



# Banco de Dados 2

014 – Data Warehouse





# Data Warehouse

- Um ***data warehouse*** é um tipo de **sistema de gerenciamento de dados** projetado para ativar e fornecer ***suporte às atividades de business intelligence (BI)***, especialmente a **análise avançada**.
- Os ***data warehouses*** destinam-se exclusivamente a realizar **consultas e análises avançadas** e geralmente contêm **grandes quantidades de dados históricos**.
- Os **dados** em um ***data warehouse*** geralmente **são derivados de uma ampla variedade de fontes**, como **arquivos de log de aplicativos e aplicativos de transações**.



# Data Warehouse

- Um ***data warehouse*** **centraliza** e **consolida** grandes quantidades de dados de várias fontes.
- Seus **recursos analíticos** permitem que as organizações obtenham **informações de negócios** úteis de seus dados para melhorar a **tomada de decisões**.
- Com o tempo, cria-se um **registro histórico** que pode ser inestimável para **cientistas de dados e analistas de negócios**.



# Data Warehouse

- Um ***data warehouse*** típico geralmente inclui os seguintes elementos:
  - Um **banco de dados relacional** para armazenar e gerenciar dados.
  - Uma **solução de extração, carregamento e transformação (ELT)** para preparar os dados para análise
  - **Análise estatística, relatórios** e recursos de **mineração de dados**
  - **Ferramentas de análise de clientes** para **visualizar e apresentar dados** aos usuários de negócios
  - Outras **aplicações analíticas** mais sofisticadas que geram informações úteis por meio de **aplicação de algoritmos de machine learning** e **inteligência artificial (IA)**



# Data Warehouse

## OLTP e OLAP

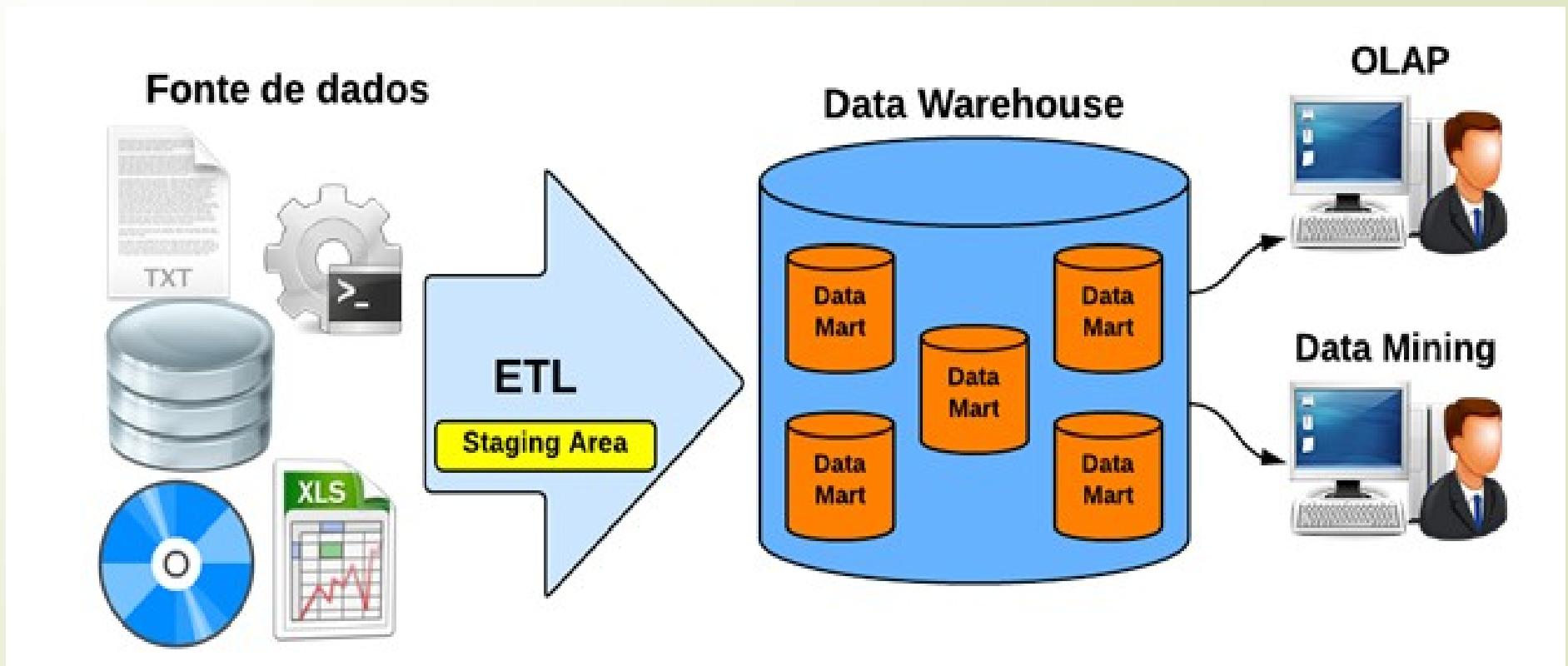
- O **OLTP**, do inglês "*On-line Transaction Processing*", é o termo usado para se referir aos **sistemas transacionais**, ou seja, os **sistemas operacionais das organizações**. São **utilizados no processamento dos dados de rotina** que são gerados diariamente através dos **sistemas informacionais da empresa** e dão **suporte às funções de execução do negócio organizacional**.
- Já o **OLAP**, do inglês "*On-line Analytical Processing*", trata da **capacidade de analisar grandes volumes de informações** nas mais **diversas perspectivas** dentro de um **Data Warehouse (DW)**. O **OLAP** também faz referência às **ferramentas analíticas utilizadas no BI** para a **visualização das informações gerenciais** e dá suporte para as funções de análises do negócio organizacional.

Os sistemas OLTP e OLAP se diferenciam em diversos outros aspectos. Vejamos:

	<b>OLAP</b>	<b>OLTP</b>
Foco	Foco no nível estratégico da organização. Visa a análise empresarial e tomada de decisão.	Foco no nível operacional da organização. Visa a execução operacional do negócio.
Performance	Otimização para a leitura e geração de análises e relatórios gerenciais.	Alta velocidade na manipulação de dados operacionais, porém ineficiente para geração de análises gerenciais.
Estrutura dos dados	Os dados estão estruturados na modelagem dimensional. Os dados normalmente possuem alto nível de sumarização.	Os dados são normalmente estruturados em um modelo relacional normalizado, otimizado para a utilização transacional. Os dados possuem alto nível de detalhes.
Armazenamento	O armazenamento é feito em estruturas de <i>Data Warehouse</i> com otimização no desempenho em grandes volumes de dados.	O armazenamento é feito em sistemas convencionais de banco de dados através dos sistemas de informações da organização.
Abrangência	É utilizado pelos gestores e analistas para a tomada de decisão.	É utilizado por técnicos e analistas e engloba vários usuários da organização.
Frequência de atualização	A atualização das informações é feita no processo de carga dos dados. Frequência baixa, podendo ser diária, semanal, mensal ou anual (ou critério específico).	A atualização dos dados é feita no momento da transação. Frequência muito alta de atualizações.
Volatilidade	Dados históricos e não voláteis. Os dados não sofrem alterações, salvo necessidades específicas (por motivos de erros ou inconsistências de informações).	Dados voláteis, passíveis de modificação e exclusão.
Tipos de permissões nos dados	É permitido apenas a inserção e leitura. Sendo que para o usuário está apenas disponível a leitura.	Podem ser feito leitura, inserção, modificação e exclusão dos dados.

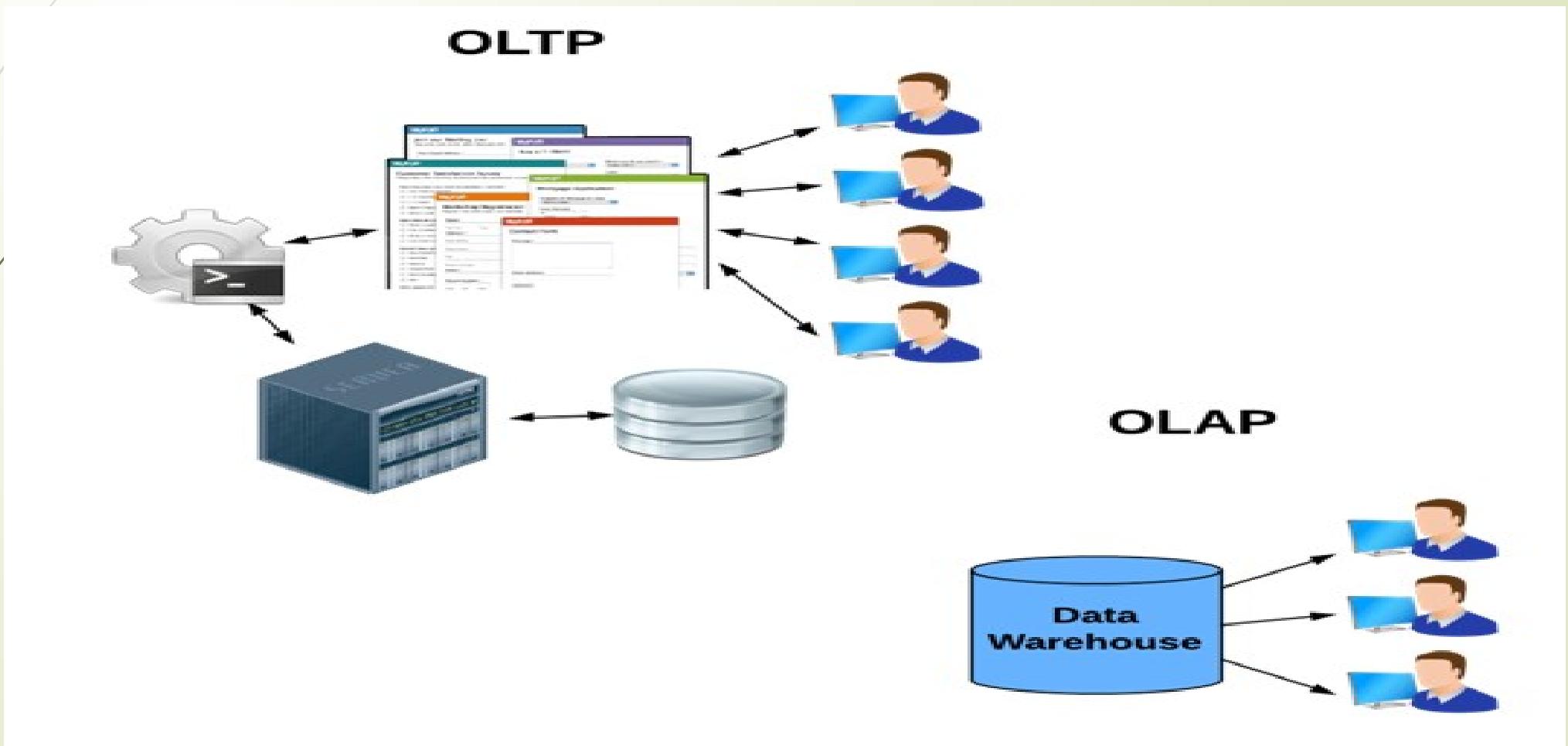
# Data Warehouse

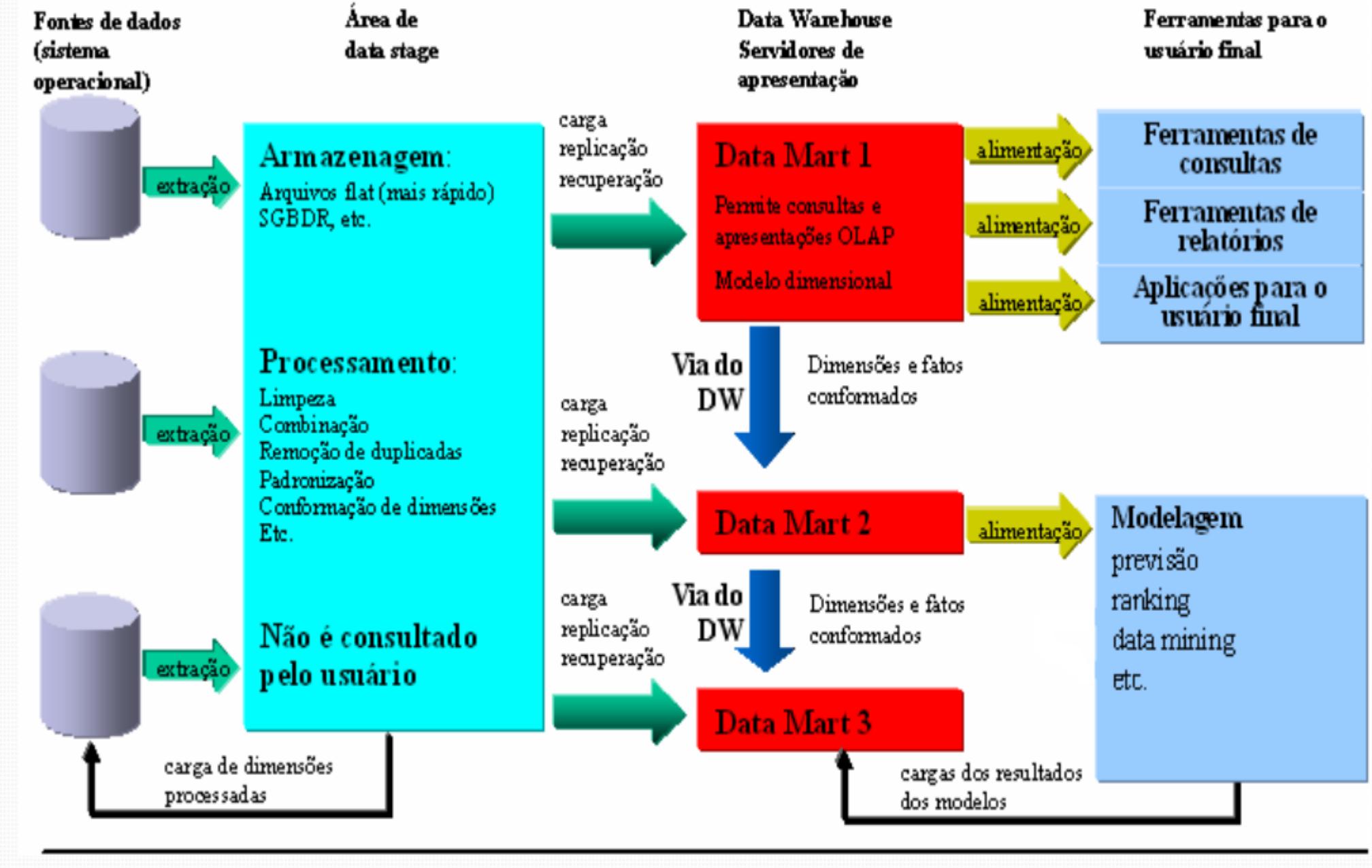
## OLTP e OLAP



# Data Warehouse

## OLTP e OLAP





# Data Warehouse

## Fonte de dados:

- **Sistemas transacionais** da empresa, pode ser composto por diversas formas de dados.
- Os **dados** podem ser **extraídos** de **bancos de dados** que suportam os **sistemas transacionais** (também chamados de **sistemas operacionais**).
- Não confundir com o **tradicional conceito de sistema operacional** (Linux, Unix, Windows etc.). São chamados **operacionais** por suportarem o **controle das operações das organizações que os utilizam** (Sistemas Financeiro, de Recursos Humanos, Vendas, Controle de Estoque etc.).

