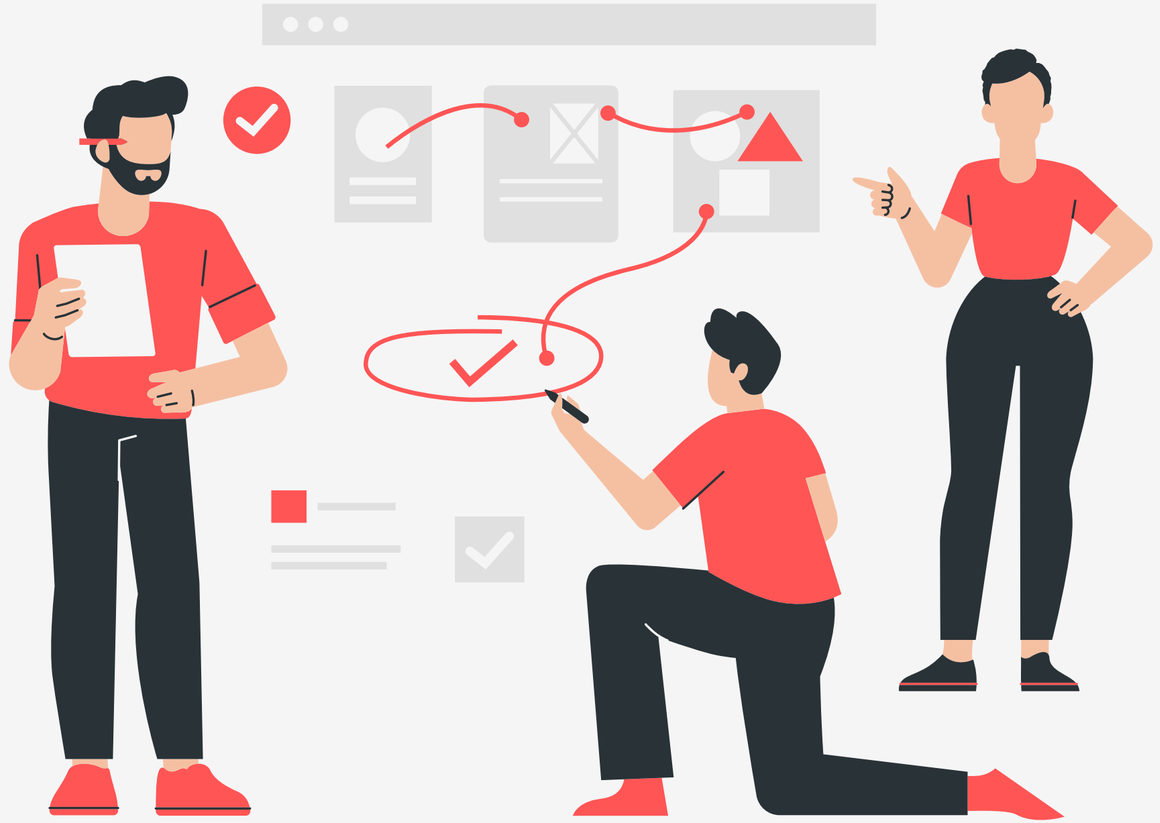
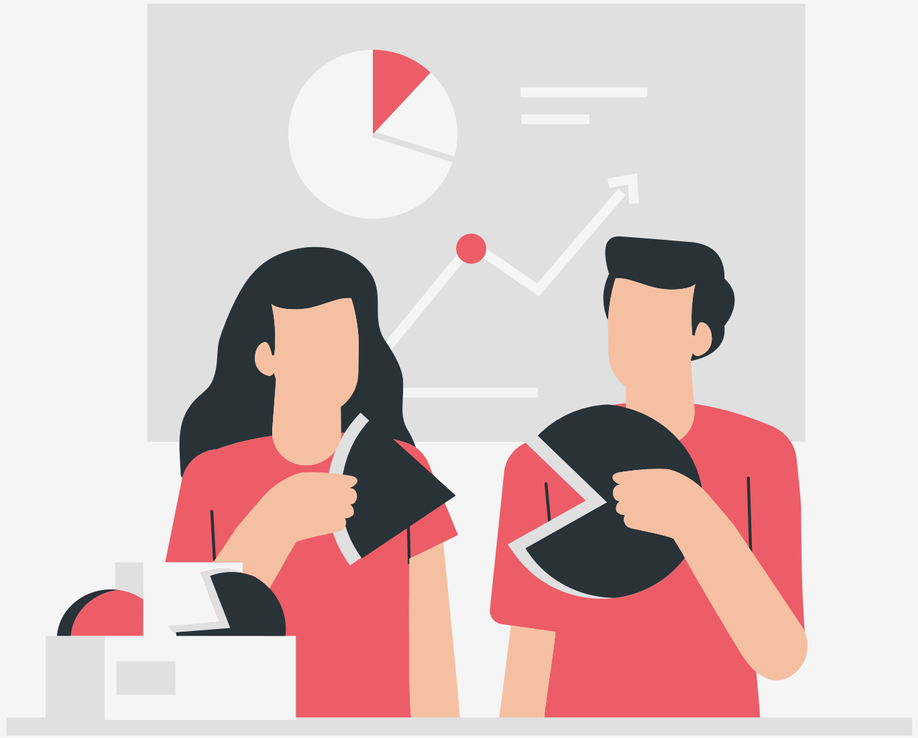
VANTAGE POINT
YOUR BUSINESS ACCELERATION TECHNOLOGY

