



KỸ SƯ DỮ LIỆU

TRẦN LAN SÓN

MỤC TIÊU NGHỀ NGHIỆP

Tôi định hướng phát triển các kỹ năng DevOps cho dữ liệu như CI/CD pipeline cho hệ thống ETL, quản lý version dữ liệu bằng DVC, và monitor hệ thống bằng Prometheus/Grafana.

THÔNG TIN CÁ NHÂN

08/08/2000

Hà Nội

thaolinh252512@gmail.com

0321262878

www.website.com

HỌC VẤN

- Khoa học dữ liệu tại Đại học Công nghệ Thông tin - ĐHQG TP.HCM - Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM

KỸ NĂNG

- BigQuery

SỞ THÍCH

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

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

+ PHÁT TRIỂN VÀ BẢO TRÌ LUỒNG DỮ LIỆU ETL SỬ

- Sưu tầm sách lập trình

- Tập gym

- Nghe nhạc

- Viết blog kỹ thuật

NGƯỜI GIỚI THIỆU

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

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

DANH HIỆU VÀ GIẢI THƯỞNG

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

- **2023** - Bằng khen về tối ưu chi phí lưu trữ dữ liệu cloud - Công ty CloudData

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

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

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

CHỨNG CHỈ

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

- **2020** - AWS Certified Data Analytics - Specialty

HOẠT ĐỘNG

- 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.

- 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 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.

DỰ ÁN

- 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

- 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

- 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

- 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

- 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