

TEXTO DE APOIO



AULA 2

Tópicos de Banco de Dados

Professor Jamilson Bispo dos Santos



Universidade Presbiteriana
Mackenzie





Sumário



APOIO À DECISÃO.....	3
DATA WAREHOUSE	5
MODELO MULTIDIMENSIONAL.....	6
TABELAS FATO.....	9
TABELAS DIMENSÃO	10
REFERÊNCIAS BIBLIOGRÁFICAS	13

MODELO MULTIDIMENSIONAL

APOIO À DECISÃO

As aplicações de banco de dados geralmente são classificadas em sistemas de processamento de transação e sistemas de apoio à decisão. Os sistemas de processamento de transação, também conhecidos como OLTP (Online Transaction Processing) são sistemas que registram informações sobre transações.

Uma transação é uma unidade de execução do programa que pode incluir, alterar ou excluir vários itens de dados de forma controlada. A integridade dos dados é garantida com os usos das seguintes propriedades das transações:

- Atomicidade: Todas as operações da transação são refletidas corretamente no banco de dados ou nenhuma.
- Consistência: A execução de uma transação preserva a consistência dos dados.
- Isolamento: Cada transação não percebe a existência de outras transações.
- Durabilidade: As mudanças efetuadas durante a execução da transação deverão ser persistidas no banco de dados, mesmo que existam falhas no sistema.

O termo processamento analítico on-line (OLAP, Online Analytical Processing) pode ser definido como o processo de criar, administrar, analisar e gerar relatórios sobre dados, que auxiliam no processo de tomada de decisão. É importante salientar que o processamento analítico exige algum tipo de agregação de dados.

Os sistemas de apoio à decisão são sistemas que auxiliam na análise de informação do negócio. A tomada de decisão exige uma visão abrangente de todos os aspectos da empresa. Para os sistemas de apoio à decisão, a consolidação de informações de vários bancos em um local único, denominado de Data Warehouse, facilitou e incentivou a utilização de ferramentas de análise dos dados.

Muitas características das consultas de apoio à decisão tornam os sistemas SQL tradicionais inadequados:

- Cláusula WHERE com muitas condições AND e OR.
- Funções estatísticas não suportadas na linguagem SQL.
- A linguagem SQL não fornece suporte adequado em consultas que envolvam condições ao longo do tempo ou que exigem agregação ao longo do tempo.
- O ambiente tradicional não está adaptado à execução em conjunto de muitas consultas relacionadas.

Os bancos de dados de apoio à decisão têm como característica principal a de ser apenas de leitura. A atualização é realizada somente por meio de carga dos dados. Não é permitido qualquer tipo de atualização na informação e a exclusão de dados é feita de forma ocasional.

As principais características dos bancos de dados de apoio à decisão são:

- Colunas são usadas em combinação.
- Integridade não essencial, é importante considerar que os dados são originais de sistemas transacionais e já foram devidamente validados.
- A inclusão de campo com características de tempo nas chaves.
- Banco de dados para tomada de decisão acumulam grande quantidade de informação em comparação aos bancos de dados tradicionais.
- Os bancos de dados são indexados.
- A redundância de informação é controlada, mas não é tão importante como nos bancos de dados tradicionais.

As características dos bancos de dados de apoio à decisão associadas às consultas conduzem a necessidade de desempenho.

