

Le Thanh Dat

lethanhdatt.me@gmail.com | 0981729676 | github.com/ltdthanhdatt

Education

Hanoi University of Science and Technology

9/2021 – 7/2025

Faculty of Mathematics and Informatics

Experience

Data Engineer Intern, Edupia

1/2025 – 5/2025

- Deployed and configured a lab environment on a bare metal server.
- Deployed and managed services using Docker Swarm, ensuring high availability and scalability of system components.
- Set up monitoring, logging, and alerting for the data infrastructure using Prometheus and Grafana, including creating and maintaining dashboards to track Oracle Tablespace memory usage.
- Implemented CI/CD pipelines using GitLab CI/CD for data workflows.
- Managed metadata using OpenMetadata to track and maintain information on data sources, tables, and data lineage.
- Built ELT pipelines using Talend and Apache Airflow to automate data processing workflows.
- Utilized Kafka and Debezium to enable real-time data synchronization between systems.
- Scheduled and orchestrated data pipelines with Airflow, ensuring stable data flow and monitoring performance.
- Configured and deployed Cloudflare Zero Trust and Tailscale to secure network connections and manage access to internal services.

Data Engineer Intern, Kim Nam Group

8/2024 – 1/2025

- Designed and implemented a Data Lakehouse architecture for data processing and storage.
- Built Raw, Curated, and Processed data layers on MinIO for efficient data management.
- Used Apache Iceberg table format in the Processed Layer.
- Developed ELT pipelines using Spark and Kafka.
- Scheduled and orchestrated pipelines using Apache Airflow, integrating with the CRM system.
- Containerized and orchestrated pipelines with Docker and Kubernetes.
- Optimized multi-dimensional data queries using Trino.
- Built a RAG (Retrieval-Augmented Generation) system using FAISS and LangChain.

Skills

DevOps & Automation: Docker, K8s, Docker Swarm, GitLab CI/CD, Terraform, Ansible, Linux

Cloud: AWS EC2, AWS Lambda, AWS S3

Networking & VPN: OpenVPN, Tailscale, SSH Tunnel, Cloudflare Zero Trust, Nginx

Data Engineering: Apache Spark, dbt, Talend, Airbyte

Data Orchestration: Apache Airflow

Database: ClickHouse, Oracle, DuckDB, MongoDB

Lakehouse: MinIO, Trino, Parquet, Iceberg, Dremio

Monitoring: Prometheus, Grafana

Web Development: Node.js, ReactJS, Bootstrap, HTMX

Web Scraping: Puppeteer, Selenium

Programming Languages: Python, JavaScript, Bash, C

Lê Thành Đạt

lethanhdad.me@gmail.com | 0981729676 | github.com/ltdthanhdad

Học vấn

Đại học Bách Khoa Hà Nội

9/2021 – 7/2025

Khoa Toán - Tin

Kinh nghiệm

Thực tập sinh Data Engineer, Edupia

1/2025 – 5/2025

- Triển khai và cấu hình môi trường lab trên server bare metal.
- Triển khai và quản lý dịch vụ bằng Docker Swarm, đảm bảo khả năng mở rộng và tính sẵn sàng cao cho các thành phần hệ thống.
- Thiết lập hệ thống giám sát, logging và alerting cho hạ tầng dữ liệu bằng Prometheus, Grafana, bao gồm việc tạo và duy trì dashboard để theo dõi bộ nhớ Oracle Tablespace.
- Triển khai CI/CD trên GitLab CI/CD cho các pipeline dữ liệu.
- Quản lý metadata dữ liệu sử dụng OpenMetadata để theo dõi và duy trì thông tin về các nguồn dữ liệu, bảng và data lineage.
- Xây dựng pipeline ELT sử dụng Talend và Apache Airflow để tự động hóa quy trình xử lý dữ liệu.
- Sử dụng Kafka và Debezium để đồng bộ dữ liệu theo thời gian thực giữa các hệ thống.
- Lập lịch và điều phối pipeline dữ liệu với Airflow, đảm bảo luồng dữ liệu ổn định và giám sát hiệu suất.
- Cấu hình và triển khai Cloudflare Zero Trust và Tailscale để bảo mật kết nối mạng, quản lý quyền truy cập và kết nối an toàn cho các dịch vụ nội bộ.

Thực tập sinh Data Engineer, Tập đoàn Kim Nam

8/2024 – 1/2025

- Thiết kế và triển khai kiến trúc Data Lakehouse để xử lý và lưu trữ dữ liệu.
- Xây dựng các lớp dữ liệu Raw, Curated, Processed trên MinIO để quản lý hiệu quả.
- Sử dụng định dạng bảng Apache Iceberg ở Processed Layer.
- Phát triển pipeline ELT bằng Spark, Kafka.
- Lập lịch và điều phối pipeline bằng Apache Airflow, tích hợp với hệ thống CRM.
- Container hóa và điều phối pipeline với Docker, Kubernetes, phân phối tài nguyên Spark Worker.
- Tối ưu truy vấn dữ liệu đa chiều bằng Trino.
- Xây dựng hệ thống RAG với FAISS, LangChain.

Kỹ năng

DevOps & Automation: Docker, K8s, Docker Swarm, GitLab CI/CD, Terraform, Ansible, Linux

Cloud: AWS EC2, AWS Lambda, AWS S3

Networking & VPN: OpenVPN, Tailscale, SSH Tunnel, Cloudflare Zero Trust, Nginx

Data Engineering: Apache Spark, dbt, Talend, Airbyte

Data Orchestration: Apache Airflow

Database: ClickHouse, Oracle, DuckDB, MongoDB

Lakehouse: MinIO, Trino, Parquet, Iceberg, Dremio

Monitoring: Prometheus, Grafana

Web Development: Node.js, ReactJS, Bootstrap, HTMX

Web Scraping: Puppeteer, Selenium

Programming Languages: Python, JavaScript, Bash, C