



KỸ SƯ DỮ LIỆU

ĐỖ GIANG CHÂU

MỤC TIÊU NGHỀ NGHIỆP

Tôi mong muốn đóng góp vào việc xây dựng hệ thống phân quyền và bảo mật dữ liệu mạnh mẽ, tuân thủ các tiêu chuẩn như GDPR, ISO 27001 trong vận hành hệ thống dữ liệu.

THÔNG TIN CÁ NHÂN

27/07/1982

Hà Nội

thaolinh252512@gmail.com

0721723463

www.website.com

HỌC VẤN

- Trí tuệ nhân tạo và Dữ liệu lớn tại
Học viện Kỹ thuật Mật mã

KỸ NĂNG

- BigQuery
- Docker
- AWS (S3, EMR, Glue)
- Snowflake
- NoSQL (MongoDB, Cassandra)

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

- **BIG DATA ENGINEER** TẠI LOGIAI (2020-2022)

+ XÂY DỰNG PIPELINE THU THẬP VÀ XỬ LÝ DỮ LIỆU

VẬN CHUYỂN TỪ THIẾT BỊ IOT

SỞ THÍCH

- Chụp ảnh

- Thử nghiệm công nghệ mới

+ SỬ DỤNG SPARK STREAMING ĐỂ XỬ LÝ DỮ LIỆU REAL-TIME

+ KẾT NỐI HỆ THỐNG VÀO ELASTICSEARCH VÀ PHÁT TRIỂN DASHBOARD GIÁM SÁT

NGƯỜI GIỚI THIỆU

- Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123

- **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

- **JUNIOR DATA ENGINEER** TẠI RETAILTECH (2020-2021)

+ PHỐI HỢP XÂY DỰNG HỆ THỐNG PHÂN TÍCH TỒN KHO THEO THỜI GIAN THỰC

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

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

DANH HIỆU VÀ GIẢI THƯỞNG

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

CHỨNG CHỈ

- **2020** - AWS Certified Data Analytics – Specialty

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

- **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera

HOẠT ĐỘNG

- **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

- 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

- 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

