



## KỸ SƯ DỮ LIỆU

TRẦN ANH TRUNG

### 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

10/09/1981

Hà Nội

thaolinh252512@gmail.com

0805365338

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

### HỌC VẤN

- Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM - Kỹ thuật phần mềm tại Học viện Công nghệ Bưu chính Viễn thông

### KỸ NĂNG

- Data Quality Validation (Great Expectations)

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

### SỞ THÍCH

### DANH HIỆU VÀ GIẢI THƯỞNG

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

- Tham gia hội thảo công nghệ

## NGƯỜI GIỚI THIỆU

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

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

- **2022** - Top nhân viên cải tiến năng suất truy vấn dữ liệu SQL tại ReportPro

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

## CHỨNG CHỈ

- **2021** - Streaming Data with Apache Kafka – Confluent Academy

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

## 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.

### - 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

### - 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