ETL to ELT - Why the Change?



The 3 stages



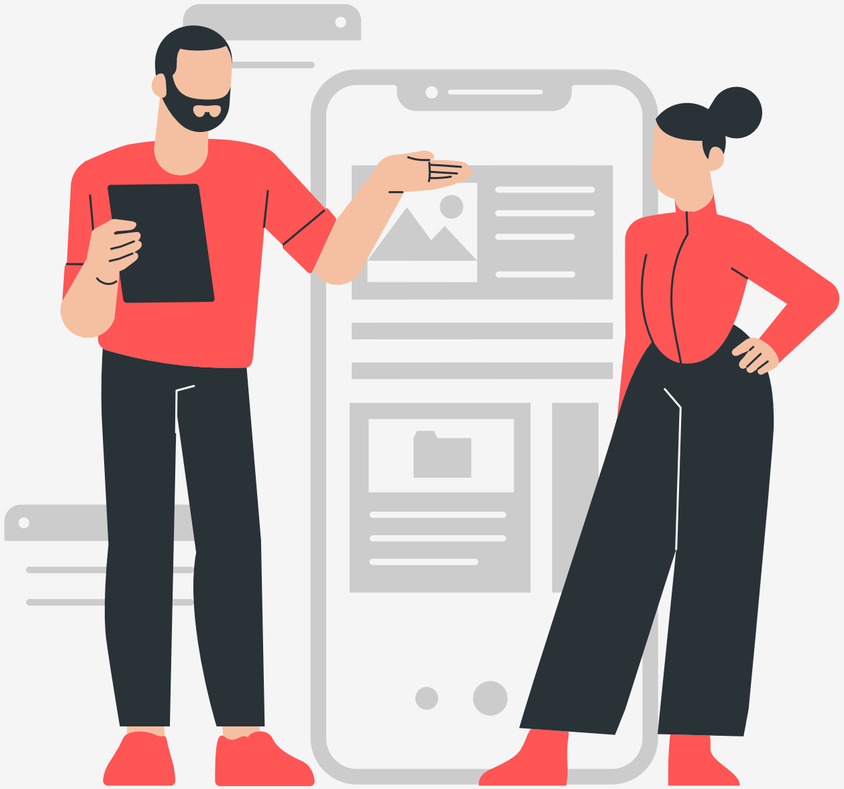
- Extract: Source data is extracted from the original data source in an unstructured format. In traditional ETL processes this data is put into a temporary staging repository.
- Transform: Additional transformation must be done to clean and model data before it can be practically useful for other analytics applications.
- Load: In the ELT model, data is copied then copied directly into the data warehouse without significant modification. In an ETL model, data would be transformed into a suitable format before loading it into the warehouse.

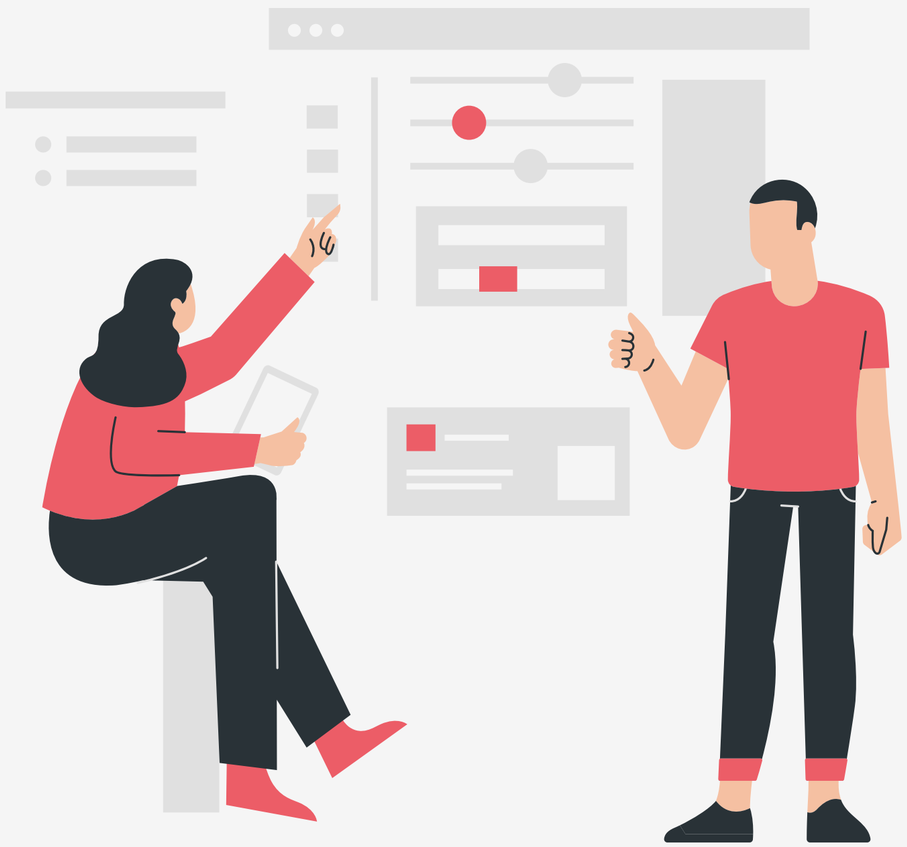
What is ETL?

