PROFILE

✉ trung280302@gmail.com

☎ 0865957707

📍 Linh Trung, Thu Duc, Ho Chi Minh

EDUCATION

UNIVERSITY:
- University of Information Technology (UIT VNU-HCM)
- Major: Information Technology
- Expected Graduation: 3/2025
- GPA: 3.3

TOEIC CERTIFICATE:
- Listening and Reading: 865
- Speaking and Writing: 290

SKILLS

PROGRAMMING:
- Python, C++ , SQL

DATA
- Spark - Kafka - DBT - Airflow - Docker
- ETL - OLAP - OLTP
- Data Analysis - Web Scraping - Data Preprocessing - Data Modeling

ADDITIONAL INFORMATION

- LinkedIn:
  https://www.linkedin.com/in/phutrungnguyen283/
- GitHub:
  https://github.com/boo283/
- Facebook:
  https://www.facebook.com/ngphtrungboo/
- Hackerrank:
  https://www.hackerrank.com/profile/Boo21521592

NGUYỄN PHÚ TRUNG
DATA ENGINEER

OBJECTIVE

Having recently completed my studies in IT, I am seeking a Data Engineer position where I can enhance my knowledge and gain practical experience. With a strong passion for data, I also look forward to making meaningful contributions to the company's operations in the long run.

EXPERIENCE

💼 **WEB CRAWLER**   //   **9/2024**   -   **Present**

Freelancer at Global Tech

Assigned Tasks:
- Developed **Python** scripts to crawl, extract, and transform data from multiple sources based on predefined structures.
- Researched and learned new tools and technologies as Apache Nifi, Neo4j, SparQL.

Technologies and skills:
- Python: Selenium, Pandas
- Data Preprocessing, Web Scraping

💼 **TRAFFIC VIOLATION DETECTION**   //   **5/2024**   -   **7/2024**

Members: 4 - Leader

My Contributions:
- Configured pretrained **YOLOv10** for custom training with new labels.
- Designed two **Kafka** topics for storing, sending and displaying frames after prediction.
- Configured **Spark** to retrieve frames from the first topic, detect violations using a user-defined function and toward the results to the next topic.
- Built a pipeline for automatically storing detected frames in **MongoDB.**

Technologies and skills:
- Python
- Kafka, Spark, MongoDB

Source:
- *https://github.com/boo283/Traffic_Violation_Detection*

💼 **DAILY AUTOMATED DATA PIPELINE**   //   **11/2024**   -   **Present**

Members: 4

My Contributions:
- Developed an API-based web crawler using **Python-Selenium** and **FastAPI** to collect job data from various sources.
- Stored and processed data in **ClickHouse**, ensuring efficient querying and analysis.
- Built a structured data model using **DBT,** transforming raw data into meaningful insights.
- Automated the daily data collection and processing workflow using **Apache Airflow** and **Docker.**

Technologies and skills:
- Python, SQL
- ClickHouse, DBT, Airflow, Docker
- ETL, Data Preprocessing