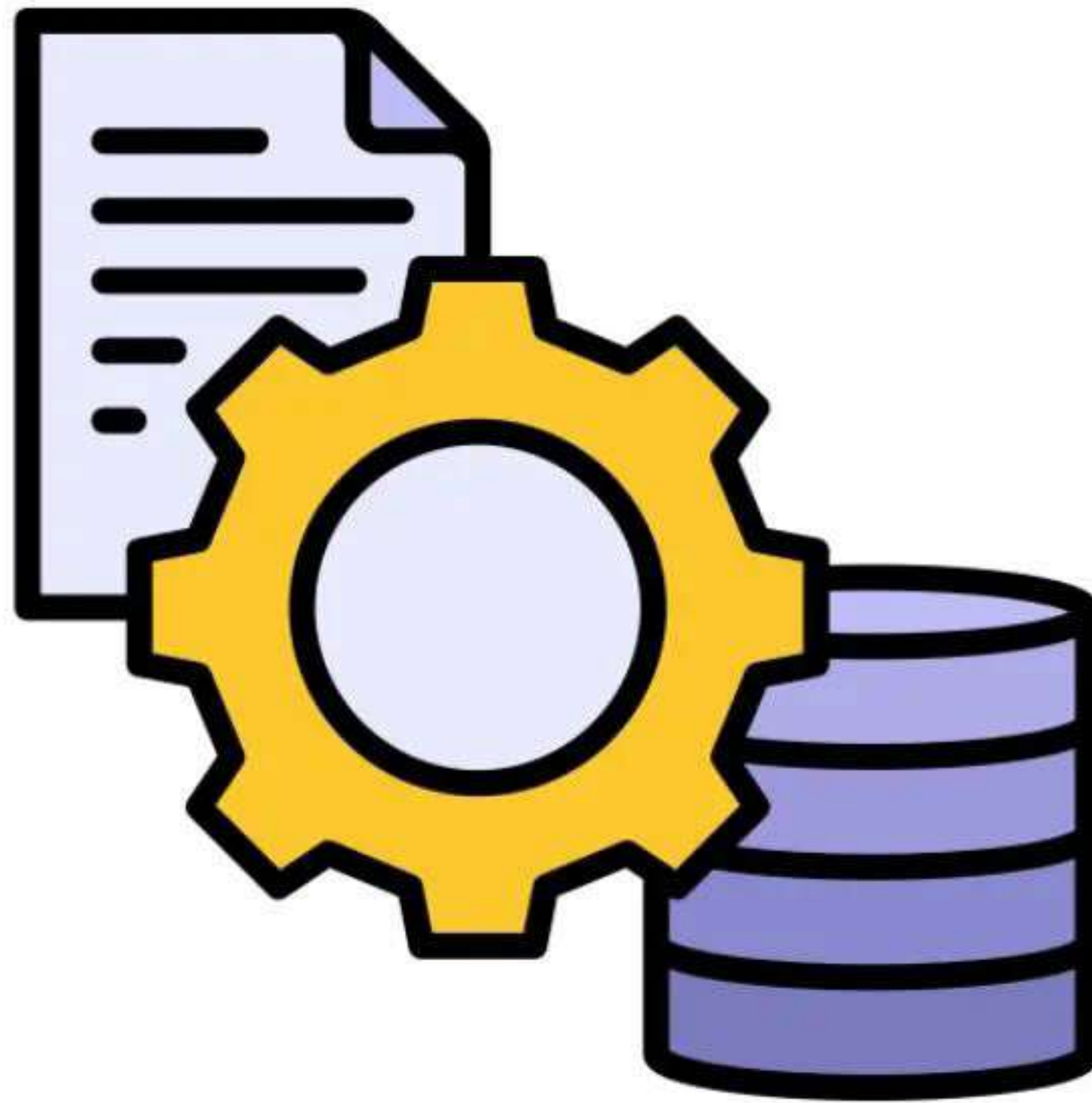
