

### PROFESSIONAL SUMMARY

**Data Engineer** with 1.5+ years delivering production data pipelines on **Azure** and **GCP** for clients across **real estate analytics** and **enterprise data management**. Specialized in **ETL/ELT architecture**, **Medallion Architecture**, and data quality frameworks — currently building **Lambda Architecture** platforms and live market analytics deployed at **khangdo.dev**.

### TECHNICAL EXPERIENCE

#### Data Engineer (Project-Based Consulting)

*Kyanon Digital*

*Project: Azure Data Lakehouse · Wood Import & Export Analytics*

**Nov 2024 — Apr 2026**

*Ho Chi Minh City, Vietnam*

*Nov 2025 — Apr 2026*

- Architected a scalable **Medallion ETL pipeline** on **Azure Synapse Analytics** with **PySpark**, enabling automated ingestion and dynamic schema mapping for large-scale datasets.
- Engineered an event-driven ingestion trigger using **Azure Logic Apps** to automatically detect new files in **ADLS Gen2** and fire downstream **Synapse pipelines**, eliminating manual intervention; integrated **Azure Monitor** with custom alert rules for real-time pipeline health tracking and incident response.

*Project: Cloud Analytics Audit · SaaS Legal Tech Platform*

*Aug 2025 — Nov 2025*

- Resolved critical **pipeline failures** between **Fivetran** and **BigQuery**; led BI tool evaluation (Hex vs. Looker Studio) and revamped **Mixpanel** dashboards to improve product analytics coverage.
- Produced a comprehensive documentation suite (validation procedures, ADRs, end-user guides), accelerating team onboarding.

*Project: CMS Automation & QC Engineering · CTI Platform Migration*

*Jun 2025 — Aug 2025*

- Built an automated testing framework with **SeleniumBase** to automate authentication flows, bulk-generate CMS content, and orchestrate page tree workflows — drastically reducing manual effort.
- Developed **QC scripts** to validate URL integrity and cross-verify content consistency across environments, ensuring **zero data loss** during migration.

*Project: Modern ELT Pipeline & Data Quality Framework · Real Estate Analytics*

*Mar 2025 — May 2025*

- Engineered a containerized **Modern Data Stack (Docker + Airflow)**, achieving **100% automated daily ELT workflows** and full decoupling of operational from analytical databases.
- Designed **Star Schema** models with **dbt** and implemented a **Data Quality framework** validating **6 dimensions** (completeness, accuracy, consistency, timeliness, uniqueness, validity) with real-time alerting.

### SKILLS

<b>Languages &amp; Query</b>	Python, SQL, PySpark
<b>Data Engineering</b>	Apache Spark, Delta Lake, dbt, Apache Airflow, Fivetran
<b>Cloud &amp; Infra</b>	Azure (Synapse Analytics, Data Factory, ADLS Gen2, Monitor, Logic Apps), Google Cloud (BigQuery), Docker, MinIO
<b>Data Architecture</b>	Medallion Architecture, ETL/ELT, Star Schema, Data Lakehouse, Incremental Load, Schema Evolution
<b>Data Quality</b>	Data Observability, Anomaly Detection, Automated Validation (6 dimensions), Data Governance
<b>BI &amp; Analytics</b>	Apache Superset, Looker Studio, Hex, Mixpanel
<b>Other</b>	SeleniumBase, CI/CD, Git, REST APIs, Technical Documentation

### PERSONAL PROJECTS

#### FinFlow — Real-Time Fraud Detection & Revenue Analytics Platform

2026

*Python, Kafka, PySpark, Airflow, dbt, PostgreSQL, Redis, FastAPI, Streamlit, Docker, Terraform*

- Built a **Lambda Architecture** processing **1,000+ transactions/minute**; **PySpark Structured Streaming** with 3 fraud rules (velocity, amount spike, geo anomaly), scoring 0–3 with **Redis** alerts and 20% circuit breaker; batch ETL via **Airflow 2.9 + dbt** (staging → intermediate → mart) with idempotent ingestion and DAG chaining.
- Observability layer (DuckDB, schema drift, SLA freshness, **Slack** alerts); **9 services** containerized via Docker Compose; infrastructure via **Terraform**; **GitHub Actions CI**; **FastAPI + Streamlit** dashboard (heatmap, choropleth, merchant charts).

### EDUCATION

#### B.E. in Information Technology

Sep 2021 — Jul 2025

*Ho Chi Minh City University of Technology and Education*

### CERTIFICATIONS

**Academy Accreditation — Databricks Fundamentals, Databricks**

Mar 2026