

MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE, MADANAPALLE

(Deemed to be University under Section 3 of UGC Act, 1956)

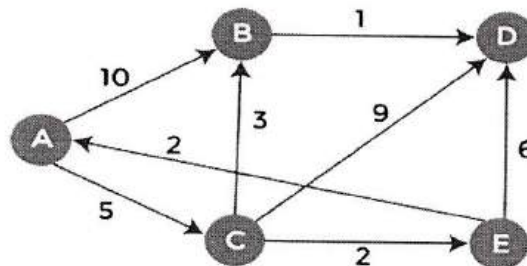
M.Tech. I Year I Semester (R25) Supplementary End Semester Examinations - July 2026
ADVANCED DATA STRUCTURES AND ALGORITHMS

(Computer Science and Engineering)

Time: 3Hrs**Max Marks: 100**

Attempt all the questions. All parts of the question must be answered in one place only.
In Q.no 1 to 5 answer either A or B only

Q.No.	Question	Marks	CO	BL
Q.1(A)	Describe in detail about the asymptotic notations and also identify the complexity class for each of the following expressions: i) x^3+100x^2+50 ii) $7\log(n^3)$ iii) $\sqrt{n} + \log n$ iv) $5 \log_2 64$ v) $3n + 2n$	20	1	3
OR				
Q.1(B)	Illustrate the Recursion Tree Method by Compare with the Substitution Method. Also Solve the following recurrence using a recursion tree: $T(n)=3T(n/2)+n$	20	1	3
Q.2(A)	Insert the following keys into an initially empty Red-Black Tree: 41, 38, 31, 12, 19, 8 i) Show the tree after each insertion. ii) Indicate all rotations and recolouring steps. iii) Verify that Red-Black properties are preserved.	20	2	3
OR				
Q.2(B)	Construct a B-Tree of order 4 by inserting the following keys: 10, 20, 5, 6, 12, 30, 7, 17 i) Demonstrate the search operation for key 17. ii) Count the number of disk accesses required. iii) Analyse the time complexity of the search operation.	20	2	2
Q.3(A)	Assume any graph to perform BFS and DFS starting from a given source vertex. Show the traversal order and BFS tree / DFS forest with in detail illustration	20	3	2
OR				
Q.3(B)	Find the shortest path from source node A, using Dijkstra's algorithm	20	3	3



Q.4(A)	Explain the Matrix Chain Multiplication problem. Derive the DP formulation and analyse its time and space complexity.	20	4	2
OR				
Q.4(B)	Compare Dynamic Programming and Greedy Algorithms with respect to design strategy, optimality, complexity, and applications. Construct a Huffman code for the characters with frequencies: {a:5, b:9, c:12, d:13, e:16, f:45} Calculate the average code length.	20	4	3
Q.5(A)	Explain the concept of Polynomial Time algorithms and examine why polynomial time is considered as efficient and give its role in complexity theory.	20	5	2
OR				
Q.5(B)	Illustrate the steps involved in proving a problem is NP-Complete. Illustrate with a suitable example.	20	5	3

*****END*****

MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE, MADANAPALLE

(Deemed to be University under Section 3 of UGC Act, 1956)

M.Tech. I Year I Semester (R25) Supplementary End Semester Examinations - July 2026**BIG DATA ANALYTICS**

(Computer Science and Engineering)

Time: 3Hrs**Max Marks: 100**

Attempt all the questions. All parts of the question must be answered in one place only.

In Q.no 1 to 5 answer either A or B only

Q.No.	Question	Marks	CO	BL
Q.1(A) i	What are the different types of data? What are the important challenges faced by analysts while handling big data? (5)	5	1	2
ii	What is Hadoop Ecosystem? Briefly explain the role of HDFS, YARN, MapReduce, Hive, and HBase.	5	1	2
iii	Consider a marketing scenario where a garment merchant promotes and sells products through his own e-commerce website and digital marketing campaigns. The website receives about 1 million customer visits (hits) per month from different regions and devices. Identify the different types of marketing data generated from this scenario. What marketing data analytics methods can be applied to analyse this data and improve customer engagement, sales conversion, and overall business performance?	10	1	4
OR				
Q.1(B) i	Explain the challenges of Big Data in terms of Volume, Velocity, Variety, Veracity, and Value.	5	1	2
ii	Explain the different types of analytics.	5	1	2
iii	Assume that you have access to credit card transaction carried out by private banking customer. There are about 50 Lakhs customers who are using the banking credit cards across India. As an analyst, you have given responsibility of the following	10	1	4
	a. Improving the usage of credit card transaction by each customer by 10% every Month.			
	b. Analyze the average month Credit Card usage by customer			
	c. Identify the fraudulent usage of credit card by the customer			
	d. Increasing the number of customers who are going to use the bank credit card.			
	Identify the appropriate analytics method for the above scenario			
Q.2(A) i	If you have an input file of 350 MB, how many input splits would HDFS create and what would be the size of each input split?	5	2	3
ii	Explain the working of Hadoop Distributed File System	10	2	2
iii	Explain how Apache Flume is used to ingest streaming data and how Apache Sqoop is used to import data from relational databases.	5	2	2
OR				
Q.2(B) I	Compare and justify the use of various HDFS interfaces for different application scenarios such as batch processing, real-time access, and remote data management.	5	2	3
ii	Explain the architecture of an Avro file. How does schema evolution work in Avro?	10	2	2
iii	What is Hadoop I/O? Briefly explain the role of compression and serialization in Hadoop.	5	2	2

Q.3(A)	i	What is MapReduce? Explain the roles of Mapper, Reducer, Combiner, and Partitioner	10	3	2
	ii	Consider that you have given data.csv file containing the details of customer credit card spending. Write a Map Reduce program to find frequency of spending by each customer and sum of total spending.	10	3	3
OR					
Q.3(B)	i	What are MapReduce data types? Explain key-value pairs and commonly used Writable data types in Hadoop.	10	3	2
	ii	Design a MapReduce program to find the top 5 most frequently occurring words in a large text dataset. Explain how the Shuffle and Sort phase helps in grouping and ordering the intermediate key-value pairs to achieve the result.	10	3	3
Q.4(A)	i	Explain the role of Pig on Hadoop. When to use or not to use Pig in Hadoop eco system?	10	4	2
	ii	Write and explain HiveQL queries to perform the following operations on a table EMPLOYEE(emp_id, name, dept, salary):	10	4	4
		a) Create the table			
		b) Insert data			
		c) Retrieve employees with salary greater than a given value			
		d) Find department-wise average salary			
OR					
Q.4(B)	i	Explain the relationship between HBase and HDFS and how HBase provides real-time read/write access.	10	4	2
	ii	Write a Pig script for the following	10	4	4
		a. Display first 3 tuples from employee relation			
		b. Display the names of the employees in department wise by descending order of employee salary.			
		c. Remove duplicate tuples of employee.			
		d. Insert new employee record.			
Q.5(A)	i	Define stream data. Explain the characteristics of the stream data model and discuss methods for sampling and filtering streaming data.	10	5	2
	ii	Critically evaluate different stream sampling methods in terms of accuracy, latency, and resource utilization.	10	5	4
OR					
Q.5(B)	i	Explain the concept of estimating moments in data streams. Describe the method for counting ones in a sliding window and explain decaying windows.	10	5	2
	ii	Compare sliding window and decaying window models. Justify the use of decaying windows for real-time analytics applications.	10	5	4

END