

# TRAN GIA PHU

Email: [trangiaphu25092003@gmail.com](mailto:trangiaphu25092003@gmail.com)

Linkedin: [linkedin.com/in/gia-phu-1b8427281](https://www.linkedin.com/in/gia-phu-1b8427281)

Phone: +(84) 898 429 072

Portfolio: [GP Portfolio](#)

DATA ENGINEER

## EDUCATION

### SGU-Saigon University

2021-Present

*Bachelor of Engineering Information Technology*

- **Relevant Coursework:** Databases, Database Management Systems, Data Structures and Algorithms, Object-Oriented Analysis and Design (OOAD), Analysis and Design of Information Systems, Algorithm Analysis and Design, Data Visualizations, NLP
- **Languages:** English (Currently studying TOEIC 800+ class (entrance test 740))

## SKILLS

- **Programming Languages:** Python, Java, SQL
- **Data Analyst:** Seaborn, Pandas, Matplotlib, Tableau, PySpark, Dash, Plotly
- **Database Management Systems:** MySQL Workbench, PgAdmin4, MongoDB Compass
- **DataOps skills:** Docker, Git, Unix/Linux
- **Web Scraping:** Scraping Data (Requests, Selenium), Playwright to automation test, scraping data
- **Data Engineering:** Dagster, Apache Spark, Apache Hadoop, Apache Airflow,

## EXPERIENCES

### International Trade Council

*Data Engineer*

HCMC, Viet Nam

Jul. 2024-Dec. 2024

- **Responsibilities:** Maintain a robust data system with gigabytes of data processed daily, scrape various data sources up to 10 million records, and process these sources into the company's unified system.
- **Technologies:** Selenium, MySQL, PostgreSQL, BeautifulSoup, Apache Airflow, Git, Apache Spark, Docker, Minio
- **Team size:** 7

## PROJECTS

### Recovered Health Website

*Backend Developer*

Jun-Jul. 2024

*Freelance*

- **Website:** <https://recoveredhealth.com/>
- **Responsibilities:** Primary responsibilities include completing the website development using the Laravel framework, ensuring robust and scalable solutions. You will work on improving the user interface (UI) of the website to enhance user experience and engagement. Additionally, you will create and maintain APIs for seamless integration with the database, enabling efficient data exchange and interaction. You will also utilize Playwright to pull data from the website, ensuring accurate and up-to-date data collection for analysis and reporting.
- **Technologies:** Laravel, PhpMyAdmin, MySQL, NodeJS, Playwright, HTML, CSS

### Pandemic ETL Pipeline

*ETL (Extract-Transform-Load) data pipeline with the Covid dataset*

Sep - Oct. 2023

- **GitHub Project:** <https://github.com/GiaPhus/Pandemic-Data-ETL>
- **Responsibilities:** Develop end-to-end data platform for Covid-19 following Lambda architecture. Building ETL pipelines for batch processing with Apache Spark. Containerized the application with Docker and orchestrated assets using Dagster DAGs.
- **Serving:** Develop a dashboard using Plotly and Dash to visualize the progression of Covid-19 death cases over time globally. Additionally, create interactive components to display the total number of confirmed deaths, recoveries, and other relevant metrics for each region.
- **Team size:** 1
- **Technologies:** Dagster, Minio, Plotly, Docker, MySQL, PostgreSQL, Apache Spark

- **Airbnb NewYork Visualization**

Apr – May. 2024

*Visualization Airbnb Dataset*

- **GitHub Project** : <https://github.com/GiaPhus/AirBnb-Visualization>
- **Responsibilities**: Utilize visualization techniques to depict insights about Airbnb listings in New York City.
- **Serving**: Utilize Plotly and Dash to create an interactive web application that allows users to interact with charts.
- **Teamsize**: 1
- **Technologies**: ,Plotly,Docker,Dash,Seaborn,Apache Spark,Bokeh,Matplotlib

## **CERTIFICATES**

---

SQL (Advanced) Certificate(hackerrank)

Python Certificate(hackerrank)

Java Certificate(hackerrank)

Fundamentals of Visualization with Tableau(Coursera)

Fudamentals DataEngineering (AIDE)