

Le Duy Kha

Ho Chi Minh city, Viet Nam | duykha0205@gmail.com | [linkedin](#) | website: <http://kha.nichedatainsight.com/>

I define myself as a Platform Engineer with nearly 4 years of experience building scalable data and AI cloud infrastructure. Expertise spans data platform architecture, DevOps automation, and AI system integration. Aspiring Solution Architect focused on delivering innovative solutions in the data and AI domain.

Work Experience:

TechX Company | HCM City
Data Engineer

Dec 2024 - Present

Synaptix AIOps Platform - Incident Detection for Banking Operations

Key Responsibilities:

- Designed and implemented data ingestion pipelines and CI/CD workflows for platform deployment
- Built infrastructure automation and application deployment flows using Helm charts with ArgoCD for automated EKS deployments
- Developed backfill logic and configured networking architecture for production operations

Technologies: Helm, AWS EKS, EMR on EKS, Terraform, ArgoCD

OLAP Reporting System - State Bank of Vietnam

Key Responsibilities:

- Built automated reporting platform for Vikki Bank to streamline regulatory and operational reporting
- Developed end-to-end data pipeline: DBT transformations → Spark processing → S3 table storage, orchestrated on EMR EC2 clusters

Technologies: DBT, Apache Spark, AWS S3, EMR on EC2

Xbot - multi agent for operation

Key Responsibilities:

- Designed and proposed the multi-agent architecture, providing technical guidance to the team during the initial development phase
- Developed a scalable automation multi-agent system using LangChain, implementing intelligent routing mechanisms and sub-agent orchestration
- Integrated Qdrant vector database for knowledge retrieval and designed the collection architecture to optimize information access
- Established comprehensive project structure, deployed infrastructure, and implemented CI/CD pipelines for end-to-end service delivery

Technologies: AWS ECS, Terraform, LangChain, Flowise, Langfuse, AWS Services

XGenAI - Internal agent chatbot platform

- Built and optimized CI/CD workflow, reduced 80% build time and deployment failures, defined workflow for code merging
- Monitoring system with cost optimization
- Monitored AWS costs, deployed Redis saving \$500/month for company, built Terraform Lambda for auto infrastructure shutdown reducing 60% monthly costs
- Developed backend with SQL API using FastAPI and integrated SQL agent using LLM for querying employee KPI and project allocation data
- **Technologies Used:** Github action, python, fastapi, langgraph, grafana, mirmir, ECR, Cloud trail, Terraform, lambda, bedrock agent

EOD - Enterprise Operation Data for banking

- Build, test, code(Kafka SMT) and config streaming platform for change data capture between multi-SQL and NoSQL database on cloud and on-prem, performance 10k message / min
- **Auto Generate config:** Build tool for automation generate config for config streaming change capture data from multi-source, reduce 1 hours to config to 1 mins
- Build data quality pipeline
- **Technologies Used:** Kafka, ArgoCD, Docker, AWS, Python, Debezium, Java, K8s, Terragrunt, MongoDB, DocumentDB, Postgres, MySQL, MSSQL, Oracle

Yes4All Company | Phu Nhuan, HCM City

Aug 2023 – Nov 2024

Data Engineer - Data Operation

- **Designed and Implemented Data Pipelines:** Architected, developed, and maintained scalable ETL pipelines to ingest, transform, and load data from diverse sources into a centralized data warehouse, ensuring high data integrity and availability.
- **Collaborated Cross-Functionally:** Worked closely with data scientists, analysts, and other stakeholders to deliver actionable data solutions tailored to business needs.
- **Optimized Data Processing:** Enhanced pipeline performance by 5x through multithreading/multiprocessing techniques and accelerated API data retrieval by 20x using asynchronous I/O.
- **Optimized Database Performance:** Improved database tables and queries using TimescaleDB, partitioning, and indexing methods to ensure efficient data storage and retrieval.
- **Streamlined System Architecture:** Proposed and led the implementation of Kafka for real-time data streaming, significantly improving resource utilization and system throughput.
- **Automated Data Operations:** Developed a bot detection bypass system for web scraping, reducing manual intervention by automating cookie generation for data crawling.
- **Implemented Automation Tools:** Designed and deployed an automation tool for dropshipping bookings, cutting down processing time and reducing manual efforts for the Sales team.
- **Technologies Used:** Python, PostgreSQL, TimescaleDB, Minio, Apache Airflow, Apache Iceberg, Apache Kafka, Google Cloud Service.

