

MODELAGEM DE DADOS

Modelagem de dados é o ato de explorar estruturas orientadas a dados. Como outros artefatos de modelagem, modelos de dados podem ser usados para uma variedade de propósitos, desde modelos conceituais de alto nível até modelos físicos de dados.

As aplicações que giram em torno do banco de dados relacionais possuem conceitos básicos sobre a modelagem de dados. Por isso, seja muito simples ou mais complexo, a correta modelagem é necessária para tornar-se uma aplicação mais robusta e eficiente.

Os modelos de dados são ferramentas que permitem demonstrar como serão construídas as estruturas de dados que darão suporte aos processos de negócios, como os dados estarão organizados e quais os relacionamentos que pretendemos estabelecer entre eles.

Esses modelos se dividem em 3 perspectivas:

Conceitual: como representação de alto nível e com foco no ponto de vista do usuário criador dos dados. É sempre o primeiro modelo a ser desenvolvido e muito fácil de ser compreendido, pois não há limitações ou tecnologias específicas.

Lógica: agrega detalhes de implementação e leva em conta as regras e algumas limitações de recursos padronizados. Aqui é possível definir atributos que serão as chaves para a estrutura.

Física: demonstra os dados fisicamente e leva em consideração todas as regras e limitações do banco de dados. Obedece a padrões e validações e a partir daqui que o modelo deve estar espelhado para o seu banco de dados final.

O modelo é uma representação conceitual dos dados, dos relacionamentos entre os dados e das regras. A própria modelagem pode conter diagramas, símbolos ou texto para representar os dados e o modo como eles se associam.

Por conta da estrutura que a modelagem de dados impõe aos dados, esse processo subsequentemente aumenta a consistência da nomenclatura, regras, semântica e segurança ao mesmo tempo em que melhora a análise de dados.

Esse processo crítico consiste em criar estruturas no sistema de armazenamento escolhido (físico ou cloud) para possibilitar a associação e o resgate de informações em um determinado padrão e momento.

A modelagem é requisito essencial no desenvolvimento de softwares, pois permite a integração de bancos de dados por sistemas independentes. Nesse contexto, vale ressaltar que um erro durante a modelagem pode comprometer toda a usabilidade do sistema e gerar retrabalho de programação e reformulação de todo o banco, o que sempre aumenta o custo do projeto.

COMO FUNCIONA A MODELAGEM DE DADOS

A modelagem de dados contempla todas as atividades relacionadas ao banco de dados para representar como os dados serão organizados e armazenados. Entre outros exemplos, a modelagem de dados:

- Identifica as entidades e atributos dos dados.
- Aplica uma convenção de nomes para padronizar ou hierarquizar os dados.
- Identifica relacionamentos e aspectos comuns entre os dados.
- Associa chaves para os dados serem requisitados com mais facilidade.
- Normaliza o banco para reduzir a redundância dos dados.
- Diversifica os dados para melhorar o desempenho.

Os modelos de dados são formados de entidades, que são os objetos ou conceitos sobre os quais se rastreiam os dados, e eles se tornam as tabelas em um banco de dados. Produtos, vendedores e clientes são exemplos de possíveis entidades em um modelo.

As entidades têm atributos, que são particularidades que queremos rastrear — você pode pensar nos atributos como as colunas de uma tabela. Se tivermos uma entidade de uma mercadoria, por exemplo, o nome dela pode ser um atributo.

As entidades não existem sozinhas. Elas estão ligadas umas às outras. As conexões entre entidades em um modelo de dados são chamadas de relacionamentos que refletem as regras de negócios. Os relacionamentos podem ser um para um, um para muitos ou muitos para muitos, como veremos mais tarde no Power BI.

MODELAGEM TRANSACIONAL E DIMENSIONAL

A modelagem transacional, também conhecida como OLTP (Online Transaction Processing), é fundamental para o suporte das operações diárias de uma empresa. Ela se concentra na captura e manipulação eficiente dos dados provenientes das interações em tempo real entre um sistema informatizado e suas fontes de dados. Esses dados transacionais são essenciais para registrar e rastrear atividades como vendas, transações financeiras, processos de compra, entre outros.

Um dos principais princípios da modelagem transacional é a normalização dos atributos, que visa reduzir a redundância e garantir a integridade dos dados durante as transações. Isso significa que os dados são estruturados de forma a minimizar a duplicação de informações, facilitando operações como inserção, atualização e exclusão de registros.

Em contraste, a modelagem informacional, também conhecida como OLAP (Online Analytical Processing), é voltada para a análise e tomada de decisões estratégicas com base em grandes volumes de dados históricos. Enquanto a modelagem transacional se concentra em dados operacionais e transacionais em tempo real, a modelagem informacional visa agregar e consolidar dados para fornecer uma visão mais abrangente e contextualizada do negócio.

