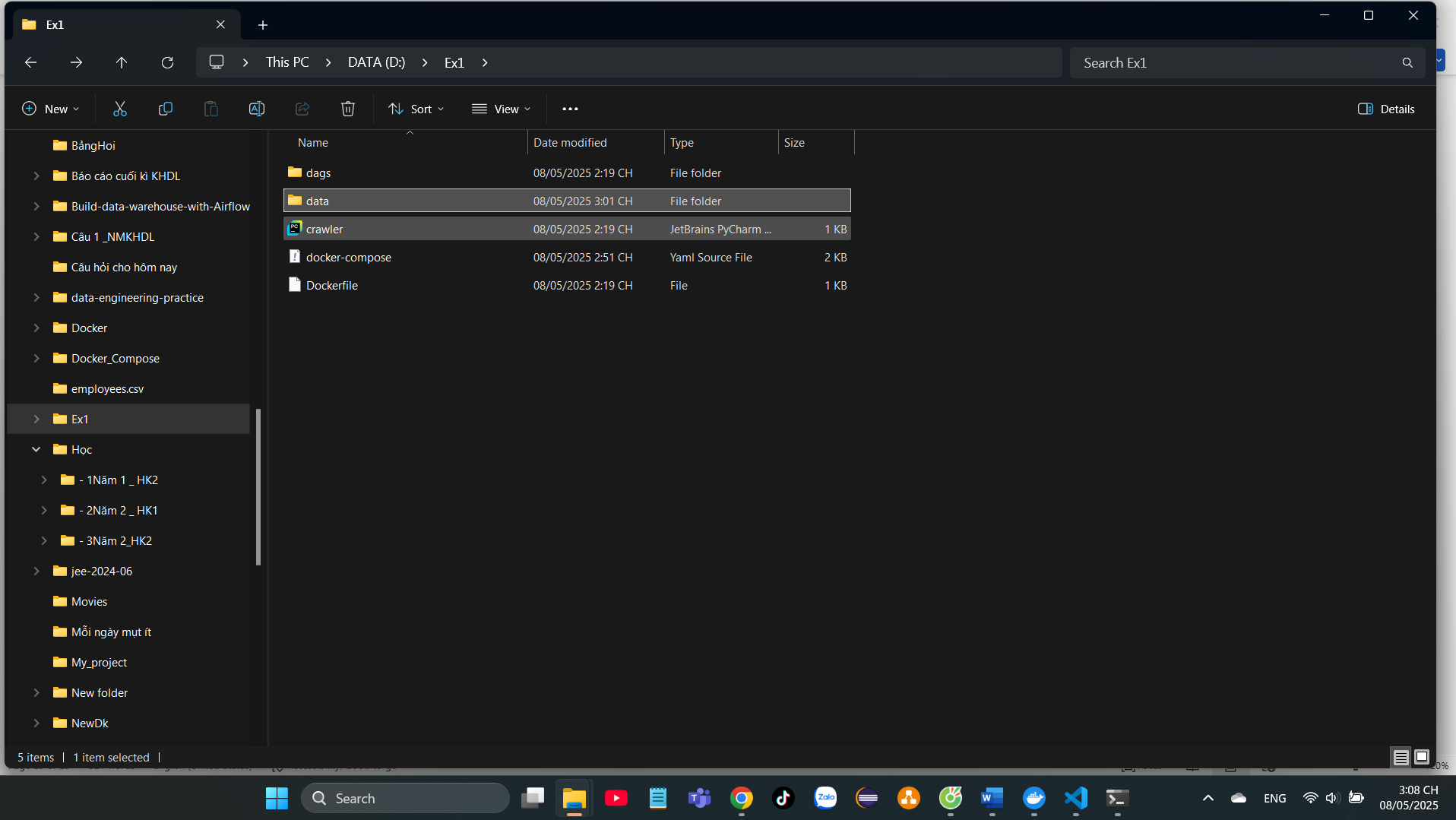
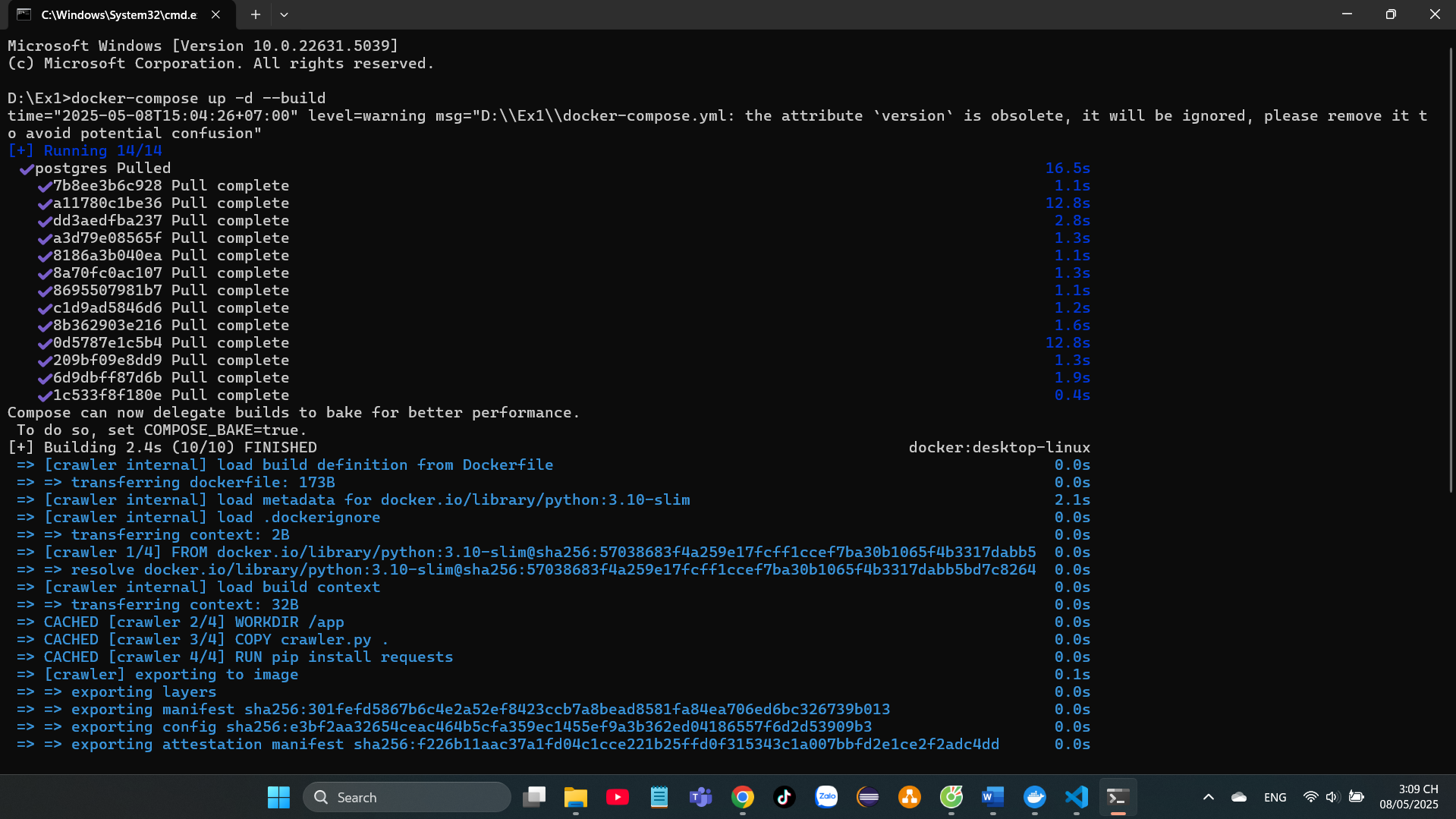
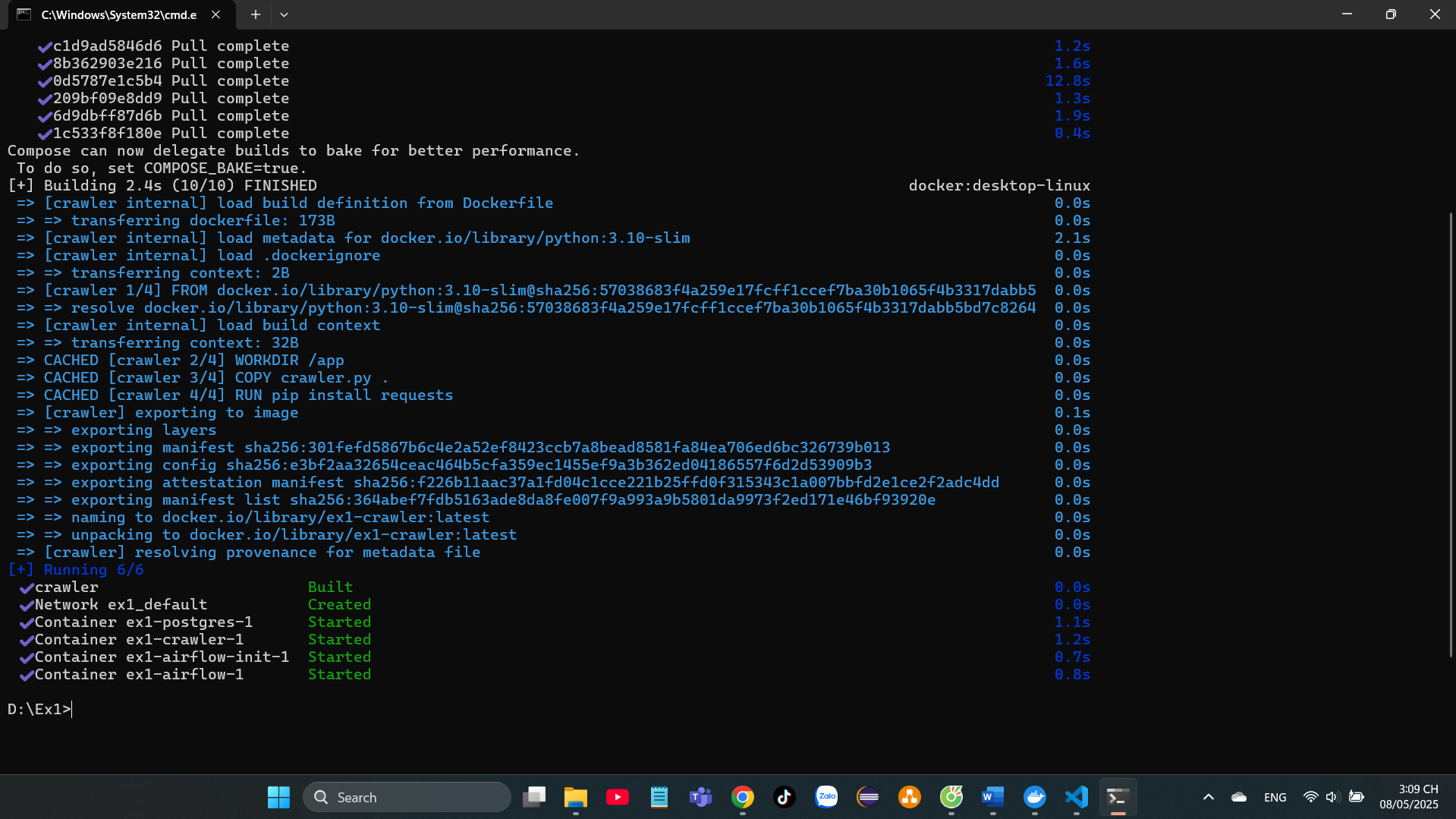
Ta xây dựng docker-compose và dags để thực hiện quá trình pipeline tự động cào và trực quan dữ liệu.

