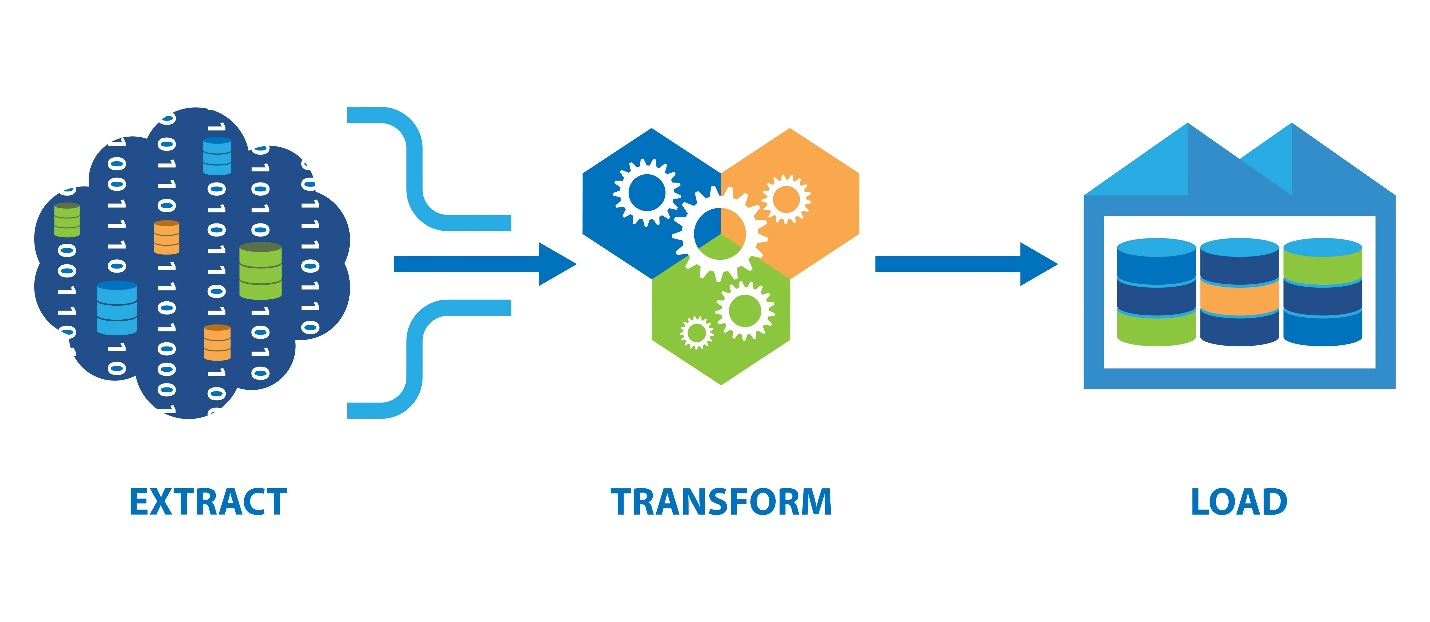
**Name: Muhammad Usman**

**Task#4**

**ETL:**

ETL stands for Extract, Transform, Load. It's a process in data engineering that involves taking data from one or more sources (Extract), cleaning, converting, or reformatting it to make it more useful and meaningful (Transform), and then loading it into a target database or data warehouse (Load) so that it can be analyzed and used for business intelligence or other purposes.

Think of it to move data from various sources such as spreadsheets, databases, and other files, clean and standardize it, and then store it in a central location for analysis and reporting. It's like taking a bunch of puzzle pieces from different boxes, cleaning them up, putting them together in a uniform way, and then storing the completed puzzle in a safe place where it can be easily accessed and used.

****

**ELT:**

ELT stands for Extract, Load, Transform. It's a process in data engineering that is similar to ETL, but with a different sequence of steps. Instead of transforming the data before loading it into a target database or data warehouse, ELT loads the data first, and then transforms it.

With ELT, data is extracted from various sources and loaded into a target database or data warehouse as is, without any major transformations. The transformations are then performed on the loaded data, using tools and techniques such as SQL queries, data pipelines, or data integration platforms. This approach allows for more flexibility in the transformation process, as it enables data analysts and scientists to work directly with the raw data and apply different types of transformations as needed.

To understand it in simpler terms, imagine you have a box of mixed toys. With ETL, you would first sort the toys by type, then clean them up and organize them, and finally store them in a toy box. With ELT, you would first dump all the toys into the toy box, then sort and organize them as needed. This allows you to work with the toys more freely and apply different types of sorting and organization based on your needs.

Graphical user interface, application

Description automatically generated

Top of Form

ELT is better than ETL when it comes to handling big amounts of data because ELT can adjust the amount of computing power needed to handle the data, making it more flexible. Also, ELT does not need extra storage spaces, which saves money and makes things run faster.

