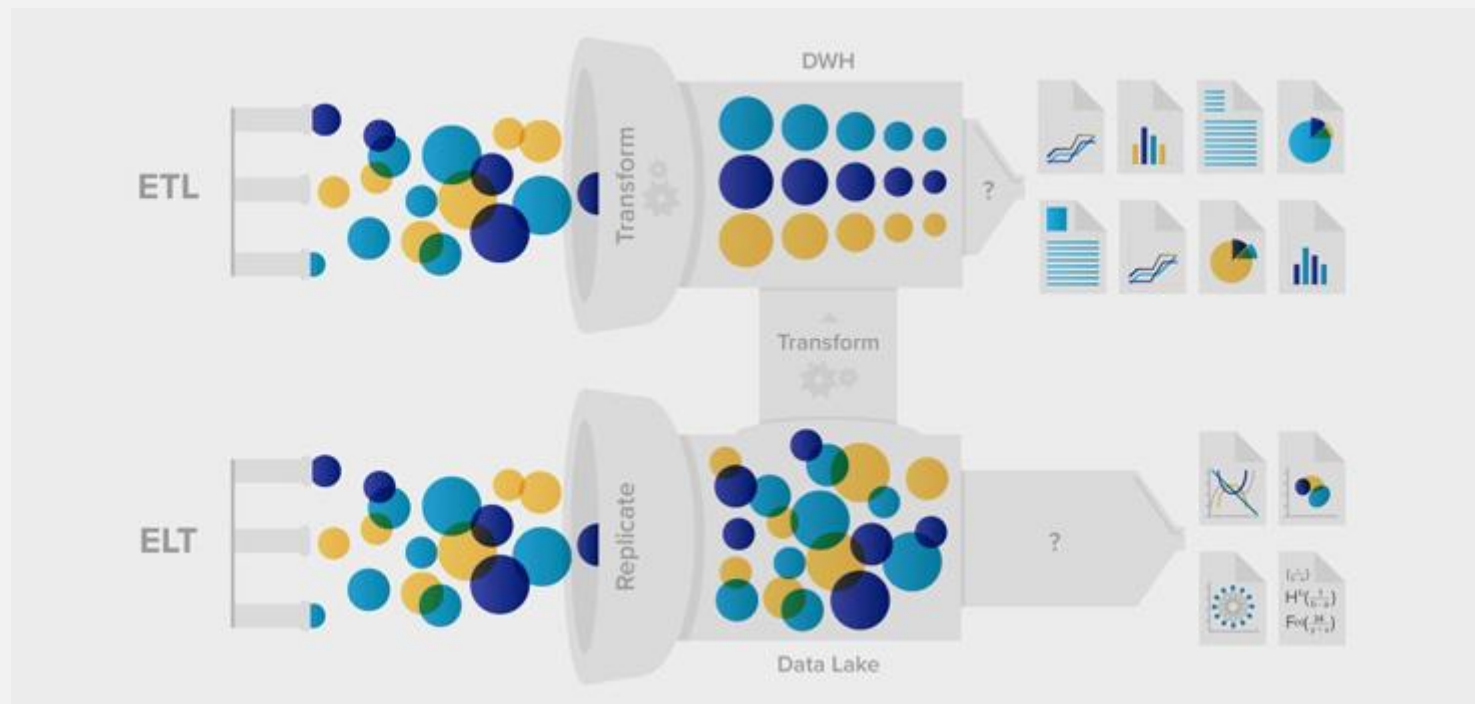


ETL vs. ELT



ETL AND ELT THE DOUBLE FACE OF DATA TRANSFORMATION

Today the data is the new oil, the business wants to analyse everything, get every insight, they want to “listen” to the data. Billions of events and transactions need to be collected, processed, enriched, modelled and transformed to extract their true value, trying to give them a “voice”.

But, which is the best architecture we should choose for our analytical projects? The main decision point is “how we do it” because there are meaningful differences in how data are managed during the load and how it is analyzed.

Data Transformation extract value from raw data applying the right transformations, gaining insights and giving to the business what they need.

Every processing stage - Extract, Transform and Load - need a deep integration between developers and data engineer, because they need to manage traditional warehouses limitations.

