MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE, MADANAPALLE
(Deemed to be University under Section 3 of UGC Act, 1956)

M.Tech. I Year I Semester (R25) Regular End Semester Examinations - January 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)	Prove using asymptotic notation that $5n^2 + 3n + 10 = O(n^2)$ using Master theorem. Identify the complexity for the following notations i) $n^2 + 7\log(n^3)$ ii) $2\log_{10} 10^{200}$ iii) 3^{n^4}	20	1	3

OR

Q.1(B)	Define and explain Big-O, Big-Ω, and Big-Θ notations with mathematical expressions and graphs. Also Solve the recurrence using the Recursion Tree Method: $T(n)=2T(n/2)+n$	20	1	3

Q.2(A)	Starting with an empty Red–Black Tree, insert the following keys in the given order: 10, 18, 7, 15, 16, 30, 25 i) Draw the Red–Black Tree after each insertion. ii) For every violation, specify the Red–Black insertion case and the corrective action (recolouring / left rotation / right rotation). iii) Show all rotations clearly with diagrams. iv) Determine the black-height of the final tree.	20	2	3

OR

Q.2(B)	Given the keys: 50, 30, 70, 20, 40, 60, 80 i) Construct the Binary Search Tree by inserting keys one by one. ii) Perform search operations for keys 40 and 90, showing each comparison. iii) Delete key 30 and redraw the BST. iv) Analyse the time complexity of each operation.	20	2	4

Q.3(A)	Explain the Prim's algorithm. Compare Prim's and Kruskal's algorithms in terms of approach, complexity, and applications.	20	3	4

OR

Q.3(B) Differentiate between Depth First Search and Breadth First Search algorithms. Also Performs the Kruskal's algorithm to find the minimal spanning tree for the following graph. 20 3 4
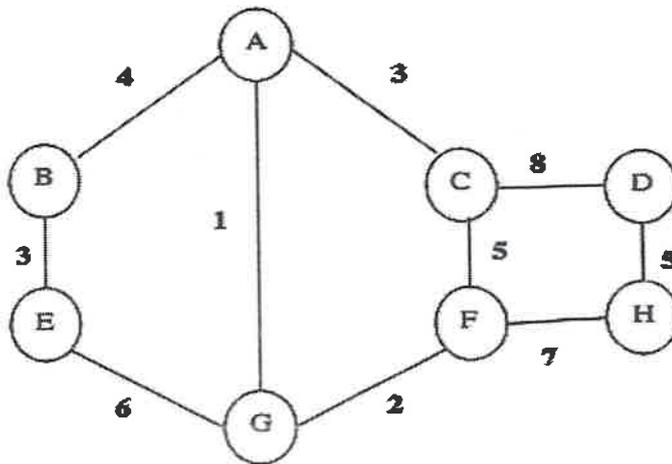
Q.4(A) The size of 5 matrices is represented as: [5, 10, 3, 12, 5, 50]. Find the optimal chain of matrix multiplication using Dynamic Programming. 20 4 3

OR

Q.4(B) Given the following activities with start and finish times, solve the Activity-Selection Problem: 20 4 3

Activity	Start	Finish
A1	1	4
A2	3	5
A3	0	6
A4	5	7
A5	8	9

Q.5(A) Explain with examples why every problem in P is also in NP, but the converse is not known. 20 5 2

OR

Q.5(B) Explain polynomial-time verification for the Clique problem. Illustrate the certificate and verification algorithm. 20 5 2

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) Regular End Semester Examinations -January 2026
MODERN DATABASE MANAGEMENT SYSTEMS
(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)	Illustrate the architecture of an RDBMS with neat diagrams including storage hierarchy file organization and record structures.	20	1	2
	OR			
Q.1(B)	Compare and contrast the differences between Tuple Relational Calculus (TRC) and Domain Relational Calculus (DRC) and also explain Relational Calculus and its types.	20	1	4
Q.2(A)	Illustrate the relational algebra transformation rules used for query optimization with suitable examples.	20	2	2
	OR			
Q.2(B)	Outline the join algorithms in query processing and explain the following nested loop, sort-merge and hash joins with examples and cost analysis.	20	2	2
Q.3(A)	Explain lock-based concurrency control protocols and give the advantages using shared and exclusive locks in two-phase locking (2PL) protocol.	20	3	4
	OR			
Q.3(B)	Explain the following protocols in detail i) Lock-based ii) Timestamp-based iii) Optimistic concurrency control protocols.	20	3	4
Q.4(A)	State and explain the CAP theorem with suitable examples and also list their impact on NoSQL database design.	20	4	4
	OR			
Q.4(B)	Demonstrate nesting and unnesting using object-relational queries on a database schema with suitable example.	20	4	3
Q.5(A)	Create an end-to-end MongoDB-based application incorporating indexing, upserts, and aggregation for enterprise application.	20	5	6
	OR			
Q.5(B)	Design a scalable MongoDB-backed application using Python and justify architectural choices and step-by-step for building such application.	20	5	5

