

Imersão Power BI mão na massa

Aula 1 - Principais erros cometidos no Power BI e como evitar ou corrigir

- Importar todo o histórico de dados para o Power BI

Para desenvolver o projeto não é necessário ter todo o histórico dos dados. Com bases muito grandes (ordem de milhões de linhas) se todo o histórico for importado para o Power BI Desktop pode haver problemas de performance no desenvolvimento do projeto.

Como evitar ou corrigir:

Criar um parâmetro para a data inicial no Power Query e aplicar um filtro na coluna de data, filtrando a partir da data informada nesse parâmetro. Posteriormente quando o relatório for publicado para o Power BI Service, é possível lá alterar a data do parâmetro para trazer todo o histórico dos dados.

- Trazer todas as colunas da fonte

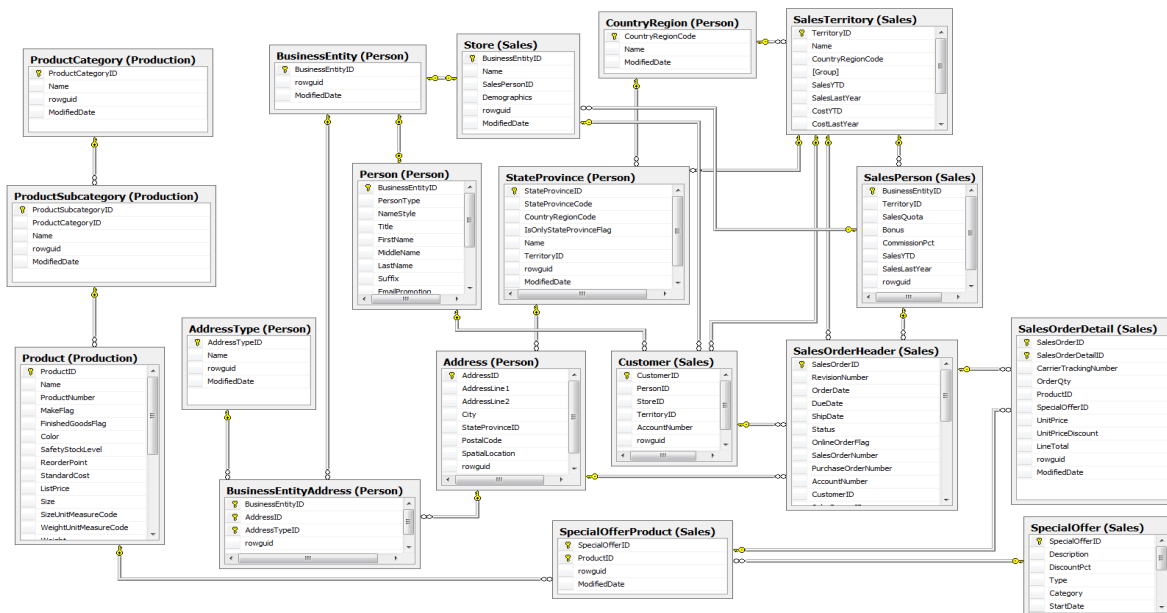
Valores distintos em colunas prejudicam a performance. Então as colunas que não vão ser usadas, não devem ser trazidas. Se trazidas, devem ser eliminadas no Power Query.

Como evitar ou corrigir:

Em caso de conexão no banco, já traga somente as colunas necessárias (caso posteriormente seja verificado que serão necessárias outras colunas, é possível alterar e trazer outras colunas). Em caso de conexão a planilhas ou Csv, assim que as tabelas forem importadas, só mantenha na consulta as colunas que serão utilizadas.

- Trazar as tabelas da fonte e sair relacionando por campos em comum

Os relacionamentos de um banco de um sistema transacional são como abaixo:

