

# Phan Thanh Hải

## Data Analysis Supervisor – Technical-oriented

(+84) 912 738 824 • phanthanhhai2411@gmail.com

[linkedin.com/in/haiphan24112000](https://www.linkedin.com/in/haiphan24112000) • [github.com/haiphan2000](https://github.com/haiphan2000)



TM

### SUMMARY

As a highly creative graduate, I am passionate about building efficient data systems and handling data pipeline challenges. With **2 years** of hands-on projects and self-directed learning experience in **Python, SQL** and **Power BI** in **data visualization models, ETL processes**, and running scalable **data pipelines**, I have a proven core ability to contribute my skills and enthusiasm to the innovative data team at **Circle K**.

### SKILLS AND ADDITIONAL

**Languages.** Conversational Proficiency in English (TOEIC LR: 775); Basic Communication in Korean

**Programming Languages.** Python, Bash (Unix shell)

**Relational DB Management System.** PostgreSQL, MySQL, Microsoft SQL Server

**Apache Software Foundation Projects.** PySpark (Python version of Apache Spark), Airflow

**Other Development Tools and Platforms.** Microsoft Power BI, Azure Data Factory, Azure Synapse,

**Soft Skills.** Problem Solving, Teamwork, Communication, Adaptability, Specialization and Generalization

### PROJECTS

#### Real Estate Data Pipeline

10/2023 – 01/2024

**GitHub Link:** [github.com/haiphan2000/real-estate-prices](https://github.com/haiphan2000/real-estate-prices)

- Created a web scraper for *Immo Scout 24*, extracting and structuring real estate data for analysis.
- Utilized **PySpark** to perform data processing, cleaning and modeling.
- Orchestrated a scalable data pipeline using **Airflow**, managing primary tasks including web scraping, **ETL** processes and machine learning model training for efficient data processing.
- Implemented data versioning with Delta tables and integrated ML model predictions into **Druid**.

#### Simple End-To-End Azure Data Engineering

05/2023 – 06/2023

**GitHub Link:** [github.com/haiphan2000/simple-end-to-end-azure-data-engineering](https://github.com/haiphan2000/simple-end-to-end-azure-data-engineering)

- Established seamless connections between **Azure Data Factory** and the local **Microsoft SQL Server**.
- Configured a copy pipeline within **Azure Data Factory**, transferring tables to **Azure Data Lake's**.
- Developed **medallion data lake** architecture (Bronze, Silver, Gold) using **Azure Databricks** and **PySpark**.
- Integrated **Azure Synapse** to serve Gold layer data via serverless SQL views for **Power BI** reporting.

#### Global Superstore Sales Analyst

04/2023 – 05/2023

**GitHub Link:** [github.com/haiphan2000/superstore-sales-dashboard](https://github.com/haiphan2000/superstore-sales-dashboard)

- Designed and optimized **PostgreSQL database schema**, implementing **normalized data models** and **indexing** strategies for efficient sales **data management**.
- Developed **SQL queries** with window functions and CTEs to analyze sales, profit, and returns.
- Built interactive **Power BI** dashboard to monitor **key business KPIs** (total sales, profit margin, product returns, customer segments, and regional performance), reducing manual reporting time.
- Conducted advanced analytics using **DAX**, including price elasticity analysis, RFM segmentation, and ABC-XYZ classification to uncover actionable customer and inventory insights.
- Used **Power Query** to transform raw sales data and implement a star schema model for streamlined reporting in **Power BI**.
- Applied **time-series analysis** to compare previous, actual, and forecast month data against a 6-month sales trend, supporting data-driven planning.

### HOBBIES

Sharing my learning and experiences in data field through writing. Featured work on LinkedIn:

- Online Assessment HackerRank Report – Database Questions For Data Engineer(ing)*. A **PostgreSQL** – focused guide with step-by-step solutions for HackerRank database assessments.
- Simple 9-Step Framework for Designing BI Dashboards*. A practical **Design Thinking**-based guide to crafting impactful and user-friendly dashboards.
- Slide Design Guideline in PowerPoint*. A clean, data-focused presentation guideline based on **IBCS standards** to enhance clarity and communication in data analysis.

### EDUCATION

#### University Of Information Technology (UIT)

Bachelor of Computer Science

GPA: 8.52/10.0

Thu Duc City, HCMC

2018 – 2023