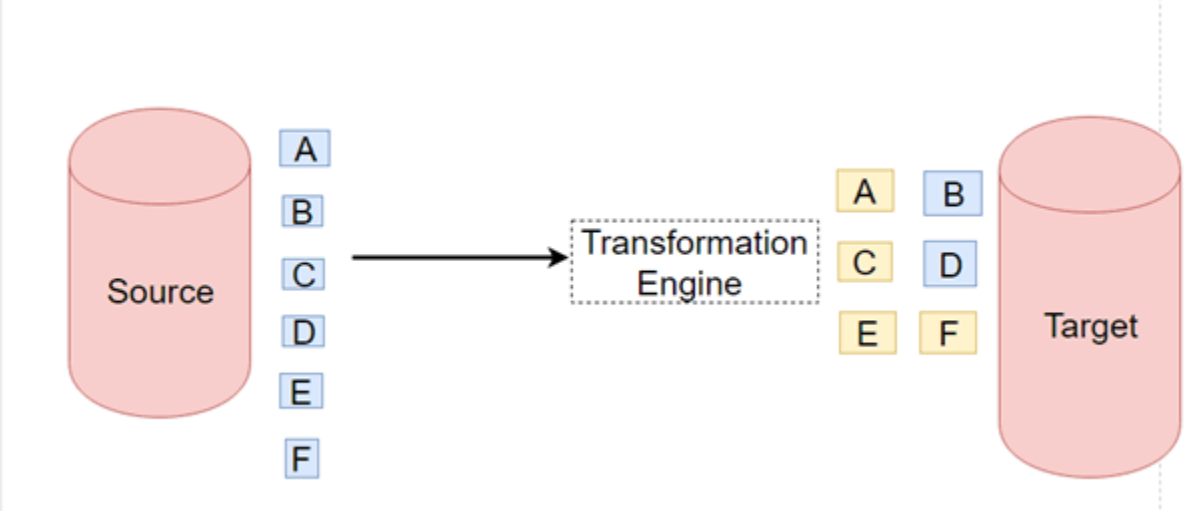
What is ETL?

ETL is an abbreviation of Extract, Transform and Load. In this process, an ETL tool extracts the data from different RDBMS source systems then transforms the data like applying calculations, concatenations, etc. and then load the data into the Data Warehouse system.

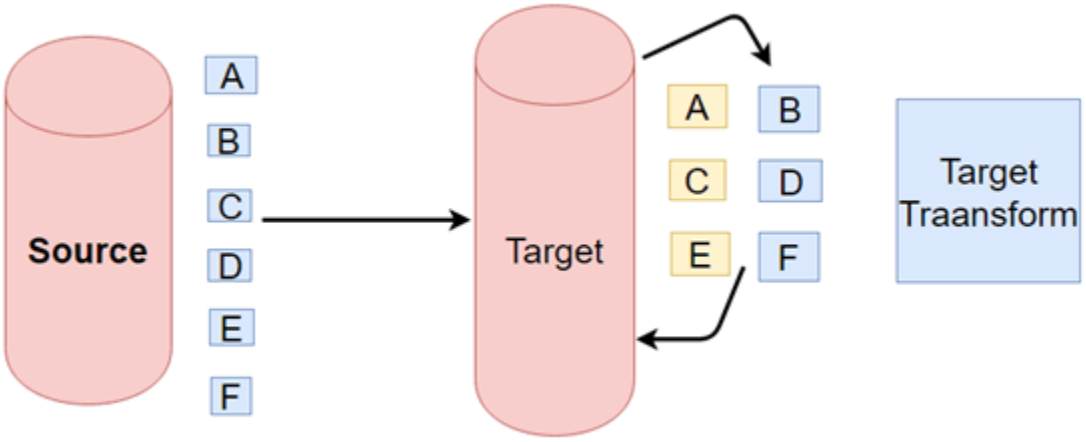
In ETL data is flows from the source to the target. In ETL process transformation engine takes care of any data changes.



What is ELT?

ELT is a different method of looking at the tool approach to data movement. Instead of transforming the data before it's written, ELT lets the target system to do the transformation. The data first copied to the target and then transformed in place.

ELT usually used with no-Sql databases like Hadoop cluster, data appliance or cloud installation.



KEY DIFFERENCE

- ETL stands for Extract, Transform and Load while ELT stands for Extract, Load, Transform.
- ETL loads data first into the staging server and then into the target system whereas ELT loads data directly into the target system.
- ETL model is used for on-premises, relational and structured data while ELT is used for scalable cloud structured and unstructured data sources.
- ETL is mainly used for a small amount of data whereas ELT is used for large amounts of data.
- ETL doesn’t provide data lake supports while ELT provides data lake support.
- ETL is easy to implement whereas ELT requires niche skills to implement and maintain.

Difference between ETL vs. ELT

ETL and ELT process are different in following parameters:

Parameters	ETL	ELT
Process	Data is transformed at staging server and then transferred to Datawarehouse DB.	Data remains in the DB of the Datawarehouse.
Code Usage	Used for • Compute-intensive Transformations • Small amount of data	Used for High amounts of data
Transformation	Transformations are done in ETL server/staging area.	Transformations are performed in the target system
Time-Load	Data first loaded into staging and later loaded into target system. Time intensive.	Data loaded into target system only once. Faster.
Time-Transformation	ETL process needs to wait for transformation to complete. As data size grows, transformation time increases.	In ELT process, speed is never dependant on the size of the data.
Time-Maintenance	It needs highs maintenance as you need to select data to load and transform.	Low maintenance as data is always available.
Implementation Complexity	At an early stage, easier to implement.	To implement ELT process organization should have deep knowledge of tools and expert skills.
Support for Data warehouse	ETL model used for on-premises, relational and structured data.	Used in scalable cloud infrastructure which supports structured, unstructured data sources.
Data Lake Support	Does not support.	Allows use of Data lake with unstructured data.

Parameters	ETL	ELT
Complexity	The ETL process loads only the important data, as identified at design time.	This process involves development from the output-backward and loading only relevant data.
Cost	High costs for small and medium businesses.	Low entry costs using online Software as a Service Platforms.
Lookups	In the ETL process, both facts and dimensions need to be available in staging area.	All data will be available because Extract and load occur in one single action.
Aggregations	Complexity increase with the additional amount of data in the dataset.	Power of the target platform can process significant amount of data quickly.
Calculations	Overwrites existing column or Need to append the dataset and push to the target platform.	Easily add the calculated column to the existing table.
Maturity	The process is used for over two decades. It is well documented and best practices easily available.	Relatively new concept and complex to implement.
Hardware	Most tools have unique hardware requirements that are expensive.	Being Saas hardware cost is not an issue.
Support for Unstructured Data	Mostly supports relational data	Support for unstructured data readily available.