

ETL VS ELT

Differences between ETL and ELT



When and where the transformations happen:

- Transformations for ETL happen within the data pipeline
- Transformations for ELT happen in the destination environment



Flexibility:

- ETL is rigid: pipelines are engineered to user specifications
- ELT is flexible: end users build their own transformations

- Differences between ETL and ELT, for one thing, the transformations happen in a different order. Transformations for ETL pipelines take place within the data pipeline before the data reaches its destination, whereas transformations for ELT are decoupled from the data pipeline and happen in the destination environment at will.
- They also differ in flexibility in how they can be used. ETL is normally a fixed process meant to serve a very specific function, whereas ELT is flexible, making data readily available for self-serve analytics.

Differences between ETL and ELT



Support for Big Data:

- Organizations use ETL for relational data, on-premise: scalability is difficult
- ELT solves scalability problems, handling both structured and unstructured Big Data in the cloud



Time-to-insight:

- ETL workflows take time to specify and develop
- ELT supports self-serve, interactive analytics in real-time

- They also differ in their ability to handle big data. ETL processes traditionally handle structured relational data, and on-premises computing resources handle the workflow. Thus, scalability can be a problem. ELT, on the other hand, handles any kind of data structured and unstructured.
- And to handle scalability problems posed by big data, ELT leverages the on-demand scalability offered by cloud-computing services.
- With regard to data discovery and time-to-insight, ETL pipelines take time and effort to modify, which means, users must wait for the development team to implement their requested changes. ELT provides more agility with some training in modern analytics applications, end users can easily connect to and experiment with the raw data, create their own dashboards, and run predictive models themselves.

The evolution of ETL to ELT

- Increasing demand for access to raw data



- In ELT, the staging area fits the description of a data lake
 - Staging areas — private ETL landing zones
 - Self-serve data platforms are the new “staging area”
-
- ELT is a natural evolution of ETL. One of the factors driving that evolution is the demand to release raw data to a wider user base for the enterprise. Traditionally, ETL processes include an intermediate storage facility called a staging area. This is a holding area for raw, extracted data where you can run processes prior to loading the resulting transformed data into a data warehouse or a data mart. This sounds a lot like an ELT process, and the staging area fits the description of a data lake, which is a modern self-serve repository for storing and manipulating raw data. A traditional staging area, however, is not something that is usually shared across the company. Its a private siloed area set aside for developing, monitoring and performance tuning the data pipeline and its built-in transformations. Along with the ever-increasing ease of use and connection capabilities of analytics tools, raw data sources have become much more accessible to less technical end users.

The shift from ETL to ELT

ETL still has its place for many applications



ELT addresses key pain points:

- Lengthy time-to-insight
 - Challenges imposed by Big Data
 - Demand for access to siloed information
- **Accordingly, the paradigm is shifting to self-service data platforms. There is still a place for conventional ETL in developing data pipelines, so ETL is not disappearing anytime soon. However, there is a trend taking place, a trend which is favoring modern ELT over conventional ETL. The trend is being driven by the pain points that ELT solves, namely, the lengthy time-to-insight, the challenges, for example, scalability imposed by big data, and the conventional siloed nature of data.**

Recap

In this video, you learned that:

- ETL and ELT differences include place of transformation, flexibility, Big Data support, and time-to-insight
- Increasing demand for access to raw data drives the evolution from ETL to ELT
- ETL still has its place
- ELT enables ad-hoc, self-service data analytics

