

Sistemas de Apoio à Tomada de Decisão

Prof^a. Silvia Inês Dallavalle de Pádua

Banco de Dados e Data warehouse

Bibliografia

- LAUDON, K.C., LAUDON, J.P. Sistemas de Informação Gerenciais. 7^a.ed. São Paulo: Pearson Prentice Hall, 2007.
- LEME FILHO, T. Business Intelligence no Microsoft Excel. Axexcel Books, 2004.
- LUCAS JR.H. C. Tecnologia da informação – tomada de decisão estratégica para administradores. Ed. LTC, 2006.
- O' BRIEN, J.A. Sistemas de Informação e as decisões gerencias na era da internet.Saraiva, 2005.
- SINGH, H.S. Data warehouse conceitos, tecnologias, implementação e gerenciamento. Makron Books, 2001.

- Modelo de Dados Relacional
- Data warehousing
- Ferramentas OLAP

Banco de Dados

- Banco de dados: conjunto de arquivos relacionados entre si que contêm registros sobre pessoas, lugares ou coisas
- Entidades e atributos
- Organizando dados em um banco de dados relacional
 - Campos, registros, campos-chave, chave primária, chave estrangeira
- Estabelecendo relacionamentos
 - Diagrama entidade-relacionamento

Banco de Dados

- Modelo de banco de dados relacional se baseia no conceito simples de tabelas, para explorar as características de linha e colunas de dados.
- Em um banco de dados relacional, essas tabelas são denominadas **relações** e o modelo é baseado na **teoria matemática de conjuntos e relações**.
- Cada linha de dados equivale a um registro, e cada coluna de dados é equivalente a um campo.

Modelo de Banco de Dados Relacional

- Na terminologia do modelo relacional, uma linha é denominada uma tupla e uma coluna um atributo
- É o modelo mais popular.

