

When to Use ETL vs ELT When Loading Data Into a Data Warehouse



Table of Contents

1. Introduction
2. What Is ETL?
3. What Is ELT?
4. ETL vs. ELT When Loading Data: What's the Difference?
5. ETL vs. ELT: 4 Reasons To Use ETL When Loading Data
6. ETL vs. ELT: 3 Reasons To Use ELT When Loading Data
7. How Integrate.io Can Help With ETL vs. ELT

1. Introduction



Here's what you need to know about ETL vs. ELT when loading data into a data warehouse:

- ETL transforms data before loading it into a repository (typically a data warehouse).
- ELT loads data into a repository (typically a data lake or data lakehouse) before transforming it.
- The ordering of the transformation and loading stages in ETL and ELT substantially affects the use cases of both techniques.
- ETL is better for businesses that work mainly with relational data or are concerned about data privacy and security.
- ELT is better for businesses that use semi-structured, unstructured, or streaming data or require flexibility when running analytical queries.

Two of the most popular methods of data integration are ETL and ELT. These techniques help companies move their information into a data warehouse, making it easier to glean insights from business intelligence and data analysis.

Regarding ETL vs. ELT, however, these two approaches differ in the order in which they load and transform data. This distinction is essential for organizations looking to improve their data workflow and data warehousing.

So which of these data integration processes—ETL or ELT—should you use for data loading into a data warehouse? This article will discuss the key differences between ETL vs. ELT when loading data into a data warehouse, including the definitions of ETL and ELT and the use cases for each approach.

2. What is ETL?

ETL is the workhorse of modern data integration, helping businesses aggregate and analyze their enterprise data. But what is ETL exactly, and how does ETL work?

What Is the ETL Process?

ETL (extract, transform, load) is a data integration method that involves three sequential stages: extracting, transforming, and loading information.



Extract: First, raw data is extracted from one or more data sources. These sources may include **flat files**, non-relational and relational databases (i.e., NoSQL and SQL databases), websites, SaaS (software as a service) applications, and more. The data sets extracted may be internal or external to the organization.



Transform: After ETL tools extract data from source systems, this source data goes through **data transformation** inside a separate staging area. These transformations may include relatively lightweight changes to improve data quality, known as **data cleansing**. More heavy-duty modifications may also be necessary to make the data compatible with the schema of the target database or to integrate multiple data sources.



Load: Finally, the transformed data is loaded into its final destination, known as the “target system.” The target system is usually a **data warehouse** optimized for data storage at scale, such as **Amazon Redshift** or **Snowflake**. Data warehouses include an **OLAP (online analytical processing)** layer to help users discover valuable insights through business intelligence and analytics.

What Is an ETL Tool?

Data integration is a highly technical process that requires **data scientists and data engineers'** expertise. What's more, many organizations work with **big data**, making it highly infeasible to perform ETL manually.

Instead of building the ETL process from scratch, companies use **ETL tools** that help automate and streamline their data integration workflow, especially for non-technical users. These solutions often come with pre-built connectors for the various data sources and targets in your IT environment, saving you the trouble of writing code to link them up manually. They also include diverse data transformations to help users clean, manipulate, and enrich their information.

The ETL Process In-Depth

The ETL process begins by identifying the data types and sources businesses want to send to their data warehouse. These may be flat files (such as XML and Excel documents), databases containing non-relational or relational data, SaaS platforms, APIs (application programming interfaces), websites, or analytics tools.

Next, users need to decide how (and how often) they want to perform the extraction stage:

- Performing a **full extraction** of all data ensures that all changes are accounted for. However, this becomes highly inefficient for all but the smallest data sets.
- Some source systems and databases can **notify users** when they have been modified, letting users extract only the relevant changes. However, this approach may not be available for all data sources.

Change data capture (CDC) is a feature offered by some ETL tools that automatically detects changes to a database or record using logs or triggers. Users can then perform fast, lightweight data extractions to enjoy updated information.

2. What is ETL?

Next, users need to decide how (and how often) they want to perform the extraction stage:

- Data cleansing is the process of removing old, inaccurate, and duplicate information from a data set to improve its quality. The transformations involved in data cleansing are generally quick and efficient.
- Businesses may perform more substantial changes, requiring more time and processing power. One example is combining information from two different databases with a join operation. Another is data summarization: creating new records by performing computations on existing records.
- The ETL process sometimes involves handling individuals' personal or sensitive data. This information may need to be anonymized or encrypted through data transformation before it reaches its destination.

Related Reading: [Data Transformation: Explained](#)

Finally, the transformed data is loaded into the target data repository, such as a data warehouse. This information may be loaded either in batches (at regular intervals) or with streaming loads, adding new data in real time.

A data warehouse is the most common destination of the ETL process, but not the only one. A [data mart](#) is essentially a mini-data warehouse for the use of a specific team or department. As we'll see in the next section on ELT, businesses may also use data warehouse variants that are better at working with unstructured data.



