

Rihua Van Steenburgh

Alpharetta, GA 30022 | 470-295-0545 | rihuavan@gmail.com

[GitHub](#) | [LinkedIn](#) | [Portfolio](#)

Summary

Data Engineer with hands-on experience building cloud data pipelines, analytics-ready datasets, and analytics reporting layers using Python, SQL, Azure, AWS, Databricks, PySpark, dbt, and Power BI. Skilled in API ingestion, ETL/ELT, data validation, dimensional modeling, workflow orchestration, and technical documentation.

Skills

- **Languages:** Python, SQL, R, JavaScript, TypeScript
- **Cloud & Platforms:** Azure, AWS, Azure Data Factory, ADLS Gen2, Databricks, AWS S3, Redshift Serverless, ECS/Fargate, EventBridge, CloudWatch
- **Data Engineering:** API ingestion, ETL/ELT, batch pipelines, orchestration, data validation, CI/CD, warehouse loading
- **Processing & Modeling:** PySpark, Delta Lake, dbt, PostgreSQL, dimensional modeling, star schema
- **BI & Reporting:** Power BI, DAX, Power Query, KPI dashboards
- **Tools:** Airflow, Docker, GitHub Actions, Git/GitHub, Jupyter Notebook, VS Code

Projects

1. NYC 311 Service Requests Lakehouse | Azure • Databricks • PySpark • Delta Lake

Mar 2026 – May 2026 [GitHub](#) | [Case Study](#)

- Built an Azure lakehouse pipeline for NYC 311 analytics, using Azure Data Factory, ADLS Gen2, and Python/PySpark in Databricks to process Bronze, Silver, and Gold layers.
- Developed reusable ingestion, cleaning, quality-check, and modeling logic to produce analytics-ready dimensions, fact tables, and Gold marts for request volume, service performance, and backlog analysis.
- Documented runbooks, architecture/data-flow notes, SQL assets, and cloud execution proof across Databricks workflows and the ADF-to-ADLS-to-Databricks handoff.

2. Cloud Flight Fare Pipeline | AWS • ECS/Fargate • S3 • Redshift • dbt

Feb 2026 – May 2026 [GitHub](#) | [Case Study](#)

- Built an end-to-end AWS batch data pipeline using Python, Docker, ECS/Fargate, EventBridge Scheduler, AWS S3 Bronze, Redshift Serverless, and dbt to ingest, load, and transform flight fare data from an external API.
- Developed SQL-based dbt staging models, dimensional marts, and data quality tests in Redshift to produce analytics-ready outputs for route trends, fare movement, and booking analysis.
- Documented cloud execution proof with CloudWatch logs, proof screenshots, CI checks, architecture diagrams, runbooks, and cost/secret safety notes for reviewer-ready validation.

3. Sumryze – AI-Powered SEO Reporting Dashboard | Next.js • TypeScript • OpenAI • APIs • Vercel

Jun 2025 – Dec 2025 [GitHub](#) | [Live Demo](#)

- Built a SaaS-style reporting dashboard for automated SEO insights, AI-generated summaries, client-facing analytics views, and reusable report workflows.
- Implemented API-based workflows, dashboard UI components, and AI-assisted reporting features using Next.js, TypeScript, and OpenAI integrations.

Experience

IT Intern — Fulton County

Jan 2024 – Apr 2024

- Standardized IT asset and network inventory records, improving data consistency and audit readiness across departments.
- Gathered and reconciled data from multiple systems to support inventory reporting, standardization, and reusable internal documentation.

Floral Assistant — Kroger

Dec 2022 – Aug 2023

- Reconciled inventory, pricing, and shrink records to improve reporting accuracy and support exception tracking.

Education

- M.S., Information Technology (Data Analytics), Middle Georgia State University — Expected Dec 2026 | GPA 4.0
- A.A.S., Information Technology (Web Development), Gwinnett Technical College — Dec 2024 | GPA 3.96

Certifications

- **Website Developer Certificate** — Gwinnett Technical College, May 2024