Tabela de banco de dados relacional

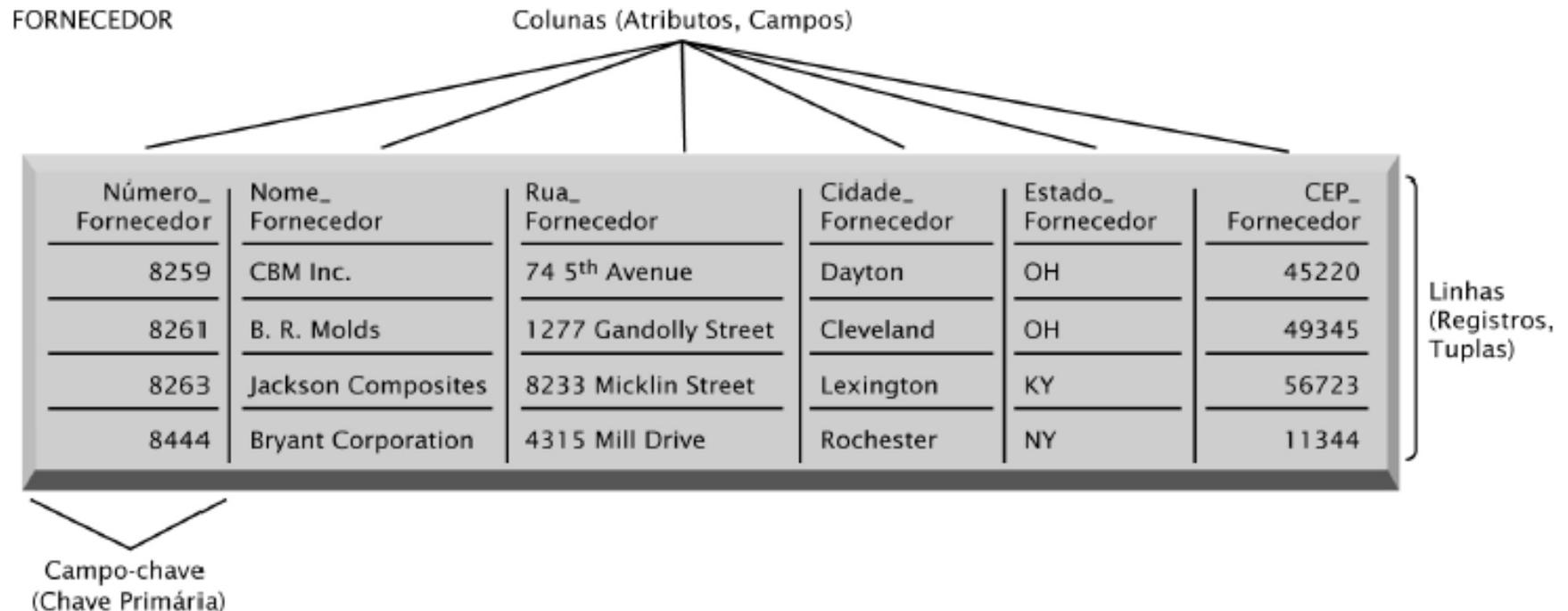


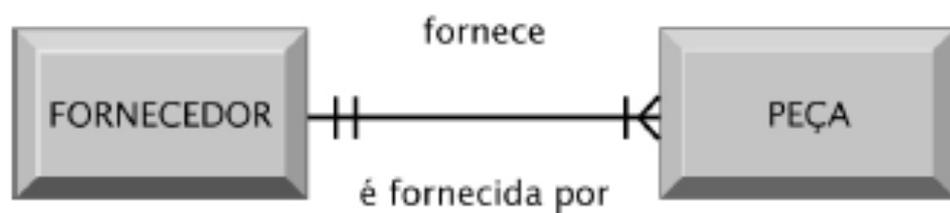
Figura 5.1

Tabela de banco de dados relacional.

Um banco de dados relacional organiza os dados na forma de tabelas bicimensionais. Temos aqui uma tabela para a entidade FORNECEDOR e vemos como ela representa a entidade e seus atributos. Número_ Fornecedor é o campo-chave.

Figura 5.3

Um diagrama entidade/relacionamento básico. Este diagrama mostra a relação entre as entidades FORNECEDOR e PEÇA.



PEÇA

Número_Peça	Descrição_Peça	Preço_Unitário	Número_Fornecedor
137	Trinco da porta	22,00	8259
145	Retrovisor externo	12,00	8444
150	Vedação da porta	6,00	8263
152	Trava da porta	31,00	8259
155	Compressor	54,00	8261
178	Maçaneta da porta	10,00	8259

Chave primária

Chave estrangeira

Figura 5.2

A tabela PEÇA.

Dados da entidade PEÇA têm sua própria tabela individual. Número_Peça é a chave primária e Número_Fornecedor é a chave estrangeira, que permite aos usuários achar, na tabela FORNECEDOR, informações correlatas sobre o fornecedor de cada peça.

Necessidade de informação

Necessidade de Informações

Negócios acumulam uma **quantidade enorme de dados** de transações como parte de suas operações de rotina.

Por exemplo, é possível solicitar uma **demonstração de vendas por região, pelo período de tempo corrente**, comparadas com o **mesmo período do ano passado**.

Após ver essa demonstração, você quer a análise **por grupo de produtos**, depois o **grupo de produtos pelos últimos seis meses** e, finalmente, os mesmos dados **por equipe de venda**.

Em vez das duas dimensões associadas ao **modelo relacional**, você está pedindo uma **análise multidimensional**.

O enfoque mudou da entrada e captura de dados através dos sistemas operacionais da empresa para o acesso e a disponibilidade de informações fornecidos pelos depósitos de dados empresariais

Data warehouse (DW) - Definição

Um data warehouse é um **conjunto de dados orientado por assunto, integrado, variável com o tempo e não volátil** (não se alteram), que fornece suporte a ao processo de tomada de decisão do negócio (Inmon, 1993)

Data warehouse (DW)

- Um DW é **orientado** com base nos principais **assuntos** empresariais, como **cliente, fornecedor, produto ou atividade**.

O data warehouse surgiu das várias tentativas realizadas por pesquisadores e organizações para criar um meio flexível, eficaz e eficiente de obter os conjuntos de dados que se tornaram os bens mais importantes e valiosos da empresa.

