

KAMARAJ COLLEGE (Autonomous)

Accredited with A+ Grade by NAAC

(Affiliated to Manonmaniam Sundaranar University, Tirunelveli)

THOOTHUKUDI – 628 003

(6 Pages)

Reg. No:

Question Code No : 25001407

Sub Code : 24PMCA32

PG Degree - End Semester Examinations, November 2025

Third Semester

M.C.A

Big Data Analytics

(For those who joined in July 2024 onwards)

Time : 3 Hours

Maximum : 75 Marks

PART- A (10 × 1 = 10 Marks)

Answer ALL Questions

Choose the correct answer:

1. Which of the following is NOT a type of digital data classification?
(a) Structured data (b) Semi-Structured data
(c) Unstructured data (d) Non-Digital data
2. What is a key difference between traditional business intelligence and big data?

- (a) Traditional BI handles larger volumes of data than big data
 - (b) Big data can process unstructured data, whereas traditional BI mainly deals with structured data
 - (c) Traditional BI uses hadoop, big data does not
 - (d) Big data is only about storing data, not analyzing it
3. Which of the following is a key component of the Hadoop framework?
- (a) Hadoop Distributed File System (HDFS)
 - (b) Structured Query Language (SQL)
 - (c) MySQL Database
 - (d) Oracle Database
4. Identify the one primary difference between RDBMS and hadoop?
- (a) RDBMS is designed for unstructured data, hadoop is not
 - (b) Hadoop supports distributed computing, while traditional RDBMS typically does not
 - (c) RDBMS uses Hadoop YARN for resource management
 - (d) Hadoop is a type of SQL database
5. In MongoDB, which of the following corresponds to a "table" in a traditional RDBMS?

- (a) Document
- (b) Field
- (c) Collection
- (d) Row

6. What is the primary role of the reducer in the MapReduce programming model?

- (a) To aggregate and summarize intermediate data from the Mapper
- (b) To partition the input data
- (c) To filter out unstructured data
- (d) To compress final output

7. Which of the following is a file format supported by Hive?

- (a) RCFile
- (b) CSVX
- (c) JSONL
- (d) XLSX

8. What is the main purpose of partitioning in Hive?

- (a) To reduce storage costs by compressing data
- (b) To split large datasets into manageable parts for efficient querying
- (c) To encrypt sensitive columns
- (d) To convert unstructured data into structured data

9. Which of the following is the scripting language used in Apache Pig?

- (a) Pig Latin
- (b) HiveQL

(c) SQL

(d) MapReduce script

10. What is the main difference between pig and Hive in terms of data flow?

(a) Hive is used for streaming data, while pig is not

(b) Pig follows a procedural data flow, whereas hive uses a declarative approach

(c) Pig supports only structured data, hive supports semi-structured data

(d) Hive is used for real-time processing, pig is for batch processing only

PART - B (5 X 5 = 25 Marks)

Answer ALL Questions choosing either (a) or (b).

Answer should not exceed 250 words.

11. (a) Discuss about the data warehouse and hadoop.

(OR)

(b) Explain in detail about the top analytics tools in big data.

12. (a) Discuss the hadoop distributed file system.

(OR)

(b) What is YARN? And explain about hadoop YARN.

13. (a) Give a Brief note on Mango DB Data Types.

(OR)

(b) Apply searching, sorting and compression in map reduce.

14. (a) Construct about the hive user defined function.

(OR)

(b) Explain the aggregations functions in hive.

15. (a) Discuss about the relational operators in pig.

(OR)

(b) Compare Pig and Hive in big data.

PART - C (5 X 8 = 40 Marks)

Answer ALL Questions choosing either (a) or (b).

Answer should not exceed 600 words.

16. (a) Briefly describe the traditional business intelligence versus big data.

(OR)

(b) Describe about the classifications of digital data.

17. (a) Discuss the difference between SQL and NOSQL.

(OR)

(b) What is distributed computing? And briefly explain the distributed computing challenges.

18. (a) Briefly describe the mongo DB query language.

(OR)

(b) Give a brief note on map reduce.

19. (a) Discuss about the hive query language statements.

(OR)

(b) Describe about serialization and deserialization.

20. (a) Write the different types of pig primitive data types.

(OR)

(b) Describe in detail about HDFS commands in pig.