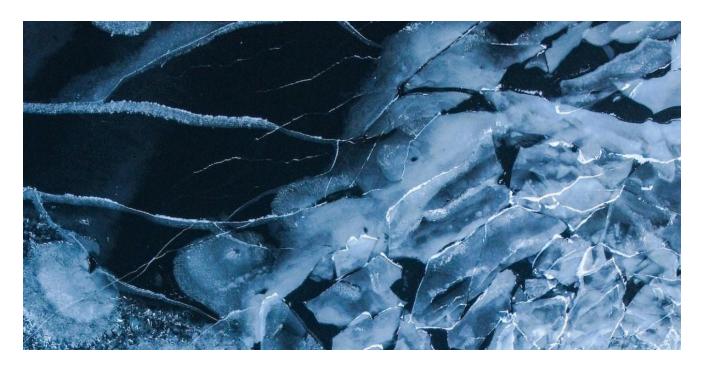# ELT vs. ETL: What's the Difference?

🌐 **ibm.com**/cloud/blog/elt-vs-etl-whats-the-difference

Similarities and differences in the definitions, benefits and use cases of ELT and ETL.

ELT (extract, load, transform) and ETL (extract, transform, load) are both data integration processes that move raw data from a source system to a target database, such as a data lake or data warehouse. These data sources can be in multiple, different repositories or in legacy systems that are then transferred using ELT or ETL to a target data location.

What is ELT (extract, load, transform)?

With ELT, unstructured data is extracted from a source system and loaded onto a target system to be transformed later, as needed. This unstructured, extracted data is made available to business intelligence systems, and there is no need for data staging. ELT leverages data warehousing to do basic data transformations, such as data validation or removal of duplicated data. These processes are updated in real-time and used for large amounts of raw data. ELT is a newer process that has not reached its full potential compared to its older sister, ETL. The ELT process was originally based on hard-coded SQL scripts. Those SQL scripts are more likely to have potential coding errors than the more advanced methods used in ETL.

What is ETL (extract, transform, load)?

With ETL, unstructured data is extracted from a source system and specific data points and potential "keys" are identified prior to loading data into the target systems. In a traditional ETL scenario, the source data is extracted to a staging area and moved into the target system. In the staging area, the data undergoes a transformation process that organizes and cleans all data types. This transformation process allows for the now structured data to be compatible with the target data storage systems. ETL was originally designed to work with relational databases, which historically dominated the market. Data engineers have worked on ETL processes since the 1970s, giving them time to refine the data science ETL processes significantly.

In the following video, Jamil Spain gives a deeper dive into ETL:

Watch the video

Benefits of ELT and ETL

ELT

The ELT approach enables *faster* implementation than the ETL process, though the data is messy once it is moved. The transformation occurs after the load function, preventing the migration slowdown that can occur during this process. ELT decouples the transformation and load stages, ensuring that a coding error (or other error in the transformation stage) does not halt the migration effort. Additionally, ELT avoids server scaling issues by using the processing power and size of the data warehouse to enable transformation (or scalable computing) on a large scale. ELT also works with cloud data warehouse solutions to support structured, unstructured, semi-structured and raw data types.

ETL

ETL takes longer to implement but results in *cleaner* data. This process is well suited for smaller target data repositories that require less frequent updating. ETL also works with cloud data warehouses by using cloud-based SaaS platforms and onsite data warehouses.

There are also many open-source and commercial ETL tools with capabilities and benefits that include the following:

- **Comprehensive automation and ease-of-use functions** that can automate the entire data flow and make recommendations on rules for the extract, transform and load process.
- **A visual drag-and-drop interface**used for specifying rules and data flows.
- **Support for complex data management** to assist with complex calculations, data integrations and string manipulation.

- **Security and compliance** that encrypt sensitive data — both in motion and at rest — and are certified compliant with industry or government regulations like HIPAA and GDPR. This provides a more secure way to encrypt, remove or mask specific data fields to protect client's privacy.

ELT and ETL use cases

ELT

An ELT process is best used in high-volume data sets or real-time data use environments.

Specific examples include the following:

- **Organizations with large volumes of data**: Meteorological systems like weather services collect, collate and use large amounts of data on a regular basis. Businesses with large transaction volumes also fall into this category. The extract, load and transform process allows for quicker transfers of source data.
- **Organizations that require instant access:** Stock exchanges generate and use large amounts of data in real-time, where delays can be harmful. Additionally, large-scale distributors of materials and components need real-time access to current data for business intelligence.

ETL

ETL is best used for synchronizing several data use environments and migrating data from legacy systems. The following are some specific examples:

- **Organizations that need to synchronize data from several sources**: Companies that are merging their ventures may have multiple consumers, supplies and partners in common. This data can be stored in separate data repositories and formatted differently. ETL works to transform the data in a unified format before loading it onto the target data location.
- **Organizations that need to migrate and update their data from legacy systems**: The legacy systems require the ETL process to transform the data into a compatible format with the new structure of the target database.

ELT vs. ETL: Main differences

The key differences between ELT and ETL are the order of operations between the two processes that make them uniquely suited for different situations. Other differences are in data size and data types that each process can handle. Although ELT and ETL sound similar, their application is different.

ELT

With ELT, the process is simplified in that it does not need "keys" or other identifiers for the data to be transferred and used. The ELT process has been refined and there are many evolved ELT tools that are used to help migrate data. Load times are shorter because the process does not have as many steps to go through. The ELT solution for business intelligence systems comes from a need to be able quickly load unstructured data. A cloud-based automated ELT solution can also be relatively low maintenance to support.

ETL

ETL data delivers more definition from the onset, which usually requires more time to transfer the data accurately. This process only requires periodic updates of information, rather than real-time updates. ETL load times are longer than ELT because of the many steps in the transformation stage that must occur before loading the data.

ELT, ETL and IBM Cloud

IBM offers several data integration services and solutions designed to support a business-ready data pipeline and give your enterprise the tools it needs to scale efficiently.

IBM, a leader in on-premises and cloud-based data integration, gives enterprises the confidence they need when managing big data projects, applications and machine learning technology. With industry-leading platforms like IBM Cloud Pak® for Data, organizations can modernize their DataOps processes while being able to use best-in-class virtualization tools to achieve the speed and scalability their business needs now and in the future.

For more information on how your enterprise can build and execute an effective data integration strategy, explore the IBM suite of data integration offerings.

Sign up for an IBMid and create your IBM Cloud account.

IBM Cloud Education

IBM Cloud Education