

Desafio 1

Sumário executivo

A Clin Corp é uma empresa de transporte brasileira que atua em uma localização estratégica, combinando facilidades "state-of-the-art" e processos eficazes através de profissionais treinados para apresentar um serviço de alta qualidade.

O setor de Controladoria (*Controller*) da Clin Corp é responsável por regularmente, coletar, analisar e disseminar vários indicadores operacionais, de estoque e informações gerenciais.

Atualmente, o esforço para coleta, consolidação e exibição dos dados provenientes do ERP da companhia, o SAP, torna-se muito custoso, visto que, as informações são transportadas para arquivos do Excel e consolidadas utilizando um banco do Access. Por conta dos arquivos em Excel permitirem sua livre manipulação, há preocupação por parte da Equipe de TI com informações que podem sofrer distorção ao longo da execução do processo manual, por não haver um local único para salvaguarda dos dados de forma gerenciada e que atenda com robustez os destinatários dos relatórios disponibilizados pela Controladoria.

Objetivo

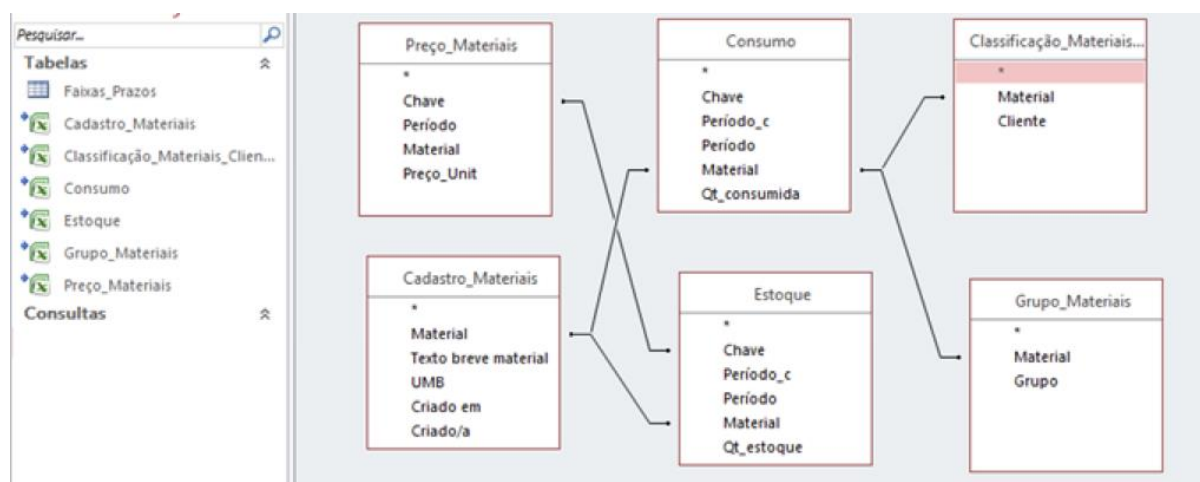
Visando melhorar o processo de preparação dos dados e confecção dos painéis que contém as informações do estoque, bem como sua posição ao longo de um período, conforme modelo enviado pelo *controller* da Clin Corp, gostaríamos das suas considerações – como consultor designado para desenvolver este projeto – sobre:

- Baseado no sumário executivo, há alguma oportunidade de negócio?
- Qual (quais) tecnologia(s) deverá (deverão) ser implementada(s)? Comente o que motivou a proposição.
- Qual (quais) ferramenta(s) deverá (deverão) ser utilizada(s)? Comente onde cada uma será aplicada.
- Quais etapas serão executadas para que seja entregue o relatório, conforme modelo encontrado no anexo 3, de forma automatizada?

Não esqueça das justificativas e esboços dos modelos propostos. A riqueza de detalhes é um fator-chave para o sucesso do projeto.

