Hong Van Pham

vanhongpham01@gmail.com • (438) 630-0695
malajvan.github.io • linkedin.com/in/vanhpham • github.com/malajvan

WORK EXPERIENCE

Distributed Digital Music Archives & Libraries Lab, McGill University — **Montreal, Canada**

Data Architect — **August 2023 – Present**

- Spearheaded the development of the data lake architecture and **ETL** framework for a new project LinkedMusic.
- 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 to secure ongoing funding and support.

Lead Developer — **May 2023 – Present**

Software Developer — **May 2022 – April 2023**

- Trained 5+ new employees, oversaw task assignments, provided feedback and code reviews for all projects.
- Optimized Docker Hub build time by 40% 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 with **Ansible**, 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 — **Sep 2019 - May 2023**

- Relevant Courses: Database Systems (COMP 421), Data Science (COMP 598)
- McHacks 10 (2023), MHCPP Volunteer, La Belle Tonki (cook)

PROJECTS

Azure Pitchfork Data Analysis — **September 2023**

- Implemented **Data Factory** pipeline for data transfer to Azure Blob Storage with Data Lake Gen2 integration.
- Utilized **Databricks** and **Spark** for efficient data cleaning and transformation, ensuring high data quality.
- Employed **Tableau** for impactful data visualizations, facilitating data-driven decision-making.

Spotify Popularity Predictions — **January 2022**

- Extracted songs' audio features (MFCC) using **librosa** and **PyTorch** and predicted songs' popularity
- Implemented **KNN** algorithm with **pandas** and **scikit-learn** to classify popular songs' common features
- Correctly predicted 89% of sampled data in comparison to Spotify's API

SKILLS & Certifications

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