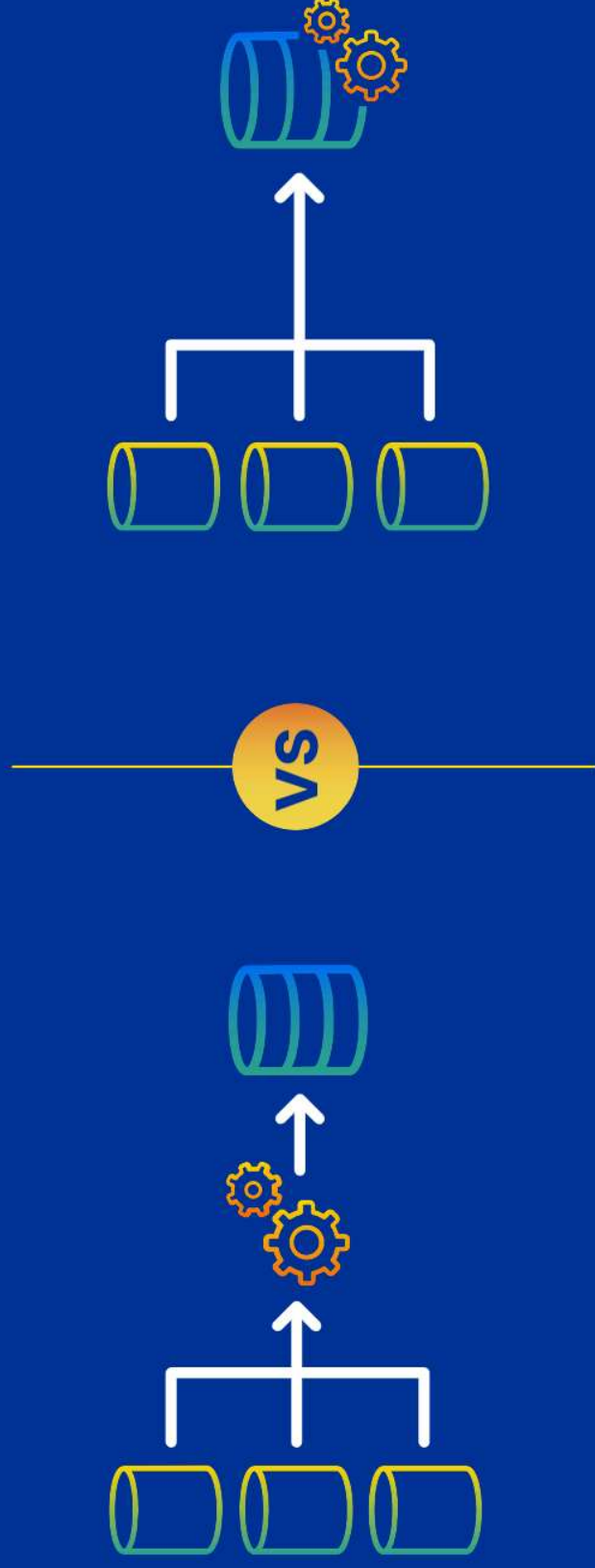


ETL vs. ELT

Which One Is Right for Your Organization?



Where the transformation happens

As part of the data integration workflow, data transformation is the process of preparing, formatting, and joining data for placement in a data repository or data warehouse for analytics. And for decades, the data transformation method of choice has been ETL (extract, transform, load). The point of data transformation is to format and structure data extracted from a source to be compatible with or to join the schema of the destination data warehouse. For data warehouses that are SQL (Structure Query Language) oriented, a data schema is necessary to ask questions of the data. Thus, transformation is performed before the data is loaded.

Today, however, thanks to the rise of new-generation cloud data warehouses, a second option has become popular: ELT (extract, load, transform). The ELT integration design pattern comes into play when the schema (and perhaps the questions you intend to ask about the data) is not known in advance. Therefore, the transformation is deferred until a later time, after the data is loaded.

