|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Hoàng Khánh TríMỤC TIÊU NGHỀ NGHIỆP **Tôi hướng tới việc thiết kế hệ thống ETL/ELT có khả năng mở rộng, dễ bảo trì và tích hợp với nhiều nguồn dữ liệu phức tạp như API, streaming và cơ sở dữ liệu phi cấu trúc.** |
| THÔNG TIN CÁ NHÂN  20/02/1980  Hà Nội  thaolinh252512@gmail.com  0831004335  www.website.com  HỌC VẤN - Cơ sở dữ liệu và khai phá dữ liệu tại Đại học Khoa học Tự nhiên – ĐHQG Hà Nội KỸ NĂNG  - ETL/ELT Design  - Apache Kafka SỞ THÍCH - Viết blog kỹ thuật  - Tập gym NGƯỜI GIỚI THIỆU - Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123  - Bà Đỗ Ngọc Trang (Senior Data Engineer - Công ty SmartRetail) - trang.do@smartretail.vn - 0922333444  - Ông Phan Thành Tâm (Big Data Architect - Công ty LogiData) - tam.phan@logidata.vn - 0955111222  - Ông Nguyễn Quốc Huy (Data Platform Lead - Công ty DataBridge) - huy.nguyen@databridge.vn - 0905123456 |  | KINH NGHIỆM LÀM VIỆC- **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 dashboard- **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- **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 DataprocDANH HIỆU VÀ GIẢI THƯỞNG - **2021** - Giải thưởng 'Dự án phân tích dữ liệu xuất sắc' tại FinData  - **2022** - Top nhân viên cải tiến năng suất truy vấn dữ liệu SQL tại ReportPro CHỨNG CHỈ - **2020** - Google Cloud Professional Data Engineer Certificate  - **2020** - AWS Certified Data Analytics – Specialty  - **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera  - **2021** - SQL for Data Engineering – Datacamp  - **2023** - Modern Data Engineering with dbt – dbt Labs 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 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.  **- 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 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 **- 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  **- 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  **- 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  **- 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 |
|  |  |  |