**SCD**

[**SCD Types Explained with SQL: A Guide for Data Engineers | by Maksim Kazartsev | Medium**](https://medium.com/@kazarmax/scd-types-explained-with-sql-a-guide-for-data-engineers-a26a07cf5c60)

**ELT Process**

Extraction, Load and Transform (ELT) is the technique of extracting raw data from the source, storing it in the [data warehouse](https://www.geeksforgeeks.org/dbms/data-warehousing/) of the target server and preparing it for end-stream users.

ELT consists of three different operations performed on the data:

1. **Extract**: Extracting data is the process of identifying data from one or more sources. The sources may include databases, files, [ERP](https://www.geeksforgeeks.org/software-engineering/introduction-to-erp/), [CRM](https://www.geeksforgeeks.org/software-engineering/customer-relationship-management-crm/), or any other useful source of data.
2. **Load**: Loading is the process of storing the extracted raw data in a data warehouse or [data lake](https://www.geeksforgeeks.org/data-engineering/what-is-data-lake/).
3. **Transform**: [Data transformation](https://www.geeksforgeeks.org/data-analysis/what-is-data-transformation/) is the process in which the raw data from the source is transformed into the target format required for analysis

A diagram of data storage server

AI-generated content may be incorrect.

In ELT, data from source systems is first loaded into the data warehouse without full transformation. Only the necessary transformations are done later, as needed for analysis. This allows raw data to be stored and accessed anytime, unlike ETL, where data is transformed before loading, and raw data may not be retained.

**ETL Process**

[ETL](https://www.geeksforgeeks.org/dbms/etl-process-in-data-warehouse/) is the traditional technique of extracting raw data, transforming it as required for the users and storing it in data warehouses. ELT was later developed, with ETL as its base. The three operations in ETL and ELT are the same, except that their order of processing is slightly different. This change in sequence was made to overcome some drawbacks.

1. **Extract**: It is the process of extracting raw data from all available data sources such as databases, files, ERP, CRM or any other.
2. **Transform:**The extracted data is immediately transformed as required by the user.
3. **Load**: The transformed data is then loaded into the data warehouse from where the users can access it.

A diagram of a stage area

AI-generated content may be incorrect.

In ETL, data from sources is first stored in a staging area, transformed there, and then loaded into the data warehouse. A major drawback is that once the data is transformed and stored, the original raw data is lost. In contrast, ELT keeps a copy of the raw data in the warehouse, allowing transformations to be done later as needed.

**Difference between ELT and ETL**

| **Category** | **ETL** | **ELT** |
| --- | --- | --- |
| **Acronym Meaning** | Extract, Transform, Load | Extract, Load, Transform |
| **Definition** | Extracts raw data, transforms it on a secondary server, then loads it into the destination. | Extracts raw data, loads it directly into the destination and transforms it there. |
| **Processing Speed** | Slower; data transformation occurs before loading. | Faster; data is loaded first and transformed in parallel. |
| **Data Volume** | Best for smaller, complex data sets like marketing data. | Suited for large data sets requiring speed, like real-time analytics. |
| **Data Output** | Primarily structured data. | Structured, semi-structured and unstructured data. |
| **Data Lake Compatibility** | Not compatible with data lakes. | Fully compatible with data lakes. |
| **Maturity** | Well-established, used for 20+ years, with extensive documentation. | Newer approach with fewer tools and less documentation. |
| **Cost Efficiency** | Higher costs due to the need for separate servers and processing infrastructure. | More cost-effective, leveraging cloud resources for scalability. |
| **Security** | Requires custom security solutions to protect sensitive data. | Built-in security features like access control and multifactor authentication. |
| **Transformation Location** | Data is transformed on a secondary server before loading. | Data is loaded as-is and transformed within the target system. |
| **Flexibility** | Best for structured data transformation. | Handles structured and unstructured data with ease. |

**Similarities Between ETL and ELT**

Both ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) are data integration processes that consolidate data from various sources into a single, unified repository for further analysis. They share several key similarities:

* **Data Extraction:** Both processes begin by extracting raw data from multiple sources like databases, files, [SaaS applications](https://www.geeksforgeeks.org/dbms/design-database-for-saas-applications/), or [IoT devices](https://www.geeksforgeeks.org/computer-networks/internet-of-things-iot-gateways/). This data can be structured, semi-structured, or unstructured.
* **Data Transformation:** While the timing of transformation differs, both ETL and ELT involve transforming the extracted data into a format that aligns with the target system's requirements. This ensures data is clean, accurate and ready for analysis.
* **Data Loading:** Both methods ultimately load the processed data into a data warehouse or data lake, providing a central repository where the data can be accessed and analyzed.
* **Unified Data Repository:** Both processes help create a single source of truth, ensuring that enterprise data is consistent, accurate and up-to-date for decision-making.

**Choosing Between ELT and ETL**

The choice between ETL and ELT depends on our specific needs and requirements.

* ETL works well for smaller datasets and [structured data](https://www.geeksforgeeks.org/dbms/what-is-structured-data/) where the data needs to be transformed immediately. It often requires special hardware and can be less flexible when handling large amounts of data.
* ELT is better for large datasets and unstructured or [non-relational data](https://www.geeksforgeeks.org/dbms/non-relational-databases-and-their-types/). It is more flexible and cost-effective, especially with cloud-based data solutions. With ELT, we can store raw data and transform it as needed.