



KỸ SƯ DỮ LIỆU

ĐỖ TRÍ KHÁNH

MỤC TIÊU NGHỀ NGHIỆP

Tôi kỳ vọng học hỏi sâu hơn về kiến trúc dữ liệu phân tán (distributed data architecture) và các nguyên lý CAP, consistency, replication, để đảm bảo tính sẵn sàng cao.

THÔNG TIN CÁ NHÂN

25/11/1995

Hà Nội

thaolinh252512@gmail.com

0833028154

www.website.com

KINH NGHIỆM LÀM VIỆC

- **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)

HỌC VẤN

- Kỹ thuật phần mềm tại Học viện Công nghệ Bưu chính Viễn thông

KỸ NĂNG

- dbt (Data Build Tool)

- Apache Kafka

- Data Lake Architecture

- Google Cloud Platform (GCP)

SỞ THÍCH

- Trồng cây
- Chơi cờ vua
- Chơi đàn guitar
- Tham gia hội thảo công nghệ

+ 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

NGƯỜI GIỚI THIỆU

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

- **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

- **2023** - Giải thưởng 'Kiến trúc dữ liệu sáng tạo' do CTO trao tặng tại DataWare

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

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

InfosecData

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

CHỨNG CHỈ

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

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

HOẠT ĐỘNG

- **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.

- **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.

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