

Descrição do desafio

DESAFIO TÉCNICO

Sua missão é construir uma pequena solução de obtenção, armazenamento e visualização de dados. Sua solução deve ter um componente que obtenha - de maneira automatizada e recorrente - dados sobre filmes e armazene-os em um banco de dados modelado por você e um componente de visualização dos dados armazenados.

O seu objetivo deve ser entregar uma solução completa, mas não desanime caso não consiga fazer alguns dos itens, todo o esforço será avaliado durante a apresentação do desafio e as dificuldades encontradas discutidas.

Disponibilize o seu trabalho em repo público no github e envie o repo para: **thiagoluiz.paula@dasa.com.br**.

Escopo do Desafio

Seu componente responsável por realizar o processo de ETL deverá utilizar como fontes de dados as APIs disponibilizadas pela The Movie Database (<https://developers.themoviedb.org/3>) para buscar e armazenar dados sobre filmes e seus respectivos atores, diretores e produtoras. Seu componente de visualização de dados pode ser construído de maneira livre: um conjunto de queries SQL parametrizadas, uma planilha em Excel conectada diretamente no banco de dados, um dashboard construído com Power BI ou qualquer outra ferramenta de visualização de dados de sua escolha. O objetivo a ser alcançado pela camada de visualização de dados é **responder às seguintes perguntas**:

- Dentre os filmes em cartaz, qual o filme com maior receita e qual sua receita e seu orçamento?
- Dentre as produtoras com filmes em cartaz ou filmes a serem lançados, qual teve o maior lucro (receita – orçamento) acumulado nos últimos 10 anos? Considere que o lucro é distribuído igualmente entre as companhias produtoras.
- Que diretor(a) dirigiu filmes com o maior orçamento acumulado dos últimos 20 anos?
- Quais os 3 gêneros com maior número de filmes nos últimos 5 anos
- Quais os países com a maior quantidade de filmes em cartaz produzidos em seus territórios?

Restrinja as pessoas e companhias produtoras somente àquelas relacionadas a filmes atualmente em cartaz ou a serem lançados em breve.

Recomendações

Não há nenhuma restrição a respeito da tecnologia ou infraestrutura de banco de dados a ser utilizada, desde que seja possível reproduzir o setup da solução. Provedores de nuvem possuem ofertas para pequenos bancos de dados e máquinas virtuais gratuitas com setup muito simples que podem ser utilizados como alternativa a um ambiente de desenvolvimento local. Para a modelagem das tabelas finais do banco de dados, recomenda-se que utilize um modelo dimensional (star-schema).

Para o desenvolvimento dos processos de ETLs também não há nenhuma restrição de tecnologia, no entanto se você possui familiaridade com Apache Airflow, dê prioridade ao uso, pois é a tecnologia que usamos no nosso dia a dia. Um simples script que seja executado por algum mecanismo de agendamento automático é uma solução válida para a construção dos processos de ETL. Apesar de um conjunto de queries ser um componente de visualização válido para esse desafio, imagine que quem quer ter as questões do desafio

respondidas seja um usuário de negócio sem conhecimentos de SQL ou qualquer outra linguagem de consulta de dados.

Informações sobre as APIs The Movie Database

Para utilizar as APIs disponibilizadas pela The Movie Database é necessário fazer um cadastro simples - <https://www.themoviedb.org/account/signup> - que será ativado instantaneamente após a validação do e-mail.

Após realizar o login com o cadastro ativo, é necessário criar uma chave de API - <https://www.themoviedb.org/settings/api/new?type=developer> - aceitar os termos de uso e preencher um formulário (não é necessário ter uma Application URL válida).

Para a obtenção dos dados necessários para responder as perguntas, recomendamos que integre sua solução nos seguintes endpoints da API:

- <https://developers.themoviedb.org/3/movies/get-now-playing> - onde são disponibilizados os dados de maneira resumida dos filmes atualmente em cartaz.
- <https://developers.themoviedb.org/3/movies/get-upcoming> - onde são disponibilizados os dados de maneira resumida dos filmes a serem lançados.
- <https://developers.themoviedb.org/3/movies/get-movie-details> - para obter as informações detalhadas dos filmes.
- <https://developers.themoviedb.org/3/movies/get-movie-credits> - para obter as pessoas relacionadas a um determinado filme. Na lista crew considere as pessoas com job "Director" como diretoras.
- <https://developers.themoviedb.org/3/people/get-person-movie-credits> - para obter a lista de filmes em que uma determinada pessoa esteve envolvida
- <https://developers.themoviedb.org/3/people/get-person-details> - para obter os detalhes de uma pessoa.

Para auxiliar no desenvolvimento a própria API disponibiliza links para Wrappers e Bibliotecas para acesso em diversas linguagens de programação, sinta-se a vontade para usar elas no desenvolvimento, elas podem ser consultas em <https://www.themoviedb.org/documentation/api/wrappers-libraries>

Sempre utilize o filtro de região com o valor "US" ao realizar requisições na API.

Modelagem

Crie uma modelagem conceitual dos dados.

Desafio Bônus (Opcional)

Construa um modelo de previsão de receita dos filmes que serão lançados em breve.