

# ETL & ELT

---

DIFFERENCES & SIMILARITIES

# ***Objects***

- What are ETL & ELT?
- Domains
- Similarities
- Differences
- Hybrid approach ETLT

# ETL Tools

In computing, extract, transform and load (ETL) refers to a process in database usage and especially in data warehousing that involves:

## **-Extraction:**

the extraction phase is one of the key stages in the overall process. It involves gathering and collecting data from multiple sources in preparation for further analysis.

During the extraction phase, data must be extracted from different data sources, which may include databases, data warehouses, spreadsheets, text files, and even web pages. The data may be collected through a variety of techniques, such as web scraping or querying a database.

## **-Transforming:**

During this phase, the data that has been extracted from various sources in the previous phase is preprocessed and manipulated to prepare it for analysis. The goals of the transforming phase include cleaning, integrating, reducing, and formatting the data so that it can be easily analyzed using statistical or machine learning techniques.

## **-Loading:**

During this phase, the cleaned and preprocessed data is loaded into a data warehouse, analytical database or other storage system for analysis and reporting purposes. The loading process typically involves inserting the transformed data from various sources into a central repository, or data warehouse, where it can be easily accessed by business intelligence tools or other analytics applications.

## ***ELT Tools***

On the other hand, in ELT, data is first extracted and then loaded directly into the target storage system without any pre-transformation. The transformation process happens within the target storage system using its built-in processing capabilities, such as SQL queries or stored procedures. This can be more efficient because it eliminates the need for a separate transformation layer.

## ***Domains***

- Data warehousing
- Business intelligence
- Data Migration
- Application Integration
- Big Data Processing

## *Similarities*

- Data Integration: Both ETL and ELT are used to integrate data from multiple sources into a single target database or data warehouse.
- Data Transformation: Both ETL and ELT involve the transformation of data from its source format into a format that can be used for analysis and reporting.
- Data Loading: Both ETL and ELT involve loading the transformed data into a target database or data warehouse.
- Automation: Both ETL and ELT can be automated using tools such as ETL/ELT software or custom scripts to improve efficiency and reduce errors.
- Scalability: Both ETL and ELT are scalable and can handle large volumes of data from various sources.
- Data Quality: Both ETL and ELT play a role in ensuring data quality by validating data as it is extracted, transformed, and loaded.

## *Differences*

ETL and ELT are two different approaches to data integration. ETL stands for Extract, Transform, and Load, while ELT stands for Extract, Load, and Transform. The main difference between the two is the order in which the data transformation step occurs.

In an ETL process, data is first extracted from various sources, then transformed into the desired format or structure, and finally loaded into the target system. This approach is useful when the source data is complex and requires significant data cleansing or transformation before it can be used in the target system.

On the other hand, ELT involves first extracting data from the source system and loading it directly into the target system. The transformation step occurs after the data has been loaded into the target system. This approach is particularly useful when the target system is capable of handling large volumes of data and can perform the necessary transformations quickly and efficiently.

The choice between ETL and ELT depends on a variety of factors, including the complexity of the source data, the capabilities of the target system, and the specific requirements of the project.

The main process that differentiates ELT from ETL is the Staging step:

Staging: The extracted data is staged in a temporary storage location, which can be a cloud-based data lake, a Hadoop Distributed File System (HDFS), or any other type of distributed storage system.

Once the data is staged, it is transformed into the desired format or structure using distributed processing frameworks like Apache Spark or Hadoop MapReduce.

After the transformation step, the data is loaded into the target system, which can be a data warehouse, a big data platform, or any other system that can store and process large volumes of data.

