

Code No.: DS702PC

R20

H.T.No.

8

R

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS

IV–B.TECH–I–Semester End Examinations (Supply) - April- 2025

BIG DATA ANALYTICS

(CSD)

[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) Define Big Data. [2M]
- b) Why can't we use databases with lots of disks to do large-scale batch analysis? [2M]
- c) Mention the areas where HDFS is not a good fit today. [2M]
- d) State the task of namenode and datanode. [2M]
- e) Name the independent entities of a job run in classic MapReduce. [2M]
- f) Label the general form of map and reduce functions. [2M]
- g) Why are RDDs resilient? [2M]
- h) What is Hive shell? [2M]
- i) Where do we use HBase? [2M]
- j) List out the types of NoSQL databases. [2M]

PART-B

(50 Marks)

2. Examine the functionality for analyzing the data with Hadoop. [10M]
- OR
3. Outline and describe the components of the Hadoop Ecosystem. [10M]
4. How the data flows between the client interacting with HDFS, the namenode and the datanodes? Show and explain the main sequence of events when reading a file. [10M]
- OR
5. Elaborate the practical aspects of developing the MapReduce application in Hadoop. [10M]
6. Show and explain the mechanism of Hadoop running a MapReduce job using the classic framework. [10M]
- OR
7. Analyze different types of input formats processed by Hadoop. [10M]
8. List out and discuss the data processing operators in Pig tool. [10M]
- OR
9. Inspect how to use various forms of the SELECT statement to retrieve data from Hive. [10M]
10. Assess various client options for interacting with an HBase cluster. [10M]
- OR
11. Label the main differences between SQL versus NoSQL databases. Summarize the Pros and cons of SQL and NoSQL. [10M]