# Data Warehouse

## □ Data Stage:

- Composta por uma **área de armazenagem** e um **conjunto de processos**.
- Sua função é **extrair os dados dos sistemas transacionais**, proceder a limpeza, a transformação, combinação, de duplicação e **preparação dos dados** para o **uso** no *Data Warehouse*.
- Estes dados **não são apresentados** ao usuário final.

# Data Warehouse

- **Servidor de apresentação:**
  - Ambiente onde os dados são organizados e armazenados para consulta direta pelos usuários finais.
  - Normalmente os dados estão **disponíveis** nestes **servidores** em **bancos de dados relacionais**, mas também podem estar armazenados em tecnologia **OLAP** (*OnLine Analytical Processing*) já que muitos ***data marts*** trabalham apenas com **dados** no modelo dimensional.



# Data Warehouse

- **Data Mart**

- Subconjunto lógico do *Data Warehouse*, geralmente divididos por departamento ou visões necessárias para os usuários.

# Data Warehouse

- **Data Mining:**
- Também conhecido como **mineração de dados**, o *Data Mining* trabalha em **grandes massas de dados** onde existem **muitas correlações entre os dados que não são perceptíveis facilmente**.
- Os *Data Warehouses* são constituídos, normalmente, de **imensa quantidade de dados**, há necessidade de uma **ferramenta para varrer automaticamente o Data Warehouse** a fim de **pesquisar tendências e padrões através de regras pré-definidas** que dificilmente seriam encontrados em uma **pesquisa comum**.

# Data Warehouse

- **Ferramentas de acesso aos dados:**
- Maneira em que os **dados** são **extraídos** e **integrados** com cada processo distinto do *Data Warehouse*. As **funções para a transformação dos dados** são:
  - **Extração:** retirada dos dados dos sistemas transacionais e armazenagem na área de ***data stage***;
  - **Carga de dimensões processadas:** realimentação do processo para garantir a representação correta dos dados em novo formato.

# Data Warehouse

- **Carga, Replicação e Recuperação:** quando pronto, o dado é carregado no ***data mart*** correspondente e são criados (ou atualizados) índices para melhorar a performance das consultas.
- **Alimentação:** apresenta as visões do ***data mart*** de acordo com as **necessidades dos usuários**.
- **Carga dos resultados dos modelos:** serve para realimentar possíveis modificações no ***data mart***, caso este não esteja adequado a aplicação que o utiliza.

# Dados, Informações e Conhecimentos

- **Dado**: é um valor relativo a descrever ...
  - um objeto de interesse,
  - uma atividade,
  - uma pessoa,
  - alguma coisa estática.
- Exemplo de **dado**:
  - Renda Mensal,
  - Despesas Mensais,
  - Pressão arterial.

# Dados, Informações e Conhecimentos

- A **informação** acrescenta algo mais que o simples dado.
- Exemplo de **Informação**:
  - **Capacidade Mensal de Endividamento** =  $100 \times (\text{Renda Mensal} - \text{Despesas Mensais}) / \text{Renda Mensal}$ .
  - Possui uma saúde cardíaca boa.
  - Vai ter um enfarto!

# Dados, Informações e Conhecimentos

- **Dados:** itens elementares, captados e armazenados por recursos da **Tecnologia da Informação (TI)**.
- **Informações:** representam dados processados, com significado e contextos bem definidos.
- **Conhecimento:** padrão ou conjunto de padrões cuja formulação pode envolver e relacionar dados e informações.

# Dados, Informações e Conhecimentos

- Exemplos de **Conhecimento**:
  - Regras de produção. Apresentam a forma **SE <condições> ENTÃO <conclusões>**.
  - **SE Capacidade Mensal de Endividamento > 60% ENTÃO Crédito = SIM.**
  - **SE valor máximo pressão arterial > 12 ENTÃO Risco de Infarto = SIM.**



# Histórico

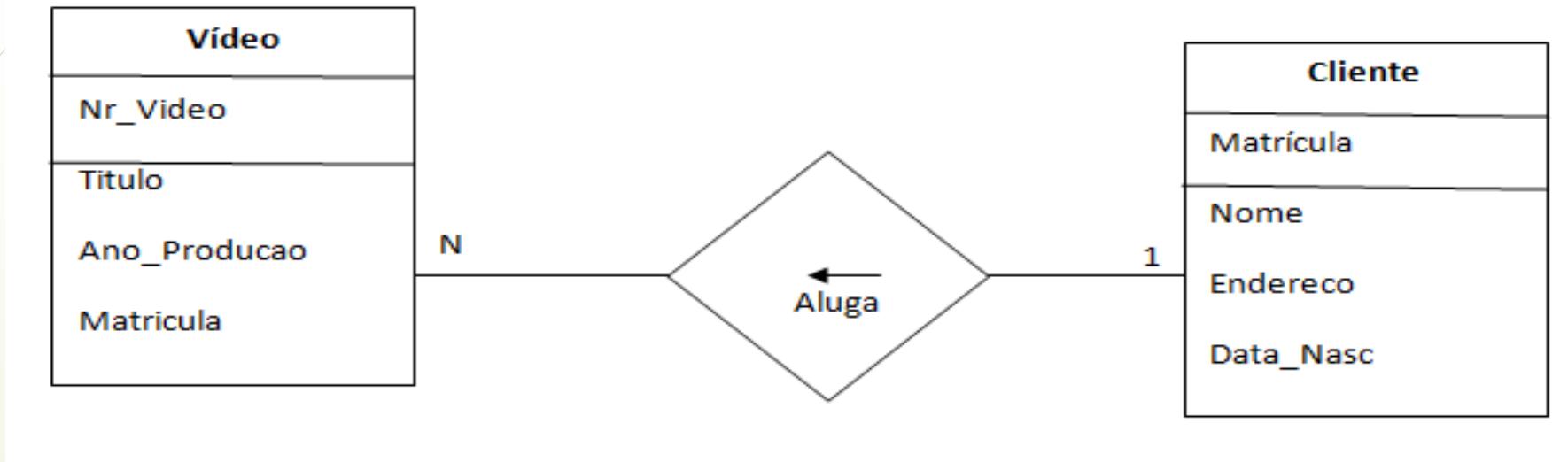
- Os primeiros **Sistemas de Informação (SI)** – em razão do **elevado custo dos dispositivos de armazenamento** – **não** mantinham **base de dados históricas** superiores ao **ano em curso** nos **meios de armazenamento e acesso rápido**.
- O objetivo era de **economizar espaço** em **meios magnéticos** de armazenamento (**disco**).



# Histórico

- Essa economia resultou em um **acesso mais difícil ao histórico das operações de negócios** das empresas.
  - **Como então tratar informações,** se normalmente elas são resultados de **processamento de séries históricas de dados?**
- 

# Histórico

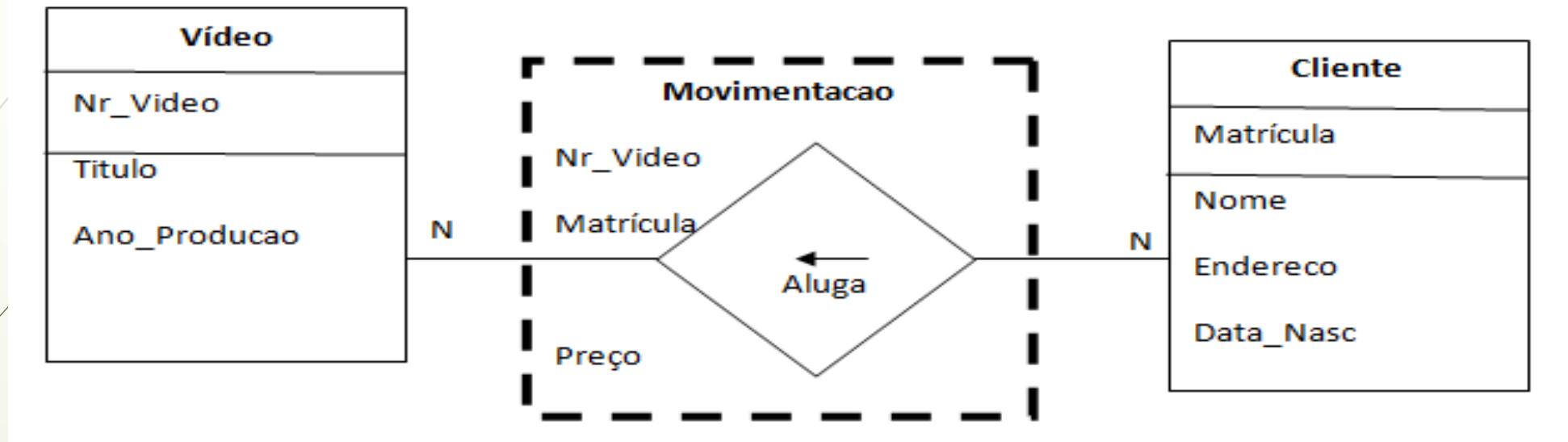


Trecho de um **SI Transacional** para uma **videolocadora** (VHS, DVD ou Blu-Ray).

**Modelagem** que **não** privilegia a manutenção de qualquer **histórico das locações** de vídeos pelos clientes. O sistema registra apenas a condição presente (atual).

A **informação** está restrita aos **vídeos que estão na locadora** (**Vídeo.Matrícula = NULO**) e Aos **vídeos que estão com os clientes no momento** (**Vídeo.Matrícula = Cliente.Matrícula**).

# Histórico



Trecho de um SI Transacional que permite o registro de um **histórico das locações feitas** Por cada **cliente** na **videolocadora**.

A tabela **Movimentação** armazena esse histórico.

Porém, os primeiros **SIs** armazenavam no máximo os dados correntes do último ano.

**Dados mais antigos** precisavam ser recuperados de *backups* (não estavam **on-line**).

# Histórico

- Nas **décadas** de 1980 e 1990, os sistemas denominados de *Executive Information System* (**EIS**).
- O **EIS** sempre foram **dirigidos para o período de tempo atual**, no gerenciamento de crises, do emergencial, do momento.
- **Não suportam os processos de análise estratégica.**

# Histórico

- ***Customer Relationship Management (CRM)*** que visa estabelecer um relacionamento diferenciado entre a empresa e seus clientes.
- “Tratar clientes diferentes de forma diferente”.
- Conhecer as preferências de cada cliente, seus dados pessoais, sua moeda de negócios etc.

# Histórico

- Instrumentos operacionais do **CRM**:
  - **Web.**
  - **Call Center**: atendimento a clientes.
  - **Automação de força de vendas**: informação por meio dos vendedores de forma automatizada e conectividade.
  - **Telemarketing**: contato com o cliente em busca de negócios.
  - **Gerenciamento de campanhas de marketing.**

# Histórico

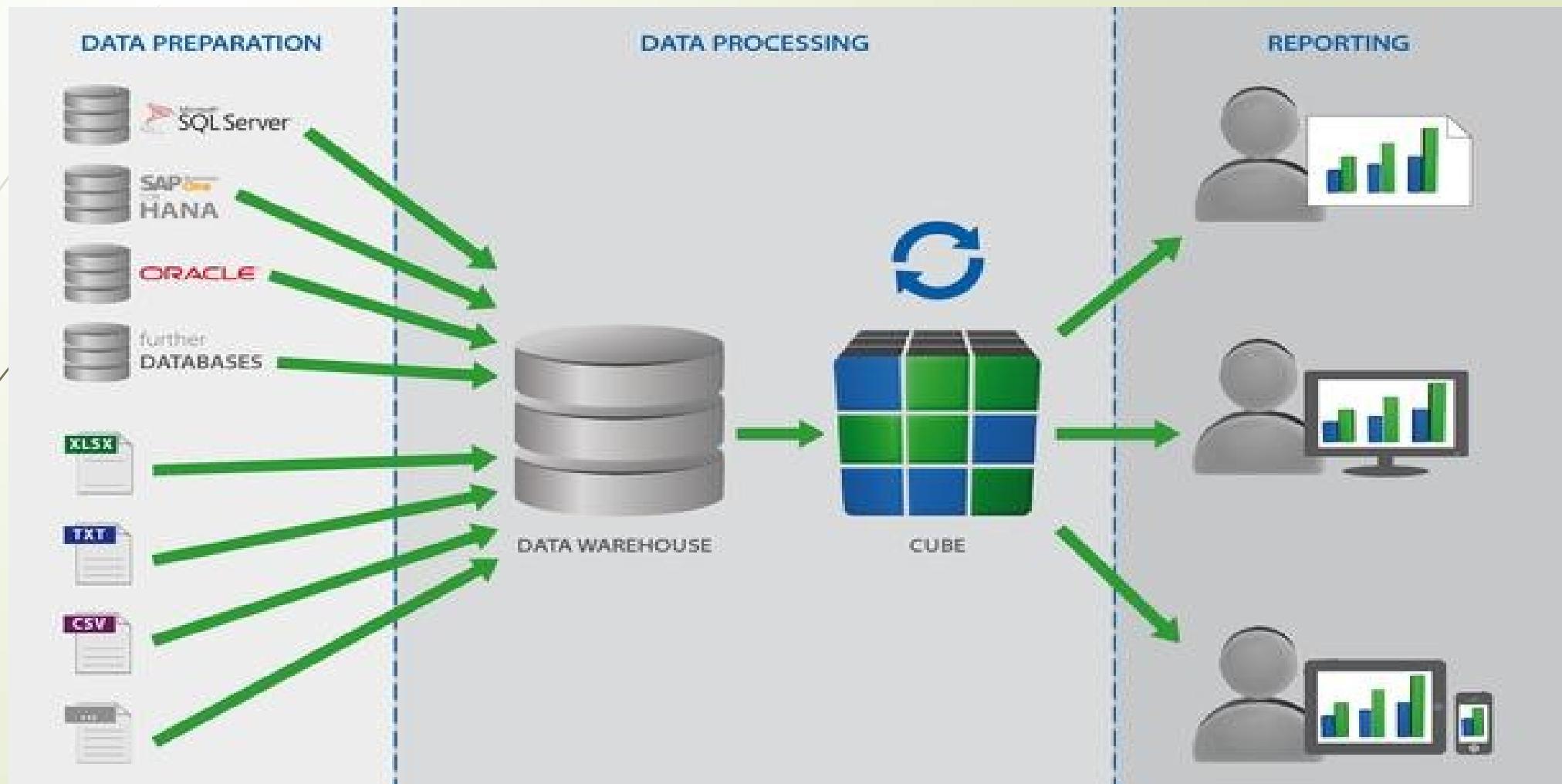
- O resultado do **CRM operacional** deve permitir a criação de **bases de informação** para a realização do que podemos denominar de **CRM analítico**.
- Um **armazém de dados** para permitir as seguintes análises:
  - **Segmentação de clientes.**
  - **Análise de campanha.**
  - **Análise de vendas.**
  - **Análise de fidelidade.**



# Histórico

- Lucratividade.
- Desempenho de negócios.
- Análise de atendimento ao cliente.
- O **CRM** é uma busca para:
  - Obter novos clientes.
  - Manter velhos clientes.
  - Desenvolver clientes.

