

Gerardo Toboso

Ingeniero de Datos

Buenos Aires, Argentina | gerardotoboso1909@gmail.com | 011 15-4045-6207
linkedin.com/in/gerardo-toboso-512a48290 | github.com/Gerardo1909

Resumen

Data Engineer especializado en el diseño de pipelines cloud-native sobre AWS, con enfoque en arquitecturas serverless, idempotencia y validación temprana de datos. Experiencia construyendo Data Lakes en Amazon S3, pipelines event-driven con AWS Lambda y almacenamiento particionado en Parquet optimizado para analítica. Fuerte orientación a calidad de datos, observabilidad (CloudWatch) y diseño costo-eficiente en entornos productivos.

Habilidades

Arquitectura de Datos & Diseño Cloud: Arquitecturas event-driven en AWS, Data Lakes en Amazon S3, diseño de pipelines idempotentes, particionamiento Hive-style (Parquet), modelado dimensional (Star Schema, SCD Type 2), Data Contracts y políticas de frescura (SLA de datos)

Ingeniería de Datos & Procesamiento: Python (pandas, PySpark, duckdb), desarrollo de ETL/ELT, procesamiento distribuido, diseño de DAGs y monitoreo en Apache Airflow, datasets versionados y segregación de datos válidos/rechazados

Bases de Datos & SQL: SQL avanzado para analítica, PostgreSQL, optimización de queries e indexación, análisis de execution plans, tablas particionadas y modelado para cargas analíticas

Cloud & DevOps (AWS): AWS Lambda, Amazon S3, IAM (principio de mínimo privilegio), CloudWatch (logs y monitoreo), EventBridge, ECR (containerización), Docker y CI/CD con GitHub Actions

Calidad de Datos & Observabilidad: Validaciones basadas en reglas de negocio, testing automatizado con Pytest, logging estructurado, detección de fallos y segregación de registros inválidos para auditoría

Idiomas: Español (Nativo) | Inglés (Fluido - C2)

Experiencia

Ingeniero de Datos, Hotel Serverless Pipeline – AWS Data Architecture (Proyecto Personal) – [GitHub](#) Ene 2026 – Feb 2026

- Diseñé e implementé un pipeline de datos completamente serverless en AWS (S3 + Lambda + EventBridge).
- Automatizé la detección de nuevos lotes en S3 y la ejecución de transformaciones mediante triggers event-driven.
- Consolidé múltiples CSV en datasets particionados en Parquet (Hive-style) optimizados para Athena.
- Implementé reglas de validación (precio, fechas, rangos de score) con segregación automática de datos rechazados.
- Diseñé el sistema bajo principios de idempotencia (reintentos seguros, sin duplicación).
- Configuré IAM con políticas de mínimo privilegio y monitoreo centralizado en CloudWatch.

Ingeniero de Datos, ETL para datos IoT (Proyecto Personal) – [GitHub](#) Dic 2025 – Ene 2026

- Diseñé arquitectura tipo Medallion (Bronze → Silver → Gold) procesando millones de eventos IoT diarios.
- Implementé transformaciones distribuidas en PySpark alcanzando ~50GB/min sin cuellos de memoria.
- Arquitecté un Data Lake en Amazon S3 con particionamiento y versionado para soportar >1TB/día.
- Definí controles de calidad y alertas de fallos logrando 99.9% de disponibilidad del pipeline.
- Modelé capa analítica dimensional (Star Schema) lista para consumo BI.

Educación

Universidad Nacional de San Martín (UNSAM), Licenciatura en Ciencia de Datos Jul 2022 – presente

- 75% completado — Promedio: 9.0 / 10