



KỸ SƯ DỮ LIỆU

ĐỒ TRÍ TÙNG

MỤC TIÊU NGHỀ NGHIỆP

Tôi mong muốn trở thành nhân tố chủ lực trong đội ngũ dữ liệu, đảm bảo dữ liệu đầu vào được làm sạch, chuẩn hóa, kiểm tra chất lượng (data validation) và tuân thủ tiêu chuẩn bảo mật.

THÔNG TIN CÁ NHÂN

28/05/1996

Hà Nội

thaolinhh252512@gmail.com

0749714633

www.website.com

HỌC VẤN

- Cơ sở dữ liệu và khai phá dữ liệu tại
Đại học Khoa học Tự nhiên – ĐHQG Hà
Nội

KỸ NĂNG

- Monitoring với Prometheus/Grafana

SỞ THÍCH

- Thiết kế sản phẩm cá nhân

KINH NGHIỆM LÀM VIỆC

- **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 TỬ HỆ THỐNG CORE

+ THAM GIA ĐÁNH GIÁ BẢO MẬT VÀ CHUẨN HÓA
SCHEMA DỮ LIỆU TÀI CHÍNH

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

- Nghe nhạc

- Du lịch

- Thể thao

+ 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

NGƯỜI GIỚI THIỆU

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

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

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

- **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 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

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

CHỨNG CHỈ

- **2022** - Big Data Engineering with Apache Spark – edX (BerkeleyX)
- **2020** - AWS Certified Data Analytics – Specialty
- **2021** - Streaming Data with Apache Kafka – Confluent Academy
- **2023** - Designing Data-Intensive Applications – O'Reilly Certification Program

HOẠT ĐỘNG

- **Thành viên nhóm tư vấn kiến trúc dữ liệu tại Công ty SmartData (2021 - 2022)**

+ Đề xuất mô hình lưu trữ dữ liệu dạng columnar cho truy vấn phân tích nhanh.

+ Tham gia thiết kế cấu trúc bảng dữ liệu trong hệ thống warehouse.

+ Xây dựng best practice về đặt tên và phân vùng dữ liệu.

- 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.

- Mentor khóa học nền tảng kỹ sư dữ liệu tại Trung tâm Đào tạo CloudTech (2023)

- + Hướng dẫn học viên triển khai hệ thống ingest dữ liệu bằng Kafka.
- + Đánh giá bài tập về xử lý dữ liệu song song với Spark.
- + Tư vấn về định hướng nghề nghiệp cho sinh viên muốn theo ngành data engineering.

- Tình nguyện viên hỗ trợ tổ chức Data Day tại Vietnam Data Community (2022)

- + Thiết lập demo pipeline dữ liệu thời gian thực bằng Kafka.
- + Hỗ trợ phần trình diễn luồng dữ liệu streaming giữa các node Spark.
- + Giới thiệu cấu trúc cơ bản của hệ thống Data Lake.

DỰ ÁN

- 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

- 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

- 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

- 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