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) Regular End Semester Examinations - January 2026
RESEARCH METHODOLOGY AND IPR
(Common to CSE and VES)

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)		Define research and discuss its importance in the contemporary world. Describe the research process in detail, illustrating the various stages with a suitable diagram.	20	1	2
		OR			
Q.1(B)		Discuss the importance of critical literature review and its uses in planning innovation research.	20	1	2
Q.2(A)		What do you mean by "Sample Design"? Under what circumstances one should use a probability sample?	20	2	2
		OR			
Q.2(B)		List the different methods of collecting data. Explain the observation method of collecting data with its merits and demerits.	20	2	2
Q.3(A)	i	What is oral presentation? What are merits and demerits of oral presentation?	10	3	2
	ii	What are the items in a research report? Explain them in brief.	10	3	2
		OR			
Q.3(B)		Describe the process of formulating a research hypothesis. Explain the objectives and significance of hypothesis formulation in research.	20	3	2
Q.4(A)	i	Define trade secrets and discuss their protection mechanisms.	10	4	2
	ii	Write a short note to summarize the all types of intellectual property rights.	10	4	2
		OR			
Q.4(B)		Discuss the relationship between IPR and biodiversity, including relevant international agreements.	20	4	2
Q.5(A)		What are the choices for patent application to be filed? Mention the patent application forms & explain the concept of claims.	20	5	2
		OR			
Q.5(B)		Explain the examination and grant process of a patent. What are the different types of patent revocation?	20	5	2

END

Hall Ticket No: [ ][ ][ ][ ][ ][ ][ ][ ][ ][ ]

MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE, MADANAPALLE
(Deemed to be University under Section 3 of UGC Act, 1956)

M.Tech. I Year I Semester (R25) Regular End Semester Examinations - January 2026
NATURAL LANGUAGE PROCESSING
(Computer Science & 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)	Explain the fundamental Natural Language Processing tasks and linguistic levels with suitable examples.	20	1	4
	OR			
Q.1(B)	Describe Text Normalization & Tokenization Techniques in detail. Give its importance in NLP pre-processing.	20	1	2
Q.2(A)	Explain the role of Context-Free Grammars (CFGs) and probabilistic CFGs in syntactic parsing with suitable examples.	20	2	2
	OR			
Q.2(B)	Explain Treebanks and Annotation Schemes used in syntactic parsing with suitable examples.	20	2	2
Q.3(A)	Explain sentiment analysis approaches and analyze challenges in discourse-level sentiment modeling.	20	3	4
	OR			
Q.3(B)	Categorize Text Classification approaches using traditional and embedding-based features.	20	3	2
Q.4(A)	Examine Maximum Likelihood Estimation in Natural Language Processing models and list its limitations.	20	4	4
	OR			
Q.4(B)	Examine fine-tuning paradigms in Neural Natural Language Processing models and their generalization issues.	20	4	4
Q.5(A)	Explain classical and Neural Machine Translation approaches and discuss their strengths and limitations.	20	5	4
	OR			
Q.5(B)	Create an end-to-end LLM-based Natural Language Processing application and justify ethical considerations.	20	5	5