DATA WAREHOUSE

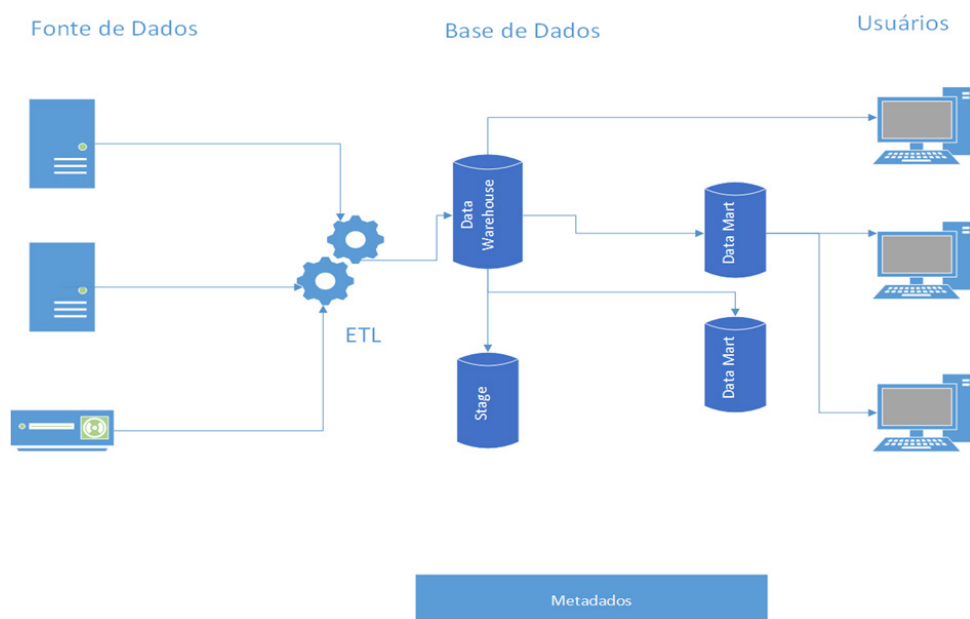
O Data Warehouse é um banco de dados que contém dados selecionados de sistemas transacionais. É estruturado para processamento de consultas, orientado por assunto, não volátil, histórico e integrados com o principal objeto de facilitar a obtenção de dados de apoio à decisão, transformando dados em informação.

Um Data Warehouse deve ter as seguintes características:

- Integração.
- Orientação por assunto.
- Variabilidade no tempo.
- Não volatilidade.

O Data Warehouse é estruturado como ilustra a Figura 1. Os dados fluem do ambiente operacional para o Data Warehouse. Na passagem do nível operacional para o nível de Data Warehouse, ocorre uma quantidade de transformações sobre os dados.

Figura 1 – Estrutura do Data Warehouse



A fonte de dados da Figura 1 representa de onde os dados são retirados para compor o Data Warehouse até a utilização pelo usuário final. Os dados passam por um processo de ETL (extração, transformação e carga) onde geralmente são enviados para uma área intermediária chamada de Stage Area e depois são carregados em Data Marts que têm informação departamental ou para o Data Warehouse.

MODELO MULTIDIMENSIONAL

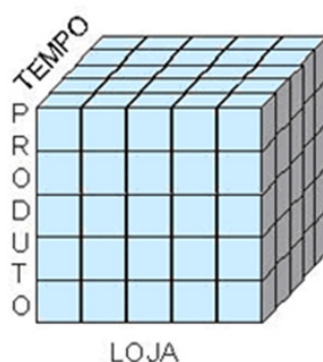
O Modelo Relacional (MR) é o modelo de dados efetivamente usado em aplicações e o ambiente operacional que apoia os usuários em suas funções do dia a dia são chamados OLTP (Online Transaction Processing).

Os sistemas OLTP são pouco flexíveis em relação à quantidade de relatórios e consultas. Para ajudar no processo de tomada de decisões, é recomendável o uso de ferramentas flexíveis e customizáveis. Os sistemas OLAP (Online Analytical Processing) permitem aos usuários navegarem entre os dados da empresa com maior facilidade, proporcionando uma visão multidimensional desses dados.

A modelagem multidimensional é a técnica que busca apresentar os dados em um modelo padronizado e intuitivo, que permita o entendimento por parte de usuários, com alto desempenho de acesso.

Modelo de dados Multidimensional são ambientes em que as informações sobre um Fato do negócio se relacionam com as Dimensões relevantes para sua análise, sendo representadas por um Cubo ou hipercubo (mais que três dimensões), conforme a figura abaixo:

Figura 2 – Cubo de três dimensões



O foco da modelagem multidimensional é o cruzamento de variáveis e a simplicidade deve ser buscada. A visão inicial deve ser conceitual para que se entenda a necessidade de negócio.

Para a criação de um modelo multidimensional, é necessário responder às seguintes questões:

- Quando?
- Onde?
- O quê?
- Quem?
- Como?

A Figura 3, na forma de estrela, ilustra esse questionamento.

