



DESAFIO TÉCNICO ENGENHARIA DE DADOS

A Panvel é uma rede de farmácias com mais de 50 anos de história, e que atualmente conta com mais de 600 filiais espalhadas pelos estados do Rio Grande do Sul, Santa Catarina, Paraná e São Paulo. Além disso, a empresa investe fortemente em sua plataforma digital, oferecendo aos clientes a conveniência de comprar produtos online com a opção de entrega em domicílio ou retirada em loja. Nossa missão é proporcionar saúde e bem-estar para as pessoas, e nossa visão é ser a melhor em produtos e serviços de saúde e bem-estar de forma sustentável e inovadora!

O time de dados da Panvel recebeu o desafio de estudar e entender profundamente o comportamento de compra dos seus clientes, com objetivo de propiciar um atendimento cada vez mais personalizado, e oferecer benefícios para eles conforme estreitam seu relacionamento com a Panvel.

A primeira etapa deste desafio começa com você, nosso(a) Engenheiro(a) de Dados, com objetivo de fornecer os dados dos clientes e das suas respectivas compras, para que o time de analistas e cientistas de dados possam trabalhar. Considere que a equipe é nova, e ainda não tem ferramentas, padrões e arquitetura definida, mas tem o desejo de contar com um Datalake. Agora cabe a você, não só entregar os dados transformados para o time, mas também projetar uma arquitetura adequada para este projeto e projetos futuros.

Ao final do desafio esperamos ver:

- **Proposta de Arquitetura:** Proponha uma arquitetura, e aborde a eficiência no processamento, escalabilidade e a fluidez dos dados entre diferentes camadas do pipeline.
- **Pipeline de Dados:** Desenvolva um pipeline de dados para extrair, transformar e carregar os dados solicitados.
- **Garantia da Qualidade dos Dados:** Apresente formas de garantir a qualidade dos dados, incluindo validações, limpezas e tratamentos necessários.
- **Apresentação:** Destaque suas propostas e implementação e justifique as suas escolhas com relação a arquitetura, ferramentas e técnicas escolhidas.

A seguir você encontra os dados mapeados pelo time de analistas de dados e um dicionário de dados com as informações sobre as origens dos dados.

DADOS ESPERADOS

Vendas:

- **codigo_filial**
- **codigo_cupom_venda**
- **data_emissao**
- **codigo_item**
- **valor_unitario**
- **quantidade**
- **codigo_cliente**
- **tipo_desconto**
(Convênio/Promoção)
- **canal_venda** (Loja/Site/App)

Clientes:

- **codigo_cliente**
- **data_nascimento**
- **idade**
- **sexo**
- **uf**
- **cidade**
- **estado_civil**
- **flag_lgpd_call**
- **flag_lgpd_sms**
- **flag_lgpd_email**
- **flag_lgpd_push**

DICIONÁRIO DE DADOS

Colunas da tabela **vendas**:

- **d_dt_vd**: Data em que a venda foi efetuada.
- **n_id_fil**: Código único da filial que efetuou a venda. (PK)
- **n_id_vd_fil**: Código identificador único da venda na filial. (PK)
- **v_cli_cod**: Código identificador do cliente da venda.
- **n_vlr_tot_vd**: Valor total da venda já com descontos.
- **n_vlr_tot_desc**: Valor total de descontos.
- **v_cpn_eml**: Se o cupom fiscal foi enviado por e-mail.
- **tp_pgt**: Tipo de pagamento da venda.

Colunas da tabela **pedidos**:

- **n_id_pdd**: Identificador único do pedido. (PK)
- **d_dt_eft_pdd**: Data efetiva do pedido.
- **d_dt_entr_pdd**: Data de entrega do pedido.
- **v_cnl_orig_pdd**: Canal de origem do pedido (Loja/Site/App).
- **v_uf_entr_pdd**: UF de entrega do pedido.
- **v_lc_ent_pdd**: Localidade de entrega do pedido.
- **n_vlr_tot_pdd**: Valor total do pedido.

Colunas da tabela **itens_vendas**:

- **n_id_fil**: Código único da filial que efetuou a venda. (PK)
- **n_id_vd_fil**: Código identificador único da venda na filial. (PK)
- **n_id_it**: Código identificador do item na venda. (PK)
- **v_rc_elt**: Se a venda do item foi com receita eletrônica.
- **v_it_vd_conv**: Se no item da venda ocorreu desconto de convênio.
- **n_vlr_pis**: Valor do imposto PIS.
- **n_vlr_vd**: Valor final do item na venda já com desconto.
- **n_vlr_desc**: Valor do desconto.
- **n_qtd**: Quantidade do item.

Colunas da tabela **pedido_venda**:

- **n_id_fil**: Código único da filial que efetuou a venda. (PK)
- **n_id_vd_fil**: Código identificador único da venda na filial. (PK)
- **n_id_pdd**: Número do pedido que originou a venda. (PK)

Colunas da tabela **clientes**:

- **v_id_cli**: Código identificador do cliente. (PK)
- **d_dt_nasc**: Data de nascimento do cliente.
- **v_sx_cli**: Gênero do cliente.
- **n_est_cvl**: Estado civil do cliente (1 Solteiro/2 Casado/3 Viúvo/4 Desquitado /5 Divorciado/6 Outros).

Colunas da tabela **clientes_opt**:

- **v_id_cli**: Código identificador do cliente. (PK)
- **b_push**: Se o cliente autoriza o recebimento de notificações via push.
- **b_sms**: Se o cliente autoriza o recebimento de notificações via SMS.
- **b_email**: Se o cliente autoriza o recebimento de notificações via e-mail.
- **b_call**: Se o cliente autoriza o recebimento de notificações via ligação.





Colunas da tabela **enderecos_clientes**:

- **v_id_cli**: Código identificador do cliente. (PK)
- **n_sq_end**: **Sequência de cadastro do endereço do cliente. (PK)**
- **d_dt_exc**: Data de exclusão do endereço.
- **v_lcl**: Localidade do endereço.
- **v_uf**: UF do endereço.

Complemento:

- **Dados públicos de gêneros por nome:**
<https://data.brasil.io/dataset/genero-nomes/nomes.csv.gz>

Boa sorte!