

(7 pages)

Reg. No. :

Code No. : 8387

Sub. Code : WCAM 32

M.C.A. (CBCS) DEGREE EXAMINATION,
NOVEMBER 2024.

Third Semester

Computer Application – Core

BIG DATA ANALYTICS

(For those who joined in July 2023 onwards)

Time : Three hours

Maximum : 75 marks

PART A — (15 × 1 = 15 marks)

Answer ALL questions.

Choose the correct answer :

1. Which of the following is an example of structured data?
(a) XML files (b) Text documents
(c) SQL databases (d) JSON files
2. The term "big data" was first coined in which era?
(a) 1990s (b) 2000s
(c) 2010s (d) 1980s

3. A data scientist is primarily responsible for which of the following?
(a) Managing servers
(b) Cleaning and analyzing data
(c) Writing SQL queries
(d) Designing websites
4. Which of the following best describes Hadoop?
(a) A type of SQL database
(b) A framework for distributed storage and processing
(c) A programming language
(d) A data visualization tool
5. What is one of the main challenges of distributed computing?
(a) High data storage cost
(b) Complexity in data analysis
(c) Network latency and fault tolerance
(d) Inability to process large volumes of data
6. Which of the following is a characteristic of NoSQL databases?
(a) Fixed schema
(b) SQL-based querying
(c) Horizontal scalability
(d) ACID transactions

7. In MongoDB, a document is equivalent to what in RDBMS?
 (a) Table (b) Row
 (c) Column (d) Database
8. In the MapReduce programming model, what is the primary function of the Mapper?
 (a) Sorting data
 (b) Partitioning data
 (c) Filtering and transforming data
 (d) Aggregating data
9. In MongoDB, which term corresponds to a 'record' in traditional RDBMS?
 (a) Collection (b) Document
 (c) Field (d) Index
10. Which of the following is a supported data type in Hive?
 (a) INTEGER (b) BOOLEAN
 (c) STRING (d) All of the above
11. Which statement is used to create a table in Hive?
 (a) CREATE TABLE.
 (b) INSERT INTO
 (c) SELECT INTO
 (d) UPDATE TABLE
12. Which type of join is NOT supported by Hive?
 (a) Inner join (b) Right join
 (c) Left join (d) None of the above
13. What is the primary component of Pig's architecture?
 (a) Pig Script
 (b) Hadoop Distributed File System
 (c) MapReduce jobs
 (d) Pig Latin
14. In which scenario would Pig be most suitable?
 (a) Real-time sentiment analysis
 (b) Online transaction processing
 (c) ETL operations on large datasets
 (d) Interactive data exploration
15. How does Pig execute jobs on a Hadoop cluster?
 (a) Sequentially
 (b) In parallel
 (c) Using MapReduce only
 (d) Using Spark only

PART B — (5 × 4 = 20 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 250 words.

16. (a) Explain the evolution, characteristics, and challenges of big data.

Or

- (b) Compare and contrast big data analytics with traditional business intelligence (BI).

17. (a) Describe the MapReduce programming model and explain the steps with an example.

Or

- (b) Discuss the role of the NameNode and DataNodes in maintaining data integrity.

18. (a) Write a short note about the concept of indexing and its purpose in MongoDB.

Or

- (b) Demonstrate why data compression is important in MapReduce workflows and describes the types of compression.

19. (a) Cover the primitive data types and complex data types.

Or

- (b) Explain the purpose of UDFs, how they can be created and used.

20. (a) Analyses the relational operators and explain its purpose and usage.

Or

- (b) Describe detail about the Piggybank and its purpose.

PART C — (5 × 8 = 40 marks)

Answer ALL questions by choosing (a) or (b).

Each answer should not exceed 600 words.

21. (a) Classify and explain different types of analytics commonly used in big data environments and mention some of the top analytics tools.

Or

- (b) Define data science and elaborate on the responsibilities in a big data context.

22. (a) Describe the core components of the Hadoop ecosystem also explain its roles and challenges.

Or

(b) Discuss the role of tools like Apache Pig and Apache Hive in simplifying data processing tasks.

23. (a) Explain about the key MongoDB terms and compare these with their counterparts in RDBMS.

Or

(b) Write a short note on the importance of indexing in MongoDB and explain its types.

24. (a) Illustrate the serialization and deserialization in the context of Hive.

Or

(b) Discuss common aggregation functions and their application.

25. (a) Describe what Apache Pig is, its purpose, and its role in big data processing.

Or

(b) List and explain the primitive data types supported by Pig.