

N.G.M.COLLEGE (AUTONOMOUS) : POLLACHI

END-OF-SEMESTER EXAMINATIONS : NOVEMBER 2025

B.Sc C. Science With D.A

MAXIMUM MARKS: 75

SEMESTER-V

TIME : 3 HOURS

PART-III

23UDA515– BIG DATA ANALYTICS

SECTION – A

(10 X 1 = 10 MARKS)

ANSWER THE FOLLOWING QUESTIONS.

1. Which of the following is structured data?

- A) Email body B) JSON file
C) Relational database table D) Video file

2. Hadoop Distributed File System (HDFS) stores data in_____

- A) Row B) Column C) Block D) File

3. In MongoDB, documents are stored in_____

- A) Table B) Collection C) Row D) Tuple

4. In Hive, the process of dividing data into parts based on values of a particular column is called_____

- A) Aggregation B) Partitioning C) Joining D) Serialization

5. Pig Latin scripts are written for_____

- A) Data warehousing B) Large-scale data transformations
C) Transaction processing D) Real-time streaming

ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES

6. Define Big Data in one sentence.

7. What does YARN stand for in Hadoop?

8. Mention any two data types supported in MongoDB.

9. What is the role of a SerDe in Hive?

10. Name any one diagnostic operator in Pig.

SECTION – B**(5 X 5 = 25 MARKS)****ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.**

11. a)Classify digital data into structured, semi-structured, and unstructured formats and give suitable real-world examples. **(OR)**
b)Explain any five key challenges that organizations face while handling Big Data.
12. a)Compare the characteristics of SQL and NoSQL databases with suitable examples.. **(OR)**
b)Explain the function of Hadoop YARN in managing resources and scheduling applications.
13. a)Differentiate between the roles of Mapper and Reducer in a MapReduce framework. **(OR)**
b)Briefly explain the different data types supported in MongoDB.
14. a)Write short notes on Hive partitions and bucketing. **(OR)**
b)Explain Hive Views with an example.
15. a)Discuss the main features and the design philosophy of Apache Pig **(OR)**
b)Write short notes on the concept of Piggy Bank and its uses.

SECTION – C**(5 X 8 = 40 MARKS)****ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.**

16. a)Explain the evolution of Big Data and describe how it differs from traditional Business Intelligence.. **(OR)**
b)What is Big Data Analytics? Explain its types of classification and discuss its importance in modern applications..
17. a)Illustrate the working of Hadoop Distributed File System for block management and data recovery. **(OR)**
b)Describe the challenges of distributed computing and explain how Hadoop effectively addresses them.
18. a)Explain the stages of a MapReduce program by using a real-world example application. **(OR)**
b)Contrast embedded documents and references in Mongo DB and explain scenarios where each would be suitable.
19. a)Discuss the architecture of Hive in detail and explain how it enables data warehousing on top of Hadoop. **(OR)**
b) Explain the use of ‘joins’ and ‘GROUP BY’ operations in Hive QL with the help of a transaction data use case.
20. a)Explain in detail the execution modes of Pig. Discuss how Pig runs in Local and MapReduce mode and illustrate with a suitable example. **(OR)**
b)How can Apache Pig be used to preprocess IoT data streams for machine learning applications? Explain with Pig’s primitive and complex data types.
