



KỸ SƯ DỮ LIỆU

NGÔ NGÀ BÌNH

MỤC TIÊU NGHỀ NGHIỆP

Tôi mong muốn phát triển hệ thống lưu trữ dữ liệu dạng time-series cho các ứng dụng theo dõi thiết bị IoT, phân tích logs và hệ thống sự kiện theo thời gian.

THÔNG TIN CÁ NHÂN

14/10/1993

Hà Nội

thaolinh252512@gmail.com

0304678697

www.website.com

HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại
Học viện Kỹ thuật Mật mã - Cơ sở dữ
liệu và khai phá dữ liệu tại Đại học
Khoa học Tự nhiên - ĐHQG Hà Nội

KỸ NĂNG

- Monitoring với Prometheus/Grafana

- SQL

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

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

SỞ THÍCH

- Đi bộ đường dài
- Nghe nhạc
- Thể thao

NGƯỜI GIỚI THIỆU

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

+ 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

- **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,
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

- **2020** - Giải thưởng 'Pipeline Ổn định nhất năm' tại TechFlow
- **2023** - Bằng khen về tối ưu chi phí lưu trữ dữ liệu cloud – Công ty CloudData

CHỨNG CHỈ

- **2021** - Streaming Data with Apache Kafka – Confluent Academy
- **2023** - Modern Data Engineering with dbt – dbt Labs
- **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 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.

- Người viết blog kỹ thuật về hệ thống dữ liệu tại Blog cá nhân datapi.dev (2022 - nay)

- + Chia sẻ kiến thức về thiết kế hệ thống ETL quy mô lớn.

- + So sánh các giải pháp lưu trữ dữ liệu: PostgreSQL, Redshift, BigQuery.

- + Viết hướng dẫn cấu hình Airflow và orchestration jobs theo lịch định kỳ.

- 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.

- 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

- 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