

Documentação do Projeto - MovieStream Analytics

Introdução

Neste projeto, fui contratado pela empresa fictícia **MovieStream Analytics**, que atua no mercado de locação e streaming de filmes. O objetivo do desafio era criar uma solução de análise de dados robusta e escalável, capaz de tratar, modelar e apresentar insights importantes para os tomadores de decisão da empresa. Utilizando dados de uma aplicação de locação de filmes, o projeto foi estruturado em várias etapas que envolvem desde a ingestão de dados até a visualização de métricas analíticas no Power BI.

Arquitetura da Solução

A arquitetura adotada para o projeto foi desenhada com o intuito de fornecer uma solução eficiente e que atenda às necessidades de análise de dados. Ela foi estruturada da seguinte forma:

1. **PostgreSQL**: A base de dados foi configurada no PostgreSQL, onde os dados da locação de filmes foram armazenados. A estrutura de tabelas incluiu informações detalhadas sobre clientes, filmes, categorias, locações e pagamentos.
2. **DBT (Data Build Tool)**: Para a transformação dos dados e criação de modelos analíticos, utilizei o DBT. O DBT permitiu que eu criasse modelos de dados limpos, bem estruturados e otimizados para análise. Ele é ideal para transformar dados brutos em informações acessíveis e úteis para decisões de negócio.
3. **Airflow**: Para garantir que os dados estivessem sempre atualizados, o Airflow foi utilizado para orquestrar o processo de ingestão e transformação de dados. Com ele, consegui agendar e executar tarefas automaticamente, garantindo que os dados estivessem sendo atualizados regularmente.
4. **Power BI**: Após a modelagem dos dados, utilizei o Power BI para criar dashboards interativos. Esse painel de visualizações permite que os stakeholders da empresa visualizem métricas importantes de forma clara e intuitiva, ajudando na tomada de decisões.

Estratégia de CDC (Change Data Capture)

Como parte da estratégia de **Change Data Capture (CDC)**, foquei nas tabelas com maior volume de dados, como **rental** e **payment**, para garantir que as alterações nessas tabelas fossem capturadas e processadas de maneira eficiente. Isso assegura que os dados utilizados para análise estejam sempre atualizados, refletindo as mudanças mais recentes e sem a necessidade de intervenção manual.

A aplicação do CDC foi crucial para a solução, pois possibilitou que as atualizações de dados fossem feitas de forma automatizada e em tempo real, sem afetar a performance do sistema e mantendo as análises consistentes.

Execução das DAGs no Airflow

O Airflow foi configurado para gerenciar o fluxo de dados no projeto, permitindo o agendamento de tarefas e a execução periódica das transformações. As DAGs (Directed Acyclic Graphs) garantiram que o processo de ETL (Extração, Transformação e Carga) fosse executado automaticamente e de forma controlada, mantendo o pipeline de dados funcionando sem problemas. Além disso, a execução periódica das DAGs assegurou que as análises estivessem sempre baseadas em dados recentes.

Modelos DBT Criados

Com o DBT, criei três modelos principais para fornecer insights valiosos para o negócio:

1. **mart_customer_lifetime_value**: Esse modelo calcula o valor total gasto por cada cliente ao longo do tempo, além de considerar o tempo desde a sua primeira locação. Isso ajuda a entender a lealdade dos clientes e identificar oportunidades para campanhas de marketing direcionadas.
2. **mart_film_popularity**: Este modelo avalia quais filmes são os mais alugados, agrupados por mês e ano. Ele permite que a empresa compreenda quais tipos de filmes estão mais em alta e quais são os preferidos dos clientes.
3. **mart_store_performance**: Aqui, analisei o desempenho de cada loja com base em métricas como quantidade de locações e receita gerada. Esse modelo fornece uma visão clara sobre quais lojas estão performando melhor e onde a empresa pode melhorar suas operações.

Visualizações no Power BI

Após processar e modelar os dados, criei um conjunto de visualizações no **Power BI**. Essas visualizações foram desenhadas para fornecer insights valiosos sobre o desempenho da empresa, com destaque para as seguintes métricas:

- **Receita por Categoria de Filme**: Este gráfico permite visualizar quais categorias de filmes estão gerando mais receita, o que pode ajudar a empresa a focar mais em determinadas categorias.
- **Locações por Mês**: Uma linha do tempo que mostra o volume de locações realizadas a cada mês, oferecendo insights sobre tendências e sazonalidades.
- **Clientes Ativos**: Uma métrica que identifica os clientes mais recentes, aqueles que fizeram locações recentemente, para ajudar a empresa a entender sua base de clientes e a focar em estratégias de retenção.

Essas visualizações ajudam os stakeholders a tomar decisões informadas sobre onde concentrar esforços e identificar áreas que precisam de mais atenção.

Conclusão

Este projeto foi uma excelente oportunidade de aplicar e combinar diversas ferramentas e tecnologias de dados. A solução final oferece uma visão clara e estratégica das operações da empresa **MovieStream Analytics**, permitindo que ela tome decisões baseadas em dados reais e sempre atualizados. Ao usar o PostgreSQL, DBT, Airflow e Power BI, consegui criar uma pipeline robusta que transforma dados brutos em insights valiosos de forma eficiente e escalável.