Assignment Questions on Introduction to Data Engineering

1. Overview of Data Engineering

- Define Data Engineering. Why is it considered essential in modern data-driven organizations? Provide examples of real-world applications of Data Engineering.

2. Importance and Applications of Data Engineering

- Discuss three key applications of Data Engineering in the finance, healthcare, and retail industries. How do these applications impact decision-making in each sector?

3. Data Engineering Lifecycle

- Describe the phases of the Data Engineering Lifecycle. How does a Data Engineer ensure smooth transitions between the Data Generation, Collection, Storage, Processing, and Analysis stages?

4. The Role of a Data Engineer

- Explain the role of a Data Engineer in an organization. How does it differ from the roles of a Data Scientist and a Data Analyst?

5. Data Collection Methods

- Compare and contrast batch processing and streaming as methods of data collection. Provide a use case where each method would be most suitable.

6. Data Extraction from APIs

- Explain the process of data extraction from APIs. Choose a public API and demonstrate how a Data Engineer would use it to collect and store data for analysis.

7. Data Modeling Concepts

- What is the importance of normalization and denormalization in data modeling? Provide an example of when denormalization would be preferred in a data engineering context.

8. Relational vs. NoSQL Databases

- Compare relational databases and NoSQL databases. In what scenarios would a Data Engineer opt for a NoSQL database over a relational one?

9. Data Warehouse vs. Data Lake

- Explain the role of a data warehouse in a data engineering project. How does it differ from a data lake, and when would each be used?

10. ETL Processes

- Describe the ETL (Extract, Transform, Load) process in detail. How can inefficient ETL processes impact the overall data pipeline in an organization?

11. In a data engineering project, how would you decide whether to use an OLTP or OLAP system? Identify the factors that should influence this decision and explain how each factor impacts the overall data architecture and business objectives.