



KỸ SƯ DỮ LIỆU

VŨ NGÀ KHANH

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

23/09/1996

Hà Nội

thaolinh252512@gmail.com

0348560796

www.website.com

HỌC VẤN

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

KỸ NĂNG

- BigQuery

SỞ THÍCH

- Thể thao

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

- Thiết kế sản phẩm cá nhân

KHO THEO THỜI GIAN THỰC

NGƯỜI GIỚI THIỆU

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

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

- 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

+ XỬ LÝ DỮ LIỆU STREAMING TỪ KAFKA VÀ LƯU TRỮ VÀO BIGQUERY

+ XÂY DỰNG CÁC VIEW LOGIC TRÊN DBT PHỤC VỤ BI DASHBOARD

- **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

CHỨNG CHỈ

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

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

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.

DỰ ÁN

- 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

- 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

- 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

- 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

