LUONG NGUYEN YEN NHI

Data Engineer

Ho Chi Minh City • (+84) 334 110 133 • lnyn.netlify.app • ynnhi1508@gmail.com

SUMMARY

Aspiring Data Engineer driven to contribute through impactful, data-driven solutions. Eager to learn, unafraid of challenges, and adaptable to any environment to continuously grow and deliver value.

- **Short-term Goal:** Secure a Data Engineer internship or an entry-level (fresher) position to gain practical experience and enhance my technical skills.
- **Long-term Goal:** Become a Senior Data Engineer within the next five years.

TECHNICAL SKILLS

- **Programming:** Python
- **Database:** PostgreSQL, MySQL
- **Big Data Tools:** Spark, Kafka
- **Orchestration:** Airflow
- **Cloud:** Azure (DP-203 Knowledge)

PROJECTS

Coffee Sales Analytics Platform | Python, MySQL, Kafka, Redis, Spark, Airflow, Minio

June 2025 *Source: Project Website*

- Goal: Built a hybrid pipeline for real-time product suggestion and batch analytics for a coffee chain.
- Used Kafka Connect to enable real-time CDC from MySQL for streaming new orders.
- Handled +1000 orders per minute with ~5ms response time.
- Scheduled daily batch jobs with Airflow and Spark, applying Medallion Architecture and SCD Type 2.
- Applied Deequ for automated data quality checks, enabling rule-based validation and anomaly detection in the batch layer.
- Integrated monitoring and alerting using Prometheus and Grafana.

Automated Job Data Pipeline | Python, Playwright, Postgres, Airflow

October 2024 *Source: Project Website*

- Goal: Automate job data scraping and streamline storage with daily updates for better job monitoring.
- Scraped job data from TOPCV using Requests and Playwright.
- Cleaned and stored the data in PostgreSQL, generating a "due date" field based on the application deadline.
- Automated updates for the "remaining time to apply" field daily (e.g., "24 days left" to "23 days left").
- Orchestrated the entire pipeline (scraping, cleaning, and storage) with Airflow for seamless automation.

Real-Time NYC Trip Analytics | Python, Postgres, Kafka, Spark, PowerBI, Docker

August 2024 *Source: Project Website*

- Goal: Improved trip monitoring and anomaly detection, providing actionable insights for NYC transport services.
- Deployed Dockerized Kafka for data streaming and Spark for real-time transformation and analytics.
- Analyzed trip trends and detected anomalies (e.g., short trips, fare mismatches).
- Stored processed data in PostgreSQL and visualized insights via an interactive PowerBI dashboard.

EDUCATION

Sai Gon University

Bachelor of English Linguistic GPA: 3.16