Question Paper

MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal

First Semester Master of Engineering - ME (Big Data Analytics) Degree Examination - November / December 2023

Architecture of Big Data Systems [BDA 5102]

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

Friday, December 1, 2023

Answer all the questions.

1)	List and explain the major challenges of Big Data Systems. (CO1) (10 marks)	(10)
2)	Differentiate between batch and speed layers. List the requirements and responsibilities of speed layer. (CO1) (5+5 marks)	(10)
3)	Explain how HDFS ensures high availability of data and services. (CO2) (10 marks)	(10)
4)	List the use and responsibilities of following tools of Hadoop eco-system HDFS FLUME SQOOP HIVE (CO2) (2.5 * 4 marks)	(10)
5)	Explain different types of failures in Map-Reduce jobs. (CO2) (10 marks)	(10)
6)	Describe how Spark manages the Cluster and Resources. (CO3) (10 marks)	(10)
7)	What are transformation and Actions in Spark? Give suitable examples. (CO3) (6+4 marks)	(10)
8)	Explain Micro-Batch Stream Processing method. List its advantages and disadvantages. Write short notes on water marking and output modes in spark streaming. (CO3) (3+3+4 marks)	(10)
9)	Write map reduce code for computing word frequency and word length from a given file. (CO4) (10 marks)	(10)
10)	The fields in bank dataset are 'Bank ID', 'Account Number', 'Transaction Date', 'Transaction Type' (credit or debit) and 'Transaction Amount'. Date format is dd-MM-yyyy. Each fields are separated by 't'. Solve the following using Spark RDD a. Count unique number of Bank ID b. Count unique number of customers per Bank ID c. Number of credit transactions for given Account Number in a given year (CO4) (3+3+4 marks)	(10)

-----End-----