(161)

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

M.Tech. (DS)-II SEMESTER | T3 EXAM - MAY 2026

COURSE CODE (CREDITS): 22M11CI213 (3) MAX. MARKS: 35

COURSE NAME: Big Data Analytics (BDA) MAX. TIME: 2 Hours

COURSE INSTRUCTORS: D. Gupta

Note: *1) All questions are compulsory. Marks and COs for each question are indicated. 2)* ***Answer the questions in the given order****. 3)* ***Be concise and write neatly.*** *4) Use of a calculator is allowed.*

Q. No.	Question	CO	Marks
Q. 1	a) Evaluate how emerging paradigms such as federated learning or edge analytics address the dual constraints of data privacy and infrastructure energy consumption. b) Explain why traditional centralized machine learning approaches fail to scale when datasets reach the zettabyte level.	1	3, 3
Q. 2	a) In the MapReduce framework, why is the shuffle and sort phase considered critical for the performance of algorithms like Word Count or Group-By? b) Describe the Market Basket Model and the Apriori algorithm for frequent itemset mining. Why is Apriori inefficient for large datasets, and how can stream-based or limited-pass approaches improve scalability?	2	3, 3
Q. 3	Discuss clustering techniques used in Big Data, with emphasis on hierarchical clustering. Explain how clustering is performed in non-Euclidean spaces. Further, describe the major challenges encountered while performing clustering in non-Euclidean spaces.	5	5
Q. 4	a) Explain how Euler's solution to the Seven Bridges of Königsberg led to the abstraction of graphs in modern network analysis? b) Compare closeness centrality and betweenness centrality. Which is more suitable for identifying gatekeepers in information flow?	4	3, 3
Q. 5	a) Explain how graph convolutional networks or graph embedding methods represent nodes using structural and attribute information. b) Analyze the computational bottleneck of performing backpropagation on a graph of this scale. How do modern frameworks like Apache Spark (GraphX) or specialized graph generative networks libraries handle the graph partitioning challenge to ensure scalability without losing the global structural context?	5	3, 3

Q. 6	a) Analyze the architectural shift from static latent factor models to deep learning and sequential recommendation frameworks (e.g., YouTubeDNN or SASRec). Explain how the introduction of multi-layer perceptrons (MLPs) and attention mechanisms allows modern systems to capture non-linear interactions and evolving user intent over time. Finally, discuss the trade-off between memorization (via wide linear layers) and generalization (via deep layers) as exemplified by Google's Wide & Deep architecture. b) What are the differences between collaborative filtering (CF) and content-based filtering (CBF)?	4	3, 3