**Cấu trúc của lab 8.1** :

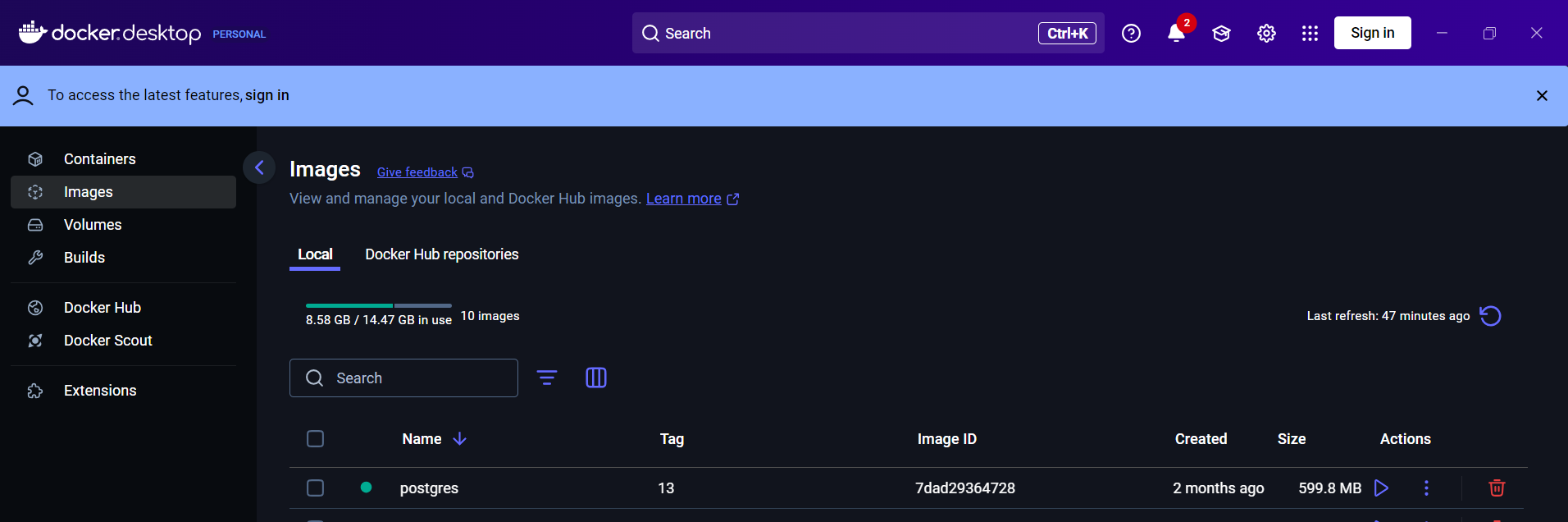


Thực hiện lệnh docker-compose up -d –build để bắt đầu build :



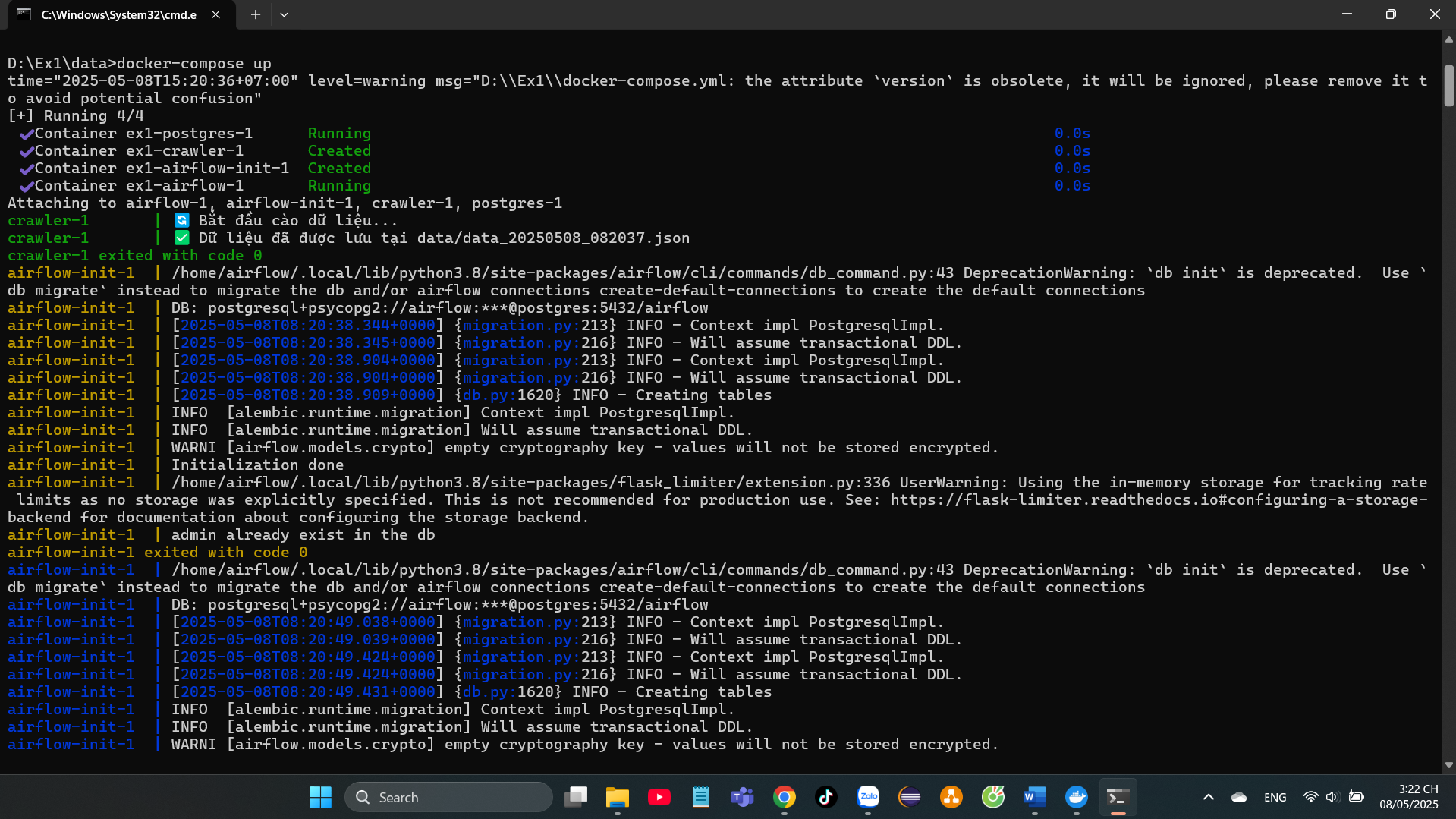


Kết quả của câu lệnh trên:

