



KỸ SƯ DỮ LIỆU

ĐẶNG KHÁNH VIỆT

MỤC TIÊU NGHỀ NGHIỆP

Tôi hướng tới khả năng xử lý dữ liệu phi cấu trúc như log, JSON, XML và tích hợp với hệ thống NoSQL như MongoDB, Elasticsearch để lưu trữ linh hoạt và truy xuất hiệu quả.

THÔNG TIN CÁ NHÂN

01/07/1987

Hà Nội

thaolinh252512@gmail.com

0319647268

www.website.com

HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại
Học viện Kỹ thuật Mật mã - Công nghệ
thông tin tại Đại học FPT

KỸ NĂNG

- CI/CD cho Data Pipeline

SỞ THÍCH

- Du lịch

KINH NGHIỆM LÀM VIỆC

- **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

- **CLOUD DATA ENGINEER** TẠI SMARTBANK (2022-2023)

+ THIẾT KẾ KIẾN TRÚC DỮ LIỆU TRÊN AWS VỚI S3,

- Tham gia hội thảo công nghệ
- Xem phim khoa học viễn tưởng
- Đọc sách

NGƯỜI GIỚI THIỆU

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

- Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666

- Ông Vũ Minh Đức (CTO - Công ty DataStreamX) - duc.vu@datastreamx.vn - 0977333444

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

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

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

+ 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

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

- **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

- **2021** - Giải thưởng 'Dự án phân tích dữ liệu xuất sắc' tại FinData
- **2022** - Vinh danh 'Nhân sự triển khai hệ thống streaming real-time đầu tiên' tại LogiTech
- **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge
- **2021** - Bằng khen về bảo mật và chất lượng dữ liệu – Công ty InfosecData
- **2022** - Top nhân viên cải tiến năng suất truy vấn dữ liệu SQL tại ReportPro

CHỨNG CHỈ

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

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

