Gerardo Toboso

Data Engineer

Buenos Aires, Argentina | gerardotoboso1909@gmail.com | 011 15-4045-6207

linkedin.com/in/gerardo-toboso-512a48290 | github.com/Gerardo1909

Summary

Data Engineer specialized in designing scalable data platforms and production-grade pipelines following modern data architecture principles. Experienced in Medallion Architecture (Bronze/Silver/Gold), dimensional modeling, distributed processing, and data quality frameworks. Strong focus on idempotent pipeline design, observability, cost-efficient storage strategies, and business-aligned data modeling. Currently expanding expertise into Azure-native data services and cloud platform optimization.

Skills

Data Architecture & Platform Design: Medallion Architecture (Bronze/Silver/Gold), Dimensional Modeling (Star Schema, SCD Type 2), Idempotent pipeline design, Data Contracts, Partitioning & Storage Optimization (Parquet), Cost-aware Data Lake design, and Data SLAs & Freshness Policies

Data Engineering & Processing: Python (pandas, PySpark, duckdb), Apache Airflow (DAG design, retries, alerting, monitoring), Distributed processing (~50GB/min throughput), ETL/ELT pipeline development, and Versioned datasets

Databases & Query Optimization: PostgreSQL, Query optimization & indexing strategies, Execution plan analysis, Partitioned tables, and Advanced SQL for analytics workloads

Cloud & DevOps: AWS S3 (Data Lake architecture), Docker, CI/CD (GitHub Actions), Infrastructure reproducibility, Monitoring & structured logging, and exploring Azure Data Factory & Azure SQL

Data Quality & Observability: Great Expectations, Automated pipeline testing (Pytest), Data lineage via dbt, Alerting & failure detection, and Freshness checks

Experience

Data Engineer, Analytics Engineering – Fintech Data Platform (dbt) (Personal Project) – GitHub Jan 2026 – Feb 2026

- Migrated 25+ fragmented SQL transformations into a structured dbt-based transformation layer.
- Implemented SCD Type 2 dimensions to preserve historical customer and investment data.
- Established automated data tests (unique, not_null, accepted_values) reducing production bugs by 80%.
- Introduced documentation-first approach using dbt Docs, reducing onboarding time from 2 weeks to 2 days.
- Modeled analytics-ready marts following scalable ELT best practices.

Data Engineer, IoT Data ETL (Personal Project) – GitHub Dec 2025 – Jan 2026

- Designed a Medallion-based architecture (Bronze → Silver → Gold) processing millions of industrial IoT events daily.
- Built distributed transformations in PySpark achieving ~50GB/min throughput with no memory bottlenecks.
- Implemented Parquet partitioning and dataset versioning to optimize query performance and reduce storage costs.
- Defined data quality checks and failure alerting ensuring 99.9% pipeline uptime.
- Architected AWS S3 data lake to support >1TB/day scalability.
- Designed dimensional analytics layer enabling BI-ready star schema models.

Education

Universidad Nacional de San Martín (UNSAM), BSc in Data Science July 2022 – present

- 75% completed — GPA: 9.0 / 10