# O QUE É UM DATAWAREHOUSE

Um data warehouse é um sistema de armazenamento de dados projetado para permitir a análise e a geração de relatórios de dados de maneira eficiente. Ele é projetado especificamente para consolidar, organizar e integrar dados de várias fontes diferentes em um único local centralizado.

A principal função de um data warehouse é fornecer um ambiente otimizado para análise de dados. Isso é alcançado por meio da estruturação dos dados de uma forma que facilite consultas complexas e análises de grandes conjuntos de dados. Os dados em um data warehouse geralmente são organizados em torno de assuntos ou áreas de interesse específicas, o que facilita a análise de tendências e padrões ao longo do tempo.

Além disso, os data warehouses frequentemente incluem recursos para limpeza, transformação e enriquecimento de dados, garantindo a qualidade e a consistência dos dados armazenados. Eles também podem suportar operações de carga de dados incremental para manter os dados atualizados e prontos para análise em tempo real.

Os data warehouses são amplamente utilizados em organizações para análise de negócios, inteligência empresarial, tomada de decisões estratégicas e geração de relatórios. Eles desempenham um papel fundamental na transformação de grandes volumes de dados em insights acionáveis que podem impulsionar o desempenho e a eficácia das organizações em diversos setores e indústrias.

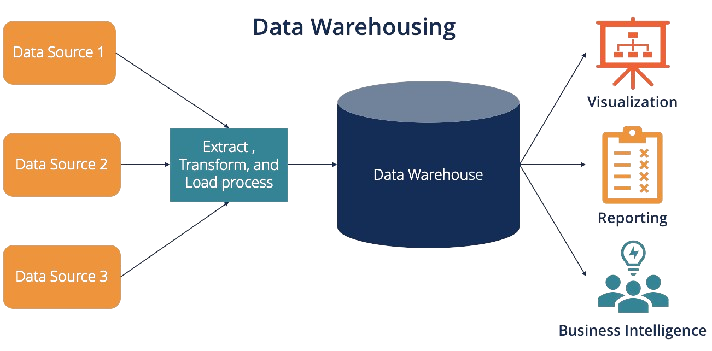
Um exemplo de data warehouse pode ser encontrado em uma empresa de varejo que opera em várias regiões geográficas. Neste cenário, a empresa pode reunir dados de suas várias fontes operacionais, como sistemas de vendas, sistemas de inventário, sistemas de compras, sistemas de gestão de clientes, entre outros, em um único repositório centralizado de dados, ou seja, o data warehouse.

Neste data warehouse, os dados são organizados e estruturados de forma a facilitar a análise e a geração de relatórios sobre as operações da empresa. Por exemplo, a empresa pode ter tabelas que armazenam informações detalhadas sobre vendas, como datas de transação, produtos vendidos, preços de venda, região geográfica da venda e informações do cliente.

Além disso, pode haver tabelas separadas que armazenam informações sobre inventário, como quantidade em estoque, datas de reposição, localização física dos produtos, etc. Outras tabelas podem conter informações sobre clientes, como histórico de compras, preferências de produtos, dados demográficos, etc.

Com todos esses dados consolidados em um único local, os analistas de negócios da empresa podem realizar análises avançadas para identificar tendências de vendas, padrões de compra, eficácia de campanhas de marketing, gerenciamento de inventário e muitos outros insights que podem impulsionar as decisões de negócios.

Essa capacidade de consolidar e analisar dados de várias fontes em um data warehouse é o que torna essa ferramenta tão valiosa para empresas que desejam obter uma compreensão mais profunda de suas operações e do comportamento do mercado.



# MODELAGEM DE BANCO DE DADOS TRANSACIONAL e MODELAGEM DE DATA WAREHOUSE

Na modelagem de banco de dados transacional, o foco está na eficiência e na consistência dos processos de transação. Os bancos de dados transacionais são projetados para dar suporte às operações diárias de uma organização, como inserção, atualização e exclusão de dados. A modelagem transacional geralmente envolve a normalização muito grande dos dados para evitar redundâncias e garantir a integridade dos dados durante as transações.

Por outro lado, na modelagem de data warehouse, o foco está na análise e geração de relatórios de dados para fins de tomada de decisões. Os data warehouses são projetados para armazenar grandes volumes de dados de várias fontes em um formato otimizado para consultas complexas e análises de dados. Isso muitas vezes envolve a denormalização dos dados para facilitar consultas rápidas e análises de tendências ao longo do tempo.

