



KỸ SƯ DỮ LIỆU

HOÀNG NGỌC YẾN

MỤC TIÊU NGHỀ NGHIỆP

Tôi kỳ vọng được làm việc trong các dự án về xây dựng data lake và data warehouse quy mô lớn, áp dụng Redshift, BigQuery hoặc Snowflake để phục vụ báo cáo phân tích đa chiều.

THÔNG TIN CÁ NHÂN

19/03/1998

Hà Nội

thaolinh252512@gmail.com

0863781984

www.website.com

HỌC VẤN

- Phân tích và xử lý dữ liệu lớn tại Đại học Phenikaa

KỸ NĂNG

- Python

- Data Lake Architecture

KINH NGHIỆM LÀM VIỆC

- **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

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

- Tham gia cộng đồng lập trình

- Học ngoại ngữ

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

NGƯỜI GIỚI THIỆU

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

- Bà Phạm Thanh Hương (Data Operations Manager - Công ty ReportPro) - huong.pham@reportpro.vn - 0912111222

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

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

CHỨNG CHỈ

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

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

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

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

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

HOẠT ĐỘNG

- **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.

- **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.

- **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.

- 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

- 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

- 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

- 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

- 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

- 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