ETL vs. ELT Comparison

	ETL	ELT
Adoption of the technology and availability of tools and experts	ETL is a well-developed process used for over 20 years, and ETL experts are readily available.	ELT is a new technology, so it can be difficult to locate experts and more challenging to develop an ELT pipeline compared to an ETL pipeline.
Availability of data in the system	ETL only transforms and loads the data that you decide is necessary when creating the data warehouse and ETL process. Therefore, only this information will be available.	ELT can load all data immediately, and users can determine later which data to transform and analyse.
Can you add calculations?	Calculations will either replace existing columns, or you can append the dataset to push the calculation result to the target data system.	ELT adds calculated columns directly to the existing dataset.
Compatible with data lakes?	ETL is not normally a solution for data lakes. It transforms data for integration with a structured relational data warehouse system.	ELT offers a pipeline for data lakes to ingest unstructured data. Then it transforms the data on an as-needed basis for analysis.

Compliance	ETL can redact and remove sensitive information before putting it into the data warehouse or cloud server. This makes it easier to satisfy GDPR, HIPAA, and CCPA compliance standards. It also protects data from hacks and inadvertent exposure.	ELT requires you to upload the data before redacting/removing sensitive information. This could violate GDPR, HIPAA, and CCPA standards. Sensitive information will be more vulnerable to hacks and inadvertent exposure. You could also violate some compliance standards if the cloud-server is in another country.
Data size vs. complexity of transformations	ETL is best suited for dealing with smaller data sets that require complex transformations.	ELT is best when dealing with massive amounts of structured and unstructured data.
Data warehousing support	ETL works with cloud-based and onsite data warehouses. It requires a relational or structured data format.	ELT works with cloud-based data warehousing solutions to support structured, unstructured, semi-structured, and raw data types.
Hardware requirements	Cloud-based ETL platforms (like Integrate.io) don't require special hardware. Legacy, onsite ETL processes have extensive and costly hardware requirements, but they are not as popular today.	ELT processes are cloud-based and don't require special hardware.
How are aggregations different?	Aggregation becomes more complicated as the dataset increases in size.	As long as you have a powerful, cloud-based target data system, you can quickly process massive amounts of data.

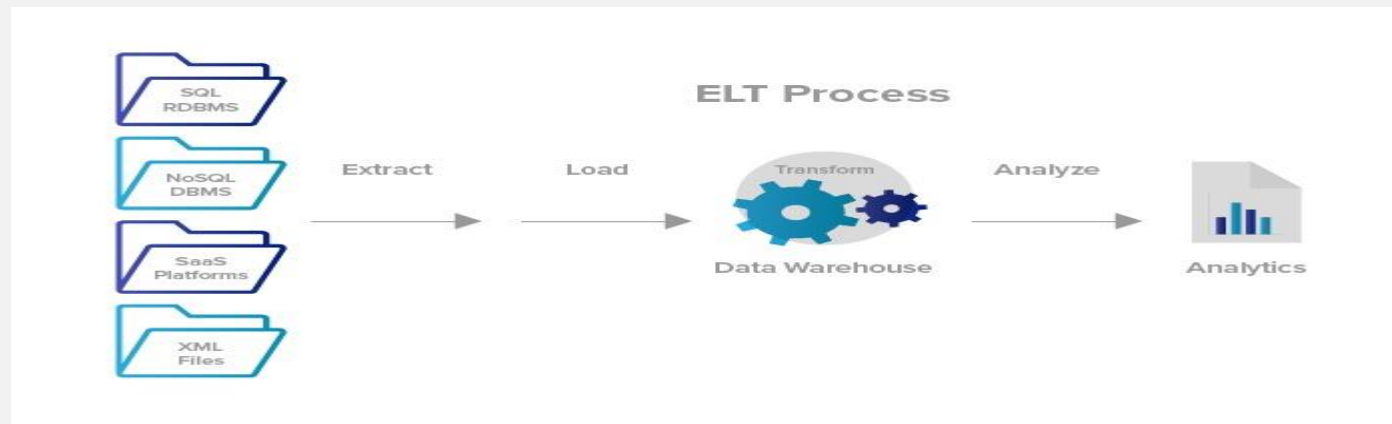
Implementation Complexity	ETL experts are easy to procure when building an ETL pipeline. Highly evolved ETL tools are also available to facilitate this process.	As a new technology, the tools to implement an ELT solution are still evolving. Moreover, experts with the requisite ELT knowledge and skills can be difficult to find.
Maintenance requirement	Automated, cloud-based ETL solutions, like Integrate.io, require little maintenance. However, an onsite ETL solution that uses a physical server will require frequent maintenance.	ELT is cloud-based and generally incorporates automated solutions, so very little maintenance is required.
Order of the extract, transform, load process	Data transformations happen immediately after extraction within a staging area. After transformation, the data is loaded into the data warehouse.	Data is extracted, then loaded into the target data system first. Only later is some of the data transformed on an “as-needed” basis for analytical purposes.
Costs	Cloud-based SaaS ETL platforms that bill with a pay-per-session pricing model (such as Integrate.io) offer flexible plans that start at approximately \$100 and go up from there, depending on usage requirements. Meanwhile, an enterprise-level onsite ETL solution like Informatica could cost over \$1 million a year!	Cloud-based SaaS ELT platforms that bill with a pay-per-session pricing model offer flexible plans that start at approximately \$100 and go up from there. One cost advantage of ELT is that you can load and save your data without incurring large fees, then apply transformations as needed. This can save money on initial costs if you just want to load and save information. However, financially strapped businesses may never be able to afford the processing power required to reap the full benefits of their data lake.

