

Alan Joffre

Data Engineer · Analytics Engineer · AI Engineer

📍 São Paulo, Brasil (GMT-3) | ✉ alanjoffre@outlook.com | 📞 +55 11 91547-2650 | in/alanjoffre | github.com/alanjoffre

RESUMO PROFISSIONAL

Engenheiro de Dados / Analytics Engineer com foco em **produtos de dados end-to-end** — da arquitetura **lakehouse** (Databricks, Unity Catalog, Delta Lake, **dbt**) à entrega de análises prontas para decisão, com pipelines orquestrados (Airflow), **Data Quality**, lineage e governança ponta a ponta. Em transição consolidada para **AI Engineer**: RAG avaliado com rigor estatístico, fine-tuning de LLM (**QLoRA/vLLM**), agentes (**LangGraph**) e **MLOps** com avaliação como gate de CI — provado no projeto open-source **RodolA**.

EXPERIÊNCIA PROFISSIONAL

Data Engineer · AI Engineer — Ser Mais Digital (Autônomo)

fev/2025 – Presente

São Paulo (Híbrido) · Cliente de mobilidade e pagamentos — produto de auditoria e simulação de vale-pedágio

- Concebi e implementei a arquitetura **Medallion** (Bronze/Silver/Gold) completa em **Databricks** (Unity Catalog + Delta Lake) modelada com **dbt**, com pipelines de retry/alertas processando dezenas de milhões de transações/mês, consolidados em camadas com bilhões de linhas derivadas.
- Reduzi em **~90%** o tempo de apuração de casos analíticos recorrentes (de até 20 dias para 2 dias), liberando o time para análise em vez de operação manual.
- Análises e relatórios automatizados entregues recorrentemente à auditoria, com governança de acesso, logging, trilha de auditoria e versionamento.
- Data Quality e governança: dbt tests/contracts, masking de PII, backup/restore versionado e KPIs auditáveis com rastreabilidade ponta a ponta.
- Portfólio público OSS espelhando esta arquitetura (dbt + Airflow + dlt + Soda, CI/CD e dashboard ao vivo) como prova verificável do stack.
- Iniciativa de IA aplicada (projetos próprios, em paralelo): RodolA** — RAG, fine-tuning QLoRA/vLLM, agente LangGraph e MLOps com gate de CI; e **Aira** — integração de LLM (Claude) com controle de custo (ver Projetos).

Stack (cliente): Python · SQL · dbt · Databricks (UC, Delta, Medallion) · PySpark · Spark · Apache Airflow · Kafka · DuckDB · Azure

Stack (IA): Python · PyTorch · Hugging Face/Transformers · QLoRA · vLLM · LangGraph · Qdrant · MLflow · DVC

Cientista de Dados — SWM Consultoria (Autônomo)

out/2023 – jan/2025

- Projetos completos de Machine Learning: classificação, clusterização, previsão e séries temporais, com impacto direto em custos e operações.
- Pipelines de experimentação e tuning (Scikit-learn, XGBoost, LightGBM) com Explainable AI (SHAP, LIME).
- Redução de **20%** em custos logísticos com séries temporais; **+18%** de acurácia em detecção de anomalias (YOLOv5, OCR) e **40%** menos esforço manual.

Stack: Python · Scikit-learn · XGBoost · LightGBM · SHAP/LIME · Pandas · Airflow · Power BI

Data Engineer — SWM Consultoria (Autônomo)

set/2022 – ago/2023

- Pipelines **ELT** escaláveis, reduzindo o tempo de ingestão/carga em até **50%**; processamento de grandes volumes com Apache Spark e AWS Redshift.
- Arquitetura resiliente com S3, Glue, Redshift e orquestração tolerante a falhas; tuning de SQL (Redshift, PostgreSQL, BigQuery) até **70%** mais rápido.
- Automação de ingestão/transformação com Python e jobs agendados, eliminando falhas manuais.

Stack: Python · SQL · Spark · PySpark · Kafka · AWS (S3, Glue, Redshift, EC2) · PostgreSQL · BigQuery · Airflow

PROJETOS EM DESTAQUE

- RodolA** (open-source) 🌟: Engenharia de IA de ponta a ponta sobre dados públicos da ANTT — RAG avaliado (hit@5, **κ humano inter-anotador**), fine-tuning QLoRA (**F1 0,13→0,77**, NER jurídico) em vLLM, agente LangGraph e MLOps com avaliação como gate de CI. Demo pública ao vivo.
- toll-analytics-platform** (open-source): plataforma de dados end-to-end — dlt → dbt (Medallion, contracts, Semantic Layer, dbt Mesh) → Airflow + Cosmos → Soda → OpenLineage → Evidence.dev; CI/CD + Terraform.
- Aira** (SaaS em produção na AWS): backend assíncrono Python/FastAPI, PostgreSQL multi-org (migração em produção sem perda), integração LLM (Anthropic Claude) com controle de custo, CI/CD com SAST e segurança ponta a ponta.

FORMAÇÃO ACADÊMICA

Pós-graduação — Engenharia de Software com Inteligência Artificial — Universidade Anhanguera (jul/2026 a jan/2027)

RAG, agentes com LangGraph, LLMs, MLOps e deploy de modelos em produção — consolidando a transição para AI Engineer.

Bacharelado em Ciência de Dados — Universidade Anhanguera (jan/2024 a jun/2026)

Foco em modelagem estatística, machine learning, engenharia de dados e visualização analítica. Disciplinas-chave: Python, SQL, estatística aplicada, big data, bancos relacionais e NoSQL, pipelines ETL/ELT.

HABILIDADES TÉCNICAS

Lakehouse & Data Eng: dbt · Databricks (Unity Catalog, Delta Lake) · Spark/PySpark · Apache Airflow · dlt · Soda · ELT/ETL · Medallion · Data Quality · Lineage

Cloud & Infra: AWS (S3, Glue, Redshift, EC2) · Azure · Docker · Terraform · GitHub Actions (CI/CD)

Dados: SQL · PostgreSQL · BigQuery · DuckDB

IA / LLM Engineering: RAG · LLM · Fine-tuning (QLoRA) · vLLM · LangChain/LangGraph · PyTorch · Hugging Face/Transformers · Qdrant · MLOps (MLflow, DVC)

Machine Learning: Scikit-learn · XGBoost · LightGBM · SHAP/LIME · MLflow · séries temporais · visão computacional

Idiomas: Português (nativo) · Inglês técnico (leitura de documentação, APIs e artigos)