Another variation of ELT is when the required transformation is known in advance of the data loaded, and the transformation is developed in the pipeline, but the execution of the transformation is pushed down (transferred) to the new-generation data warehouse.

Does that mean ELT is the more favorable process? Not always. Is ELT right for you? It depends. The choice between ETL and ELT depends on several factors that are unique to each organization, including data schema requirements, transformation complexity, performance, and budget constraints, just to name a few. And that's what we've outlined in this guide.

We'll cover:

- The **ETL** approach, its benefits, and common challenges
- The **ELT** approach, its benefits, and common challenges
- A side-by-side comparison of functionality between ETL and ELT
- Considerations for which approach is best for your organization
- How to simplify the data transformation process with either approach

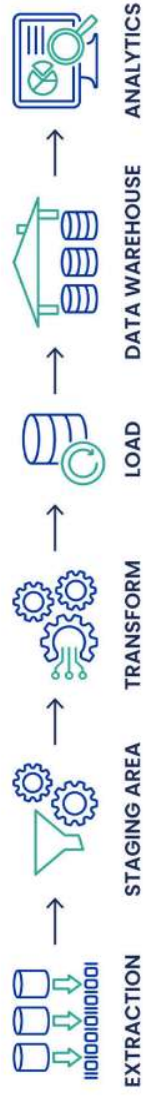
Whether you're a data practitioner or a technology executive, read on to learn the functional differences between the two processes and explore their respective considerations and use cases.



What is ETL?



ETL stands for **Extract, Transform, and Load**. In this integration design pattern, data is extracted from one or more sources and then moved temporarily to a staging environment. Once in the staging environment, the data is cleansed and transformed *before* being loaded into a data warehouse (or other data store) for storage and analysis or joined with other data.



The driving goal of ETL is to establish a schema for the incoming data or to match the schema of the incoming data with the schema of an existing database in the data warehouse, to which the incoming data will be inserted. The ETL integration design process can be implemented with data that is targeted for data warehouses that are either cloud or on-premises based.

ETL: Data transformations are handled in the data pipeline



Extract

Data is extracted from source(s) and moved to a staging area. Data sources can also include non-relational data files like JSON documents.

Upon extraction, data is validated to ensure its accuracy. Data that fails validation rules will be rejected, then discarded.

Common data-source formats include:

- Relational databases
- JSON
- Flat files
- XML

Transform

The validated data is cleansed in the staging area. This crucial part of the data transformation process involves identifying corrupt, duplicate, irrelevant, or misrepresentative data and then replacing, modifying, or deleting it.

Other transformations occur so the data can be stored in a useful form. Sensitive data is scrubbed, encrypted, and protected before it is exposed to business users.

Common transformations include:

- Sorting and filtering
- Merging data from multiple sources
- Combining or splitting rows and columns
- Normalizing schemas
- Performing basic calculations

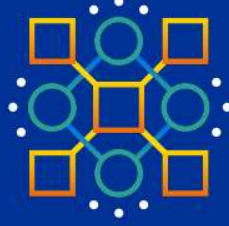
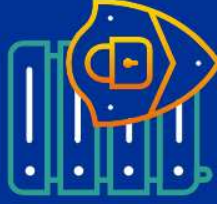
Load

The data is loaded to its end target for storage and analytics. The process for loading the data varies according to organizational needs.

Constraints defined within the database may also trigger upon load, further filtering the data. The database may filter out duplicates that already exist in the database, reject data that is missing mandatory fields, or perform other actions based on the parameters set by the organization. The stored data is now ready for further analysis.

Popular data analytics tools include:

- Tableau
- Microsoft Power BI
- BusinessObjects (SAP)



Benefits of ETL

+ Convenience and simplicity

Handling transformations early in the integration cycle, within the pipeline, is a natural motion for data integration specialists and for those who work with data.

+ Early masking for compliance

Sensitive data sets can be masked or encrypted earlier in the data integration process, making it easier to satisfy GDPR, HIPAA, CCPA, and other compliance standards. Transforming data before it gets to the data warehouse or other destination reduces the risk of exposing sensitive data.