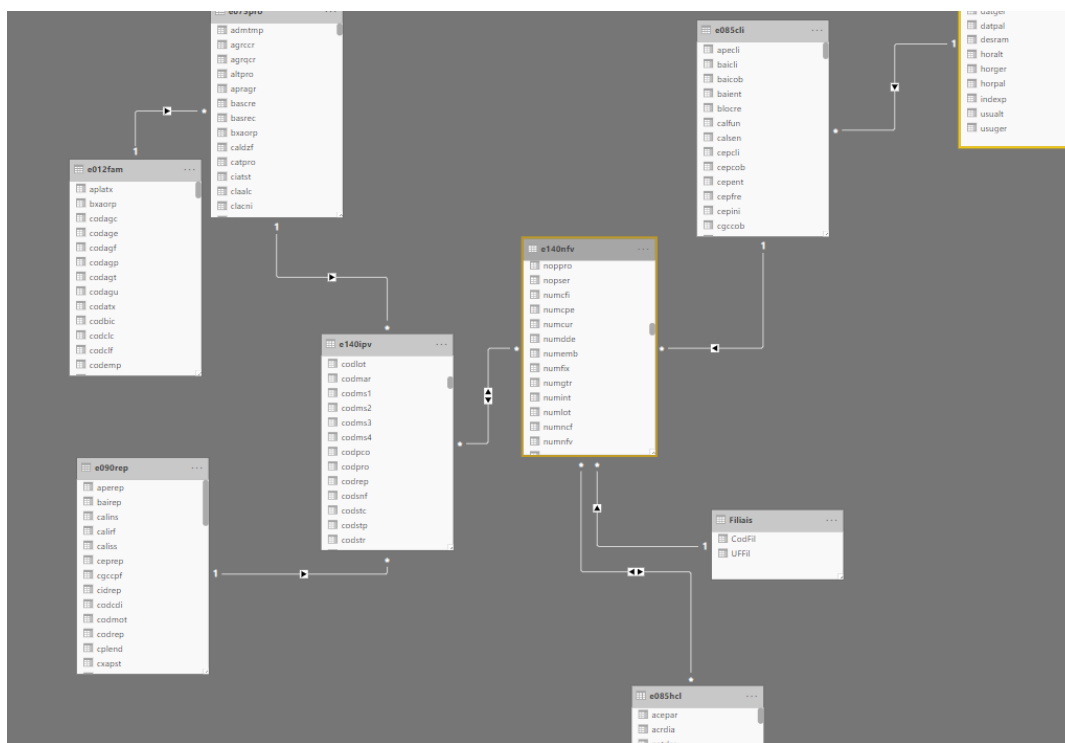


É um modelo **altamente normalizado**, ou seja, quebrado em diversas tabelas.

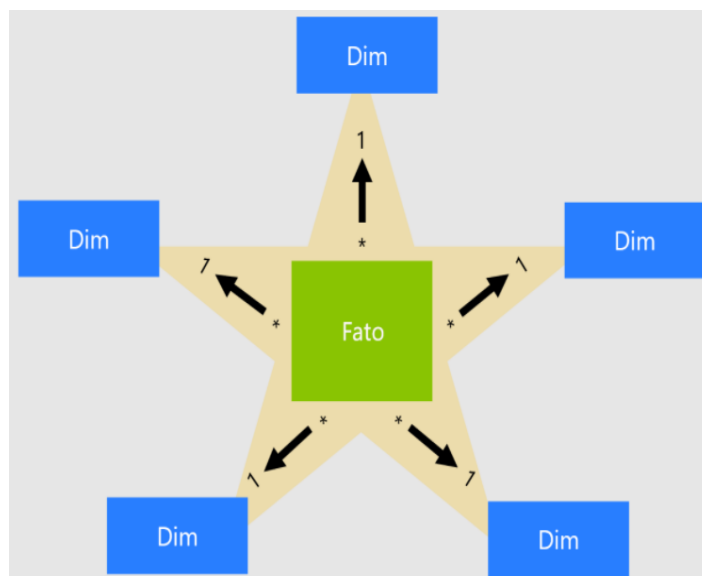
OLTP é focado na velocidade de escrita e não de leitura.

OLAP é foca na leitura.

Errado:



Esquema estrela:



