

Rafaella Duarte

 [rafaelladuarte](#) |  [rafaella-duarte-carvalho](#) |  [rafaelladuarte.github.io](#)
|  rafaella.d.d.carvalho@gmail.com |

RESUMO

Graduada em Ciências Econômicas, com 3 anos de experiência em engenharia e análise de dados utilizando Python. Atuo em todo o ciclo de dados, incluindo coleta via web scraping, rotinas ETL/ELT, criação de APIs REST e integração com bancos de dados relacionais e não relacionais. Atualmente, trabalho como engenheira de dados júnior CLT na Legal Data da E-xyon e desenvolvo projetos pessoais para estudos focados em Big Data, explorando tecnologias como Spark, Airflow, ecossistema Hadoop e computação em nuvem. Tenho também familiaridade com RabbitMQ, Google Cloud Storage, BigQuery e Amazon S3.

SKILLS

Linguagens	Python, Javascript, Bash, SQL
Banco de Dados	PostgreSQL, SQL Server, MongoDB, Elasticsearch, Redis e Neo4j
Cloud	Google Cloud Storage, Google Cloud BigQuery, Amazon S3
Ferramentas	Git, RabbitMQ, Docker, Apache Airflow, FastAPI, Selenium
Metodologias	Web Scraping, ETL/ELT, API Rest, RAG, Integração com LLMs

EDUCAÇÃO

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

EXPERIÊNCIAS

Legal Data, E-xyon - Engenheira de Dados Junior Jan 2025 - atual
CLT Tempo Integral - Remoto

Flow2Go - Desenvolvedora Python ETL Jan 2024 - Dez 2024
Freelancer - Remoto

- Desenvolveu Web Scraping com resolução de captchas para coleta de dados jurídicos (CNA e CNSA).
- Criou scripts Python para integração e saneamento de dados no MongoDB.
- Realizou integração e enriquecimento de dados com bases públicas (CNPJ, IBGE, CEP, SFN).
- Desenvolveu API Python para sumarização e criação de dashboards de análise.

Linx, StoneCo - Engenheira de Dados Junior Nov 2022 - Out 2023
CLT Tempo Integral - Remoto

- Automatizou e expandiu rotinas de Web Scraping para múltiplos marketplaces (Amazon, Mercado Livre, Magalu, etc.).
- Desenvolveu bibliotecas internas em Python para otimização de processos.
- Replicou dados do Elasticsearch para BigQuery, viabilizando análises de catálogo de produtos.
- Monitorou filas no RabbitMQ e pipelines em Kubernetes (Cron Jobs).
- Implementou ingestão de dados no Google Cloud Storage e Amazon S3.

Nodis, StoneCo - Analista de Dados Jan 2021 - Nov 2022
Estágio Meio Período - Remoto

- Criou e manteve rotinas de Web Scraping em Python para marketplaces.
- Desenvolveu pipelines ETL para PostgreSQL, MongoDB e Elasticsearch.

- Integrou dados com a plataforma VTEX via API.
- Produziu dashboards no Google Data Studio para análise de KPIs internos.

PROJETOS

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

Desenvolvimento de pipeline ELT para tratamento e análise da Base CNPJ, com automação, web scraping e integração em BigQuery. Foco em limpeza, estruturação e visualização de dados. **2º lugar**. Tecnologias: Python, Google BigQuery, Google Storage.

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

Aplicação com pipeline RAG e interface Streamlit para identificar aves a partir de descrições textuais. Integração de LLM via API para recuperação contextual de informações. Tecnologias: Python, Streamlit, RAG, APIs de LLMs.

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

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

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

Pipeline ETL para coleta e análise de imóveis em Uberlândia-MG, com carga em PostgreSQL e MongoDB e dashboard no Power BI. Tecnologias: Python, Airflow, ETL, PostgreSQL, MongoDB, Power BI.