CAP HUU ANH TRI

INTERN DATA ENGINEER

CONTACT

📞 +84941218057

✉️ trihx2003@gmail.com

📍 Thu Duc city, Ho Chi Minh city

🔗 https://github.com/CapHaTri

EDUCATION

University of Information Technology (UIT) (Sep 2021 - Present)

- Bachelor of Computer Science (Expected graduation: November 2024)
- GPA: 3.23/4.00
- TOEIC : 730/990

SKILLS

Technical skills:

- **Programming Languages :**
  - Python
  - Javascript
  - SQL
- **Database Management:**
  - mySQL
  - postgreSQL
- **Data Visualization:**
  - Apache Superset
- **Big Data:**
  - Apache Hadoop
  - PySpark
  - Kafka
  - Airflow
- **Microsoft Excel**
- **Tools:**
  - Docker
  - Web Scraping
- Languages:
  - Vietnamese
  - English

Soft skills:

- Effective Communication
- Teamwork
- Critical Thinking & Problem Solving
- Time Management

ABOUT ME

As a fourth-year Computer Science student with a solid foundation in Data Engineering and Big Data technologies, I am eager to secure an internship. Though lacking practical experience, I am hardworking, willing to learn, and made an effort to contribute effectively to the team.

- **Short-term Goal:** Secure an internship as a Data Engineer to gain hands-on experience and apply my knowledge to real-world projects.
- **Long-term Goal:** Develop into a proficient Data Engineer, mastering advanced tools and contributing to impactful data-driven initiatives.

PROJECT

REAL TIME ANOMALY DETECTION IN WEB SERVER   Link

Real-time detection and visualization of anomalies in web server logs.

- **Real-time Data Ingestion**: Stream web server logs from **MySQL** using **Apache Kafka**
- **Efficient Storage:** Use **Apache Pinot** for storing and querying time series data.
- **Anomaly Detection**: Implement a machine learning model (**Autoencoder with Regression Model**) to predict anomalies.
- **Interactive Visualization:** Display results and insights using **Streamlit**.
- **Docker Compose Setup:** Streamlines deployment and management of services including Apache Kafka, Apache Pinot, Streamlit, and MySQL

Technologies used: Python, MySQL, Apache Kafka, Apache Pinot, Streamlit, Docker

AUTOMATED ETL AND VISUALIZATION OF PREMIER LEAGUE MATCH RESULTS   Link

Utilize ETL to extract match results for analysis and visualization

- **Web Srapy:** Using **BeautifulSoup** to get data from EPL source, store in **MySQL** as datalake
- **Automated ETL Pipeline**: Using **Apache Airflow** to automate the process of collecting data into **MySQL**, extracting it from **MySQL**, transforming into a usable format, and loading into a **PostgreSQL** database. After that, using **Smtplib** to email about results
- **Data Visualization:** Leveraging **Apache Superset** to create interactive dashboards and visualizations for analyzing Premier League match results.
- **Docker Compose Setup**: Streamlines deployment and management of services including Apache Airflow, PostgreSQL, Redis, MySQL, and Apache Superset

Technologies used: Python, BeautifulSoup, MySQL, Apache Airflow, Smtplib, PostgreSQL, Docker

BOOKING MEDICINE WEB APPLICATION PROJECT   Link

Creates an online prescription system.

- **Web App Design**: Design admin and user interfaces for online prescription booking
- **UI Development:** Build interfaces using **ReactJS** for a modern user experience.
- **Database Setup**: Establish a **MySQL** database to store prescription data efficiently.
- **Backend Implementation**: Develop backend functionality using the **Express framework** and **RESTful APIs.**

Technologies used : JavaScripts, MySQL, ReactJS, NodeJS, RESTful APIs, Express framework