



## KỸ SƯ DỮ LIỆU

### ĐỒ BÌNH HIỂU

### 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

02/04/1980

Hà Nội

thaolinhh252512@gmail.com

0996048596

www.website.com

## HỌC VẤN

- Toán - Tin ứng dụng tại Đại học Kinh tế Quốc dân - Công nghệ thông tin tại Đại học FPT

## KỸ NĂNG

- Google Cloud Platform (GCP)

- Data Lake Architecture

## SỞ THÍCH

## 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 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 TỶ

- Thử nghiệm công nghệ mới
- Nấu ăn
- Đọc sách

## KHO THEO THỜI GIAN THỰC

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

## NGƯỜI GIỚI THIỆU

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

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

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

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

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

+ 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

- **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

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

- **2022** - Top 3 nhân viên cải tiến hệ thống ETL tại SmartRetail

## CHỨNG CHỈ

- **2022** - Big Data Engineering with Apache Spark – edX (BerkeleyX)

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

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

Academy

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

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

## HOẠT ĐỘNG

### - 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.

### - 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.

### - 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.

### - 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 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.

## **DỰ ÁN**

### **- 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