3. What is ELT?

First developed in the 1970s, ETL remains the dominant form of data integration. In recent years, however, ETL alternatives such as ELT have been created to better deal with data size and variety changes.

What Is the ELT Process?

As the name suggests, [ELT \(extract, load, transform\)](#) switches the order of the last two stages in the ETL data pipeline. Instead of changing data before loading and storing it, the ELT process loads data into the target repository as soon as possible. This data can then be transformed on-demand during BI and analytics workloads.

Unlike ETL, ELT does not initially perform data cleansing or standardization—much less heavy transformations. Thus, ELT does not require a separate staging area for data transformations, as ETL does.

Instead, these changes occur inside the final data repository as necessary once these records are used. This means that the ELT process is relatively low-maintenance versus ETL. With ETL, businesses must decide which transformations to perform ahead of time since large-scale changes can't be made to the data once it's inside the repository.

Integrate.io is an ETL and ELT data integration platform with more than 140 pre-built connectors. Integrate.io's user-friendly, drag-and-drop interface makes it easy for anyone to build ETL and ELT workloads. [Speak with our team of data experts today](#) to start your 14-day pilot of Integrate.io.

What Is an ELT Tool?

Compared with ETL, ELT is still relatively new. However, several ELT solutions have emerged to help companies perform the ELT process. These data integration tools either support ELT exclusively or can perform both ELT and ETL. However, ELT is a more novel approach to data integration, and many ELT tools and systems lack the features and functionality of their full-scale ETL counterparts. Users should look for a robust ELT solution that can match their business needs and objectives.

The ELT Process In-Depth

The rise of ELT is primarily thanks to the recent growth in cloud computing, which has enabled unprecedented levels of storage and scalability. Without the technological advances of cloud computing, ELT might prove too costly for many businesses since it typically demands more storage and processing power than ETL. This is because ELT loads data indiscriminately without first removing duplicate or out-of-date records.

One major ETL vs. ELT difference is that the target repository in ELT often has a different structure than a data warehouse. Data warehouses have a relational structure in which information is organized into tables, rows, and columns. However, much of today's enterprise data consists of semi-structured or [unstructured data](#)—e.g., images, text, audio, and video—that cannot fit neatly into a relational database. Instead, unstructured data is stored using alternative database structures such as graphs and key-value stores.

A [data lake](#) is a repository for storing large amounts of data, particularly semi-structured and unstructured data (although they can also store structured relational data). Data lakes, rather than data warehouses, are typically the final destination of the ELT process. Another option is the [data lakehouse](#), which attempts to combine the best of both data lakes and data warehouses.

4. ETL vs. ELT When Loading Data: What's the Difference?

On the surface level, the main difference between ETL vs. ELT is the order of the transformation and load processes. ETL transforms data before it reaches the data warehouse, while ELT transforms it after loading, often on the fly.

This simple change to the order of extract, transform, and load has a significant effect:

- **In ETL**, data is already transformed once it reaches the target system and generally cannot be further changed inside the repository. **In ELT**, data has not been transformed before reaching the target system.
- **In ETL**, the target repository is typically a data warehouse optimized for storing relational data. **In ELT**, the target repository is typically a data lake that can hold semi-structured and unstructured data.

What's more, these two differences in the ETL vs. ELT process have widespread implications for the use cases of these two approaches. In the following two sections, we'll explore some of the biggest reasons businesses might use ETL or ELT when loading data.



5. ETL vs. ELT: 4 Reasons To Use ETL When Loading Data

Despite the rise of ELT and other alternatives, ETL remains a viable option for many data integration use cases.

1. You have lots of complex or relational data

ETL is optimized for working with relational data that fits into an organized database schema with tables, rows, and columns. Because a relational data structure is predictable, OLAP data warehouses can process and analyze it much more quickly.

ETL pre-processes the data, transforming and cleansing it before it reaches the data warehouse. In ETL, records only need to be changed once before they can be used in BI and analytics workloads an unlimited number of times.

As a result, ETL is especially well-suited for data sets that require complex transformations. Performing these changes once during the transformation phase of ETL saves you from having to repeat them every time you want to analyze the data, as you would with ELT.

2. You need a more straightforward implementation

As the decades-long industry standard for data integration, ETL is a well-known and well-established technique with multiple benefits. For one, finding data science and data integration experts skilled in ETL is easier. There are also more ETL tools on the market, and the ones available are generally more feature-rich and mature.

On the other hand, ELT is a newer concept that has only gained speed in the last decade as technology can accommodate it. Finding data integration professionals with specific expertise in ELT will likely be more difficult. The ELT tools on the market are also typically less advanced, and the ELT tech ecosystem still has not caught up with ETL.

3. You have concerns about regulatory compliance

Companies that process large amounts of data may be subject to several privacy and security regulations. These regulations govern how businesses collect, process, and store sensitive data.

[HIPAA](#), for example, is a U.S. federal law that requires healthcare organizations to protect information about patients' health. Meanwhile, the European Union's [GDPR \(General Data Protection Regulation\)](#) limits how companies doing business in the EU can handle personal data.

