**Documento de Arquitetura Técnica e Integração com ML**

**1. Visão Geral**

O Tech Challenge Scrap Book é uma aplicação desenvolvida para realizar web scraping no site Books to Scrape, coletando informações de livros e categorias. Através de uma API RESTful, o usuário pode acessar diferentes endpoints para obter dados organizados sobre os livros disponíveis no site, filtrando por critérios como categoria, título e classificação.

**2. Pipeline de Dados**

**Fluxo Geral:**

**Diagrama

O conteúdo gerado por IA pode estar incorreto.2.1 Ingestão**

* Fontes: Raspagem realizada do website https://books.toscrape.com/.
* Ferramentas: Script Python.
* Frequência: Em tempo real.

**2.2 Processamento**

* Tecnologias: Python.
* Etapas: limpeza, transformação, enriquecimento, feature engineering.
* Armazenamento: Arquivo CSV.

**2.3 API**

* Frameworks: Flask
* Autenticação: JWT.
* Endpoint: Conforme descrito no arquivo readme.md

**2.4 Consumo**

* Hospedado no Render, pode ser acessível via API.
* Dashboards (Power BI, Tableau, Streamlit)
* Aplicações web/mobile
* Cientistas de Dados via notebooks

**3. Arquitetura com Escalabilidade**

**3.1 Componentes**

* Microsserviços: Pode ser containerizado com Docker e orquestrados com Kubernetes
* Mensageria: Não utiliza mensageria
* Armazenamento: Arquivo CSV

**3.2 Escalabilidade**

Possibilidades para escalonamento

* Auto Scaling com Kubernetes
* Balanceamento de carga com NGINX ou Load Balancer da cloud

**4. Cenário de Uso para Cientistas de Dados / ML**

**4.1 Acesso aos Dados**

* Acesso controlado via APIs

**4.2 Ferramentas**

* Postman
* Linux (comando curl)
* Repositório Git para versionamento de notebooks