LUONG NGUYEN YEN NHI

github.com/lnynhi02

Skills

Python, Postgres, MySQL, Spark, Kafka, Airflow, Docker, Azure

Projects

New York Trips Stream Processing (Kafka, Spark, Postgres, PowerBI, Docker)
- Implemented Docker to run Kafka and Kafka UI.
- Employed Spark to transform and compute streaming data from Kafka, performing real-time analytics.
- Calculated average hourly revenue, identified trip discrepancies between boroughs, and detected trips with unusual durations (e.g., under 2 minutes) or abnormal fare amounts (e.g., total amount not matching the sum of individual fees).
- Transferred data and computations to PostgreSQL for monitoring.
- Visualized and analyzed trip trend insights more easily with a PowerBI dashboard.

Web Scrapping Pipeline (Python, Postgres, Airflow)
- Used BeautifulSoup and Requests to scrape data from TOPCV.
- Cleaned the data and stored it in Postgres.
- Generated a "due date" field, representing the application deadline based on the remaining time to apply from the scraped data.
- Developed a procedure to update the "remaining time to apply" daily, e.g., from "24 days left to apply" today to "23 days left to apply" tomorrow.
- Utilized Airflow to orchestrate the data scraping, cleaning, and storage processes.

Education

Sai Gon University   Ho Chi Minh City
Bachelor of English Linguistic   Aug, 2020