

(Time: 2½ Hours)

[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 labeled diagrams wherever necessary.
 (6) Use of Non-programmable calculators is allowed.

1. Attempt any Three of the following: 15

- How has the role of a data engineer changed over the years?
- What is meant by data maturity? Discuss the various stages in the data maturity model for a company.
- Explain in brief the skills that a data engineer should possess.
- Distinguish between batch and stream data ingestion.
- Explain the Reverse ETL process with the help of examples.
- What are 'undercurrents' in the data engineering life cycle? Explain Discoverability and Accountability with respect to Data Governance.

2. Attempt any Three of the following: 15

- List and discuss the following two principles of good architecture.
 - Choose common components wisely
 - Make reversible decisions
- Define and explain how Scalability, Availability, Reliability, Elasticity are inter-related.
- With the help of a diagram, explain the Lambda architecture.
- Explain how team size and its capabilities affect choosing data technologies across the data engineering lifecycle.
- Explain factors to consider with a commercial OSS project.
- Briefly discuss the impact of DataOps on choosing technologies.

3. Attempt any Three of the following: 15

- Discuss in detail the following source systems that generate the data for the rest of the data engineering lifecycle:
 - Files and Unstructured Data
 - Database logs
- List the major considerations for understanding database technologies.
- Explain the various areas to consider for data management of source systems.
- Write in brief about the following two raw storage ingredients:
 - Serialization
 - Compression
- Distinguish between eventual and strong consistency.
- Elaborate on the following as data engineering storage abstractions:
 - Data warehouse
 - Data lake

4. **Attempt any Three of the following:** 15
- a Briefly elaborate the factors to consider when designing your ingestion architecture.
 - b Discuss the two types of batch ingestion.
 - c Explain how security and software engineering impact the ingestion phase.
 - d List and explain the common types of windows used for streaming queries and processing.
 - e State the Kimball approach to data modelling.
 - f Highlight the differences between micro-batch and true streaming.
5. **Attempt any Three of the following:** 15
- a. What are the common ways to serve data for analytics and Machine Learning?
 - b. Define analytics. List and explain the types of analytics.
 - c. Discuss how each of the three – people, process and technology are related to security.
 - d. Briefly state the features of the live data stack.
 - e. Explain the various serialization formats.
 - f. What is a cloud network topology? What are availability zones and regions?
-