

# Rafaella Duarte

 rafaelladuarte |  rafaella-duarte-carvalho |  rafaelladuarte.github.io  
|  rafaella.d.d.carvalho@gmail.com |

## RESUMO

---

Engenheira de Dados Júnior com 3 anos de experiência em pipelines de dados, ETL/ELT e web scraping em Python. Atuação end-to-end no ciclo de dados, com integração a bancos relacionais e NoSQL, orquestração com Apache Airflow e uso de cloud (GCP e AWS). Experiência em APIs REST, mensageria e projetos com IA aplicada (RAG e LLMs).

## SKILLS

---

Linguagens	Python, Javascript, Bash, SQL
Engenharia de Dados	ETL/ELT, Web Scraping, APIs REST, Data Ingestion, RAG, Integração com LLMs
Banco de Dados	PostgreSQL, SQL Server, MongoDB, Elasticsearch, Redis e Neo4j
Cloud	Google Cloud Storage, Google Cloud BigQuery, Amazon S3
Orquestração/Infra	Apache Airflow, Linux Cron (crontab), Docker, RabbitMQ
Ferramentas/Frameworks	Git, FastAPI, Selenium

## EDUCAÇÃO

---

2019 - 2025 Bacharelado em Ciências Econômicas **Universidade Federal de Uberlândia**

## EXPERIÊNCIAS

---

**Engenheira de Dados - Legal Data, E-xyon (Remoto)** Jan 2025 - atual

- Modelagem e parametrização de pipelines no SQL Server, com procedures e jobs para processamento de dados.
- Enriquecimento de dados via APIs públicas de processos jurídicos.
- Desenvolvimento de APIs REST e atuação em pipelines analíticos no BigQuery (datalake).
- Manutenção de pipeline de OCR em Python com uso de LLMs.

**Desenvolvedora Python ETL - Flow2Go (Remoto)** Jan 2024 - Dez 2024

- Desenvolvimento de pipelines ETL em Python com web scraping e resolução de captchas para coleta de dados jurídicos.
- Integração, saneamento e enriquecimento de dados em MongoDB a partir de bases públicas.
- Criação de APIs REST em Python para summarização de dados e suporte a dashboards analíticos.

**Engenheira de Dados Júnior - Linx, StoneCo (Remoto)** Nov 2022 - Out 2023

- Automação e escalabilidade de pipelines de web scraping em Python para múltiplos marketplaces.
- Replicação de dados do Elasticsearch para Google BigQuery e ingestão em Google Cloud Storage e Amazon S3.
- Monitoramento de pipelines de dados e filas RabbitMQ em produção, com execução de jobs via Linux Cron.

**Analista de Dados - Nodis, StoneCo (Estágio - Remoto)** Jan 2021 - Nov 2022

- Criação de pipelines ETL e rotinas de web scraping em Python para marketplaces.
- Integração de dados em PostgreSQL, MongoDB e Elasticsearch.

- Integração de dados com a plataforma VTEX via APIs e desenvolvimento de dashboards no Google Data Studio.

## PROJETOS

---

### **Woman A3Data Challenge: Hackathon Engenharia de Dados - [Link da Solução](#)**

Pipeline ELT para tratamento e análise da base CNPJ, com automação, web scraping e carga em Google BigQuery. **2º lugar**. Tecnologias: Python, Google BigQuery, Google Storage.

### **Assistente IA com RAG para Identificação de Aves - [Link da Solução](#)**

Aplicação em Streamlit com pipeline RAG e integração com APIs de LLMs para recuperação contextual de informações a partir de descrições textuais. Tecnologias: Python, Streamlit, RAG, APIs de LLMs.

### **BookTrack API: Gerenciamento de Livros com FastAPI - [Link da Solução](#)**

API REST para gerenciamento de catálogo de livros, com CRUD, autenticação e documentação automática via Swagger. Tecnologias: Python, FastAPI, SQLAlchemy, Alembic, PostgreSQL, Docker.

### **Web Scraping de Imobiliárias com Apache Airflow - [Link da Solução](#)**

Pipeline ETL para coleta e análise de dados imobiliários, com orquestração via Airflow e carga em PostgreSQL e MongoDB. Tecnologias: Python, Apache Airflow, PostgreSQL, MongoDB.