Orientado por assunto

- **Orientado por assunto** significa que foca as principais entidades do negócio; no caso do ensino superior, assuntos como alunos, cursos, departamentos e professores.
- **Integrado** significa que os dados estão armazenados em um formato consistente (ou seja, especificando convenções, restrições de domínios, atributos físicos e medições).
- **Variável com o tempo** significa que os dados estão associados a um ponto no tempo (semestre, ano fiscal e período de pagamento).

- Os dados no depósito são armazenados em um **único formato especificado**, mesmo quando as aplicações operacionais subjacentes armazenam dados de modo diferente
 - Por exemplo
Uma aplicação operacional pode armazenar a data como ano-mês-dia, e outra pode armazená-la como mês-dia-ano.

No DW, data terá um formato sempre consistente (mês-dia-ano, por exemplo)

Data warehouse (DW)

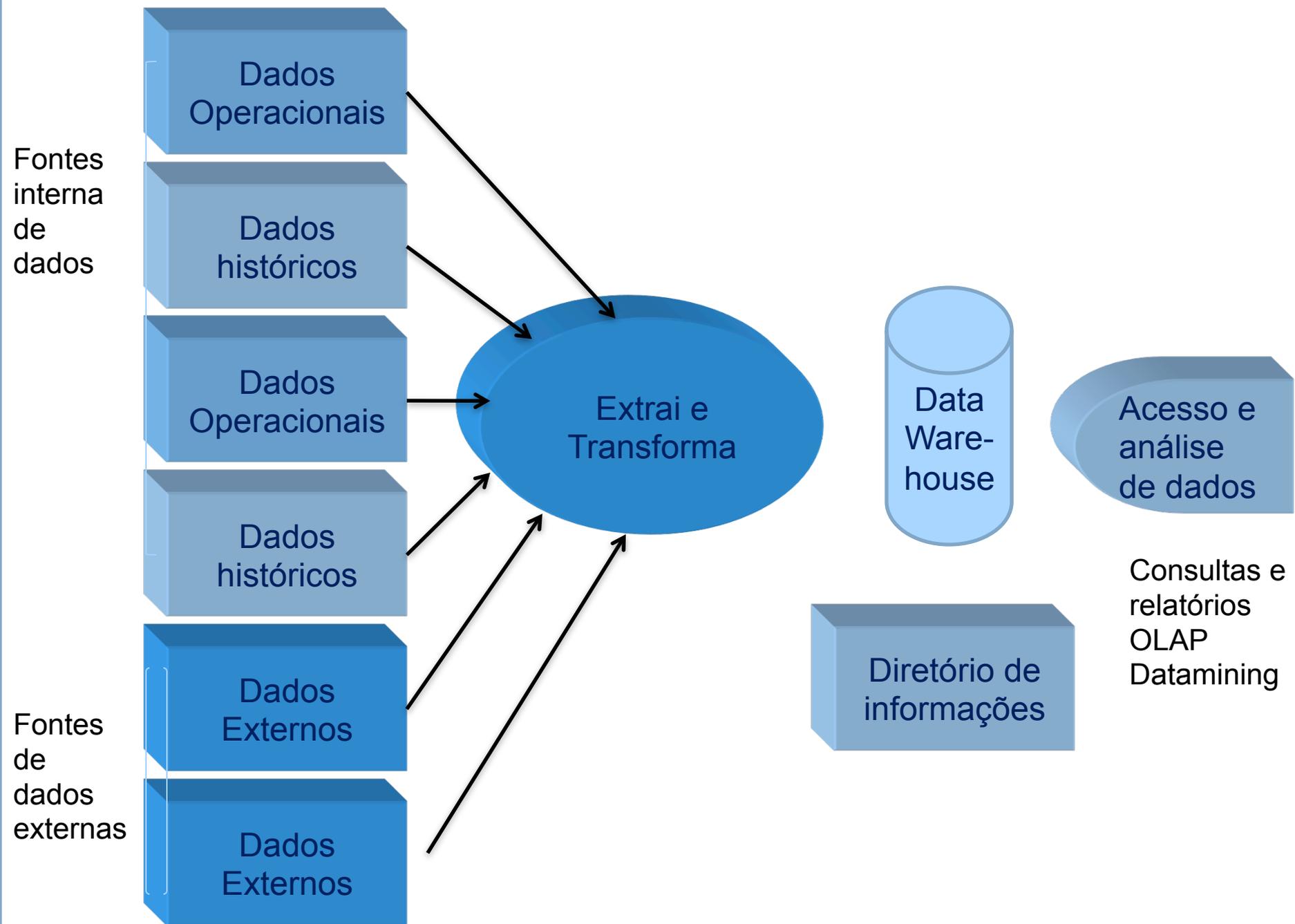
- O DW transforma os dados em um recurso mais útil:
 - **agrupando-os** de modo mais adequado para os usuários finais,
 - colocando-os em **formatos mais úteis**, permitindo a sua análise,
 - e **distribuindo-os** para os grupos de trabalhos pertinentes, para aumentar a disponibilidade e acessibilidade

