



## KỸ SƯ DỮ LIỆU

PHẠM TÙNG LAN

### MỤC TIÊU NGHỀ NGHIỆP

Tôi muốn học hỏi sâu hơn về kiến trúc dữ liệu hiện đại, áp dụng mô hình Data Mesh và DataOps nhằm tăng khả năng cộng tác và tự động hóa quá trình triển khai dữ liệu.

### THÔNG TIN CÁ NHÂN

23/04/1980

Hà Nội

thaolinh252512@gmail.com

0338819135

www.website.com

### KINH NGHIỆM LÀM VIỆC

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

+ 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

### HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại  
Học viện Kỹ thuật Mật mã - Khoa học  
dữ liệu tại Đại học Công nghệ Thông tin  
- ĐHQG TP.HCM

### KỸ NĂNG

- Snowflake

- Google Cloud Platform (GCP)

- Data Quality Validation (Great Expectations)

- Python

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

- **ETL DEVELOPER** TẠI FINDATA (2019-2021)

+ PHÁT TRIỂN VÀ BẢO TRÌ LUỒNG DỮ LIỆU ETL SỬ DỤNG TALEND VÀ PYTHON

## SỞ THÍCH

- Sưu tầm sách lập trình
- Trồng cây
- Đi bộ đường dài
- Thiết kế sản phẩm cá nhân
- Tham gia cộng đồng lập trình

+ VIẾT SCRIPT KIỂM TRA CHẤT LƯỢNG DỮ LIỆU VÀ CẢNH BÁO KHI DỮ LIỆU SAI LỆCH

+ TÍCH HỢP HỆ THỐNG PHÂN QUYỀN DỮ LIỆU THEO VAI TRÒ NGƯỜI DÙNG

- **BIG DATA ENGINEER** TẠI LOGIAI (2020-2022)

## NGƯỜI GIỚI THIỆU

- Bà Nguyễn Thị Lệ (Analytics Lead - Công ty InsightTech) -  
le.nguyen@insighttech.vn - 0966999888

- Ông Vũ Minh Đức (CTO - Công ty DataStreamX) - duc.vu@datastreamx.vn  
- 0977333444

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

+ XÂY DỰNG PIPELINE THU THẬP VÀ XỬ LÝ DỮ LIỆU VẬN CHUYỂN TỪ THIẾT BỊ IOT

+ SỬ DỤNG SPARK STREAMING ĐỂ XỬ LÝ DỮ LIỆU REAL-TIME

+ KẾT NỐI HỆ THỐNG VÀO ELASTICSEARCH VÀ PHÁT TRIỂN DASHBOARD GIÁM SÁT

## DANH HIỆU VÀ GIẢI THƯỞNG

- **2023** - Giải thưởng 'Kiến trúc dữ liệu sáng tạo' do CTO trao tặng tại DataWare

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

- **2021** - Giải thưởng 'Dự án phân tích dữ liệu xuất sắc' tại FinData

## CHỨNG CHỈ

- **2021** - SQL for Data Engineering – Datacamp
- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera

## HOẠT ĐỘNG

### - Tình nguyện viên hỗ trợ tổ chức Data Day tại Vietnam Data Community (2022)

- + Thiết lập demo pipeline dữ liệu thời gian thực bằng Kafka.
- + Hỗ trợ phần trình diễn luồng dữ liệu streaming giữa các node Spark.
- + Giới thiệu cấu trúc cơ bản của hệ thống Data Lake.

### - Thành viên nhóm phát triển hệ thống phân tích log tại Công ty SecureLog (2021)

- + Xây dựng pipeline thu thập và phân tích log từ nhiều nguồn hệ thống.
- + Triển khai hệ thống cảnh báo bất thường dựa trên mô hình thống kê.
- + Sử dụng ELK stack để trực quan hóa dữ liệu log.

### - 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.

## DỰ ÁN

### - Data pipeline xử lý log giao dịch hệ thống ngân hàng (Data Pipeline Engineer, BankDataTech) 2020

Lưu trữ, phân tích và phát hiện bất thường từ log giao dịch ATM và mobile banking.

- + Sử dụng Fluentd thu thập log và gửi về Kafka
- + Phân tích real-time qua Spark Streaming
- + Lưu dữ liệu đã xử lý vào Elasticsearch và trực quan hoá bằng Kibana

### - Kiến trúc dữ liệu cho nền tảng giáo dục trực tuyến (Cloud Data Engineer, LearnHub) 2022

Chuẩn hoá kiến trúc lưu trữ và xử lý dữ liệu học viên để phục vụ dashboard học tập.

- + Thiết kế hệ thống lưu trữ với BigQuery và Data Studio
- + Xây dựng Dataflow jobs để xử lý dữ liệu sự kiện học tập
- + Tạo luồng dữ liệu từ Firebase tới Google Cloud

### - 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