# Histórico



# Histórico

- **Data Warehouse:** armazéns de dados onde a história da empresa, seus clientes, fornecedores e operações se mantenham disponíveis e acessíveis para **consultas** e **análises**.
- Para que se possibilitassem as **consultas**, surgiram **ferramentas** de conceito **OLAP (On Line Analytical Processing)**.
- Os *data warehouse* surgiram como conceito acadêmico na década de 1980.





# Características Gerais de Data Warehouse

- O **OLAP** permite:
  - A análise dos dados e descoberta das informações antes inimagináveis,
  - Disponibilizar a visualização dos dados sob diversas perspectivas,
  - A capacidade de navegar no nível de detalhe da informação.



# Características Gerais de Data Warehouse

- Não se pode projetar um novo ambiente de **Data Warehouse** pensando em **consistência entre dados** que entram; **integridade referencial** já não é mais a nossa maior preocupação.
- Em primeiro lugar devemos **recuperar os dados históricos da empresa**.
- Estes dados devem ser buscados e copiados das **bases de dados operacionais (transacionais)**.



# Características Gerais de Data Warehouse

- Devemos descobrir as **necessidades de informação, indicadores de negócio** da empresa.
- O resultado de um projeto de **Data Warehouse (DW)** é:
  - Informação disponível para gestão.
  - Visão de curvas de comportamento.
  - Agilidade de ferramentas para apoio à decisão.
  - Segurança de informações para decisão.
  - Maior abrangência de visão de **indicadores**.

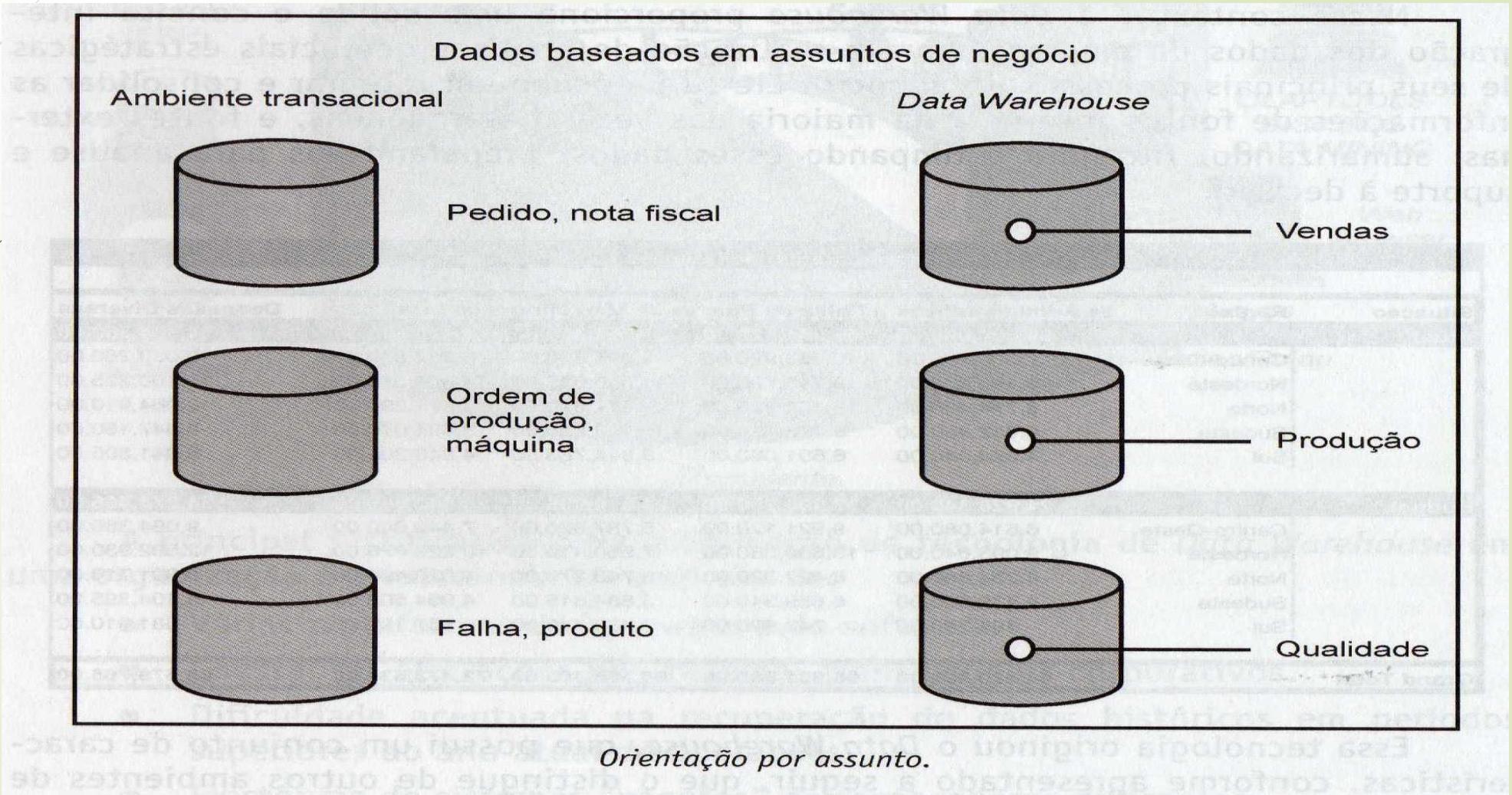


# Características Gerais de Data Warehouse

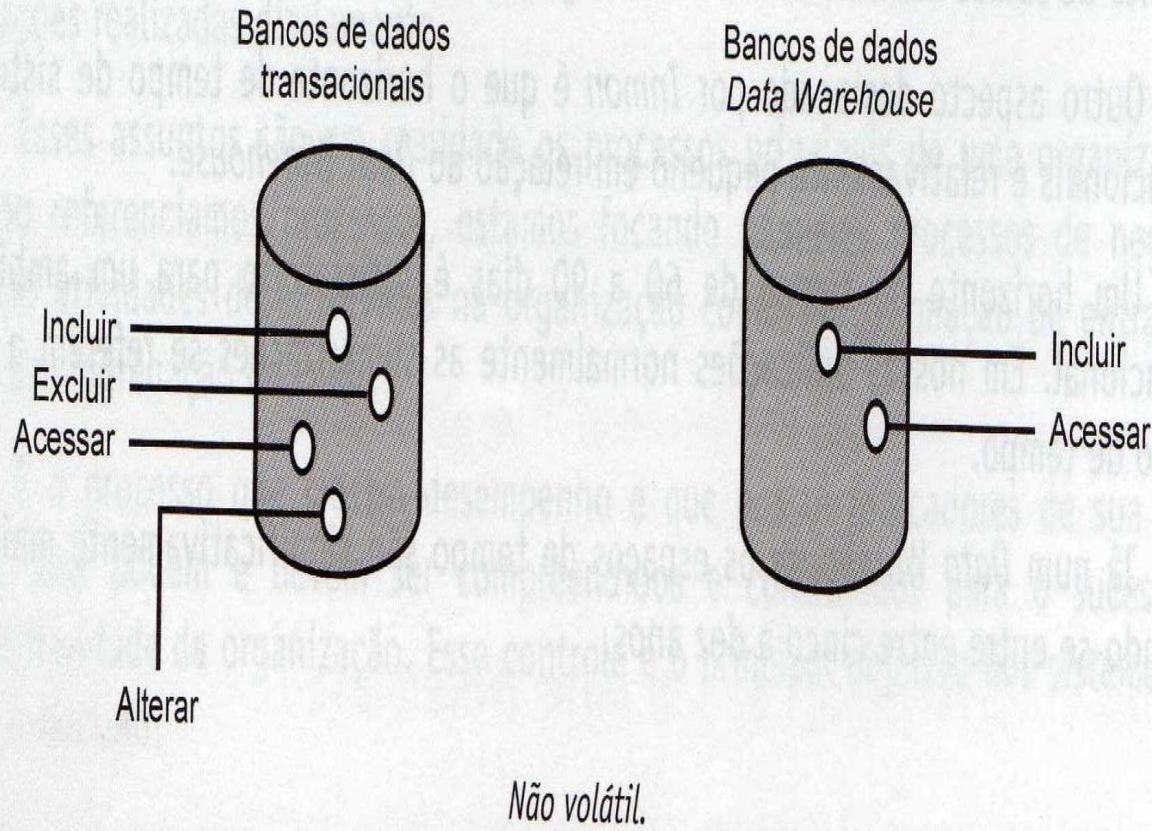
- Recursos mais abrangentes para análise de negócios.
- Necessidades de expectativas executivas atendidas por tecnologia da informação (TI).
- **Modelagem Multidimensional:** é uma técnica estruturada desenvolvida para obtenção de **modelos de dados de simples entendimento e alta performance de acesso aos dados.**

<b>Características</b>	<b>Bancos de dados Operacionais</b>	<b>Data Warehouse</b>
<b>Objetivo</b>	Operações diárias do negócio	Analisar o negócio
<b>Uso</b>	Operacional	Informativo
<b>Tipo de processamento</b>	OLTP	OLAP
<b>Unidade de trabalho</b>	Inclusão, alteração, exclusão.	Carga e consulta
<b>Número de usuários</b>	Milhares	Centenas
<b>Tipo de usuário</b>	Operadores	Comunidade gerencial
<b>Interação do usuário</b>	Somente pré-definida	Pré-definida e ad-hoc
<b>Condições dos dados</b>	Dados operacionais	Dados Analíticos
<b>Volume</b>	Megabytes – gigabytes	Gigabytes – terabytes
<b>Histórico</b>	60 a 90 dias	5 a 10 anos
<b>Granularidade</b>	Detalhados	Detalhados e resumidos
<b>Redundância</b>	Não ocorre	Ocorre
<b>Estrutura</b>	Estática	Variável
<b>Manutenção desejada</b>	Mínima	Constante
<b>Acesso a registros</b>	Dezenas	Milhares
<b>Atualização</b>	Contínua (tempo real)	Periódica (em batch)
<b>Integridade</b>	Transação	A cada atualização
<b>Número de índices</b>	Poucos/simples	Muitos/complexos
<b>Intenção dos índices</b>	Localizar um registro	Aperfeiçoar consultas

# Características Gerais de Data Warehouse



# Características Gerais de Data Warehouse



Depois de **carregado**, um **DW** somente possui operações de **consulta**.

No **DW** os **dados** são, portanto, **Não Voláteis**.

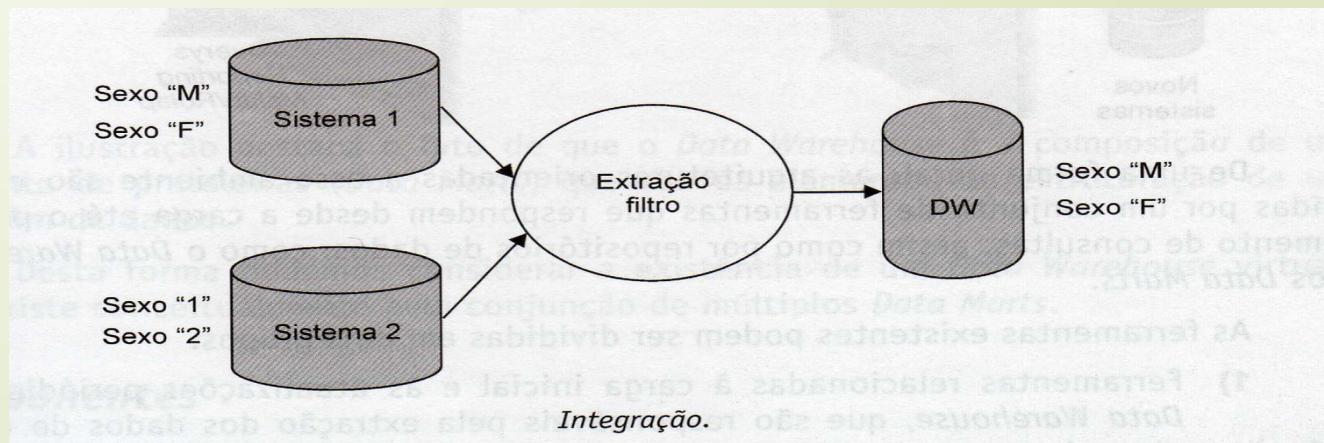


# Características Gerais de Data Warehouse

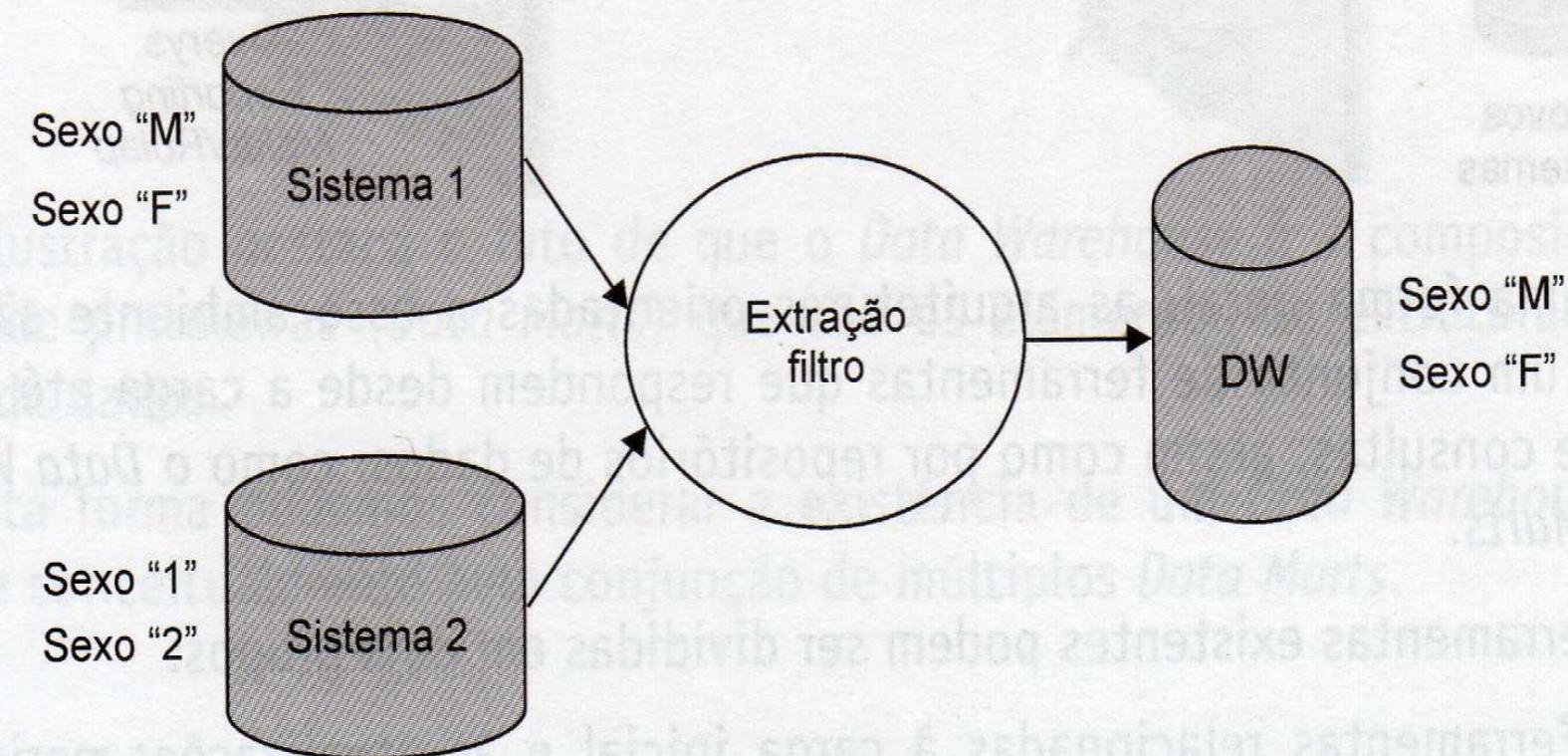
- Os **dados operacionais** de um **sistema transacional** podem ou não fazer uso de **datas** – conforme sua conveniência.
- Em um **DW**, porém, a **data** é um **elemento essencial e componente-chave**.
- Enquanto o **horizonte de tempo** em **sistemas transacionais** situa-se entre **60** e **90** dias (é curto), em um **DW** ele fica entre **5** e **10** **anos** (é mais longo).

