

CMR ENGINEERING COLLEGE: : HYDERABAD

UGC AUTONOMOUS

III-B.TECH-II-Semester End Examinations (Supply) - June- 2025

BIG DATA ANALYTICS

(AI&DS)

[Time: 3 Hours]

[Max. Marks: 70]

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 20 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART-A**(20 Marks)**

1. a) Which data is known to be Big Data? [2M]
- b) Differentiate RDBMS and MapReduce. [2M]
- c) State the task of nodes in a Hadoop cluster. [2M]
- d) Mention the file permissions in HDFS. [2M]
- e) What is the role of a Job Tracker? [2M]
- f) How Hadoop runs a MapReduce job using YARN? [2M]
- g) Name the execution modes of Pig. [2M]
- h) Why RDD is fault-tolerant in Spark? [2M]
- i) Where do we use HBase? [2M]
- j) List out the types of NoSQL databases. [2M]

PART-B**(50 Marks)**

2. Label the Hadoop Ecosystem and explain the purpose of each component in the Hadoop Ecosystem. [10M]

OR

3. With a neat diagram, Assess the MapReduce data flow with single reduce task and multiple reduce tasks. [10M]

- 4.a) Examine the following statement: HDFS is a filesystem designed for storing very large files with streaming data access patterns, running on clusters of commodity hardware. [5M]
- b) Discuss the areas where HDFS is not a good fit. [5M]

OR

5. Describe the anatomy of a file read in HDFS. [10M]

6. How Hadoop runs a MapReduce job using the classic framework? Describe in detail. [10M]

OR

- 7.a) Analyze the failure of the task, the application master, the node manager, and the resource manager in YARN. [5M]
- b) Demonstrate the procedure for Shuffle and sort in MapReduce [5M]

- 8.a) Provide the high-level comparison of SQL and HiveQL. [5M]
- b) List out and describe the Hive data types with literal examples. [5M]

OR

- 9.a) Inspect the usage of user-defined functions in Pig tool. [5M]
- b) Illustrate the functionality of data processing operators in Pig tool. [5M]

10. Elaborate the features of various types of NoSQL databases. [10M]

OR

11. Infer various client options for interacting with an HBase cluster. [10M]
