Jorge Torres-Homem Lira Filho

Engenheiro de Dados

jorgelira542502@gmail.com · linkedin.com/in/jorge-lira-dev · github.com/jorgelira123 · Portfólio · Maceió, AL — Brasil (Remoto)

RESUMO

Engenheiro de Dados com 2,6 anos de atuação em projetos de engenharia e arquitetura de dados em GCP para clientes enterprise, com background sólido em desenvolvimento de software. Experiência em integração de dados de ERP e APIs, pipelines escaláveis (Apache Airflow/Cloud Composer, Dataflow, BigQuery), arquitetura Medalhão, modelagem relacional e dimensional, qualidade de dados e otimização de custo em nuvem (FinOps) — redução de ~83% no custo mensal de rotinas críticas em BigQuery para cliente de varejo (saving anual projetado de ~US$ 112 mil) e migração de 449M de registros para GCP. Experiência prática com RAG em Vertex AI e Gemini e com LLM function calling em produção num projeto pessoal. Perfil consultivo, traduzindo requisitos de negócio em soluções de dados e apresentando resultados a stakeholders.

EXPERIÊNCIA PROFISSIONAL

Leega — Engenheiro de Dados **mai/2024 – atual**
São Paulo, Brasil · Remoto · Consultoria de dados

Projetos de engenharia e arquitetura de dados em GCP para clientes enterprise, com foco em pipelines escaláveis, modelagem, qualidade de dados, operação em produção e otimização de custo.

Empresa de software ERP — Integração de ERP e Sustentação de Pipelines · 2026 – atual

- Responsável técnico único pela integração de dados do ambiente de ERP em Oracle para a plataforma de dados do cliente no BigQuery, abrangendo mais de 150 tabelas distribuídas em 8 tenants.
- Sustentação e evolução de pipelines de ingestão pelos conectores do cliente, incluindo modelagem de views, deduplicação de registros versionados e construção de consultas analíticas para relatórios.
- Diagnóstico e resolução de falha crítica de replicação que mantinha registros excluídos no ERP como ativos no destino, afetando a apuração de valores contábeis. Identificação da causa raiz em condição de trigger que impedia a captura de exclusões lógicas, com impacto estimado em dezenas de milhões de registros.
- Definição de padrão de deduplicação por chave de negócio, substituindo o critério anterior que não tratava reingestões e gerava duplicidade em consultas.
- Otimização de consultas em BigQuery, com redução significativa de volume processado por execução através de ajustes de filtros, particionamento e seleção de colunas.
- Interface técnica com times de ERP, banco de dados e conector, além de suporte a analistas do cliente na construção de consultas e interpretação do modelo de dados.

Empresa de varejo — FinOps e otimização de BigQuery · 2026

- Liderança técnica no mapeamento e diagnóstico dos maiores ofensores de consumo de processamento (slots) e volume de dados, em time de 4 engenheiros.
- Refatoração e otimização de queries críticas, com ganhos superiores a 90% em tempo de execução e dados processados.
- Redução de ~83% nos custos mensais das rotinas mapeadas, com saving anual projetado de ~US$ 112,4 mil.
- Aplicação de boas práticas de SQL e arquitetura de dados para transformar pipelines de alto custo em rotinas ágeis e escaláveis.

Indústria — Assessment e Modernização de Dados · 2026

- Assessment de dados em 9 áreas da empresa, com atuação ativa em 6 delas, das reuniões de levantamento ao entendimento das necessidades de cada área.
- Apoio técnico do projeto junto à especialista de negócio; suporte e mentoria a uma engenheira júnior.
- Mapeamento do modelo as-is e definição do modelo to-be e do ROI com gerente e coordenador; entrega de ~12 PoCs (cerca de 2 por área) para validação técnica.

Indústria — Assessment e Modernização de Dados · 2025

- Assessment de dados em 12 áreas da empresa, com mapeamento completo de processos de 3 e proposta de melhorias de arquitetura em GCP.

Empresa de telecom — Migração de dados para GCP · 2025

- Migração de 449M de registros de Teradata e Oracle para o Google Cloud Platform.
- Criação de DAGs em Python (Apache Airflow) para orquestração, processamento via Dataflow e carga final em BigQuery — janela de carga reduzida de 24h para 12h.
- Implementação de Cloud Logging e LogSink para monitoramento e disponibilização de logs ao cliente.

