



## KỸ SƯ DỮ LIỆU

VŨ DƯƠNG HẠNH

### MỤC TIÊU NGHỀ NGHIỆP

Tôi hướng tới mục tiêu trong 3 năm tới trở thành một Senior Data Engineer, có khả năng thiết kế toàn bộ kiến trúc dữ liệu, dẫn dắt đội nhóm và đề xuất cải tiến cho chiến lược dữ liệu doanh nghiệp.

### THÔNG TIN CÁ NHÂN

15/06/1984

Hà Nội

thaolinh252512@gmail.com

0356990989

www.website.com

### HỌC VẤN

- Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM - Khoa học dữ liệu tại Đại học Công nghệ Thông tin - ĐHQG TP.HCM

### KỸ NĂNG

- Monitoring với Prometheus/Grafana
- Python

### KINH NGHIỆM LÀM VIỆC

- **CLOUD DATA ENGINEER** TẠI SMARTBANK (2022-2023)

+ THIẾT KẾ KIẾN TRÚC DỮ LIỆU TRÊN AWS VỚI S3, GLUE VÀ REDSHIFT

+ TỰ ĐỘNG HÓA VIỆC INGESTION LOG GIAO DỊCH NGÂN HÀNG TỬ HỆ THỐNG CORE

+ THAM GIA ĐÁNH GIÁ BẢO MẬT VÀ CHUẨN HÓA SCHEMA DỮ LIỆU TÀI CHÍNH

- **DATA ENGINEER** TẠI CÔNG TY DATAFLOW VIỆT NAM (2021-2023)

## SỞ THÍCH

- Du lịch
- Thể thao

+ THIẾT KẾ VÀ TRIỂN KHAI PIPELINE ETL SỬ DỤNG APACHE AIRFLOW

+ TÍCH HỢP DỮ LIỆU TỪ NHIỀU NGUỒN NHƯ MYSQL, GOOGLE ANALYTICS VÀ API BÊN THỨ BA

## NGƯỜI GIỚI THIỆU

- Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666

- Ông Phan Thành Tâm (Big Data Architect - Công ty LogiData) - tam.phan@logidata.vn - 0955111222

- Bà Phạm Thanh Hương (Data Operations Manager - Công ty ReportPro) - huong.pham@reportpro.vn - 0912111222

+ TỐI ƯU HOÁ QUY TRÌNH XỬ LÝ BẰNG SPARK TRÊN GCP DATAPROC

## DANH HIỆU VÀ GIẢI THƯỞNG

- **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge

- **2021** - Bằng khen về bảo mật và chất lượng dữ liệu - Công ty InfosecData

## CHỨNG CHỈ

- **2021** - SQL for Data Engineering - Datacamp

- **2021** - IBM Data Engineering Professional Certificate - Coursera

- **2020** - Google Cloud Professional Data Engineer Certificate

## HOẠT ĐỘNG

- **Tình nguyện viên hỗ trợ nghiên cứu học thuật tại Khoa Công nghệ Thông tin (2020)**

+ Chuẩn hóa dữ liệu nghiên cứu sinh viên từ nhiều năm học.

+ Phân tích và tổng hợp xu hướng đăng ký môn học.

- + Viết báo cáo kỹ thuật trình bày cho giảng viên hướng dẫn.

#### **- Cộng tác viên dự án dữ liệu mở tại OpenData Vietnam (2020)**

- + Thu thập và chuẩn hóa dữ liệu dân số từ các tỉnh thành.
- + Tích hợp dữ liệu vào hệ thống cơ sở dữ liệu MongoDB phục vụ phân tích.
- + Tạo dashboard đơn giản để trực quan hóa dữ liệu bằng Plotly Dash.

#### **- Thành viên nhóm tư vấn kiến trúc dữ liệu tại Công ty SmartData (2021 - 2022)**

- + Đề xuất mô hình lưu trữ dữ liệu dạng columnar cho truy vấn phân tích nhanh.
- + Tham gia thiết kế cấu trúc bảng dữ liệu trong hệ thống warehouse.
- + Xây dựng best practice về đặt tên và phân vùng dữ liệu.

#### **- Người viết blog kỹ thuật về hệ thống dữ liệu tại Blog cá nhân datapi.dev (2022 - nay)**

- + Chia sẻ kiến thức về thiết kế hệ thống ETL quy mô lớn.
- + So sánh các giải pháp lưu trữ dữ liệu: PostgreSQL, Redshift, BigQuery.
- + Viết hướng dẫn cấu hình Airflow và orchestration jobs theo lịch định kỳ.

#### **- Thành viên nhóm xây dựng pipeline dữ liệu tại Phòng Lab Dữ liệu - Đại học Bách khoa (2021 - 2023)**

- + Phát triển luồng dữ liệu ETL sử dụng Apache Airflow.
- + Thiết lập hệ thống ingestion dữ liệu từ API và cơ sở dữ liệu nguồn.

- + Phối hợp với nhóm phân tích dữ liệu để tối ưu schema lưu trữ.

## DỰ ÁN

### - Data Lake cho hệ thống quản trị khách hàng (CRM) (Data Engineer, CRMPro) 2022

Xây dựng nền tảng lưu trữ dữ liệu tập trung phục vụ phân tích hành vi khách hàng.

- + Tạo pipeline ingestion từ Salesforce, Google Ads và Facebook API

- + Lưu trữ dữ liệu theo mô hình phân vùng S3 Data Lake

- + Sử dụng Airflow để lập lịch và monitor luồng dữ liệu hàng ngày

