Data Engineering Notes

Chukwuemeka O. Okoli

30 October, 2023

Introduction to Data Engineering

In the simplest possible terms, the field of Data Engineering concerns itself with the mechanics for the flow and access of data. And its goal is to make quality data available for fact-finding and data-driven decision making.

The field of Data Engineering involves:

1. Collecting source data
    - Extracting, integrating, and organizing data from disparate sources
        - Develop tools, workflows, and processes that help data acquisition from multiple sources
        - Design, build, and maintain scalable data architecture for storing source data
2. Processing data
    - Cleaning, transforming, and preparing data to make it usable
        - Implement and maintain distributed systems for large-scale data processing
        - Design pipelines for the extraction, transformation, and loading of data into data repositories
        - Design solutions for validating and safeguarding quality, privacy, and security of data
        - Performance optimization
        - Adherence to compliance guidelines
3. Storing data
    - Storing data for reliability and easy availability of data
        - Data stores for storage of processed data
        - Scalable systems
        - Ensuring data privacy, security, compliance, monitoring, backup, and recovery
4. Making data available to users securely
    - APIs, services, and programs for retrieving data for end-users
    - User access through interfaces and dashboards
    - Checks and balances to ensure data security