|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Nguyễn Nga HạnhMỤC TIÊU NGHỀ NGHIỆP **Tôi định hướng tạo ra nền tảng dữ liệu mở (open data platform) cho các bộ phận nội bộ dễ dàng truy cập, khai thác và sử dụng dữ liệu theo cơ chế phân quyền rõ ràng.** |
| THÔNG TIN CÁ NHÂN  09/04/1990  Hà Nội  thaolinh252512@gmail.com  0325703752  www.website.com  HỌC VẤN - Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM - Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội KỸ NĂNG  - PostgreSQL  - Apache Kafka  - Google Cloud Platform (GCP) SỞ THÍCH - Thể thao  - Nấu ăn  - Chơi đàn guitar  - Chơi cờ vua NGƯỜI GIỚI THIỆU - Bà Trần Thị Mai (Technical Manager - Công ty CloudData) - mai.tran@cloudata.vn - 0933666777  - Bà Lê Thùy Linh (Data Infrastructure Manager - Công ty CloudOps) - linh.le@cloudops.vn - 0944555666  - Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123 |  | 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- **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- **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- **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 chínhDANH HIỆU VÀ GIẢI THƯỞNG - **2021** - Bằng khen về bảo mật và chất lượng dữ liệu – Công ty InfosecData  - **2022** - Top nhân viên cải tiến năng suất truy vấn dữ liệu SQL tại ReportPro  - **2023** - Giải thưởng 'Kiến trúc dữ liệu sáng tạo' do CTO trao tặng tại DataWare CHỨNG CHỈ - **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera  - **2022** - Data Engineering on Google Cloud Platform – Coursera  - **2021** - Streaming Data with Apache Kafka – Confluent Academy HOẠT ĐỘNG **- 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 **- 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  **- 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  **- 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 |
|  |  |  |