# Características Gerais de Data Warehouse

- Em um DW todos os dados têm um alto nível de integração.
- As convenções de nomes, valores de variáveis, tais como sexo masculino e feminino, e outros atributos físicos de dados como *data types* são formalmente unificados e integrados nessa base única.



# Características Gerais de Data Warehouse



*Integração.*



# Características Gerais de Data Warehouse

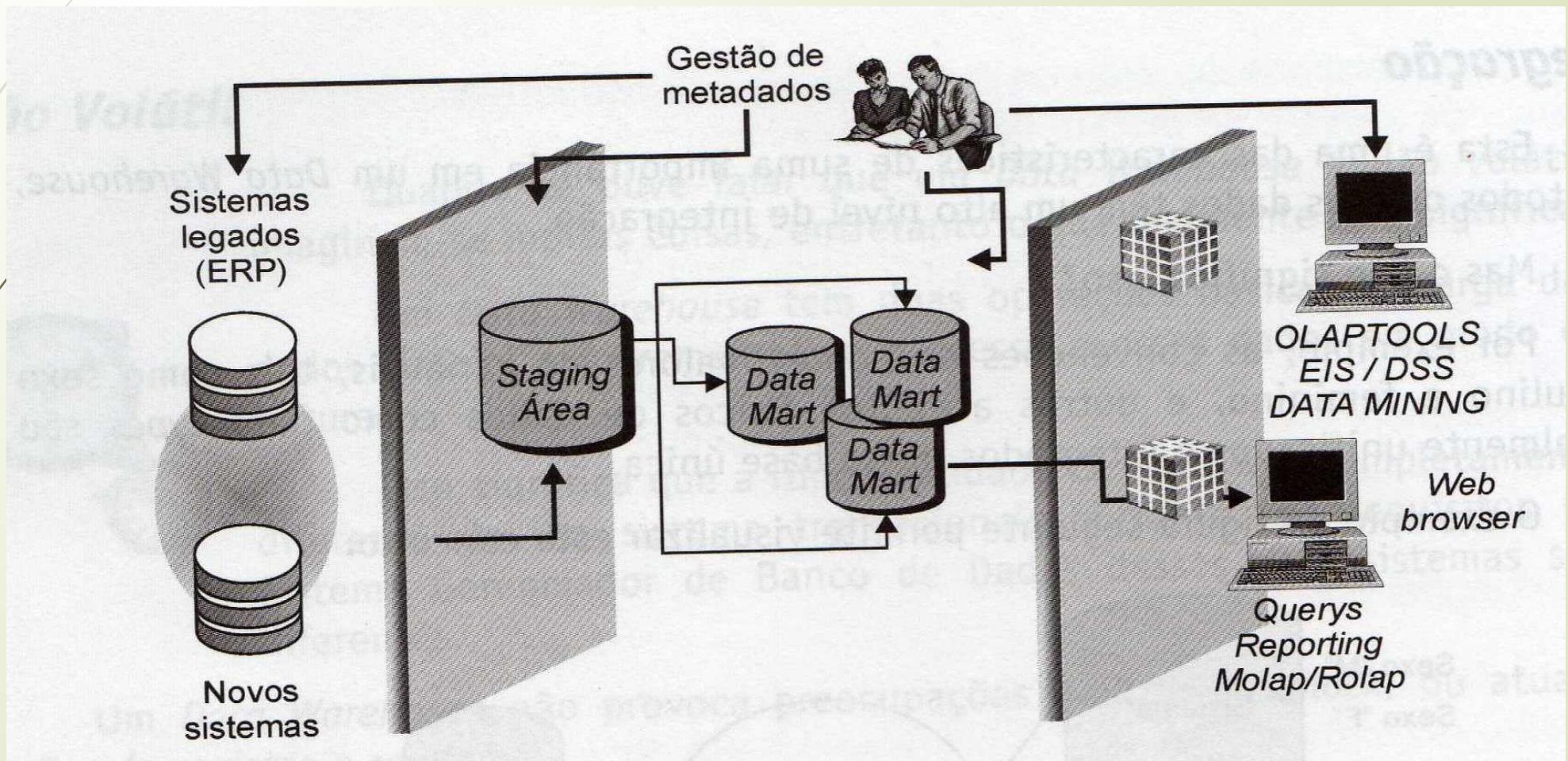
- Em **ambientes de múltiplas plataformas sistêmicas**, a característica de **integração** se torna fundamental, pois necessitamos de **unicidade de informações**.
- A existência de **sistemas mais antigos** com **padrões de codificação de dados variados** exige – durante a **carga dos dados** pelo **DW** – por um **processo de filtragem**.



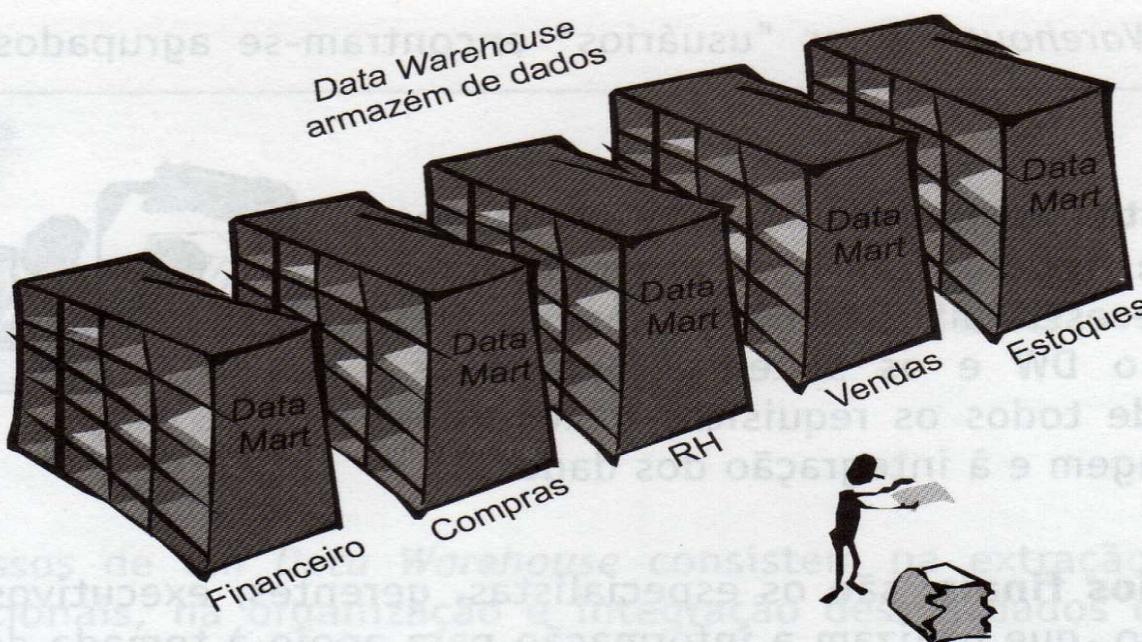
# Características Gerais de Data Warehouse

- O DW funciona como um grande conjunto de todos os dados.
- Os *Data Marts* (DMs) são as **prateleiras** desse **armazém de dados**, que permitem uma **visão mais direcionada de um problema**.
- Os DMs funcionam como **repositórios menores orientados a áreas específicas**.

# Características Gerais de Data Warehouse

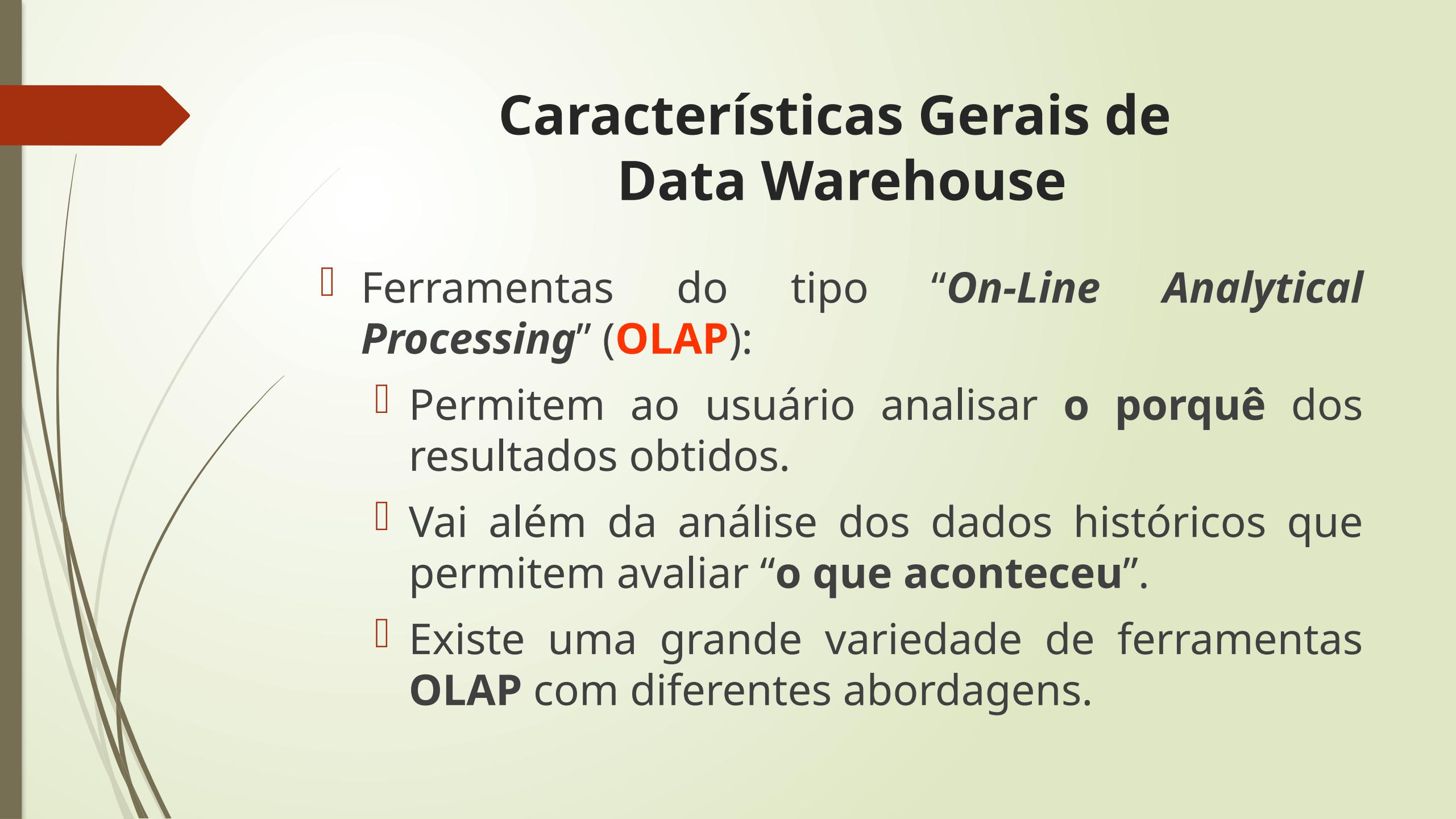


# Características Gerais de Data Warehouse



A ilustração destaca o fato de que o *Data Warehouse* é a composição de um conjunto de prateleiras (*Data Marts*) que são os elementos de estruturação de um armazém de dados.

Desta forma podemos considerar a existência de um *Data Warehouse* virtual, que existe conceitualmente pela conjunção de múltiplos *Data Marts*.

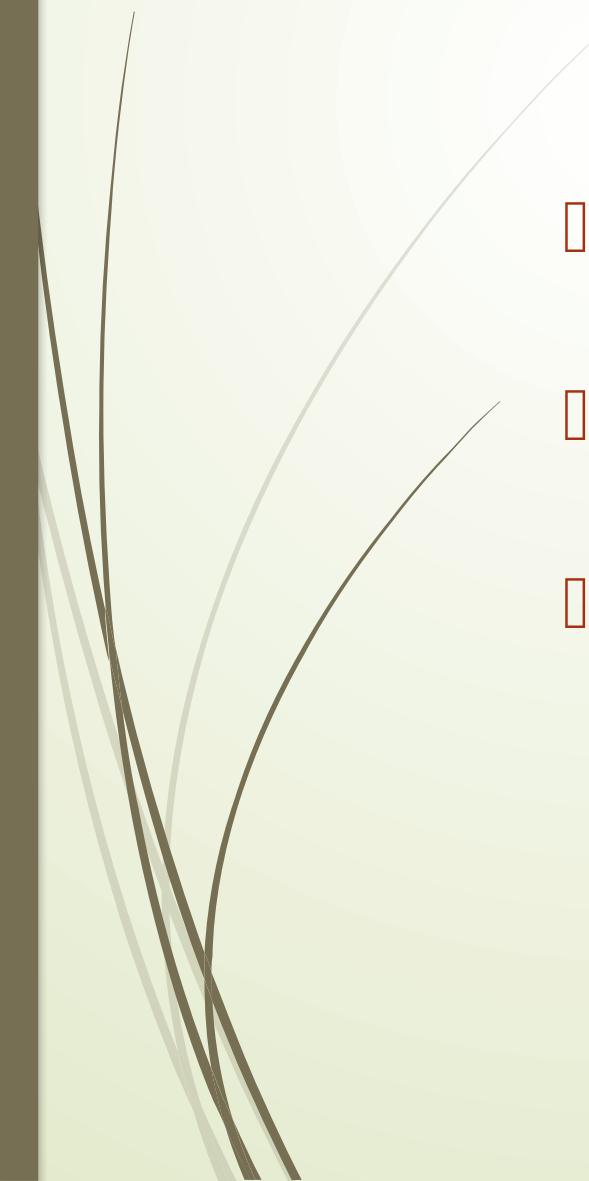


# Características Gerais de Data Warehouse

- Ferramentas do tipo "*On-Line Analytical Processing*" (**OLAP**):
  - Permitem ao usuário analisar **o porquê** dos resultados obtidos.
  - Vai além da análise dos dados históricos que permitem avaliar "**o que aconteceu**".
  - Existe uma grande variedade de ferramentas **OLAP** com diferentes abordagens.

