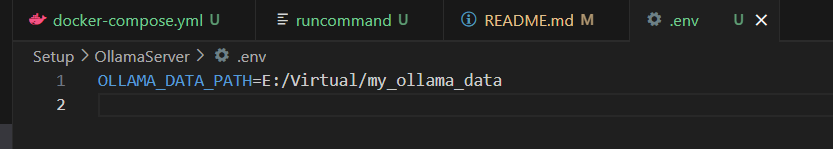
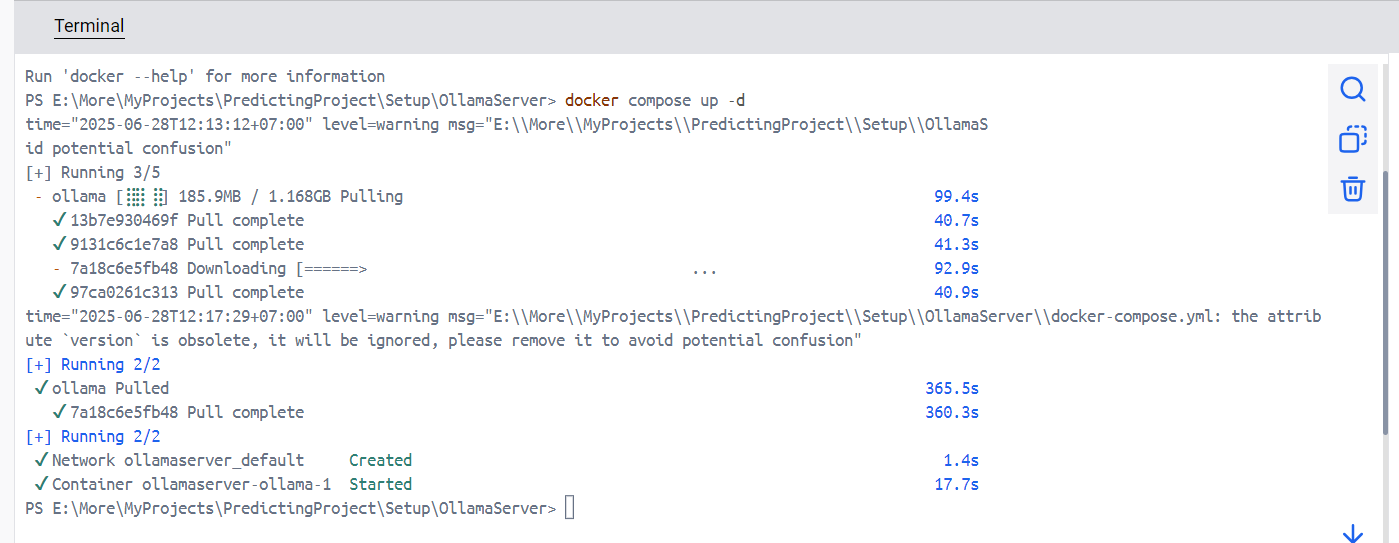
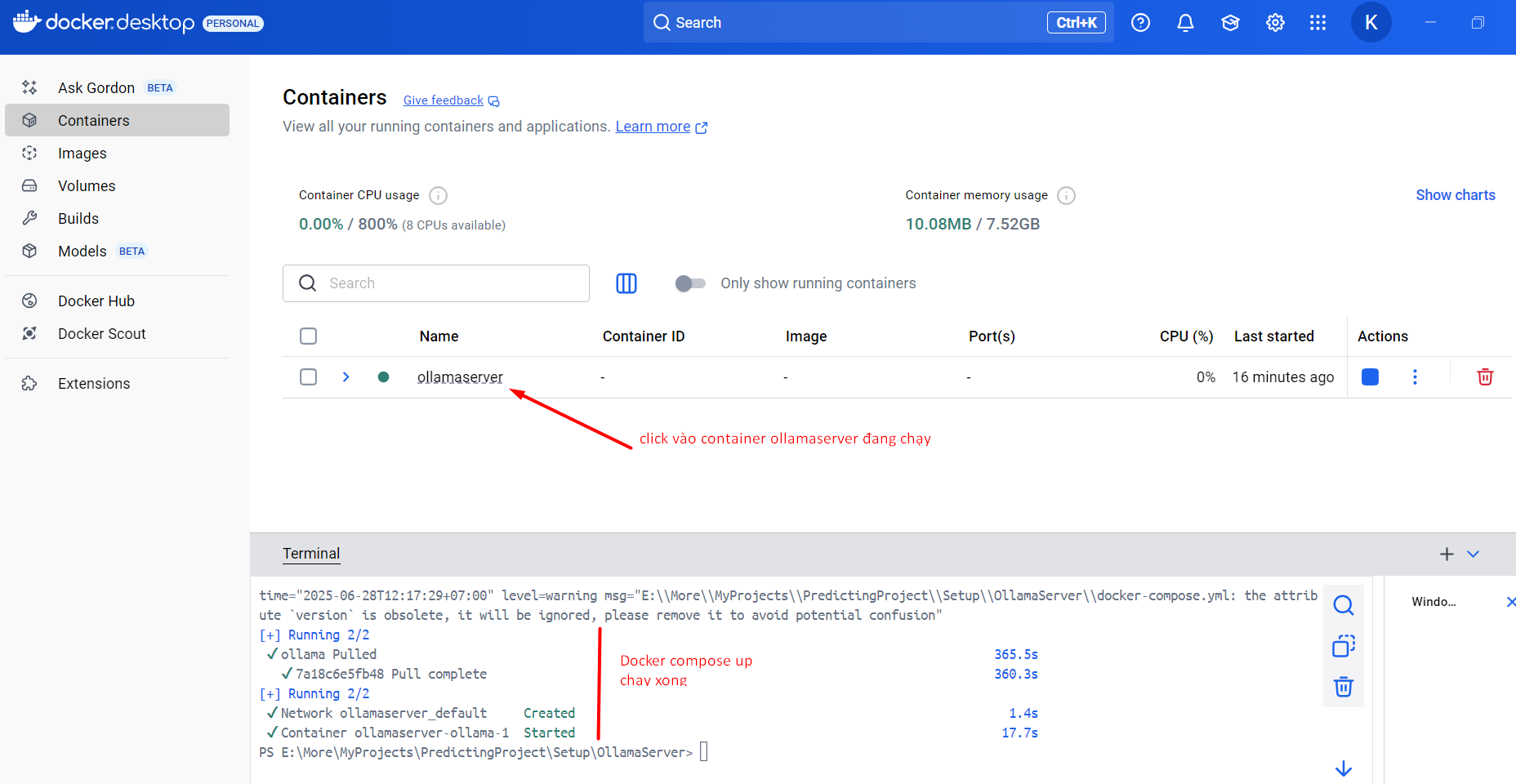
Cách setup như sau. - Thay đổi đường dẫn **OLLAMA\_DATA\_PATH** trong file **.env** bằng đường dẫn của một thư mục trong bất kỳ (tránh đặt trong ổ C)