O **data warehouse** consolida e padroniza as informações oriundas de diferentes banco de dados operacionais, de modo que elas possam ser utilizadas por toda a empresa para análise gerencial e tomada de decisões.

Próxima figura: **Componentes de um data warehouse.**

Fonte: Laudon, K.C., Laudon, J.P. (2007), p. 150

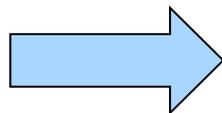
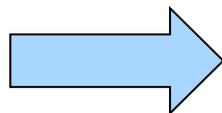
Os **dados** estão disponíveis a todos para acesso conforme a necessidade, **mas não podem ser alterados.**



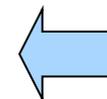
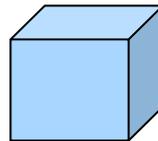
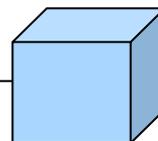
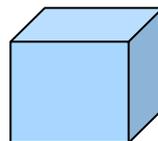
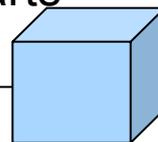
Origem dos dados

ETL

Extração, Transformação e Carga



Data marts



- Na **camada origem dos dados** é o ambiente operacional onde se encontram aplicativos de gestão e os consumidores e fornecedores interagem diretamente com os sistemas administrativos.
- Na **camada ETL** (Extração, Transformação e Carga) os dados não integrados dos aplicativos de gestão são convertidos para uma estrutura integrada, onde poderão sofrer severa transformação.

Os dados poderão passar por um processo de **transformação durante a carga**, quando são então **filtrados** para evitar duplicação, **calculados** para gerar indicadores, **agrupados e divididos**, entre outras funções.

Dados podem ser inseridos

Metas, por exemplo,
uma vez que muitos sistemas de gestão não oferecem esta
funcionalidade.

Neste caso, telas de inserção são criadas para que o BI possa trabalhar com metas estabelecidas de vendas, faturamento, lucro, etc.

Qualidade das informações

Depende fortemente da alimentação e da administração dos dados nos sistemas transacionais.

É muito comum também as empresas **possuírem vários cadastros de um mesmo cliente**, e todos com muito “lixo”. Há casos de todos os tipos: as vezes, um simples acento pode gerar dois cadastros para uma mesma pessoa, que pode também ser cadastrada pelo apelido num determinado momento, e noutra, pelo nome completo, gerando até mesmo três cadastros.

Extraction, Transformation and Loading - ETL

- Quase todos projetos de Data warehouse envolve o uso de uma ou mais ferramentas de ETL.
- A ferramenta ETL deve fornecer todas as definições dos modelos de dados do sistema de gestão e todos os programas de cargas prontos.
- Alguns dos principais fornecedores de banco de dados, como Oracle e Microsoft, oferecem ferramentas ETL

