

Ricardo Júnio Santos Marinho

Engenheiro de Dados

+55 61 99265-5144 | ricardomarinho.vaga@gmail.com | Brasília, DF
linkedin.com/in/ricardojunioo | rmarinho-port.vercel.app

RESUMO

Estudante de Engenharia de Software atuando como **Engenheiro de Dados**, com experiência em **modernização de estruturas de dados, pipelines em nuvem** e automação de **ingestão, transformação e disponibilização de dados**. Integra **sistemas legados à nuvem** via **APIs** e modela **dados relacionais e não estruturados** com **Python e SQL**, apoiando projetos de **ciência de dados e machine learning** em ambiente cloud. Inglês avançado (C2 – EF SET).

EXPERIÊNCIA PROFISSIONAL

Estágio em Engenharia de Dados

02/2025 – Presente

Ministério das Cidades

Brasília, DF

- Integrei **sistemas legados** migrando **100%** de uma base de **Python 2 para Python 3.8**, modernizando os fluxos de dados e elevando a estabilidade do sistema
- Construí e automatizei **pipelines de ETL de ingestão, transformação e disponibilização** de dados relacionais e não estruturados, tornando o processamento **~30%** mais rápido
- Desenvolvi aplicação desktop em Python que automatizou os testes do banco de dados, economizando **~50%** do tempo de validação da equipe
- Documentei** os processos de engenharia de dados executados, padronizando a rotina do time
- Stack:** Python, SQL, Pandas, APIs REST

Especialista de Sistemas (Freelancer)

02/2025 – Presente

Workana

Brasília, DF

- Arquiteto de software** da livraria **Livoretto: modelagem de dados** (diagramas EER) e integração completa do banco de dados de grande porte, com **30+ entidades**
- Modelei e otimizei o banco de dados** (SQL) de uma plataforma médica (prontuários e fluxos de prescrição)
- Construí **integrações via API** entre sistemas distintos (incluindo **sistemas ERP**) e automatizei **100%** das rotinas manuais de cadastro dos clientes
- Stack:** TypeScript, PHP, SQL, Python

PROJETOS

Índice de Inserção Urbana de Empreendimentos Habitacionais (MCMV/Novo PAC)

- Geocodifiquei empreendimentos habitacionais contra o **CNEFE 2022** do IBGE (**~90 milhões de endereços**) via **record linkage difuso**, cruzando setores censitários, escolas (INEP), saúde (CNES), rede viária (OpenStreetMap) e GTFS de transporte para gerar um **score composto de acesso urbano** por empreendimento. **Stack:** Python, PostgreSQL/PostGIS, dbt, Airflow, Docker

Pipeline de Dados com Apache Airflow

- Orquestrei **mais de 20 fluxos** via DAGs, **integrando sistemas legados à nuvem** (Web APIs → Databricks/SharePoint), com camadas bruto → processado → pronto e backups automatizados. **Stack:** Airflow, Python, APIs REST

Qualidade de Dados e ETL (Minha Casa Minha Vida)

- Criei rotinas de **validação e monitoramento** de ingestões em base governamental, elevando a qualidade dos dados e gerando relatórios para gestores. **Stack:** PostgreSQL, ETL

Anonimização de Microdados em R (PNAD Contínua – IBGE)

- Pipeline para processar **13 anos** de microdados do IBGE (2012–2024) com **Controle de Sigilo Estatístico (SDC)**, anonimizando dados e mantendo o valor analítico. **Stack:** R, Python

FORMAÇÃO ACADÊMICA

Universidade Católica de Brasília

08/2024 – 12/2028 (em andamento)

Bacharelado em Engenharia de Software

HABILIDADES

Linguagens: Python, SQL, R, Java, TypeScript, JavaScript

Bancos de Dados: PostgreSQL/PostGIS, MySQL, MongoDB, Oracle APEX

Engenharia de Dados: Pipelines ETL/ELT, ingestão de dados, integração via APIs REST e sistemas ERP, modelagem de dados, dados geoespaciais, Airflow, dbt, PySpark

Cloud & Ferramentas: GCP (BigQuery, Cloud Storage), Databricks, Azure (Data Factory, Azure Functions), Microsoft Fabric, AWS (S3), Power BI, Pandas, Docker, Git, CI/CD (GitHub Actions, Cypress)

Idiomas: Inglês (C2 – EF SET), Espanhol

CERTIFICAÇÕES

DIO: Formação Databricks Data Engineer; Ecossistema de Big Data: Databricks e PySpark; Big Data em Ambiente em Nuvem (Amazon S3); Aceleração Microsoft – Gestão de Dados & IA (Data Factory); Introdução à Nuvem com Azure; SQL Database Specialist; Conceitos e Melhores Práticas com Bancos de Dados PostgreSQL; Análise de Dados com Python e Pandas; Python para Machine Learning na Prática; Dashboards Interativos com Power BI; CI/CD com Cypress | **HashTag Treinamentos:** Curso Intensivo de Power BI