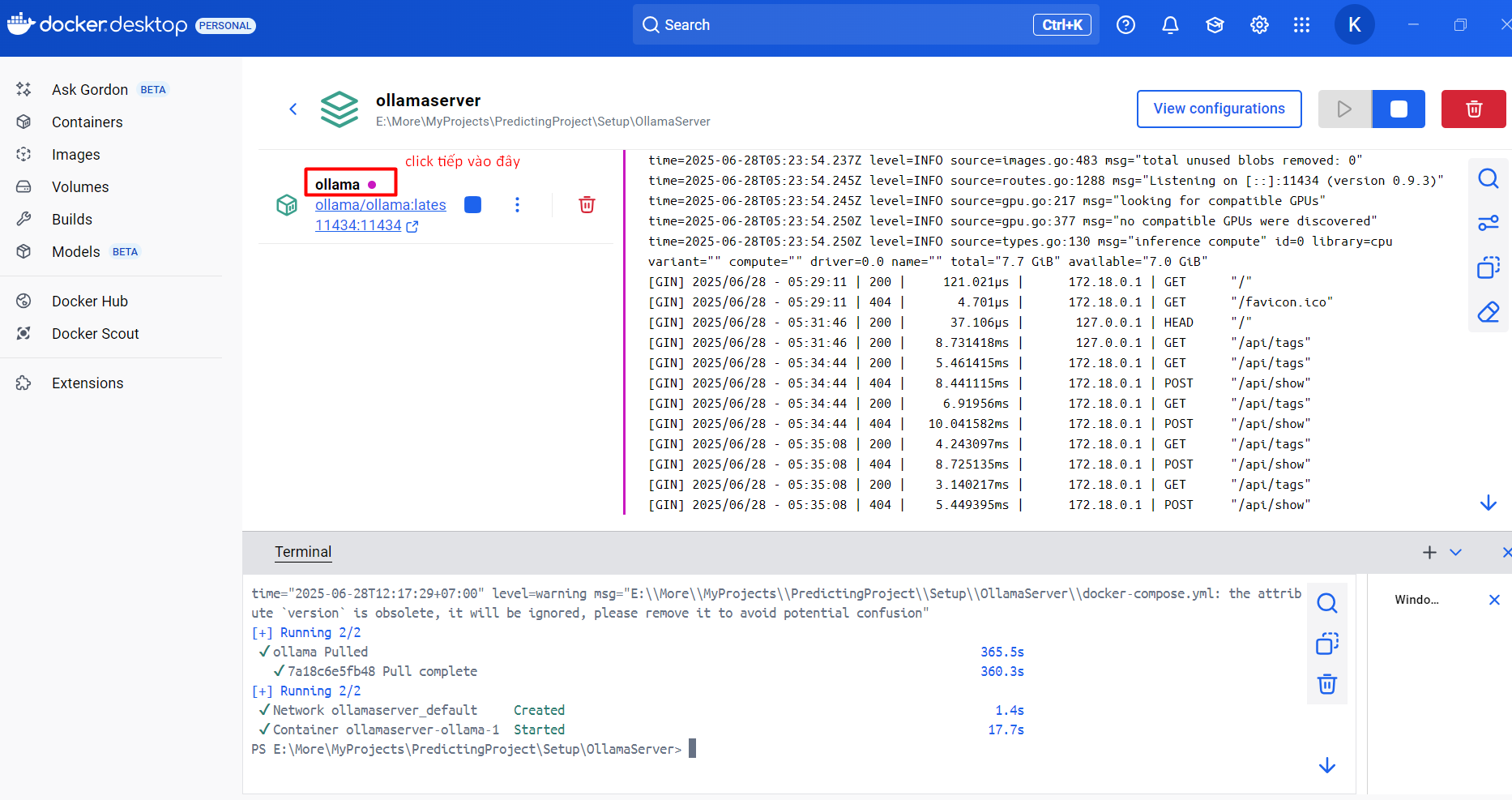


- Trong Terminal của Docker desktop, cd directory đến thư mục chứa file **docker-compose.yml** - chạy câu lệnh "**docker compose --env-file .env up -d**"



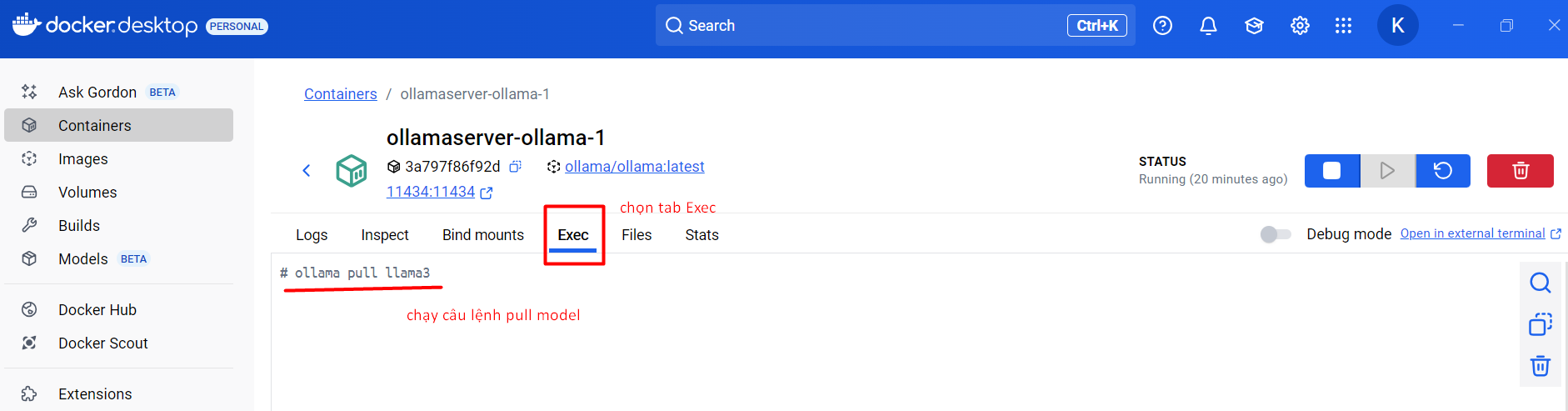
Sau khi docker-desktop chạy xong câu lệnh docker compose up - vào bên trong Container ollamaserver đang chạy

