



# TRAN QUOC THANG

## DATA ENGINEER

### EDUCATIONS

- **Ton Duc Thang University – Ho Chi Minh City, Vietnam**

Bachelor of Science in Statistics and Mathematics

Graduated: 2024

Relevant coursework: Probability, Statistical Modeling, Data Analysis, Linear Algebra, Calculus

- **Cambridge English B1 Certificate**

Issued by Cambridge Assessment English

Demonstrates intermediate English proficiency (CEFR Level B1)

### 🏆 AWARDS & ACHIEVEMENTS

- 🏆 **Completed SQL (Advanced) Certification – HackerRank**

Verified advanced SQL skills with hands-on challenges and top-tier performance on HackerRank.

- 🏆 **Completed Data Engineer Bootcamp – Cole Community (2024)**

Gained practical experience building real-world data pipelines using Airflow, AWS S3, SQL, and Python.

### CONTACT INFO

📞 0908221175

✉️ [thangtranquoc1805@gmail.com](mailto:thangtranquoc1805@gmail.com)

🌐 [github.com/thangtran1805](https://github.com/thangtran1805)

📍 Ho Chi Minh City

### SUMMARY

Motivated and detail-oriented Data Engineer with strong proficiency in Python and SQL, and a solid understanding of Big Data concepts and distributed systems. Experienced in designing and deploying end-to-end ETL/ELT pipelines using tools like Apache Airflow, PySpark, and DuckDB. Proficient in working with Amazon S3 for data storage and integration, and in managing source code and collaboration via Git and GitHub. Demonstrated ability to transform raw data into structured analytical models through hands-on projects in stock market and YouTube trending analytics. Passionate about building scalable, cloud-ready data pipelines to support data-driven decision-making.

### PROJECTS

- 2015 – 2017

- 📁 **Stock Market Data Pipeline**

Designed a fully automated ETL pipeline to collect daily stock data from multiple APIs (Alpha Vantage, Polygon.io, SEC).

Processed and normalized raw data using Pandas, modeled into dim/fact tables using DuckDB.

Stored raw data in MySQL, transformed data in Amazon S3, and scheduled workflows using Apache Airflow.

Tools: Python, Airflow, DuckDB, MySQL, boto3, SQLAlchemy.

- 2017 – 2019

- 📁 **YouTube Trending Data Warehouse**

Built an ETL pipeline to clean and transform YouTube trending data from CSV and JSON formats.

Normalized schema into dimensional tables: dim\_channel, dim\_video, dim\_category, fact\_views.

Handled null values and foreign key constraints, and automated workflows using Airflow.

Tools: Python, Airflow, DuckDB, Pandas

- 2019 – 2021

- 📁 **Retail Sales Data Analytics with SQL**

Designed and implemented a mini Data Warehouse from a raw retail dataset using PostgreSQL.

Created normalized dim\_customer, dim\_product, and fact\_sales tables to support OLAP-style queries.

Wrote complex SQL queries to answer business questions such as sales by region, top-performing products, and customer segmentation.

Tools: PostgreSQL, SQL, ERDPlus, dbdiagram.io

### 🔧 TECHNICAL SKILLS

- Programming: Python, SQL, Bash
- Data Engineering: Airflow (DAGs, scheduling), ETL/ELT pipelines, Data Warehousing
- Databases: PostgreSQL, MySQL, DuckDB, SQLite
- Big Data Tools: PySpark, SparkSQL
- Cloud & Storage: AWS S3, EC2 (basic), boto3
- Data Modeling: Dimensional modeling (Star Schema), ERD
- Data Processing: Pandas, NumPy, SQLAlchemy
- Version Control & CI/CD: Git, GitHub Actions
- Other Tools: VS Code, Jupyter Notebook