



KỸ SƯ DỮ LIỆU

BÙI HÙNG NGÂN

MỤC TIÊU NGHỀ NGHIỆP

Tôi muốn làm việc với các nhóm Data Scientist và BI Analyst để đảm bảo dữ liệu được xử lý đúng định dạng, đúng ngữ nghĩa và cung cấp kịp thời cho phân tích chuyên sâu.

THÔNG TIN CÁ NHÂN

21/08/1995

Hà Nội

thaolinhh252512@gmail.com

0726548312

www.website.com

HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại
Học viện Kỹ thuật Mật mã

KỸ NĂNG

- SQL

- NoSQL (MongoDB, Cassandra)

- Python

- Snowflake

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

- Chơi đàn guitar

NGƯỜI GIỚI THIỆU

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

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

- Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) -
linh.le@cloudops.vn - 0944555666

- Ông Vũ Minh Đức (CTO - Công ty DataStreamX) - duc.vu@datastreamx.vn
- 0977333444

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

- **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

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

+ 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

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

- + PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỒN KHO THEO THỜI GIAN THỰC

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

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

DANH HIỆU VÀ GIẢI THƯỞNG

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

- **2022** - Top nhân viên cải tiến năng suất truy vấn dữ liệu SQL tại ReportPro

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

- **2021** - Giải thưởng 'Dự án phân tích dữ liệu xuất sắc' tại FinData

CHỨNG CHỈ

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

- **2022** - Data Engineering on Google Cloud Platform – Coursera

- **2021** - Streaming Data with Apache Kafka – Confluent Academy

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

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

HOẠT ĐỘNG

- **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.

- **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.

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