Transformation process	Transformations happen within a staging area outside the data warehouse.	Transformations happen inside the data system itself, and no staging area is required.
Unstructured data support	ETL can be used to structure unstructured data, but it can't be used to pass unstructured data into the target system.	ELT is a solution for uploading unstructured data into a data lake and make unstructured data available to business intelligence systems.
Waiting time to load information	ETL load times are longer than ELT because it's a multi-stage process: (1) data loads into the staging area, (2) transformations take place, (3) data loads into the data warehouse. Once the data is loaded, analysis of the information is faster than ELT.	Data loading happens faster because there's no waiting for transformations and the data only loads one time into the target data system. However, analysis of the information is slower than ETL.
Waiting time to perform transformations	Data transformations take more time initially because every piece of data requires transformation before loading. Also, as the size of the data system increases, transformations take longer. However, once transformed and in the system, analysis happens quickly and efficiently.	Since transformations happen after loading, on an as-needed basis—and you transform only the data you need to analyse at the time—transformations happen a lot of faster. However, the need to continually transform data slows down the total time it takes for querying/analysis.

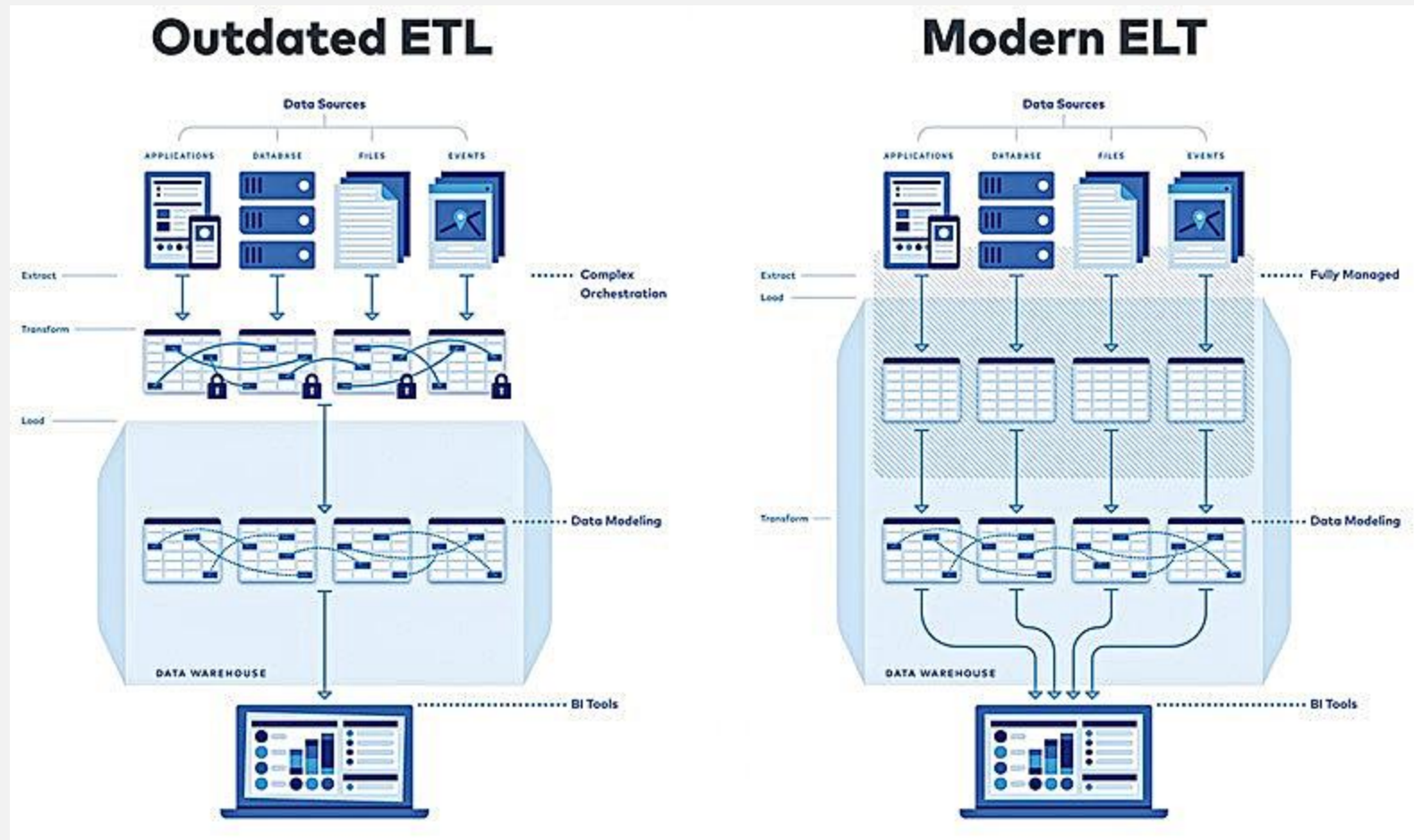
The five critical differences of ETL vs ELT:

- ETL is the Extract, Transform, and Load process for data. ELT is Extract, Load, and Transform process for data.
- In ETL, data moves from the data source to staging into the data warehouse.