Modelagem dos Dados

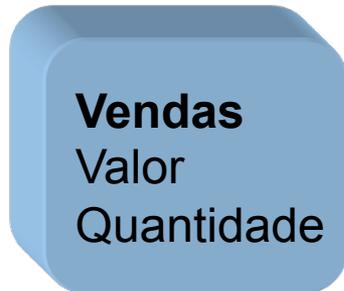
- Em um **sistema transacional**, uma transação desencadeia a outra, como a **emissão de uma nota fiscal** que por sua vez gera um **lançamento no estoque** e outro no módulo de **contas a receber**.
- **Sistemas de apoio à tomada de decisão** contemplam outras necessidades com grandes volumes de consultas, inviabilizando assim a modelagem tradicional

- A **modelagem dos dados** resulta da **compreensão do negócio** por parte do analista e da capacidade deste para empregar seu entendimento na formatação de um banco de dados (**aderência dos sistemas ao negócio da empresa**).
- A modelagem utilizada para data warehouse é a dimensional, mais voltada para muitos pontos de entradas específicos (**dimensões**) e menos para dados granulares (**fatos**)

Exemplo de dimensão: TEMPO



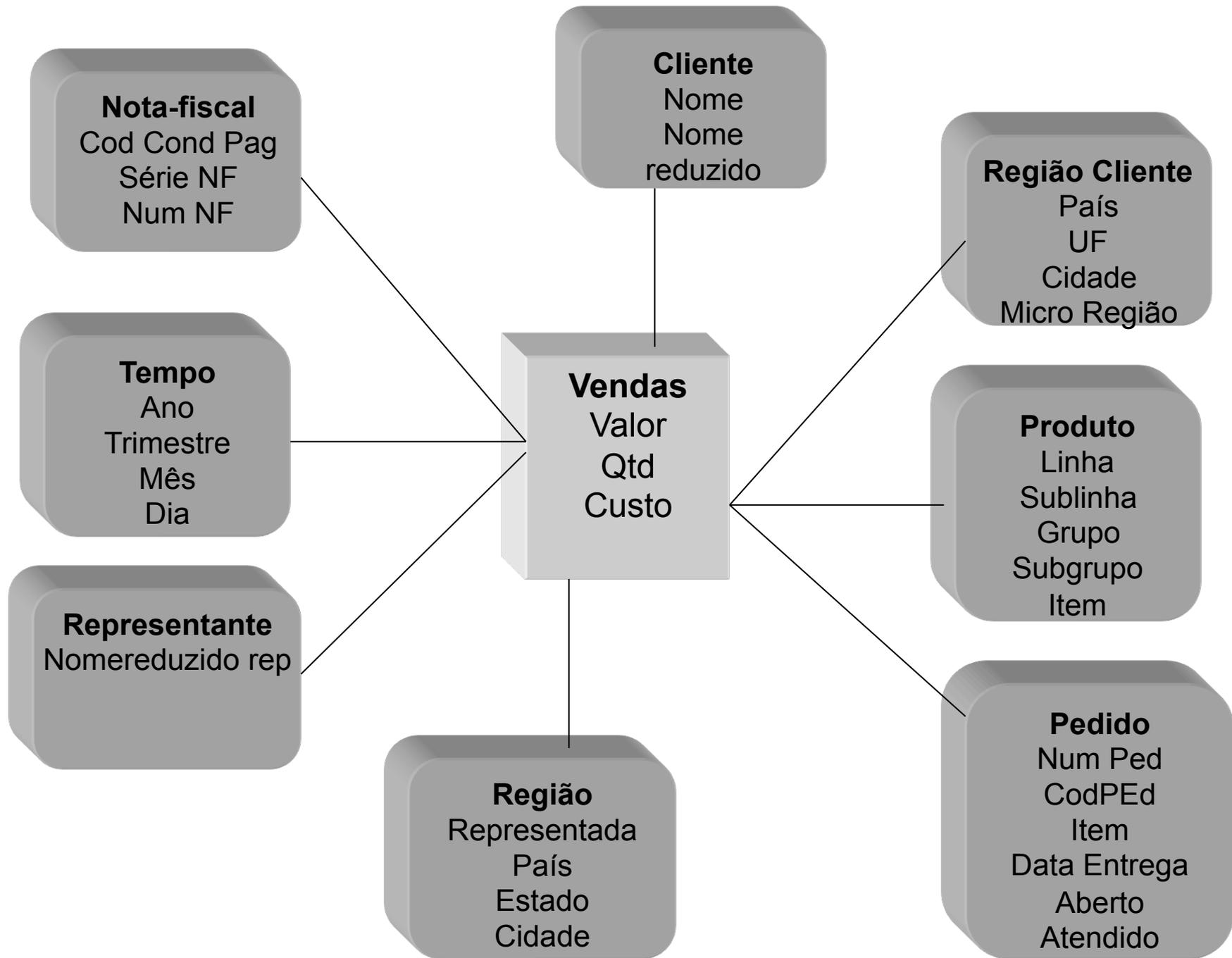
Exemplo de fato: VENDAS



Por exemplo:

- Dentro da dimensão tempo, é possível ter uma hierarquia representando os níveis anos, meses, e dias.
- Da mesma forma, a dimensão região poderá ter os níveis país, região, estado e cidade.
- Assim, um usuário visualizando dados em um modelo irá navegar para cima (drill up) ou para baixo (drill down) entre níveis para visualizar informação com maior ou menor nível de detalhe sem a menor dificuldade.

Em uma estrutura dimensional os dados estarão dispostos de forma quase estelar (daí o termo Star Schema), onde **várias tabelas de entrada (dimensões)** se relacionarão com algumas **tabelas de informações (fatos)**.



A modelagem dimensional permite ao usuário perceber e visualizar os dados de forma similar a sua compreensão, ou seja: em várias perspectivas, entre tempo e espaço.

Apesar de cumprir um papel relativamente menor no conjunto de componentes do data warehouse, **ela é certamente um dos principais motivos de insucesso para muitas soluções de BI**, uma vez que um desenvolvedor de sistemas desatento às diferenças entre modelagem tradicional e a dimensional pode comprometer a qualidade e a vida útil dos sistemas.

Benefícios

- Os usuários finais podem acessar dados **de forma rápida e facilmente por navegadores Web**, pois estão localizados em um lugar.
- Os usuários finais podem realizar **uma extensa análise com os dados**, de maneiras que não eram possíveis antes.
- Os usuários finais possuem uma **visão consolidada dos dados organizacionais**

- Que tipo de pergunta um DW pode responder?

Responde