# Características Gerais de Data Warehouse

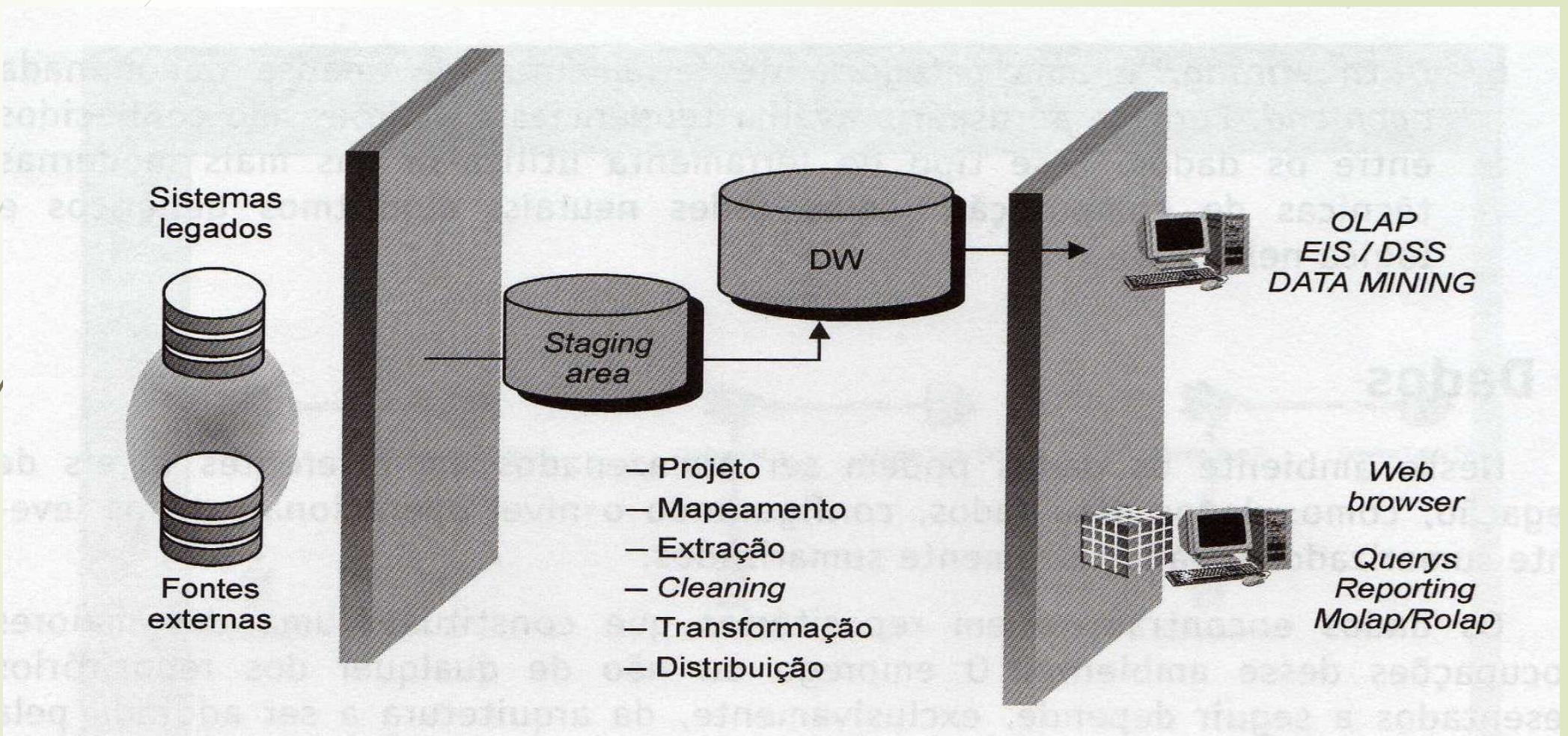
- **ROLAP (OLAP Relacional)**: ferramentas **OLAP** que acessam bancos de dados relacionais.
- **MOLAP (OLAP Multidimensional)**: ferramentas **OLAP** que acessam bancos de dados multidimensionais por meio de cubos e hipercubos.
- **HOLAP (OLAP Híbrida)**: ferramentas **OLAP** que permitem acesso tanto aos bancos de dados relacionais como aos multidimensionais.
- **DOLAP (OLAP Desktop)**: ferramentas **OLAP** voltadas para computadores pessoais – geralmente empregadas em **BDs** individuais para análises mais específicas do que as relaizadas no **DM**. Os dados normalmente são carregados a partir de um **DM**.

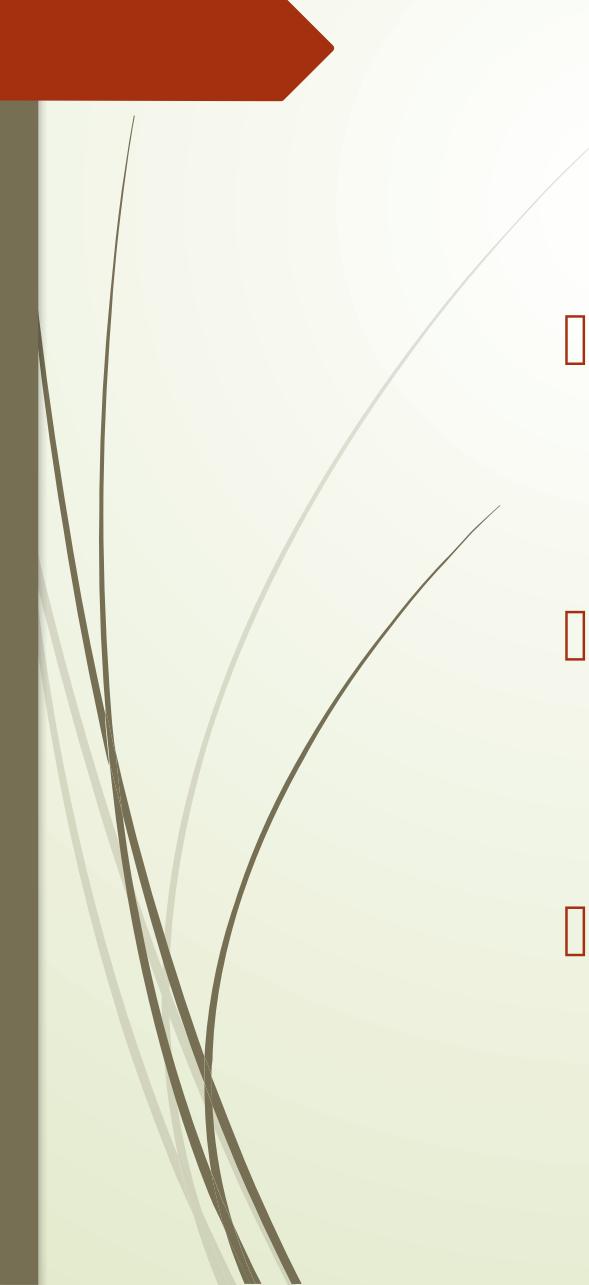


# Características Gerais de Data Warehouse

- Staging Area também é conhecido como **Operational Data Storage (ODS)**.
- Representa um **armazenamento intermediário** dos **dados** – facilitando a **integração** destes.
- Originalmente era um **repositório temporário** – uma **cópia** dos ambientes de **sistemas transacionais** existentes na empresa.

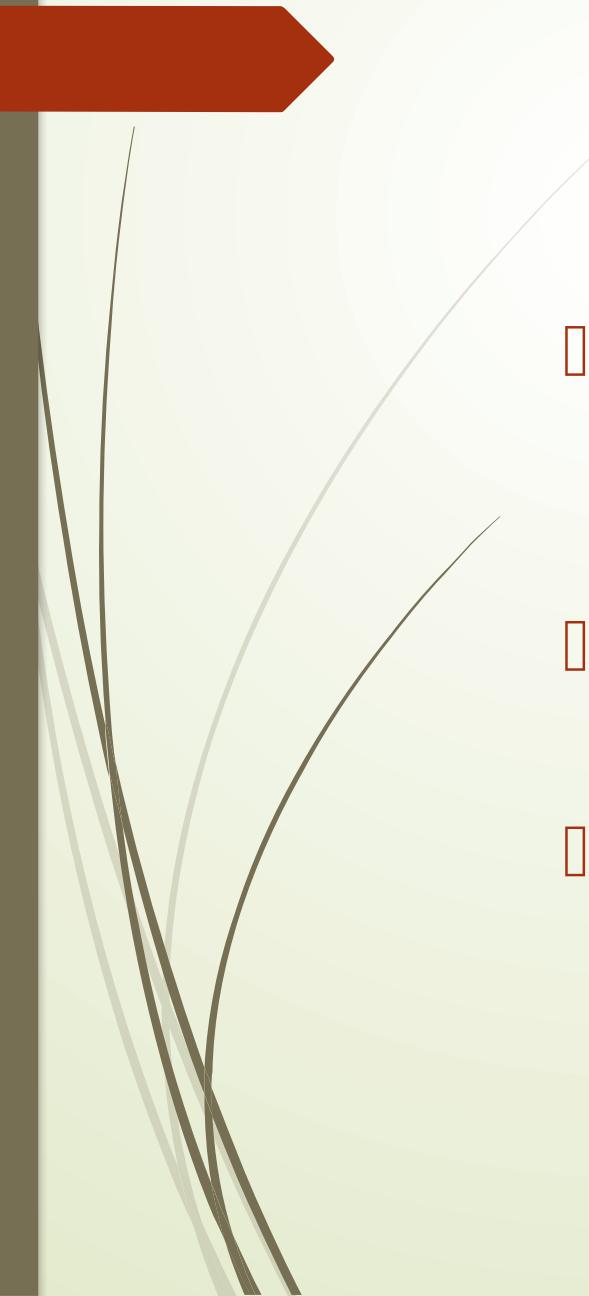
# Características Gerais de Data Warehouse





# Características Gerais de Data Warehouse

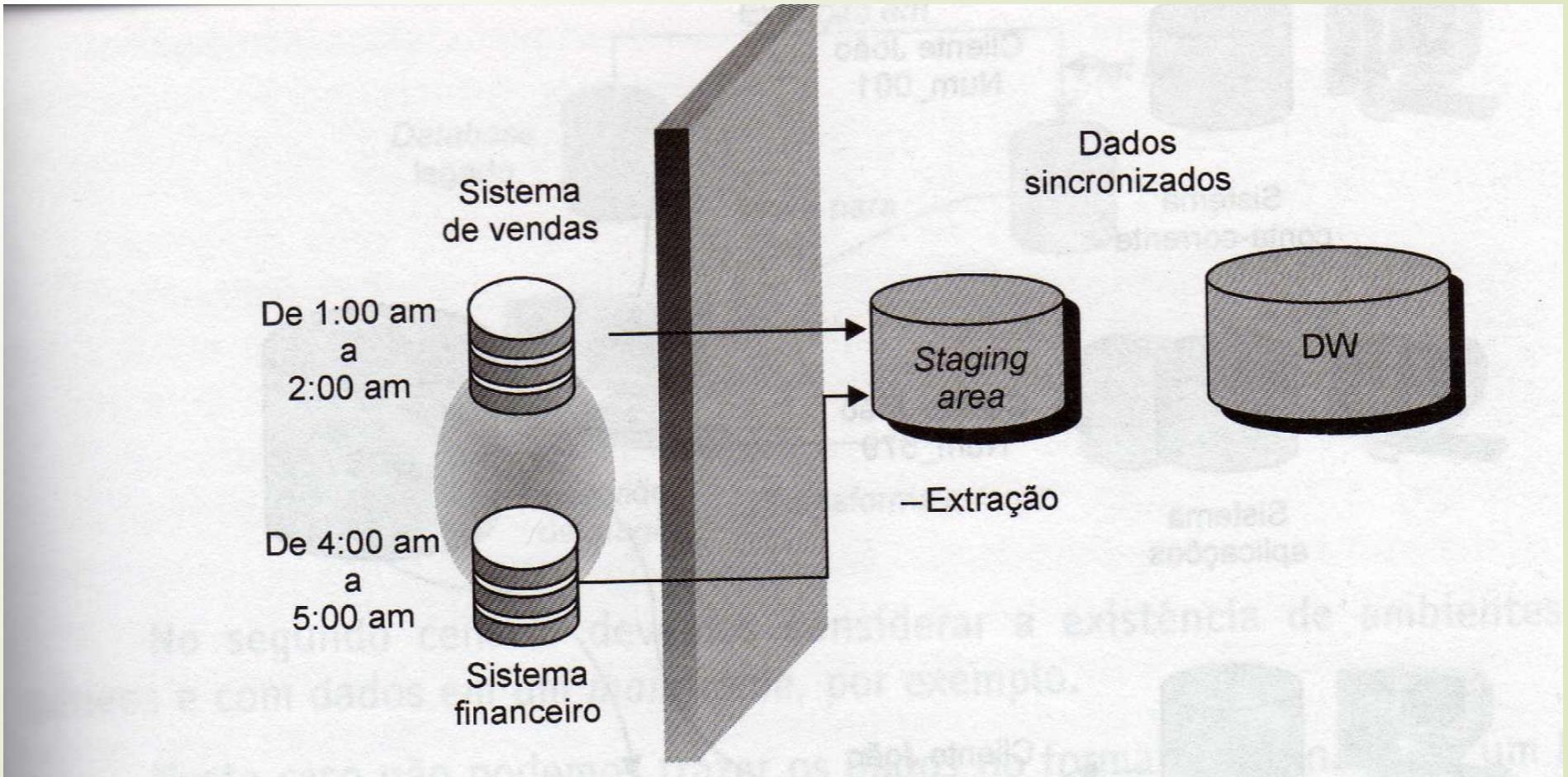
- Uma *staging area* permite ao administrador do **DW** extrair os dados no momento em que estão disponíveis e posteriormente integrá-los.
- Isso facilita as **extrações** dos **sistemas operacionais** durante períodos **fora de pico de operações**.
- **Exemplo:** os **dados de vendas** podem estar **disponíveis** para **extração** apenas **entre 1:00 a.m. e 2:00 a.m.**



# Características Gerais de Data Warehouse

- O **processamento das vendas** é **terminado** e os dados estão em um **estado estável** e **sincronizado**.
- Já os **dados financeiros** estão **disponíveis** apenas entre **4:00 a.m.** e **5:00 a.m.**.
- Uma ***staging area*** permite que o processo de **conversão dos dados** seja **separado** do processo de **transformação**.

