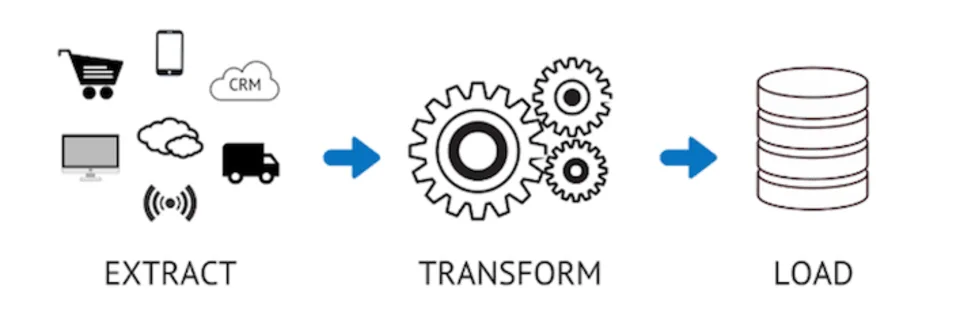
What is ETL?

ETL stands for Extract, Transform and Load which refers to a data integration process that combines data from multiple data sources into a single and compatible data store called a data warehouse. ETL process plays a vital role in producing business intelligence and executing broader data management strategies.

ETL provides the ground for data analytics and machine learning workstreams. Through a series of business rules, ETL cleanses and organize data through which we can solve business intelligence needs, like monthly reporting and predicting the outcome of business decision.



How does ETL works?

ETL process works in three steps:

1. Extract the relevant data from data sources.
2. Transform the data for data analysis and business intelligence.
3. Load the data into target data base.

Extract:

Data extraction is the process of obtaining raw data from a source and replicating that data somewhere else. The raw data can come from various sources, such as a database, Excel spreadsheet, a SaaS platform, web scraping, or others. It can then be replicated to a [destination](https://www.stitchdata.com/integrations/destinations/), such as a [data warehouse](https://www.stitchdata.com/resources/data-warehouse). This can also include unstructured data or simply a data which is poorly organized. Once the data has been proceeded now it waits for transformation.

Transform:

Data transformation is the process of converting data from one format to another, typically from the format of a source system into the required format of a destination system. This step includes filtering, cleansing, de-duplication, validating, normalizing and authenticating the data.

Load:

Data loading is the process to move the transformed data from the staging area into the target data warehouse. For most organizations that use ETL, the process is automated, well-defined, continuous and batch driven. Typically, ETL takes place during off-hours when traffic on the source systems and the data warehouse is at its lowest.

Why is ETL important?

1. Data Integration: ETL plays a crucial role in integrating data from multiple data sources into a centralized repository, such as data warehouse or data lake. This enables organization to have a overall view of the data and make informed decisions based on accurate and consistent information.
2. Data Quality: ETL processes help to ensure data quality by transforming the data as it is extracted from source systems. This improves the accuracy and consistency of the data, which is critical for decision making and analysis.
3. Performance: ETL processes can help to improve the performance of data processing and analysis by aggregating and summarizing data, reducing data duplication, and optimizing the data model for faster querying and reporting.

By applying the process of extract, transform, and load (ETL), individual raw datasets can be prepared in a format and structure that is more consumable for analytics purposes, resulting in more meaningful insights. For example, online retailers can analyze data from points of sale to forecast demand and manage inventory. Marketing teams can integrate CRM data with customer feedback on social media to study consumer behavior.