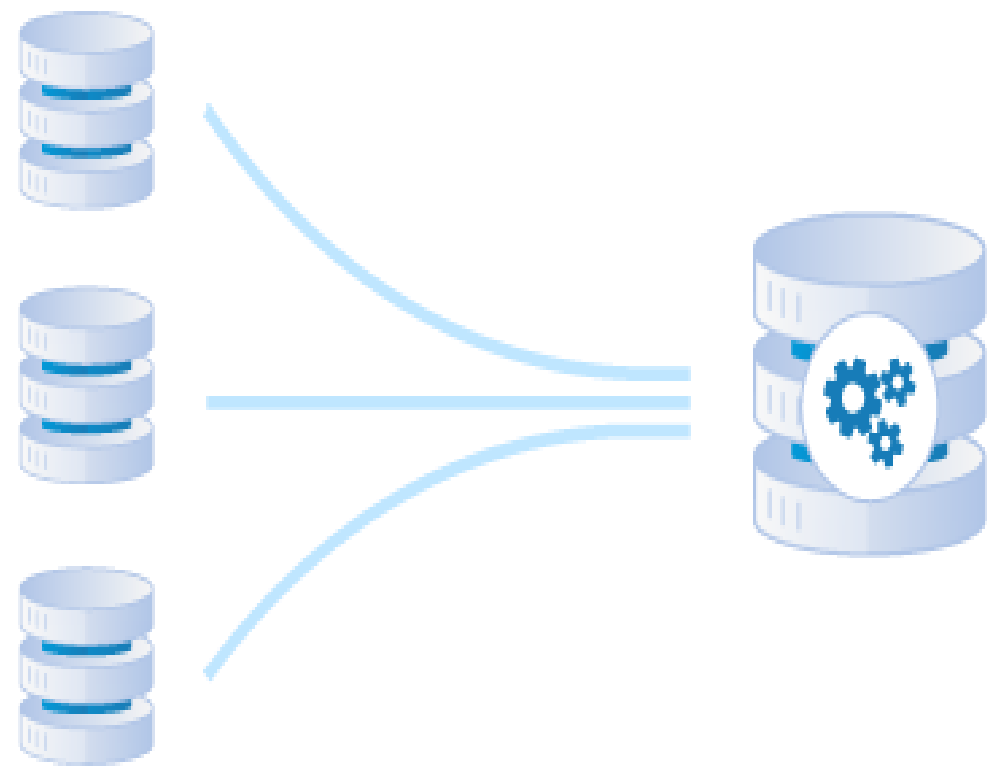


## ETL



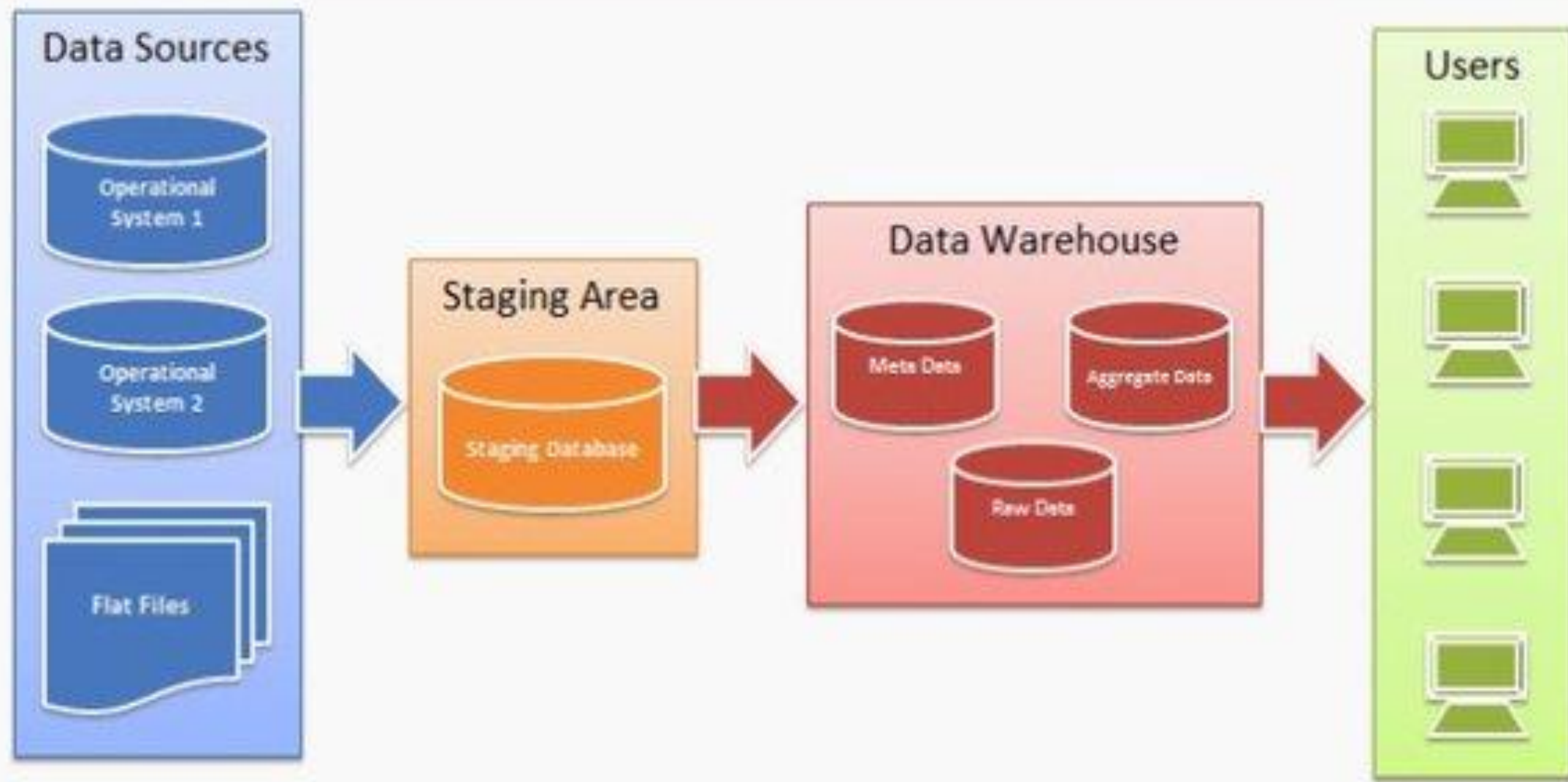
EXTRACT → TRANSFORM → LOAD

## ELT



EXTRACT & LOAD → TRANSFORM

## Data Warehouse with Staging Area Architecture



## ***A Hybrid Approach To Data Movement***

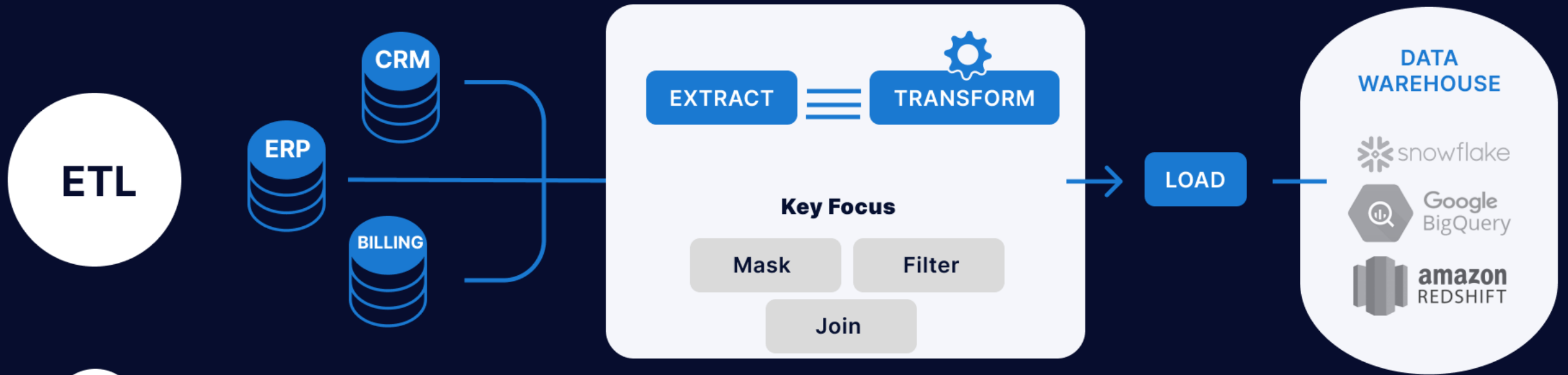
some scenarios benefit from employing a combination of ETL and ELT. For example, you might want to take advantage of ELT's rapid ingesting to give your data analysts immediate data access with the flexibility to change analytics data models on the fly.

At the same time, you may need to meet data security and compliance requirements that require you to mask, remove, or encrypt PHI and PII before moving data into the data warehouse. ETLT (extract, transform, load, transform) is an emerging framework that combines the best of both worlds.

ETLT integrates data into the data warehouse or data lake by:

- Extracting raw data from source applications and databases and loading it into a staging area.
- Lightly transforming the data in the staging area to remove, mask, and encrypt sensitive data. These transformations occur quickly because they only transform one source at a time.
- Loading the data into the data warehouse.
- Transforming and integrating the data more completely within the data warehouse, using database and SQL commands to process the transactions. This second transformation step performs the tasks necessary to consolidate data from multiple sources.

Taken together, this combination process enables IT teams to flexibly meet any need.



VS

