

NGUYEN DANG HONG HUY

Data Engineer

✉ ndhhuy1320@gmail.com 📍 Hochiminh City
☎ 0905263087 🌐 ndhhuy2k 📷 ndhhuy2k



Keywords: Data Warehouse, ETL/ELT, Distributed Systems, Data Modeling, Lakehouse Architecture, Medallion Architecture, Microservices, CI/CD.

SUMMARY

Data Engineer with 4+ years of experience designing and building scalable data platforms, data warehouses, and large-scale ETL/ELT pipelines. Experienced in processing billions of records using modern Big Data technologies including Apache Spark, Airflow, Kafka, and cloud-based data architectures. Skilled in implementing **Medallion Data Platforms (Bronze–Silver–Gold)**, optimizing distributed data pipelines, and building enterprise reporting systems that support high-performance analytics and data-driven decision making. Passionate about **distributed systems, real-time data processing, and modern Data Lakehouse architectures**, with a strong focus on performance optimization, scalability, and reliable data infrastructure.

WORK EXPERIENCE

📌 Galaxy Pay – Data Engineer

📅 January 2026 – Present | Ho Chi Minh City

- Architected and developed a modern **Medallion Data Platform** to support scalable analytics and high-volume data processing.
- Designed end-to-end **Bronze–Silver–Gold** architecture using **Amazon S3, Apache Spark, Delta Lake, Apache Airflow, and Snowflake**.
- Built high-performance ETL/ELT pipelines for data ingestion, validation, transformation, and serving.
- Developed real-time **Data Quality & Observability dashboards** using **FastAPI + ReactJS** for monitoring pipeline health and SLA tracking.
- Optimized processing workflows, reducing **data latency by 95%** and lowering **infrastructure costs by 60%**.
- Collaborated with cross-functional teams to enable self-service analytics and scalable reporting capabilities.

EDUCATION

• Master Computer Science - HCMUS

📅 2023-Ongoing GPA: 8.20

• B.S. Information Systems - HCMUS

📅 2018-2022 GPA: 8.33

TECHNICAL SKILLS

- **Data Architecture:** Data Lake & Lakehouse, Data Warehouse, Data Modeling, ETL/ELT & Orchestration, Data Governance, Data Pipeline
- **Big Data & Processing:** Spark Hadoop, HDFS, Kafka, Airflow, Databricks, Delta lake, Apache Iceberg
- **Storage:** S3, HDFS, MinIO; Oracle, PostgreSQL, MySQL; MongoDB, Cassandra
- **Programming:** Python, SQL, JS
- **Tools & Visualization:** PowerBI, Superset


HDBank – Data Engineer

 17 March 2024 – December 2025 | Ho Chi Minh

City

- Designed and implemented **enterprise ETL/ELT pipelines** for Data Warehouse and reporting systems.
- Built a dynamic **Report Builder platform** enabling on-demand report creation without redeploying.
- Developed **Excel/PDF generation services, storage service, and cache service** for large-scale reporting.
- Implemented advanced **permission management** for modules, menus, APIs, and report categories.
- Migrated multiple legacy SQL procedures to scalable **Spark jobs**, improving performance by **30–70%** depending on workload.

Fujinet Systems – Data Engineer & Full-stack Developer

 17 Oct 2021 – Oct 2023 | Ho Chi Minh City

- Built a real-time sales analytics platform using **Kafka, Spark Streaming, and Cassandra** for streaming data processing.
- Designed and validated batch-processing pipelines on **Hadoop/HDFS** for large-scale data workloads.
- Developed inventory forecasting and analytics features using **Python** integrated with **AWS Cloud** services.
- Built full-stack applications using **ReactJS, FastAPI, NodeJS, and PostgreSQL**.
- Worked in Agile/SCRUM teams of 2–5 members, collaborating closely with developers, QA engineers, and business stakeholders.

SOFT SKILLS

Teamwork · Problem-solving · Ownership · Communication · Leadership · English · Multitasking

- **Cloud:** AWS (ECS, EKS, RDS, Kinesis, Glue); Databricks
- **Devops:** Docker, Git, CI/CD (GitHub Actions, Jenkins)

PROJECTS

Optimizing ETL Processes for Data Warehouse Using Big Data Technologies

Tech: Spark, Airflow, Oracle, Python
Designed and optimized ELT pipelines for enterprise DWH using Spark. Migrated legacy Oracle ETL to scalable Spark jobs and automated workflows with Airflow.

Achievements: Reduced ETL runtime by 40–60%, improved data reliability, and lowered compute cost by ~25%.

Real-Time Sales Analytics – Streaming Data Pipeline

Tech: Kafka, Spark Streaming, Cassandra, Python, VueJS
Built a real-time data pipeline for processing sales events with low latency. Implemented Kafka → Spark Streaming → Cassandra architecture and developed dashboards for real-time monitoring.

Achievements: Achieved 3–5s end-to-end latency and improved real-time sales visibility for business teams.

CERTIFICATIONS

Databricks

- AWS- Azure Databricks Platform Architect - Issued Oct 2025
- Databricks with AI Agent Fundamentals - Issued Oct 2024

Google Certificates

- Google Data Analytics Specialization - Issued Jan 2023