



KỸ SƯ DỮ LIỆU

LÊ TRUNG TRÍ

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

07/08/1997

Hà Nội

thaolinhh252512@gmail.com

0367813878

www.website.com

HỌC VẤN

- Kỹ thuật phần mềm tại Học viện Công nghệ Bưu chính Viễn thông - Công nghệ thông tin tại Đại học FPT

KỸ NĂNG

- Data Quality Validation (Great Expectations)

- Airflow

- AWS (S3, EMR, Glue)

- 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)

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

SỞ THÍCH

- Trồng cây
- Viết blog kỹ thuật
- Tham gia hội thảo công nghệ
- Chụp ảnh

+ 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

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

- Ông Vũ Minh Đức (CTO - Công ty
DataStreamX) - duc.vu@datastreamx.vn
- 0977333444

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

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

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

DANH HIỆU VÀ GIẢI THƯỞNG

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

CHỨNG CHỈ

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

- **2020** - AWS Certified Data Analytics – Specialty
- **2021** - SQL for Data Engineering – Datacamp
- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera

HOẠT ĐỘNG

- 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

- 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