Um dos principais objetivos da modelagem informacional é a criação de estruturas dimensionais que permitem a análise multidimensional dos dados. Isso envolve a organização dos dados em torno de

dimensões-chave, como tempo, localização, produto e cliente, permitindo uma exploração mais eficaz das relações e tendências dentro dos dados.

Embora a modelagem transacional e informacional tenha objetivos distintos, elas frequentemente trabalham em conjunto para fornecer suporte abrangente às necessidades de informação de uma organização. Os dados transacionais capturados pelo sistema OLTP alimentam os sistemas OLAP, onde são agregados, consolidados e analisados para fornecer insights valiosos para a tomada de decisões estratégicas.

De forma resumida, a modelagem transacional projetada para lidar com transações em tempo real, focando na captura e manipulação eficiente de dados provenientes de interações de um sistema informatizado com suas fontes de dados. Prioriza a normalização dos atributos e é essencial para as operações diárias de uma empresa (OLTP - Online Transaction Processing).

Enquanto a modelagem dimensional tem como objetivo a análise e tomada de decisões estratégicas com base em grandes volumes de dados históricos. Organiza os dados em torno de dimensões-chave, permitindo uma análise multidimensional dos dados para fornecer insights valiosos para tomada de decisão (OLAP - Online Analytical Processing).

RELACIONAMENTOS

No Power BI, os relacionamentos desempenham um papel fundamental na integração de dados de várias fontes e na realização de análises avançadas. Esses relacionamentos referem-se à conexão entre diferentes conjuntos de dados com base em campos comuns, permitindo que seja possível combinar informações de maneira eficaz para criar visualizações e análises significativas.

Ao trabalhar com conjuntos de dados que possuem campos compartilhados, os usuários podem estabelecer relacionamentos entre as tabelas com base nesses campos. Por exemplo, ao ter uma tabela de vendas com uma coluna de ID do cliente e outra tabela com detalhes dos clientes também contendo uma coluna de ID do cliente, é possível estabelecer um relacionamento entre essas duas tabelas. Isso permite que haja a combinação facilmente das informações das duas tabelas nas análises e visualizações.

Quando os relacionamentos são criados no Power BI, a ferramenta utiliza essas conexões para calcular automaticamente os resultados corretos ao fazer consultas e criar visualizações. Isso significa que é possível explorar e analisar os dados de forma eficiente e precisa, sem a necessidade de manipulação manual extensa dos dados.

Ao utilizar relacionamentos no Power BI, é possível otimizar a análise de dados, economizando tempo e esforço. A capacidade de integrar facilmente dados de várias fontes e combiná-los de maneira inteligente permite insights mais profundos e decisões mais informadas.

Além disso, os relacionamentos no Power BI oferecem flexibilidade e escalabilidade. Os usuários podem criar relacionamentos simples entre duas tabelas ou estabelecer conexões mais complexas envolvendo várias tabelas. Isso permite lidar com conjuntos de dados de diferentes tamanhos e complexidades, garantindo que o Power BI seja uma ferramenta poderosa para análise de dados em uma variedade de cenários.

No Power BI, atualmente, não é possível criar relacionamentos de auto-relacionamento (onde uma tabela se relaciona com ela mesma) ou relacionamentos ternários (onde três tabelas estão diretamente conectadas entre si) usando a funcionalidade de relacionamentos padrão.

Relacionamentos de auto-relacionamento geralmente ocorrem em cenários onde uma tabela tem uma relação hierárquica consigo mesma, como em uma tabela de funcionários onde cada funcionário tem um gerente que

também é um funcionário na mesma tabela. No entanto, no Power BI, é possível simular relacionamentos de auto-relacionamento usando técnicas como a criação de uma tabela auxiliar para representar a hierarquia.

Relacionamentos ternários, por outro lado, envolvem a interconexão direta de três tabelas. Embora seja possível modelar relacionamentos ternários em um modelo de dados, o Power BI não oferece suporte direto para definir esses relacionamentos usando a funcionalidade de relacionamentos padrão.

No entanto, apesar dessas limitações, muitas vezes é possível contornar esses cenários complexos usando técnicas avançadas de modelagem de dados, como a criação de tabelas auxiliares ou o uso de funções DAX para manipulação de relacionamentos.

RELACIONAMENTOS BIDIRECIONAIS

Relacionamentos bidirecionais, permitem a navegação em ambas as direções entre duas tabelas. Essa funcionalidade permite seguir o relacionamento de uma tabela para outra e, em seguida, retornar à tabela original usando o mesmo relacionamento.

