



KỸ SƯ DỮ LIỆU

VŨ LAN HÙNG

MỤC TIÊU NGHỀ NGHIỆP

Tôi đặt mục tiêu xây dựng hệ thống quản trị chất lượng dữ liệu (Data Quality Management) và giám sát tự động giúp cảnh báo sớm khi có dữ liệu bất thường hoặc thiếu hụt.

THÔNG TIN CÁ NHÂN

14/08/1999

Hà Nội

thaolinh252512@gmail.com

0982315898

www.website.com

HỌC VẤN

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

KỸ NĂNG

- Data Lake Architecture
- AWS (S3, EMR, Glue)
- Apache Spark
- Data Quality Validation (Great Expectations)

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

- **BIG DATA ENGINEER** TẠI LOGIAI (2020-2022)

+ XÂY DỰNG PIPELINE THU THẬP VÀ XỬ LÝ DỮ LIỆU

VẬN CHUYỂN TỪ THIẾT BỊ IOT

SỞ THÍCH

- Du lịch

+ SỬ DỤNG SPARK STREAMING ĐỂ XỬ LÝ DỮ LIỆU REAL-TIME

+ KẾT NỐI HỆ THỐNG VÀO ELASTICSEARCH VÀ PHÁT TRIỂN DASHBOARD GIÁM SÁT

NGƯỜI GIỚI THIỆU

- Bà Nguyễn Thị Lệ (Analytics Lead - Công ty InsightTech) -
le.nguyen@insighttech.vn - 0966999888

- Bà Trần Thị Mai (Technical Manager - Công ty CloudData) -
mai.tran@cloudata.vn - 0933666777

DANH HIỆU VÀ GIẢI THƯỞNG

- **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

CHỨNG CHỈ

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

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

- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka - Coursera

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

HOẠT ĐỘNG

- 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.

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