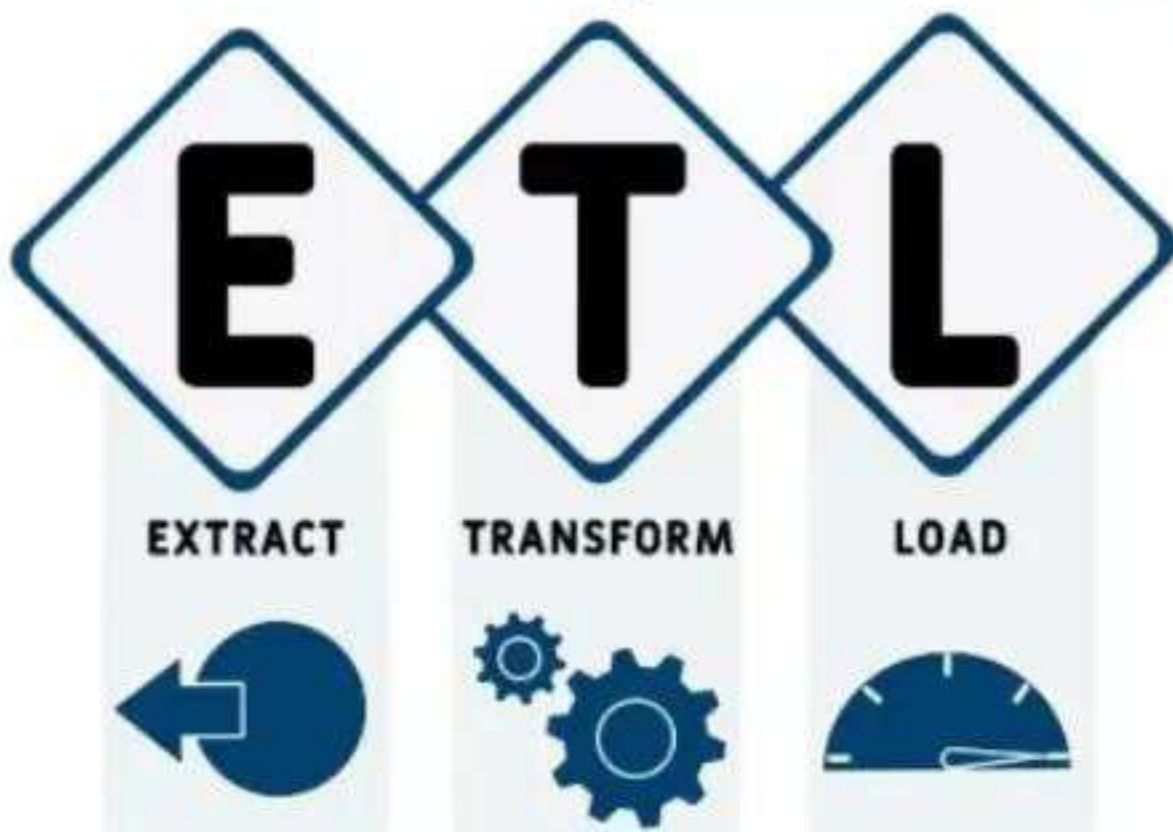