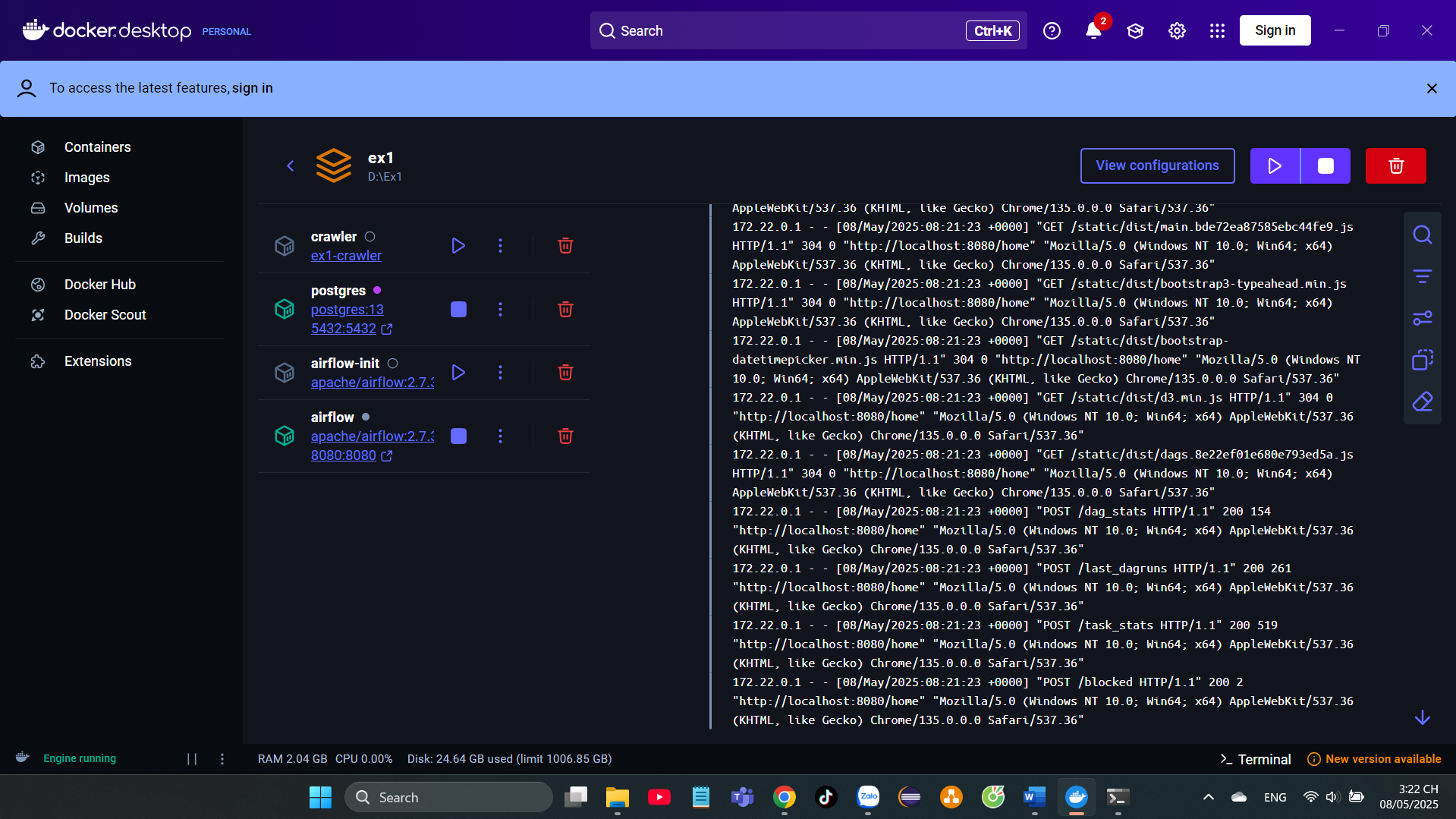


Sau khi đã build xog chúng ta thực hiện câu lệnh docker-compose up:

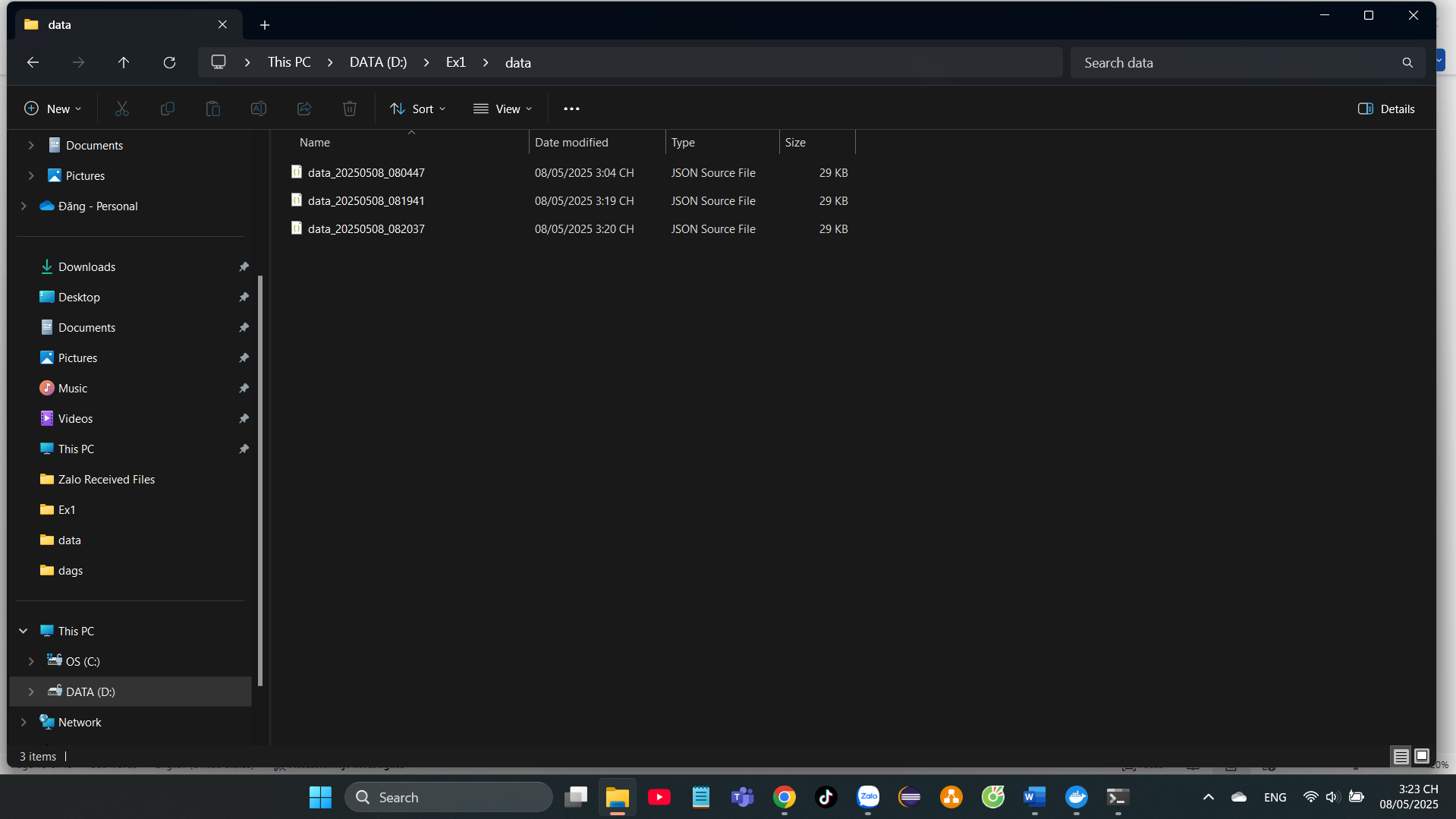
Lúc này sẽ hiện ra giao diện



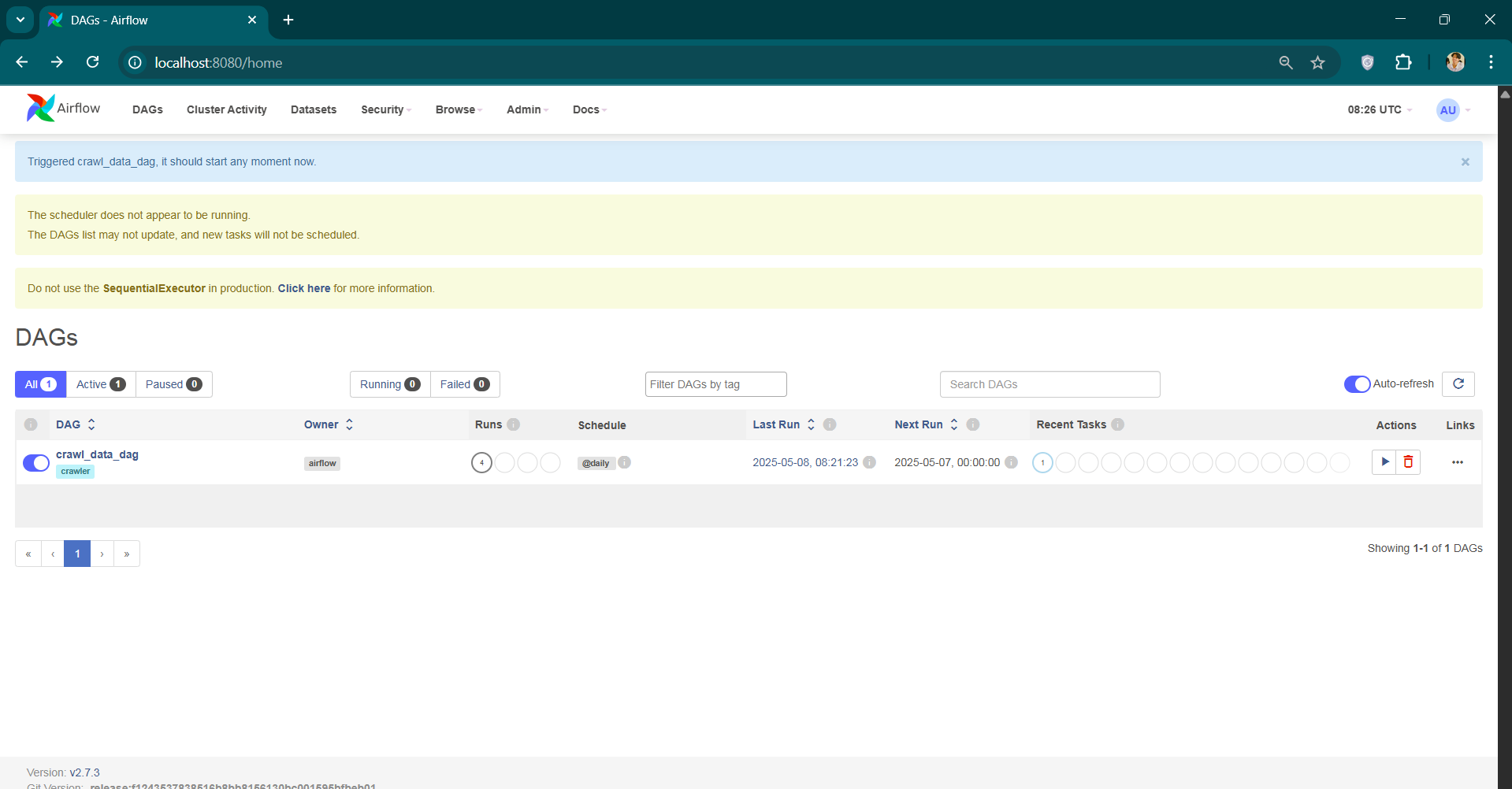
Chọn v để xem bên trong :



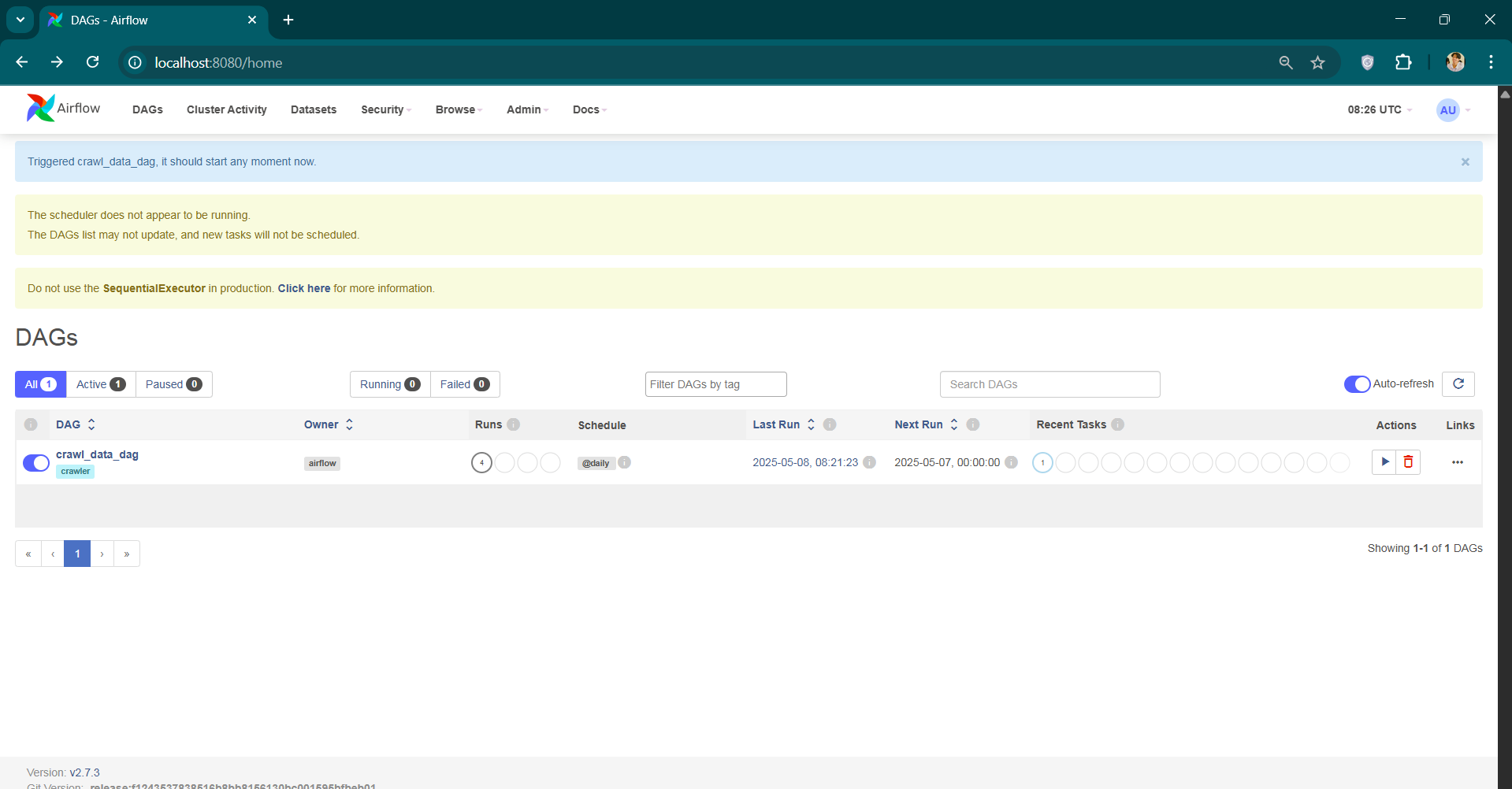
Sau khi ấn chạy airflow-init thì data sẽ đc cào vào folder data :

