



ETL

Database Integration



ETL



Extract

Pull data from source systems and moving to a staging area.



Transform

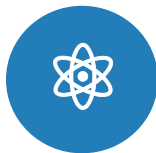
Prepare the data to be loaded into the target system



Load

Load the data into the target system

Extract



Source Systems

On-Premises ERPs, cloud systems like Salesforce, or raw data sources like CSV files and SQL databases.



Read Data from Source and Write into Staging

Avoid impacting performance on source system.
Write into staging area that can be used for next step in the pipeline.



Change Data Tracking

Although it's possible to pull the entire source data set each time, it's more efficient to only pull changes since the last batch.

Transform



Data Cleansing

Cleansing the source data so that bad data is not loaded into the target system. Common tasks: eliminating bad records, removing duplicates, or fixing formatting problems.



Data Enrichment

Supplementing the data with additional information that is not present in the source system but necessary to load into the target system.



Data Munging

Data manipulation required to format the data in a way that is ready for loading. Includes selecting certain columns, encoding values, deriving calculated values, aggregating data, or transposing data.

Load



Load into Target Database

Once the data is loaded and transformed, it's ready for loading into the target database.



Performance Impact

Loading large amounts of data into a database using conventional SQL can be extremely slow. Use bulk data load tools when possible.



Master Data

Reference information that is used across many tables in a database. Handle foreign key constraints.



Slowly Changing Dimensions

A standard data warehousing practice that stores changes to master data over time. For example, a customer address change.