

Truong Tien Anh

+ (84) 982 619 731 | truongtienanh16@gmail.com | [linkedin.com/in/trgtanh](https://www.linkedin.com/in/trgtanh) | github.com/trgtanh04

EDUCATION

University of Science - VNUHCM

Bachelor of Science in Data Science and Computer Science

Oct. 2022– Present

Current GPA: 3.7/4.0

CAREER OBJECTIVE

I'm a third-year Computer Science student passionate about data science, machine learning, and big data systems. Seeking an internship or job opportunity as a Data Scientist or Data Engineer to apply my skills to real-world problems and grow through hands-on experience.

SKILLS

Programming Languages: Python, SQL, JavaScript, C/C++

Data Science & ML: Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, Matplotlib, Seaborn, Plotly

Data Engineering: Apache Hadoop, Spark, Airflow, Kafka, PostgreSQL

Cloud & DevOps: AWS (S3, Glue, Redshift, Athena), Docker, Git/GitHub, MongoDB Atlas, Neon Cloud

CERTIFICATION & AWARD

TOEIC: 810/990

Dong Hanh Scholarship - 2024

PROJECTS

End to End MovieDB

February 2025 – Present

- **Key Technologies:** Python, Apache Airflow, Spark, Kafka, HDFS, PostgreSQL, Streamlit
- **Description:**
 - * Crawled movie data from a film website and transformed it into structured JSON format.
 - * Stored raw JSON data in a distributed Data Lake using HDFS.
 - * Used Apache Spark to read from HDFS, perform data cleaning, and basic analysis, with Kafka used to trigger the ETL process.
 - * Created fact and dimension tables, then loaded cleaned data into PostgreSQL.
 - * Deployed PostgreSQL database to the cloud using Neon for easier access and scalability.
 - * Built a prediction model to estimate movie prices based on key features.
 - * Developed a Streamlit web app for interactive movie recommendations based on user preferences.
- **GitHub:** <https://github.com/trgtanh04/End-to-End-MovieDB-Data-Engineering>

Mobile AWS Pipeline Engineering

March 2025 – Present

- **Key Technologies:** AWS (S3, Glue, Athena, Redshift), Docker, Apache Kafka, Apache Spark, Python, PostgreSQL
- **Description:**
 - * This project was created to achieve a 30–50% performance boost in data processing compared to the project above by leveraging Kafka and AWS, and to make the system more scalable.
 - * Crawled mobile phone data from Mobile City and ingested it into Kafka staging.
 - * Logged incoming data to PostgreSQL for monitoring and backup.
 - * Used Apache Spark to consume Kafka streams, process data, and upload to AWS S3.
 - * Triggered AWS Glue jobs to transform data and store it in Athena and Redshift.
 - * Built dashboards to visualize insights from the processed data.
- **GitHub:** <https://github.com/trgtanh04/Mobile-AWS-Pipeline-Engineering>

Movie Recommender System

December 2024 – Present

- **Key Technologies:** Python, Matplotlib, Scikit-learn, Streamlit
- **Description:**
 - * Loaded and explored a movie dataset from Kaggle.
 - * Performed data cleaning, preprocessing, and feature extraction.
 - * Applied cosine similarity to compute movie-to-movie similarities.
 - * Built a recommendation engine using collaborative filtering.
 - * Developed a Streamlit app to allow users to get personalized movie suggestions.
- **GitHub:** <https://github.com/trgtanh04/Movie-Recommendation-System>