Com um relacionamento bidirecional, os filtros aplicados em uma tabela afetam automaticamente a outra tabela relacionada e vice-versa. Isso é especialmente útil em cenários específicos onde é desejável que as interações em uma visualização tenham impacto nos resultados em outras visualizações conectadas por meio do relacionamento.

A capacidade de utilizar relacionamentos bidirecionais no Power BI oferece flexibilidade adicional na criação de análises interativas e dinâmicas, permitindo uma exploração mais profunda e uma compreensão mais completa dos dados. No entanto, é importante usar essa funcionalidade com cuidado, pois relacionamentos bidirecionais podem aumentar a complexidade das análises e afetar o desempenho do modelo de dados se não forem gerenciados adequadamente.

PROPAGAÇÃO DE FILTRO

A propagação de filtro, no contexto do Power BI e outras ferramentas de análise de dados, refere-se à capacidade de aplicar filtros em uma visualização e ter esses filtros automaticamente refletidos em outras visualizações conectadas por meio de relacionamentos de dados.

Por exemplo, quando um filtro é aplicado para exibir apenas as vendas de um produto específico em uma visualização, se outra visualização estiver conectada por meio de um relacionamento entre as tabelas de produtos e vendas, ela será automaticamente filtrada para mostrar informações relevantes apenas para esse produto específico.

Essa funcionalidade é fundamental para análises interativas, permitindo que os usuários explorem os dados de diferentes maneiras sem a necessidade de aplicar manualmente os mesmos filtros em várias visualizações. Além disso, a propagação de filtro torna as análises mais dinâmicas e intuitivas, facilitando a descoberta de insights importantes.

No entanto, é crucial garantir a integridade dos relacionamentos no modelo de dados para que a propagação de filtro funcione conforme desejado. Se os relacionamentos não estiverem configurados corretamente, os filtros podem não ser propagados como esperado ou podem resultar em resultados incorretos.

Suponhamos que existam duas visualizações em um relatório do Power BI:

Visualização de Vendas por Região: Esta visualização mostra o total de vendas por região em um gráfico de barras.

Visualização de Lucro por Produto: Esta visualização mostra o lucro por produto em uma tabela.

Quando um filtro é aplicado na visualização de vendas para mostrar apenas as vendas de um produto específico, como um smartphone, a propagação de filtro é ativada:

A visualização de vendas é filtrada automaticamente para mostrar apenas as vendas do smartphone selecionado.

Ao mesmo tempo, a visualização de lucro por produto é automaticamente filtrada para exibir o lucro associado apenas ao smartphone selecionado na visualização de vendas. Dessa forma, ao explorar os dados em uma visualização e aplicar filtros específicos, esses filtros são automaticamente refletidos em outras visualizações conectadas por meio de relacionamentos, permitindo uma análise coesa e interativa dos dados.

HIERARQUIAS NO POWER BI

No Power BI, a hierarquia é uma estrutura organizada de dados que nos permite analisar informações de maneira mais detalhada ou resumida, navegando por diferentes níveis de agregação. Isso é especialmente útil quando se trabalha com conjuntos de dados que possuem dados organizados em diferentes níveis de granularidade, como datas, regiões geográficas, categorias de produtos, entre outros.

Por exemplo, suponha que tenhamos dados de vendas que incluem datas de transação, mês, trimestre e ano. Podemos criar uma hierarquia onde o ano contenha trimestres, que por sua vez contêm meses, e assim por diante. Com essa hierarquia, podemos visualizar os dados em níveis mais altos, como o total de vendas por ano, e depois aprofundar a análise para ver os trimestres ou meses individuais.

No Power BI, podemos criar hierarquias facilmente arrastando e soltando os campos desejados na área de hierarquia da visualização. Uma vez criada, a hierarquia pode ser usada em diferentes visualizações para explorar dados em diferentes níveis de detalhe.

As hierarquias são úteis para simplificar a análise de conjuntos de dados complexos, permitindo focar nos detalhes relevantes e navegar pelos dados de maneira intuitiva. Isso facilita a identificação de padrões, tendências e insights significativos nos dados.

MEDIDAS

Medidas são expressões ou fórmulas que realizam cálculos sobre os dados de um conjunto de dados específico. Elas são fundamentais para a criação de análises e visualizações mais avançadas, permitindo extrair informações úteis e relevantes a partir dos dados brutos.

As medidas podem executar uma variedade de operações, como soma, contagem, média, mínimo, máximo, entre outras, dependendo das necessidades de análise. Além disso, as medidas podem ser criadas usando funções específicas do Power BI, como SUM, AVERAGE, COUNT, MAX, MIN, entre outras, ou usando expressões DAX (Data Analysis Expressions), a linguagem de fórmula usada no Power BI.

