



## KỸ SƯ DỮ LIỆU

PHẠM TRUNG THỊ

### MỤC TIÊU NGHỀ NGHIỆP

Tôi muốn chia sẻ kiến thức qua việc viết blog, tổ chức hội thảo kỹ thuật hoặc đóng góp tài liệu chính thức để thúc đẩy cộng đồng Data Engineering trong nước phát triển mạnh mẽ.

### THÔNG TIN CÁ NHÂN

11/08/1990

Hà Nội

thaolinhh252512@gmail.com

0701620150

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

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

### HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã

### KỸ NĂNG

- Apache Spark

### SỞ THÍCH

- Tham gia hội thảo công nghệ

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

- Chơi đàn guitar
- Thử nghiệm công nghệ mới
- Du lịch

+ 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

## NGƯỜI GIỚI THIỆU

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

- Ông Hồ Trung Kiên (Engineering Manager - Công ty ETLTech) - kien.ho@etltech.vn - 0933888777

- Bà Trần Thị Mai (Technical Manager - Công ty CloudData) - mai.tran@cloudata.vn - 0933666777

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

- **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 TỪ HỆ THỐNG CORE

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

- **JUNIOR DATA ENGINEER** TẠI RETAILTECH (2020-2021)

+ PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỒN KHO THEO THỜI GIAN THỰC

+ XỬ LÝ DỮ LIỆU STREAMING TỪ KAFKA VÀ LƯU TRỮ VÀO BIGQUERY

+ XÂY DỰNG CÁC VIEW LOGIC TRÊN DBT PHỤC VỤ BI DASHBOARD

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

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

+ 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

## DANH HIỆU VÀ GIẢI THƯỞNG

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

- **2020** - Giải thưởng 'Pipeline ổn định nhất năm' tại TechFlow

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

- **2021** - Kỹ sư dữ liệu xuất sắc quý I tại Công ty DataBridge

## CHỨNG CHỈ

- **2023** - Modern Data Engineering with dbt – dbt Labs

- **2021** - Streaming Data with Apache Kafka – Confluent Academy

## HOẠT ĐỘNG

### - Diễn giả hội thảo 'Big Data Architecture' tại Data Talks Vietnam (2022)

- + Trình bày kiến trúc hệ thống thu thập và xử lý dữ liệu đa nguồn.
- + Phân tích ưu nhược điểm của Data Warehouse vs Data Lakehouse.
- + Giới thiệu các công cụ phổ biến như Airflow, dbt, Snowflake.

### - 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.

### - 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.

#### **- Mentor khóa học nền tảng kỹ sư dữ liệu tại Trung tâm Đào tạo CloudTech (2023)**

- + Hướng dẫn học viên triển khai hệ thống ingest dữ liệu bằng Kafka.

- + Đánh giá bài tập về xử lý dữ liệu song song với Spark.

- + Tư vấn về định hướng nghề nghiệp cho sinh viên muốn theo ngành data engineering.

## **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

#### **- ETL tự động hóa cho nền tảng tài chính (ETL Developer, FinTrack) 2021**

Tự động hoá quá trình tổng hợp báo cáo từ nhiều hệ thống tài chính nội bộ.

- + Viết ETL pipeline bằng Python và Pandas

- + Tích hợp với hệ thống email để gửi báo cáo tự động hàng tuần

- + Quản lý lịch ETL với crontab và giám sát bằng Prometheus

