



KỸ SƯ DỮ LIỆU

NGUYỄN TÙNG PHƯƠNG

MỤC TIÊU NGHỀ NGHIỆP

Tôi định hướng trở thành một Kỹ sư Dữ liệu chuyên nghiệp, có khả năng xây dựng các pipeline xử lý dữ liệu mạnh mẽ và tối ưu nhằm phục vụ các hệ thống phân tích và học máy trong doanh nghiệp.

THÔNG TIN CÁ NHÂN

16/09/1993

Hà Nội

thaolinh252512@gmail.com

0373075505

www.website.com

HỌC VẤN

- Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM - Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã

KỸ NĂNG

- Docker

- Data Quality Validation (Great Expectations)

- BigQuery

- CI/CD cho Data Pipeline

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

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

+ PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỶ

- dbt (Data Build Tool)

KHO THEO THỜI GIAN THỰC

+ XỬ LÝ DỮ LIỆU STREAMING TỪ KAFKA VÀ LƯU TRỮ VÀO BIGQUERY

SỞ THÍCH

- Viết blog kỹ thuật

- Xem phim khoa học viễn tưởng

- Trồng cây

+ XÂY DỰNG CÁC VIEW LOGIC TRÊN DBT PHỤC VỤ BI DASHBOARD

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

NGƯỜI GIỚI THIỆU

- Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666

+ 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

- **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge

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

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

- **2022** - Vinh danh 'Nhân sự triển khai hệ thống streaming

real-time đầu tiên' tại LogiTech

CHỨNG CHỈ

- **2023** - Designing Data-Intensive Applications – O'Reilly Certification Program
- **2022** - Big Data Engineering with Apache Spark – edX (BerkeleyX)
- **2020** - AWS Certified Data Analytics – Specialty
- **2020** - Google Cloud Professional Data Engineer Certificate

HOẠT ĐỘNG

- Thành viên nhóm xây dựng pipeline dữ liệu tại Phòng Lab Dữ liệu - Đại học Bách khoa (2021 - 2023)

- + Phát triển luồng dữ liệu ETL sử dụng Apache Airflow.
- + Thiết lập hệ thống ingestion dữ liệu từ API và cơ sở dữ liệu nguồn.
- + Phối hợp với nhóm phân tích dữ liệu để tối ưu schema lưu trữ.

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

bảng Kibana