



KỸ SƯ DỮ LIỆU

ĐẶNG SƠN THU

MỤC TIÊU NGHỀ NGHIỆP

Tôi muốn trở thành chuyên gia về tối ưu hóa truy vấn SQL, sử dụng chỉ mục thông minh và partitioning trong các hệ thống dữ liệu lớn nhằm rút ngắn thời gian phản hồi truy vấn.

THÔNG TIN CÁ NHÂN

10/04/1991

Hà Nội

thaolinh252512@gmail.com

0868392612

www.website.com

HỌC VẤN

- Công nghệ thông tin tại Đại học FPT

KỸ NĂNG

- Apache Spark

- NoSQL (MongoDB, Cassandra)

- dbt (Data Build Tool)

- Apache Kafka

- ETL/ELT Design

KINH NGHIỆM LÀM VIỆC

- **ETL DEVELOPER** TẠI FINDATA (2019-2021)

+ PHÁT TRIỂN VÀ BẢO TRÌ LUỒNG DỮ LIỆU ETL SỬ DỤNG TALEND VÀ PYTHON

+ VIẾT SCRIPT KIỂM TRA CHẤT LƯỢNG DỮ LIỆU VÀ CẢNH BÁO KHI DỮ LIỆU SAI LỆCH

+ TÍCH HỢP HỆ THỐNG PHÂN QUYỀN DỮ LIỆU THEO VAI TRÒ NGƯỜI DÙNG

- **JUNIOR DATA ENGINEER** TẠI RETAILTECH (2020-2021)

+ PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỶ

KHO THEO THỜI GIAN THỰC

SỞ THÍCH

- Chơi đàn guitar
- Học ngoại ngữ
- Đọc sách
- Nghe nhạc

+ XỬ LÝ DỮ LIỆU STREAMING TỪ KAFKA VÀ LƯU TRỮ VÀO BIGQUERY

+ XÂY DỰNG CÁC VIEW LOGIC TRÊN DBT PHỤC VỤ BI DASHBOARD

NGƯỜI GIỚI THIỆU

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

- **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

DANH HIỆU VÀ GIẢI THƯỞNG

- **2021** - Kỹ sư dữ liệu xuất sắc quý I tại Công ty DataBridge

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

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

CHỨNG CHỈ

- **2021** - SQL for Data Engineering – Datacamp
- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera
- **2021** - Streaming Data with Apache Kafka – Confluent Academy
- **2023** - Designing Data-Intensive Applications – O'Reilly Certification Program

HOẠT ĐỘNG

- 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 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.

- 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.

- 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.

- 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

- 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

- 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