



KỸ SƯ DỮ LIỆU

TRẦN HÀ KHÁNH

MỤC TIÊU NGHỀ NGHIỆP

Tôi mong muốn đóng góp vào việc xây dựng hệ thống phân quyền và bảo mật dữ liệu mạnh mẽ, tuân thủ các tiêu chuẩn như GDPR, ISO 27001 trong vận hành hệ thống dữ liệu.

THÔNG TIN CÁ NHÂN

03/06/1987

Hà Nội

thaolinh252512@gmail.com

0378733973

www.website.com

HỌC VẤN

- Công nghệ thông tin tại Đại học FPT

KỸ NĂNG

- BigQuery

SỞ THÍCH

- Thử nghiệm công nghệ mới

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

DANH HIỆU VÀ GIẢI THƯỞNG

- **2022** - Vinh danh 'Nhân sự triển khai hệ thống streaming real-time đầu tiên' tại LogiTech

NGƯỜI GIỚI THIỆU

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

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

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

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

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

CHỨNG CHỈ

- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera

- **2021** - Streaming Data with Apache Kafka – Confluent Academy

HOẠT ĐỘNG

- **Thành viên nhóm phát triển hệ thống phân tích log tại Công ty SecureLog (2021)**

+ Xây dựng pipeline thu thập và phân tích log từ nhiều nguồn hệ thống.

+ Triển khai hệ thống cảnh báo bất thường dựa trên mô hình thống kê.

+ Sử dụng ELK stack để trực quan hóa dữ liệu log.

- **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.

- **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.

- Người viết blog kỹ thuật về hệ thống dữ liệu tại Blog cá nhân datapi.dev (2022 - nay)

- + Chia sẻ kiến thức về thiết kế hệ thống ETL quy mô lớn.
- + So sánh các giải pháp lưu trữ dữ liệu: PostgreSQL, Redshift, BigQuery.
- + Viết hướng dẫn cấu hình Airflow và orchestration jobs theo lịch định kỳ.

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

- 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

