SCHOOL OF DATA
by Springboard

Data Engineering Career Track
Guided Capstone Step One: Design and Setup

Learning objective

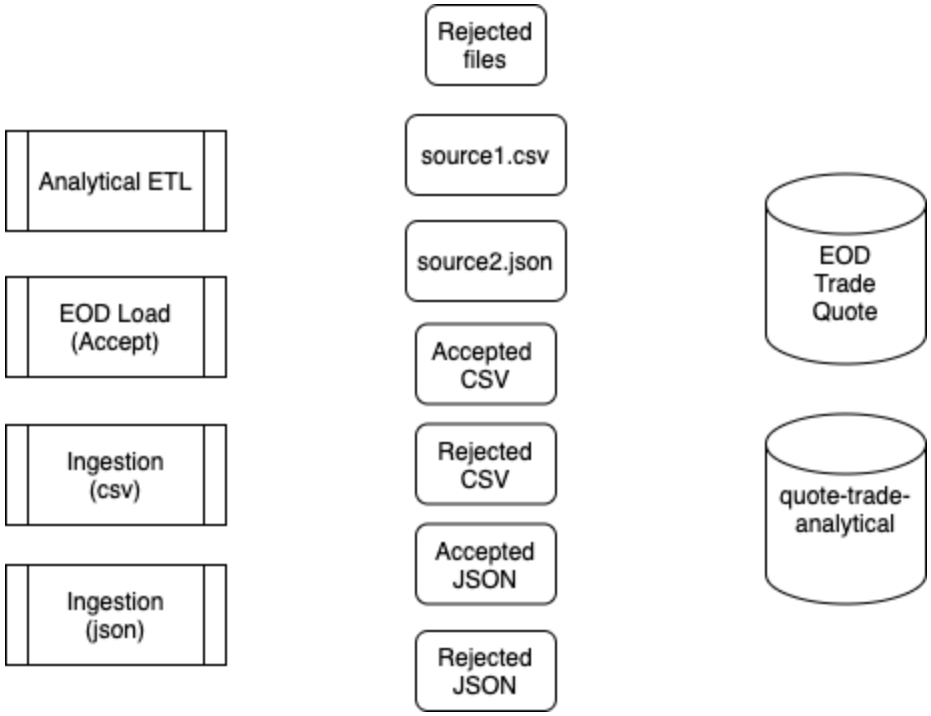
Every time you develop a data pipeline, it is essential to have a high-level architecture and data flow before starting your development work. Usually, developers draw flow diagrams as part of the thought process from beginning to the end of the pipeline. Flow Diagrams are used as a key guideline for implementing every single component later on. Such a skill is critical for a developer who wants to build a system independently.

1. Design the architecture diagram

You must come up with the entire logical data flow to use as a guideline before you begin doing actual development work. One way to do this is to create a sample architecture flow connecting major components of the system.
Your diagram should follow some commonly used rules.
When you design the diagram, ask yourself these questions:

- How would you break down the end-to-end flow into smaller components?
- What are the dependencies between each component?
- What are the input and output datasets for each component?
- What are the storage formats for each dataset - flat files, database tables, etc.?

In this project, we are breaking up the end-to-end pipeline into several components. As a guided step to draw the diagram, the components and their input/output datasets are below. Based on your understanding, you need to connect them using arrows to complete the flow diagram.



You can use Draw.io to draw a diagram. You'll find a short tutorial below.

https://www.youtube.com/watch?v=rd5Icj1i_ng&ab_channel=CarlArrowsmith

Once you complete the diagram, please upload it and share with your mentor to review.

2. Setup your IDE

In this project, we will use Python as the coding language. There are a plethora of IDEs for Python. Here are some popular ones. Please note, some of these IDEs will require you to install packages for them to support Python development.

- Jupyter Notebook
- PyCharm
- Spyder
- Sublime
- Intellij
- VS Code

3. Create a GitHub repository for this project and push your diagram to it.