# Características Gerais de Data Warehouse

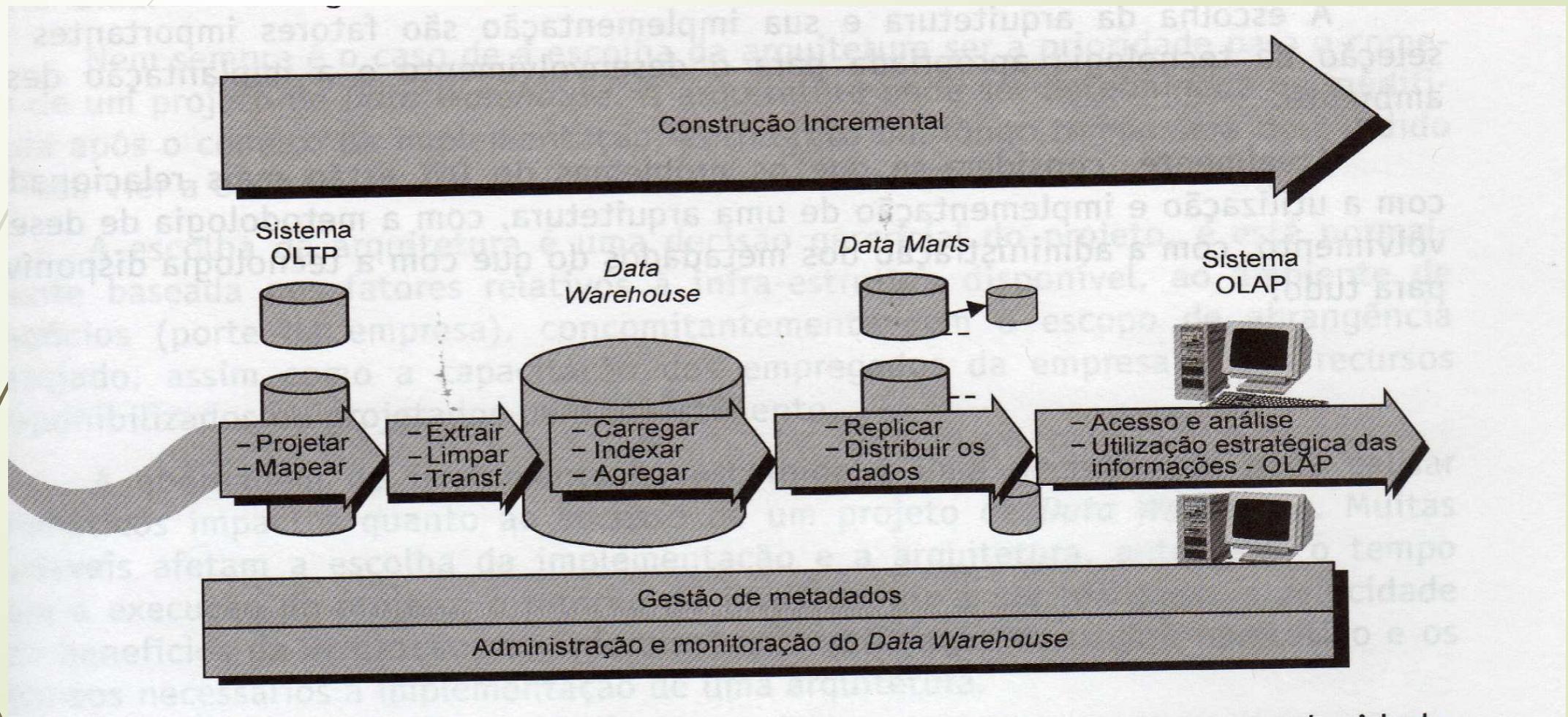




# Características Gerais de Data Warehouse

- Objetivo da *Staging Area*:
- Criar um ambiente intermediário de armazenamento e processamento dos dados oriundos de aplicações **OLTP** e outras fontes, para o processo de extração, transformação e carga (**ETL**), possibilitando o seu tratamento, e permitindo sua posterior integração em formato e no tempo, evitando problema após a criação do DW e a concorrência com o ambiente transacional no consumo de recursos.

# Características Gerais de Data Warehouse







# Granularidade de Dados

- A **granularidade** de dados refere-se ao **nível de summarização** dos elementos e de **detalhe** disponíveis nos **dados**.
- Quanto **mais detalhe** existir, mais **baixo** será o **nível de granularidade**.
- Quanto **menos detalhe** existir, mais **alto** será o **nível de granularidade**.



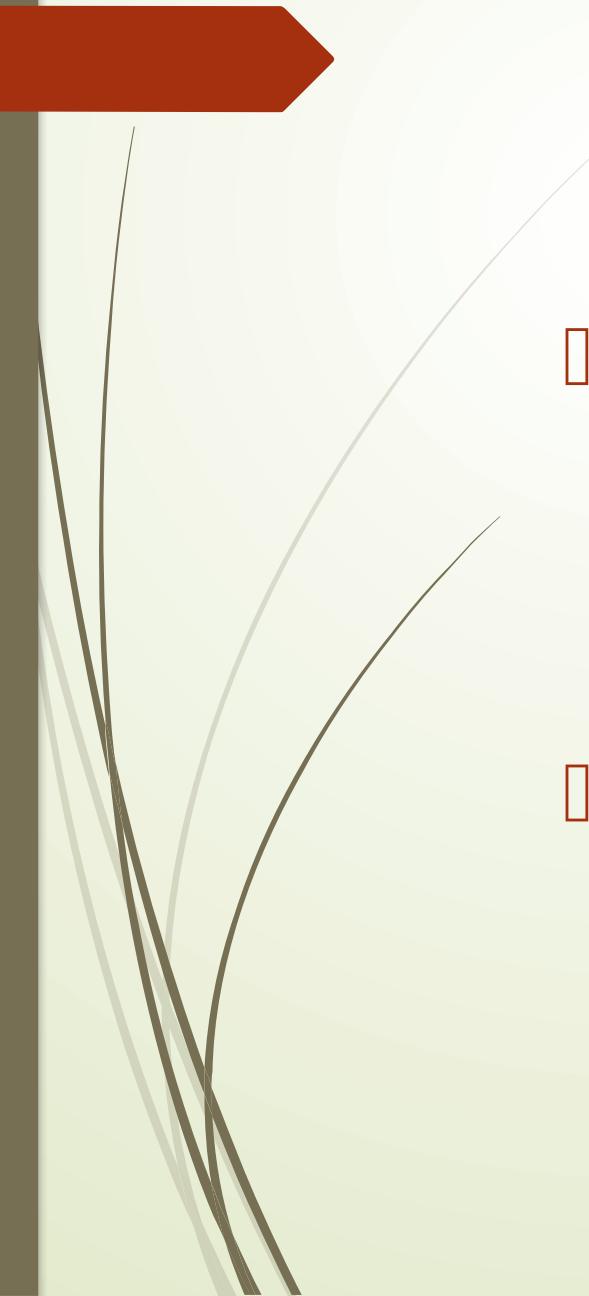
# Granularidade de Dados

- **Granularidade** de dados é uma **questão crítica** nos **projetos de Data Warehouse (DW)**.
- Nos **sistemas transacionais/operacionais** é tida como **certa**: os **dados** são **detalhados** e armazenados no **nível mais baixo de granularidade**.



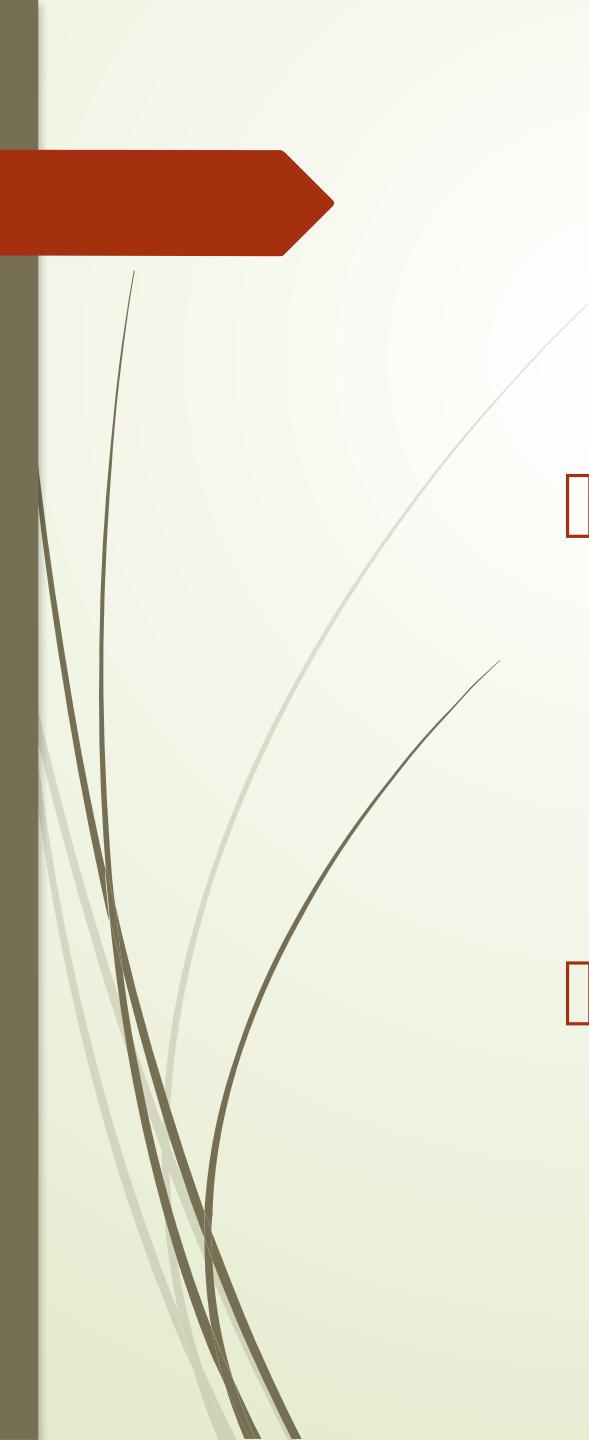
# Granularidade de Dados

- No DW a **granularidade** não é um pressuposto.
- A **granularidade** afeta profundamente o **volume de dados** que reside no DW e também afeta o **tipo de consulta** que pode ser realizada.
- Quanto **maior o volume de dados menor a performance**.



# Granularidade de Dados

- Quando se tem um **nível de granularidade muito alto**, o **espaço em disco** e o **número de índices** necessários se tornam **bem menores**.
- Porém há uma correspondente **diminuição da possibilidade** de utilização dos dados para atender a **consultas detalhadas**.



# Granularidade de Dados

- Considere um ***Data Mart*** de **vendas** com as **vendas** realizadas de **todos os produtos** da empresa nos **últimos 5 anos**.
- Qual seria a utilidade desse ***Data Mart*** para as **estratégias de negócios**?

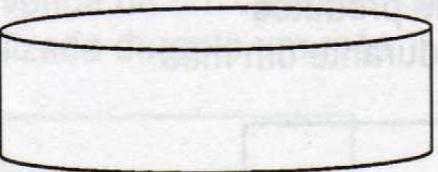


# Granularidade de Dados

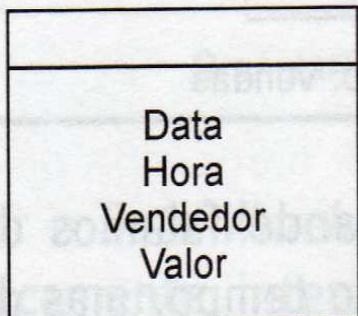
- Manter uma **granularidade baixa** de transação de vendas, tratando-as **consolidadas mês a mês?**
- Ou **alta granularidade**, tratando as vendas por **cada operação realizada?**
- Interessaria para alguém conhecer a **evolução das vendas no dia 5 de cada mês nos últimos cinco anos?** Acredito que não.

# Granularidade de Dados

Alto nível de detalhes  
Baixo nível de granularidade

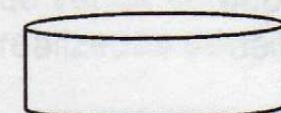


Exemplo  
Detalhe de cada transação  
de venda de um vendedor  
realizada durante um mês

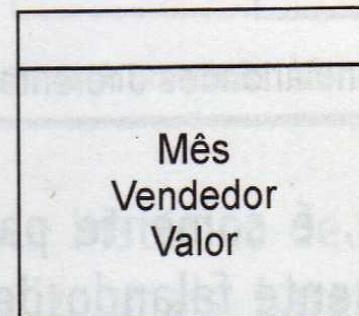


50 registros por mês

Baixo nível de detalhes  
Alto nível de granularidade



Exemplo  
Sumário das transações  
de venda de um vendedor  
realizadas durante um mês



1 registro por mês

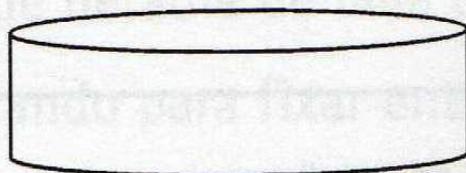


# Granularidade de Dados

- Mas poderíamos ter uma **granularidade relativa** ao **agrupamento de dias**, como, por exemplo, período de **Natal**, período de **Páscoa**, **dia das mães** etc.
- Um fato são as **vendas realizadas por vendedores, independente de produto**.
- Outro fato são as vendas de produto no mês.

