

# GYANMANJARI INNOVATIVE UNIVERSITY

GYANMANJARI INSTITUTE OF TECHNOLOGY

M.Tech.-Mid Semester Examination (MSE)-S2026

Enrollment No.: \_\_\_\_\_

Subject Code: METAI12511

Subject Name: Big Data Analytics

Time: 10:30 AM To 12:30 PM

Date: 23-03-2026

Semester: 02

Total Marks: 60

Instructions:

1. Question No. 1 is compulsory.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

	Marks
Q.1 (a) Define big data technologies and explain the role of Hadoop and cloud computing in big data ecosystems.	10
(b) Explain the concept of schema-less databases and compare key-value, document, and aggregate data models with suitable examples.	10
Q.2 (a) Explain the applications of big data analytics in healthcare, finance, and digital advertising. Discuss how big data improves decision making in these domains.	10
<b>OR</b>	
(a) A distributed database system needs high availability and scalability. Design a suitable replication and sharding strategy using master-slave or peer-to-peer models. Justify your choice.	10
(b) Explain Hadoop data flow and demonstrate how serialization and compression improve storage and processing efficiency in distributed systems.	10
<b>OR</b>	
(b) Demonstrate how MapReduce computation can be used to process large-scale data aggregation (e.g., counting frequency of transactions). Explain partitioning and combining steps.	10
Q.3 (a) Analyze the design of Hadoop Distributed File System (HDFS). Explain how it ensures data integrity, scalability, and fault tolerance.	10
(b) Explain the structure of a MapReduce job execution pipeline. Analyze the role of shuffle and sort in improving processing efficiency.	10
<b>OR</b>	
Q.3 (a) Evaluate the advantages and limitations of YARN over traditional MapReduce architecture in large-scale data processing environments.	10
(b) Design a complete MapReduce workflow for analyzing large-scale log data. Specify input format, processing stages, job scheduling, and output format.	10