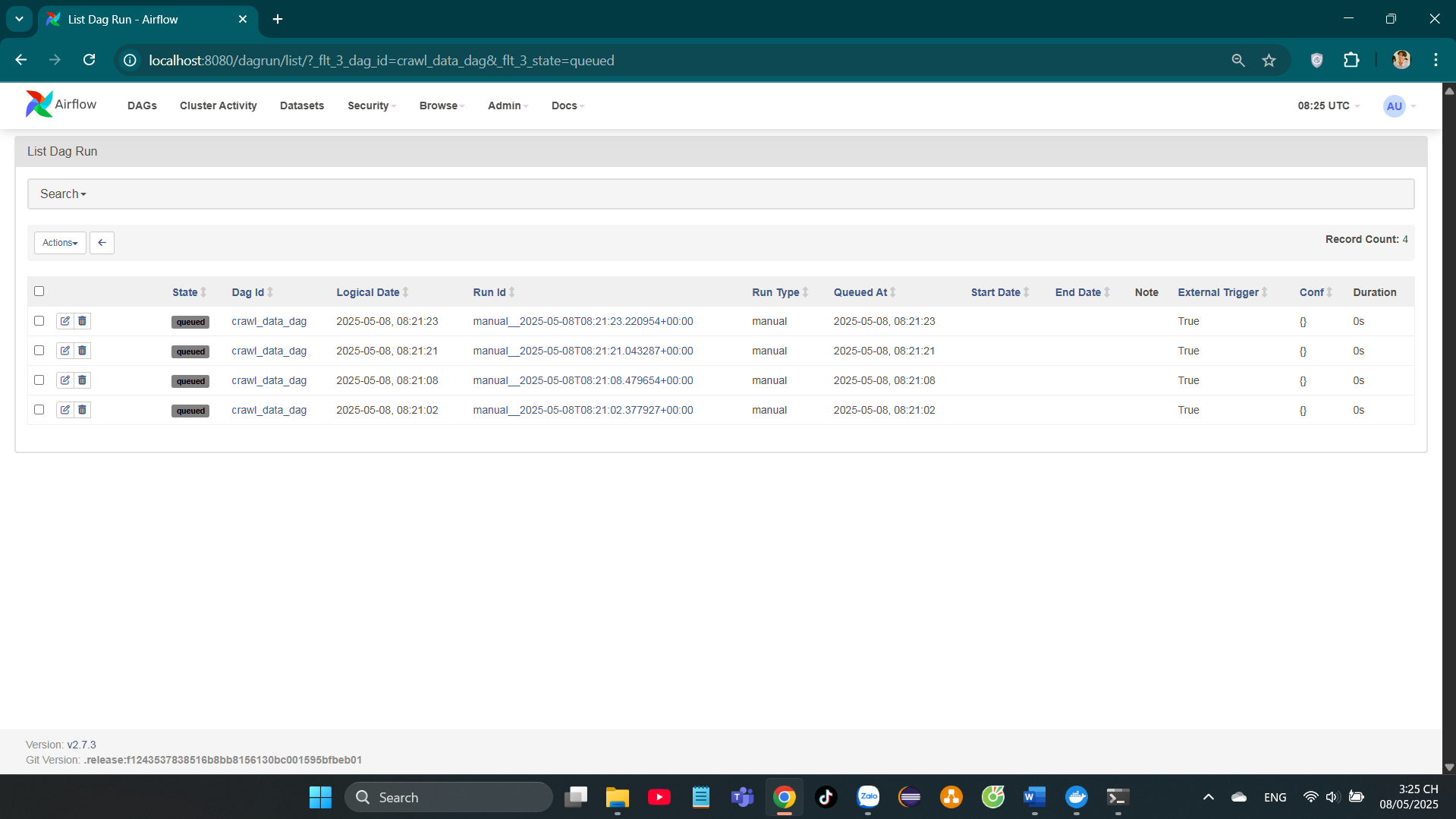


Truy cập localhost sẽ hiện lên kết quả như sau



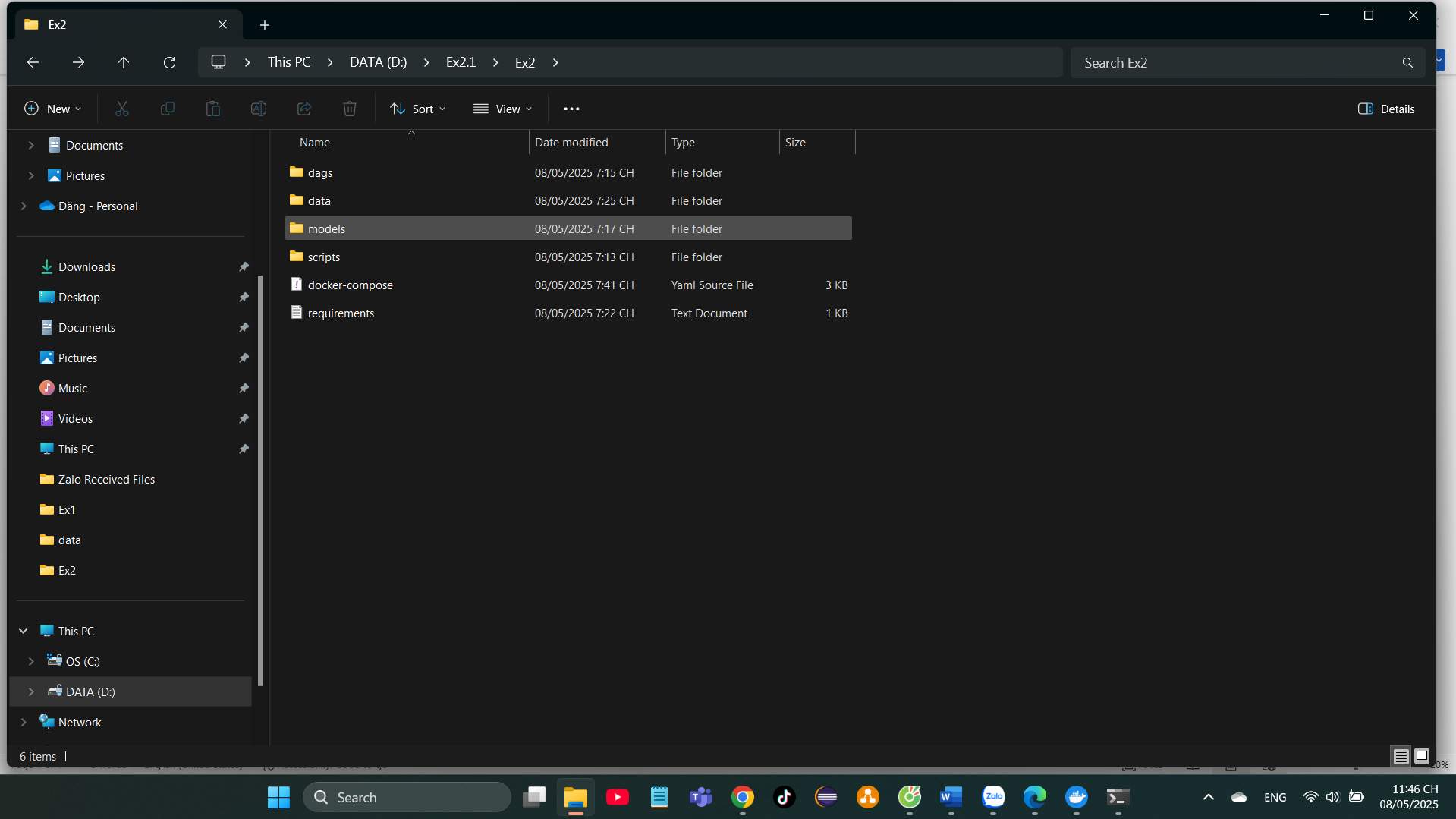
Ấn vào như ảnh dưới :



