CMR INSTITUTE OF TECHNOLOGY BANGALORE

CBCS SCHEME

USN | | | | | | | | | | |

18CS72

Seventh Semester B.E. Degree Examination, Dec.2025/Jan.2026
Big Data Analytics

Time: 3 hrs. Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

Module-1

1 a. Briefly explain the uses of massive parallel processing and cluster computing in Big data scenario. **(10 Marks)**

b. Justify how big data analytics would benefit the General Electric (GE) organization in a scenario where as it retains large data which is not evaluated or analyzed, will not benefit the organization. **(10 Marks)**

OR

2 a. With a neat block diagram, show sources of structured, semi structured multi structured and unstructured data for finance application. **(10 Marks)**

b. List and explain the functions of each layer in Big data architecture with a neat diagram. **(10 Marks)**

Module-2

3 a. Bring out the salient fractures of Hadoop and explain hadoop ecosystem component with a neat diagram. **(10 Marks)**

b. With a neat diagram contrast the features and applications of HDFS components. **(10 Marks)**

OR

4 a. How does YARN resource manager do the following :
i) Keep track of the active node managers and available resources.
ii) Allocate the containers to the appropriate sub tasks and monitors the application master. **(10 Marks)**

b. Explain Apache Sqoop with import and export method with neat block diagram. **(10 Marks)**

Module-3

5 a. List and explain the differences between NOSQL and SQL databases with suitable example for each. **(10 Marks)**

b. Briefly explain the design considerations when using column – family data stores life Cassandra? List and explain usages of data types built into Cassandra. **(10 Marks)**

OR

6 a. List and explain any ten examples where NOSQL data stores are required. **(10 Marks)**

b. What are the features of Shared Nothing (SN) architecture? How does Big data store SN system Partition? How does a Partition process the different quires? **(10 Marks)**

Module-4

7 a. List and explain the features of the Map Reduce Programming Model? How does Map Reduce program enable parallel processing. **(10 Marks)**

b. Describe the Hive architecture components. Why are Hive QL, SQL – like scripts used in pace of RDBMs, such as MySQL for Big data? **(10 Marks)**

OR

8 a. Explain in detail aggregation commands provisioned in Hive QL. What are the partitioning commands? **(10 Marks)**

b. Discuss in detail the Pig data types and operators with an examples for each :
i) Group ii) Join iii) Filter
iv) Limit v) Order by **(10 Marks)**

Module-5

9 a. Describe the approaches used in Linear multivariate and multiple regression algorithms. **(10 Marks)**

CMRIT LIBRARY
BANGALORE - 560 037

b. List the steps in classifying web documents into categories, identifying similar pages across different web documents to classify them as web pages of a University. **(10 Marks)**

OR

10 a. List the steps for finding the association , association rules, filtering the frequent item sets with suitable examples. **(10 Marks)**

b. Write the steps in performing bipartite weighted graph matching in Social Network graph analysis. **(10 Marks)**

* * * * *