



KỸ SƯ DỮ LIỆU

NGÔ QUANG HIẾU

MỤC TIÊU NGHỀ NGHIỆP

Tôi hướng tới mục tiêu trong 3 năm tới trở thành một Senior Data Engineer, có khả năng thiết kế toàn bộ kiến trúc dữ liệu, dẫn dắt đội nhóm và đề xuất cải tiến cho chiến lược dữ liệu doanh nghiệp.

THÔNG TIN CÁ NHÂN

08/01/1985

Hà Nội

thaolinh252512@gmail.com

0723198115

www.website.com

HỌC VẤN

- Toán - Tin ứng dụng tại Đại học Kinh tế Quốc dân

KỸ NĂNG

- CI/CD cho Data Pipeline
- Data Quality Validation (Great Expectations)
- Data Lake Architecture

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

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

+ PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỶ

SỞ THÍCH

- Thiết kế sản phẩm cá nhân
- Tham gia hackathon
- Nghe nhạc
- Tham gia cộng đồng lập trình
- Viết blog kỹ thuật

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

NGƯỜI GIỚI THIỆU

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

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

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

- **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

- **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

- **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

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

- **2022** - Top 3 nhân viên cải tiến hệ thống ETL tại SmartRetail

CHỨNG CHỈ

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

HOẠT ĐỘNG

- 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.

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

- 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

- 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

- 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