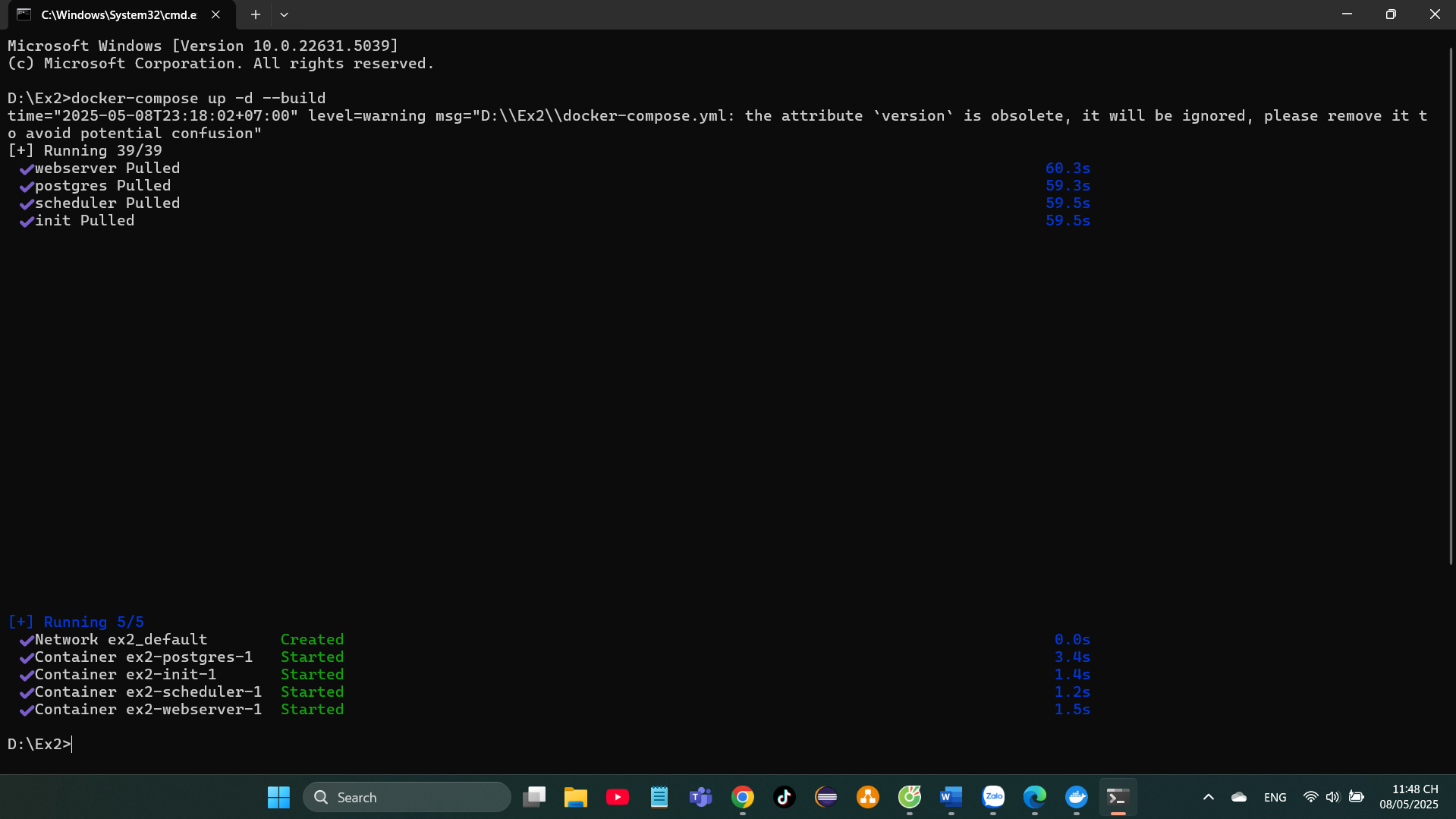
Các dữ liệu được hiện lên như sau 

**Lab8.2:**

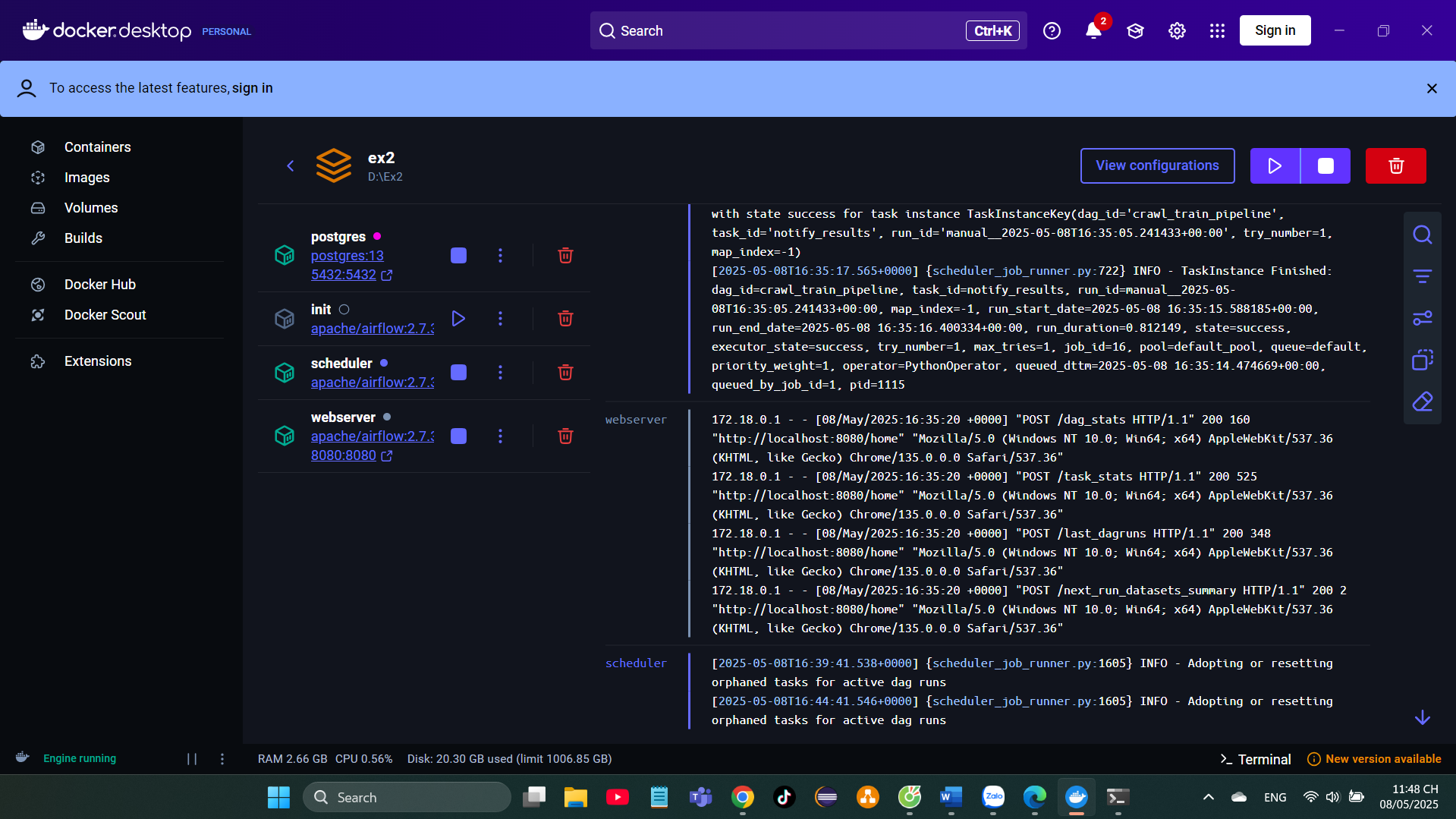
Sau khi hoàn thiện thư mục chứa docker có cấu trúc như sau:



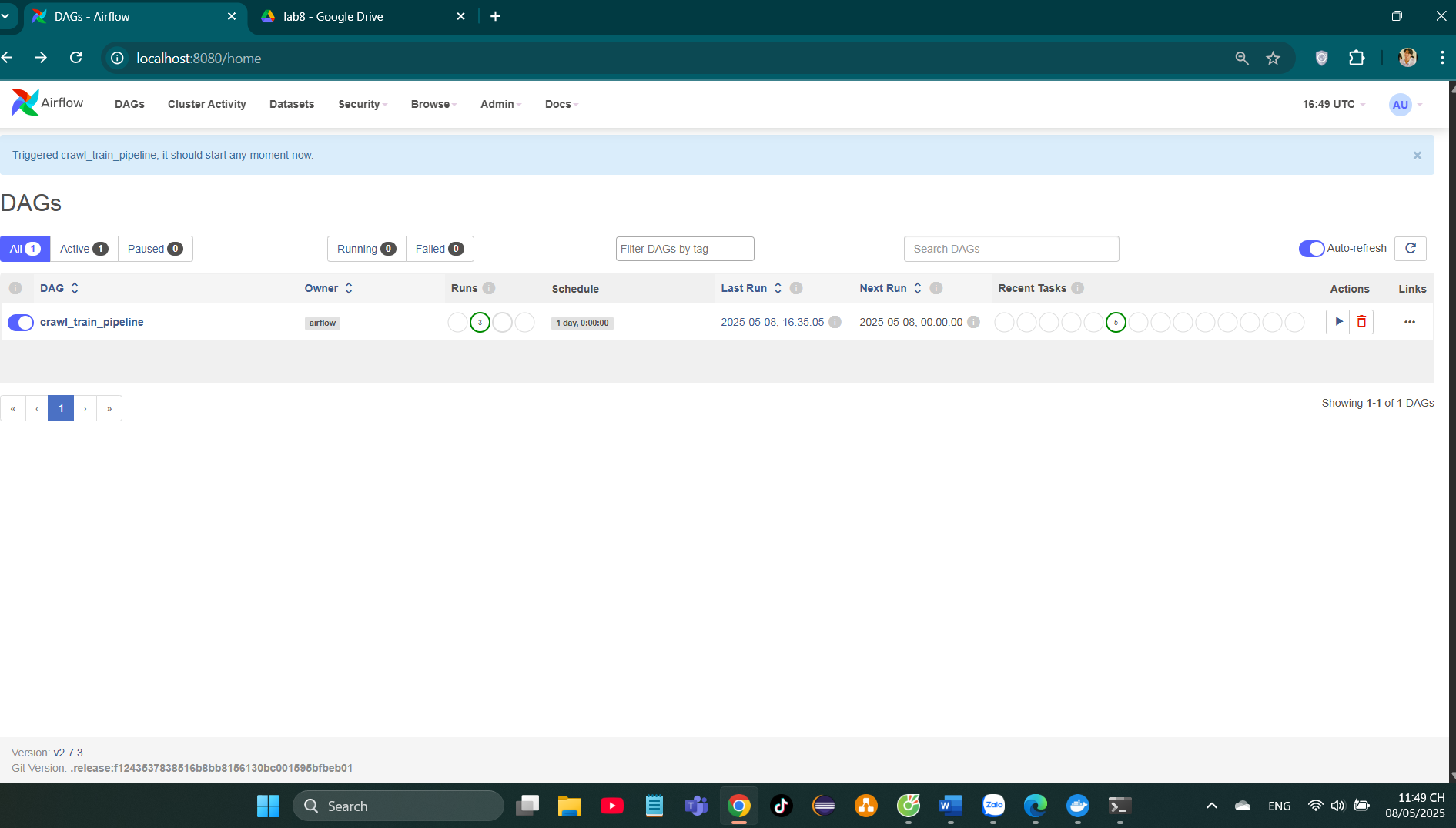
Chạy lệnh docker-compose up -d –build trong thư mục để build:

