



## KỸ SƯ DỮ LIỆU

NGUYỄN HẠNH NGÂN

### 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

11/05/1986

Hà Nội

thaolinh252512@gmail.com

0881873037

www.website.com

### HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại  
Học viện Kỹ thuật Mật mã - Cơ sở dữ  
liệu và khai phá dữ liệu tại Đại học  
Khoa học Tự nhiên - ĐHQG Hà Nội

### KỸ NĂNG

- Apache Kafka

### SỞ THÍCH

### KINH NGHIỆM LÀM VIỆC

- **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

- **CLOUD DATA ENGINEER** TẠI SMARTBANK (2022-2023)

+ THIẾT KẾ KIẾN TRÚC DỮ LIỆU TRÊN AWS VỚI S3,

- Tham gia cộng đồng lập trình
- Du lịch
- Chụp ảnh
- Nghe nhạc

## NGƯỜI GIỚI THIỆU

- Ông Nguyễn Quốc Huy (Data Platform Lead - Công ty DataBridge) - huy.nguyen@databridge.vn - 0905123456
- Ông Hồ Trung Kiên (Engineering Manager - Công ty ETLTech) - kien.ho@etltech.vn - 0933888777
- Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123
- Bà Trần Thị Mai (Technical Manager - Công ty CloudData) - mai.tran@cloudata.vn - 0933666777

## 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

- **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

## DANH HIỆU VÀ GIẢI THƯỞNG

- **2021** - Bằng khen về bảo mật và chất lượng dữ liệu - Công ty InfosecData
- **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge

## CHỨNG CHỈ

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

- **2023** - Designing Data-Intensive Applications – O'Reilly Certification Program

- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera

## HOẠT ĐỘNG

- **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.

## DỰ ÁN

- **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

- **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

**- 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

