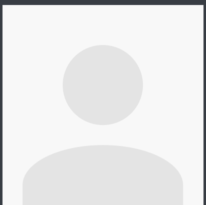

Pham Duc Thanh

DATA ENGINEER

☎ 0345678123

🎂 10/11/2003

✉ duybaoandinh@gmail.com

📍 Chung cu 4S Linh Dong, TP Hồ Chí Minh

🌐 github.com/thanhphamdata

About me

Data Engineer with strong skills in data pipeline design, ETL, and real-time streaming. Passionate about building scalable data infrastructures.

Education

Ho Chi Minh University of Technology

06/2022 - 06/2026 | Bachelor - Data Engineer

Skill

Core Skills

Docker (< 1 year) **Google Cloud** (< 1 year) **Kafka** (< 1 year) **Python** (< 1 year)

Spark (< 1 year) **SQL** (< 1 year)

Work Experience

01/2025 - 08/2025

HEALTHCARE ETL PLATFORM | Ho Chi Minh University of Technology

PROJECT:

Healthcare ETL Platform I 01/2025 - 08/2025

- **Description:** Built data transformations and data marts for medical record analytics.
- **Role:** ETL Developer
- **Responsibilities:**
  - The platform implemented automated ETL pipelines for structured and semi-structured healthcare data ingestion into Snowflake.
  - DBT was adopted for modular data transformation, model versioning, and documentation.
  - Data quality validation and lineage tracking were integrated to ensure compliance with healthcare data standards.
  - Audit mechanisms were included to monitor data freshness, transformation success rates, and anomaly detection.
  - Data marts were designed to support downstream predictive modeling and analytical workloads.
  - Continuous integration and delivery were handled through CI/CD pipelines for seamless deployment and testing.
- **Tech stack:** Python, DBT, Snowflake
- **Team size:** 3 members

03/2024 - 12/2024

RETAIL DATA WAREHOUSE | Ho Chi Minh University of Technology

PROJECT:

Retail Data Warehouse I 03/2024 - 12/2024

- **Description:** Designed ELT pipelines on GCP using Airflow and BigQuery for a retail analytics platform, reducing reporting latency by 60%.
- **Role:** Data Engineer
- **Responsibilities:**
  - The project established automated ELT workflows using Apache Airflow for multi-source data ingestion and integration.
  - BigQuery was utilized as the central data warehouse, optimized with partitioned tables and clustering to enhance query performance.
  - The architecture followed a modular pipeline structure separating extraction, staging, transformation, and serving layers.
  - DBT models were applied to create standardized fact and dimension tables for analytical reporting.
  - Data visualization was implemented using Looker Studio to provide sales and customer behavior insights.
  - Cloud Storage buckets were used for intermediate data staging and archival layers in GCP.
- **Tech stack:** Python , Airflow , BigQuery
- **Team size:** 4 members

07/2023 - 01/2024

LOGSTREAM | Ho Chi Minh University of Technology

PROJECT:

Streaming Engineer I 07/2023 - 02/2024

- **Description:** Implemented real-time log processing with Kafka Streams and Spark, enabling real-time analytics dashboards.
- **Role:** Your role in this project
- **Responsibilities:**
  - The system ingested logs from distributed services through Apache Kafka for high-throughput message streaming.
  - Spark Structured Streaming processed and aggregated data streams in near real-time to detect anomalies and generate metrics.
  - Data persistence was managed using PostgreSQL with time-series partitioning for efficient storage and retrieval.
  - Avro schemas were introduced to maintain schema consistency across microservices and message producers.
  - Monitoring and alerting were configured using Prometheus and Grafana dashboards to track system performance.
  - The overall pipeline achieved high fault tolerance and horizontal scalability through containerized deployment.
- **Tech stack:** Kafka , Spark , PostgreSQL.
- **Team size:** 5 members

Certificate

Google CLoud Data Engineer

02/2025 | **Google**