|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Đỗ Bình HiếuMỤC TIÊU NGHỀ NGHIỆP **Tôi mong muốn nâng cao kỹ năng xử lý dữ liệu streaming sử dụng Kafka và Spark Structured Streaming để phục vụ các bài toán real-time như giám sát hành vi người dùng hoặc phát hiện gian lận.** |
| THÔNG TIN CÁ NHÂN  02/04/1980  Hà Nội  thaolinh252512@gmail.com  0996048596  www.website.com  HỌC VẤN - Toán – Tin ứng dụng tại Đại học Kinh tế Quốc dân - Công nghệ thông tin tại Đại học FPT KỸ NĂNG  - Google Cloud Platform (GCP)  - Data Lake Architecture SỞ THÍCH - Thử nghiệm công nghệ mới  - Nấu ăn  - Đọc sách NGƯỜI GIỚI THIỆU - Ông Vũ Minh Đức (CTO - Công ty DataStreamX) - duc.vu@datastreamx.vn - 0977333444  - Bà Trần Thị Mai (Technical Manager - Công ty CloudData) - mai.tran@cloudata.vn - 0933666777  - Ông Hồ Trung Kiên (Engineering Manager - Công ty ETLTech) - kien.ho@etltech.vn - 0933888777 |  | 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- **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 dashboard- **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- **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- **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 dùngDANH HIỆU VÀ GIẢI THƯỞNG - **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge  - **2022** - Top 3 nhân viên cải tiến hệ thống ETL tại SmartRetail CHỨNG CHỈ - **2022** - Big Data Engineering with Apache Spark – edX (BerkeleyX)  - **2021** - IBM Data Engineering Professional Certificate – Coursera  - **2021** - Streaming Data with Apache Kafka – Confluent Academy  - **2020** - Google Cloud Professional Data Engineer Certificate  - **2020** - AWS Certified Data Analytics – Specialty HOẠT ĐỘNG **- Mentor khóa học nền tảng kỹ sư dữ liệu tại Trung tâm Đào tạo CloudTech (2023)**  + Hướng dẫn học viên triển khai hệ thống ingest dữ liệu bằng Kafka.  + Đánh giá bài tập về xử lý dữ liệu song song với Spark.  + Tư vấn về định hướng nghề nghiệp cho sinh viên muốn theo ngành data engineering.  **- Thành viên nhóm tư vấn kiến trúc dữ liệu tại Công ty SmartData (2021 - 2022)**  + Đề xuất mô hình lưu trữ dữ liệu dạng columnar cho truy vấn phân tích nhanh.  + Tham gia thiết kế cấu trúc bảng dữ liệu trong hệ thống warehouse.  + Xây dựng best practice về đặt tên và phân vùng dữ liệu.  **- 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.  **- 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 **- 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 |
|  |  |  |