Javier García

Senior Data Engineer | Cloud Data Platforms & GenAI Data Infrastructure

Barcelona, Spain | Open to remote and relocation | EU work authorization
+34 695 718 903 | edjagalo@gmail.com | linkedin.com/in/edjagalo | github.com/javigarlop

Summary

Senior Data Engineer building AWS and GCP data platforms and the data infrastructure behind AI products; 7+ years in data, 4+ in data engineering. At IAG, one of two data engineers building a contract intelligence platform from scratch (Airflow on EKS, dbt, PostgreSQL/pgvector) whose Gold tables feed AI agents flagging aircraft-maintenance invoice overbilling across 3 airlines. At L'Oréal, led an advertising data platform on BigQuery and a GenAI analytics agent used by 500+ marketers. Builds idempotent, tested pipelines with Terraform and CI/CD; mentors, reviews PRs and interviews candidates. Two MScs in data and business analytics.

Skills

Programming & Backend: Python, SQL, Go, FastAPI, SQLAlchemy, scikit-learn, pytest, REST APIs
Data Engineering & Orchestration: Apache Airflow, dbt, Dataform, Apache Beam (Dataflow), PySpark (AWS Glue), Pub/Sub, Alembic, medallion architecture, SCD2, batch and streaming pipelines, data quality testing
Cloud & Infrastructure: AWS (S3, EKS, Aurora/RDS, Lambda, Glue, Athena, IAM, CloudWatch), GCP (Cloud Build, Workflows, Cloud Functions), Terraform, Docker, Kubernetes, GitHub Actions, Grafana
Databases & Warehousing: PostgreSQL, pgvector, BigQuery, Redshift, Snowflake, SQL Server
AI & Machine Learning: Amazon Bedrock, OpenAI APIs, Vertex AI, LLMs, RAG, embeddings, vector search, AI agents, LLM evaluation, NLP, classification, forecasting
Analytics & Semantic Modeling: star schema, LookML/Looker semantic layers, Power BI, Tableau, KPI design
Languages: Spanish (native), English (C1), Catalan (A2)

Experience

Capitole Consulting — Barcelona, ES

Senior Data Engineer — Jun 2026 – Present

For airline group IAG (Iberia, Vueling, Aer Lingus; 10-person squad, 2 data engineers): contract intelligence flagging overbilling and contract non-compliance in aircraft maintenance (MRO) invoices.

- Co-designed and built the platform's data foundation, from an empty repository to UAT across 4 releases in 4 months: an S3-to-PostgreSQL/pgvector medallion architecture (Bronze/Silver/Gold) checking 450 contracts against millions of invoices and work orders from Snowflake, with 6 Airflow DAGs on EKS, 57 dbt models and 2,150 tests.
- Built the founding Python ingestion pipeline (boto3, SQLAlchemy, Alembic) and made it idempotent on a SHA-256-based document identity (payload hash, type, airline, source ID): corrections become new versions, retries are no-ops and a re-run never pays twice for the same LLM call.
- Delivered typed Gold models (contracts, price books, scope rules, turnaround commitments, billing lines) linking each invoice to its contract, as the source for the reasoning team's knowledge graph and compliance agents (3 patterns in the MVP, e.g. unhonored warranties), with Bedrock embeddings and SCD2 history; raised compliance-rule coverage from 83.2% to 91.4% through a detailed rule review with business SMEs.
- Owned source integrations from Snowflake, external APIs and CSV/XLSX vendor feeds into airline-scoped Gold tables, and integrated a document-AI extraction gateway into the contract registry FastAPI service.
- Secured the stack with no long-lived credentials (IRSA, short-lived RDS IAM tokens, TLS verify-full) and per-airline row-level security, and made migrations provably reversible with a CI round-trip test on real PostgreSQL.

Devoteam — Barcelona, ES

Senior Data Engineer — Aug 2025 – Jun 2026

Consultant for L'Oréal's European marketing division, leading two projects and one engineer: an advertising data platform and a GenAI agent.

