



KỸ SƯ DỮ LIỆU

NGUYỄN KHANH DŨNG

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

10/02/1980

Hà Nội

thaolinh252512@gmail.com

0382068840

www.website.com

HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại
Học viện Kỹ thuật Mật mã - Công nghệ
thông tin tại Đại học FPT

KỸ NĂNG

- NoSQL (MongoDB, Cassandra)

- Amazon Redshift

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 TỬ 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)

SỞ THÍCH

- Đọc sách

- Du lịch

- Chụp ảnh

- Tham gia cộng đồng lập trình

+ 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

NGƯỜI GIỚI THIỆU

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

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

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

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

DANH HIỆU VÀ GIẢI THƯỞNG

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

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

CHỨNG CHỈ

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

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

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

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.

- 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

- 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

- 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