



## KỸ SƯ DỮ LIỆU

NGÔ THU DƯƠNG

### MỤC TIÊU NGHỀ NGHIỆP

Tôi kỳ vọng được làm việc trong các dự án về xây dựng data lake và data warehouse quy mô lớn, áp dụng Redshift, BigQuery hoặc Snowflake để phục vụ báo cáo phân tích đa chiều.

### THÔNG TIN CÁ NHÂN

22/12/1988

Hà Nội

thaolinh252512@gmail.com

0859343954

www.website.com

### HỌC VẤN

- Khoa học máy tính tại Đại học Quốc tế - ĐHQG TP.HCM - Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã

### KỸ NĂNG

- PostgreSQL

### SỞ THÍCH

### KINH NGHIỆM LÀM VIỆC

- **JUNIOR DATA ENGINEER** TẠI RETAILTECH (2020-2021)

+ PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỔN 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,

- Tập gym

- Đọc sách

- Thể thao

- Học ngoại ngữ

- Thử nghiệm công nghệ mới

## 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

## NGƯỜI GIỚI THIỆU

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

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

- Bà Phạm Thanh Hương (Data  
Operations Manager - Công ty  
ReportPro) - huong.pham@reportpro.vn  
- 0912111222

+ THIẾT KẾ VÀ TRIỂN KHAI PIPELINE ETL SỬ DỤNG  
APACHE AIRFLOW

- Bà Đỗ Ngọc Trang (Senior Data  
Engineer - Công ty SmartRetail) -  
trang.do@smartretail.vn - 0922333444

+ TÍCH HỢP DỮ LIỆU TỪ NHIỀU NGUỒN NHƯ MYSQL,  
GOOGLE ANALYTICS VÀ API BÊN THỨ BA

- Ông Hồ Trung Kiên (Engineering  
Manager - Công ty ETLTech) -  
kien.ho@etltech.vn - 0933888777

+ TỐI ƯU HOÁ QUY TRÌNH XỬ LÝ BẰNG SPARK TRÊN  
GCP DATAPROC

- Ông Phan Thành Tâm (Big Data  
Architect - Công ty LogiData) -  
tam.phan@logidata.vn - 0955111222

## DANH HIỆU VÀ GIẢI THƯỞNG

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

## CHỨNG CHỈ

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

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

- **2020** - Google Cloud Professional Data Engineer Certificate

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

## HOẠT ĐỘNG

- **Tình nguyện viên hỗ trợ nghiên cứu học thuật tại Khoa Công nghệ Thông tin (2020)**

- + Chuẩn hóa dữ liệu nghiên cứu sinh viên từ nhiều năm học.
- + Phân tích và tổng hợp xu hướng đăng ký môn học.
- + Viết báo cáo kỹ thuật trình bày cho giảng viên hướng dẫn.

- **Diễn giả hội thảo 'Big Data Architecture' tại Data Talks Vietnam (2022)**

- + Trình bày kiến trúc hệ thống thu thập và xử lý dữ liệu đa nguồn.
- + Phân tích ưu nhược điểm của Data Warehouse vs Data Lakehouse.
- + Giới thiệu các công cụ phổ biến như Airflow, dbt, Snowflake.

## DỰ ÁN

- **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

