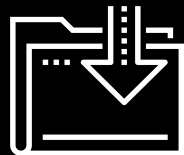




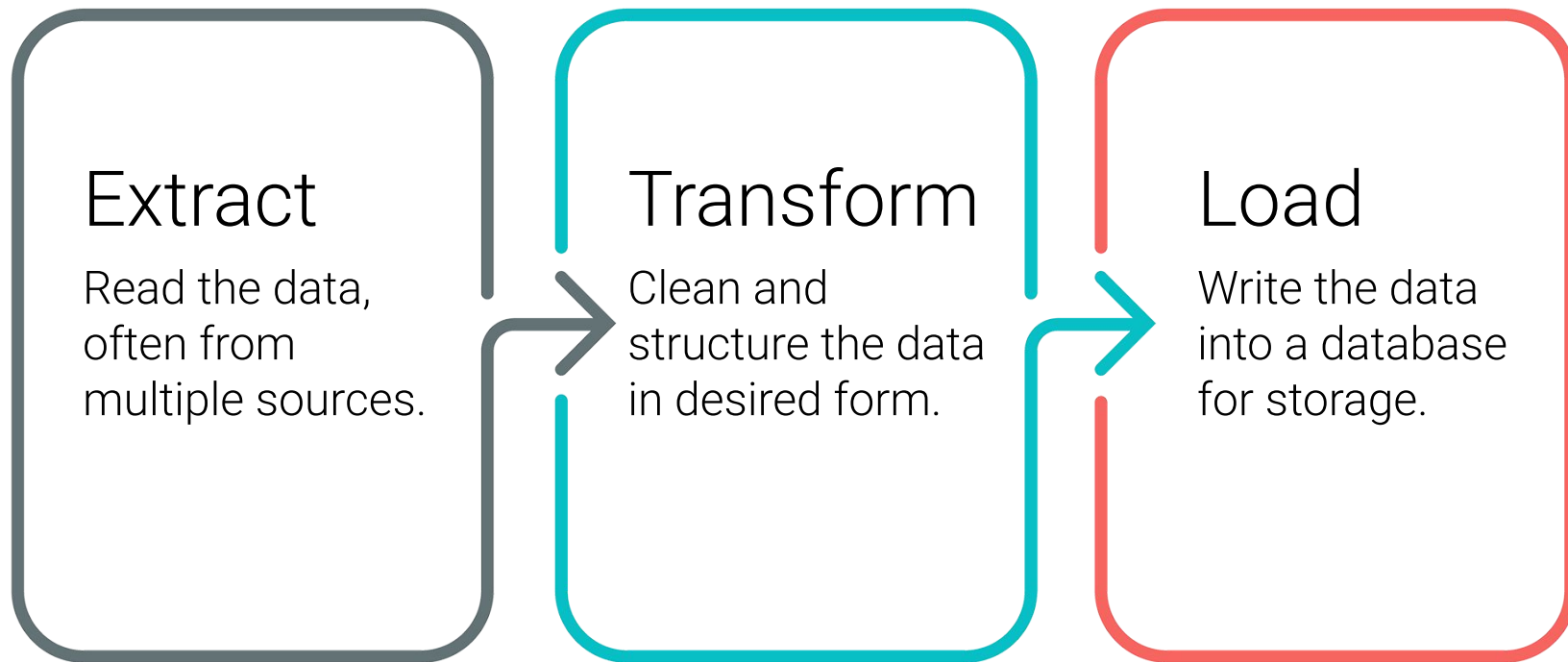
ETL: Extract, Transform, Load

Data Boot Camp
Lesson 13.1



ETL

Data integration is an important part of working with data.



Extract

Data may come from disparate sources, such as:

