

Desafio para a vaga de Data Engineer

Bemol Digital

Olá, parabéns por ter chegado a essa nova fase. Nós elaboramos um desafio bem bacana para conhecermos melhor a sua forma de trabalho.

Considere que você trabalha em uma empresa de varejo onde um grande volume de dados é gerado diariamente. Parte do seu time trabalha remoto. Na próxima semana eles iniciarão um projeto que necessita de alguns dados dos sistemas de vendas.

Devido a algumas mudanças na política de segurança da empresa, no que diz respeito a dados, agora só é permitido consumir dados via API. Uma outra restrição é que os times não tem acesso direto aos banco de dados dos sistemas de vendas. É necessário solicitar ao departamento de informática os arquivos em lotes do período de vendas desejado, eles podem vir em diferentes formatos (.xlsx, .txt, .csv, .html).

Dentro da rede da empresa você pode consumir esses dados em lotes e fazer suas análises, porém o time externo não. Logo, você estaria ferindo um política de segurança se repassasse esses dados via E-mail, Google Drive, FTP, etc.

Partindo do ponto em que você já passou pela processo de conseguir os dados e tem eles em mão, **sua missão é disponibilizar esses dados para seu time remoto/externo via API.**

Não temos restrição quanto a *stack*, mas gostamos do básico, algo como: Python, SQL, Node.js. Sinta-se livre para utilizar sua *stack* preferida.

Alguns diferenciais no seu projeto:

- Mostrar os artefatos gerados durante o processo de construção (códigos, arquitetura, programas utilizados)
- API documentada
- Projeto versionado usando Git (para API)
- Metadados da base de dados (pode ser um .docx, planilha online, etc)
- Disponibilizar essa aplicação em alguma nuvem: Azure, Google Cloud, AWS (eles possuem *trial*)
- Mostrar um exemplo de consumo da API, poderia ser no jupyter(lab/notebook)

Dica: Menos é mais, *keep it simple* :)