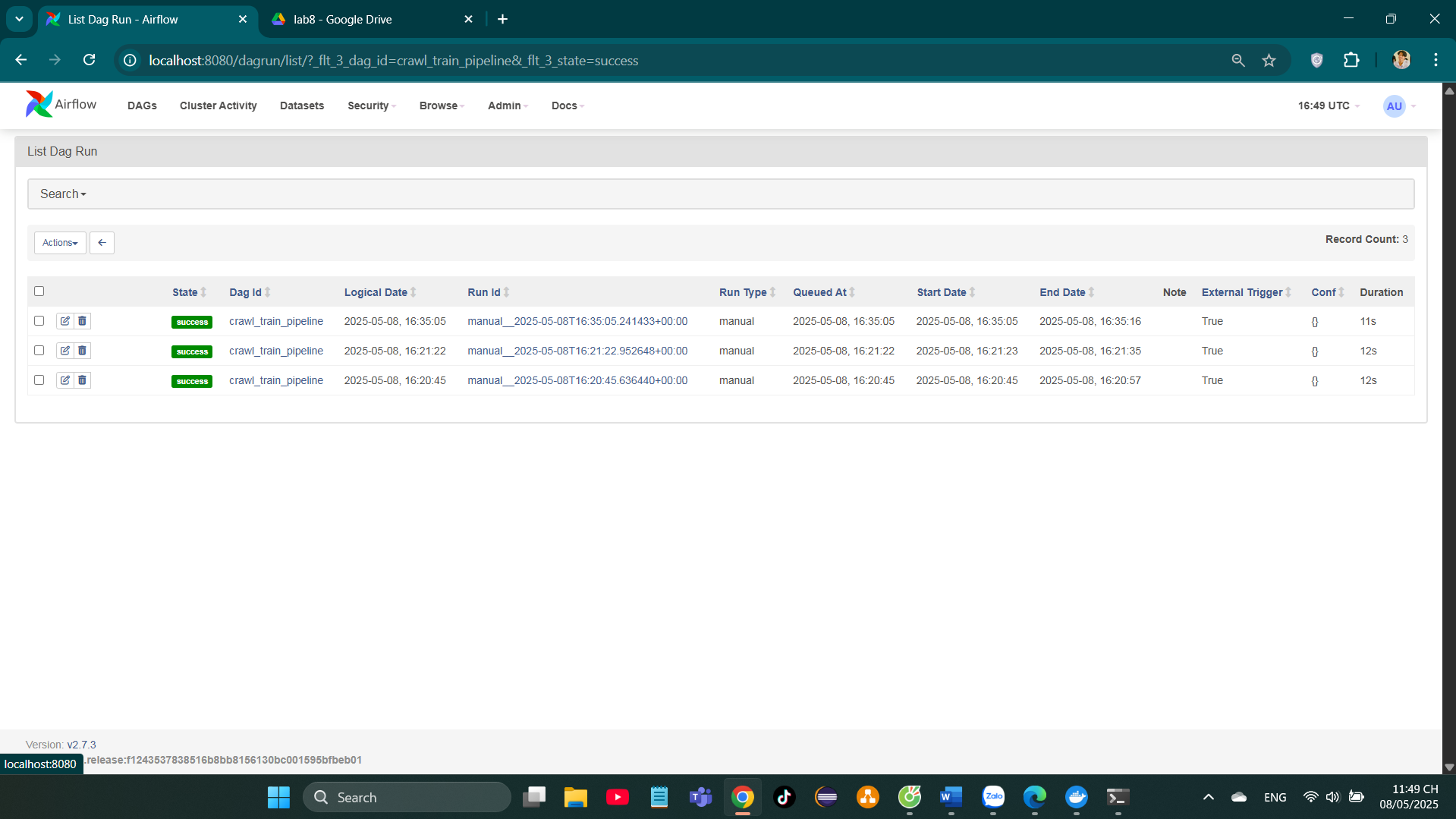


Kết quả container đc build:



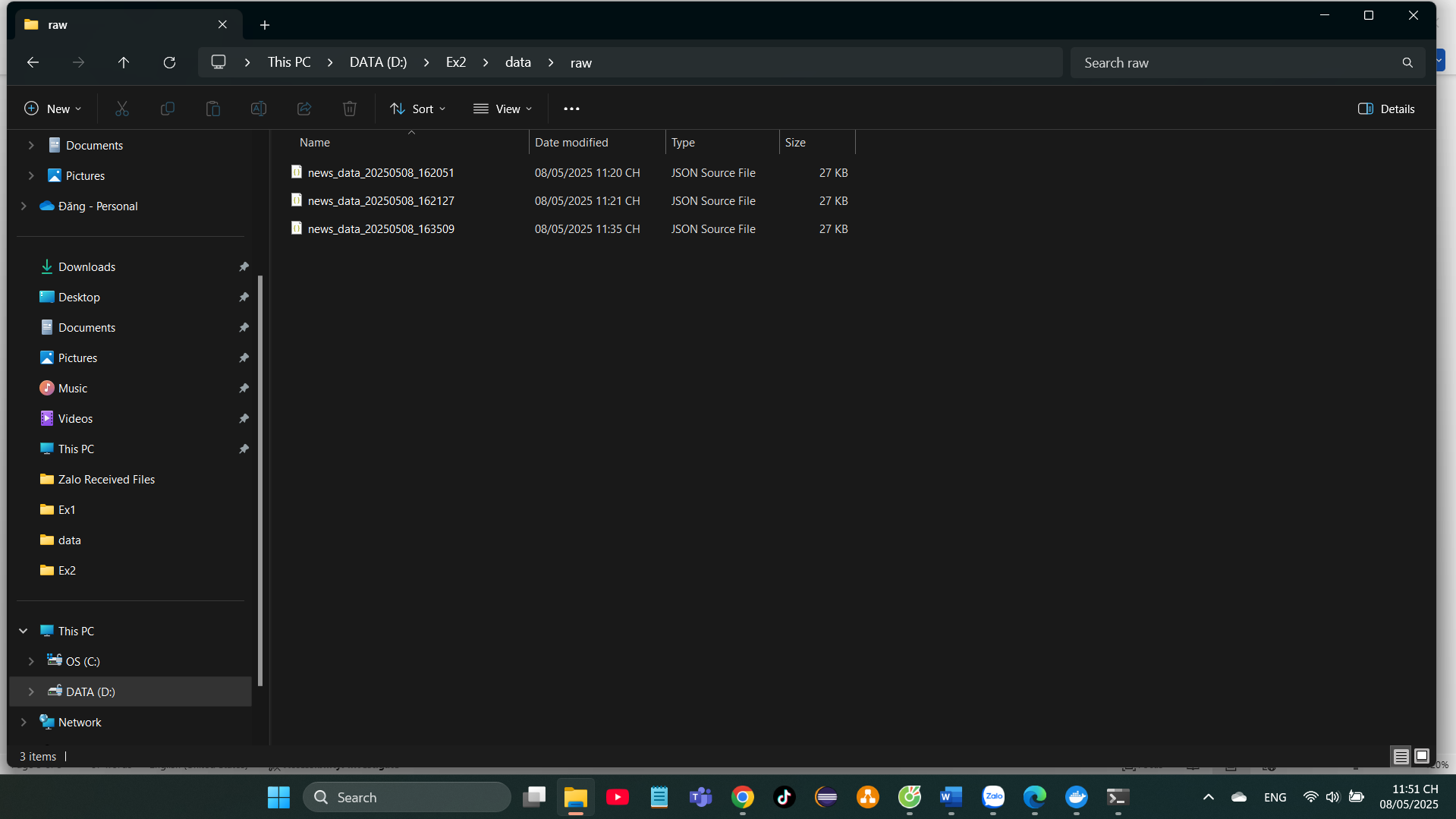
Tiếp theo truy cập locahost ta sẽ có:



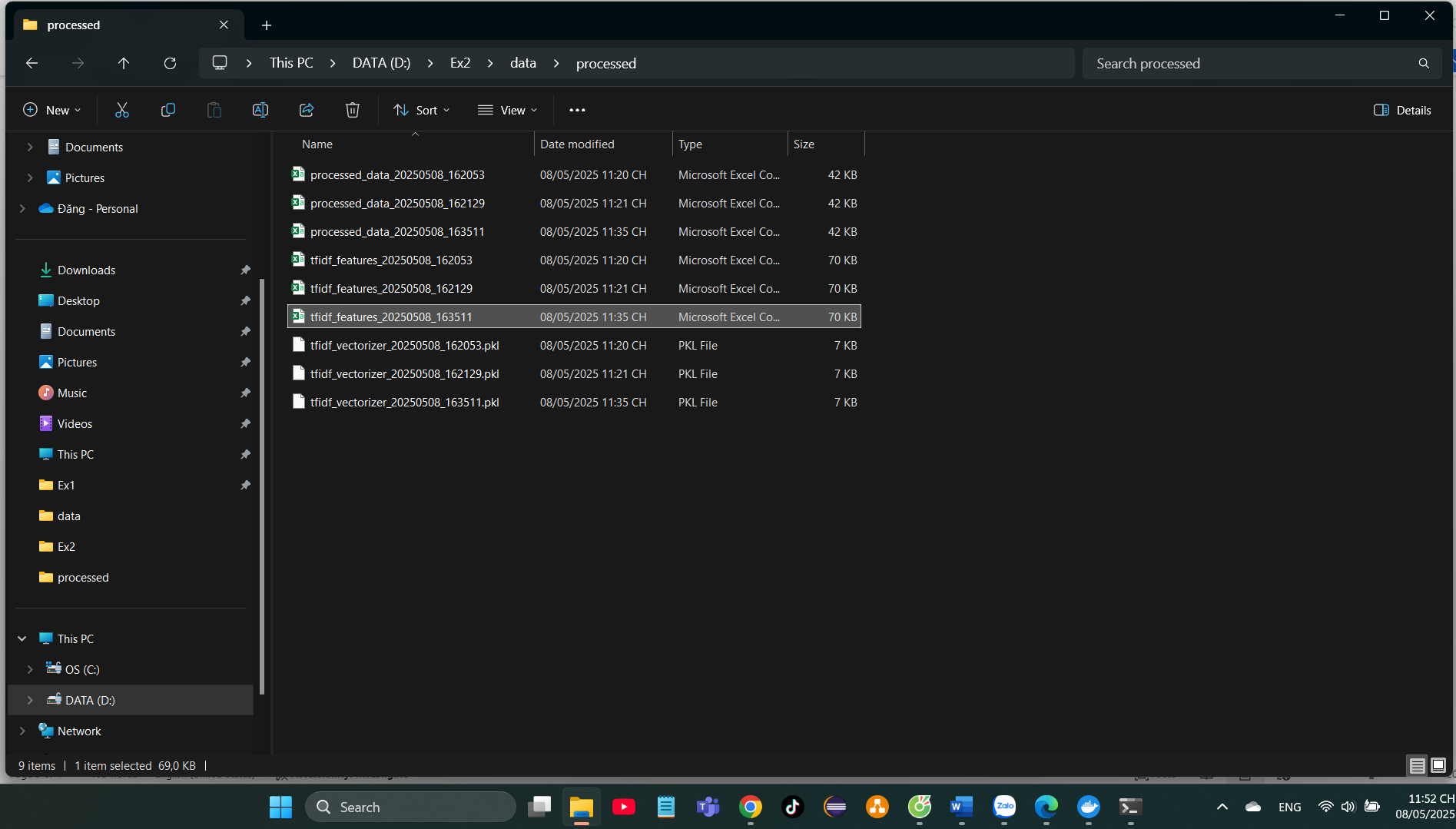


Mục tiêu của pipeline là craw được dữ liệu thô sau đó chuyển sang csv và ta sẽ có được mô hình ứng với từng dữ liệu. Ứng với các bộ dữ liệu ở trên ta sẽ có file csv và mô hình cụ thể như hình:

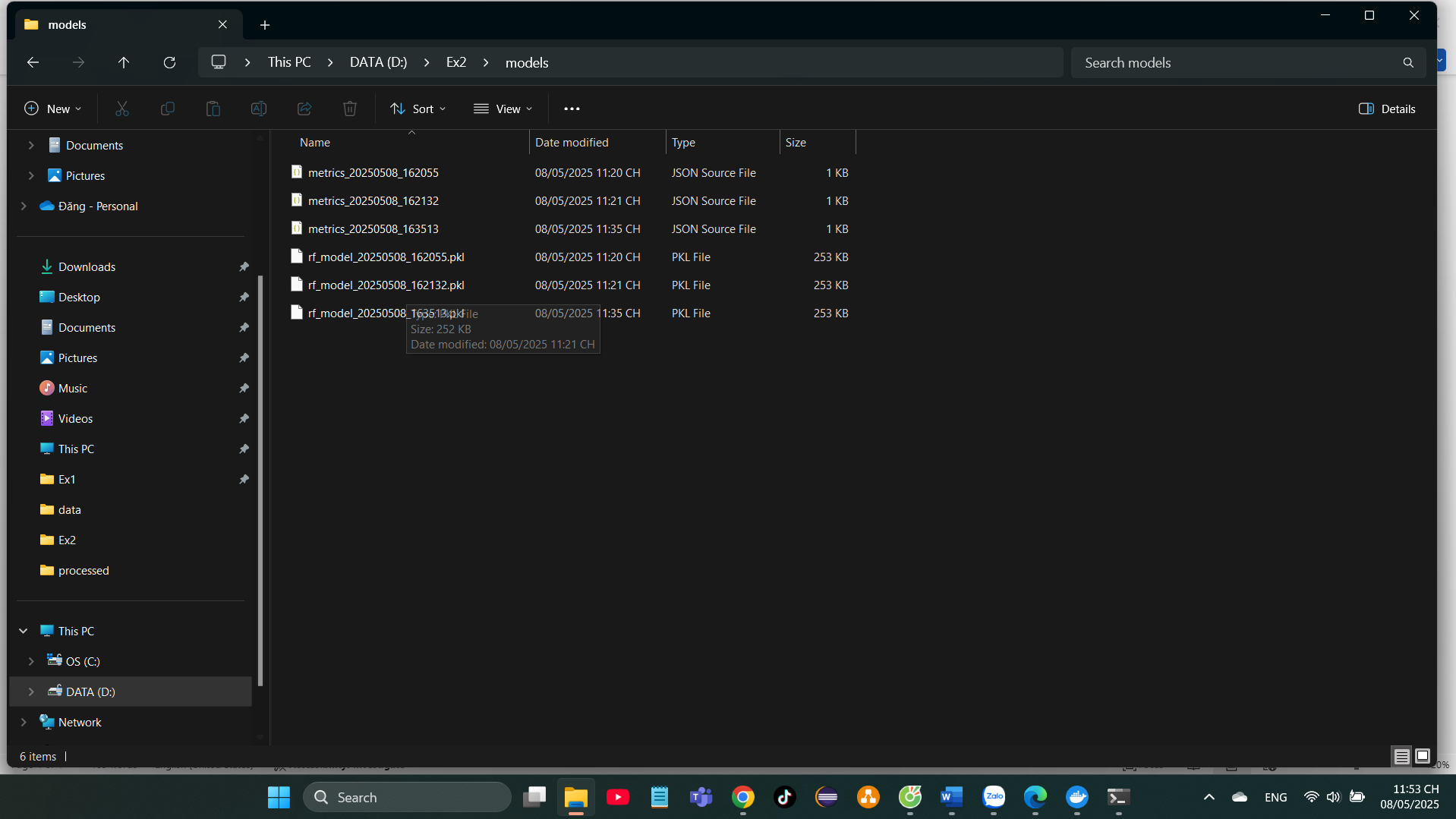
Dữ liệu craw được nằm trong thư mục raw



Dữ liệu chuyển sang file csv nằm trong thư mục processed:



Các mô hình của dữ liệu nằm trong thư mục models:



Và đó là kết quả của Case-Study 2: Xây dựng pipeline để tự động cào dữ liệu, huấn luyện mô hình