**STT7 - Nguyễn Văn Hùng - 23668291**

**Đề bài: gold\_pipeline *(*** *10 điểm****, CLO 1)***

a. (2 điểm) Viết script python để Crawl tin tức từ 1 trang web

https://vnexpress.net/cong-nghe/ai

A screenshot of a computer

AI-generated content may be incorrect.

b. (2 điểm) Viết script python để làm sạch dữ liệu: loại bỏ các thẻ HTML, lấy các

thông tin cần thiết

A screenshot of a computer

AI-generated content may be incorrect.

c. d. (2 điểm) Viết script python để lưu vào PostgreSQL (có script tạo bảng, kết nối DB)

A screenshot of a computer

AI-generated content may be incorrect.

(2 điểm) Viết Dockerfile + docker-compose.yml để đóng gói và chạy ứng dụng và

chạy app và PostgreSQL

A screenshot of a computer

AI-generated content may be incorrect.

e. (2 điểm) Viết DAG Airflow để chạy pipeline trên lúc 10h sáng mỗi ngày

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.