Namiq Company | Tan Binh, HCM City

May 2022 – Apr 2023

Python Engineer

- **Developed GraphQL Data Tools:** Researched, designed, and implemented a GraphQL-based database tool, creating efficient CRUD APIs to facilitate seamless data interactions for the Housee mobile application.
- **Implemented Speech AI Models:** Designed and deployed Convolutional Neural Network (CNN) models using mel-spectrogram features for accurate speech classification, enhancing voice-driven data processing.
- **Built NLP Models for Intent Classification:** Collected and preprocessed large datasets, implemented BERT-based models, and deployed them using TorchServe for prediction serving.
- **Designed Robust Data Schemas:** Created optimized database schemas and implemented CRUD APIs with comprehensive unit testing for the Vilik livestream web application, ensuring reliable data management.
- **Technologies Used:** Unity, C#, TypeScript, Hasura, TorchServe, Python, PyTorch, Torch Audio, Flask, PostgreSQL, SQLAlchemy.

Projects:

Mernva technologies Company | Remote part time - India

Jul 2023 – Nov 2024

As a Platform Engineer, I integrate AI-driven data processing models and pipelines into backend systems, collaborating remotely with international teams to enhance data engineering processes.

- **Integrated AI Models into Data Pipelines:** Seamlessly incorporated AI-driven models into backend systems, enhancing data processing capabilities and enabling intelligent data transformations.
- **Chat Bot Document Summary | LLM Model:** Conducted research and development on a chatbot model using Retrieval-Augmented Generation (RAG) and APIs for efficient data retrieval and summary generation.
- **Technologies Used:** AWS Cloud (EC2, K8s), Python, FastAPI, Docker, Kubernetes, LangChain, Hugging Face.

Data Platform Project - Reddit data pipeline | 2024

- This project implements an automated data pipeline that fetches data from APIs, processes it, and makes it available for analysis using various AWS services. The pipeline is orchestrated using Apache Airflow running on Kubernetes.

- I am developing a project that uses Kubernetes (microk8s), Helm, and Docker to deploy an Airflow job that retrieves data from an API, extracts it, and uploads it to Amazon S3. Once the data is uploaded to S3, it triggers an AWS Lambda function to invoke a Glue job for data transformation. We then use Glue Crawler, Athena, and Redshift for querying, and employ CloudWatch for log monitoring.
- Technologies Used: Airflow (K8s, Microk8s, Helm chart), AWS S3, AWS Lambda, AWS Glue Job, AWS Glue Crawler, AWS Athena and Redshift, AWS CloudWatch
- Source code: [LINK](#)

LLM crash course| MLOpsVN| 2024

- Developed a monitoring platform (including logging, tracing, and application/system monitoring) for LLM applications.
- Technologies: Python, Docker, Grafana, Prometheus, Tempo, Loki, Open Telemetry

Sample MLOps Marathon | MLOps solution | 2023

- Responsibilities included testing workflow of the sample solution, and setting up the environment on linux server for team members ([github](#)).
- Technologies Used: Torch, Python, Bash Script, Makefile, Linux, Docker, Nginx

Photorealistic image to image translation | Bachelor thesis | 2023

- Developed a model using normalization, activation, and attention techniques that trained 100x faster while matching the default model's performance.

Human Activity Recognition based on Wi-Fi CSI Data | Support Master thesis | 2023

- Improved classification accuracy from 40% to 80% using RNNs, normalization, and one-hot encoding techniques.

Education:

Bachelor degree | VNUHCM – University of science | Computer science - GPA: 8.4 / 10

High school | LUONG THE VINH HIGH SCHOOL FOR THE GIFTED | 2016 – 2019

Skills:

- Soft skills: Teamwork, communication, listening, leadership.
- **Technical skill:**
 - **Programming Languages:** Python, C++, Rust, Golang, Type script
 - **Frameworks:** PySpark, Selenium, FastAPI, PyTorch, React
 - **Databases:** PostgreSQL, Redis, MySQL, MongoDB, Redis, MinIO, Iceberg
 - **Skills:** Git, Gitlab, Linux, Bash Script, AWS Cloud, Kubernetes
 - **Language:** English (TOEIC 765), VietNamese (Native)

Certifications:

- Top 2 in MedTech domain, HackYouth 2023 Vin Uni, 2023
- Machine learning in production, online course(coursera), 2022
- SAT Math level 2 (800 / 800), SAT exam, 2021
- Deep learning specialization, online course(coursera), 2021

EXTRA-CURRICULAR ACTIVITIES:

Charity Work | Volunteer at [HOPE](#) | 2024 - Now

[MLOpsVN](#) | Coordinator - Community for Machine Learning Engineer | 2023

[SAB in HCMUS](#) | Head of department - Student Organization | 2020 – 2022