# Granularidade de Dados

Alto nível de detalhes  
Baixo nível de granularidade

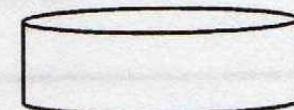


Exemplo  
Detalhe de cada transação  
de venda de um vendedor  
realizada durante um mês

Data
Hora
Vendedor
Valor

50 registros por mês

Baixo nível de detalhes  
Alto nível de granularidade



Exemplo  
Sumário das transações  
de venda de produtos  
realizadas durante um mês

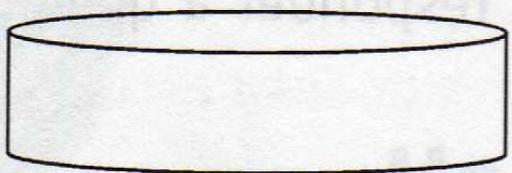
Mês
Vendedor
Valor

1 registro por mês

Granularidades diferentes em um mesmo assunto: Vendas

# Granularidade de Dados

**Alto nível de detalhes**  
**Baixo nível de granularidade**



**Exemplo**  
Detalhe de cada transação  
de venda de um vendedor  
realizada durante um mês

Data
Hora
Estado
Vendedor
Valor

50 registros por mês

**Baixo nível de detalhes**  
**Alto nível de granularidade**

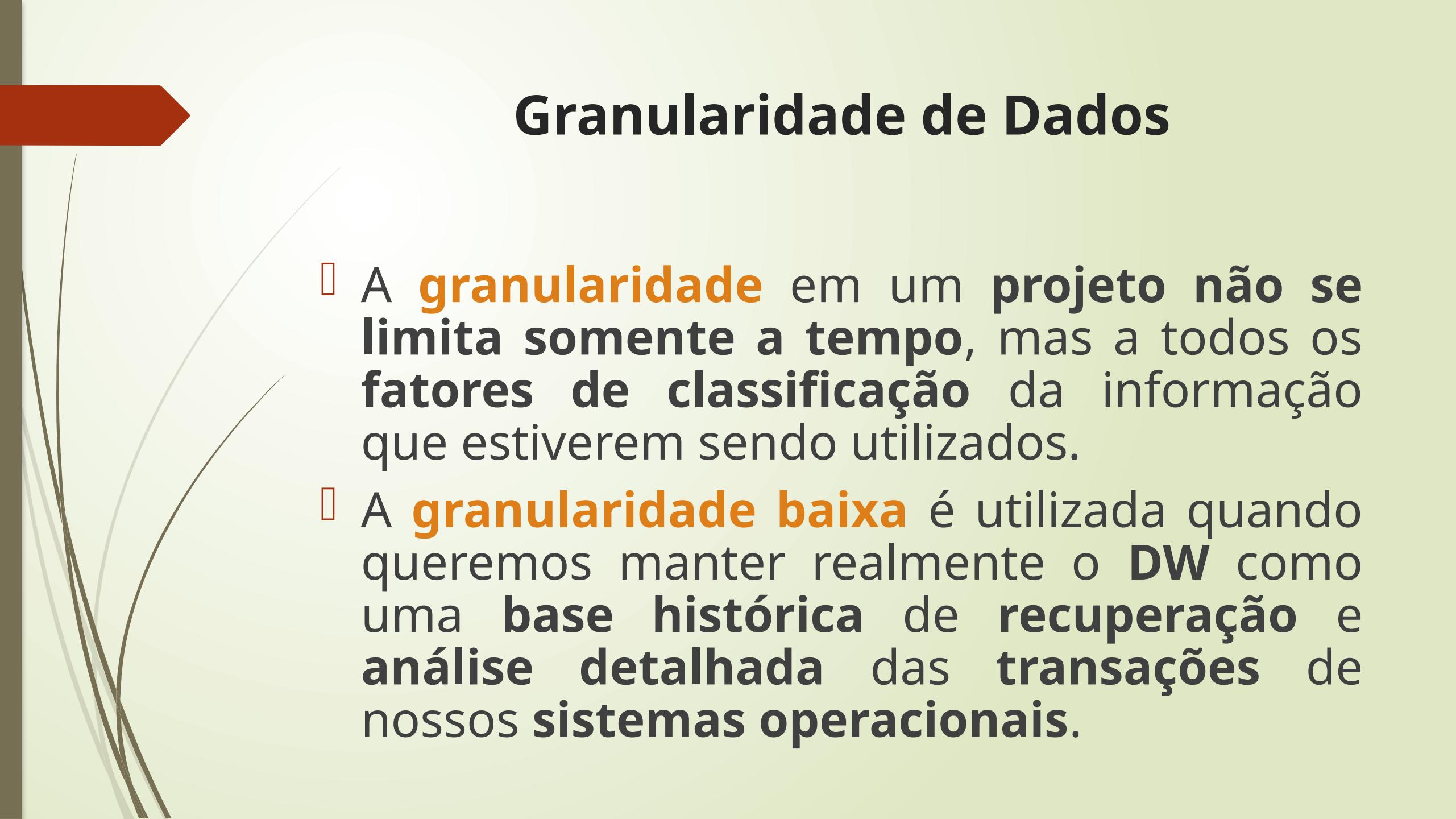


**Exemplo**  
Sumário das transações  
de venda de produtos  
realizadas durante um mês

Mês
Região
Produto
Valor

1 registro por mês

Granularidades diferentes em um mesmo assunto: Vendas



# Granularidade de Dados

- A **granularidade** em um projeto não se limita somente a tempo, mas a todos os **fatores de classificação** da informação que estiverem sendo utilizados.
- A **granularidade baixa** é utilizada quando queremos manter realmente o **DW** como uma **base histórica** de **recuperação** e **análise detalhada** das **transações** de nossos **sistemas operacionais**.



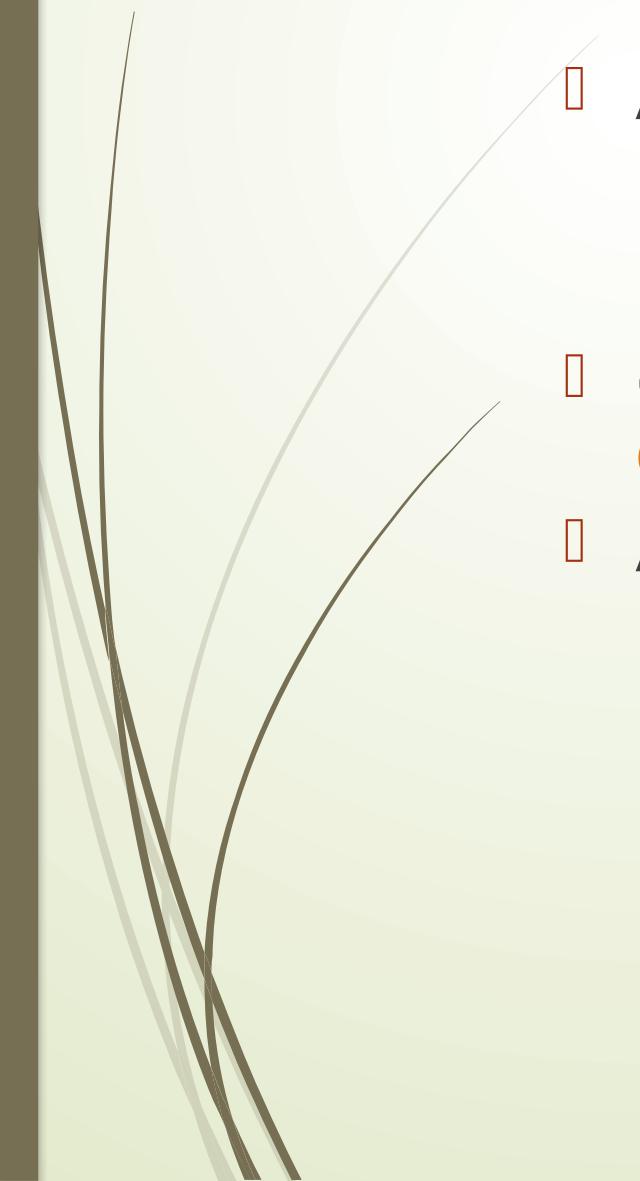
# Granularidade de Dados

- Normalmente **nem todos os campos** dos sistemas legados são transportados para o **DW**.
- Utilizar o **DW** como **ambiente histórico** de dados **não** é uma **decisão economicamente interessante**.
- É mais **barato** manter **informações históricas** para **recuperação** em **fitas magnéticas**.



# Modelagem de Dados

- Um **modelo** é uma **abstração** e uma **reflexão** do mundo **real**.
- **Modelar** é a forma de **visualizarmos** o que queremos **realizar**.
- A **modelagem Entidade e Relacionamento** (**modelagem ER**) é a essência técnica da arquitetura de **Data Warehouse** (**DW**).

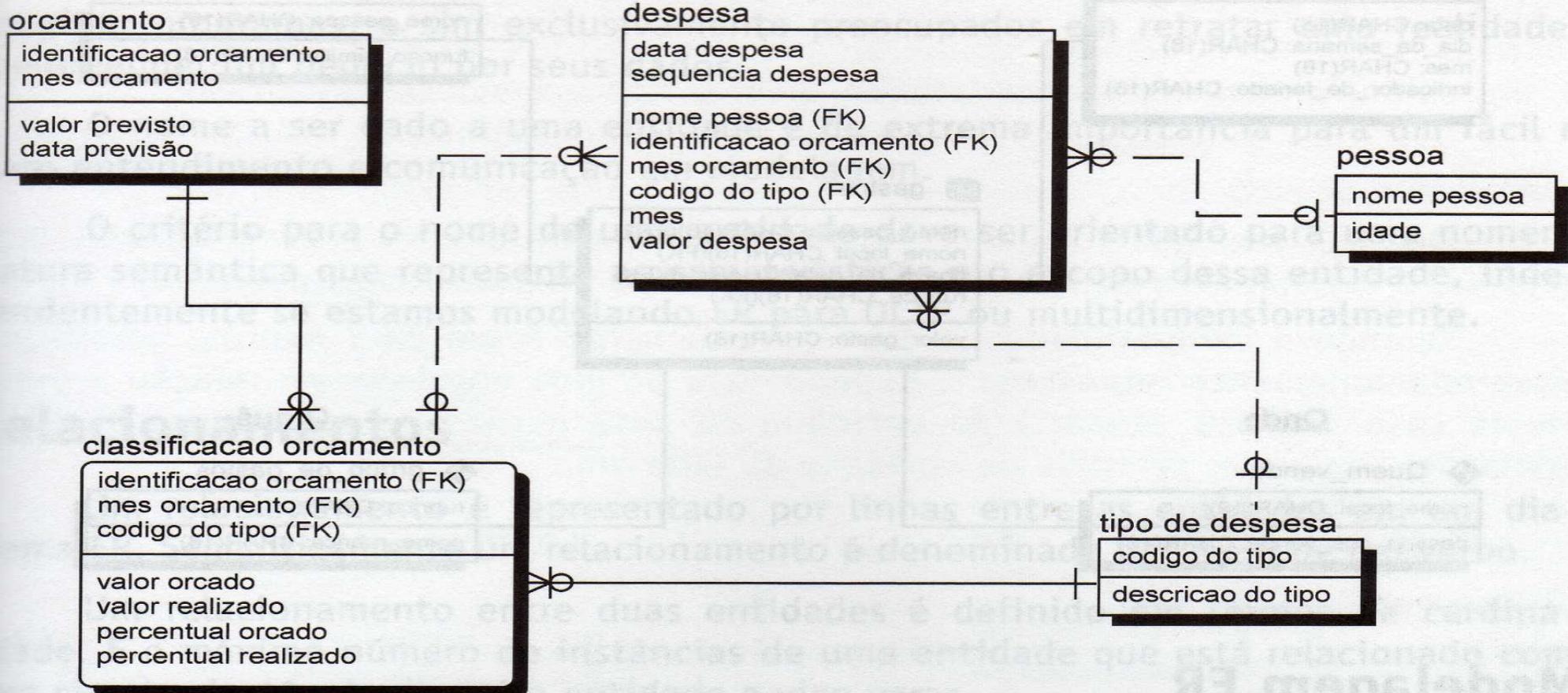


# Modelagem de Dados

- A **modelagem ER** tem como foco primário:
  - A eliminação de redundância de dados.
  - Manutenção de consistência dos dados.
- Considere o **exemplo** de um **sistema para controle dos gastos familiares**.
- A **operacionalização**:
  - Realizamos despesas com o orçamento.
  - Devemos controlar as despesas em relação ao orçamento do mês.
  - Devemos saber quais os produtos ou serviços relacionados com essas despesas.

# Modelagem de Dados

Para atender às necessidades de operacionalização do negócio, obtivemos o modelo da figura:





# Modelagem de Dados

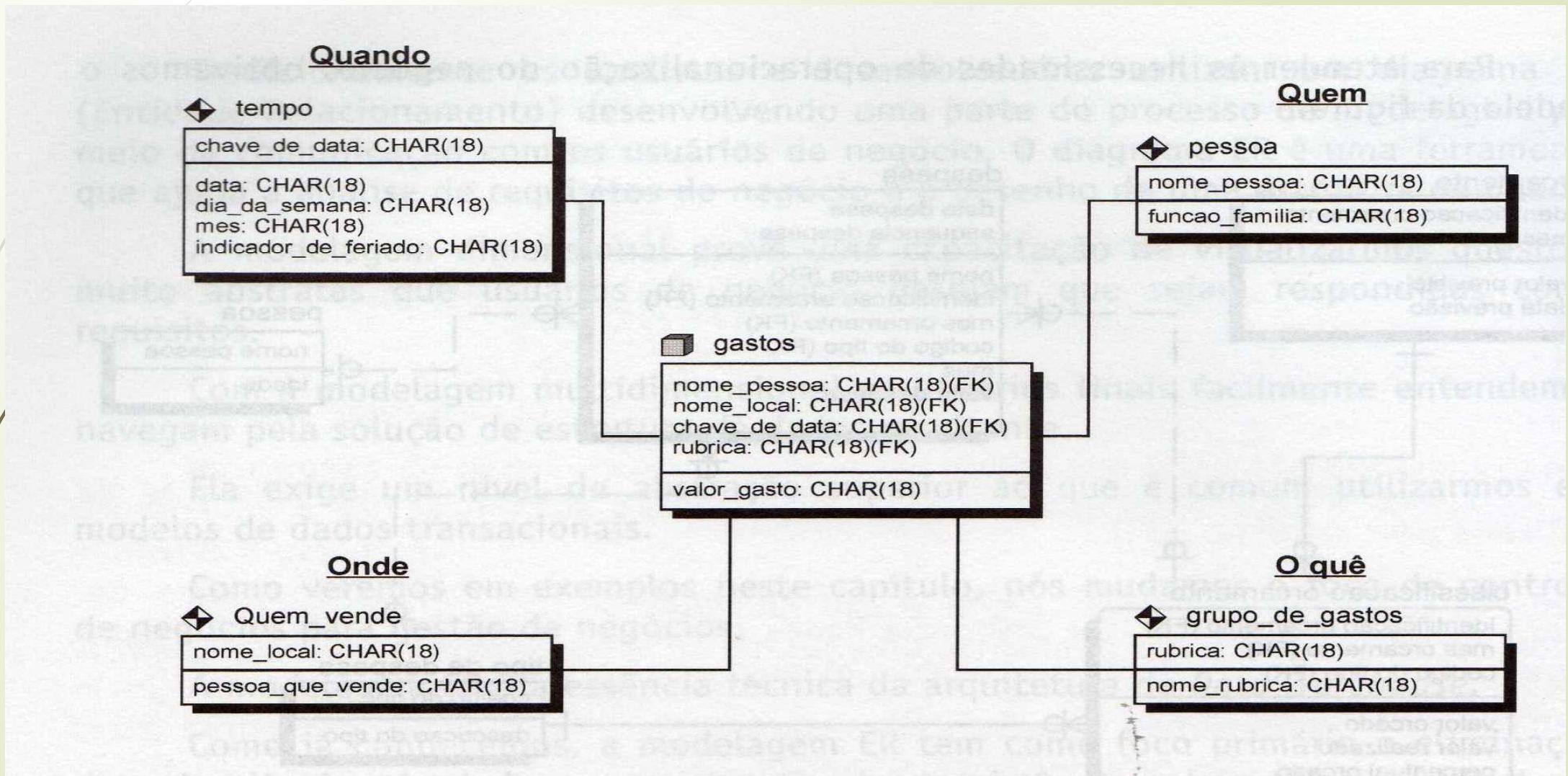
- **Gestão:**
  - Como está a evolução das despesas durante o ano?
  - Como está a relação percentual da despesa por tipo de gasto em relação ao orçamento?
  - Quais são os produtos ou serviços que possuem evolução histórica de aumento de valores em relação à variação e crescimento do orçamento?
  - Como a redução dos percentuais de destinação do orçamento relativos a um tipo de gasto afeta a performance do caixa doméstico?



