

Introduction to ETL

Extract

Data may come from disparate sources, such as:

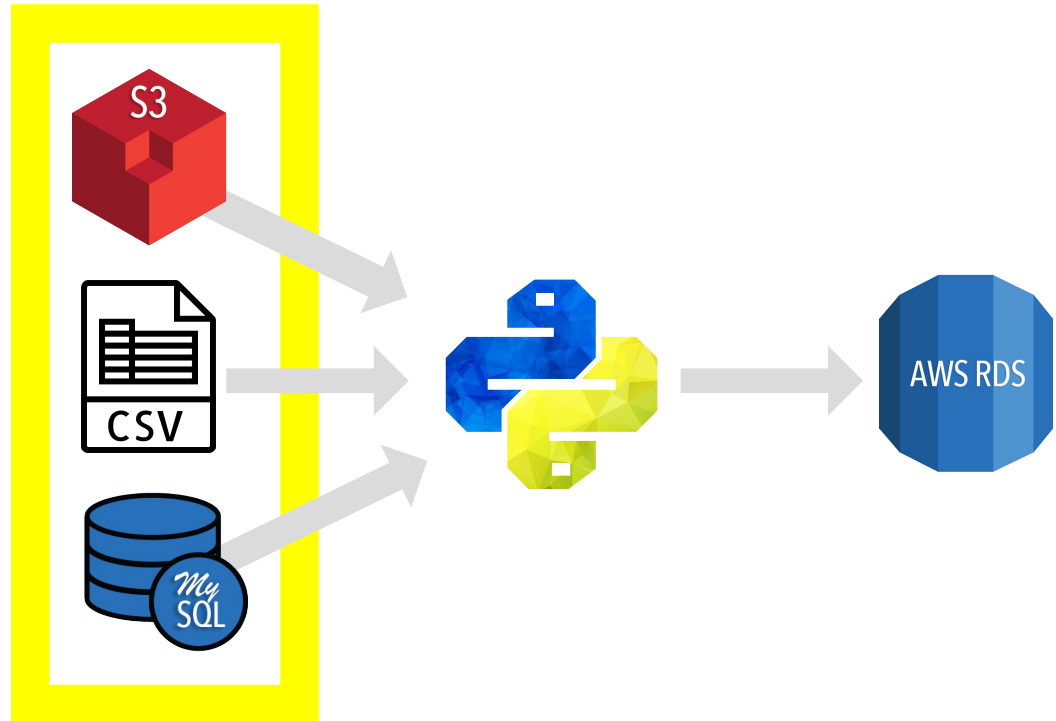
CSV files

JSON files

HTML tables

SQL databases

Spreadsheets



Extract

Introduction to ETL

Transform

Transform the data to suit business needs.
This may include:

Data Cleaning

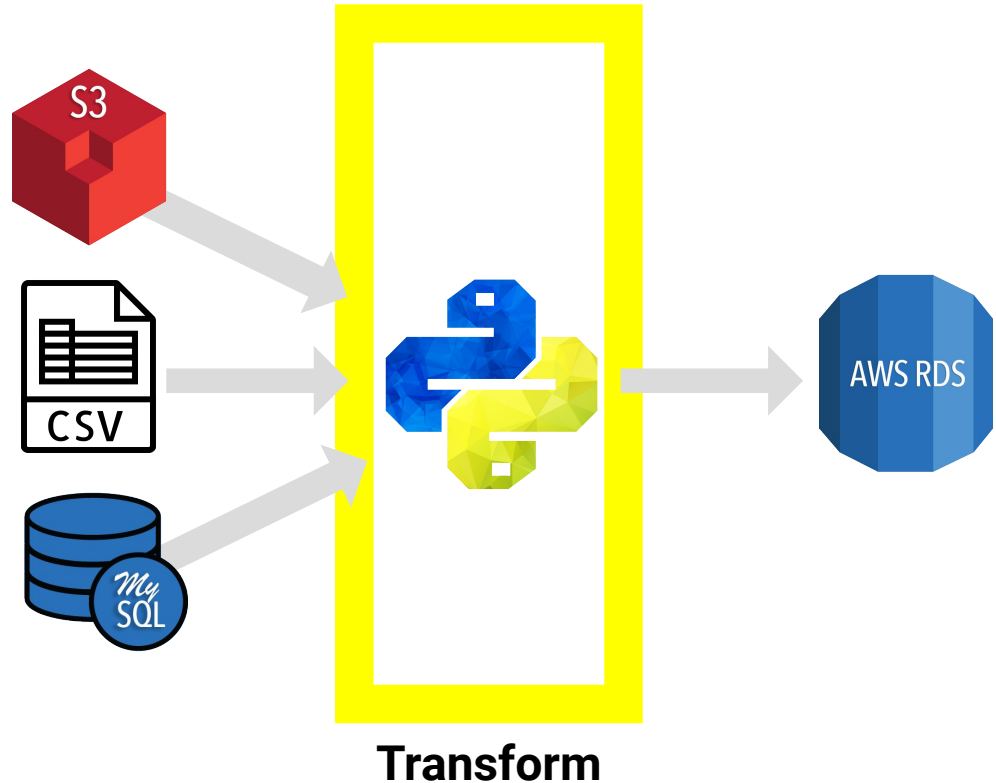
Summarization

Selection

Joining

Filtering

Aggregating



Introduction to ETL

Load

Load the data into a final database that can be used for future analysis or business use.

Can be a relational or non-relational database

Can be local or in the cloud

Can be a data lake or data warehouse