- ELT leverages the data warehouse to do basic transformations. There is no need for data staging.
- ETL can help with data privacy and compliance by cleaning sensitive and secure data even before loading into the data warehouse.
- ETL can perform sophisticated data transformations and can be more cost-effective than ELT.

ETL Process:



Outdate ETL VS Modern ETL:



WHICH ONE WE HAVE TO CHOOSE?

- ETL:
 - Pros:
 - More horizontal scalability: **ETL can balance the load and divide it between RDBMS and the Data Transformation Engine**
 - **Parallel Processing: it is independently from the data model, database layout, database technology**
 - Streaming Processing **between source and target**
 - **SMP, MPP (Massively Parallel Processing) and Cloud Solution architectures**
 - Cons:
 - **Optimizations can be “tricky” data can be processed row-by-row in some functions and tools, with less degree of parallelism and of course, efficiency. (Sometimes this constraint is necessary and it is chosen)**
 - **Network Latency**
- ELT:
 - Pros:
 - **Data are read, processed and wrote in the same RDBMS, no network latency if source and target are in the same server**
 - **Native parallelism, I/O operations is generally optimized at DB engine level for high throughput**
 - **Can have 3x-4x higher throughput rate compared with an MPP (Massively Parallel Processing) RDBMS optimized platform**
 - Cons:
 - **Scalable and fast as much as the chosen RDBMS**
 - **Transformations are limited to database functionalities**
 - **In case of low-performance database or medium-small systems, there’s a chance to collide with the operational use, slowdowns can occur**
 - **If there are many sources, data must be migrated first, then transformed**

Ref :

[Architectures 001: Datawarehouse: ETL vs ELT \(curiousmind.name\)](#)