



R20 Regulation

Subject code: 3P7HA

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VII Semester Regular Examinations, December 2024

BIG DATA ANALYTICS

(CSE (DATA SCIENCE))

Maximum Marks: 70

Date: 04.01.2025

Duration: 3 hours

- Note:**
1. This question paper contains two parts A and B.
 2. Part A is compulsory which carries 20 marks. Answer all questions in Part A.
 3. Part B consists of 5 Units. Answer any one full question from each unit which carries 10M.
 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 (10X2M=20 Marks)		CO	Bloom Tx
1	What are the main components of Big Data Analytics?	1	L1
2	What is visualization in big data?	1	L2
3	What are the different types of MapReduce patterns?	2	L1
4	How does a Column Family Database differ from traditional relational databases?	2	L2
5	What are the various considerations for data acquisition, which will drive the choice of Big Data tools or frameworks?.	3	L2
6	List characteristics of HDFS?	3	L2
7	Explain Map Reduce programming model?	4	L2
8	List Apache Spark Tools?	4	L2
9	What is a NoSQL Database?	5	L1
10	What is the architecture of Django?	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		CO	Bloom Tx
11	A. Identify and Illustrate with an example the big data challenge in the domain of finance and medicine? [5M] B. Describe briefly about 5 V's. [5M]	1	L4 L4
OR			
12	A. Elucidate how the big data stack is used for analysis of weather data? [5M] B. Discuss about analytical flow for big data. [5M]	1	L4 L2
13	A. Describe any two architectural components and design styles for big data systems and applications? [5M] B. Discuss briefly about Map reduce pattern with an example. [5M]	2	L2 L2
OR			
14	A. Explain about load balancing with multiple consumers and leader election with neat diagram. [5M] B. Write Python program for computing inner join with MapReduce? [5M]	2	L2 L3

15	A. Discuss HDFS Architecture ? [5M] B. Describe Publish - Subscribe Messaging Framework? [5M]	3	L2	L2
	OR			
16	A. Elucidate how various considerations for data acquisition will drive the choice of the tools or frameworks? [5M] B. Discuss any one Big Data Collection System? [5M]	3	L4 L2	
17	A. Describe some examples of batch data analysis with Pig? [5M] B. Discuss the Components of Apache Spark cluster? [5M]	4	L2 L2	
	OR			
18	A. Elucidate Hadoop MapReduce workflow for batch analysis of weather data? [5M] B. Explain the Apache Oozie workflow comprising a Hadoop streaming MapReduce job action and Email actions that notify the success or failure of the job? [5M]	4	L4 L3	
19	A. Explain the pros and cons of Non-Relational Databases? [5M] B. Describe an example of using DynamoDB for a reference application that retrieves and stores weather data for different cities? [5M]	5	L2 L3	
	OR			
20	A. Explain the pros and cons of Relational Databases? [5M] B. Compare and Contrast Key-Value Databases, Document Databases, Column Family Databases, and Graph Databases? [5M]	5	L2 L4	