



## KỸ SƯ DỮ LIỆU

PHẠM DŨNG TRÍ

### MỤC TIÊU NGHỀ NGHIỆP

Tôi kỳ vọng học hỏi sâu hơn về kiến trúc dữ liệu phân tán (distributed data architecture) và các nguyên lý CAP, consistency, replication, để đảm bảo tính sẵn sàng cao.

### THÔNG TIN CÁ NHÂN

15/10/1984

Hà Nội

thaolinh252512@gmail.com

0387461145

www.website.com

### HỌC VẤN

- Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội - Công nghệ thông tin tại Đại học FPT

### KỸ NĂNG

- Google Cloud Platform (GCP)
- Apache Spark
- Monitoring với Prometheus/Grafana

### KINH NGHIỆM LÀM VIỆC

- **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

- **DATA ENGINEER** TẠI CÔNG TY DATAFLOW VIỆT NAM (2021-2023)

## SỞ THÍCH

- Đi bộ đường dài
- Thể thao
- Chơi đàn guitar
- Tham gia hackathon

+ 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

## NGƯỜI GIỚI THIỆU

- Ông Nguyễn Quốc Huy (Data Platform  
Lead - Công ty DataBridge) -  
huy.nguyen@databridge.vn -  
0905123456

- 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

- Bà Nguyễn Thị Lệ (Analytics Lead -  
Công ty InsightTech) -  
le.nguyen@insighttech.vn - 0966999888

## DANH HIỆU VÀ GIẢI THƯỞNG

- **2020** - Giải thưởng 'Pipeline ổn định nhất năm' tại TechFlow

## CHỨNG CHỈ

- **2021** - SQL for Data Engineering - Datacamp

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

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

- **2022** - Data Engineering on Google Cloud Platform -  
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

