



KỸ SƯ DỮ LIỆU

NGÔ DƯƠNG 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

21/07/1984

Hà Nội

thaolinh252512@gmail.com

0851139938

www.website.com

HỌC VẤN

- Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM - Cơ sở dữ liệu và khai phá dữ liệu tại Đại học Khoa học Tự nhiên - ĐHQG Hà Nội

KỸ NĂNG

- Airflow

- SQL

- Python

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 TỪ HỆ THỐNG CORE

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

- **BIG DATA ENGINEER** TẠI LOGIAI (2020-2022)

+ XÂY DỰNG PIPELINE THU THẬP VÀ XỬ LÝ DỮ LIỆU

VẬN CHUYỂN TỪ THIẾT BỊ IOT

SỞ THÍCH

- Tham gia hội thảo công nghệ

+ SỬ DỤNG SPARK STREAMING ĐỂ XỬ LÝ DỮ LIỆU REAL-TIME

NGƯỜI GIỚI THIỆU

- Ông Phan Thành Tâm (Big Data Architect - Công ty LogiData) - tam.phan@logidata.vn - 0955111222

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

- 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

+ KẾT NỐI HỆ THỐNG VÀO ELASTICSEARCH VÀ PHÁT TRIỂN DASHBOARD GIÁM SÁT

DANH HIỆU VÀ GIẢI THƯỞNG

- **2023** - Giải thưởng 'Kiến trúc dữ liệu sáng tạo' do CTO trao tặng tại DataWare

- **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)

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

HOẠT ĐỘNG

- **Cộng tác viên dự án dữ liệu mở tại OpenData Vietnam (2020)**

+ Thu thập và chuẩn hóa dữ liệu dân số từ các tỉnh thành.

+ Tích hợp dữ liệu vào hệ thống cơ sở dữ liệu MongoDB phục vụ phân tích.

+ Tạo dashboard đơn giản để trực quan hóa dữ liệu bằng

Plotly Dash.

- 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.

- 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.

- 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.

- 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.

DỰ ÁN

- 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