- Led the advertising data platform consolidating campaign performance and social engagement (sales, reactions, likes, comments) for global product campaigns from 7 platforms (DV360, Meta, Google Ads, TikTok, Amazon DSP, Snapchat, offline media) into BigQuery, with Dataform ELT deployed continuously.
- Designed a GenAI analytics agent on an internal ChatGPT deployment (OpenAI APIs), with a RAG data model in Vertex AI that routes marketing questions to BigQuery queries; in production for 500+ users across the division.
- Refactored the data model into a star schema of 45+ fact and dimension tables with a rule-based matching engine for quarterly benchmarks, cutting query latency by 40% and dashboard load time from 5 minutes to 1.
- Provisioned dev/qa/prod on GCP with Terraform (Cloud Build CI/CD, Workflows, Cloud Scheduler), running daily automated Power BI refreshes.
- Built a LookML semantic layer with 50+ marketing KPIs, enabling self-service analytics across Europe.
- Built the offline-media ETL pipeline (TV, sponsorship) and led the migration adding age/gender breakdowns.

Evolutio Cloud Enabler Barcelona, ES

Senior Data Engineer Aug 2024 – Jun 2025

Consultant in the cyberdefense area, building a data platform for insurers including Mapfre.

- Co-built the Python/Terraform pipeline ingesting millions of rows per hour from an API, in batch and real time.
- Migrated 2 years of historical data from an S3/Athena/Glue data lake to Redshift for the front-end team, cutting processing time by 50%, and reduced weekly data incidents by 25% (PySpark on Glue, EKS, Grafana).
- Defined the technical roadmap for ML and generative AI applications, and built customer features and time-series models (ARIMA, XGBoost) forecasting hourly call volumes on Redshift and SQL Server.

NTT Data Barcelona, ES

AI Engineer Oct 2023 – May 2024

Consultant on two engagements: Inditex and European Commission research data.

- Fine-tuned and deployed Inditex's online-store clothing recommendation system on Vertex AI across 13 European countries, resolving stability and performance issues in its pipelines (BigQuery, Dataflow, Pub/Sub).
- Built a multi-label text-classification pipeline over the European Commission's 130,000+ CORDIS research projects (7 scientific domains), benchmarking five classical ML models against zero-shot GPT-3.5/GPT-4o on 13,473 hand-labelled projects; classical models won on label error (Hamming loss about 10% vs 15%).

Nexllence, powered by Glintt Madrid, ES

Senior Software Developer (Data Engineering) Aug 2022 – May 2023

Consultant for Cofares on pharmaceutical digital projects.

- Designed the data model for the Welnia API and delivered the Airflow pipeline on GCP behind its front end.
- Delivered Looker analytical sources, standardized data quality protocols and interviewed engineering candidates.

Mercado Libre Bogotá, CO

Senior Data Engineer Jan 2022 – Aug 2022

Central data engineering department; left to relocate to Spain for the MSc at Oviedo.

- Built terabyte-scale batch and real-time warehouses on GCP and AWS (S3, BigQuery, Dataflow, Lambda).
- Built and maintained Go and Python microservices on AWS, including replicated Go consumers ingesting real-time event streams and landing unstructured data in S3.
- On critical-systems on-call shifts, owned the daily run of 3,000+ heterogeneous processes, debugging unfamiliar code and shipping fixes to minimize data loss.

Colombo American Center Bogotá, CO

Business Intelligence Lead Mar 2021 – Jan 2022

- Led a squad of data analysts and designed the data architecture integrating CRM systems and custom software APIs, with ETL pipelines and Tableau/Power BI KPI dashboards.
- Cut business process times by 85% with workflow applications, and built a linear programming model for course-to-student ratios, reducing that process from 2 days to half a day.

Spectrum Brands Inc. Bogotá, CO

Business Intelligence Specialist Jan 2019 – Mar 2021

- Led data-driven digital transformation projects, reducing purchasing costs by 26% and stock levels by 30%.
- Improved forecast accuracy with time-series and predictive models (RMSE down by over 15%), and built KPIs and Power BI dashboards for business decision-making.

Education

University of Oviedo Oviedo, ES

MSc in Data Analytics for Business Intelligence 2022 – 2025

Thesis: classical ML vs zero-shot GPT for multi-label classification of EU research projects (NTT Data project).

Pontifical Javeriana University Bogotá, CO

MSc in Business Analytics 2018

Thesis (team of 3): k-means driver segmentation and a geospatial ride-recommendation model for SmartTaxi.

Industrial University of Santander Bucaramanga, CO

BSc in Industrial Engineering 2015

Thesis: automatic ARIMA forecasting to improve SYGLA's supply chain.