Time: 2½ hrs **[Total Marks: 75]**

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 labelled diagrams** wherever **necessary**.
(6) Use of **Non-programmable** calculators is **allowed**.

Q1. Attempt the three of the following. **15 Marks**
a. List and explain the different data stores used in data science.
b. State and explain the four critical steps to avoid data swamps.
c. Why is it necessary to train the data science team.
d. Explain the fundamental data science process steps
e. Give an overview of the Drum-buffer-rope Method

Q2. Attempt the three of the following. **15 Marks**
a. Explain the Retrieve Super step.
b. Explain Data Lakes and Data Swamps.
c. Explain in detail the function of Operational Management Layer
d. List & Explain the data structures in the functional layer of the ecosystem
e. List the super steps for processing the data lake.

Q3. Attempt three of the following. **15 marks**
a. State and explain the five steps of data science
b. How to Complete the 5 Whys.
c. Write a note on ANOVA.
d. Explain CLustering techniques.
e. What is Simple Linear Regression? Explain.

Q4. Attempt three of the following. **15 marks**
a. Explain the process superstep.
b. Explain concept of data valut.
c. What is an event? Explain explicit and implicit events
d. What is Directed Acyclic Graph Scheduling.
e. Write a short note on the Rapid Information Factory Ecosystem.

Q5. Attempt three of the following. **15 marks**
a. Explain the transform superstep.
b. Write a note on Decision Trees.
c. What is feature engineering? What are the common feature extraction techniques.
d. Explain the use of correlation and forecasting in data science.
e. What are the general rules for data source catalog.

****