. (CO2) (BL 3) (10 marks)	(10)
6)	Design python application to scrape job data related to "Cloud Computing" from https://better-opportunities/technical. Data to be scrapped are job description and link to apply. All job listings are posted inside tag **div** and **id as "Tech-Jobs".** Job description is inside tag **p**. The text corresponding to link to apply jobs is "Apply". Store the scrapped data in pandas dataframe. (CO3) (BL 4) (10 marks)	(10)
7)	Design python application to scrape Universities data from https://en.wikipedia.org/wiki/List_of_universities_in_India. Data need to be scraped are "State Name", "Central", "State", "Deemed" ,"Private" and "Total". **Class_** name for table to scrape data is **"wikitable sortable"**.	(10)
8)	With examples explain different classification of Visualization. (CO4) (BL 4) (10 marks)	(10)
9)	Explain the following rules for better visualizations:**Do Not Trust the Defaults, Use Color Effectively** and **Do Not Mislead the Reader** (CO4) (BL 4) (10 marks)	(10)
10)	Design python application to create subplots with four charts. Four charts need to be drawn are, line graph, bar chart, scatter plot and pie chart. Assume suitable data for each chart. Individual chart should have proper title. (CO4) (BL 4) (10 marks)	(10)

-----End-----