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) Regular End Semester Examinations - January 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	Explain how data analysis is performed using Hadoop. Describe the key components of the Hadoop Ecosystem used for data analytics.	10	1	2
	ii	Consider a healthcare scenario where a hospital provides online appointment booking, telemedicine services, and health awareness campaigns through its own website and digital platforms. The website receives about 1 million patient visits per month from different regions, age groups, and devices. Identify the different types of healthcare data generated from this scenario. What healthcare data analytics methods can be applied to analyze this data and improve patient engagement, service quality, treatment outcomes, and overall hospital performance?	10	2	4
		OR			
Q.1(B)	i	What is Hadoop Streaming? Explain its working and how it enables running MapReduce jobs using non-Java programs.	10	1	2
	ii	Consider a fraud detection scenario where a financial institution monitors transactions through its online platform and mobile app. The system receives about 1 million transactions per month from different regions, devices, and customer profiles. a) Identify the different types of data generated from this scenario. b) What data analytics methods can be applied to analyze this data and improve fraud detection accuracy, real-time alerting, and overall financial security?	10	2	4
Q.2(A)	i	In Hadoop HDFS, the default block size is 128 MB (or 64 MB, depending on configuration). Suppose you have an input file of 350 MB stored in HDFS. 1. Calculate the number of input splits that HDFS will create for this file. 2. Determine the size of each input split. 3. Explain the concept of input split and how it is related to HDFS block size in Hadoop MapReduce.	10	2	3
	ii	Critically examine the challenges faced by the NameNode in HDFS. Suggest solutions such as High Availability (HA) and Federation, and explain how they improve scalability.	10	2	4
		OR			
Q.2(B)	i	Differentiate between Writable and Serializable in Hadoop. Explain why Hadoop uses Writable.	10	2	3
	ii	A company receives real-time clickstream data from its website and daily customer records from its SQL database. Analyse how Flume and Sqoop can be used together in a data ingestion pipeline. Evaluate the strengths and limitations of each tool for this use case.	10	2	4

Q.3(A)	i	Explain the importance of job scheduling in Hadoop. Discuss how scheduling improves cluster performance and resource utilization. Given a dataset of student marks (student_id, subject, marks), design a MapReduce pipeline to compute: • Total marks per student • Rank students based on total marks • Output top 10 students	10	3	2
	ii	Explain the composition of MapReduce jobs and provide pseudo-code.	10	3	4
		OR			
Q.3(B)	i	Explain how map and reduce tasks are executed in a Hadoop cluster with a neat diagram.	10	3	2
	ii	Design a modified MapReduce approach to reduce shuffle overhead.	10	3	4
Q.4(A)	i	Compare Apache Pig with traditional databases. Explain the differences in terms of data processing, query language, and architecture.	10	4	2

	ii	You are given a dataset of customer transactions stored in HDFS at /data/transactions.csv with the following schema:
		transaction_id, customer_id, amount, category, transaction_date
		Write a Pig Latin script to perform the following tasks:
		• Load the dataset into Pig.
		• Filter transactions with amount greater than 5000.
		• Create a new field discount using a User Defined Function (UDF) which applies:
		  o 10% discount if amount > 10000
		  o 5% discount if amount is between 5000 and 10000
		  o 0% discount otherwise
			10	4	4

		OR			
Q.4(B)	i	Explain how Hive uses data types and file formats for efficient storage and query processing in Hadoop.	10	4	2
	ii	You are given two Hive tables: Orders order_id STRING, customer_id STRING , order_date DATE, amount DOUBLE, status STRING Customers customer_id STRING, customer_name STRING, city STRING, signup_date DATE Write HiveQL queries to perform the following: a) Find the top 5 customers by total order amount. b) Calculate the monthly sales trend (total amount per month) for the last 6 months. c) Find customers who have not placed any orders. d) Get the average order amount by city for customers who joined in the last year. e) Create a partitioned table orders_partitioned by order_date and insert data into it.	10	4	4
Q.5(A)	i	Explain the concept of stream sampling and stream filtering. Describe different techniques used for sampling and filtering streaming data with examples.	10	5	2
	ii	A telecom company needs to compare two data streams representing customer call patterns from two different regions. Each stream contains millions of call records per day. The company wants to measure the distance between these two streams to detect behavioural differences and anomalies.	10	5	4

a) Explain the concept of counting distance between two data streams and why exact computation is not feasible in real-time.
b) Design a streaming algorithm (e.g., Count-Min Sketch or Hamming distance approximation) to estimate the distance between two streams.
c) Analyse the trade-offs between accuracy and memory usage in your proposed algorithm.
d) Evaluate how the estimated distance can be used to detect anomalies or changes in customer behavior.

OR

Q.5(B)	I	Explain the concept of elements in a data stream. Describe the characteristics of streaming data and list the types of elements found in a stream with examples.	10	5	2
	ii	A real-time analytics system monitors credit card transactions to detect fraud. The system receives a continuous stream of transactions and needs to compute a fraud risk score that gives more importance to recent transactions than older ones. Explain the concept of decaying windows and how they differ from sliding and tumbling windows. Design a decaying window model (e.g., exponential decay) to compute the weighted sum of transaction amounts.	10	5	4

END