- Como o produto A vendeu no mês passado? Como essa quantidade se compara com as vendas no mesmo mês nos últimos cinco anos? Como o produto vendeu por filial, região e território?
- Este produto vendeu melhor em determinadas regiões? Há tendências regionais?

Responde

- Clientes devolveram o Produto A no ano passado? A devolução foi devido a defeitos no produto? A empresa fabricou o produto em uma fábrica específica?
- As comissões e a definição de preços influenciaram a postura dos vendedores?
- Determinados vendedores executaram um trabalho melhor na venda do produto?

O problema da segurança

- Aumenta a medida que o data warehouse torna-se disponível a um número maior de usuários, a menos que sejam observados os seguintes requisitos:
 - Definição de segurança para cada usuário e grupo de trabalho.
 - Vínculo da segurança ao banco de dados.
 - Vínculo da segurança às dimensões do negócio.
 - Fácil manutenção da segurança

- Este banco de dados possui:
 - **Contas** possui quatro membros: Vendas, CMV (Custo da mercadoria Vendidas), Margem e Margem %
 - **Tempo** possui quatro membros trimestrais (a figura mostra apenas o primeiro trimestre)
 - **Cenário** possui dois membros-filho: Orçamento para valores de orçamento e Real para valores reais.

- É o componente mais importante do data warehouse
- Contém dados sobre dados
- São todas as coisas relacionadas ao conteúdo dos dados que permitem o entendimento de como o dado foi criado e como é mantido

Metadado é usado como:

- Um diretório que ajuda a localizar o conteúdo do data warehouse
- Um guia para o mapeamento dos dados à medida que estes são transformados do ambiente operacional para o ambiente do data warehouse

Metadado contém informações sobre

- A estrutura dos dados
- Algoritmos usados para a sumarização
- Mapeamento do ambiente operacional para o ambiente do data warehouse

Metadado pode conter

- Histórico da extração e transformação de dados
- Estatística de uso de dados
- Tamanho de tabelas do data warehouse
- Identificadores de colunas

Processos de Transformação

- Os processos de transformação e movimentação dos dados são executados sempre que se deseja atualizar o data warehouse
- Deve haver capacidade para gerenciar e automatizar os processos requeridos para executar essas funções

- Definição: data warehouse pequeno, projetado para uma **unidade estratégica de negócios (UEN)** ou um departamento.
- Fornecem uma **resposta mais rápida** por conter menos informações.
- São mais fáceis de entender e navegar.
- Apóiam o controle local em vez de central, conferindo poder ao grupo de usuários.

