TRAN NHAT DUONG

DATA ENGINEER

+84 825 687 941
nhatduong01012005@gmail.com
Github
Linkedin

ABOUT ME

My name is Duong, a Third-year Information Technology student at VNUHCM-University of Science with a deep passion for Data Engineer and Software Development. I thrive on solving complex problems, uncovering insights from data, and building intelligent systems that drive innovation.

EDUCATION

University of Science, VietNam National University HCM *2023 - 2027 (expected)*
Bachelor of Computer Science
GPA: 3.74/4.0

TECHNOLOGY

Programing Language: Python, C++, SQL, JavaScript
ETL / Orchestration: Airflow, dbt
Big Data Processing: PySpark
Cloud Platform: Microsoft Azure (ADLS Gen2, Databricks)
DevOps/Tools: Docker, Git, Linux.
Machine Learning: Scikit-learn

PROJECT

VietNamworks DE Pipeline **Jan 2026 - Feb 2026**

***Role**: Data Engineer*

***Tech**: Python, dbt, Airflow, PostgresSQL (Neon), Docker, Azure*

***Github**: VietNamworks_DE_Pipeline*

Description: An automated, scalable End-to-End ETL Pipeline designed to crawl, store, and transform job market data from VietnamWorks

Key feature:

- Medallion ELT: Architected a Raw-Silver-Gold pipeline using Azure Data Lake Gen2 and PostgreSQL.
- Scalable Ingestion: Engineered cloud ingestion with Python (adlfs) and Airflow, leveraging HNS for big data optimization.
- Modular dbt: Orchestrated SQL transformations with dbt Core.
- Docker Containerization: Containerized the entire infrastructure (Airflow, Redis, Postgres) for seamless Dev-to-Prod deployment.

Vietnamese-Chinese Corpus Pipeline **Nov 2025 - Dev 2025**

***Role**: Data Engineer, Core Developer*

***Tech**: Python, PyTorch (LABSE), Vecallign*

Github: Vie_Chn align pipeline

Description: An e2e pipeline serves for cleaning, and segmenting raw bilingual text (JSON) into high-quality parallel datasets (CSV).

- Achieved over 90% accuracy in bilingual sentence alignment.
- Implemented LaBSE embeddings and the Vecalign algorithm to resolve complex sentence mismatches (1-N, N-1) based on vector cosine similarity.
- Integrated cross-platform hardware acceleration support (NVIDIA CUDA & Apple MPS), significantly reducing processing latency for large-scale text embedding.

CERTIFICATIONS IELTS: 6.5 (2022 - 2024)