



KỸ SƯ DỮ LIỆU

PHẠM NAM TRÍ

MỤC TIÊU NGHỀ NGHIỆP

Tôi định hướng trở thành một Kỹ sư Dữ liệu chuyên nghiệp, có khả năng xây dựng các pipeline xử lý dữ liệu mạnh mẽ và tối ưu nhằm phục vụ các hệ thống phân tích và học máy trong doanh nghiệp.

THÔNG TIN CÁ NHÂN

17/01/1991

Hà Nội

thaolinh252512@gmail.com

0722545240

www.website.com

HỌC VẤN

- Khoa học máy tính tại Đại học Quốc tế - ĐHQG TP.HCM - Toán - Tin ứng dụng tại Đại học Kinh tế Quốc dân

KỸ NĂNG

- Docker

- Data Lake Architecture

- PostgreSQL

- Monitoring với Prometheus/Grafana

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)

SỞ THÍCH

- Học ngoại ngữ

+ 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

NGƯỜI GIỚI THIỆU

- Bà Phạm Thanh Hương (Data
Operations Manager - Công ty
ReportPro) - huong.pham@reportpro.vn
- 0912111222

+ 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 CHÍNH

DANH 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
- **2020** - Top 5 kỹ sư xử lý dữ liệu lớn trong cuộc thi nội bộ AI Challenge

CHỨNG CHỈ

- **2021** - SQL for Data Engineering – Datacamp
- **2020** - AWS Certified Data Analytics – Specialty
- **2022** - Big Data Engineering with Apache Spark – edX (BerkeleyX)
- **2021** - IBM Data Engineering Professional Certificate – Coursera
- **2023** - Designing Data-Intensive Applications – O'Reilly Certification Program

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.

- 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.

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

- 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

- 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

- 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