|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Bùi Quang TríMỤC TIÊU NGHỀ NGHIỆP **Tôi mong muốn phát triển hệ thống lưu trữ dữ liệu dạng time-series cho các ứng dụng theo dõi thiết bị IoT, phân tích logs và hệ thống sự kiện theo thời gian.** |
| THÔNG TIN CÁ NHÂN  09/01/1986  Hà Nội  thaolinh252512@gmail.com  0954835273  www.website.com  HỌC VẤN - Kỹ thuật máy tính tại Đại học Sư phạm Kỹ thuật TP.HCM - Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã KỸ NĂNG  - ETL/ELT Design  - SQL  - dbt (Data Build Tool) SỞ THÍCH - Tham gia hội thảo công nghệ  - Tham gia hackathon  - Chơi đàn guitar NGƯỜI GIỚI THIỆU - Bà Trần Thị Mai (Technical Manager - Công ty CloudData) - mai.tran@cloudata.vn - 0933666777 |  | 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- **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 dashboardDANH HIỆU VÀ GIẢI THƯỞNG - **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge CHỨNG CHỈ - **2020** - Google Cloud Professional Data Engineer Certificate  - **2022** - Big Data Engineering with Apache Spark – edX (BerkeleyX)  - **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – 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 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. 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  **- 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  **- 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 |
|  |  |  |