TRẦN DUY TUẤN
DATA ENGINEER

MY CONTACT

- (+84) 762162640
- duytuan2412k4@gmail.com
- 24/01/2004
- Tien Thang, Ha Noi
- github.com/Trantuan24

EDUCATION

Information Technology
2022 - Present
Thuy loi University
GPA: 3.51

SKILLS

- **Programmin**g: Python (numpy, pandas, asyncio, requests), Playwright, FastAPI, REST APIs
- **Data Engineering**: Apache Airflow, Apache Kafka, Apache Flink, ETL/ELT, CDC, SSIS
- **Databases**: PostgreSQL, MySQL, SQL Server, DuckDB, SQLAlchemy, JSONB/JSONL
- **DevOps & Cloud**: Docker, Docker Compose, Linux, Git, Google Cloud Platform (GCP)
- **Monitoring & Reliability**: Prometheus, Grafana, Logging, alerting, retries, performance tuning

CERTIFICATIONS

- IBM Data Engineering Professional Certificate – Coursera 🔗
- IBM Data Warehouse Engineer Professional Certificate — Coursera 🔗
- SQL (Advanced) Certificate - Hackerrank 🔗

LANGUAGE

English

ABOUT ME

Aspiring Data Engineer Intern with hands-on experience designing and deploying production-grade data pipelines using Python, Airflow, Docker, and PostgreSQL. Skilled in transforming messy web/API data into reliable datasets, and optimizing pipeline reliability with monitoring, CDC, and automated recovery. Fast learner, strong communicator, and adaptable team player.

PROJECTS

May 2025 - July 2025
JobInsight Data Pipeline (Personal Project)

- Built a robust ETL pipeline for job market data with **Python, Airflow, Playwright, Docker, PostgreSQL, and DuckDB**, supporting **automated web crawling, data standardization, star schema modeling, and SCD Type 2.**
- Developed **resilient scraping** with **anti-bot measures**, **CAPTCHA handling, concurrent workers**, and **circuit breaker logic**.
- Delivered a **high-performance data warehouse** (PostgreSQL COPY ~170 records/sec, 7.5s ETL, 26-field validation) with **CDC audit trails, backup/recovery, Docker orchestration,** and **monitoring**.

Source: github.com/Trantuan24/JobInsight_Data_Pipeline

Aug 2025
StreamPulse v1 – Single-Source Metrics (Personal Project)

- Developed a **production-grade real-time streaming pipeline** with **Kafka, Flink (Java), Redis,** and **FastAPI**, processing **10K+ events/sec** with **sub-3s latency** (**Dockerized microservices**).
- **Implemented event-time processing, watermarks, exactly-once delivery, RocksDB state,** and **automated checkpointing** for **fault tolerance** and **zero data loss.**
- **Built end-to-end monitoring** with **Prometheus, Grafana, Alertmanager,** and **JMX exporters, achieving 99%+ system reliability** and **proactive alerting.**

Source: github.com/Trantuan24/StreamPulse-v1-Single-Source-Metrics

GOAL

Short-term (3–6 months)

- Excel in Data Engineering internship: deliver reliable ETL/ELT pipelines, document processes clearly.
- Achieve strong performance and target conversion to part-time/full-time.

Long-term (1–3 years)

- Grow into a proficient Junior Data Engineer building production-ready data systems.
- Deepen expertise in cloud data platforms (AWS / Azure / GCP), and achieve cloud certification.