WHAT IS ETL?



Explained in fun and easy way



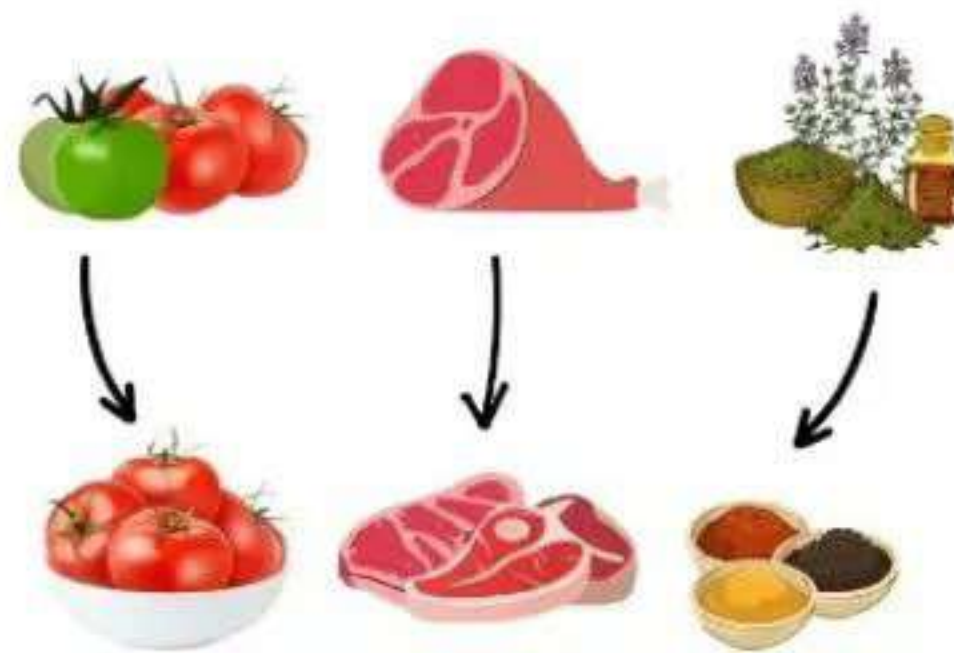
Swipe →

If you are working with data, you probably have heard the term "ETL" thrown around in relation to data, data warehousing, and analytics.



ETL is a data processing method used to gather (Extract) information from various sources, modify it (Transform) into a consistent format, and Load it into a destination where it can be easily accessed and analyzed.

