U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

October 2024 Supplementary Examinations

Programme: B.E. **Semester: VI**
Branch: Computer Science and Engineering **Duration: 3 hrs.**
Course Code: 22CS6PEBDA **Max Marks: 100**
Course: Big Data Analytics

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	CO	PO	Marks
1	a)	Identify the primary issues associated with managing and analyzing unstructured data, and suggest strategies that could effectively deal with the such data.	CO 1	PO1	10
	b)	Construct the analytics flow for a weather data analysis application, detailing the sequential steps involved in extracting valuable insights from meteorological data.	CO	PO1	10
		UNIT - II			
2	a)	Analyze the implications of the CAP theorem influence the selection and implementation of NoSQL databases and evaluate the trade-offs involved in engineering such systems.	CO1	PO1	5
	b)	Imagine you are developing a social media platform focused on music where users can share songs, comment on tracks, and like their favorite tunes. Each song may have multiple comments and likes. How would you devise an optimized MongoDB schema to manage this platform effectively?	CO2	PO2	5
	c)	Perform the following DB operations using MongoDB. i) Create a collection "Student" with the following attributes Rollno, Age, ContactNo, Email-Id. ii) Insert appropriate values. iii) Write query to update Email-Id of a student with rollno 10. iv) Replace the student's name from "ABC" to "FEM" of rollno 11. v) Drop the table.	CO2	PO2	10
		UNIT - III			
3	a)	Illustrate the process of read and write operations in Cassandra using a clear diagram.	CO1	PO1	10
	b)	Write CQL queries to manage a database for a bookstore:	CO2	PO2	10

			a. Create a keyspace named Bookstore with suitable replication settings. b. Create a table named Bookswith the following attributes: BookID, Title, Author, Genre, and PublicationDate. c. Insert two records into the Bookstable. d. Update the table to change the Genre of the book with BookID ISBN123456 to Science Fiction. e. Sort the books in ascending order of their titles.			
			UNIT - IV			
	4	a)	Imagine you are managing a data storage system for a research institution. The system stores various types of data, including research papers, experimental results, and administrative documents. Discuss how you would design a data storage solution using HDFS (Hadoop Distributed File System) to efficiently manage and organize this diverse range of data. Outline the steps you would take to set up the HDFS environment, including the creation of directories, list the complete directories and files, file copying, managing access control lists, and organizing files within the system.	***CO2***	***PO2***	**10**
		b)	Write a map and reduce program to count the number of words in a given input text file.	***CO2***	***PO2***	**10**
			OR			
	5	a)	Describe the architecture of YARN along with its components, using a well-organized diagram.	***CO1***	***PO1***	**10**
		b)	Write a map and reduce program to find out sum of Even and Odd numbers for the given input text file.	***CO3***	***PO3***	**10**
			UNIT - V			
	6	a)	Write a Scala program that takes an integer input from the user and calculates the factorial of that number using recursion. Ensure the program handles non-positive integer inputs gracefully, providing appropriate error messages.	***CO3***	***PO3***	**10**
		b)	Examine and contrast the features, capabilities, and performance of Hadoop and Spark.	***CO1***	***PO1***	**5**
		c)	Demonstrate the use of the following transformation function with a suitable example: map(), flatmap(), filter(), uinion() and distinct().	***CO2***	***PO2***	**5**
			OR			
	7	a)	Discuss how Spark unified stack enhance data processing capabilities in modern analytics architectures?	***CO1***	***PO1***	**10**
		b)	Develop a solution to calculate the frequency of each word in a file and generate a list of words appearing more than four times using RDD and FlatMap in Spark.	***CO3***	***PO3***	**10**

******