Jorge Torres-Homem Lira Filho

Data Engineer

jorgelira542502@gmail.com · linkedin.com/in/jorge-lira-dev · github.com/jorgelira123 · Portfolio · Maceió, Brazil (Remote)

PROFESSIONAL SUMMARY

Data Engineer with 2.6 years building and operating data platforms on GCP for enterprise clients, backed by a solid software development background. Experienced in ERP and API data integration, scalable pipelines (Apache Airflow/Cloud Composer, Dataflow, BigQuery), Medallion architecture, relational and dimensional modeling, data quality and cloud cost optimization (FinOps) — ~83% monthly cost reduction on a retail client's critical BigQuery routines (~US$112K projected annual savings) and 449M records migrated to GCP. Hands-on with RAG on Vertex AI and Gemini and with LLM function calling in a production side project. Consultative profile: translates business requirements into data solutions and presents outcomes to stakeholders.

PROFESSIONAL EXPERIENCE

Leega — Data Engineer **May 2024 – Present**

São Paulo, Brazil · Remote · Data consultancy

Data engineering and architecture projects on GCP for enterprise clients, focused on scalable pipelines, modeling, data quality, production operations and cost optimization.

ERP software company — ERP Data Integration and Pipeline Operations · 2026 – Present

- Sole technical owner of data integration from an Oracle-based ERP environment into the client's data platform on BigQuery, covering 150+ tables across 8 tenants.
- Operate and evolve ingestion pipelines through the client's connectors, including view modeling, deduplication of versioned records and analytical queries for reporting.
- Diagnosed and resolved a critical replication failure that kept records deleted in the ERP as active in the destination, affecting accounting figures. Traced the root cause to a trigger condition that blocked capture of logical deletions, with an estimated impact of tens of millions of records.
- Defined a business-key deduplication standard, replacing the previous criterion that did not handle re-ingestions and produced duplicates in queries.
- Optimized BigQuery queries, significantly reducing data processed per run through filter tuning, partitioning and column selection.
- Technical interface with ERP, database and connector teams; supported client analysts in building queries and interpreting the data model.

Retail company — FinOps and BigQuery Optimization · 2026

- Led the technical mapping and diagnosis of the largest offenders in slot consumption and data volume, in a team of 4 engineers.
- Refactored critical queries, achieving 90%+ gains in execution time and data processed.
- Cut monthly cost of the mapped routines by ~83%, for ~US$112.4K in projected annual savings.
- Applied SQL and data architecture best practices to turn high-cost pipelines into agile, scalable routines.

Manufacturing company — Data Assessment and Modernization · 2026

- Data assessment across 9 business areas, actively leading 6 of them from discovery meetings to each area's requirements.
- Technical lead alongside the business specialist; supported and mentored a junior engineer.
- Mapped the as-is model and defined the to-be model and ROI with the project manager and coordinator; delivered ~12 PoCs (about 2 per area) for technical validation.

Manufacturing company — Data Assessment and Modernization · 2025

- Data assessment across 12 business areas, with full process mapping of 3 and data architecture improvements proposed on GCP.

Telecom company — Data Migration to GCP · 2025

- Migrated 449M records from Teradata and Oracle to Google Cloud Platform.
- Built Python DAGs in Apache Airflow for orchestration, processing via Dataflow and final load into BigQuery — load window cut from 24 hours to 12.
- Implemented Cloud Logging and LogSink for monitoring and log delivery to the client.

Leega (internal) — RAG Chatbot PoC for Résumé Analysis · 2025

- Implemented a RAG architecture with Vertex AI, Gemini models and Cloud Storage for document-grounded answers.
- Extracted and structured skills and experience from résumés; validated technical feasibility for candidate screening.

ERP software company — Data Platform Engineering · 2024 – 2025

- Helped build Python connectors for Zendesk, Jira and Jira Service to standardize data ingestion.
- Contributed to the Medallion architecture (Bronze/Silver/Gold) on the client's data platform, applying data lake, data warehouse and pipeline concepts.
- Supported the team in defining the data architecture and automating processes on the platform.

