|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Ngô Nga BìnhMỤC 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  14/10/1993  Hà Nội  thaolinh252512@gmail.com  0304678697  www.website.com  HỌC VẤN - Trí tuệ nhân tạo và Dữ liệu lớn tại Học viện Kỹ thuật Mật mã - Cơ sở dữ liệu và khai phá dữ liệu tại Đại học Khoa học Tự nhiên – ĐHQG Hà Nội KỸ NĂNG  - Monitoring với Prometheus/Grafana  - SQL SỞ THÍCH - Đi bộ đường dài  - Nghe nhạc  - Thể thao NGƯỜI GIỚI THIỆU - 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 - **2020** - Giải thưởng 'Pipeline ổn định nhất năm' tại TechFlow  - **2023** - Bằng khen về tối ưu chi phí lưu trữ dữ liệu cloud – Công ty CloudData CHỨNG CHỈ - **2021** - Streaming Data with Apache Kafka – Confluent Academy  - **2023** - Modern Data Engineering with dbt – dbt Labs  - **2022** - Data Engineering on Google Cloud Platform – Coursera HOẠT ĐỘNG **- 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.  **- 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.  **- Người viết blog kỹ thuật về hệ thống dữ liệu tại Blog cá nhân datapi.dev (2022 - nay)**  + Chia sẻ kiến thức về thiết kế hệ thống ETL quy mô lớn.  + So sánh các giải pháp lưu trữ dữ liệu: PostgreSQL, Redshift, BigQuery.  + Viết hướng dẫn cấu hình Airflow và orchestration jobs theo lịch định kỳ.  **- Tình nguyện viên hỗ trợ tổ chức Data Day tại Vietnam Data Community (2022)**  + Thiết lập demo pipeline dữ liệu thời gian thực bằng Kafka.  + Hỗ trợ phần trình diễn luồng dữ liệu streaming giữa các node Spark.  + Giới thiệu cấu trúc cơ bản của hệ thống Data Lake.  **- 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 **- 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 |
|  |  |  |