(2 hours)

[Total Marks: 50]

N. B.: (1) **All** questions are **compulsory**.
(2) Make **suitable assumptions** wherever necessary and **state the assumptions** made.
(3) Answers to the **same question** must be **written together**.
(4) Numbers to the **right** indicate **marks**.
(5) Draw **neat labeled diagrams** wherever **necessary**.
(6) Use of **Non-programmable** calculators is **allowed**.

1. Attempt *any two* of the following: **10**

a. Discuss the rules of European Union General Data Protection Regulation (GDPR).
b. Enumerate the general rules for data source catalog.
c. Explain hypothesis testing, t-test and chi-square test with respect to data science.
d. Explain the organize superstep with suitable example.

2. Attempt *any two* of the following: **10**

a. How will you use MoSCoW prioritization technique in Data Science Projects?
b. If a Data Science ecosystem does not have a proper structure for data storage, then which eco-system will be preferred, Schema–on–Write or Schema–on–Read? Give reasons.
c. Explain Cross-Industry Standard Process for Data Mining (CRISP-DM).
d. What do you mean by slowly changing dimensions? Explain different types of slowly changing dimensions with suitable examples.

3. Attempt *any two* of the following: **10**

a. Explain the different types of watchers.
b. State and explain the five fundamental steps of the data science process.
c. Explain the retrieve superstep.
d. How will you avoid data swamps? Explain four critical steps.

4. Attempt *any two* of the following: **10**

a. Explain local time and Universal Coordinated time.
b. What is Fishbone Diagram? Explain with example.
c. Explain Person Hub in Time-Person-Object-Location-Event Data Vault. Which are the different Person Links created?
d. Create the following Sun Models:
   a. Person-to-Time
   b. Person-to-Object
   c. Person-to-Location
   d. Person-to-Event

5. Attempt *any two* of the following: **10**

a. How are the results of data science summarized? Explain.
b. Explain Andrews' curves with their use in data science.
c. What do you mean by Clustering? Explain any two types of Clustering?
d. Differentiate between Univariate, Bivariate and Multivariate Analysis.
