|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Ngô Thu NgọcMỤC TIÊU NGHỀ NGHIỆP **Tôi muốn làm việc với các nhóm Data Scientist và BI Analyst để đảm bảo dữ liệu được xử lý đúng định dạng, đúng ngữ nghĩa và cung cấp kịp thời cho phân tích chuyên sâu.** |
| THÔNG TIN CÁ NHÂN  20/09/1991  Hà Nội  thaolinh252512@gmail.com  0899054097  www.website.com  HỌC VẤN - Phân tích và xử lý dữ liệu lớn tại Đại học Phenikaa - Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội KỸ NĂNG  - SQL  - Linux  - Apache Spark SỞ THÍCH - Tham gia cộng đồng lập trình  - Tập gym  - Thử nghiệm công nghệ mới  - Chơi đàn guitar NGƯỜI GIỚI THIỆU - Ông Phan Thành Tâm (Big Data Architect - Công ty LogiData) - tam.phan@logidata.vn - 0955111222 |  | KINH NGHIỆM LÀM VIỆC- **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- **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  - **2022** - Vinh danh 'Nhân sự triển khai hệ thống streaming real-time đầu tiên' tại LogiTech CHỨNG CHỈ - **2023** - Designing Data-Intensive Applications – O’Reilly Certification Program  - **2021** - SQL for Data Engineering – Datacamp HOẠT ĐỘNG **- Tình nguyện viên hỗ trợ nghiên cứu học thuật tại Khoa Công nghệ Thông tin (2020)**  + Chuẩn hóa dữ liệu nghiên cứu sinh viên từ nhiều năm học.  + Phân tích và tổng hợp xu hướng đăng ký môn học.  + Viết báo cáo kỹ thuật trình bày cho giảng viên hướng dẫn.  **- 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. DỰ ÁN **- 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  **- 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 |
|  |  |  |