Code No.: R22AI732PE	R22	H.T.No.			8	R						

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
IV–B.TECH–I–Semester End Examinations (Regular) - December- 2025
BIG DATA ANALYTICS
(CSM)

[Time: 3 Hours] **[Max. Marks: 60]**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 10 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 **(10 Marks)**

1. a)	What is data architecture in Big Data systems?	[1M]
b)	List any two data pre-processing techniques.	[1M]
c)	Define HDFS and its purpose.	[1M]
d)	Mention any two advantages of using Spark over Hadoop.	[1M]
e)	What is data analytics cycle?	[1M]
f)	Differentiate between descriptive and predictive analytics.	[1M]
g)	Define training data and testing data.	[1M]
h)	What is model evaluation and why is it important?	[1M]
i)	What are the advantages of visual analytics?	[1M]
j)	List any two charts used for visualization of categorical data.	[1M]

PART-B **(50 Marks)**

2.a)	Explain the concept of data governance and data quality management.	[5M]
b)	Discuss methods for handling duplicate and missing values in datasets.	[5M]
	OR	
3.	Discuss the role of data architecture in Big Data analytics and list the main phases of data management cycle.	[10M]
4.a)	Explain the working of Hadoop and MapReduce framework with a neat diagram.	[5M]
b)	Describe how data is transferred between nodes in Spark.	[5M]
	OR	
5.	Discuss ETL process stages – Extraction, Transformation and Loading – with suitable examples.	[10M]
6.a)	Explain the need for statistical analysis in Big Data.	[5M]
b)	Discuss techniques for detecting and eliminating outliers.	[5M]
	OR	
7.	Illustrate the Big Data analytics life cycle and explain each phase in detail.	[10M]
8.a)	Explain the steps involved in building a predictive model for a dataset.	[5M]
b)	Compare supervised and unsupervised learning algorithms.	[5M]
	OR	
9.	Discuss cross-validation and overfitting issues in machine learning models.	[10M]
10.a)	Explain how visualization enhances data-driven decision making.	[5M]
b)	Demonstrate the process of designing interactive dashboards using Qlik View or Tableau.	[5M]
	OR	
11.	Describe various chart types and their suitability for different data scenarios in visualization.	[10M]

****