Let's see that on example. Imagine you're a master chef preparing a scrumptious feast. First, you gather the finest ingredients (Extract) from various suppliers, carefully selecting the ripest tomatoes, juiciest meats, and freshest herbs.



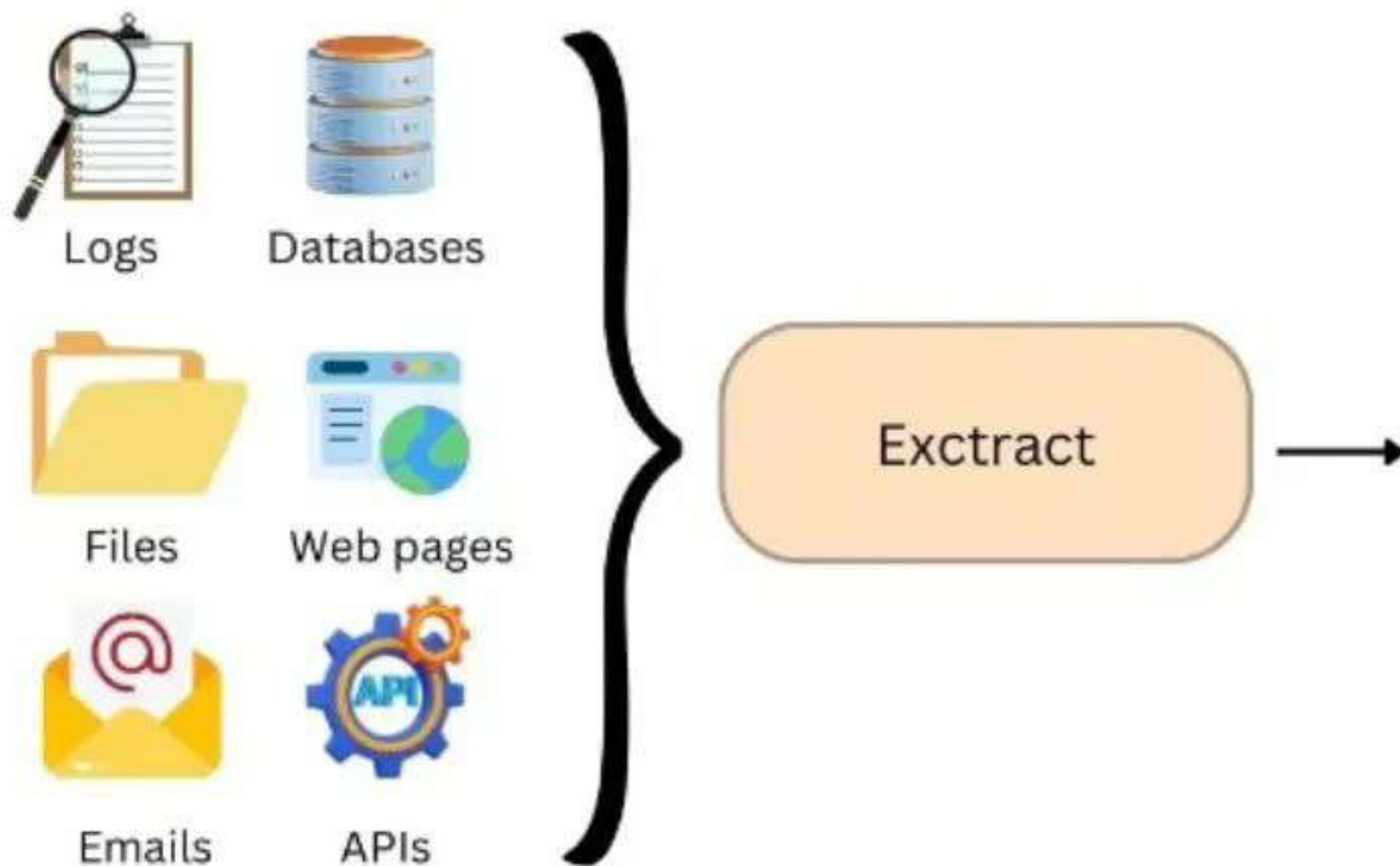
Then, in your culinary laboratory, you work your magic (Transform) - slicing, dicing, marinating, and blending the elements together, turning them into delectable dishes with unique flavors.

