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 (Even) Semester Examination June 2025

Roll no.....

Name of the Course and semester: BCA, Fourth Semester
Name of the Paper: Big Data Analytics
Paper Code: TBC405-(1)
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 Big Data and discuss the evolution of Big Data. **(CO1)**
b. Discuss the characteristics and sources of Big Data
c. Describe HDFS and its components, also why it is suitable for storing for massive data.

Q2. (2X10=20 Marks)

a. Differentiate between supervised and unsupervised learning. **(CO2)**
b. The data set given below describes whether the substance is poisonous or not. Using Naïve Bayes Classifier predict for the new input data
{ Color = Green, Toughness = Soft, Fungus = Y, Appearance = Wrinkled}
whether the substance is poisonous or not.

Color	Toughness	Fungus	Appearance	Poisonous
Green	Hard	No	Smooth	No
Green	Hard	Yes	Smooth	No
Brown	Soft	No	Wrinkled	No
Orange	Hard	No	Wrinkled	Yes
Green	Soft	Yes	Smooth	Yes
Green	Hard	Yes	Wrinkled	Yes
Orange	Hard	No	Wrinkled	Yes

c. Create a decision tree from the following dataset. Predict for the input data **{Age=Middle Age, Gender=Female, Blood Pressure=Normal}**

Age	Gender	Blood Pressure	Drug
Young	Female	High	Drug A
Young	Female	High	Drug A
Middle-Age	Female	High	Drug B
Middle Age	Male	Low	Drug B
Senior	Male	Normal	Drug B
Senior	Female	Normal	Drug B
Senior	Male	High	Drug A

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 (Even) Semester Examination June 2025

Q3. (2X10=20 Marks)

a. Explain association Rule and describe its contribution in data analytics. **(CO3)**

b. Find the frequent itemset and association Rule for the following dataset where minimum threshold for support is 33% and confidence is 55%.

Transactions	Item
T1	Milk, Butter, Bread
T2	Milk, Chocos Egg
T3	Chocos, Milk, Butter
T4	Egg, Bread, Butter
T5	Egg, Bread
T6	Milk, Bread, Chocos
T7	Chocos, Milk
T8	Bread, Butter, Milk, Egg
T9	Bread Butter

c. Explain the various phases of Hadoop MapReduce Processing and discuss its implementation for finding occurrence of every word in a file with a diagram.

Q4. (2X10=20 Marks)

a. Explain Stream and Stream Computing? Discuss the various real time applications of Stream Computing.

b. Discuss the various components of Stream Computing Model **(CO4)**

c. Explain Graph Databases and differentiate between Graph Database and Traditional Databases.

Q5. (2X10=20 Marks)

a. Discuss the various Data Model of No-SQL. **(CO5)**

b. Differentiate between Schema less and No-Schema Databases

c. Explain Object Store Model also discuss its advantages and disadvantages.