|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Lê Ngọc SơnMỤC TIÊU NGHỀ NGHIỆP **Tôi hướng tới mục tiêu trong 3 năm tới trở thành một Senior Data Engineer, có khả năng thiết kế toàn bộ kiến trúc dữ liệu, dẫn dắt đội nhóm và đề xuất cải tiến cho chiến lược dữ liệu doanh nghiệp.** |
| THÔNG TIN CÁ NHÂN  11/05/1988  Hà Nội  thaolinh252512@gmail.com  0338833937  www.website.com  HỌC VẤN - Công nghệ thông tin tại Đại học FPT KỸ NĂNG  - Amazon Redshift SỞ THÍCH - Tham gia hội thảo công nghệ  - Tham gia hackathon  - Thể thao  - Chơi đàn guitar  - Trồng cây NGƯỜI GIỚI THIỆU - Ông Phan Thành Tâm (Big Data Architect - Công ty LogiData) - tam.phan@logidata.vn - 0955111222  - Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123  - Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666 |  | 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- **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- **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 dùngDANH HIỆU VÀ GIẢI THƯỞNG - **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge  - **2020** - Giải thưởng 'Pipeline ổn định nhất năm' tại TechFlow CHỨNG CHỈ - **2020** - Google Cloud Professional Data Engineer Certificate  - **2023** - Modern Data Engineering with dbt – dbt Labs HOẠT ĐỘNG **- Diễn giả hội thảo 'Big Data Architecture' tại Data Talks Vietnam (2022)**  + Trình bày kiến trúc hệ thống thu thập và xử lý dữ liệu đa nguồn.  + Phân tích ưu nhược điểm của Data Warehouse vs Data Lakehouse.  + Giới thiệu các công cụ phổ biến như Airflow, dbt, Snowflake.  **- 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 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. 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  **- 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 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  **- 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 |
|  |  |  |