

Caio Martim Barros
Case Técnico Estágio Engenharia de Dados
EloGroup

1. Apresente as características do modelo de dados fornecido, destacando sua visão geral acerca do modelo de dados e tecendo críticas e comentários.;

Esse modelo de dados proposto é classificado como um star-schema, que é um modelo multidimensional. Dele, conseguimos analisar que é composto por dois tipos de tabelas, uma principal chamada de 'tabela fato' dado pelo nome no modelo de dados: 'FT_vendas' e o outro tipo de tabela chamada de 'tabela dimensão', que são as tabelas restantes desse modelo.

O esquema apresentado possui uma tabela fato FT_Vendas que contém as chaves estrangeiras que estarão relacionadas com as tabelas dimensões, que possuem as chaves primárias para fazer essa relação. Portanto, a tabela FT.Vendas armazena as medidas que serão analisadas de acordo com as dimensões que queremos realizar, interagindo, assim, a tabela fato com o restante das tabelas com simples metodologias como um inner join para esse acesso.

Nesse contexto, no ponto de vista geral, o star-schema é um modelo importante para a análise de dados e fácil de ser manipulado e lido, esse esquema possui características importantes para a realização de análises com propósito de visualização por períodos, se tratando com dados não normalizados e está comumente relacionado com a metodologia dos sistemas OLAP. Além disso, esse modelo de dados garante uma boa performance pelo fato das queries serem simples para serem carregadas e são usadas em serviços de análise de dados, como PowerBI e podem ser usados também na construção de Data Warehouses e Data Mart.

Todavia, o star-schema atua no formato 'one-to-many', a qual a tabela fato se relaciona com o restante das tabelas dimensões, então para uma arquitetura do formato 'many-to-many', ela não é recomendada. Nesse âmbito, outra crítica seria no conceito desse esquema usar dados não normalizados, o que indica em uma limitação da análise dos dados, não permitindo visualizações mais complexas, além de que se tratando de sistema OLAP, não se pode adicionar ou atualizar os dados pois pode comprometer a análise, dessa forma, o star-schema é usado mais na metodologia 'read-only'.

2. Demonstre seu conhecimento em SQL, inserindo código abaixo de cada solicitação:

(2.1) Quantidade vendida no período entre 01/04/2020 a 01/04/2021;

```
1  SELECT
2  SUM(Qtde_vendida) AS Quantidade_Vendida
3  FROM FT_Vendas
4  INNER JOIN DIM_Tempo
5      ON DIM_Tempo.ID_tempo = FT_Vendas.ID_tempo
6      WHERE Dia = 1 AND Mes = 4 AND (Ano = 2020 OR Ano = 2021);
```

(2.2) Quantidade de clientes únicos do sexo feminino e estado civil divorciada;

```
1  SELECT
2  COUNT(*) AS Quantidade_Clientes
3  FROM DIM_Cliente
4  WHERE Estado_civil = "Divorciada" AND Sexo = "Feminino";
```

(2.3) Buscar os produtos com código 55, 120, 142 e que tiveram receita maior que R\$ 120,00;

```
1 SELECT
2   Cod_produto ,
3   Receita_venda
4 FROM FT_Vendas
5 INNER JOIN DIM_Produto
6     ON DIM_Produto.ID_produto = FT_Vendas.ID_produto
7     WHERE (DIM_Produto.Cod_produto = 55 OR DIM_Produto.Cod_produto = 120
8           OR DIM_Produto.Cod_produto = 142) AND Receita_venda > 120;
```

(2.4) Verificar a receita da loja com o código 32 no período igual ou superior a 01/04/2021;

```
1 SELECT
2   SUM(Receita_venda) AS Receita_total
3 FROM FT_Vendas
4 INNER JOIN DIM_Loja
5     ON DIM_Loja.ID_loja = FT_Vendas.ID_loja
6     WHERE DIM_Loja.Cod_loja = 32 AND Dia >= 1 AND Mes >= 4 AND Ano >= 2021
7 GROUP BY DIM_Loja.Cod_loja;
```

3. Por fim, descreva o que você melhoraria nesse modelo;

Primeiramente, temos uma tabela fato: FT_Vendas, e 4 tabelas dimensão: Produto, Cliente, Tempo e Loja. Na tabela DIM_Tempo, por exemplo, eu aconselharia a colocar novos atributos para indicar o período dessa compra, se foi feita a realização da compra de manhã, tarde, noite ou madrugada, pois poderíamos utilizar essa informação aliada a uma maior quantidade de anúncios, promoções, etc nesse período a fim de os clientes comprarem mais artigos juntos. Da mesma forma, poderia atribuir também uma *flag* para indicar se esse produto está sendo vendido em período de feriado ou em finais de semana, com o mesmo intuito indicado anteriormente.

Na tabela DIM_Cliente, poderíamos colocar um atributo para indicar o tipo de pagamento que está sendo realizado na hora de comprar o produto. Por exemplo, se para o e-commerce o pagamento via PIX é muito bom no quesito de lucro, então poderíamos utilizar desse novo atributo para saber qual são os tipos mais usados na hora do pagamento e assim trabalhar para tentar aumentar a utilização do pagamento via PIX. Dessa forma, poderíamos fazer um desconto de 5%, por exemplo, para clientes que efetuarem pagamentos via PIX e assim ao longo do tempo iríamos verificar se essa curva com relação ao pagamento por PIX está valendo a pena ou não.

Nesse âmbito, a fim de analisar melhor os produtos que estão sendo vendidos, recomendaria a colocação de mais informações na tabela dimensão DIM_Produto, inserindo a marca, cor, tamanho, tipo do produto, para coletarmos mais informações do que está sendo mais vendido especificamente nesse e-commerce e assim podermos agir a respeito.

Por fim, poderíamos também criar uma nova tabela dimensão chamada DIM_Promoção ou fazer uma indicação se há promoção do produto com uso de uma flag para poder analisar se os clientes estão mais inclinados a comprar os produtos se estes estiverem em promoção e também indicar o valor, em porcentagem, dessa promoção realizada no produto. Aliado a isso, poderíamos também criar um atributo para a visualização dos cupons de desconto, onde normalmente esse está realizado com marketing fora do site, então é importante também fazer uma análise mais precisa para perceber se está compensando.