



## DESAFIO DE ARQUITETURA

Você como Pessoa Engenheira de Dados foi escolhida para liderar a criação de um produto de dados para o GloboPlay, a fim de ajudar a área de negócio a entender o comportamento dos usuários na plataforma. Uma das primeiras tarefas será a de desenhar uma arquitetura de alto nível para o produto, a qual deve suportar a captura, tratamento e a disponibilização de dados seguindo boas práticas de Data/Software Engineering e Data Mesh.

Importante ressaltar que o GloboPlay tem como premissa técnica ser resiliente e escalável para suportar grandes eventos como Copa do Mundo e BBB, por exemplo, então o mesmo deve ser aplicado ao produto de dados.

Então, considere os seguintes pontos para o desenho de arquitetura:

1. Os dispositivos com o Globoplay instalado (mobile, TV, computadores etc) enviarão eventos a cada ação do usuário na plataforma (assistir um vídeo, pausar o vídeo, favoritos etc). Pense em uma solução escalável que suporte muitas requisições por segundo.
2. A solução deve ser capaz de receber esses eventos e armazená-los em near real-time.
3. Uma arquitetura de dados, com camadas bem definidas e boas práticas, deverá ser pensada e desenhada.
4. Considere que picos de acesso acontecem em determinadas horas do dia e também em período de grandes eventos.
5. Uma abordagem para orquestração deverá ser considerada.
6. Uma Engine SQL que permita que Pessoas Cientistas de dados e Analistas de Dados façam as suas explorações.
7. Governança de dados:
  - a. Ter uma solução para catálogo
  - b. Seguir boas práticas de segurança para que seja possível armazenar dados PII e garantir que apenas pessoas autorizadas tenham acesso.
8. Monitoramento e possibilidade de testes de qualidade.
9. Por fim, um dashboard, desenvolvido por outro time de dados, consumirá o produto a cada 5min.



## DESAFIO DE CONSUMO DE DADOS

### Objetivo:

Avaliar as habilidades técnicas de um engenheiro de dados em relação a:

- ☐ Capacidade de consumir dados de uma API pública
- ☐ Capacidade de realizar análises exploratórias de dados
- ☐ Capacidade de gerar insights a partir de dados

### Requisitos:

- ☐ Conta no GitHub
- ☐ Conhecimentos básicos de Python
- ☐ Familiaridade com APIs públicas

### Instruções:

1. Crie um repositório no GitHub para o seu desafio.
2. Complete o desafio de acordo com as instruções.
3. Envie pull request para o seu repositório.
4. Envie o link do repositório para avaliação.

### Recursos:

- ☐ <https://swapi.dev/>
- ☐ <https://swapi.dev/documentation>

### Desafio:

O desafio consiste em consumir dados da API pública do Star Wars: <https://swapi.dev/>. Os dados fornecidos pela API incluem informações sobre personagens, planetas, naves espaciais, armas e muito mais.

Você deve realizar as seguintes tarefas:

- ☐ Consumir os dados da API e armazená-los em um banco de dados.
- ☐ Realizar análises exploratórias de dados para identificar padrões e tendências.
- ☐ Criar fluxos de dados.
- ☐ Gerar insights a partir dos dados, como:
  - Qual é o personagem que apareceu em mais filmes de Star Wars?
  - Quais são os planetas mais quente do universo de Star Wars?
  - Quais são as naves espacial mais rápida do universo de Star Wars?
  - Qual é a arma mais poderosa do universo de Star Wars?

### Avaliação:

A avaliação do desafio será baseada nos seguintes critérios:

- ☐ Correção técnica das soluções propostas.
- ☐ Clareza e organização do código.
- ☐ Organização no uso da ferramenta de versionamento.
- ☐ Qualidade dos insights gerados.