Hong Van Pham

vanhongpham01@gmail.com • (438) 630-0695

malajvan.github.io • linkedin.com/in/vanhpham • github.com/malajvan

WORK EXPERIENCE

TEC Energy — Montreal, Canada
Data Engineer — February 2024 – Present

- Developed backend data pipelines and ingestion workflows using the **AWS stack (Lambda, S3, CDK)**, enabling users to self-serve custom data imports.
- Reduced **RDS PostgreSQL** database bloat by over 35% using pg_repack and table migrations, supporting the transition from a monolithic DB to a more modular architecture, cutting costs and improving performance.
- Led zero-downtime transition of critical backend data sources, maintaining continuous access to high-volume, revenue-generating data and preventing pipeline disruption during infrastructure updates.
- Developed an ETL pipeline in **Python** for ingesting weather forecast data using **pandas**, **xarray**, and **geopandas**, supporting analytics and machine learning applications.
- Built a proof-of-concept BI stack using **Apache Superset** on **ECS** with **Docker**, **Amazon Cognito**, and load balancer/networking configs.
- Built and maintained internal **APIs** for data access, transformation, and trading operations, used across trading and analytics teams.
- Authored detailed documentation and contributed to architecture discussions, promoting shared understanding and reliability across the team.

Distributed Digital Music Archives & Libraries Lab, McGill University — Montreal, Canada
Data Engineer — August 2023 – February 2024

- Spearheaded the development of the **data lake** architecture and **ETL** framework for a new project LinkedMusic.
- Implemented graph database technology, utilizing **SPARQL** and Openlink **Virtuoso** for effective data virtualization.
- Leveraging **JSON-LD** and **OpenRefine** to facilitate seamless integration with **linked data** sources, enhancing interoperability and data relationships for 5+ database clients.
- Collaborated with grant co-investigators and delivered project presentations before the grant review board.

Software Developer — May 2022 – August 2023

- Optimized **Docker** Hub build time by 30% for **Django**-based workflow management software.
- Improved workflow run times by 15-30% and reduced crash rate by 60% by migrating codebase to **Python3**.
- Mitigated security risks by 30% and facilitated testings by setting up staging server, reduced deployment time by 50% through standardized optimized feature deployment process.

Vivas Technology — Hanoi, Vietnam
Data Science Intern — January - April 2021

- Eliminated 10% of office manual data entry tasks by implementing reusable ETL pipeline from factories to **ClickHouse** with **Apache Nifi** using **SQL**.
- Built performance metrics BI dashboards in **Apache Superset** for the board of directors' decision-making.
- Automated data loading tasks and cut data analysis time by an additional 30% .

EDUCATION

McGill University — Montreal, Canada
Bachelor in Mathematics and Computer Science, 3.62/4.0 GPA — September 2019 - May 2023

- McHacks 10 (2023), MHCPP Volunteer, La Belle Tonki (cook)

SKILLS & Certifications

- **Certifications:** Microsoft AZ-900: Azure Fundamentals (August 2023)
- **Technologies & Skills:** Python, SQL, Docker, Git, Microsoft Azure, AWS, Server Management, Cloud Computing