

Franklin Ferreira

Staff Data Engineer

Brazil · Remote/Global

Data Architecture · Big Data · Governance · FinOps · ML & Generative AI

franklins390@gmail.com | linkedin.com/in/franklins390 | github.com/franklin390 | franklin390.github.io

SUMMARY

Staff Data Engineer and data architect. I design the data architecture a whole company runs on, from warehouse to lakehouse, from batch to streaming, from classic machine learning to generative AI: more than 50% lower platform cost and ~3x fewer incidents. Still hands-on, with a technical team of 15+ people across Brazil and Romania.

>50%	~3x	~50	10+	100+	15+
lower data platform cost	fewer incidents and less rework	sources unified in one platform	Unity Catalog domains	Airflow pipelines at Clicksign	people led

CAREER PATH

○	○	○	○	○	●
Oct 2020	Jun 2021	Oct 2021	Aug 2024	Jul 2025	Jul 2026
Data Scientist	Data Architect	Data Team Lead	Data Engineer	Data Engineer – Tech Lead	Staff Data Engineer
Clicksign	Clicksign	Clicksign	Flutter Brazil	Flutter Brazil	Flutter Brazil

CORE COMPETENCIES

- Architecture & platforms:** Data architecture, dimensional modeling, data modeling, data warehouse, data lake, lakehouse, Databricks, Delta Lake, Apache Spark, Databricks Workflows, Amazon Redshift, streaming
- Engineering:** Python, PySpark, SQL, dbt, Apache Airflow, data pipelines, AWS, Terraform, Docker, Kubernetes, GitHub Actions, CI/CD
- Governance & reliability:** Data governance, data catalogs (Unity Catalog, AWS Glue Data Catalog, Hive Metastore), data quality, LGPD, observability, FinOps
- Data apps & AI:** Streamlit, Metabase, Google Analytics 4, machine learning, generative AI (Lovable)
- Leadership:** Technical team leadership, architecture & ways of working, priority setting

EXPERIENCE

- **Flutter Brazil** Aug 2024 – Present
 - **Staff Data Engineer** Jul 2026 – Present

A Databricks data platform used by thousands of people across the whole company, with a team of 15+ people (engineering, analytics and data science) across Brazil and Romania.

 - Set the architecture and ways of working for a team of 15+ people (10 as Tech Lead), with a wider scope across the whole platform.
 - Still hands-on in architecture, modeling and platform decisions. When the team needs it, I set priorities and unblock business decisions.

Technologies: Databricks, Unity Catalog, Delta Lake, Apache Spark, dbt, SQL, Python, AWS, Terraform, GitHub Actions
 - **Data Engineer – Tech Lead** Jul 2025 – Jun 2026

Led a team of 10 people (engineering, analytics and data science) on a Databricks platform of tens of TB.

 - Cut the data platform's cost by more than 50% with a FinOps effort (clusters, jobs, storage, governance policies), including cost forecasting models. Annual spend: six figures (USD).
 - Reduced incidents and rework by ~3x by rolling out CI/CD with trunk-based development.
 - Delivered decision-support dashboards, platform-wide observability and near-real-time business metrics monitoring (latency of a few minutes), and restructured Unity Catalog by domain (10+).
 - Defined the architectures and ways of working for almost the whole data team. Helped create the company's LTV and payback models.

Technologies: Databricks, Unity Catalog, Delta Lake, Apache Spark, dbt, SQL, Python, AWS, Terraform, Streamlit, Metabase
 - **Data Engineer** Aug 2024 – Jul 2025
 - Led the migration of the company's entire data infrastructure to Databricks: ~50 sources (Google Analytics 4, internal systems and third-party tools) in a single platform, with pipelines going from hours to minutes.

- Built the company's dimensional model and data lake from scratch with Python, dbt and SQL. Every area still uses that architecture today. Also contributed to the regulatory process that allowed the company to operate legally in Brazil.

Technologies: Python, dbt, SQL, Databricks, Google Analytics 4

● Clicksign

Oct 2020 – Aug 2024

○ Data Team Lead

Oct 2021 – Aug 2024

- One of the first engineers on Clicksign's data team. Grew the group from 4 to ~20 people across data engineering, data analysis and data quality assurance, and built the processes that sustained that growth.
- Led the data engineering, data analysis and data quality assurance teams and was directly responsible for the reliability of the data used in company decisions.

Technologies: Amazon Redshift, dbt, Apache Airflow, dimensional modeling

○ Data Architect

Jun 2021 – Aug 2024

- Architected practically every data structure the company had in that period, including 100+ Airflow pipelines (modeling and ingestion).
- Led the implementation of dbt on Redshift, which brought quality control, governance, easier modeling, observability and simpler orchestration.
- Built 10+ APIs for sending data to partners and third parties, and architected the data protection processes (in compliance with LGPD and aligned with GDPR) and the continuity of data delivery.

Technologies: Apache Airflow, dbt, Amazon Redshift, dimensional modeling, data governance

○ Data Scientist (Python Developer)

Oct 2020 – Jun 2021

- Business Intelligence team: scripts and integrations between databases, APIs and REST web services with Python, PostgreSQL and Apache Airflow, plus predictive analyses and modeling.

Technologies: Python, PostgreSQL, Apache Airflow, REST APIs, predictive modeling

● Let's Code from Ada *alongside Clicksign*

Jun 2022 – Jul 2024

○ Adjunct Instructor

Jun 2022 – Jul 2024

- Taught Big Data and Data Engineering courses to hundreds of students: database modeling, data warehouses, data lakes and lakehouses, software architecture and cloud development.

— EDUCATION

XP Educação — MBA in Data Engineering

Jul 2022 – Jul 2023

PUCRS — Specialization (Lato Sensu) in Data Science and Artificial Intelligence

Jun 2021 – Jul 2022

UEZO — Bachelor's degree in Computer Science

2015 – 2019

— CERTIFICATIONS

Data Analyst Training — Data Science Academy

Jan 2023

Advanced SQL, data analysis in Python and Power BI on data warehouses and data lakes.

Astronomer Certification: DAG Authoring for Apache Airflow — Astronomer

Dec 2022

Data Architect Training — Data Science Academy

Dec 2022

Data governance and LGPD, relational, NoSQL and data store modeling, ETL/ELT pipeline architecture, and automation with Spark, Airflow, Kafka and AWS Glue.

Astronomer Certification: Apache Airflow Fundamentals — Astronomer

Nov 2022

Data Scientist Training — Data Science Academy

Oct 2021

Applied machine learning with R, Python and Spark on the Hadoop ecosystem, with a focus on data visualization and storytelling.

Selected courses: AWS Cloud Computing Design and Administration; Data Pipeline Design and Automation; Data Engineering Bootcamp.

— TALKS & RESEARCH

Talk at TDC: Airflow — who said you can't process data in it?

Dec 2021

Undergraduate research at UEZO: recognition and detection techniques; edge-detection optimization for cells

Aug 2018 – Aug 2019

— LANGUAGES

Portuguese — Native or bilingual

English — Full professional

Spanish — Professional working