



Case Mr. Health

Tarefa teste

Processo seletivo DataLakers

DataLakers

Tarefa





A DataLakers foi contratada pela MR. HEALTH para ajudar na modernização de seu Data Warehouse. A empresa forneceu a DataLakers um breve resumo sobre seu negócio.

Leia o cenário abaixo para ter um entendimento do que a empresa faz, quem são pessoas chaves e do que se trata o projeto a ser executado. Ao final, encontrará mais instruções sobre os entregáveis do exercício para seu processo de recrutamento na DataLakers.

Cenário





Mr. Health: um sonho

A **MR HEALTH** é uma rede de “**Slow-Food**” presente na região sul e com aproximadamente 50 unidades.

Foi fundada **há 5 anos pelo chef João Silva**, e desde lá vem crescendo de forma exponencial. Com a pandemia, a rede teve que se adequar e expandir sua atuação on-line.





A **MR HEALTH** é o sonho de João Silva em poder oferecer **alimentação saudável e acessível a todos**.

Porém após ter passado o período da pandemia, vem refletindo sobre **como escalar seu negócio para outros estados**, já que necessita ter as **informações para tomada de decisão** de forma mais rápida em suas mãos e poder adotar **modelos estatísticos para gestão dos estoques de suas unidades**.

Ricardo Martins

Ricardo Martins é o **braço direito de João Silva**, responsável pela **área de operações**, o que o torna conhecedor de todos os **processos internos da empresa**.



Wilson Luiz

Wilson Luiz é o **Gerente de TI**, e o maior defensor da necessidade de **levar os dados para nuvem para estruturação de um Datalake/Data Warehouse**, onde possa centralizar as informações das diversas unidades e utilizar os recursos oferecidos em cloud para modernização do negócio.



Atualmente





Cada unidade envia as informações, toda a meia noite, sobre o fechamento das vendas diárias (D-1). Os arquivos são gerados em formato CSV e possuem a seguinte estrutura:

ARQUIVO PEDIDO.CSV

Id_Unidade
Id_Pedido
Tipo_Pedido (Loja Online ou Loja Física)
Data_Pedido
Vlr_Pedido
Endereco Entrega
Taxa_Entrega
Status (Pendente, Finalizado ou Cancelado)

ARQUIVO ITEM_PEDIDO.CSV

Id_Pedido
Id_Item_Pedido
Id_Produto
Qtd
Vlr_Item
Observacao



Atualmente, a equipe de Ricardo Martins tem a responsabilidade por **consolidar estes dados recebidos em planilhas**, para que possam gerar uma visão da venda realizada no dia anterior, o que acaba **envolvendo diversas pessoas** e além de gerar a possibilidade de **introduzir erros nos dados**.

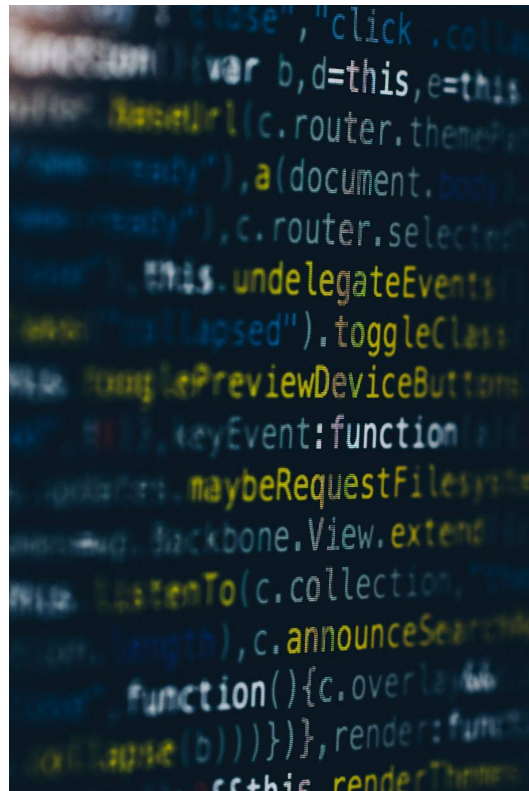
Futuramente





Desejos de João

João quer que os dados possam ser **consolidados de forma automática ao serem recebidos**, permitindo que Ricardo possa se concentrar na análise destas informações e na **tomada de ações** junto aos fornecedores e gestores das unidades. Ele ainda quer que o sistema tenha algum tipo de **inteligência que gere alertas ou recomendações** de forma automática.



Entregáveis





1. Mapeamento do Processo Atual:

- Utilize qualquer técnica desejada e **indique os pontos problemáticos no processo.**

2. Mapeamento do Processo Futuro

- Utilize qualquer técnica desejada e avalie se os pontos problemáticos foram **resolvidos ou minimizados**;
- Considere que os **dados serão levados para nuvem** (não se preocupe neste momento com o Cloud escolhido).

3. Proponha um Diagrama de Arquitetura para realizar a ingestão e consolidação dos dados.

- Pode ser considerado **qualquer um dos ambientes**: Google, Azure ou Amazon, bem como os serviços que melhor se adequarem.
- - Fique à vontade para utilizar o ambiente que mais estiver à vontade.
- Considere que além dos arquivos recebidos das unidades, a matriz da MR HEALTH possui um **banco de dados PostgreSQL** com as seguintes tabelas que também necessitam serem ingeridas:



TABELA PRODUTO

Id_Produto

Nome_Produto

TABELA UNIDADE

Id_Unidade

Nome_Unidade

Id_Estado

TABELA ESTADO

Id_Estado

Id_Pais

Nome_Estado

TABELA PAIS

Id_Pais

Nome_Pais



4. Descreva como os dados serão armazenados no datalake em **termos de camadas e qual a estrutura adotada para cada uma delas**. Considere que além dos arquivos recebidos, os dados oriundos do banco de dados da matriz da MR HEALTH.
5. Elabore uma relação de atividades que serão necessárias para **implementação do processo de ingestão e consolidação de dados**. Estas informações serão necessárias para realizar o entendimento das atividades com os demais membros da equipe e poder **elaborar o “Product Backlog” do projeto**.

Observações:

- Não esqueça de colocar uma breve descrição sobre a atividade.
- As atividades devem estar organizadas na ordem de sua execução.



DataLakers

The Big Data Company



Boa Sorte!