

GUSTAVO RIBEIRO DOS SANTOS

Senior Data Engineer | Senior Cloud Engineer | Machine Learning Engineer | Google Developer Groups Organizer | TESU Alumni | Microsoft Student Ambassador | GitHub Program Member | Google Cloud AI Trusted Tester | Google Cloud Arcade Facilitator

ABOUT

Senior Data and Cloud Engineer with 10 years of experience in Data Engineering, specializing in Advanced SQL, Spark, Python, and multi-cloud providers (Google Cloud, Azure, AWS, OCI, IBM Cloud). Expertise includes Data Lakes/Lakehouses orchestration (medallion model), Data modeling, and developing automated batch/streaming pipelines (Lambda, Delta Lake, ETL/ELT). Multi-cloud expert with 6 years of experience, holding AWS certification and practice (S3, Glue, Lambda, DynamoDB, Athena, EventBridge), plus knowledge in GCP and Azure.

Specialist in Artificial Intelligence, with solid experience in Machine Learning and LLMs, focusing on MLOps for model development and cloud deployment (Vertex AI, Claude, Gemini).

Extensive experience working in global teams across the USA and Europe, with a proven track record in various sectors (CI&T, Pluxee, C6 Bank, Ford, Cognizant).

Leader of communities like GDG Organizer (Google Developer Groups, representing Google Cloud in Brazil) and Google Cloud AI Trusted Tester (early access to releases for feedback). International speaker at events including The Developer's Conference, Campus Party, Microsoft (USA), Build with AI (Peru), Google DevFest, Google IO Extended, and Google Developers Group Summit LATAM 2025 (Peru).

EXPERIENCE

CI&T — Senior Data Engineer — New York, New York

As a Data and artificial intelligence specialist, his main role is to ensure the robustness and innovation of the organization's ecosystem on Google Cloud, Microsoft Azure and AWS. He is responsible for designing and building cloud solutions that are not only scalable and highly available but also have artificial intelligence as a central element (AI-First).

His experience extends to the development of complex machine learning models for the banking, healthcare, and technology sectors, using advanced architectures such as Retrieval-Augmented Generation (RAG) and Reinforcement Learning, to deliver accurate and high-value results. Expert in building robust data ecosystems using Apache Airflow for advanced orchestration and Databricks for Big Data processing. Proven experience in operationalising AI models with Vertex AI and implementing DataOps practices (CI/CD, IaC), ensuring agility, governance, and scalability in delivering value through data. I complement this background with strategic expertise in designing solutions for high-load environments, focusing on technical mentoring of teams and optimising the performance and costs of distributed pipelines with Python

and PySpark. I seek to apply my six years of experience with AWS services (Lambda, Glue, DynamoDB) to drive AdTech platforms, ensuring reliable, monitorable, and secure data flows in large-scale architectures.

Keeggo — Cloud FinOps Engineer — São Paulo, Brasil

August 2024 - September 2025

Experienced professional in cost optimization and compliance, I add strategic value with expertise in data analysis, AI, and automation. I have a history of driving efficiency and reducing expenses in multi-cloud environments (Azure, GCP, AWS). My work stands out for my ability to identify and implement innovative solutions that minimize costs and ensure adherence to regulations. I participated in a crucial project for the Federal Government, managing sensitive data from the Official Gazette. I used data analysis and AI to extract information, identify patterns, and optimize acquisition processes, ensuring transparency. My approach combines strategic vision with technical execution, allowing organizations to achieve financial and operational objectives sustainably.

Cognizant — Data Engineer — São Paulo, Brasil

June 2021 - June 2024

Leadership in large-scale data processing on Google Cloud and Microsoft Azure, building ETL/ELT pipelines for batch and streaming data extraction, treatment, and cleaning. Data orchestration with Airflow, and design and development of scalable Data Lakes architectures. Experience in optimizing complex SQL query performance in BigQuery and data automation with Alteryx. Collaboration with business teams for Data Lakes validation and testing. Pioneer in promoting a Data-Driven culture, especially in the LATAM data team. Experience in cybersecurity applied to data, ensuring the security of Data Lakes/Lakehouses, identifying and mitigating vulnerabilities and critical information leakage risks. Responsible for leading and instructing training for interns and new employees in data (analysis, science, and engineering) and cloud computing (Google Cloud) areas. Definition, planning, and execution of the initial phase of an automated testing system using Generative AI (Vertex AI and Gemini), including monitoring and managing resources on Google Cloud.

