**ETL Pipeline**

* ETL Pipeline stands for Extract, Transform and Load Pipeline.
* In an ETL process we extract the data from a data source
* Once the data is extracted we Transform the data according to our requirement
* After the data is in our desired state we load it into a data warehouse or a datalakehouse
* An ETL pipeline allows us to provide context, consolidation and productivity to our data.

**ELT Pipeline**

* ELT stands for Extract, Load and Transform
* In ELT we transform the data after it has been moved from the source unlike ETL where we transform the data as it is being moved from the source.
* ELT enables us to deal with huge amount of data since it is able to harness the power of a database tool once the data has been loaded unlike ETL processes which suffer from bottleneck as the size of the data increases.
* ELT process are massively used in cloud computing where existing data can be loaded into cloud and the processing power of the cloud can be leveraged.

**3 Tier Architecture in Data Engineering**

The three tier architecture can be define in 3 layers:

1. **Presentation layer**

This is the topmost layer that interacts with users and other systems. It presents data to users through a graphical user interface (GUI) or an application programming interface (API). The presentation layer is responsible for displaying data in a user-friendly format and collecting user inputs.

1. **Application Layer**

This is the middle layer that handles the application logic and business rules. It receives requests from the presentation layer and processes them. The application layer is responsible for data validation, authentication, authorization, and data manipulation. It also communicates with the data layer to retrieve or store data

1. **Data Layer**

This is the bottommost layer that stores and retrieves data. It can be a database, a data warehouse, a data lake, or any other data storage system. The data layer is responsible for ensuring data integrity, security, and availability.

**3 ETL Tools**

Following are the 3 tools used in data engineering

1. **Apache Kafka:**

Apache Kafka is a distributed messaging system used in data engineering for real-time data streaming. It is designed to handle large amounts of data across multiple systems, making it ideal for building scalable, fault-tolerant data pipelines. Kafka allows data to be published, subscribed to, and processed in real-time, making it a popular choice for data ingestion, processing, and analytics.

1. **Azure Data Factory:**

Azure Data Factory is a cloud-based data integration service provided by Microsoft Azure. It is used in data engineering for building and managing data pipelines that extract data from various sources, transform it, and load it into different destinations. With Azure Data Factory, data engineers can create data workflows that integrate with various Azure services, such as Azure Blob Storage, Azure SQL Database, and Azure Data Lake Storage, as well as third-party sources like Salesforce, Oracle, and Amazon S3. The tool provides an intuitive graphical interface for designing and monitoring data pipelines and includes advanced features such as data masking and transformation, fault tolerance, and data lineage tracking. Azure Data Factory is an excellent choice for enterprises looking to build scalable, cost-effective, and reliable data integration solutions in the cloud.

1. **AWS Glue:**

AWS Glue is a fully managed ETL (Extract, Transform, Load) service provided by Amazon Web Services. It is used in data engineering for building and managing data pipelines that automate the process of extracting data from various sources, transforming it, and loading it into various destinations. With AWS Glue, data engineers can create data workflows that integrate with various AWS services, such as Amazon S3, Amazon RDS, and Amazon Redshift, as well as third-party sources like Salesforce, Oracle, and MongoDB. The tool provides an intuitive graphical interface for designing and monitoring data pipelines and includes advanced features such as automatic schema discovery, data deduplication, and data quality validation. AWS Glue is an excellent choice for enterprises looking to build scalable, cost-effective, and reliable data integration solutions in the cloud.