

(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. Explain spark as a data science processing tool.
 b. State and explain the four critical steps to avoid data swamps.
 c. Explain the Person section of Time-Person-Object-Location-Event (TPOLE).
 d. Explain Report Superstep.
2. Attempt any two of the following: 10
 a. What is data science? Explain Data lake.
 b. Explain the functional layer R-A-P-T-O-R framework.
 c. Write short note on Hub, and Satellite with data vault context..
 d. Explain Akka What are the data science processing tools.
3. Attempt any two of the following: 10
 a. Explain the training the trainer model.
 b. Explain five fundamental data science process steps.
 c. Explain the different ways to deal with errors.
 d. Explain the Drum-Buffer-Rope scheduling methodology.
4. Attempt any two of the following: 10
 a. Explain Person-to-Time Sun Model.
 b. Explain Overfitting and Underfitting.
 c. Explain Hypothesis testing with the help of T-test.
 d. Why data has missing values? Explain missing value treatment in transforming with data science.
5. Attempt any two of the following: 10
 a. Describe the Linear regression and logistic regression.
 b. Explain ANOVA in data science..
 c. Write short note on clustering techniques.
 d. Explain organize superstep in data science.