# Desafio Técnico – Analytics Engineer Sênior

## 🎯 Pipeline de Dados com Airflow + Snowflake

Este desafio técnico simula uma situação real onde uma empresa precisa consolidar dados de vendas, CRM e marketing diariamente, utilizando uma arquitetura escalável e confiável.

Seu objetivo será construir uma DAG no Airflow para orquestrar esse pipeline e realizar cargas diárias no Snowflake.

## 🧪 Descrição do Desafio

Você trabalhará com dados simulados em formato CSV ou JSON representando três fontes distintas: CRM, faturas e campanhas de e-mail.

Sua responsabilidade é garantir a ingestão, transformação e carga desses dados em Snowflake, com validação de sucesso e alertas em caso de falha.

## 🔧 Entregáveis Técnicos

- DAG no Apache Airflow com as seguintes tarefas:

- Ingestão dos arquivos (pode ser leitura local ou bucket)

- Limpeza e validação dos dados (pode usar Pandas ou SQL)

- Carga final no Snowflake em tabelas normalizadas

- Implementar lógica de dependência entre as tarefas (sequencial e paralela onde fizer sentido)

- Simular alertas por e-mail/log em caso de falhas

- Tabelas criadas devem conter schemas padronizados (nomes, tipos e chaves primárias/estrangeiras)

## ✨ Extras que contam pontos

- Uso de task groups para organização

- Uso de macros ou templates no Airflow

- Validação de schema dos arquivos antes de carga

- Documentação das tarefas no código e no README

- Pipeline executável via Docker Compose

## ⚠ Importante

Você pode mockar os dados (ex: CSVs gerados manualmente ou com Faker).

Não é necessário realizar deploy real em ambiente de produção ou configurar autenticação robusta. Foque em:

- Clareza da DAG

- Estrutura de código e organização

- Boas práticas de engenharia de dados

## 📦 Entrega esperada

- Repositório no GitHub com:

- Código do Airflow

- Scripts de transformação se aplicável

- Arquivo docker-compose.yml (opcional)

- README.md com instruções de execução e estrutura de tabelas

- Prints ou link de dashboard (Airflow UI ou estrutura Snowflake)

## 🕒 Instruções Finais

Prazo para entrega: 3 dias corridos

Tempo estimado de execução: entre 6 a 8 horas

Nível de dificuldade: Sênior

Este teste visa avaliar sua habilidade de estruturar pipelines confiáveis, com boa organização de tarefas e visão de arquitetura de dados.

Capriche no README e aproveite a chance de mostrar sua senioridade! 🚀