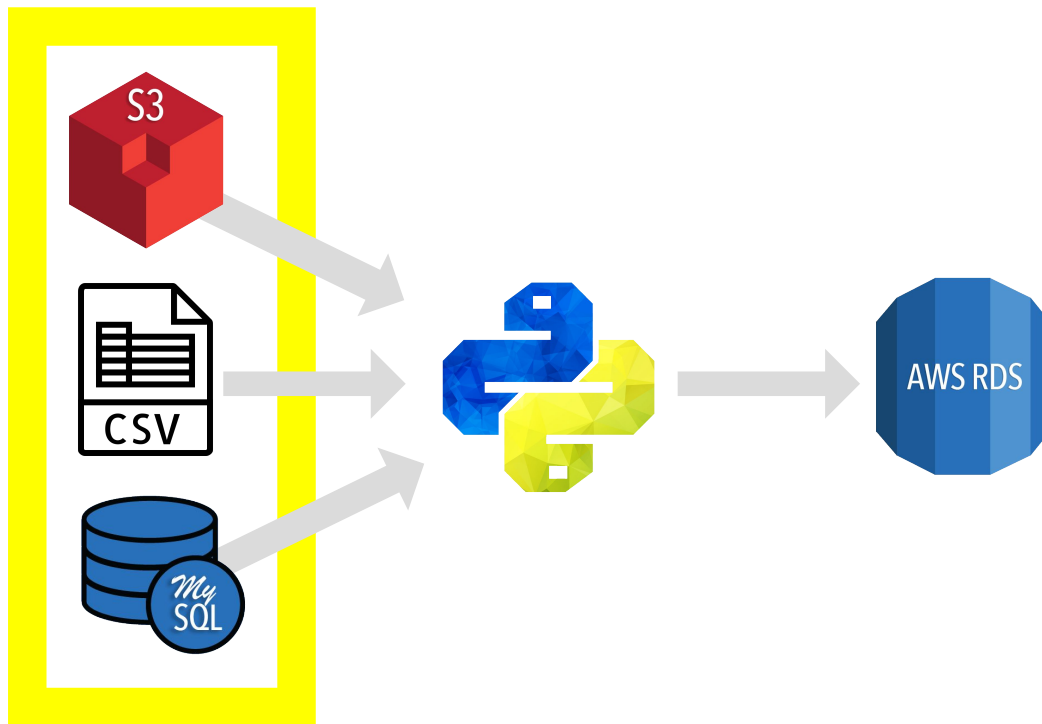
CSV files

JSON files

HTML tables

SQL databases

Spreadsheets



Extract

Transform

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

Data Cleaning

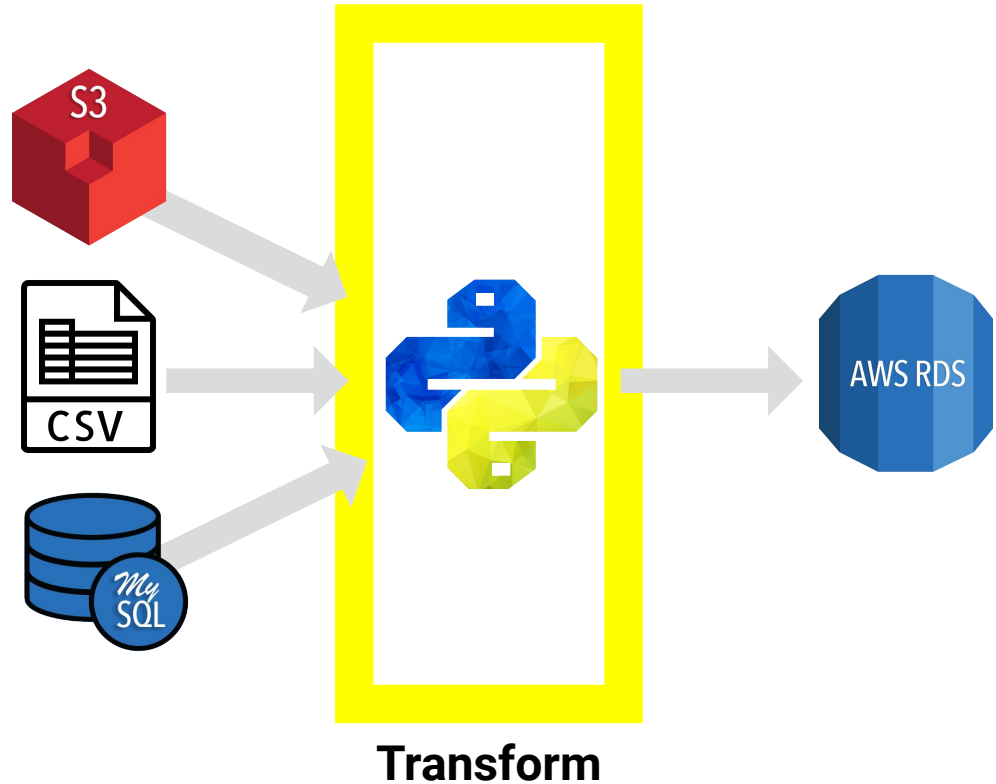
Summarization

Selection

Joining

Filtering

Aggregating





Note: We will use Python and pandas for transformation, which can also be done with SQL or a specialized ETL tool.

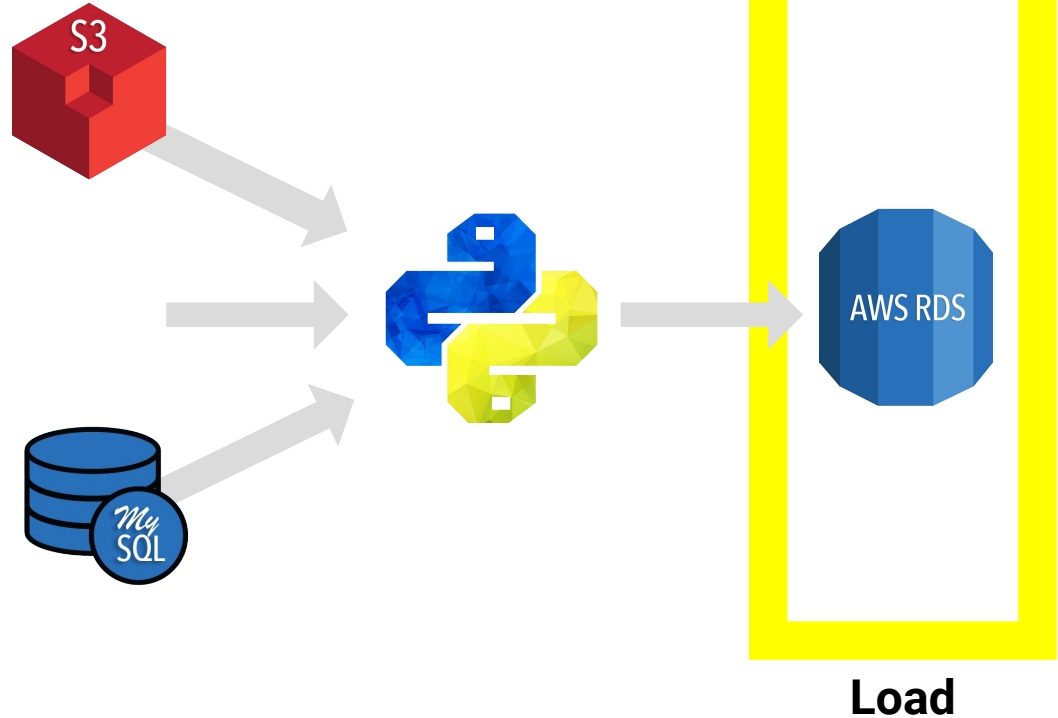
Load

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





Questions?