GUSTAVO MELO

Brasília, Brazil
+55 (61) 98289-6001 | gustavorsdemelo@gmail.com
github.com/gusrberto | linkedin.com/in/gusrberto

Software Engineering student focused on **Data Engineering and Analytics**. I have hands-on experience building end-to-end data pipelines — from ingestion via SFTP, APIs, and web scraping to modeling in **dbt** and delivery for analysis — including work in real production environments, handling sensitive data under **LGPD** (Brazil's data protection law) compliance and scalable data lake architectures (**Airflow**, MinIO, DuckDB, PostgreSQL). I combine this technical foundation in **Python/SQL** with familiarity in Big Data (**Spark**) and an analytical mindset to turn raw data into strategic insight.

PROFESSIONAL EXPERIENCE

DATA ENGINEERING INTERN — LabLivre · GovHub Project (Repository / Documentation)

BRASÍLIA, BRAZIL · FEB 2026 – PRESENT

Part of the GovHub data team, a LabLivre project that consolidates municipal datasets from different government agencies into a unified data lake, ensuring data quality, LGPD compliance, and availability for analysis.

- Implemented the dbt pipeline (bronze → silver → gold) for **Minha Casa, Minha Vida** (Brazil's federal affordable-housing program, FAR modality) developments, modeling construction physical and financial progress to power a national map, state-level dashboards, and project fact sheets used by public managers.
- Built ingestion DAGs in Apache Airflow (Python) to bring municipal datasets from SFTP into a data lake on MinIO, implementing file-hash idempotency and parallel multipart download/upload for files up to 37 GB over a low-bandwidth VPN.
- Built API and web scraping ingestion DAGs (Python/Requests) for FGV economic indicators (ICST, INCC-M), replicating the authentication flow of a legacy OutSystems/ASP.NET system to automate a previously manual collection process.
- Implemented a sensitive-data masking pipeline (CPF, ZIP code, address) using deterministic HMAC-SHA256 across 1,600+ files, preserving join integrity across datasets while ensuring LGPD compliance.
- Designed the lake's bronze layer using typed Parquet (Python + DuckDB), exposing data as PostgreSQL views and replacing roughly 2,000 untyped VARCHAR tables with a single incremental, idempotent architecture validated in production with zero data loss.

Skills: Python, Apache Airflow, dbt, MinIO (S3), PostgreSQL, DuckDB, REST APIs, Web Scraping, Docker, Git, Linux.

ACADEMIC PROJECTS

SCINEWSAI — Scientific Dissemination Platform with RAG · UnB (Repository / Documentation)

AUG 2025 – DEC 2025

Platform that monitors arXiv papers and uses Generative AI/RAG to translate them into accessible content via blog and newsletter; I worked on data ingestion, preprocessing, and pipeline orchestration.

- Built automated collection pipelines (scraping and APIs) and metadata preprocessing/enrichment to feed the paper catalog.
- Integrated retrieval and generation components (RAG) and orchestrated the flow with Docker/docker-compose across scrapers, the AI pipeline, and web services.

Skills: Python, Pandas, Selenium, BeautifulSoup, Docker.

EDUCATION

Software Engineering

UnB – University of Brasília
July 2021 – Present

Advanced English

FISK – Fundação Richard Hugh Fisk
January 2014 – December 2018