



KỸ SƯ DỮ LIỆU

NGÔ PHƯƠNG KHANH

MỤC TIÊU NGHỀ NGHIỆP

Tôi muốn chia sẻ kiến thức qua việc viết blog, tổ chức hội thảo kỹ thuật hoặc đóng góp tài liệu chính thức để thúc đẩy cộng đồng Data Engineering trong nước phát triển mạnh mẽ.

THÔNG TIN CÁ NHÂN

18/05/1987

Hà Nội

thaolinh252512@gmail.com

0982740001

www.website.com

HỌC VẤN

- Cơ sở dữ liệu và khai phá dữ liệu tại Đại học Khoa học Tự nhiên – ĐHQG Hà Nội - Kỹ thuật phần mềm tại Học viện Công nghệ Bưu chính Viễn thông

KỸ NĂNG

- Monitoring với Prometheus/Grafana
- Apache Kafka
- PostgreSQL
- ETL/ELT Design

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

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

+ THIẾT KẾ VÀ TRIỂN KHAI PIPELINE ETL SỬ DỤNG
APACHE AIRFLOW

SỞ THÍCH

- Sưu tầm sách lập trình

+ TÍCH HỢP DỮ LIỆU TỪ NHIỀU NGUỒN NHƯ MYSQL,
GOOGLE ANALYTICS VÀ API BÊN THỨ BA

NGƯỜI GIỚI THIỆU

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

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

- Bà Đỗ Ngọc Trang (Senior Data
Engineer - Công ty SmartRetail) -
trang.do@smartretail.vn - 0922333444

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

- Ông Phan Thành Tâm (Big Data
Architect - Công ty LogiData) -
tam.phan@logidata.vn - 0955111222

+ TỐI ƯU HOÁ QUY TRÌNH XỬ LÝ BẰNG SPARK TRÊN
GCP DATAPROC

- **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

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

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

- **2021** - Bằng khen về bảo mật và chất lượng dữ liệu - Công ty

InfosecData

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

CHỨNG CHỈ

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

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

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

HOẠT ĐỘNG

- **Mentor khóa học nền tảng kỹ sư dữ liệu tại Trung tâm Đào tạo CloudTech (2023)**

+ Hướng dẫn học viên triển khai hệ thống ingest dữ liệu bằng Kafka.

+ Đánh giá bài tập về xử lý dữ liệu song song với Spark.

+ Tư vấn về định hướng nghề nghiệp cho sinh viên muốn theo ngành data engineering.

- **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 tư vấn kiến trúc dữ liệu tại Công ty SmartData (2021 - 2022)

- + Đề xuất mô hình lưu trữ dữ liệu dạng columnar cho truy vấn phân tích nhanh.
- + Tham gia thiết kế cấu trúc bảng dữ liệu trong hệ thống warehouse.
- + Xây dựng best practice về đặt tên và phân vùng dữ liệu.

DỰ ÁN

- 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

- 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