Figura 3– Perguntas a serem respondidas para elaboração de um modelo multidimensional genérico

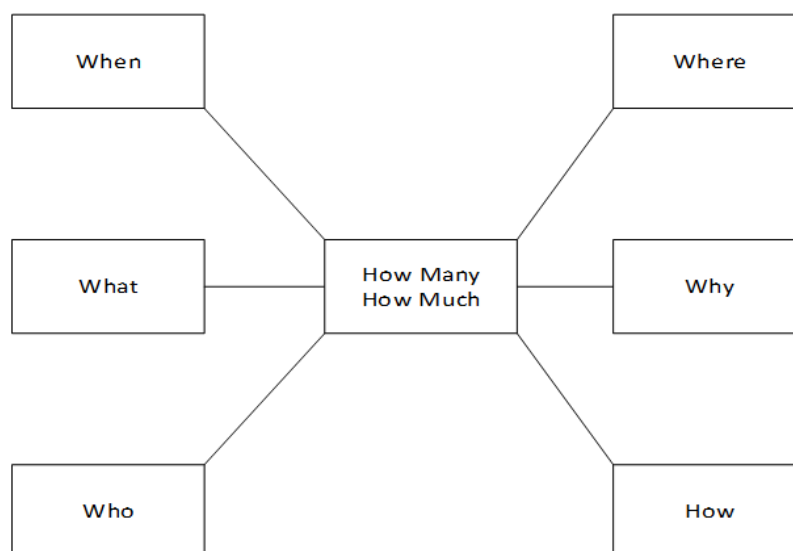


Figura 4 – Questões para um sistema de vendas

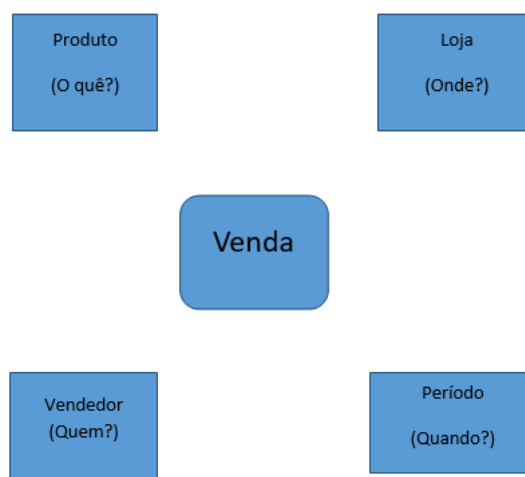


Figura 5 – Questões para um sistema de compras

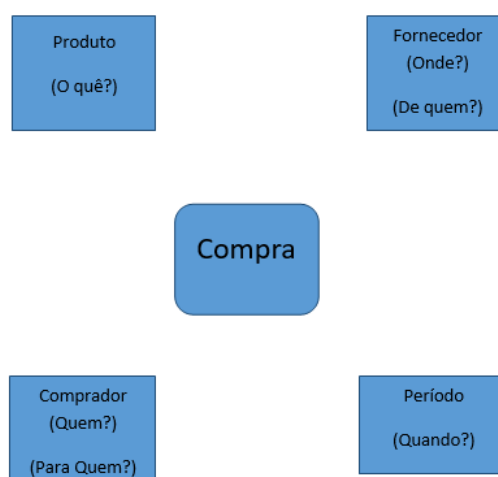
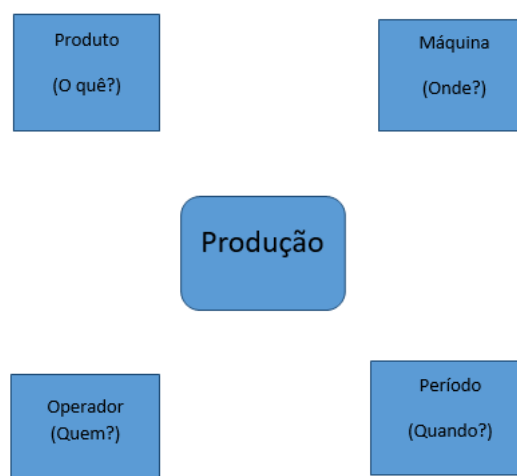


Figura 6 – Questões para um sistema de produção



Outro item necessário na elaboração do modelo multidimensional são as tabelas Fato e Dimensão.

TABELAS FATO

Fatos são as informações importantes para a gestão do negócio, são números e valores que podem ser somados, agregados para indicar o desempenho da empresa, unidade de negócio ou departamento. São identificados durante a fase de Análise de Requisitos do projeto.

Tabelas Fato têm as seguintes características:

- Armazenam os dados que medem o processo que estamos modelando.
- Compõe-se de Chaves e Métricas.

Figura 7 – Tabela Fato identificado como Fato_Vendas

Para identificar as Tabelas Fato, responda às seguintes questões:

Que processo está modelando?

