



KỸ SƯ DỮ LIỆU

ĐỖ DƯƠNG LAN

MỤC TIÊU NGHỀ NGHIỆP

Tôi đặt mục tiêu thành thạo việc xây dựng kiến trúc dữ liệu hiện đại trên nền tảng cloud như AWS, GCP hoặc Azure, đặc biệt tập trung vào việc tối ưu hóa chi phí và hiệu suất lưu trữ.

THÔNG TIN CÁ NHÂN

26/10/1997

Hà Nội

thaolinh252512@gmail.com

0339330827

www.website.com

HỌC VẤN

- Khoa học máy tính tại Đại học Quốc tế - ĐHQG TP.HCM - Khoa học dữ liệu tại Đại học Công nghệ Thông tin - ĐHQG TP.HCM

KỸ NĂNG

- Monitoring với Prometheus/Grafana
- Data Quality Validation (Great Expectations)
- Docker

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

SỞ THÍCH

- Nấu ăn
- Tham gia hội thảo công nghệ
- Trồng cây
- Sưu tầm sách lập trình
- Du lịch

NGƯỜI GIỚI THIỆU

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

- Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123

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

- Ông Hồ Trung Kiên (Engineering Manager - Công ty ETLTech) - kien.ho@etltech.vn - 0933888777

+ 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

- **2020** - Giải thưởng 'Pipeline Ổn định nhất năm' tại TechFlow

CHỨNG CHỈ

- **2023** - Designing Data-Intensive Applications - O'Reilly Certification Program

- **2020** - Google Cloud Professional Data Engineer Certificate

- **2022** - Big Data Engineering with Apache Spark - edX (BerkeleyX)

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

- **2023** - Modern Data Engineering with dbt - dbt Labs

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 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 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.

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

- 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

