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

Term Evaluation (ODD) Semester Examination September 2025

Roll no.....

Name of the Course: MTech(CSE)
Semester: I
Name of the Paper: Big Data Analytics
Paper Code: MCS 129
Time: 1.5-hour **Maximum Marks: 50**

Note:
(i) Answer all the questions by choosing any one of the sub-questions
(ii) Each question carries 10 marks.

Q1. (10 Marks)
a. Propose how sampling techniques can adapt classical statistical methods to big data. (CO2)

OR

b. Illustrate how the map, shuffle, and reduce stages operate using the logical workflow, and explain how it enables parallel processing. (CO1)

Q2. (10 Marks)
a. Define the Zettabyte Era. What does it signify about our data-driven world and why does it underscore big data's importance now. (CO1)

OR

b. How do techniques like biased reservoir sampling and concise sampling select representative subsets from a stream? (CO2)

Q3. (10 Marks)
a. Outline three benefits that big data processing brings over conventional approaches, such as real-time decision-making, richer analytics, predictive capabilities, and personalization also provide examples.(CO2)

OR

b. How do moment-based metrics (like frequency moments) help characterize a data stream, and what algorithms estimate them efficiently. (CO1)

Q4. (10 Marks)
a. Define statistical modeling and machine learning. How do they function within big data analytics, and what challenges arise when scaling them (e.g., high dimensionality, noise). (CO1)

OR

b. Compare MapReduce with alternative big data frameworks such as Apache Spark. Highlight the architectural strengths and limitations of both models. (CO2)

Q5. (10 Marks)
a. Compare and contrast ordinary data processing with big data approaches. Focus on architecture, tools, performance, and result scope.(CO1)

OR

b. How do limitations like monoid properties or shuffle costs constrain algorithm design, and what theoretical models help explain these. (CO2)