

Tô Gia Bảo

DATA ENGINEER

Thu Duc, Ho Chi Minh City, Vietnam

+84 96 196 4353 | baokdl2226@gmail.com | ToGiaBaoKDL | togiabao

Summary

Data Engineer building Python and SQL pipelines, orchestration workflows, cloud data platforms, and conversational analytics tools. At Timo Digital Bank, delivered MCP-powered Slack integrations, operational automation, and analytics solutions with dbt and Lightdash. Hands-on with Airflow, Amazon EMR, AWS Glue, Amazon Athena, Amazon S3, PostgreSQL, MongoDB, Docker, and modern AI and lakehouse technologies, with a focus on reliable, maintainable, analytics-ready systems.

Experience

Timo Digital Bank by BVBank

Ho Chi Minh City, Vietnam

DATA ENGINEER

Aug. 2025 – Present

- Fulfilled on-demand data requests by using Amazon EMR and AWS Glue to sync MariaDB data from the Timo platform into the landing layer, applying PII masking based on field sensitivity before the curated layer, and using dbt to populate the analytics layer.
- Implemented a FastMCP server for Slack, grounding conversational analytics in 3 context sources: dbt docs, Lightdash dashboards, and semantic-layer metadata.
- Replaced 2 recurring manual checks for MongoDB backup validation and general ledger lookups with automated workflows.
- Built GenBI dashboards with Lightdash, dbt, and Airflow for DAG, Athena, and S3 monitoring; migrated the Timo Metrics observability stack from Loki/Grafana to SigNoz, consolidating logs and metrics.
- Built Tracking247 for the Customer team, accepting tracking codes in Streamlit and using Airflow to sync shipment status to MongoDB, centralizing updates and reducing manual follow-up.

Projects

Mini Lakehouse

GitHub

LAKEHOUSE & DATA PLATFORM

Jul. 2026

- Built an AWS lakehouse with Terraform-provisioned S3, IAM, KMS, ECR, and EMR Serverless, hosting Airflow on a private Oracle Cloud ARM instance and managing Iceberg tables through AWS Glue Data Catalog.
- Orchestrated Spark, dbt, and OCR workflows for GitHub activity analytics and arXiv document review, protecting Airflow and the inspector with Cloudflare Zero Trust Access and Tunnel.

VN Market Pulse

GitHub

AI RESEARCH & CONTENT PLATFORM

Jul. 2026

- Built a Vietnamese web-research pipeline that plans LLM queries, searches across Brave, Tavily, Serper, and Exa, deduplicates and reranks sources, extracts bounded context, and generates source-backed Facebook posts.
- Integrated PydanticAI with an OpenAI-compatible Timo gateway for typed output, validation, and usage tracking; added retrieval budgets, retries, and CLI and Streamlit interfaces.

Technical Skills

Programming Python, SQL, C++

Data Engineering Airflow, dbt, PySpark, Pylceberg, Polars, Dagster

Cloud & Infrastructure AWS (S3, Athena, EMR, Glue), OCI, Cloudflare Zero Trust, SigNoz, Terraform, Docker, Git

Lakehouse & Databases Apache Iceberg, AWS Glue Data Catalog, PostgreSQL, MariaDB, MongoDB

AI & Analytics Model Context Protocol (MCP), FastMCP, PydanticAI, semantic layer design, Lightdash, Streamlit, Slack

Education

University of Science, VNU-HCM (HCMUS)

Ho Chi Minh City, Vietnam

BACHELOR OF DATA SCIENCE | GPA: 3.78/4.00

2022 – 2026

Awards & Certifications

- 2025 **Top 25**, AIC Challenge
- 2025 **Recipient**, Ninety Eight Scholarship
- 2024 **Top 10**, RMIT Data Analytics Competition
- 2024 **Google Business Intelligence**, Professional Certificate
- 2024 **Google Data Analytics**, Professional Certificate

Publication

GalaxyAssistant: An Intelligent Assistant for Multimedia Event Retrieval

SOICT

CO-AUTHOR

2025

Languages

English TOEIC Listening & Reading: 910 | TOEIC Speaking & Writing: 350