R22 Regulation *Subject code:4E7GB*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

(Autonomous, Accredited by NAAC with 'A+' Grade)

B.Tech VII Semester Regular/Supplementary Examinations, October 2026

BIG DATA ANALYTICS

(CSE(AIML))

Maximum Marks: 60 Date: 28.10.2026 Duration: 3 hours

Note: **1.This question paper contains two parts A and B.**
2. Part A is compulsory which carries 10 marks. Answer all questions in Part A.
3. Part B consists of 5 Units. Answer any one full question from each unit.
4. Each question carries 10 marks and may have a, b, c, d as sub questions.

Part-A

All the following questions carry equal marks (10X1M=10 Marks)		Marks	CO	BTL
1.a)	List any two characteristics of Big Data.	1M	1	L1
b)	Name the major types of analysis used in Big Data.	1M	1	L1
c)	Mention the purpose of Load Balancing with Multiple Consumers.	1M	2	L1
d)	List the main elements involved in the Scheduler Agent Supervisor pattern.	1M	2	L1
e)	State the role of a subscriber in Publish-Subscribe messaging.	1M	3	L1
f)	Describe any two examples of Big Data Collection Systems.	1M	3	L1
g)	Identify the Big Data technologies suitable for processing news articles in batch mode.	1M	4	L2
h)	A MapReduce application divides a large task into smaller tasks before processing. Identify the first phase.	1M	4	L1
i)	An application stores customer profiles as JSON-like documents. Which type of NoSQL database is suitable?	1M	5	L1
j)	A social networking application stores relationships between users. Which NoSQL database type is suitable?	1M	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL
2	a) An organization collects temperature, rainfall, humidity, and wind-speed data from weather stations. Explain how a Big Data Stack can be used to collect, process, analyze, and visualize this weather data. b)Explain the role of Data Collection and Data Preparation in the Big Data analytics cycle.	6M 4M	1	L3 L2
	OR			
3	a) A healthcare organization collects patient records from multiple sources. Develop a suitable Big Data analytics flow for this application. b) Explain the characteristics of Big Data with suitable domain-specific examples.	6M 4M	1	L3 L2

4	Describe Load Balancing with Multiple Consumers and its role in distributed systems.	10M	2	L2
	OR			
5	a) A Big Data application receives a large number of requests simultaneously. Select suitable patterns to handle the workload and explain your choice. b) Explain the MapReduce pattern and its major processing stages.	6M 4M	2	L3 L2
6	Describe the working of Apache Flume for collecting data from multiple sources and storing it in a centralized data store.	10M	3	L4
	OR			
7	Describe HDFS architecture and explain its major components.	10M	3	L2
8	Compare Hadoop MapReduce, Pig, Oozie, and Spark based on their roles in batch analysis.	10M	4	L4
	OR			
9	a) Explain Hadoop MapReduce architecture and its working process. b) A system collects thousands of news articles from multiple websites every day. Explain how you would design a batch analysis process to identify frequently occurring topics and keywords.	4M 6M	4	L2 L3
10	a) A social networking application needs to store user relationships. Select a suitable NoSQL database type and justify your choice. b) Explain Graph databases use in real time applications	6M 4M	5	L3 L2
	OR			
11	a) Write a python script to Writing data to DynamoDB table b) Describe the steps involved in converting an ER model into a relational model, including the mapping of entities, attributes, and relationships.	5M 5M	5	L3 L2