Como fazer para que um Data warehouse atenda a empresa?

Como fazer para que um Data warehouse atenda a empresa

- Os primeiros passos para a construção de um data warehouse envolvem a execução de uma análise completa dos requisitos de informação que poderiam ser satisfeitos. Estas são algumas perguntas que os gerentes devem fazer:
 - **Usuários:** quais são os usuários primários do data warehouse? Quais são os níveis da organização e funções empresariais que representam?
 - **Propriedade:** que grupo (ou grupos) organizacional é o proprietário dos dados? Quem é responsável pela manutenção deles? Quem está autorizado a acessar os dados?

- Estas são algumas perguntas que os gerentes devem fazer (continuação):
 - **Requisitos de informação:** que tipos de relatórios e consultas o data warehouse deve suportar? Quais são os dados utilizados nesses relatórios e consultas? Os relatórios requisitam dados detalhados e resumidos?
 - **Fontes de dados:** quais são as fontes de dados requeridas pelos relatórios? Quais dados vêm de sistemas internos de transação e de outros sistemas? Quais vêm de fontes externas à empresa? Como a empresa pode obter os dados para esses relatórios?

- Estas são algumas perguntas que os gerentes devem fazer (continuação):
 - **Atualidade:** Com que frequência os dados de um data warehouse precisam ser atualizados? Por quanto tempo os dados históricos devem ser mantidos no armazém?
 - **Padrão dos dados:** Aplicativos elaborados para dar suporte a diferentes funções ou unidades organizacionais podem usar o mesmo **termo de diferentes maneiras**. Essas **discrepâncias** devem ser identificadas de modo que cada elemento de dado seja definido e utilizado do mesmo modo no sistema de data warehouse. **Todos concordam com a definição de cada dado e com o modo que será utilizado?**

- Estas são algumas perguntas que os gerentes devem fazer (continuação):
 - **Expectativas de qualidade:** que nível de precisão e integridade dos dados do data warehouse é suficiente para atender as necessidades da empresa?
 - **Benefícios:** exatamente que benefícios a construção desse data warehouse traz para a empresa? Até que ponto esses benefícios podem ser quantificados? Eles são maiores do que os custos?

- Estas são algumas perguntas que os gerentes devem fazer (continuação):
 - **Mudança do processo de negócios:** a empresa precisa mudar seus processos de negócio para poder utilizar efetivamente as informações contidas no data warehouse. Que grau de mudança será necessário?