Finally, you serve your culinary creations to eager guests (Load) who savor the harmonious fusion of tastes.



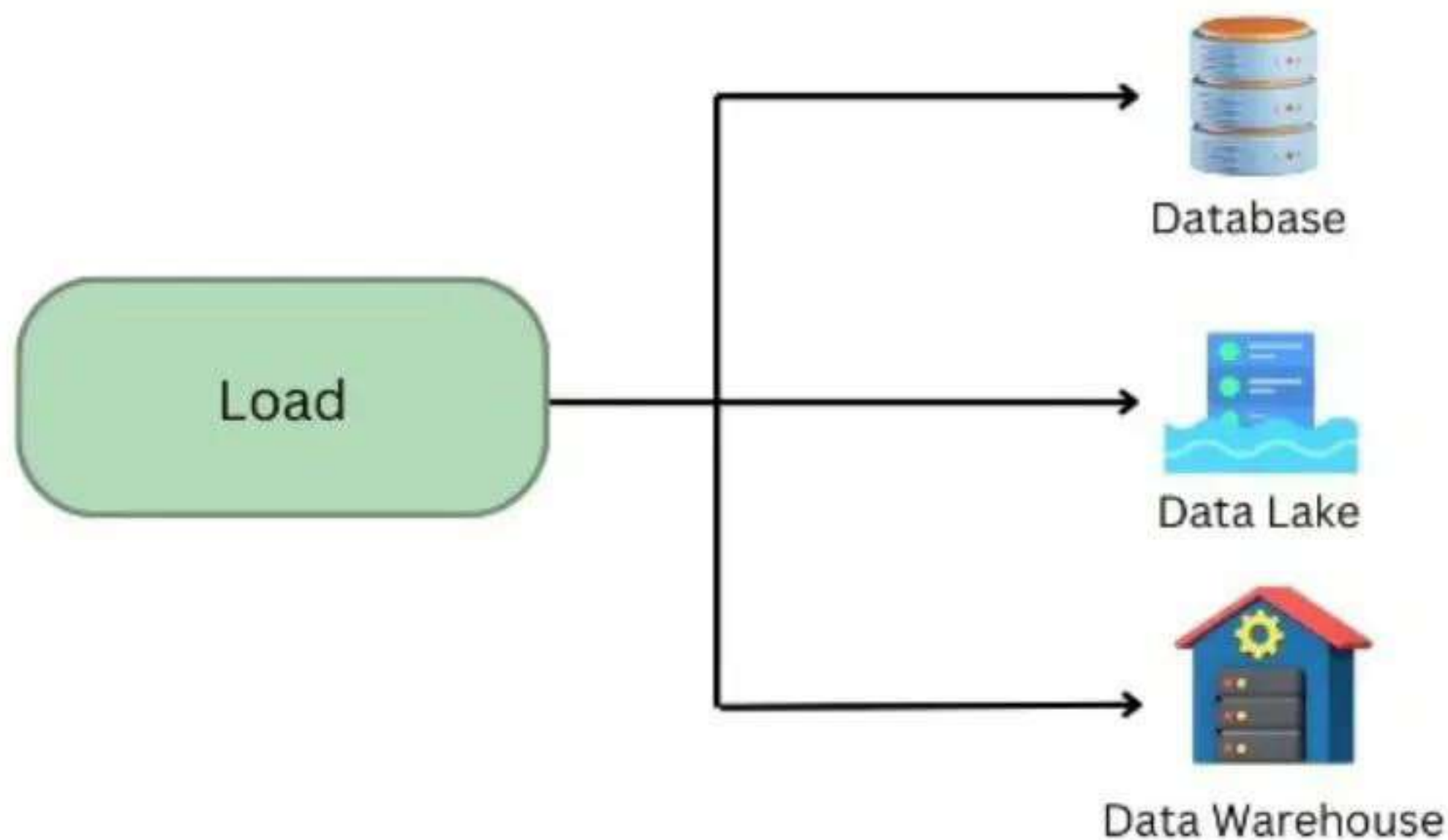
In the data world, ETL does something similar, taking raw data, transforming it into useful insights, and delivering it to hungry data analysts, just like you serve your gourmet dishes to delighted diners.

