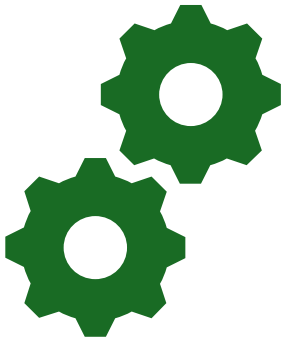
Description of the steps needed to design your pipeline

Final Project

Design your pipeline



Technical outline

Setups in Snowflake

STEP 1: Create your raw data

- Python
  - Use the generator to generate data : customers, orders, product
  - (Optional upload the data to an SQL database)
  - Extract : Connect to the SQL database to retrieve your data
  - Connect to Snowflake
  - Upload the data to Snowflake

- Snowflake
  - Create a warehouse
  - Create a Database
  - Create three schema : prod, staging, raw
  - Create your destination table

```
--- CREATE WAREHOUSE FOR THE CLASS ---
CREATE WAREHOUSE IF NOT EXISTS TEACH_WH
  WITH WAREHOUSE_SIZE = 'XSMALL'
  AUTO_SUSPEND = 60                 -- pause after 1 min of inactivity
  AUTO_RESUME = TRUE;

-- Create you database on snowflake
CREATE DATABASE IF NOT EXISTS RETAIL_LAB;

--Create you schema
CREATE SCHEMA IF NOT EXISTS RETAIL_LAB.STG;
CREATE SCHEMA IF NOT EXISTS RETAIL_LAB.DWH;
CREATE SCHEMA IF NOT EXISTS RETAIL_LAB.RAW;
```

Technical STEP 2 : Create your event

- Python
  - Setup your docker with redpanda, your producer and your consumer
  - Write the code to produce event messages
  - Write the code to consume messages and send them to snowflake
- Snowflake
  - Create your event table

PROJECT STEP 3 : Automate the update of your table

- Python
  - Send new events to snowflake
- Snowflake
  - Create a stream on your events table
  - Create and schedule your update task

PROJECT STEP 4 : Step Monitoring

- Snowflake
  - Select what entity you want to monitor
  - Create your analytics tables
  - Call those tables in an App Streamlit

What is expected

- A full repo with all your project scripts.
- A schema to outline the steps of your process.
- Your snowflake SQL scripts that you can share via a link : the automation scripts with stream and task and validation scripts to check if your data was well processed.
- Screenshots of your snowflake tables to validate that you ingested the data.
- A video of the data streaming with kafka (events created, and their ingestion in Snowflake).
- **Or** a video of automation process in snowflake.
- A screenshot/a video of your monitoring APP.