

SISTEMA DE RECOMENDAÇÃO DE FILMES

Integrantes: Pedro Jersey, Gabriel Queiroz, Matheus Lobo

CONTEXTO DE USO

Uma empresa de streaming de filmes e séries enfrenta grandes desafios em melhorar a experiência dos seus usuários. O sistema de recomendação atual não consegue atender com precisão aos gostos específicos de cada assinante, o que prejudica a descoberta de novos conteúdos interessantes. Isso impacta diretamente a satisfação e o engajamento dos usuários, que podem sentir dificuldade em encontrar filmes e séries relevantes. Em um mercado altamente competitivo, a empresa está comprometida em implementar um sistema mais assertivo, que facilite a busca pelo conteúdo ideal e torne a experiência de assistir ainda mais prazerosa e personalizada.



OBJETIVO DO PROJETO

O objetivo deste projeto é construir um sistema de recomendação mais eficaz, focado em resolver os desafios enfrentados pela empresa, mas com impacto direto na experiência do usuário. As análises e melhorias serão orientadas para oferecer recomendações mais precisas e personalizadas aos assinantes, ajudando a empresa a aumentar a satisfação e o engajamento na plataforma.



FERRAMENTAS UTILIZADAS



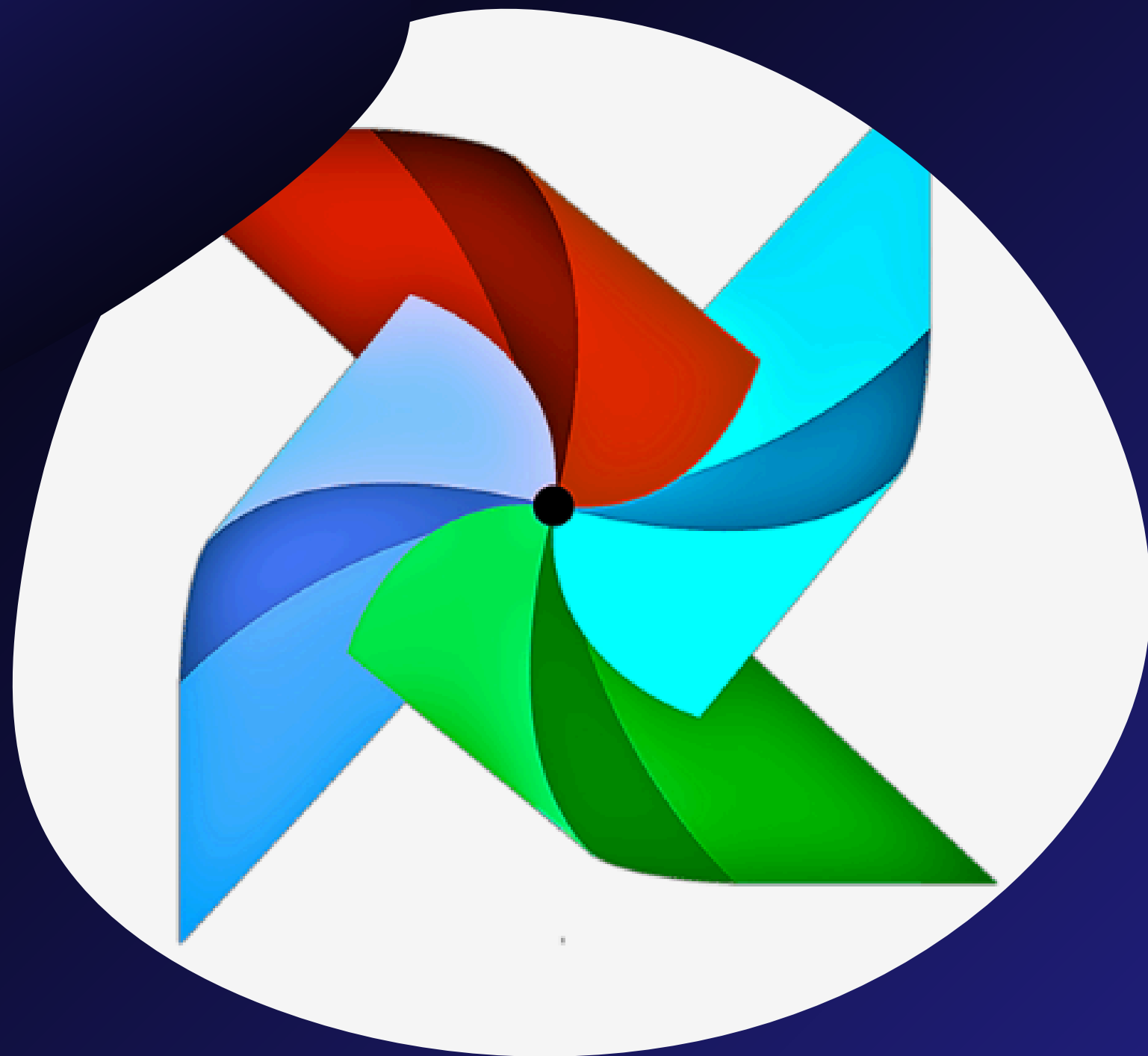
MOVIE LENS

A base de dados MovieLens, fornecida pelo GroupLens, contém informações detalhadas sobre filmes e avaliações de usuários. Ela inclui dados como títulos, gêneros, notas atribuídas, além de timestamps que permitem análises temporais. Essa base é amplamente utilizada para estudos acadêmicos e no desenvolvimento de sistemas de recomendação, sendo uma excelente escolha para treinar algoritmos e realizar análises aprofundadas sobre preferências e padrões de comportamento dos usuários.



CLICKHOUSE

Com o ClickHouse, teremos um banco de dados poderoso e altamente otimizado para análise de grandes volumes de informações. Sua arquitetura orientada a colunas oferece um desempenho excepcional para consultas analíticas, sendo ideal para processar e agrupar dados como popularidade de filmes, médias de avaliações e tendências de gênero. A capacidade de lidar com milhões de registros em tempo real permitirá uma experiência fluida e ágil para o dashboard, enquanto sua escalabilidade assegura que o sistema possa crescer conforme as demandas aumentem.



