

EUGENE SONG

Los Angeles, CA | esong28@outlook.com | [linkedin.com/in/eugenessong](https://www.linkedin.com/in/eugenessong) | github.com/EugenSong | eugensong.github.io

Data Engineer with 3 years running production ETL pipelines on 20M+ healthcare claims and eligibility records in a HIPAA-regulated environment with deep SQL and Teradata expertise; led an on-prem → Google BigQuery migration with zero data loss.

TECHNICAL SKILLS

- **Languages:** SQL, Python, Java
- **Data & Pipelines:** Teradata (BTEQ), ETL/ELT pipeline design, data modeling, data validation, BigQuery, Snowflake, PostgreSQL, dbt, Databricks
- **Cloud & Infrastructure:** AWS (S3, Lambda, RDS, CloudWatch), Terraform, Docker, CI/CD
- **Orchestration & Tools:** Airflow, Control-M, Git, REST APIs, Power BI

WORK EXPERIENCE

enGen (Highmark Health)

May 2023 – Present

Software Engineer, Data & Analytics

- Migrated on-prem Teradata workloads to Google BigQuery with zero data loss, validating full parity between legacy and target systems to maintain continuity during transition.
- Build and maintain two-stage Teradata ETL pipelines (source → stage → final) processing over 20M claims and eligibility records across dev, QA, and prod, with controlled promotion ensuring auditable, HIPAA-compliant data delivery.
- Cleaned and modeled raw enterprise warehouse data, consolidating 40+ source tables into a single curated, analysis-ready layer for analytics and operations teams.
- Serve as primary data-quality contact for 15+ analysts across 5 teams, triaging lineage issues to root cause in the ETL layer and cutting repeat incidents reaching downstream reporting.
- Orchestrate pipeline execution and dependencies through enterprise Control-M workflows with automated retries and error handling, delivering fresh data on a daily schedule.

PROJECTS

Stack Overflow Analytics – dbt + BigQuery

- Built a multi-layer dbt project (staging → intermediate → marts) modeling 23M+ rows of public data on BigQuery, using source() and ref() to manage dependencies and auto-generate lineage, with materialization tuned per layer and data quality tests across 7 models and 4 sources.

Healthcare Analytics ELT Pipeline – Airflow + Snowflake + dbt

- Built a production-style batch ELT pipeline over CMS Medicare synthetic claims using a Python extract and load into Snowflake, dbt transformations from raw through staging to marts, orchestrated end-to-end with Apache Airflow via Docker Compose.

The Cloud Resume Challenge

- Deployed a serverless portfolio application on AWS (S3, CloudFront, Lambda, DynamoDB) with Terraform-managed infrastructure and GitHub Actions CI/CD, achieving sub-100ms global latency and demonstrating end-to-end cloud architecture proficiency.

EDUCATION

Oregon State University

Graduated December 2023

Bachelor of Science in Computer Science

California State University, Long Beach

Graduated June 2021

Bachelor of Science in Health Science

CERTIFICATIONS

- AWS Certified Solutions Architect – Associate
- CompTIA Security+