**A BEGINNER’S GUIDE TO DATA ENGINEERING CONCEPT, TOOLS AND RESPONSIBILITY.**

Data engineering is the development, implementation, and maintenance of systems and processes that take in raw data and produce high-quality, consistent information that supports downstream use cases, such as analysis and machine learning (Reins. J & Housely. M, 2022). In other word this field involves designing and building systems for collecting, storing, and analysing data at scale.

The role of data engineers are as follows;

1. Acquire datasets that align with business needs, develop algorithms to transform data into useful, actionable information,
2. Build, test, and maintain database pipeline architectures
3. Collaborate with management to understand company objectives
4. Create new data validation methods and data analysis tools
5. Ensure compliance with data governance and security policies.

Data engineering cycle involves the following stage;

1. Data acquisition

This is stage where data is created from various platform such as transactional database, an IoT device, flat files (CSV, XML), RSS feeds, web services and many more

1. Data ingestion

This stage involve incorporates data from generating sources into the processing system.

1. Data transformation

This stage involves refines raw data through operations that enhance its quality and utility. At this stage, data is normalized, gaps are filled where data might be missing, converts between data types. In short this standardize data for analysis stage.

1. Data Serving

This stage ensures that that data is timely, reliable, and accessible to support various analytical, reporting, and operational needs of an organization

1. Storage

This stage is where data is stored through various data engineering stages. Storage merges the disparate sets to offer a cohesive and consistent data view.

The data engineer’s uses different tool for different purpose and commonly, data engineer uses following tools;

1. Relational Databases/Data Warehouses which store and organize large amounts of structured data.
2. ETL Tools which make the movement of data easier.
3. Data Orchestration Platforms which automate and manage complex workflows involving multiple data sources and processes, ensuring efficient data flow and integration.
4. Data Governance Tools which add data quality and governance solutions to the list, aiming at maintaining data integrity and compliance.

It is requirement that data engineer must have a knowledge of coding since the software like python is used alongside SQL, PostgreSQL etc.

This field is very reward as one can work as software engineers or business intelligence analysts, data architect, solutions architect, or machine learning engineer and many more.