ETL requires the transformations to happen before the loading process. Data is extracted from data sources and then deposited into a staging area. Data is then cleaned, enriched, transformed and finally loaded into the data warehouse.

Data teams have to predict all the uses cases for the data before any analysis is even performed and then create transformations accordingly.

Because Engineering teams typically own the extract and transform process, analysts have little visibility into the logic that has been used for the transformation process.





What is ELT?

ELT is a modern variation of ETL

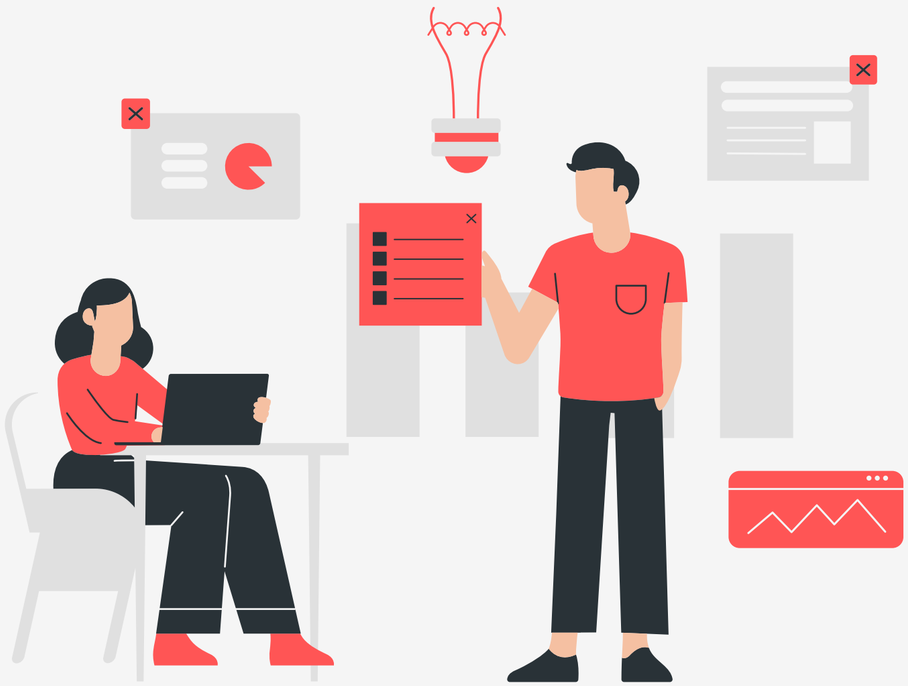
- Data cleaning, enrichment and transformation happen after the loading process.
- Modern cloud data warehouses are extremely scalable and separate storage from compute resources.

So essentially the main difference between ETL and ELT is the order that these steps take place.

With the transformation happening in the warehouse, it’s typically defined using SQL.

From ETL to ELT

- ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) are processes for moving data from one system to another
- The fundamental difference is how the raw data is managed, at which stage it is loaded into the warehouse and how analysis is then performed.
- In ETL data moves from the data source, to staging and then into the warehouse. Transformations are performed before the data is loaded
- ELT loads data into the warehouse before transforming it



Why is ELT the better choice?

- Agility: all the data is stored in the warehouse and readily available to use.
- Simplicity: Transformations in the data warehouse are generally written in SQL, a language that the entire data team (data engineers, data scientists, data analyst) understands.
- Self service analytics: If all of your raw data is within the warehouse, you can use BI tools to drill down from aggregated summary statistics to the raw data underlying them .
- Fixing bugs: If you find errors in your transformation pipeline, you can fix the bug and re-run just the transformations to fix your data.