Graphic Era HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Odd) Semester Examination November 2025

Roll no.....

Name of the Course and semester: M. Tech. I
Name of the Paper: Big Data Analytics
Paper Code: MCS129
Time: 3 hour Maximum Marks: 100

Note:

(i) All the questions are compulsory.
(ii) Answer any two sub questions from a, b and c in each main question.
(iii) Total marks for each question is 20 (twenty).
(iv) Each sub-question carries 10 marks.

Q1. (2X10=20 Marks)

a. Explain the evolution of Big Data technologies over time. Highlight key milestones and their impact on data-driven decision-making. CO1

b. Compare and contrast statistical modeling and machine learning approaches for data analysis. When is each method preferable? CO1

c. Why is statistical modeling important in the Big Data era? Explain with examples how statistical principles guide Big Data analysis and interpretation. CO1

Q2. (2X10=20 Marks)

a. Explain the architecture and working of a Distributed File System (DFS). Discuss how data replication and fault tolerance are handled in DFS with examples. CO2

b. Describe the MapReduce programming model. Explain with a suitable example how the map and reduce functions work together to process large-scale data. CO2

c. Discuss filtering streams using Bloom Filters. Explain how Bloom Filters help in approximate membership queries and analyze their error probability. CO2

Q3. (2X10=20 Marks)

a. Explain the PageRank algorithm in detail. Discuss how it models the importance of web pages and outline the iterative process used for its computation. CO3

b. Describe the Apriori Algorithm with example. Explain its working with an example and discuss how it helps in identifying frequent item sets in large transactional datasets. CO3

c. What are Limited-Pass Algorithms? Discuss their role and importance in mining frequent item sets from massive data streams. Explain with example. CO3

Q4. (2X10=20 Marks)

a. What are non-Euclidean spaces? Explain the challenges of performing clustering in non-Euclidean spaces and discuss algorithms suitable for such data. CO4

b. Explain how distance measures are used in clustering non-Euclidean data. Provide examples of their applications in real-world datasets. CO4

c. Compare parallel K-Means and parallel hierarchical clustering approaches in terms of scalability, accuracy, and computational efficiency. CO4

Q5. (2X10=20 Marks)

a. Explain how social networks can be represented as graphs. Discuss the roles of nodes, edges, and weights with suitable real-world examples from online platforms. CO5

b. Discuss the importance of clustering in social network graphs. Explain any one clustering algorithm used to identify groups of closely connected users. CO5

c. Define overlapping communities in social networks. Discuss why overlapping community detection is important and describe one algorithm used for finding such communities. CO5