**Three- Tier Architecture:**

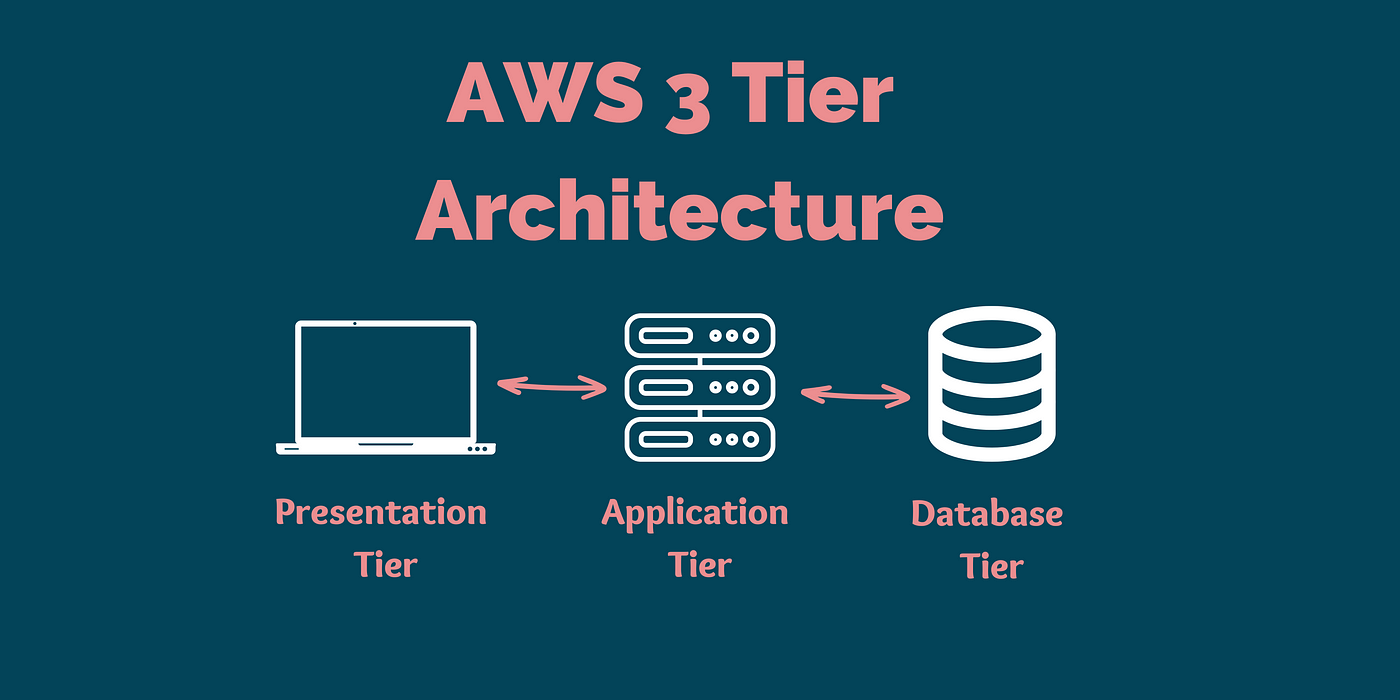
The three-tier architecture in data engineering is a way to organize the components and processes of a data system into three layers, each with a specific purpose.

The first tier is the presentation tier, also known as the user interface layer. This is where users interact with the system, input data, and receive output. Examples of this tier include web applications or mobile apps.

The second tier is the application or business logic layer. This layer handles the processing and manipulation of data. It is responsible for performing calculations, enforcing business rules, and managing workflows. Examples of this tier include application servers or batch processing systems.

The third tier is the data storage or data access layer. This layer is where data is stored, retrieved, and managed. It includes databases, data lakes, and other data storage systems. This layer is responsible for ensuring data integrity, security, and availability.

By separating these functions into three distinct tiers, data engineers can create a more modular, scalable, and maintainable system. Changes or upgrades to one tier can be made without affecting the other tiers, allowing for greater flexibility and agility in data engineering projects.



**ETL Tools (any 3)**

There are many ETL (Extract, Transform, Load) tools available, but here are three popular ones:

1. Apache NiFi: A free and open-source ETL tool that allows users to design and automate data flows between systems. It has a drag-and-drop interface for designing workflows and supports a variety of data sources and destinations.
2. Talend: A commercial ETL tool that offers a wide range of features, including data profiling, data quality, and data governance. It supports a variety of data sources and targets, including cloud-based systems like AWS and Azure.
3. Informatica PowerCenter: A commercial ETL tool that offers a scalable and high-performance data integration platform. It supports a variety of data sources and destinations, including on-premises and cloud-based systems, and offers advanced features like real-time data integration and data masking.

**Task#5**

**Three Loads in ETL**

Historical load, full load, and incremental load are three methods used in data engineering to transfer data from one system to another, such as from a database to a data warehouse.

**Historical load** refers to the process of transferring all historical data from the source system to the target system. This is typically done when setting up a new data warehouse or when there are major changes in the source system. It can be a time-consuming process as it involves transferring large amounts of data.

**Full load** refers to the process of transferring all data from the source system to the target system, regardless of whether it has changed or not. This is typically done on a regular basis, such as daily or weekly. It can also be time-consuming, especially if the source system contains a lot of data.

**Incremental load** refers to the process of transferring only the changes or updates that have occurred in the source system since the last transfer. This is typically done on a regular basis, such as hourly or daily, and is faster than a full load. Incremental load requires a way to track changes in the source system, such as using timestamps or unique identifiers.

**References:**

[**https://www.youtube.com/watch?v=VyvTabQHevw**](https://www.youtube.com/watch?v=VyvTabQHevw)

<https://www.youtube.com/watch?v=xRfi1mVk4uU>