

## O que já foi feito:

## Ideias e Sugestões para o EmpatiA

- **1. Pipeline de Ingestão de Dados (Opções de Arquitetura):**
  - **O quê (Opção A - Centralizado):** Implementar um pipeline de dados que centralize as informações dos vários bancos e servidores dos clientes em um único Data Warehouse. As cargas podem ser semanais (em horários de pouco fluxo) usando uma ferramenta como o **Airbyte**, com um processo de ETL ou ELT.
  - **Por quê:** Para unificar todos os dados em uma única fonte da verdade, simplificando a análise global e a manutenção.
  - **O quê (Opção B - DW por Cliente):** Como alternativa, criar um Data Warehouse local para cada cliente. Cada DW seria independente, mas todos seguiriam um esquema conceitual em comum para padronização (similar à arquitetura usada no Redu).
  - **Por quê:** Para garantir um maior isolamento e segurança dos dados de cada cliente, potencialmente simplificando a lógica de ingestão individual.
- **2. Estrutura do Data Warehouse:**
  - **O quê:** A modelagem do DW deve ser estruturada com duas tabelas fato principais: uma para **interações** (fonte: **statuses**) e outra para **assignments** (fonte: **user\_assignments**). *Esta estrutura deve ser o padrão aplicado, seja no DW centralizado (Opção A) ou em cada DW individual (Opção B).*
  - **Por quê:** Para criar um modelo de dados eficiente e padronizado para monitorar o engajamento dos alunos, independentemente da arquitetura de ingestão escolhida.
- **3. Criação e Armazenamento do **cubo\_engajamento**:**
  - **O quê:** A partir das tabelas fato, criar o **cubo\_engajamento** (via view ou function). O destino final deste cubo, após ser classificado, seria uma estrutura à parte (ex: um schema dedicado) dentro do mesmo banco de dados do DW correspondente.
  - **Por quê:** Para centralizar a arquitetura, simplificar a infraestrutura, reduzir custos e garantir acesso performático aos dados de engajamento já processados.
- **4. Evolução com Inteligência Artificial (Modelos de Classificação):**
  - **O quê (Abordagem de Modelagem):** Para classificar os dados do cubo, treinar múltiplos modelos (um para cada nível hierárquico x agregação de tempo) ou desenvolver um *ensemble* de classificadores.
  - **Por quê:** Para garantir que a classificação do engajamento seja precisa e adaptada às diferentes granularidades de análise.

- **O quê (Opção - Não Supervisionado):** Implementar um algoritmo de clusterização (K-Means ou similar) para descobrir segmentos e perfis de usuários de forma automática.
- **Por quê:** Para revelar padrões de comportamento inesperados sem depender de rótulos pré-existentes.
- **5. Repositório de Referência**
  - Além disso, todas as informações do que foi produzido e possíveis mudanças, incluindo esse documento, encontram-se no seguinte repositório
  - [https://github.com/matheussilvaviitra/Classificador\\_engajamento.git](https://github.com/matheussilvaviitra/Classificador_engajamento.git)