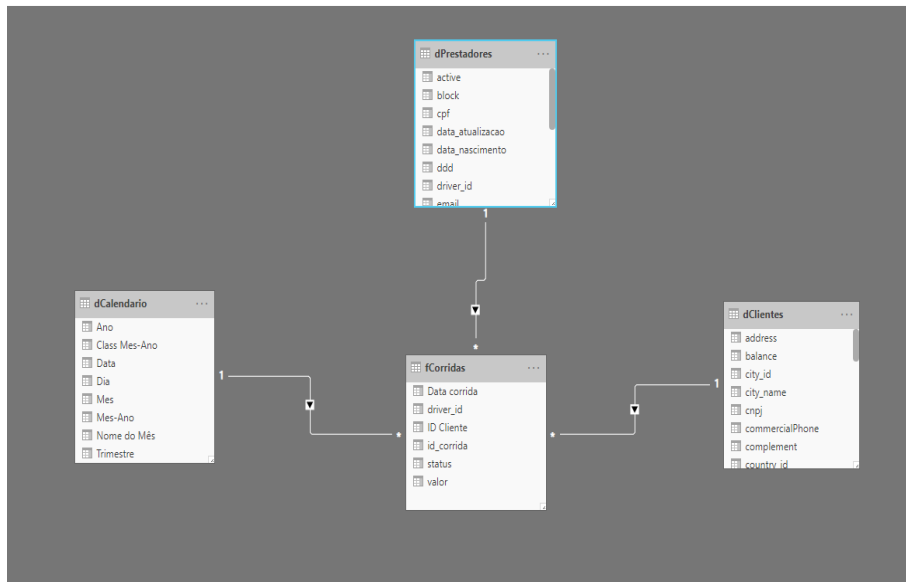


Tabela fato:

- Corresponde a um acontecimento específico e deve conter só dados relacionados a esse acontecimento (exemplo: vendas, acidentes de trabalho, emissão de notas fiscais, contratações, lucro...).
- Contêm valores numéricos e que podem ser agregados.
- Cada linha representa um evento (exemplo: uma venda, um recebimento, um pagamento, um acidente, uma infecção por COVID-19, uma morte...)
- Geralmente essa tabela possui muitas linhas (eventos) e poucas colunas.
- De um modo geral, não devem ser armazenadas informações textuais na tabela fato. A menos que o texto seja único para cada linha, ele pertence a uma dimensão.
- Essa tabela não deve conter, por exemplo, o endereço do cliente (esse dado estará contido na tabela dimensão de clientes).

Tabela dimensão

- Contêm informações que descrevem os fatos
- Possuem valores únicos em cada linha.
- Possuem menos linhas que a tabela fato e mais colunas.
- Descrevem o porquê, quem, onde, quando, o quê e como os acontecimentos da tabela fato foram registrados.

Como evitar ou corrigir (3 formas, em ordem de performance):

1 - Criar views no banco para as tabelas fatos e dimensões e importar essas views já prontas para o Power BI. No Power Query praticamente não vai ser necessário fazer tratamento. Fazer os relacionamentos no Power BI.

2 - Importar os fatos e dimensões por meio de script SQL. Nesse caso não há query folding, então não deve ser feito nenhum tratamento no Power Query (ou fazer o mínimo). Fazer os relacionamentos no Power BI.

3 - Importar as tabelas originais e criar os fatos e dimensões dentro do Power Query por meio de mesclagens. Fazer os relacionamentos no Power BI. Essa é a situação de pior performance.

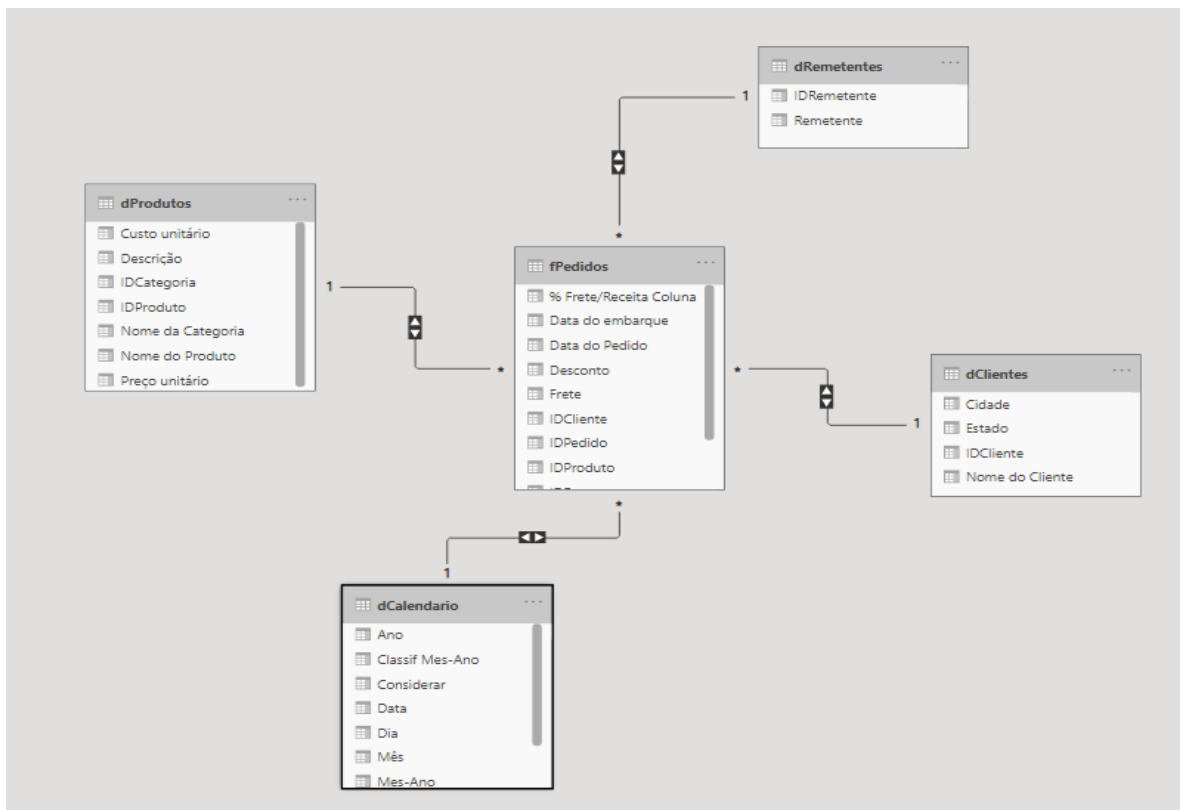
- Fazer relacionamentos com sentido duplo