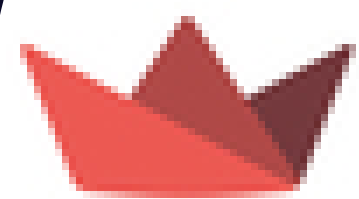
APACHE AIRFLOW

O Apache Airflow será utilizado para orquestrar e monitorar os fluxos de dados necessários para alimentar o sistema. Ele automatiza e organiza as tarefas de extração, transformação e carregamento (ETL), garantindo que as informações sejam atualizadas regularmente no Data Lake e estejam sempre disponíveis para o dashboard e a API. Com seu modelo de fluxo visual, o Airflow proporciona fácil rastreamento de processos e identificação de falhas, garantindo a confiabilidade e eficiência das operações.



FAST API

Para garantir uma comunicação eficaz entre o Banco e o dashboard, utilizaremos o FastAPI para desenvolver uma API REST. Essa API será responsável por realizar operações de leitura e atualização no Data Lake, permitindo que o dashboard obtenha dados de maneira rápida e em tempo real. Além disso, o FastAPI garante a segurança e a escalabilidade do sistema, oferecendo um bom desempenho para manipulação de grandes volumes de dados.



Streamlit

STREAMLIT

Com o Streamlit, seremos capazes de desenvolver visualizações dinâmicas e interativas, que facilitarão a análise de métricas importantes como a popularidade dos filmes, notas médias, e variações de avaliação por gênero.



DOCKER

O Docker será a base para toda a aplicação, permitindo a criação de containers isolados para cada componente do sistema. Isso garante portabilidade, consistência e facilidade na implantação, independentemente do ambiente. Com o Docker, será possível gerenciar a API, o dashboard, o banco de dados e o Apache Airflow de forma integrada, simplificando atualizações e escalabilidade. Além disso, o uso de containers otimiza recursos, assegurando desempenho e flexibilidade no desenvolvimento e na manutenção do sistema.

ESTRUTURA DO PROJETO
