Problem Statement:

Imagine you are part of a data team that wants to bring in daily data for COVID-19 test occurring in New York state for analysis. Your team has to design a daily workflow that would run at 9:00 AM and ingest the data into the system.

API:
https://health.data.ny.gov/api/views/xdss-u53e/rows.json?accessType=DOWNLOAD

By following the ETL process, extract the data for each county in New York state from the above API, and load them into individual tables in the database. Each county table should contain following columns :

- ❖ Test Date
- ❖ New Positives
- ❖ Cumulative Number of Positives
- ❖ Total Number of Tests Performed
- ❖ Cumulative Number of Tests Performed
- ❖ Load date

Implementation options:

1. Python scripts to run a daily cron job
   a. Utilize SQLite in memory database for data storage
   b. You should have one main standalone script for a daily cron job that orchestrates all other remaining ETL processes
   c. Multi-threaded approach to fetch and load data for multiple counties concurrently
2. Airflow to create a daily scheduled dag
   a. Utilize docker to run the Airflow and Postgres database locally
   b. There should be one dag containing all tasks needed to perform the end to end ETL process
   c. Dynamic concurrent task creation and execution in Airflow for each county based on number of counties available in the response

Implement unit and/or integration tests for your application

Solution should be implemented using Object Oriented design and Python3, with Readme describing the steps on how to run the application

As you have noticed by now, this is not a typical take home exercise. It’s pretty open ended with no single right or wrong solution. Many requirements are unknown or ambiguous. IRL, while working in a team, you would be asking questions to your team, architects, and business Stakeholders for clarity. For this exercise, make assumptions to the best of your understanding. Also, you may not be able to complete all the requirements, and that’s okay. We are hiring for many levels: Associate to Senior level Engineers. So, give your best shot with the time you have. Give your imagination free rein. All the best and we are excited to discuss your implementation in the interview.

Note: Leveraging any Cloud Services is out of scope for building this solution.