



## KỸ SƯ DỮ LIỆU

ĐẶNG HÙNG BÌNH

### MỤC TIÊU NGHỀ NGHIỆP

Tôi mong muốn nâng cao kỹ năng xử lý dữ liệu streaming sử dụng Kafka và Spark Structured Streaming để phục vụ các bài toán real-time như giám sát hành vi người dùng hoặc phát hiện gian lận.

### THÔNG TIN CÁ NHÂN

17/09/1980

Hà Nội

thaolinh252512@gmail.com

0917586578

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

- Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội - Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã

### KỸ NĂNG

- Google Cloud Platform (GCP)

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

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

### SỞ THÍCH

- Tham gia hội thảo công nghệ
- Đọc sách
- Học ngoại ngữ
- Tập gym

+ 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 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

+ 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

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

## CHỨNG CHỈ

- **2020** - AWS Certified Data Analytics – Specialty

## HOẠT ĐỘNG

- 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.

#### **- Thành viên nhóm kỹ thuật dữ liệu tại Dự án E-Government (2023)**

- + Thiết kế hệ thống thu thập dữ liệu hành chính từ nhiều bộ ngành.

- + Thiết lập hệ thống kiểm tra chất lượng dữ liệu tự động.

- + Triển khai pipeline đồng bộ dữ liệu hằng ngày với độ trễ thấp.

#### **- 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.

#### **- 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.

**- 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