Magna Sistemas — Senior Data Engineer — São Paulo, Brasil

June 2021 - June 2022

Development of web applications using JavaScript, Angular, and Liferay DXP, also having experience in integration and APIs using InterSystems (IRIS, Caché, Ensemble, COS, Zen, CSP), REST APIs with Python, Java 8+, and Spring Framework. Database management such as PostgreSQL, SQL Server, and Oracle, and with solid experience in quality and delivery, including unit and integration tests, and CI/CD.

OSM — Software Engineer — Brasília, Brasil

April 2020 - January 2021

Support, code review, and application development in legal and healthcare systems

for large Brazilian public agencies, using SQL, PL/SQL, Python, JavaScript, InterSystems (IRIS, Caché, ObjectScript, CSP). Development of ETL pipelines, data analysis, and integration with IRIS Data Platform and MDX multidimensional language, in addition to DBMS maintenance (PostgreSQL, SQL Server, Redis, Oracle, MongoDB).

Banco do Brasil — Software Engineering Intern — Brasília, Brasil

March 2018 - February 2020

Development of databases with MySQL and IBM DB2, data analysis and statistics using SAS (Enterprise Guide and Data Explorer Management), creation of dashboards with PowerBI and Spotfire, implementation of scripts for task execution and automation with Python and Go, assistance in application creation using Java and Spring Framework for the legal intranet and assistance with management information in web applications using IONIC, React, HTML, CSS, JavaScript, jQuery.

E D U C A T I O N

Thomas Edison State University

Associate's degree, Architecting Cloud Solutions (2024 - Present)

Universidade Católica de Brasília

Bachelor's degree in Software Engineering (2018-2023)

C E R T I F I C A T I O N S

- DP-900: Azure Data Fundamentals
- AZ-900: Azure Fundamentals
- CLF-C02: AWS Certified Cloud Practitioner
- Astronomer Certification for Apache Airflow Fundamentals
- Astronomer Certification for Apache Airflow Fundamentals
- Astronomer Certification DAG Authoring for Apache Airflow
- GitHub Foundations
- IBM Cloud Essentials

L A N G U A G E S

- **English** — IELTS (General Training) Score 8.5
- **French** — DELF Level B2.
- **Spanish** — Beginner

S K I L L S

- **Languages & Fundamentals:** Python (focus on *data manipulation*, code optimization), SQL (advanced), ETL/ELT, REST APIs.
- **Cloud & MLOps:** Google Cloud Platform (GCP), VertexAI, AWS, Microsoft Azure, Multi-cloud, DevOps, CI/CD.
- **Data & Workflow:** Data Engineering and Science, AI and Machine Learning, Gemini (Generative AI), Apache Airflow (Orchestration), Databricks, Alteryx, SAS.
- **Infrastructure & Automation:** Automation, Optimization, FinOps, Cybersecurity.
- **Databases:** SQL Server, MongoDB, IBM DB2, Cloud SQL, DynamoDB, Google BigQuery, Azure Data Lake Gen2, InterSystems IRIS/Cache.

ADDITIONAL INFORMATION

- Speaker at the Google Developer Groups Summit LATAM 2025, in Lima, Peru.
- Awarded the titles of Airflow Champion, in 2024, GDG Organizer, in 2023, for contributions made to the development and evolution of technologies;
- Coordinator of TDC (The Developer's Conference) in the 2024 editions in São Paulo, in the Cloud track, and in Brasília, in the Generative Artificial Intelligence track;
- Speaker at TDC (The Developer's Conference) in Brasília, 2024, with the theme "Cloud Computing and Sustainability: Reducing Costs and Increasing Efficiency in the Public and Private Sector";
- Speaker at the Campus Party editions of 2023 (CPBS 5, CP Goiás 3, CPBR 15), in the 2024 editions (CP Goiás 4, CBPSB 6, and CPBR 16) and in 2025, at CPBR 17;
- Organizer and licensed by Google, with the Google Developer Groups;
- Member of the Google Cloud AI Trusted Tester, Google Cloud Innovators, Google AI Developers, Microsoft Student Ambassador, Apple Developer Program programs;
- Leader of training in data areas (Analytics, Engineering, Visualization, and Science) for interns and new employees at Ford Motor Company;
- Professor and instructor of Google Cloud Platform, Artificial Intelligence - with a focus on VertexAI, AutoML, Gemini, and OpenAI, and Machine Learning for senior developers at Cognizant;
- Actively participated in the development of Ford Motor Company's first Data Lake in all of Latin America;
- Participation as an exhibitor at CBIE 2019 - Brazilian Congress on Informatics in Education;
- Participation in IBM Cloud Discovery and IBM's Bluehack Hackathon, in São Paulo, October 2018;