



## KỸ SƯ DỮ LIỆU

ĐẶNG TÙNG KHANH

### 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

16/03/1992

Hà Nội

thaolinhh252512@gmail.com

0724821957

www.website.com

### HỌC VẤN

- Công nghệ thông tin tại Đại học FPT

### KỸ NĂNG

- Snowflake

### SỞ THÍCH

- Tham gia hội thảo công nghệ

- Thiết kế sản phẩm cá nhân

### 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

### DANH HIỆU VÀ GIẢI THƯỞNG

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

## NGƯỜI GIỚI THIỆU

- Bà Trần Thị Mai (Technical Manager - Công ty CloudData) -

mai.tran@cloudata.vn - 0933666777

- Bà Nguyễn Thị Lệ (Analytics Lead - Công ty InsightTech) -

le.nguyen@insighttech.vn - 0966999888

- Bà Đỗ Ngọc Trang (Senior Data Engineer - Công ty SmartRetail) -

trang.do@smartretail.vn - 0922333444

## CHỨNG CHỈ

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

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

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

## HOẠT ĐỘNG

- **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.

- **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.

- **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 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.

## **DỰ ÁN**

#### **- 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

#### **- 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