Extract: In this initial stage, data is extracted from various heterogeneous sources like databases, files, APIs, or web services.



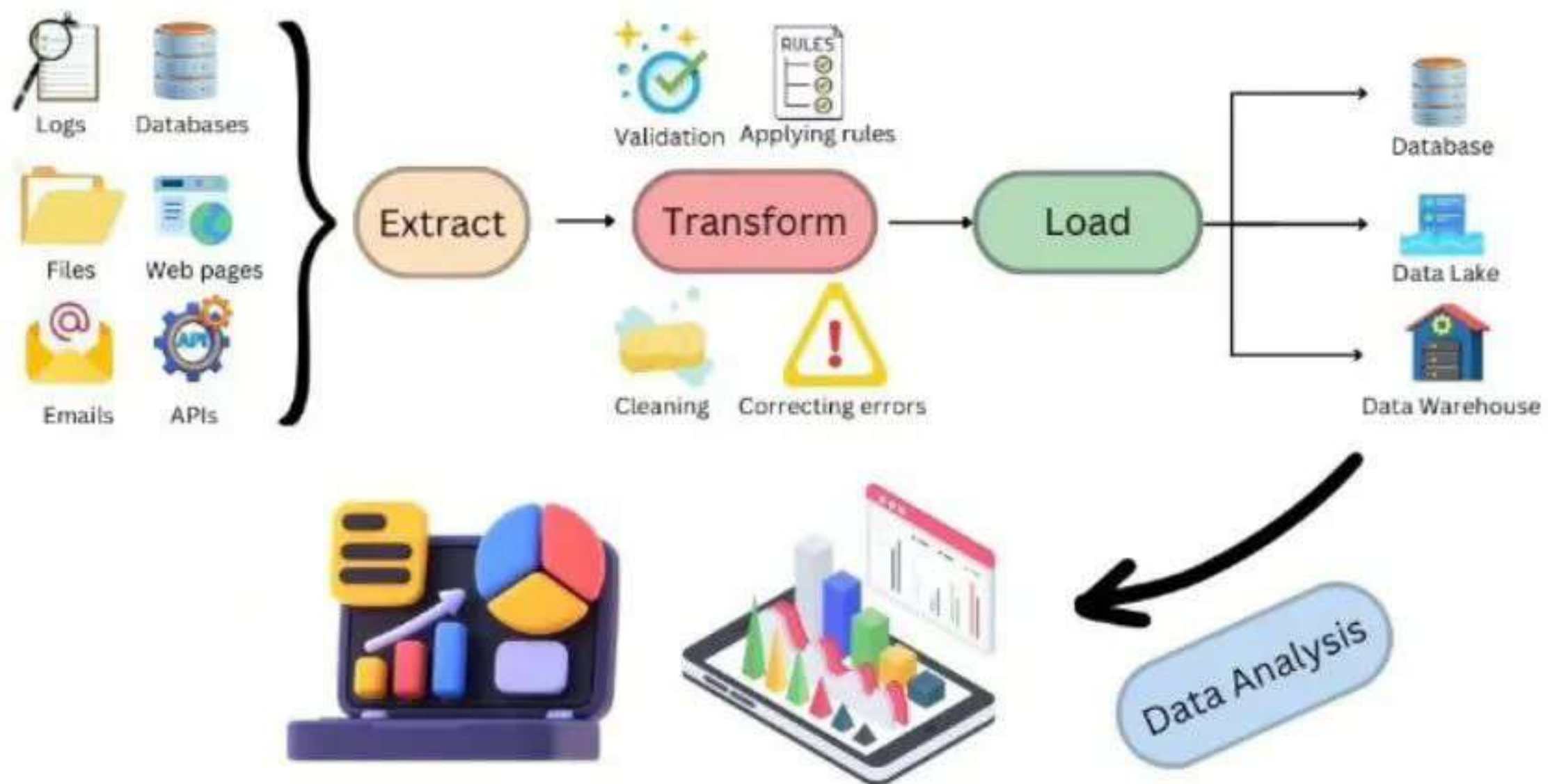
The goal is to gather all relevant data needed for analysis, reporting, or storage in a data warehouse.

Load: After the data is transformed and optimized for analysis, it's loaded into the target data repository, such as a data warehouse or a data mart.

