



R22 Regulation

Subject code:4E6HA

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Supplementary Examinations, November 2025

BIGDATA ANALYTICS

(CSE(DS))

Maximum Marks: 60

Date: 19.11.2025

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	Bloom Tx
1.a)	What are the main challenges of traditional data processing before Big Data emerged?	1M	1	1
b)	List out the characteristics of big data.	1M	1	1
c)	Define the concept of Load Leveling with Queues	1M	2	2
d)	How Lambda architecture helps to handling Big Data?	1M	2	2
e)	List out various methods of Big Data acquisition	1M	3	2
f)	Compare Big Data storage systems and relational data base systems.	1M	3	2
g)	Describe Apache Spark and its significance in batch data processing.	1M	4	1
h)	Differentiate Hadoop and MapReduce in batch analysis	1M	4	4
i)	How can R language be used to analyze big datasets from NOSQL databases?	1M	5	2
j)	When should Nosql databases should be used	1M	5	3

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	Bloom Tx
2	Explain in detail about the features of big data. Compare and contrast the Big Data Stack with traditional data processing systems	10M	1	3
OR				
3	Explain various real-world use cases of Big Data across different industries.	10M	1	2
4	Discuss about MapReduce patterns in Big Data analytics. Explain how MapReduce helps to handling large datasets and processing them in parallel.	10M	2	4
OR				
5	Explain the significance of CAP theorem in Big Data systems. How does it affect data consistency, availability and partition tolerance in distributed systems?	10M	2	5
6	Explain the architecture of HDFS in detail. How does it enable scalable and fault-tolerant storage for Big Data?	10M	3	2

	OR			
7	What is SCALA and how is it used in Big Data analytics with Apache Spark?	10M	3	5
8	Explain how Apache Oozie helps in scheduling and managing workflows in batch processing environments. Why is it crucial for Big Data systems?	10M	4	4
	OR			
9	Compare Apache Spark with Hadoop MapReduce in terms of batch processing. Highlight their strengths and weaknesses in handling large-scale data analytics.	10M	4	4
10	Explain the advantages and disadvantages of using Django as a web framework for Big Data applications. Include key features that make Django suitable for building scalable and secure web applications.	10M	5	4
	OR			
11	a) Compare and contrast tabular stores and object data stores in NoSQL databases. b) Discuss NoSQL databases design, advantages, and specific use cases where they are most effective.	5M 5M	5	4