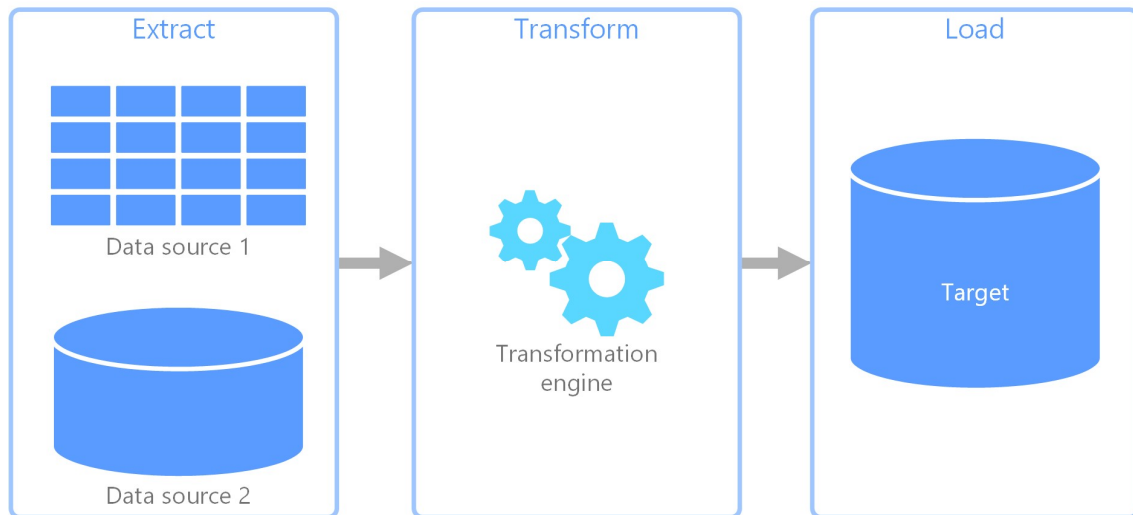


ETL Architecture



Extract, Transform, Load: Extract, transform, and load (ETL) is a data pipeline used to collect data from various sources. It then transforms the data according to business rules, and it loads the data into a destination data store. The transformation work in ETL takes place in a specialized engine, and it often involves using staging tables to temporarily hold data as it is being transformed and ultimately loaded to its destination.

The data transformation that takes place usually involves various operations, such as filtering, sorting, aggregating, joining data, cleaning data, deduplicating, and validating data.

Often, the three ETL phases are run in parallel to save time. For example, while data is being extracted, a transformation process could be working on data already received and prepare it for loading, and a loading process can begin working on the prepared data, rather than waiting for the entire extraction process to complete.