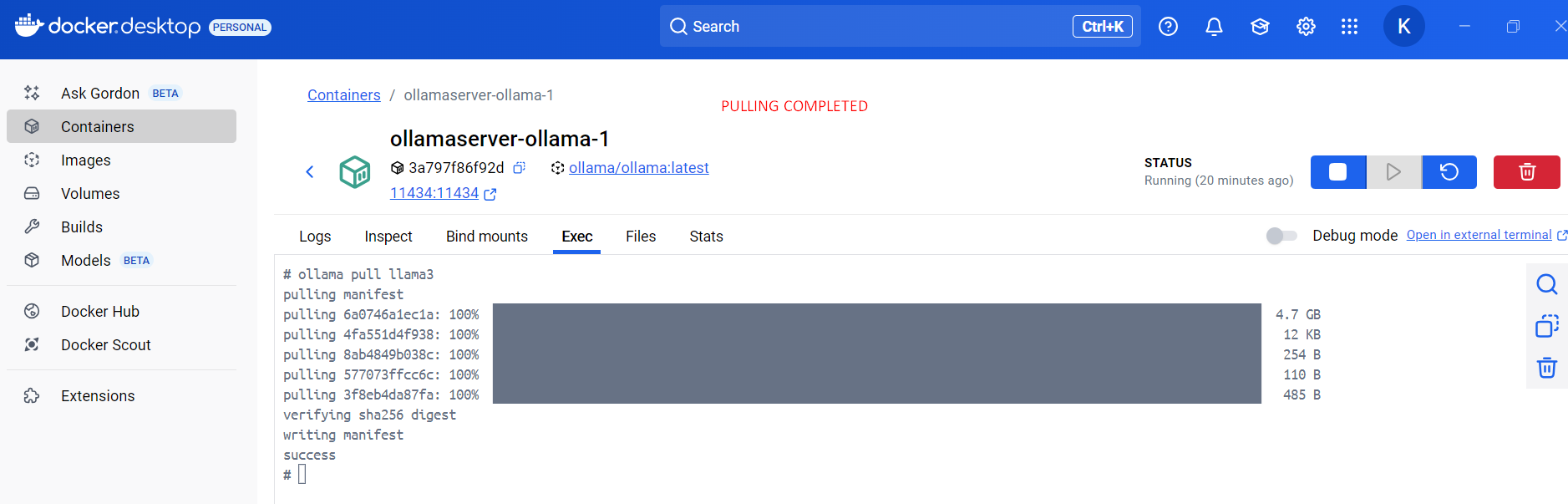




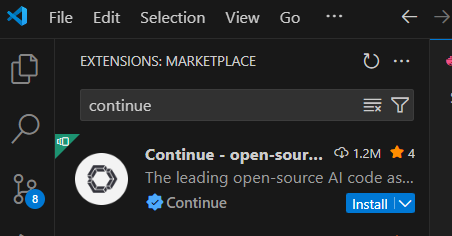
- Chọn một hoặc nhiều Model với dung lượng phù hợp để chạy trên Ollama server. Mình chọn **llama3 (4.9G)** và **deepseek-coder:6.7b (3.8G)**. List các Model có thể tìm ở <https://ollama.com/search>

- Vào tab **Exec** gõ câu lệnh “**ollama pull llama3”** hoặc “**ollama pull deepseek-coder:6.7b**”, tùy theo model được chọn.

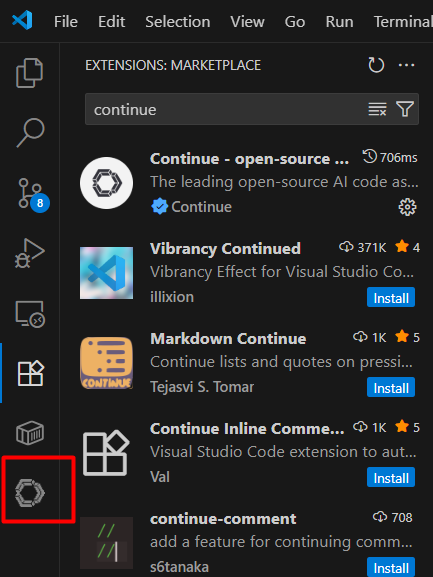




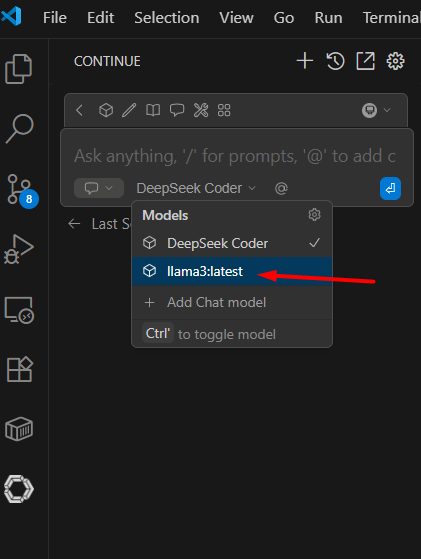
**Mở vscode, cài đặt extension Continue**



Sau khi cài đặt, biểu tượng của Extension Continue trên thanh công cụ bên phải



Kêt nối Continue với Ollama server và chọn model vừa được cài đặt (Cách kết nối có thể … hỏi ChatGPT – mình quên cách r 😊)



More documents for Continue extension: <https://docs.continue.dev/>