

Brahmani Thota

Windsor Mill, Maryland, 21224 | 908-290-7971 | brahmanithota98@gmail.com | [Loom Self-Introduction](#)

SKILLS

Programming & Cloud: Python, SQL, PySpark, Apache Spark, Amazon Redshift, SQL, AWS (S3, Glue, Lambda, Athena, RDS)
Data Engineering & Architecture: ETL/ELT Pipeline, Snowflake, dbt, Databricks, Batch & Streaming Data Processing, Kafka (CDC pipelines), Airflow Orchestration, Dimensional Modeling (Star Schema), Distributed Data Systems
Data Platforms & Governance: Data Warehousing, Data Lakes, Lakehouse Architecture, Data Quality Frameworks, Data Observability, Data Lineage, Data Governance, Schema Design, Data Reliability Engineering
Visualization & DevOps: Power BI, Tableau, Amazon QuickSight, CI/CD (GitHub Actions), Git Version Control, Infrastructure as Code (Terraform), Pipeline Monitoring, Workflow Automation, Cost Optimization

EXPERIENCE

Centene Corporation

USA

Data Engineer II – Production team

Mar 2025 - Present

- Designed scalable **AWS S3 data lake ingestion** architecture using **Glue**-based ETL pipelines, improving raw data availability for analytics workloads and reducing ingestion latency by 35% across enterprise datasets.
- Built distributed PySpark transformation pipelines on **Databricks Lakehouse** using **Delta Lake**, improving large-scale batch processing performance and reducing financial data processing time by 40%.
- Implemented **Kafka-based CDC** streaming architecture for near real-time ingestion into cloud data platforms, improving data freshness for operational reporting by 30% across downstream systems.
- Developed modular **dbt transformation** models on **Amazon Redshift**, standardizing ELT workflows and improving data consistency across reporting layers, reducing reconciliation issues by 28% for finance analytics.
- Built enterprise **Power BI dashboards** on curated cloud datasets, enabling business users to analyze revenue leakage and operational KPIs, reducing decision turnaround time by 40%.
- Established **data quality**, validation, and observability frameworks across batch and streaming pipelines using **SQL** and **Python**, improving data accuracy by 32% and reducing production data incidents by 25%.
- Built end-to-end **CI/CD** data engineering workflows using **Git** and **Terraform**, automating deployment of pipelines and infrastructure, improving release reliability and reducing deployment failures by 38%.

Data Engineer II – Snowflake Transition Team

Jan 2024 – Feb 2025

- Developed automated financial reporting pipelines using **AWS Glue**, **Snowflake** stored procedures, and VBA, reducing Medicare and Medicaid report generation time by 60% for finance stakeholders.
- Migrated** enterprise healthcare datasets from **Palantir to Snowflake**, rewriting **PySpark** and **SQL** pipelines into optimized Snowflake SQL, improving performance and enabling scalable compliance analytics workloads.
- Converted legacy **SQL** and **Teradata** datasets into **Python** and **PySpark** transformation pipelines, consolidating Medicare and Medicaid data assets, reducing redundant datasets by 25% across enterprise systems.
- Designed and optimized **Snowflake dimensional data models** with transformation and validation logic, improving financial reporting accuracy and enabling faster downstream analytics for actuarial and business teams.
- Built **Tableau** dashboards on Snowflake curated datasets, enabling KPI monitoring for migration validation, improving operational visibility and supporting finance leadership decision-making during enterprise data transition.

Data Engineer I – Palantir Team

Aug 2021 – Dec 2023

- Improved data deployment workflows through **Git version control** and automated validation checks, improving release traceability and reducing manual deployment errors across shared data assets.
- Assisted ingestion architecture planning exploring **event driven ingestion patterns**, supporting long term design for near real time financial data availability across enterprise reporting systems.
- Built distributed extraction pipelines from **Teradata** to external **APIs** using **Python** multiprocessing with fault tolerant batching and retry handling, reducing runtime to 80% and improving downstream dataset availability.
- Performed **enterprise table mapping** and **metadata reconciliation** supporting migration into centralized reporting platforms, ensuring parity between legacy outputs and enterprise regulatory reporting datasets.

IBM

USA

Python Developer Intern - Quantum Computing

June 2020 to May 2021

- Developed quantum learning applications using **Python** and **Qiskit**, enabling global learners to experiment with quantum circuits through guided simulation modules.
- Built reusable hybrid optimization modules using structured **Python OOP design**, improving experimentation speed and enabling reuse across IBM Quantum training programs.

PROJECTS

Delta Analytics Fellowship — Data Fellow (Multiple Projects)

2023 – 2025

- Built survey analytics dashboards connecting **Salesforce** → **Power BI**, enabling impact measurement for veteran family programs across US retreat locations.
- Developed predictive models for behavioral risk using **PySpark** + **SageMaker**, deploying ML pipelines and building **QuickSight dashboards** supporting clinical intervention decision workflows.

EDUCATION

University of Maryland Baltimore County

Majors: Data Science | GPA: 4.0

Graduation Date: 2021