|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Bùi Hạnh TùngMỤC TIÊU NGHỀ NGHIỆP **Tôi định hướng trở thành một Kỹ sư Dữ liệu chuyên nghiệp, có khả năng xây dựng các pipeline xử lý dữ liệu mạnh mẽ và tối ưu nhằm phục vụ các hệ thống phân tích và học máy trong doanh nghiệp.** |
| THÔNG TIN CÁ NHÂN  11/10/1990  Hà Nội  thaolinh252512@gmail.com  0356492912  www.website.com  HỌC VẤN - Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội KỸ NĂNG  - CI/CD cho Data Pipeline  - Apache Spark  - Google Cloud Platform (GCP)  - ETL/ELT Design SỞ THÍCH - Tham gia cộng đồng lập trình  - Chụp ảnh NGƯỜI GIỚI THIỆU - Bà Phạm Thanh Hương (Data Operations Manager - Công ty ReportPro) - huong.pham@reportpro.vn - 0912111222  - Ông Hồ Trung Kiên (Engineering Manager - Công ty ETLTech) - kien.ho@etltech.vn - 0933888777  - Bà Nguyễn Thị Lệ (Analytics Lead - Công ty InsightTech) - le.nguyen@insighttech.vn - 0966999888 |  | 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 DataprocDANH HIỆU VÀ GIẢI THƯỞNG - **2023** - Giải thưởng 'Kiến trúc dữ liệu sáng tạo' do CTO trao tặng tại DataWare  - **2021** - Giải thưởng 'Dự án phân tích dữ liệu xuất sắc' tại FinData  - **2021** - Bằng khen về bảo mật và chất lượng dữ liệu – Công ty InfosecData  - **2020** - Giải thưởng 'Pipeline ổn định nhất năm' tại TechFlow  - **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge CHỨNG CHỈ - **2022** - Data Engineering on Google Cloud Platform – Coursera HOẠT ĐỘNG **- Thành viên nhóm xây dựng pipeline dữ liệu tại Phòng Lab Dữ liệu - Đại học Bách khoa (2021 - 2023)**  + Phát triển luồng dữ liệu ETL sử dụng Apache Airflow.  + Thiết lập hệ thống ingestion dữ liệu từ API và cơ sở dữ liệu nguồn.  + Phối hợp với nhóm phân tích dữ liệu để tối ưu schema lưu trữ.  **- 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 phát triển hệ thống phân tích log tại Công ty SecureLog (2021)**  + Xây dựng pipeline thu thập và phân tích log từ nhiều nguồn hệ thống.  + Triển khai hệ thống cảnh báo bất thường dựa trên mô hình thống kê.  + Sử dụng ELK stack để trực quan hóa dữ liệu log.  **- Cộng tác viên dự án dữ liệu mở tại OpenData Vietnam (2020)**  + Thu thập và chuẩn hóa dữ liệu dân số từ các tỉnh thành.  + Tích hợp dữ liệu vào hệ thống cơ sở dữ liệu MongoDB phục vụ phân tích.  + Tạo dashboard đơn giản để trực quan hóa dữ liệu bằng Plotly Dash.  **- 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. 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 |
|  |  |  |