As medidas são especialmente úteis quando se trabalha com conjuntos de dados complexos que exigem cálculos personalizados ou agregações específicas. Por exemplo, pode-se criar uma medida para calcular a receita total de vendas, o lucro líquido, a taxa de crescimento, entre outras métricas importantes para a análise.

Uma vez criadas, as medidas podem ser usadas em diversas visualizações dentro do relatório ou painel, permitindo analisar os dados de diferentes perspectivas e em diferentes contextos. Elas também podem ser combinadas com filtros, slicers e outras funcionalidades interativas do Power BI para oferecer uma análise mais dinâmica e personalizada.

Em resumo, as medidas são ferramentas poderosas no Power BI que permitem realizar cálculos personalizados e extrair insights valiosos dos dados, contribuindo para uma análise mais profunda e informada.

NORMALIZAÇÃO

A normalização é um processo utilizado para organizar dados de forma eficiente, minimizando a redundância e a inconsistência. O objetivo principal da normalização é eliminar a redundância de dados, garantindo que cada informação seja armazenada apenas uma vez, o que ajuda a reduzir o espaço de armazenamento necessário e a evitar problemas de inconsistência.

Existem várias formas normais (1NF, 2NF, 3NF, BCNF, 4NF, 5NF, entre outras), cada uma com critérios específicos que um banco de dados deve atender para ser considerado normalizado nesse nível. Geralmente, o processo de normalização envolve dividir grandes tabelas em tabelas menores e mais coesas, estabelecendo relacionamentos entre elas por meio de chaves primárias e estrangeiras.

Por exemplo, suponha que você tenha uma tabela de clientes que inclui informações como nome, endereço e número de telefone, e outra tabela de pedidos que inclui informações sobre os produtos pedidos pelos clientes. Ao invés de repetir as informações de clientes em cada linha da tabela de pedidos, você pode criar uma tabela separada para clientes e estabelecer uma relação entre as tabelas usando uma chave estrangeira, como o ID do cliente. Isso reduz a redundância e facilita a manutenção dos dados.

Em resumo, a normalização é um processo importante no design de bancos de dados que visa garantir a eficiência, integridade e consistência dos dados, facilitando o gerenciamento e a análise das informações armazenadas.

INTEGRIDADE REFERENCIAL NO POWER BI

No Power BI, a integridade referencial é um conceito importante, embora a ferramenta não execute verificações automáticas como um banco de dados relacional faria.

É responsabilidade do modelador de dados ou do analista que está criando o relatório garantir que os relacionamentos entre tabelas estejam corretamente configurados para manter a integridade referencial.

Na prática, isso significa que, ao modelar os dados no Power BI, é necessário garantir que os relacionamentos entre as tabelas estejam configurados adequadamente, especialmente entre tabelas de dimensão e tabelas de fato. Embora o Power BI não sinalize explicitamente erros em caso de valores órfãos na tabela de dimensão na tabela de fato, a ausência desses valores pode levar a resultados imprecisos em análises e visualizações.

Portanto, ao trabalhar com o Power BI, é fundamental que seja verificada a integridade referencial e que seja tomada medidas proativas para garantir que os relacionamentos estejam configurados corretamente e que os dados estejam consistentes, garantindo análises precisas e confiáveis.

Em um cenário típico, duas tabelas são comumente encontradas em um modelo de dados no Power BI: uma tabela de vendas (tabela de fato) e uma tabela de clientes (tabela de dimensão).

Na tabela de vendas, são registradas informações como ID do cliente, ID do produto, data da venda e valor da compra. Na tabela de clientes, são armazenadas informações como ID do cliente, nome, endereço e e-mail.

Para assegurar a integridade referencial, é essencial estabelecer um relacionamento entre a tabela de vendas e a tabela de clientes com base no ID do cliente. Isso viabiliza a análise dos dados de vendas e a possível inclusão de informações complementares dos clientes, como nome e endereço, em análises e visualizações.

Considerando que o relacionamento entre essas tabelas tenha sido configurado corretamente, a inclusão de um novo registro de venda na tabela de vendas com um ID de cliente inexistente na tabela de clientes poderia resultar em um valor "órfão". Nesse contexto, um registro de venda estaria referenciando um cliente que não consta na tabela de clientes.

No Power BI, caso o ID do cliente seja utilizado como dimensão em uma visualização e ocorra um valor órfão na tabela de vendas, tal cliente não terá informações associadas, como nome, endereço, etc. Isso pode impactar a precisão e interpretação das análises.

Portanto, para garantir a integridade referencial e evitar valores órfãos, é crucial que os modeladores de dados verifiquem e mantenham a consistência dos dados em seus modelos do Power BI, certificando-se de que todos os valores de referência estejam presentes e corretamente relacionados entre as tabelas.