Processamento analítico online

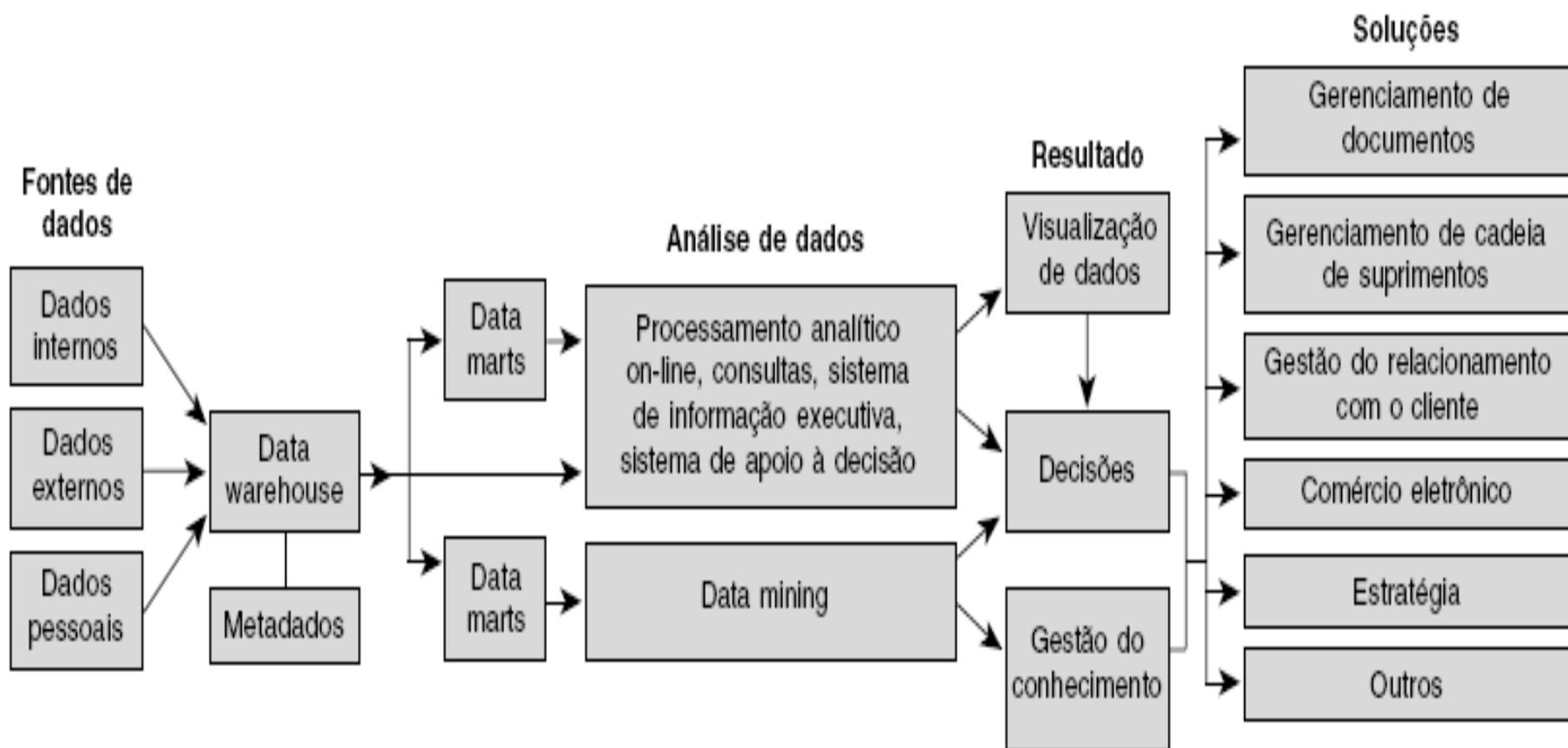
OLAP – On line Analytical Processing

O **processamento analítico online** é a capacidade dos sistemas de informação gerencial, de apoio à decisão e de informação executiva que permite aos gerentes e analistas **examinarem e manipularem interativamente enormes quantidades de dados detalhados e consolidados, a partir de múltiplas perspectivas.**

- OLAP apresenta informações para usuários via um modelo de dados natural e intuitivo.
- Através de um simples estilo de navegação e pesquisa, usuários finais podem rapidamente analisar inúmeros cenários, gerar relatórios "ad-hoc", e descobrir tendências e fatos relevantes independente do tamanho, complexidade, e fonte dos dados corporativos.

O processamento analítico online envolve diversas **operações analíticas básicas**, incluindo:

- **Consolidação** – envolve a agregação de dados. Isto pode envolver simples anexações ou agrupamentos complexos envolvendo dados interrelacionados.
- **Drill-Down** – o OLAP pode seguir na direção inversa e automaticamente exibir os dados detalhados que compõem os dados consolidados.
- **Slicing and Dicing** (“fatiar em cubos”) – referem à possibilidade de considerar os bancos de dados a partir de diferentes pontos de vista. O slicing and dicing geralmente é executado ao longo de um eixo de tempo a fim de analisar tendências e descobrir padrões



- História do BI
 - <http://www.youtube.com/watch?v=GfL9JeVaJa4>