Organizations must take regulations such as HIPAA and GDPR seriously when working with confidential information or face harsh penalties and legal repercussions. However, one unintended consequence of using ELT instead of ETL is that it may violate these regulations.

Suppose, for instance, that your business processes the personal information of EU citizens. If you use ELT to load data into a cloud-based data warehouse, this information will not yet have gone through transformation before it is stored in the cloud. Unlike on-premises data, data "in the cloud" must be located on a remote physical server, which may be outside the EU.

The combination of ELT and cloud data warehouses thus risks breaching the GDPR's ["third countries"](#) rule. This rule only permits companies to transfer personal data to a limited number of countries outside the EU. The laws of these countries have been deemed to offer an "adequate level of protection" for personal data, similar to EU law.

Related Reading: [What It Actually Means to Be HIPAA Compliant](#)

5. ETL vs. ELT: 4 Reasons To Use ETL When Loading Data

4. You want to protect data privacy

Even if there are no formal regulations against using ELT, it can expose your business to data privacy issues. If sensitive information travels to the target database before being transformed (as in ELT), it may be disclosed to unauthorized parties while in transit. Businesses that fail to protect the privacy of their data may suffer significant legal, financial, and reputational damage.

During the data integration process, sensitive data may appear in system logs that are visible to administrators, unintentionally causing a privacy breach. Moreover, malicious actors who have gained access to your network may be able to eavesdrop and exfiltrate confidential information.

Therefore, it may be necessary to transform data by anonymizing or encrypting it before it is loaded into the repository. This is only possible with ETL, which includes an explicit transformation stage. Businesses have also used alternatives such as ETLT to deal with these issues and get the best of both worlds. ETLT is a variant of ETL that performs the transformation process twice: once before loading and once after. The first transformation stage involves relatively light changes while data is in a staging area, and the second involves more thorough integrations of multiple data sources.

Related Reading: [Security and ELT — A Tragedy](#)



6. ETL vs. ELT: 3 Reasons To Use ELT When Loading Data

Despite the advantages of ETL above, there are several good reasons ELT might win the ETL vs. ELT battle.

1. You have lots of unstructured data

ETL struggles to incorporate semi-structured and unstructured data inside a traditional relational data format. If these data types are a significant portion of your enterprise data, then ELT is strongly preferred. The ELT process typically sends data to a target “data lake”: a variant of the data warehouse that can handle large quantities of raw information. Unstructured data does not have to be transformed before it enters the data lake, as it would with an ETL data warehouse. ELT allows businesses to save any information without organizing it and giving it structure.

2. You want to load lots of data fast

ELT is almost always faster than ETL regarding how long it takes for input data to be loaded. Because ELT completely omits the transformation stage before loading, your data has a straight path from source to target.

The reordering of data transformation makes ELT a good choice for streaming and real-time use cases, where massive amounts of data continuously arrive faster than they can be processed and transformed. Whereas ETL would be incapable of handling this constant firehose of information, ELT can load it into the data lake unmodified and apply changes later.

3. You don't know how you want to use the data

ETL requires all data transformations to be done before reaching the warehouse. Performing these changes before loading improves the speed and efficiency of BI and analytics but makes the ETL process less flexible and poorly suited for ad hoc queries.

If you aren't sure in advance what kind of analytical queries you will run on the data, ELT is a superior choice. ELT makes your data integration processes more elastic, enabling you to run different transformations at different times.



7. How Integrate.io Can Help with ETL vs. ELT



Whatever you decide on ETL vs. ELT, you need a tool to help you achieve your data integration goals.

Integrate.io is a powerful, feature-rich, user-friendly solution for structured data integration using ETL and ELT. The Integrate.io platform has been designed from the ground up for the needs of E-commerce businesses.

Using Integrate.io, it's never been easier for users to build production-ready data pipelines between their source and target systems. The Integrate.io platform includes a simple, user-friendly, drag-and-drop visual interface for users of all technical levels. Integrate.io comes with more than 140 pre-built connectors and integrations and can link to any service using a REST API.

Integrate.io is packed with valuable features and functionality to fit a variety of data integration workflows. For example, Integrate.io's **CDC (change data capture)** automatically detects when certain records have been modified or deleted so that your next integration job only includes these records. Integrate.io also supports **reverse ETL**: moving data out of a centralized data warehouse and into third-party systems. Reverse ETL makes it easier for non-technical users to access and analyze this information.

Ready to learn how Integrate.io can help make your business more efficient and data-driven?

Get in touch with our team of data integration experts today to discuss your business needs and objectives. You can also sign up for a 14-day pilot of the Integrate.io platform.

**We hope that you have enjoyed
reading this white paper.**

If so, be sure to check out our other [White papers](#) related to
"When to Use ETL vs ELT When Loading
Data Into a Data Warehouse".

**Learn more about how Integrate.io can help
make your data work for you!**

GET A LIVE DEMO!

