

Gaston Echenique

Senior Data Engineer | ML & AI

SUMMARY

Data Engineer with 4+ years designing, building, and owning cloud data architecture end-to-end — with growing hands-on focus on production ML and agentic AI systems — from ingestion to production SQL-driven analytics. Deep hands-on experience with AWS (S3, Glue, Athena, Lambda), Apache Spark, and large-scale SQL modeling and query optimization across multi-country, multi-terabyte data platforms in healthcare, energy, and pharma. Strong ownership mentality: I design the architecture, build the pipeline, and debug it in production. Comfortable working directly with both engineering and business stakeholders to translate requirements into scalable data solutions.

CORE STACK

AWS (S3, Glue, Athena, Lambda, SageMaker, Bedrock) · Apache Spark / PySpark · SQL (advanced) · Python · ML Pipelines / Agentic AI · Data Lakes & ETL/ELT · Data Architecture · Airflow · Docker · NoSQL (MongoDB) · PostgreSQL · Azure (Synapse, ADF)

PROFESSIONAL EXPERIENCE

Data Engineer | The Black Puma AI (AWS Partner)

Sept 2022 – Present

- Own the end-to-end data architecture on AWS (S3, Glue, Athena, Lambda) for multi-country production pipelines serving healthcare (Intramed), energy (EDEMSEA), and pharma (Roemmers/IQVIA) clients — designing for scalability, availability, and cost efficiency.
- Drive complex SQL-based data modeling and query optimization on Athena/Glue for large-scale verification and compliance datasets, resolving data quality issues and schema inconsistencies across hundreds of thousands of records with full ownership from investigation to fix.
- Built and optimized high-throughput Glue-based ETL pipelines processing and consolidating large volumes of data (e.g., ~110M records), improving performance, deduplication, and data integrity end-to-end.
- Designed and maintained Azure Synapse/ADF pipelines for ingesting and consolidating multi-country financial and operational data, partnering directly with technical and business stakeholders to define requirements.
- Migrated a research-based ML fraud-detection system to AWS SageMaker for EDEMSEA Energy, building and deploying the production ML pipeline end-to-end, optimizing for cost and enterprise-scale deployment.
- Built enterprise-scale geomarketing solutions using H3 spatial indexing, combining POI and customer data at scale for business decision-making.
- Automated recurring data-ops workflows with n8n and Docker, removing daily manual validation work and reducing time-to-delivery for downstream teams.

Data Modeling Engineer | Google X (Moonshot Factory)

Jul 2021 – Aug 2022

- Rebuilt the data platform for a stalled agricultural analytics project, transforming 4 years of disconnected, inconsistent datasets into production-ready pipelines feeding predictive models.
- Designed a GCP data lake (BigQuery, Dataflow) with YAML-driven configuration management, enabling reliable, repeatable ingestion for crop-yield prediction models.
- Built automated data-quality validation into multi-year historical ETL workflows and interactive visualizations that supported cross-functional decision-making with the data science and client teams.

AI & ML Instructor | UNTREF / GeoDojo

Mar 2023 – Present

- Design and deliver a Python, SQL, and Machine Learning curriculum focused on practical, production-oriented data engineering skills.

TECHNICAL SKILLS

Data Architecture & Warehousing: AWS data lakes, ETL/ELT design, data modeling, dimensional modeling, scalability and performance tuning

AWS: S3, Glue, Athena, Lambda, SageMaker, Bedrock

Processing & Orchestration: Apache Spark, PySpark, Airflow, Docker, n8n

SQL & Databases: Advanced SQL, query optimization, PostgreSQL, MongoDB (NoSQL)

Languages: Python, SQL

Other cloud ecosystems: Azure (Synapse, ADF, Delta Lake), GCP (BigQuery, Dataflow)

CERTIFICATIONS & EXECUTIVE EDUCATION

- AWS Certified AI Practitioner (AIF-C01) — 2025
- Executive Education Program — Agentic AI | Universidad de San Andres (UdeSA) — 2026: end-to-end program covering AI agents, Agentic RAG, tool use / function calling, MCP, multi-agent architectures, evaluation and testing, production deployment on AWS, guardrails, and monitoring.

EDUCATION

- Postgraduate Diploma in Big Data and Spatial Territorial Intelligence — FLACSO (2022)
- MSc in Environmental Management — UNSAM (2017)
- BA in Naval Resource Management — UNDEF (2009)