+ Familiarity

ETL has been around for decades; there are numerous ETL tools in the Data Integration space and ETL knowledge is a common skill set among integration specialists.

Challenges of ETL

- Defining the schema in advance

Transforming data into a schema implies knowing the schema in advance. For new, fresh data, the questions that need to be asked of the data may not yet be known, creating a challenge to define the schema in advance.

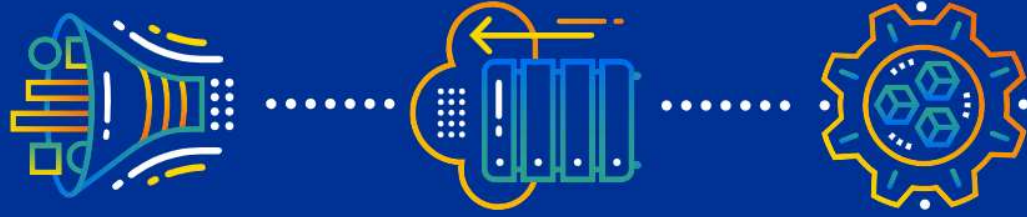
- Potential performance strain

Larger and more complex transformations may require additional processing resources to boost performance.



[Read this ebook to learn more](#)





What is ELT?



ELT stands for **Extract, Load, and Transform**. With this data integration design pattern, data is loaded into a data warehouse before any transformations take place.

The ELT integration design pattern may be necessary if knowledge about the data to be loaded is not known and it is therefore uncertain what questions to ask of the data, thereby inhibiting creation of a data schema in advance. After the data has reached its target destination and it is explored, it can be transformed as needed.



New generation cloud data warehouse technology that feature data staging capabilities make the ELT data integration design pattern possible. Platforms like Amazon Redshift-Spectrum, Google BigQuery with Google Cloud Storage, Snowflake, Databricks and others can accommodate loading of data into a storage area and then transforming the data later. Note however, some platforms enable this capability more easily than others.

ELT: When necessary data transformations are handled in the **data warehouse**



Extract

Data is extracted from source(s) and moved to a staging area. Unlike ETL, the data does not undergo a validation process at this stage.

In the extraction process, data is pulled from sources and the frequency and trigger of refresh are set.

Common data-source formats include:

- Relational databases
- XML
- JSON
- Flat files

Load

The data is immediately loaded in its raw format to the data lake, where it will be warehoused.

Popular data lake solutions include:

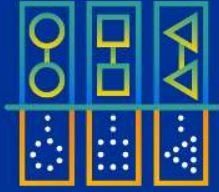
- Amazon S3
- Azure Data Lake Storage
- Google Cloud Storage
- Hadoop HDFS

Transform

The data is transformed on an as-needed basis. This saves time in the long run because people won't be applying unnecessary data transformations.

Popular data transformation tools include:

- [SnapLogic](#)
- Python
- MATLAB
- DBT



Benefits of ELT

+ Speed

Because it relies on cloud storage in data lakes, which can handle a massive amount of raw, unstructured, and semi-structured data, ELT provides faster data availability and data loading for quicker analytics.

+ Precision

ELT is ideal for smaller datasets with simple transformations that have minimal impact on cloud data warehouse processing utilization.

+ Less maintenance

ELT is cloud-based and does not require extensive, costly hardware and maintenance.

+ Scalability

Since ELT is flexible and unconstrained by size, the solution scales very easily.

Challenges of ELT

- Cost

ELT tools and resources can be more expensive than ETL tools because more information is processed in less time.

- Complexity

ELT can also require more skill than an ETL solution, which may necessitate hiring ELT experts.

- Security

With ELT, all data gets sent to the warehouse, potentially exposing organizations to HIPAA, CCPA or GDPR violations. However, it's possible to protect sensitive data during the process with encryption and proper data governance.