Leega (interno) — PoC de chatbot RAG para análise de currículos · 2025

- Implementação de arquitetura RAG com Vertex AI, modelos Gemini e Cloud Storage para respostas fundamentadas em documentos.
- Extração e estruturação de habilidades e experiências dos currículos; validação da viabilidade técnica para triagem de candidatos.

Empresa de software ERP — Engenharia de Dados em plataforma · 2024 – 2025

- Participação no desenvolvimento de conectores em Python para Zendesk, Jira e Jira Service, padronizando a ingestão de dados.
- Participação na criação da arquitetura Medalhão (Bronze/Silver/Gold) na plataforma de dados do cliente, aplicando conceitos de data lake, data warehouse e pipelines.
- Apoio ao time na definição da arquitetura de dados e automação de processos na plataforma.

Indústria de papel e celulose — Análise de Dados e Modernização de BI · 2024

- Migração e modernização de dashboards Power BI e indicadores; adequação de tabelas e validação de fontes para desempenho e qualidade dos dados.

JLira — Desenvolvedor Full Stack (Freelance) **ago/2023 – mai/2024**
Maceió, Brasil

- Desenvolvimento de aplicações e sites com Python/Django, Django REST Framework, React, Node.js e PostgreSQL, incluindo integração de front-end com APIs.

Grupo In4 — Desenvolvedor Full Stack (Delphi) **ago/2021 – jul/2023**
Maceió, Brasil

- Desenvolvimento front-end e back-end em Delphi para sistemas corporativos em produção, com gestão e priorização de tickets, versionamento (Git) e colaboração em equipe.

PROJETOS

Fuel Data Platform (ANP) — Plataforma de dados medalhão em GCP · Concluído
Python · Apache Beam · PySpark · Dataproc · Cloud Storage · Cloud Composer · BigQuery · Looker Studio · Docker

- Dataflow (Apache Beam) consome a API pública da ANP com paginação e grava Parquet no Cloud Storage (Bronze); Spark no Dataproc limpa e normaliza (Silver) e sobre isso são geradas métricas agregadas (Gold); tabelas externas e views no BigQuery alimentam dashboards no Looker Studio; orquestração com Cloud Composer.
- Particionamento Hive-style para partition pruning no BigQuery, clusters Dataproc efêmeros para controle de custo e isolamento de service accounts.

Clairemind — Assistente de IA pessoal no WhatsApp · 2026 – atual
Node.js · Claude Haiku 4.5 · function calling · Whisper · Evolution API · n8n · Google Calendar API · SQLite · Docker · GCP

- Assistente conversacional rodando inteiramente no WhatsApp: gerencia agenda, tarefas e lembretes, entende mensagens de voz e negocia reuniões com terceiros de ponta a ponta, em linguagem natural.
- Construído para confiabilidade: as ações da LLM passam por tools e as confirmações são geradas pelo código, então ele não confirma uma ação que não executou; a negociação assíncrona é uma máquina de estados explícita. Roda em GCP (nginx/SSL, SSH só via IAP, backups diários com restore validado, rate limit, teto de custo); ~41% de economia na LLM com prompt caching. Documentado publicamente no LinkedIn.

dbt no BigQuery · Em andamento

- dbt Core 1.12 sobre o dataset público thelook_ecommerce: modelos staging, intermediate e marts, tabela fato incremental, snapshot SCD tipo 2, data tests e freshness das sources, com dbt build no GitHub Actions.

COMPETÊNCIAS TÉCNICAS

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, integração de dados de ERP (Oracle), conectores de API (Zendesk, Jira), ingestão e transformação de dados
Linguagens	Python, SQL, PySpark, JavaScript
Dados	Modelagem relacional e dimensional, arquitetura Medalhão (Bronze/Silver/Gold), Data Lake, Data Warehouse, Data Mart, qualidade e governança de dados, FinOps, otimização de queries e custo, dbt (aprendendo)
Operação	Resolução de incidentes, análise de causa raiz, monitoramento e logging, documentação técnica
IA	RAG (Vertex AI, Gemini, Cloud Storage), LLM function calling, prompt caching
Ferramentas	Git, Docker, Jupyter Notebooks, Power BI, Looker Studio, Jira, JSON, npm/pip

CERTIFICAÇÕES

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)

FORMAÇÃO E IDIOMAS

Bacharelado em Engenharia Civil — Universidade Presbiteriana Mackenzie
Inglês — Avançado (uso profissional: leitura, escrita, reuniões e documentação técnica) · Português — Nativo