



KỸ SƯ DỮ LIỆU

NGÔ THÀNH TÚ

MỤC TIÊU NGHỀ NGHIỆP

Tôi đặt mục tiêu thành thạo việc xây dựng kiến trúc dữ liệu hiện đại trên nền tảng cloud như AWS, GCP hoặc Azure, đặc biệt tập trung vào việc tối ưu hóa chi phí và hiệu suất lưu trữ.

THÔNG TIN CÁ NHÂN

01/02/1988

Hà Nội

thaolinh252512@gmail.com

0718222732

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

- Công nghệ thông tin tại Đại học FPT
- Kỹ thuật phần mềm tại Học viện Công nghệ Bưu chính Viễn thông

KỸ NĂNG

- dbt (Data Build Tool)

SỞ THÍCH

- Sưu tầm sách lập trình

DANH HIỆU VÀ GIẢI THƯỞNG

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

- Tham gia cộng đồng lập trình
- Thiết kế sản phẩm cá nhân
- Du lịch

NGƯỜI GIỚI THIỆU

- Ông Phan Thành Tâm (Big Data Architect - Công ty LogiData) - tam.phan@logidata.vn - 0955111222
- Bà Nguyễn Thị Lệ (Analytics Lead - Công ty InsightTech) - le.nguyen@insighttech.vn - 0966999888
- Ông Nguyễn Quốc Huy (Data Platform Lead - Công ty DataBridge) - huy.nguyen@databridge.vn - 0905123456

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

CHỨNG CHỈ

- **2020** - Google Cloud Professional Data Engineer Certificate
- **2022** - Big Data Engineering with Apache Spark - edX (BerkeleyX)
- **2023** - Modern Data Engineering with dbt - dbt Labs

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.

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

- 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 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

