



KỸ SƯ DỮ LIỆU

NGUYỄN HÀ TRÍ

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

09/05/1997

Hà Nội

thaolinh252512@gmail.com

0909672111

www.website.com

HỌC VẤN

- Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM - Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội

KỸ NĂNG

- Google Cloud Platform (GCP)
- Monitoring với Prometheus/Grafana
- BigQuery
- ETL/ELT Design

KINH NGHIỆM LÀM VIỆC

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

+ PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỒN 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,

GLUE VÀ REDSHIFT

SỞ THÍCH

- Tham gia hackathon

- Du lịch

+ 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

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

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

- Bà Trần Thị Mai (Technical Manager -
Công ty CloudData) -
mai.tran@cloudata.vn - 0933666777

- **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

DANH HIỆU VÀ GIẢI THƯỞNG

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

CHỨNG CHỈ

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

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

Coursera

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

DỰ ÁN

- **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

- **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

**- 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

- 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

