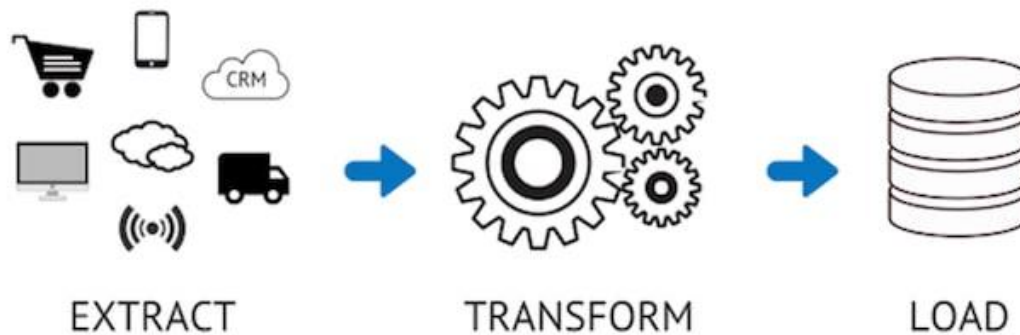**ETL** stands for **Extract ,Transform, and Load**.

Definition-A data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse

EXTRACT   TRANSFORM   LOAD

- **EXTRACT data from its original source**
- **TRANSFORM data by deduplicating it, combining it, and ensuring quality, to then**
- **LOAD data into the target database**

ETL tools enable data integration strategies by allowing companies to gather data from multiple data sources and consolidate it into a single, centralized location. ETL tools also make it possible for different types of data to work together.

A typical ETL process collects and refines different types of data, then delivers the data to a data lake or data warehouse such as Redshift, Azure, or BigQuery.

How ETL works

The ETL process is comprised of 3 steps that enable data integration from source to destination: data extraction, data transformation, and data loading.

Step 1: Extraction

 In this first step of the ETL process, structured and unstructured data is imported and consolidated into a single repository. Volumes of data can be extracted from a wide range of data sources, including:

- Existing databases and legacy systems
- Cloud, hybrid, and on-premises environments
- Sales and marketing applications
- Mobile devices and apps
- CRM systems
- Data storage platforms
- Data warehouses
- Analytics tools

Although it can be done manually with a team of data engineers, hand-coded data extraction can be time-intensive and prone to errors. ETL tools automate the extraction process and create a more efficient and reliable workflow.

Step 2: Transformation

During this phase of the ETL process, rules and regulations can be applied that ensure data quality and accessibility. You can also apply rules to help your company meet reporting requirements.

The process of data transformation is comprised of several sub-processes:

- **Cleansing** — inconsistencies and missing values in the data are resolved.
- **Standardization** — formatting rules are applied to the dataset.
- **Deduplication** — redundant data is excluded or discarded.
- **Verification** — unusable data is removed and anomalies are flagged.
- **Sorting** — data is organized according to type.
- **Other tasks** — any additional/optional rules can be applied to improve data quality.

Transformation is generally considered to be the most important part of the ETL process. Data transformation improves data integrity — removing duplicates and ensuring that raw data arrives at its new destination fully compatible and ready to use.

Step 3: Loading

The final step in the ETL process is to load the newly transformed data into a new destination fyf(data lake or data warehouse.) Data can be loaded all at once (full load) or at scheduled intervals (incremental load).

Full loading — In an ETL full loading scenario, everything that comes from the transformation assembly line goes into new, unique records in the data warehouse or data repository. Full loading produces datasets that grow exponentially and can quickly become difficult to maintain. -Research purposes

Incremental loading — A less comprehensive but more manageable approach is incremental loading. Incremental loading compares incoming data with what's already on hand, and only produces additional records if new and unique information is found. This allows smaller, less expensive data warehouses to maintain and manage business intelligence.