Pulp and paper company — Data Analysis and BI Modernization · 2024

- Migrated and modernized Power BI dashboards and KPIs; adjusted database tables and validated sources for performance and data quality.

JLira — Full Stack Developer (Freelance) **Aug 2023 – May 2024**
Maceió, Brazil

- Built applications and websites with Python/Django, Django REST Framework, React, Node.js and PostgreSQL, including front-end integration with APIs.

Grupo In4 — Full Stack Developer (Delphi) **Aug 2021 – Jul 2023**
Maceió, Brazil

- Front-end and back-end development in Delphi for corporate systems in production, with ticket management and prioritization, version control (Git) and team collaboration.

PROJECTS

Fuel Data Platform (ANP) — Medallion data platform on GCP · Completed
Python · Apache Beam · PySpark · Dataproc · Cloud Storage · Cloud Composer · BigQuery · Looker Studio · Docker

- Dataflow (Apache Beam) ingests the paginated public API of Brazil's National Petroleum Agency into Cloud Storage as Parquet (Bronze); Spark on Dataproc cleans and normalizes it (Silver) and aggregated metrics are built on top (Gold); BigQuery external tables and views feed Looker Studio dashboards; orchestrated by Cloud Composer.
- Hive-style partitioning for partition pruning in BigQuery, ephemeral Dataproc clusters for cost control and isolated service accounts.

Clairemind — Personal AI assistant on WhatsApp · 2026 – Present
Node.js · Claude Haiku 4.5 · function calling · Whisper · Evolution API · n8n · Google Calendar API · SQLite · Docker · GCP

- Conversational assistant running entirely on WhatsApp: manages calendar, tasks and reminders, understands voice messages and negotiates meetings with third parties end to end, in natural language.
- Built for reliability: LLM actions are gated through tools and confirmations are generated by code, so it cannot confirm an action it did not execute; asynchronous negotiation is an explicit state machine. Runs on GCP (nginx/SSL, IAP-only SSH, daily backups with validated restores, rate limits, cost cap); ~41% LLM cost saved with prompt caching. Documented publicly on LinkedIn.

dbt on BigQuery · In progress

- dbt Core 1.12 on the public thelook_ecommerce dataset: staging, intermediate and mart models, an incremental fact table, an SCD Type 2 snapshot, data tests and source freshness, with dbt build in GitHub Actions.

TECHNICAL SKILLS

Cloud	Google Cloud Platform (BigQuery, Dataflow, Cloud Composer, Cloud Storage, Dataproc, Cloud Logging, Vertex AI), AWS, Databricks
Pipelines	Apache Airflow / Cloud Composer, Dataflow, ETL/ELT, ERP data integration (Oracle), API connectors (Zendesk, Jira), data ingestion and transformation
Languages	Python, SQL, PySpark, JavaScript
Data	Relational and dimensional modeling, Medallion architecture (Bronze/Silver/Gold), Data Lake, Data Warehouse, Data Mart, data quality and governance, FinOps, query and cost optimization, dbt (learning)
Operations	Incident resolution, root cause analysis, monitoring and logging, technical documentation
AI	RAG (Vertex AI, Gemini, Cloud Storage), LLM function calling, prompt caching
Tools	Git, Docker, Jupyter Notebooks, Power BI, Looker Studio, Jira, JSON, npm/pip

CERTIFICATIONS

Google Cloud: Professional Cloud Architect (2026) · Professional Data Engineer (2025) · Cloud Digital Leader (2024) · **Databricks:** Certified Data Engineer Associate (2025)
AWS: Solutions Architect Associate (2026) · Cloud Practitioner (2023) · **Cisco:** Python Essentials 2 (2023) · CCNA Introduction to Networks (2023)

EDUCATION AND LANGUAGES

Bachelor's Degree in Civil Engineering — Mackenzie Presbyterian University
English — Advanced (professional use: reading, writing, meetings and technical documentation) · Portuguese — Native