



KỸ SƯ DỮ LIỆU

BÙI NGÀ TRÍ

MỤC TIÊU NGHỀ NGHIỆP

Tôi mong muốn nâng cao kỹ năng xử lý dữ liệu streaming sử dụng Kafka và Spark Structured Streaming để phục vụ các bài toán real-time như giám sát hành vi người dùng hoặc phát hiện gian lận.

THÔNG TIN CÁ NHÂN

07/02/1984

Hà Nội

thaolinh252512@gmail.com

0861651068

www.website.com

HỌC VẤN

- Khoa học dữ liệu tại Đại học Công nghệ Thông tin - ĐHQG TP.HCM - Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã

KỸ NĂNG

- Data Quality Validation (Great Expectations)

- dbt (Data Build Tool)

- Docker

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

DANH HIỆU VÀ GIẢI THƯỞNG

- **2022** - Top 3 nhân viên cải tiến hệ thống ETL tại SmartRetail

SỞ THÍCH

- Du lịch
- Tập gym
- Đọc sách
- Chơi đàn guitar
- Tham gia cộng đồng lập trình

NGƯỜI GIỚI THIỆU

- Ông Hồ Trung Kiên (Engineering Manager - Công ty ETLTech) - kien.ho@etltech.vn - 0933888777
- Bà Phạm Thanh Hương (Data Operations Manager - Công ty ReportPro) - huong.pham@reportpro.vn - 0912111222

- **2020** - Giải thưởng 'Pipeline Ổn định nhất năm' tại TechFlow
- **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge
- **2023** - Giải thưởng 'Kiến trúc dữ liệu sáng tạo' do CTO trao tặng tại DataWare
- **2021** - Kỹ sư dữ liệu xuất sắc quý I tại Công ty DataBridge

CHỨNG CHỈ

- **2021** - SQL for Data Engineering – Datacamp

HOẠT ĐỘNG

- 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.

- 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.

- 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 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

