

GUILHERME XAVIER

Staff Data Engineer

Brazil • contact@guilhermexavier.dev • linkedin.com/in/guilherme-xavier • github.com/xavierguilherme

SUMMARY

Staff Data Engineer with 5+ years building cloud data platforms across fintech, gaming, and real estate. Deep experience with Snowflake, Airflow, dbt, and Python on GCP, AWS, and Azure, with a track record of automating revenue-critical workflows and cutting compute costs at scale. AWS Certified Data Engineer (DEA-C01) and SnowPro Core certified.

WORK EXPERIENCE

Riot Games

Jun 2026 – Present

Staff Data Engineer (Contractor)

- Designed and delivered an end-to-end console partner funnel pipeline on Airflow, dbt, and Databricks, giving business teams full visibility from wishlist to gameplay and surfacing conversion improvement opportunities.
- Cut the team's main daily job runtime from 7 hours to 30 minutes (93%) by adopting the dbt Fusion engine and converting models to incremental runs.
- Reduced dbt compute costs by 60% by redesigning model materialization strategies and implementing appropriate incremental logic across the project.

Mission Lane

Apr 2024 – Present

Senior Data Engineer (Contractor)

- Automated onboarding for 15+ debt settlement partners into a centralized data platform, streamlining payment and settlement workflows supporting \$700K+ in monthly revenue.
- Architected and delivered 5 end-to-end projects for collections and recoveries, integrating Snowflake with internal APIs to automate customer workflows and eliminate 80+ hours of manual work per week.
- Migrated legacy Dataproc jobs to a modern Airflow and dbt architecture, cutting monthly compute costs by 15% and improving query performance by 65% through optimized table design.
- Optimized the internal Snowflake-to-Pub/Sub export tool with batch messaging, reducing execution time from 30 minutes to 40 seconds (98%) across 30+ pipelines.
- Developed Airflow ETL pipelines processing 50K+ daily transactions into Google Pub/Sub, powering real-time payment handling for 100+ downstream consumers.
- Built a REST API serving layer between Snowflake and 10 core microservices, enabling low-latency real-time decisioning with stable data contracts and operational observability.
- Partnered with InfoSec to automate ingestion of sensitive customer data into the exfiltration detection platform, reducing threat detection time by 60%.

Cherre

Jun 2025 – Dec 2025

Senior Data Engineer (Fixed 6-Month Contract)

- Designed and delivered 30+ production data pipelines using Python, dbt, Airflow, Snowflake, and BigQuery to ingest and model real estate datasets for analytics and product use cases.
- Reduced Snowflake costs by 20% by optimizing warehouse usage, SQL performance, and data model design across high-volume analytical workloads.
- Improved Kubernetes infrastructure for data services, increasing deployment reliability, resource efficiency, and operational maintainability.

Nelogica

Jan 2022 – Apr 2024

Data Engineer

- Designed and built an in-house customer communications orchestrator processing 70K+ daily messages, cutting third-party costs by 65%, saving 80 engineering hours per week, and enabling full marketing self-service.
- Built and scaled a real-time event ingestion API processing 3M+ daily events from mobile, web, and desktop applications at 30-second latency, delivering transformed data to OpenSearch for analytics.
- Built and scaled 100+ Airflow pipelines integrating MySQL, SQL Server, ClickHouse, S3, OpenSearch, and third-party APIs (Jira, Zendesk, Mailchimp, Typeform).

- Built a REST API enabling teams to define and consume 120+ KPIs and data models across the company intranet, establishing the foundation for a data-driven culture organization-wide.
- Architected the company's first Data Lake, establishing a unified foundation for analytics consumption and data model experimentation across teams.
- Managed a 5TB OpenSearch cluster, implementing log ingestion pipelines and building 50+ visualizations for operational monitoring and analytics.

MG Info

Jun 2021 – Jan 2022

Data Engineer

- Led migration of 50 on-premises data sources (10 TB) to Azure for a mining company, delivering a modern cloud data platform on time.
- Developed 20 data pipelines on Azure Data Factory, Synapse, and Databricks, integrating SQL Server, REST APIs, and SharePoint into daily processing workflows.
- Implemented Delta Lake architecture on ADLS, consolidating 10 TB of data and improving query performance by 40%.
- Established CI/CD practices on Azure Pipelines across 5 repositories, ensuring code quality through automated testing.

FOUNDER EXPERIENCE

Soli, soli.vectorgrid.com.br: Territorial-intelligence platform for the Brazilian rural land market: geospatial analysis of 8M+ mapped rural properties combining environmental, land-registry, logistics, and market data (PostGIS, vector tiles, remote sensing).

Vitrini, vitrinicreative.com: AI-powered creative-generation platform producing brand-consistent social media and ad assets across formats (posts, stories, carousels, ads) from a one-time brand setup.

- Built both products end to end as a full-stack engineer: Nuxt.js (Vue) frontends, FastAPI backends, REST APIs with OAuth 2.0/OIDC, and PostgreSQL/PostGIS data layers.
- Own and operate the full production stack: Hetzner Cloud VMs running Dockerized services behind a Caddy reverse proxy, with automated CI/CD via GitHub Actions.
- Developed an MCP (Model Context Protocol) server for Vitrini, exposing brand-consistent creative generation to AI assistants and agent workflows.

TECHNICAL SKILLS

Programming Languages: Python, SQL, TypeScript, Java, Rust

Data Engineering & Processing: Apache Spark/PySpark, Databricks, Apache Airflow, dbt, Apache Kafka, Delta Lake, Apache Iceberg, DuckDB, Polars

Data Platforms & Databases: Snowflake, BigQuery, ClickHouse, PostgreSQL, MySQL, Redis, DynamoDB, OpenSearch

Backend & APIs: FastAPI, REST APIs, OAuth 2.0/OIDC, MCP (Model Context Protocol)

Cloud & Infrastructure: AWS, Google Cloud, Azure, Hetzner Cloud, Linux, Docker, Kubernetes, Terraform, Caddy

CI/CD & Observability: GitHub Actions, GitLab CI, Azure Pipelines, Grafana

Frontend: Nuxt.js (Vue), TypeScript

EDUCATION & CERTIFICATIONS

BSc, Computer Engineering, Federal University of Pelotas (2022)

Certifications: AWS Certified Data Engineer – Associate (DEA-C01) · SnowPro Core · Palantir Foundry Aware