

Contato

rafaella.d.d.carvalho@gmail.com

www.linkedin.com/in/rafaella-duarte-carvalho (LinkedIn)
github.com/elladarte (Portfólio)

Principais competências

ETL (Extração, transformação e carregamento)

Processamento de dados

Pipeline de dados

Certifications

Challenge Woman Hackathon
Engenharia de Dados

Formação Modelagem de Dados

Python Fundamentos para Análise de Dados

SQL Completo

Excel - Básico | Intermediário | Avançado

Rafaella Duarte

Data Engineer Júnior | Python | Web Scrapping | Big Data | SQL/NoSQL | ETL/ELT | RabbitMQ | APIs Rest

Uberlândia, Minas Gerais, Brasil

Resumo

Graduanda em Ciências Econômicas, entusiasta em Big Data e apaixonada por processamento e infraestrutura de dados. Atuo há dois anos com processamento de dados com a linguagem de programação Python. Possuo experiência de 3 anos em desenvolvimento e manutenção de rotinas de ETL/ELT, web scraping para coleta de informações na internet; APIs Rest para uso interno e externo; pipelines de dados para ingestão de dados em buckets e Bancos de dados relacionais e não-relacionais. Além disso, já tive contato com atividades de provisionamento de VMs na Embratel e Brascloud, e deploy de Cron Jobs no Kubernetes. Por fim, sou artesã (Instagram: @ella_darte) nas horas vagas, apaixonada por corrida de rua e xadrez, e amo ensinar o pouco que sei em Matemática e Física para crianças.

SKILLS PRIMARIAS: Python, Web Scrapping, ETL/ELT, Git, Bash, PostgreSQL, MongoDB, Elasticsearch, Google Big Query, RabbitMQ

SKILL SECUNDÁRIAS: Redis, Google Cloud Storage, Google DataStudio, Microsoft Power BI, Microsoft Excel, API Rest, Javascript, GraphQL, , Amazon S3, Docker

SKILL ACADÊMICAS: Stata, EViews, LaTeX, QGIS, R

INTERESSES: Hadoop, Spark, Airflow, Kafka, Ansible, AWS, GCP

Site pessoal: <https://rafaelladuarte.github.io/>

Experiência

e-Xyon

Engenharia de Dados Júnior

janeiro de 2025 - Present (2 meses)

Flow2go - Brasil

Engenharia de Dados Júnior

janeiro de 2024 - Present (1 ano 2 meses)

- Desenvolveu projetos de Web Scraping com resolução de captchas, coletando dados de advogados e associações de advogados nos sites da CNA e CNSA.
- Criou scripts Python para integração e correção de bases de dados no MongoDB, garantindo a qualidade e consistência dos dados.
- Integração de bases de dados públicas do governo, como CNPJ, IBGE, CEP e SFN, para análise e enriquecimento de dados.
- Desenvolveu API Python para sumarização de dados e criação de dashboards para visualização e análise.

Linx

Engenharia de Dados Junior

novembro de 2022 - outubro de 2023 (1 ano)

- Desenvolveu e manteve rotinas de Web Scraping em Python, expandindo a coleta de dados para novos marketplaces (Amazon, Mercado Livre, Magalu, Americanas, Casas Bahia, Droga Raia).
- Criou bibliotecas internas em Python para automatizar operações repetitivas nas rotinas de Web Scraping.
- Replicou dados do catálogo de produtos do Elasticsearch para o BigQuery, utilizando rotinas de carga em Python para análises.
- Monitorou e gerenciou filas no RabbitMQ, garantindo o fluxo contínuo de dados entre os sistemas.
- Desenvolveu e manteve APIs REST internas em Python, integrando sistemas e processos de dados.
- Realizou ingestão de dados no Google Cloud Storage e Amazon S3, otimizando a gestão de grandes volumes de dados.
- Gerenciou e manteve Cron Jobs no Kubernetes, garantindo a execução contínua das pipelines de dados.
- Provisionou VMs na Embratel e Brascloud, escalando o processamento das rotinas de Web Scraping.

Nodis

Estagiário de Análise de Dados

janeiro de 2021 - novembro de 2022 (1 ano 11 meses)

- Desenvolveu e manteve rotinas de Web Scraping em Python para extrair dados de marketplaces (Amazon, Mercado Livre, Magalu) e enriquecer o catálogo de produtos.
- Implementou e gerenciou processos de ETL em Python, atualizando bases de dados no PostgreSQL, MongoDB e Elasticsearch, e realizando cargas de dados.

- Monitorou e gerenciou filas no RabbitMQ, garantindo o fluxo adequado dos dados entre os processos de ETL.
- Desenvolveu e manteve APIs REST internas em Python, integrando sistemas e processos de dados.
- Gerenciou e atualizou bases de dados em PostgreSQL, MongoDB e Elasticsearch, além de realizar integração de dados com a plataforma VTEX via API.
- Desenvolveu dashboards no Google Data Studio, visualizando dados do PostgreSQL e Elasticsearch para análise interna.

CNPq

Iniciação Pesquisa Científica - Bolsista

agosto de 2021 - julho de 2022 (1 ano)

Uberlândia, Minas Gerais, Brasil

Tema: A Dinâmica da Taxa de Inflação no Brasil e o Papel das Políticas Monetária e Fiscal

- Extraí e tratei dados macroeconômicos do IPEA DATA (CSV), utilizando Excel para deflacionamento, normalização e cálculo de médias móveis.
- Estimou Modelos Autorregressivos de Defasagens Distribuídas (ARDL) no EViews, analisando a relação entre inflação e desemprego sobre as perspectivas de políticas monetária e fiscal.
- Realizou testes estatísticos (multicolinearidade, heterocedasticidade e autocorrelação) para validar os modelos.
- Interpretou os resultados e elaborou relatórios estatísticos sobre a dinâmica da inflação no Brasil.

Formação acadêmica

Universidade Federal de Uberlândia - UFU

Bachelor's degree, Economics · (2019 - 2023)

Universidade Federal de Uberlândia

Bachelor's degree, Aeronautical Engineering · (2015 - 2018)