

## Project Summary

### **Healthcare Metrics Project Summary**

The **Healthcare Metrics Project** was a full-stack data pipeline and analytics solution designed to ingest, transform, and visualize nursing home performance data using a modern data stack. The goal was to enable dynamic exploration of key healthcare quality and staffing metrics at the provider and state level.

### **End-to-End Architecture Overview**

The project followed a modular, cloud-native architecture built around secure ingestion, scalable transformation, and interactive visualization:

#### **Data Ingestion (Bronze Layer)**

- **Google Drive** was the initial source of truth, where raw CSVs were stored.
- **AWS Lambda** fetched only new or changed files using a secure and incremental approach:
  - **AWS Secrets Manager**: Retrieved Google Drive API credentials
  - **AWS KMS**: Decrypted secrets and encrypted files uploaded to **Amazon S3**
  - **Lambda Layer**: Added external Python dependencies (e.g., google-api-python-client)
  - **DynamoDB**: Tracked processed file IDs to ensure incremental ingestion

✅ *This ensured controlled, redundant-free ingestion and encrypted data handling.*

#### **Centralized Storage and Transformation (Silver Layer)**


- **Amazon S3** served as the raw file landing zone.
- **Snowflake** acted as the centralized data warehouse. Raw tables were loaded from S3 and further refined.
- **dbt (Data Build Tool)** powered the transformation pipeline:
  - **Bronze Models** ingested CSVs from S3 into Snowflake raw tables (via custom macro)

- **Silver Models** cleaned, joined, and enriched the data into analysis-ready tables





⚙️ *The use of dbt enabled version-controlled, modular SQL transformations.*

## Visualization and Insights (Gold Layer)

- **Streamlit** was used to build a polished dashboard that connected directly to Snowflake.
- The app allowed users to:
  - Filter data by **state**, **provider**, or **ownership**
  - View trends in **nurse staffing**, **overtime**, **readmission rates**, and **survey deficiencies**
  - Explore **top and bottom performing providers** via interactive bar and line charts

 *The UI was built with streamlit + snowflake-connector-python, leveraging built-in charting and UI components for filtering.*

## Key Accomplishments

-  Secure, incremental ingestion pipeline built with AWS and Python
-  Raw-to-Silver transformation logic in dbt using macros and standardized models
-  Real-time Snowflake integration with a user-friendly Streamlit frontend
-  Final dashboard answered business questions around staffing, quality, and compliance metrics