As setas indicam as direções dos filtros nos relacionamentos. O filtro deve se propagar da dimensão para a fato (vendas por produto, por cliente, por vendedor...).

Sentido duplo gera ambiguidade no modelo, pois uma tabela está filtrando outra e sendo filtrada ao mesmo tempo e ao criar medidas você poderá obter resultados indesejados e não conseguir entender quem está filtrando quem no modelo. Ou seja, o comportamento do modelo fica complexo e imprevisível.

Situações que fazem o desenvolvedor deixar os filtros duplos:

- a) Para sincronizar filtros
- b) Para uma dimensão filtrar outra dimensão



Como evitar ou corrigir:

De um modo geral, utilize apenas relacionamentos com sentido único no modelo, onde a dimensão filtra a fato.

Para sincronizar filtros, arraste a medida para o painel de filtros do filtro para a parte "Filtros nesse visual" e selecione "Não está em branco"

Para alterar o sentido do relacionamento internamente na medida utilize a função CROSSFILTER.

- Usar colunas calculadas e medida implícita

Medida implícita é simplesmente arrastar uma coluna da tabela para um visual no relatório. Você pode escolher o tipo de agregação que será feita.

Medida explícita é criar a medida usando Dax e definir explicitamente o que está sendo feito por trás da medida (agregações, filtros..).

Medida implícita pode resultar em valores errados, principalmente quando se agrega percentuais. Também dificulta a posterior manutenção do relatório, pois você não sabe o que cada medida implícita está fazendo por trás sem ter que verificar uma por uma.

A medida explícita é a forma correta de criar medidas, pois você sabe exatamente o que a medida está fazendo. Facilita a reusabilidade da medida e a manutenção, pois se a medida for alterada vai valer para todo o relatório.

Coluna calculada: usada nos eixos dos visuais, no segmentador (filtro) ou para classificar uma coluna pela outra.

Medida: usada no campo de valor dos visuais.

Como evitar ou corrigir:

Para efetuar cálculos, dê preferência a medidas explícitas.

- Encher a página com muitos visuais e informação em excesso

As vezes o desenvolvedor prefere pecar por excesso do que por falta. Muito cuidado com isso.

Como evitar ou corrigir:

Utilize até 5 (no máximo 6) visuais em uma página. Só mostre as informações que são essencialmente necessárias.

- Não usar a regra do Z

Muitas vezes o desenvolvedor distribui os cartões e visuais de forma aleatória na página.

Como evitar ou corrigir:

A regra do Z consiste no seguinte: Estudo realizado concluiu que os consumidores de painéis e relatórios leem a página da esquerda para direita e da parte superior para a inferior. Então, a informação mais importante da página de estar do lado esquerdo superior, seguido pelo lado direito superior, lado esquerdo inferior e lado direito inferior.