Anexos

1 - Modelo de dados no Access



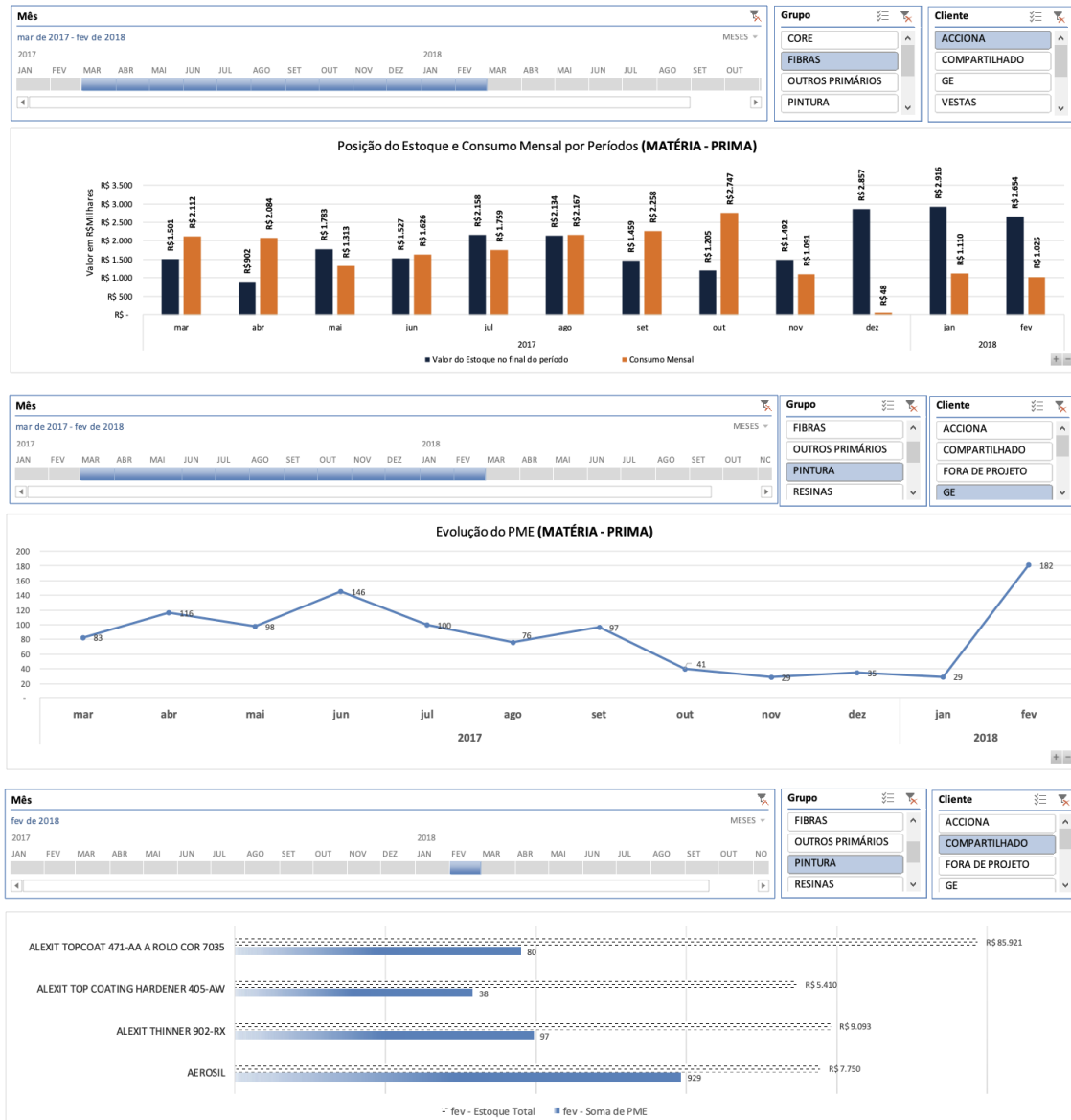
2 – Fórmulas

ValorTotalConsumo = Qt_consumida * Preço_Unit

ValorTotalEstoque = Qt_estoque * Preço_Unit

PME = $\sum \text{ValorTotalConsumo} * 30 / \sum \text{VlrTotalEstoque}$

3 - Modelo de relatório



Desafio 2

Desafio Técnico: Transformação de Dados para Data Lake

Contexto:

Você faz parte de uma equipe de ciência de dados em uma empresa que coleta dados de vendas de uma cadeia de lojas. A equipe decidiu migrar esses dados para um Data Lake para facilitar análises futuras.

Tarefas:

- Faça as seguintes consultas em SQL antes de transformas os dados:
 - o Pergunta: Quantos pedidos foram feitos por cada cliente?
 - o Pergunta: Qual é o produto mais vendido em termos de quantidade?
 - o Pergunta: Qual é o cliente que gastou mais dinheiro em compras?
 - o Pergunta: Quais são os produtos que foram vendidos em todos os pedidos?
 - o Pergunta: Qual é o dia da semana com o maior número de pedidos?
- Crie um script SQL que transforme os dados brutos do banco de dados de vendas em um formato mais performático e adequado para armazenamento em um Data Lake. O banco de dados de vendas agora possui as seguintes tabelas:

Esquema dos dados

- **orders:** Contém informações sobre cada pedido, como ID do pedido, data do pedido e ID do cliente.
 - o Colunas: order_id, customer_id, order_date
- **products:** Contém informações sobre os produtos vendidos, como ID do produto, nome do produto, tipo de produto e preço.
 - o Colunas: product_id, product_name, product_type, price
- **order_items:** Contém informações sobre os itens de cada pedido, como ID do item, ID do pedido associado, ID do produto e quantidade.
 - o Colunas: item_id, order_id, product_id, quantity

Requisitos:

- o **Extração e Transformação de Dados:** Escreva consultas SQL para extrair os dados das tabelas **orders**, **products** e **order_items** e transformá-los em um formato adequado para análises posteriores. Considere as seguintes transformações:
 - Agregação dos dados para obter informações relevantes de cada pedido, como o valor total do pedido, o número de itens, etc.
 - Manipulação dos dados conforme necessário para atender aos requisitos de análise da equipe de ciência de dados.
- o **Performance:** Certifique-se de que sua consulta SQL seja eficiente em termos de desempenho, levando em consideração o volume de dados que será processado.

Observação:

- Documente sua abordagem, explicando as etapas da transformação e qualquer consideração especial feita em relação ao desempenho.