U.S.N.										

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

Autonomous Institute Affiliated to VTU

June 2025 Semester End Main Examinations

Programme: B.E. **Semester: VI**

Branch: Computer Science and Engineering **Duration: 3 hrs.**

Course Code: 23CS6PCBDA /22CS6PEBDA /20CS6PEBDA **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.

			CO	PO	Marks
		UNIT - I			
1	a)	Discuss the categories of digital data with examples, and analyze how computer program manages to understand each of this data.	CO1	PO1	**8**
	b)	Big data has provided more opportunities for business. Analyze the added advantages provided by big data over traditional business intelligence approach.	CO2	PO2	**8**
	c)	When a set of digital data is provided to you, what features of digital data has to be looked into, in order to identify if the data is small data or big data.	CO2	PO2	**4**
		OR			
2	a)	Different forms of analytics may provide varying amounts of value to a business, Discuss various types of analytics along with its goals.	CO1	PO1	**8**
	b)	Analyze the tools and frameworks from big data stack that can be used for Weather Data Analysis and explain the flow with a neat diagram.	CO2	PO2	**8**
	c)	Define Big Data. List out the sources from which big data gets generated.	CO1	PO1	**4**
		UNIT - II			
3	a)	Discuss the motivation for the development of NoSQL. Illustrate with a neat diagram and example the different types of NoSQL database.	CO1	PO1	**7**
	b)	Consider the below "Users" document: { "_id": ObjectId("..."), "name": "John Doe", "age": 28,	CO1	PO1	**7**

			"email": "john@example.com", "city": "New York", "interests": ["music", "travel", "sports"], "isActive": true, "createdAt": ISODate("2024-02-01T10:00:00Z") } Write the MongoDB queries for the following: a. Find users who do not have an email field b. Find the 3 youngest users c. Delete users created before January 1, 2024 d. Aggregate the average age of users grouped by city			
		c)	Write SQL and MongoDB queries for the following. a. Find all users who are older than 25 b. Update a user's email by name c. Delete users younger than 18	*CO1*	*PO1*	**6**
			OR			
	4	a)	Explain the CAP theorem, with a neat diagram. Assume you are designing a messaging app where message delivery is critical, but you can tolerate slight delays. Should you prioritize consistency or availability? Why?	*CO1*	*PO1*	**7**
		b)	You are managing a MongoDB collection named Orders, which stores customer purchase records. Each document contains fields like customerName, orderDate, items (an array of purchased items), totalAmount, and status (e.g., "shipped", "pending"). Write MongoDB queries to do the following: 1. Count how many orders have a status of "shipped". 2. Retrieve the latest 5 orders sorted by orderDate in descending order. 3. Find the top 3 highest-value orders by totalAmount. 4. Use the aggregation framework to calculate the total sales (sum of totalAmount) grouped by status	*CO2*	*PO2*	**7**
		c)	Consider a MongoDB collection named sales where each document contains the fields product, quantity, and price. Write an aggregation query to calculate the total revenue, i.e., quantity × price, generated for each product.	*CO2*	*PO2*	**6**
			UNIT - III			
	5	a)	Explain any seven key features of Apache Cassandra that contribute to its performance and scalability in distributed database environments.	*CO1*	*PO1*	**7**
		b)	Write CQL commands for the following: i. Create a user_profile table with the following fields : user_id(VARCHAR,primary key), name(TEXT),	*CO2*	*PO2*	**7**

			emails(SET), phone_numbers(LIST), preferences(MAP<TEXT,TEXT>) ii. Insert a user with the following data: name = 'Alice', emails = {'alice@example.com'}, phone_numbers = ['+1234567890', '+0987654321'], preferences = {'theme': 'dark', 'language': 'en'} iii. Append a phone number to the phone_numbers list: iv. Replace the first phone number in the list.			
		c)	Create a column family LIBRARY with the following fields: Book_ID int, Book_Name text, Student_Name text, Book_taken_count counter Demonstrate the usage of counters by adding the necessary primary key into the table, insert the required values and display the student names who have taken a book more than once.	*CO3*	*PO3*	**6**
			OR			
	6	a)	Compare Hadoop and traditional RDBMS by explaining any seven key differences between them.	*CO2*	*PO2*	**7**
		b)	Create a Messaging application with the following fields: id int PRIMARY KEY, message text, message_by text, time timestamp. Insert appropriate values and show the usage of TTL.	*CO3*	*PO3*	**7**
		c)	You are developing a sports event tracker. Create a column family called Cycling_Calendar to store information about cycling races. Each race has: • race_id (int) - unique identifier • race_name (text) • race_start_date (timestamp) • race_end_date (timestamp) Write queries to show the following by inserting appropriate values: 1. Automatically remove races from the table, 1 day (86400 seconds) after their race_end_date. 2. Keep track of how many times the race information has been viewed using a counter in a separate table.	*CO2*	*PO2*	**6**
			UNIT - IV			
	7	a)	With neat diagram provide the file read operation in Hadoop Distributed File Systems.	*CO1*	*PO1*	**6**
		b)	Given a large dataset consisting of multiple text files stored in a distributed file system like HDFS, design a MapReduce program to compute the number of occurrences of each distinct word across the entire dataset.	*CO3*	*PO3*	**8**

		c)	Explain the importance of Name node and Secondary name node in HDFS architecture	*CO1*	*PO1*	**6**
			OR			
	8	a)	Discuss the key considerations followed in the development of Hadoop framework. Explain Hadoop components, considering master slave structure with a neat diagram.	*CO1*	*PO1*	**6**
		b)	Design a MapReduce program to compute the average temperature per year. Explain the mapper and reducer tasks. Assume the input is the form of (Date, Location, Temperature). Sample input is given below: 2015-01-03,New York,5 2015-02-10,New York,-2 2016-03-20,Los Angeles,18 2015-07-25,New York,30 2016-08-11,Los Angeles,35	*CO3*	*PO3*	**8**
		c)	Differentiate between Fair scheduler and Capacity scheduler in Hadoop system	*CO2*	*PO2*	**6**
			UNIT - V			
	9	a)	Write a scala program to perform the following: 1. Create an array buffer fruits with the elements "Apple","Banana", "Mango" and "Orange" . 2. Append ("Strawberry","Pinapple") to the end of the arraybuffer. 3. Sort the array buffer in descending order. 4. Remove last two elements from the arraybuffer. 5. Convert the arraybuffer to an array	*CO2*	*PO2*	**7**
		b)	Point out the constraints applied while using fold() and reduce() functions in RDDs. Demonstrate how aggregate() operation is beneficial in overcoming the usage of this constraint with an example program	*CO2*	*PO2*	**7**
		c)	Discuss the importance of transformations in Spark and explain the types with examples.	*CO1*	*PO1*	**6**
			OR			
	10	a)	Design a scala function to perform Search of an item in the given collection. The arguments passed to the function are the collection of items and item to be searched. Assume the type of the argument for the function suitably. The return type is to be Boolean.	*CO3*	*PO3*	**7**
		b)	Write Spark code that will read a file with comma separated numbers and prints the sum of all the numbers.	*CO2*	*PO2*	**7**
		c)	Illustrate with a neat diagram the architecture of Spark and explain how an application is executed.	*CO1*	*PO1*	**6**

****