Fato_Vendas
idTempo
IdLoja
IdProduto
IdCliente
QtdeVendida
VIVenda

- Por exemplo: Vendas.
- O que se usa para medir esse processo?
- Por exemplo: Valor das vendas, quantidade vendida.
- A que nível esta métrica é significativa? Granularidade

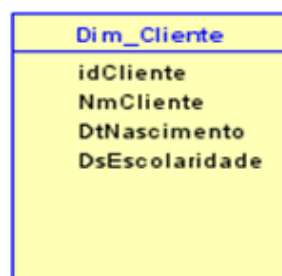
Por exemplo: Vendas no trimestre.

TABELAS DIMENSÃO

As dimensões são as formas como os gestores querem visualizar o negócio. São informações, pois dão significado aos números (medidas, valores, métricas)

As dimensões são identificadas durante a fase de Análise de requisitos. A Tabela Dimensão é uma perspectiva pela qual se pretende observar as métricas relativas ao processo que estamos modelando. Compõe-se de um ou mais atributos.

Figura 8 – Tabela Dimensão identificada como Dim_Cliente



Descobrimos tabela dimensão:

- Geralmente, as dimensões são derivadas das perguntas do usuário após entendimento da métrica.
- Reescreva a pergunta do usuário utilizando a seguinte forma: Métrica POR Dimensão POR Dimensão.

Por exemplo:

“Quais foram as vendas no período de junho a julho deste ano na região sul?”

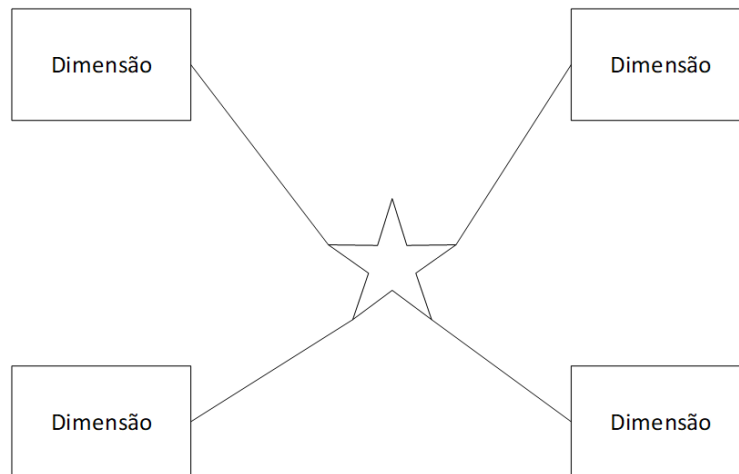
Seria: Valor das vendas POR período POR região

Existem diversas técnicas de modelagem de Data Warehouse. As mais utilizadas são o esquema Estrela (Star Schema) e o Floco de Neve (Snowflake); será comentado o esquema Estrela.

No modelo Estrela, um único FATO (tabela fato) é colocado no centro e ligado às DIMENSÕES (tabelas dimensões) como uma estrela.

Cada dimensão é representada por uma única tabela. A chave primária (PK) de cada dimensão está relacionada com uma chave estrangeira (FK) da tabela fato. E todas as medidas da tabela Fato estão relacionadas com todas as Dimensões, isto significa que elas têm o mesmo nível de Granularidade.

Figura 9 – Esquema Estrema



Fato

Medidas

Granularidade

Dimensão

Atributos

Hierarquias

Na Figura 10, está ilustrado um exemplo do esquema Estrela.

Figura 10 – Exemplo esquema Estrela



Fato

Os dados foram consolidados por dia e por venda (ETL)

Valor da venda por dia e por vendedor

Dimensões

Tempo

Vendedor

REFERÊNCIAS BIBLIOGRÁFICAS

DATE, C. J. Introdução a Sistemas de Banco de Dados. 7. ed. São Paulo: Campus, 2000.

- Capítulo 21: Apoio à decisão

ELMASRI, R.; NAVATHE, S. Sistema de banco de dados. 6. ed. São Paulo: Pearson, 2011 | Biblioteca Virtual Universitária — Pearson.

- Capítulo 25: Data Warehousing e Apoio à Decisão.

INMON, W. H.; WELCH, J. D; GLASSEY, K. L. Gerenciando Data Warehouse. São Paulo: Makron Books, 1999.

- Capítulo 1: Monitorando a Atividade do Data Warehouse
- Capítulo 2: Monitorando Dados do Data Warehouse

SILBERCHATS, A., KORTH, H. F.; SUDARSHAN, S. Sistemas de Banco de Dados. São Paulo: Elsevier, 2012.

- Capítulo 18: Mineração e Análise de Dados