|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Đỗ Châu ChâuMỤC TIÊU NGHỀ NGHIỆP **Tôi hướng tới việc thiết kế hệ thống ETL/ELT có khả năng mở rộng, dễ bảo trì và tích hợp với nhiều nguồn dữ liệu phức tạp như API, streaming và cơ sở dữ liệu phi cấu trúc.** |
| THÔNG TIN CÁ NHÂN  07/10/1982  Hà Nội  thaolinh252512@gmail.com  0933487608  www.website.com  HỌC VẤN - Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã - Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội KỸ NĂNG  - NoSQL (MongoDB, Cassandra)  - Docker SỞ THÍCH - Đi bộ đường dài NGƯỜI GIỚI THIỆU - Bà Nguyễn Thị Lệ (Analytics Lead - Công ty InsightTech) - le.nguyen@insighttech.vn - 0966999888  - Bà Phạm Thanh Hương (Data Operations Manager - Công ty ReportPro) - huong.pham@reportpro.vn - 0912111222  - Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666  - Ông Nguyễn Quốc Huy (Data Platform Lead - Công ty DataBridge) - huy.nguyen@databridge.vn - 0905123456 |  | KINH NGHIỆM LÀM VIỆC- **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- **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)+ 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- **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- **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 sátDANH HIỆU VÀ GIẢI THƯỞNG - **2021** - Kỹ sư dữ liệu xuất sắc quý I tại Công ty DataBridge CHỨNG CHỈ - **2022** - Data Engineering on Google Cloud Platform – Coursera  - **2021** - SQL for Data Engineering – Datacamp  - **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.  **- Thành viên nhóm kỹ thuật dữ liệu tại Dự án E-Government (2023)**  + Thiết kế hệ thống thu thập dữ liệu hành chính từ nhiều bộ ngành.  + Thiết lập hệ thống kiểm tra chất lượng dữ liệu tự động.  + Triển khai pipeline đồng bộ dữ liệu hằng ngày với độ trễ thấp.  **- 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 **- 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  **- 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  **- 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  **- 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  **- 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 |
|  |  |  |