# 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: Artificial Intelligence and Machine Learning**    **Duration: 3 hrs.**

**Course Code: 24AM6PCBDA**                    **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.

<div style="writing-mode: vertical">**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.</div>

| | | | | CO | PO | Marks |
|---|---|---|---|---|---|---|
| | | | **UNIT - I** | *CO* | *PO* | **Marks** |
| 1 | a) | | In today's digital age, company ABC seeks to boost customer satisfaction by utilizing Big Data Analytics. Explain how a data specialist would classify the collected customer data to drive business growth and improve services? Justify with suitable example. | *CO1* | *PO2* | **10** |
| | b) | | Describe different types of analytics which provides meaningful insights to the business with an example. | *CO1* | *PO1* | **10** |
| | | | **OR** | | | |
| 2 | a) | | Define Big Data Analytics. Explain the various sources of big data. | *CO1* | *PO1* | **8** |
| | b) | | Describe the various tools and techniques in big data analytics. | *CO1* | *PO1* | **7** |
| | c) | | List and explain the merits and demerits of big data. | *CO1* | *PO1* | **5** |
| | | | **UNIT - II** | | | |
| 3 | a) | | Distinguish between the row oriented and column oriented file formats supported in Big Data Analytics. | *CO2* | *PO2* | **6** |
| | b) | | Describe the Avro's schema evolution capabilities that contributes in efficient data management and interoperability within big data ecosystems. Highlight its advantages over traditional data storage formats. | *CO2* | *PO2* | **6** |
| | c) | | Elucidate the features of Zip and Snappy compression techniques. | *CO2* | *PO2* | **8** |
| | | | **OR** | | | |
| 4 | a) | | Describe the various types of data compression techniques with relevant examples. | *CO2* | *PO1* | **8** |

| | | | | CO2 | PO1 | 8 |
|---|---|---|---|---|---|---|
| | | b) | Pictorially represent Parquet and ORC file formatting techniques. | CO2 | PO1 | 8 |
| | | c) | Compare and contrast Avro and ORC file formats with relevant examples. | CO2 | PO1 | 4 |
| | | | **UNIT - III** | | | |
| 5 | | a) | Depict with a neat diagram the basic interaction of components of Hadoop Architecture. | CO2 | PO3 | 10 |
| | | b) | Elucidate the handshaking process between YARN, MapReduce and HDFS in execution of an application, with a visual representation. | CO2 | PO3 | 10 |
| | | | **OR** | | | |
| 6 | | a) | Illustrate with a neat diagram the mechanism of Hadoop Distributed File System. | CO2 | PO3 | 10 |
| | | b) | Design a program to count the number of occurrences of each word by applying the concept of MapReducer. | CO2 | PO3 | 10 |
| | | | **UNIT - IV** | | | |
| 7 | | a) | With an example, illustrate how different types of joins and sub queries are performed in HiveQL. | CO3 | PO2 | 10 |
| | | b) | With a neat diagram explain the architecture of Hive. | CO3 | PO2 | 10 |
| | | | **OR** | | | |
| 8 | | a) | Given a banking database in HIVE containing a table named transactions with the following schema:<br>　　　transaction_id INT,<br>　　　account_number STRING,<br>　　　transaction_date DATE,<br>　　　transaction_type STRING,<br>　　　amount DECIMAL (10,2)<br>Write a HIVE query to:<br>i. Group the transactions by account_number and calculate the total transaction amount for each account.<br>ii. Sort the results by the total transaction amount in descending order. | CO3 | PO2 | 6 |
| | | b) | Consider two tables in a HIVE database for an employee management system:<br>employees table:<br>　　　employee_id INT,<br>　　　name STRING,<br>　　　department_id INT,<br>　　　salary DECIMAL (10,2)<br>departments table:<br>　　　department_id INT,<br>　　　department_name STRING<br>Write a HIVE query to:<br>i. Join the employees table with the departments table on department_id.<br>ii. Calculate the average salary for each department. | CO3 | PO2 | 8 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | iii. Display the department_name and the calculated average_salary. | | | |
| | | c) | Describe the various features of HIVE. | CO3 | PO1 | **6** |
| | | | **UNIT - V** | | | |
| | 9 | a) | Elucidate how Apache Spark's architecture can handle and process large-scale data from multiple sources. | CO3 | PO2 | **10** |
| | | b) | Describe the mechanism of Apache Zookeeper in cluster monitoring and management in distributed systems. Highlight key features that enable effective coordination and synchronization among cluster nodes. | CO3 | PO2 | **10** |
| | | | **OR** | | | |
| | 10 | a) | Demonstrate the working of "Hello World" in Spark using Java programming. | CO3 | PO2 | **10** |
| | | b) | Illustrate the working of Apache Pig in distributed parallel processing. | CO3 | PO2 | **10** |

**\*\*\*\*\*\***