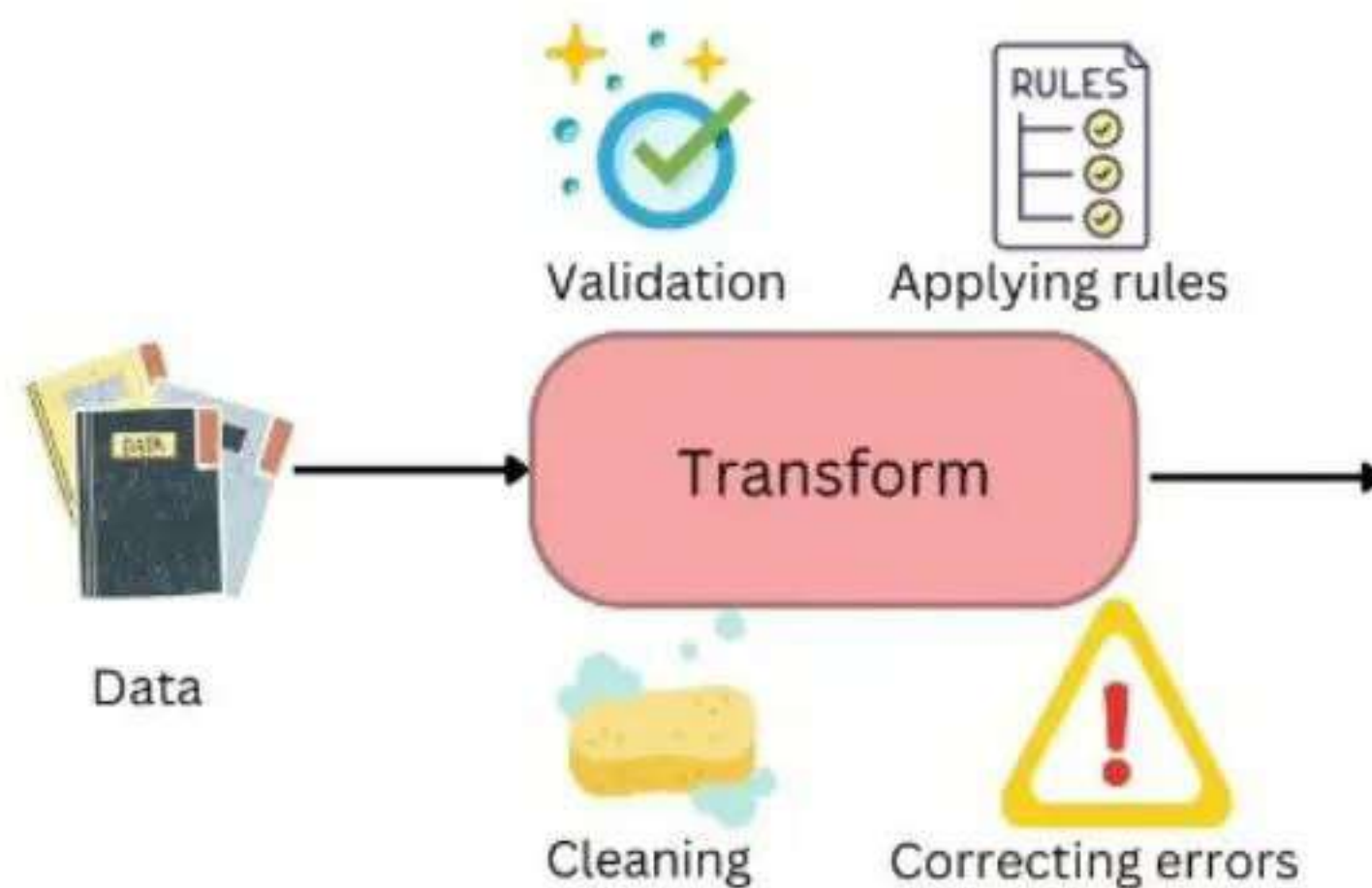


The loading process involves storing the data in a structured manner that facilitates efficient querying and reporting.

By performing these three steps, ETL processes help organizations access high-quality data that is consistent, actionable, and relevant, thus enabling data-driven insights and strategic decision-making.



Transform: Once the data is extracted, it often requires cleansing, validation, and manipulation to ensure its quality and consistency.



This transformation process involves converting data into a standardized format, correcting errors, handling missing values and applying business rules.. It may also involve aggregating or disaggregating data as required.

Traditional ETL (Extract, Transform, Load): In this approach, data is first extracted from the source systems, then transformed in a separate ETL tool or engine, and finally loaded into the target data repository, like a data warehouse or data mart. This method is suitable for batch processing and works well when dealing with large volumes of data that require complex transformations before being stored for analysis.

ETLT (Extract, Transform, Load, Transform): ETLT is a hybrid approach that includes two transformation steps. Data is initially extracted from the source systems, transformed in an intermediary staging area or a temporary data store, then further transformed before being loaded into the final target data repository. This approach can be helpful in scenarios where additional complex transformations are required after the initial load but before storing data for analysis.

ELT (Extract, Load, Transform): ELT is a modern variation of ETL where the extraction and loading steps remain the same, but the transformation step is shifted to the target data repository. In other words, the data is extracted from source systems, directly loaded into the target system, and then the transformation is applied using the processing power of the target system. ELT is often associated with big data platforms like Hadoop, where the storage system also serves as a compute engine, enabling faster processing of large-scale data.

Each type of ETL has its strengths and weaknesses, and the choice of which one to use depends on factors like data volume, complexity of transformations, target systems, and performance requirements. As technology evolves, new variations and approaches to ETL continue to emerge to better address the diverse data integration needs of modern businesses.