THANH VU

Data Engineer | SQL · Python · Microsoft Fabric · ETL

Lancaster, PA | vulethanh07112002@gmail.com | (717) 869-9502 | linkedin.com/in/thanh25 | github.com/thanhvu911 | Portfolio

SUMMARY

Data Engineer and Microsoft Certified Fabric Data Engineer Associate who builds ETL/ELT pipelines, Medallion warehouses, and star-schema models with SQL, Python, and Microsoft Fabric, automating data loads that save teams 15+ hours a month.

EDUCATION

Franklin & Marshall College — Lancaster, PA

Bachelor of Arts in Computer Science | *GPA: 3.72/4.0* — May 2026

- Magna Cum Laude; Honors List (5 semesters); Dean's List (Spring 2022)
- Coursework: Data Science, Algorithms, Artificial Intelligence, Machine Learning, Statistical Modeling, Theory of Computer Science

EXPERIENCE

Data Analyst Fellow | EDC Lancaster County (Center for Regional Analysis), *Lancaster, PA* — Aug 2026 – Present

- Automated ingestion of 25+ federal economic series from BLS and BEA public APIs into a star-schema Power BI semantic model with 20 years of history and daily scheduled refresh, saving the team **15+ hours per month**.
- Developed a Google Apps Script web app on a Google Sheets data store with cross-sheet imports, nightly refresh jobs, sheet locking, and a Gemini AI assistant, letting **6+ non-technical staff** maintain data without code.
- Delivered **4+ dashboards and data products** for partner organizations and internal teams, designing data-intake structures, dataset documentation, and data dictionaries that standardize how data is submitted.

Data Science Intern | KCC Capital Partners, *Sheridan, WY* — Jan 2026 – May 2026

- Integrated data on **100+ precious-metal and mining companies**, markets, and transactions into one dataset and enriched **500+ contact records** with Seamless.AI, improving data quality for deal sourcing.
- Screened mandate data with the firm's built-in AI tools, shortlisting **100+ likely buyers** for outreach.
- Launched automated weekly deal-team reports and dashboards, replacing manual spreadsheets and saving **5+ hours per week**.

Data Engineer Intern | KPIM Joint Stock Company, *Hanoi, Vietnam* — Jan 2021 – Apr 2021

- Engineered ETL pipelines in Python (Pandas, NumPy) and SQL Server to automate legacy client data workflows, reducing processing time by **30%**, eliminating manual entry errors, and feeding validated Power BI reports.

PROJECTS

Can Mail Data Predict Housing Vacancy? | *Microsoft Fabric (Lakehouse, Dataflow Gen2), Power BI, DAX* — Sep 2026

- Architected an end-to-end Fabric pipeline that lands USPS quarterly and Census ACS files in a Lakehouse, consolidates them with Dataflow Gen2 (Power Query M), and serves a semantic model covering 6 counties and 15 years.
- Fixed two silent data-quality bugs (inconsistent source headers and a combine step that dropped the county key) and wrote DAX measures showing a pooled **r = 0.77** but only **57.7%** directional agreement.

Grocery Transaction Data Warehouse | *SQL Server, T-SQL, Medallion Architecture* — Jul 2025 – Aug 2025

- Designed a Bronze–Silver–Gold warehouse loading **131M+ transaction lines** with BULK INSERT, stored-procedure ETL, load-time logging, and TRY/CATCH error handling; query tuning cut processing time by **25%**.
- Modeled a Gold-layer star schema and wrote **21 automated data-quality checks** that filtered 10M raw transactions to 3.3M completed sales and flagged 1.58M orders whose header totals did not match their line items.

Sales Planning Analysis | *Power BI, DAX, SQL Server* — Aug 2025

- Unified sales and monthly plan data from SQL Server into a galaxy-schema model (two fact tables on shared store and date dimensions) over **58K+ transactions**, revealing district plan completion of 89%–98%.

SKILLS & CERTIFICATIONS

Languages: SQL (T-SQL, Spark SQL), Python (Pandas, NumPy, PySpark), KQL, DAX, Power Query M, R, JavaScript (Google Apps Script)
Microsoft Fabric: Lakehouse, Warehouse, OneLake, Delta Lake, Data Factory, Dataflow Gen2, Notebooks, semantic models
Data Engineering: ETL/ELT pipelines, API ingestion, Medallion architecture, dimensional modeling, stored procedures, data testing
Tools: SQL Server, MySQL, Power BI, Power Automate, Excel, Git/GitHub
Certifications: Microsoft Certified: Fabric Data Engineer Associate (DP-700), Oct 2026; HackerRank SQL (Advanced), Sep 2025

LEADERSHIP

Technology Committee | Diplomatic Congress, Franklin & Marshall College — Aug 2021 – Aug 2022

- Cut classroom technology costs by **18%** by leading IT policy work and centralized license management across **15+ departments**.