|  |  |  |
| --- | --- | --- |
| Decorative |  | Kỹ sư Dữ liệu |
|  | Trần Khánh ChâuMỤC TIÊU NGHỀ NGHIỆP **Tôi kỳ vọng được làm việc trong các dự án về xây dựng data lake và data warehouse quy mô lớn, áp dụng Redshift, BigQuery hoặc Snowflake để phục vụ báo cáo phân tích đa chiều.** |
| THÔNG TIN CÁ NHÂN  20/06/1992  Hà Nội  thaolinh252512@gmail.com  0828970548  www.website.com  HỌC VẤN - Hệ thống thông tin quản lý tại Đại học Bách Khoa Hà Nội KỸ NĂNG  - Data Lake Architecture  - SQL  - Google Cloud Platform (GCP) SỞ THÍCH - Học ngoại ngữ NGƯỜI GIỚI THIỆU - Ông Lê Hoàng Dương (Head of Engineering - Công ty FinData) - duong.le@findata.vn - 0988999123 |  | KINH NGHIỆM LÀM VIỆC- **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- **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 CHỨNG CHỈ - **2022** - ETL and Data Pipelines with Shell, Airflow and Kafka – Coursera  - **2023** - Modern Data Engineering with dbt – dbt Labs 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.  **- 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 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 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.  **- Diễn giả hội thảo 'Big Data Architecture' tại Data Talks Vietnam (2022)**  + Trình bày kiến trúc hệ thống thu thập và xử lý dữ liệu đa nguồn.  + Phân tích ưu nhược điểm của Data Warehouse vs Data Lakehouse.  + Giới thiệu các công cụ phổ biến như Airflow, dbt, Snowflake. 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  **- 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 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  **- 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 |
|  |  |  |