[Read this ebook to learn more](#)



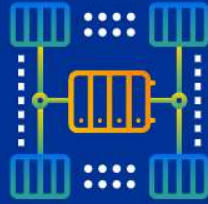
ETL vs. ELT: side-by-side

Oftentimes the choice of performing an ELT versus an ETL design pattern comes down to the integration infrastructure deployed, environment preferences, performance, and cost. The following table outlines important considerations.

Consideration	ETL	ELT
Destination schema	Whether utilizing an integration platform or relying on manual coding, the ETL design pattern requires mapping into a known schema or must support schema on read of raw data before transferring to the destination	The target destination system or data warehouse must support schema on read of received data or must accept data in native format for transformation later
Complex transformation performance	Generally, favors light to heavy transformation complexity against small to medium dataset sizes and is dependent on the integration platform size, configuration, and performance capabilities	Generally, favors medium to heavy transformation complexity against larger data sets and is dependent on the data warehouse size, configuration, and performance capabilities
Environment preference	Integration specialists/data engineers may prefer to develop transformations within an integration platform environment	Data warehouse specialists/data engineers may prefer to work within a data warehouse environment
Cost	May be best if the integration platform is already a sunk cost, particularly if it offers unlimited data processing at a flat pricing (non-consumption) structure	Care must be taken to evaluate cost, particularly if the target data warehouse/system is consumption price oriented

Which is best, ETL or ELT?

At the end of the day, the answer is, it depends. As detailed throughout this eBook, ETL and ELT are both viable solutions for mobilizing and transforming data. The above table outlines the important considerations; however, a proof of concept may be required to assess the best approach based on your particular use case.



Make data transformations easy with SnapLogic

Trustworthy migration and transformation processes fuel trustworthy data. Whether you choose ETL or ELT, the ability to trust and readily access your data is critical to the success of your business. SnapLogic delivers all of these data movement and transformation design patterns in a proven and powerful cloud-based platform that addresses the unique needs and goals of your organization.

Our [iPaaS](#) (integration platform as a service) can help you simplify, standardize, and deploy data and application integration and automate business processes. The iPaaS uses API endpoints that allow different applications within your organization to share information across a common communication channel, enabling more collaborative workflows.

Versatile and scalable, SnapLogic's data integration solution enables organizations to connect and load data from on-premises-based or cloud-based sources using either ETL or ELT design patterns. Furthermore, after you've enriched your data in a cloud data warehouse, you can connect to all of your applications with a reverse ETL design pattern all within the same SnapLogic platform, thereby eliminating data silos and preventing costly and complex technology sprawl with additional tools.

Featuring a graphical "drag-and-snap" user interface and AI-augmented data pipeline design assistance, SnapLogic is also simple enough for IT and non-technical users alike. Our AI-driven, next-step recommendation engine technology, [AutoSuggest](#), is backed with more than seven years and petabytes of metadata. You can create pipelines in 50% of the time compared to other visual approaches, and even faster compared to manual coding, making your integration environment leaner and fit to scale.

Taking it even further, we've introduced [SnapGPT](#), the first-ever generative integration tool that translates business intent into pipelines and workflows using natural language commands—making integrations easy for everyone.

Explore the entire [SnapLogic AI Suite](#), and sign up for a [free trial](#) of the SnapLogic Intelligent Integration Platform today.



Want to give it a spin? Sign up for a **free trial** today!



SnapLogic is the leader in **generative integration**. As a pioneer in AI-led integration, the SnapLogic Platform accelerates digital transformation across the enterprise and empowers everyone to integrate faster and easier. Whether automating business processes, democratizing data, or delivering digital products and services, SnapLogic enables you to simplify your technology stack and take your enterprise further. Thousands of enterprises around the globe rely on SnapLogic to integrate, automate and orchestrate the flow of data across their business.

Join the **generative integration** movement at snaplogic.com.

©2023 SnapLogic Inc. All rights reserved. | info@snaplogic.com | snaplogic.com

EE202311-L