Em resumo, enquanto a modelagem de banco de dados transacional prioriza a eficiência e a consistência das transações diárias, a modelagem de data warehouse é focada na análise e geração de relatórios de dados para suportar a tomada de decisões estratégicas. Ambos os tipos de modelagem têm suas próprias considerações e técnicas específicas para atender às necessidades de suas respectivas aplicações.

**O Power BI é uma ferramenta de análise de dados e não de amarzenamento de dados. Para grandes volumes de dados, o ideal é construtir um DW e usá-lo como fonte no Power BI**

# MODELO STAR SCHEMA

O modelo star schema é um dos modelos de design de banco de dados mais comuns e é amplamente utilizado para organizar e estruturar dados em um data warehouse. Ele é composto por uma tabela de fatos central, cercada por várias tabelas de dimensão.

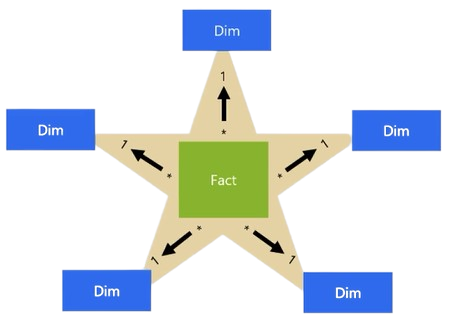
Na estrutura do modelo star schema, a tabela de fatos contém métricas e medidas que representam os eventos ou transações que estão sendo analisados, como vendas, pedidos ou transações financeiras. Essa tabela de fatos geralmente contém chaves estrangeiras que se relacionam com as tabelas de dimensão.

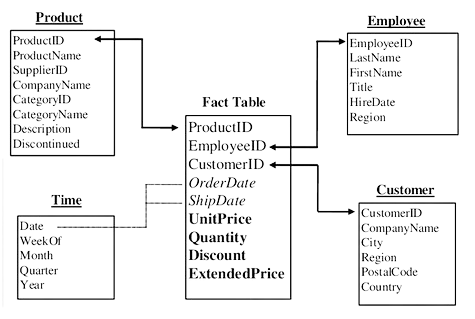
Por sua vez, as tabelas de dimensão contêm atributos descritivos que fornecem contexto e detalhes sobre os dados na tabela de fatos. Esses atributos podem incluir informações como data, localização, produto, cliente, entre outros.

A tabela de fatos está conectada diretamente às tabelas de dimensão através de chaves estrangeiras, formando uma estrutura de estrela quando visualizada graficamente, com a tabela de fatos no centro e as tabelas de dimensão ao redor dela, como os raios de uma estrela.

Esse modelo de design é chamado de "star schema" devido à sua aparência visual, onde a tabela de fatos está no centro, como a estrela, e as tabelas de dimensão são como os raios que irradiam dela.

O modelo star schema é preferido em muitos casos devido à sua simplicidade e facilidade de compreensão, o que torna as consultas e análises de dados mais diretas e eficientes. Ele também oferece excelente desempenho para consultas analíticas, especialmente ao trabalhar com grandes volumes de dados.





# MODELO SNOWFLAKE

O modelo snowflake é uma variação do modelo star schema. Enquanto o modelo star schema é caracterizado por ter apenas uma tabela de dimensão para cada dimensão, o modelo snowflake normaliza ainda mais as tabelas de dimensão, dividindo-as em subdimensões.

Na estrutura do modelo snowflake, as tabelas de dimensão são normalizadas, o que significa que os atributos de uma dimensão são armazenados em várias tabelas relacionadas. Por exemplo, em vez de ter uma única tabela de dimensão de produto, o modelo snowflake pode ter uma tabela para categorias de produtos, outra para subcategorias e outra para detalhes específicos do produto.

Essa estrutura hierárquica de tabelas de dimensão resulta em um esquema que se assemelha a um floco de neve, daí o nome "snowflake" (floco de neve) para o modelo.

O modelo snowflake é útil em situações em que a normalização adicional das tabelas de dimensão pode economizar espaço de armazenamento e facilitar a manutenção dos dados, especialmente em casos onde há muitos atributos de dimensão repetidos ou onde certos atributos são compartilhados por múltiplas dimensões.

No entanto, o modelo snowflake também pode complicar as consultas e análises de dados, pois requer que várias tabelas sejam unidas em consultas para obter uma visão completa dos dados. Isso pode afetar o desempenho das consultas, especialmente em ambientes com grandes volumes de dados.

Portanto, a escolha entre o modelo star schema e o modelo snowflake depende das necessidades específicas do projeto, incluindo considerações de desempenho, complexidade da consulta e facilidade de manutenção dos dados.

