



## KỸ SƯ DỮ LIỆU

### TRẦN HẠNH THẢO

#### MỤC TIÊU NGHỀ NGHIỆP

Tôi muốn làm việc với các nhóm Data Scientist và BI Analyst để đảm bảo dữ liệu được xử lý đúng định dạng, đúng ngữ nghĩa và cung cấp kịp thời cho phân tích chuyên sâu.

#### THÔNG TIN CÁ NHÂN

20/05/1980

Hà Nội

thaolinhh252512@gmail.com

0937038764

www.website.com

#### HỌC VẤN

- Khoa học máy tính tại Đại học Quốc  
tế - ĐHQG TP.HCM

#### KỸ NĂNG

- NoSQL (MongoDB, Cassandra)
- Apache Kafka
- Data Quality Validation (Great Expectations)
- Data Lake Architecture
- Apache Spark

#### KINH NGHIỆM LÀM VIỆC

- **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

- **2021** - Bằng khen về bảo mật và chất lượng dữ liệu - Công ty  
InfosecData

## SỞ THÍCH

- Du lịch
- Chơi đàn guitar
- Đi bộ đường dài

## NGƯỜI GIỚI THIỆU

- Bà Đỗ Ngọc Trang (Senior Data Engineer - Công ty SmartRetail) - trang.do@smartretail.vn - 0922333444
- Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666
- Ông Vũ Minh Đức (CTO - Công ty DataStreamX) - duc.vu@datastreamx.vn - 0977333444
- Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123

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

- **2021** - Giải thưởng 'Dự án phân tích dữ liệu xuất sắc' tại FinData

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

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

## CHỨNG CHỈ

- **2020** - Google Cloud Professional Data Engineer Certificate
- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka - Coursera
- **2020** - AWS Certified Data Analytics - Specialty
- **2023** - Designing Data-Intensive Applications - O'Reilly Certification Program

## HOẠT ĐỘNG

- **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.

## 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

### **- 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

### **- 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

### **- Realtime Analytics cho hệ thống bán lẻ toàn quốc (Big Data Engineer, Retail360) 2023**

Xây dựng hệ thống xử lý dữ liệu bán hàng theo thời gian thực để hỗ trợ ra quyết định tức thì.

- + Sử dụng Kafka để thu thập dữ liệu từ các chi nhánh toàn quốc
- + Xử lý dữ liệu streaming bằng Spark Structured Streaming

- + Gửi dữ liệu về Redshift và hiển thị trên dashboard BI

**- 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