



## KỸ SƯ DỮ LIỆU

BÙI YẾN THU

### MỤC TIÊU NGHỀ NGHIỆP

Tôi muốn chia sẻ kiến thức qua việc viết blog, tổ chức hội thảo kỹ thuật hoặc đóng góp tài liệu chính thức để thúc đẩy cộng đồng Data Engineering trong nước phát triển mạnh mẽ.

### THÔNG TIN CÁ NHÂN

28/04/1993

Hà Nội

thaolinh252512@gmail.com

0971590136

www.website.com

### HỌC VẤN

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

### KỸ NĂNG

- Airflow

### SỞ THÍCH

- Nghe nhạc

### 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

- **2022** - Vinh danh 'Nhân sự triển khai hệ thống streaming  
real-time đầu tiên' tại LogiTech

- Du lịch
- Tham gia hackathon
- Thử nghiệm công nghệ mới

## NGƯỜI GIỚI THIỆU

- Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666
- Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123
- Bà Nguyễn Thị Lệ (Analytics Lead - Công ty InsightTech) - le.nguyen@insighttech.vn - 0966999888

- **2023** - Bằng khen về tối ưu chi phí lưu trữ dữ liệu cloud - Công ty CloudData

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

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

## CHỨNG CHỈ

- **2023** - Modern Data Engineering with dbt - dbt Labs
- **2020** - AWS Certified Data Analytics - Specialty
- **2020** - Google Cloud Professional Data Engineer Certificate
- **2021** - SQL for Data Engineering - Datacamp
- **2022** - Data Engineering on Google Cloud Platform - Coursera

## 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.

## 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

#### **- 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

**- 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