

# TRIET LE

ben.le.data@gmail.com • (+84)-825-717-449

• LinkedIn: [://trietbenle](https://www.linkedin.com/in/trietbenle) • Ho Chi Minh City, Vietnam

## SKILLS

---

- Language: Python, Scala, JavaScript, SQL, Bash, HTML, CSS
- Technology: BigQuery, DuckDB, dbt, Pandas, Airflow, Spark, GCP, Docker, Ansible, Vagrant, Terraform, Flask
- Skills: ETL, Data Warehouse, Business Intelligence, Data Modeling, Cloud, RDBMS, Machine Learning, Functional Programming

## EXPERIENCE

---

### Leading Wealth Management Firm (Name Withheld)

Kansas, US (Remote)

Data Engineer (Part-time)

Jun 2024 - Nov 2024

- Migrated a legacy on-premise wealth management data platform to **Azure**, improving scalability and reducing operational overhead.
- Designed and implemented ETL pipelines copying operational data from **Redshift** and **S3** to **Azure Data Lake** using **Azure Data Factory**.
- Applied the Medallion architecture for data processing with **Spark** on **Databricks**
- Developed APIs with **.NET** to expose system data from the Silver layer for seamless integration with downstream applications

### Holistics

Ho Chi Minh, Vietnam

Data Engineer

Aug 2022 - Nov 2024

- Redesign a legacy **Data Warehouse** architecture with various modeling paradigms(Dimensional Modeling, Event Schema) using **dbt**, enabling scalable support for advanced business analytics use cases.
- Develop and orchestrate **~100 ETL data pipelines** with **Prefect** from **~20** data sources to a central Data Warehouse, providing robust analytics insights and capabilities.
- Implement a comprehensive **Data Catalog** in **dbdocs** with over **100** tables to streamline data accessibility, resulting in a **40%** increase in weekly active SQL Data Warehouse users.
- Design and implement a data usage monitoring system with **Holistics** to pinpoint expensive query run on production. The system ultimately reduces redundant **BigQuery** usage by **70%** compared to its peak.
- Setup **Snowplow event tracking** to measure product analytics, offering visibility into key success metrics and help Product team make data-informed decisions to accelerate iterative development.
- Build a **data observability system** in **Grafana** for **~1500** data products (dashboards, data models and data pipelines). This system has enabled prompt responses to data pipeline failures with **60%** decrease in failure rates and reducing the mean time to recovery
- Collaborate with the DevOps team to automate **GitHub Actions CI/CD** pipelines for internal data projects, reducing manual deployments and enabling faster, error-free releases.

Technical Writer

Aug 2023 - Jan 2024

- Managed a team of **3** technical writers and data analysts, publishing **20+** analytics best practice articles and tutorials. These articles enhanced customers understand on the Holistics capabilities which increases product adoptions.
- Researched and documented BI use cases to enhance user understanding and adoption of analytics solutions.

### Troy University

Troy, Alabama, US

Computer Science Tutor and Lab Assistant

Jul 2021 - May 2022

- Tutored college students in Python, C++, Java with more than 200+ hours of coding lessons
- Assisted instructor in the daily assessment of student progress and inspected student assignment

### Kids Code Camp

Ha Noi, Vietnam (Remote)

Python Instructor

Jul 2021 - Sep 2021

- Taught introductory Python for a group of ~150 international secondary/high-school students
- Helped students develop programmatical thinking and inspired the new generation to pursue a career in tech

## EDUCATION

---

**Troy University, Troy, AL**

*B.S Computer Science*

**Graduation Date: May 2022**

***GPA: 3.5/4.0***

## CERTIFICATES

---

- Data Engineer with Python (DataCamp)
- DataTalks Data Engineer Zoomcamp
- Big Data with Spark and Hadoop (IBM)
- Python for Data Science and Machine Learning Bootcamp (Udemy)
- SQL Advanced (HackerRank)