# Modelagem de Dados

- O enfoque mudou de **registrar** e **visualizar** as **informações** para uma **avaliação do desempenho** de determinados **assuntos** ao longo de um **tempo histórico**, além de buscar **simulações táticas** (cenários) para **análise estratégica de decisões**.

# Modelagem Multidimensional



# Modelagem Multidimensional

- Um modelo multidimensional é formado por três elementos básicos:
  - Fatos;
  - Dimensões;
  - Medidas (variáveis).
- **Fatos:**
  - É uma **coleção de itens de dados**, composta de **dados de medidas** e de **contexto**.

# Modelagem Multidimensional

## □ Fatos:

- Cada fato representa um **item**, uma **transação** ou um **evento de negócio**.
- É tudo aquilo que **reflete a evolução dos negócios** do dia-a-dia de uma organização.
- Um **fato** é representado por **valores numéricos** e implementado em tabelas denominadas **tabelas de fato** (*fact tables*).

# Modelagem Multidimensional

## □ Dimensões:

- São as possíveis **formas de visualizar os dados.**
- São os “por” dos dados:
  - “**por mês**”;
  - “**por país**”;
  - “**por produto**”;
  - “**por região**”.

# Modelagem Multidimensional

## □ Dimensão:

- As dimensões determinam o contexto de um assunto de negócios.
- Exemplo: as **dimensões** que participam do **fato vendas de produtos** geralmente são:
  - Tempo;
  - Localização;
  - Clientes;
  - Vendedores;
  - Cenários (realizados, projetados).

# Modelagem Multidimensional

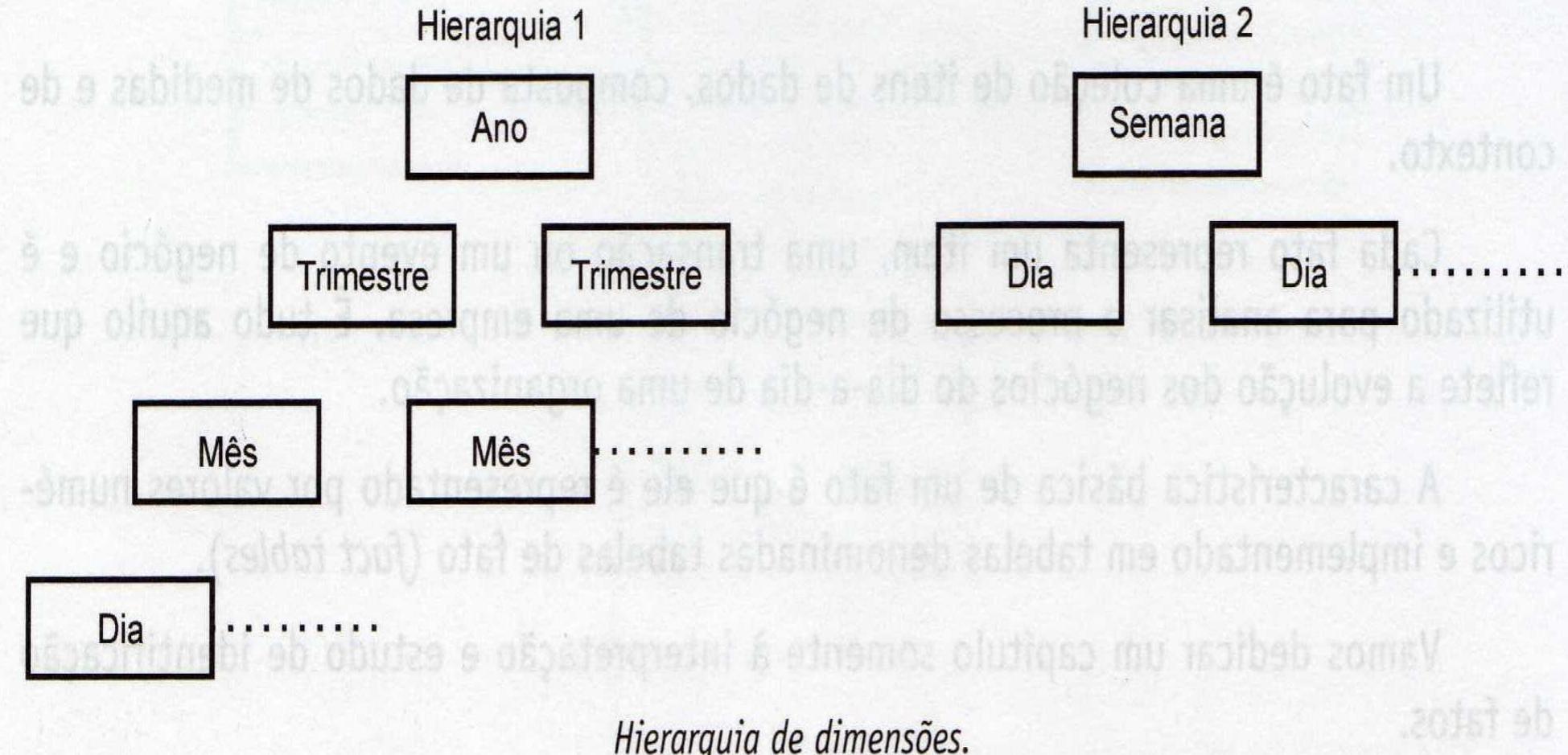
## □ Dimensões:

- Normalmente **não** possuem **atributos numéricos**, pois **são** somente **descritivas** e **classificatórias** dos **elementos** que participam de um **fato**.
- Uma **dimensão** pode conter muitos **membros**.
- Um **membro de dimensão** é um **nome diferente** utilizado para **determinar a posição** de um **item de dado**.

# Modelagem Multidimensional

- **Dimensão:**
  - **Exemplo 1:** todas as ocorrências de **ano, trimestre e mês** fazem a **dimensão tempo**.
  - **Exemplo 2:** todas as **cidades, estados e regiões** fazem a **dimensão geográfica**.
  - **Hierarquia de uma dimensão** é uma **classificação** de dados **dentro** de uma **dimensão**.

# Modelagem Multidimensional



# Modelagem Multidimensional

## □ Dimensão:

□ No exemplo anterior definimos duas hierarquias de dimensão pois uma mesma semana pode pertencer a dois meses.

## □ Medidas (Variáveis):

□ São os atributos numéricos que representam um fato, a performance de um indicador de negócios relativo às dimensões que participam desse fato.

# Modelagem Multidimensional

- **Medidas:**

- Exemplos:

- Valor em reais das vendas;

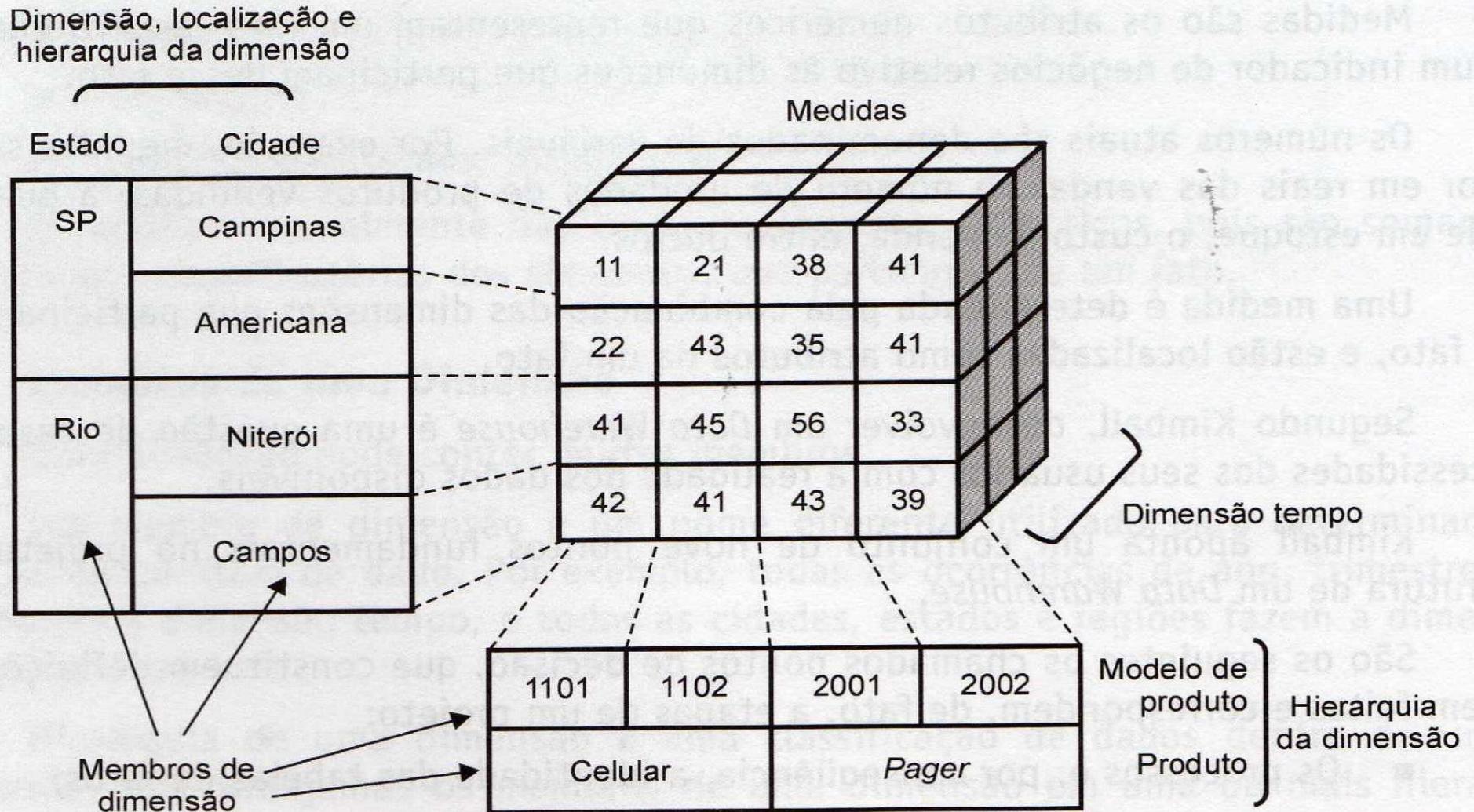
- Número de unidades de produtos vendidas;

- Quantidade em estoque de um produto;

- Custo de venda, etc.

- O caminho mais popular para **visualizar** um **modelo dimensional** é o desenho de um **cubo**.

# Modelagem Multidimensional





# Modelagem Multidimensional

- Podemos representar um **modelo tridimensional** por um **cubo**.
- Usualmente um **modelo dimensional** consiste em **mais de três dimensões**, o que é definido como um **hipercubo**.
- É muito **difícil visualizar graficamente** um **hipercubo**.

# Modelagem Multidimensional

- Conforme o exemplo anterior, a **medida** é o **volume de venda**.
- O **volume de venda** é determinado pela **combinação** de **três dimensões**:
  - Localização,
  - Produto e
  - Venda.
- As **dimensões localização e produto** possuem 2 níveis de hierarquia cada uma.

# Modelagem Multidimensional

- A **dimensão localização** tem o nível **estado** (membros: **SP** e **Rio**) e o nível **cidade** (membros: **Campinas**, **Americana**, **Niterói** e **Campos**).
- A **dimensão tempo** neste exemplo teria o número dos **anos de venda**, tais como **1997**, **1998** e **1999**.

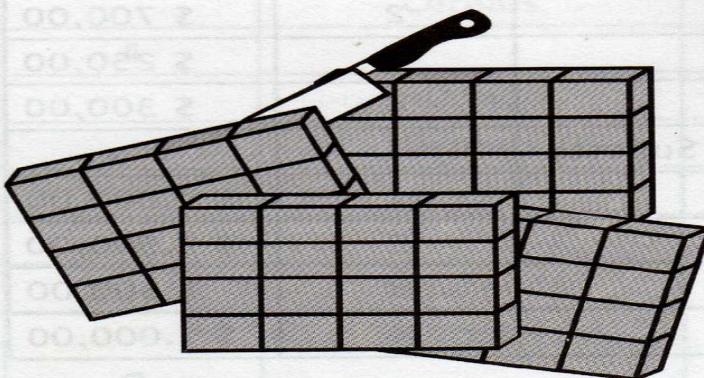
# Modelagem Multidimensional

- Em um **período específico de tempo** (não expresso em nosso exemplo), na cidade de **Campinas** no estado de **SP** foram vendidos **11.000 telefones celulares** do **modelo 1101** e **38.000 pagers** do **modelo 2001**.

# Modelagem Multidimensional

Imagine você analisando um cubo!

Você fatia os dados e as dimensões que quiser analisar, observar ou estudar.



Para melhor entendimento, então vamos buscar entender em exemplos de dados de um negócio a visão de cubos.

Vamos analisar que queremos saber as vendas totais classificadas por região de venda.

Região	Venda
Sul	\$ 1.500,00
Sudeste	\$ 5.000,00
Nordeste	\$ 2.350,00
Norte	\$ 1.890,00
Centro-Oeste	\$ 1.732,00
<b>Total de Vendas</b>	<b>\$ 11.472,00</b>

# Modelagem Multidimensional

Podemos ver que o fato analisado Venda associou uma dimensão de análise, no nosso caso a dimensão Região.

Observe que estamos utilizando somente uma dimensão.

Vamos expandir a análise.

Digamos que um usuário deseja ver os dados de venda (fato) por Região, porém dividido em trimestres de negócio.

Teríamos então:

Região	Trimestre	Venda
Sul	1	\$ 250,00
	2	\$ 700,00
	3	\$ 250,00
	4	\$ 300,00
Sudeste	1	\$ 1.000,00
	2	\$ 2.500,00
	3	\$ 500,00
	4	\$ 1.000,00

Agora a este resultado associamos mais uma dimensão, que é tempo. Temos então duas dimensões associadas a um fato.

# Modelagem Multidimensional

- ***On-line Analytical Processing*** (OLAP).
- É o **conjunto de ferramentas** que possibilita efetuar a **exploração dos dados** de um **Data Warehouse** (DW).
- A **análise** denominada **multidimensional** representa os **dados** como **dimensões** em vez de **tabelas**.
- **Slice** e **Dice** são formas de mudança das **dimensões** a serem **visualizadas**.

# Modelagem Multidimensional

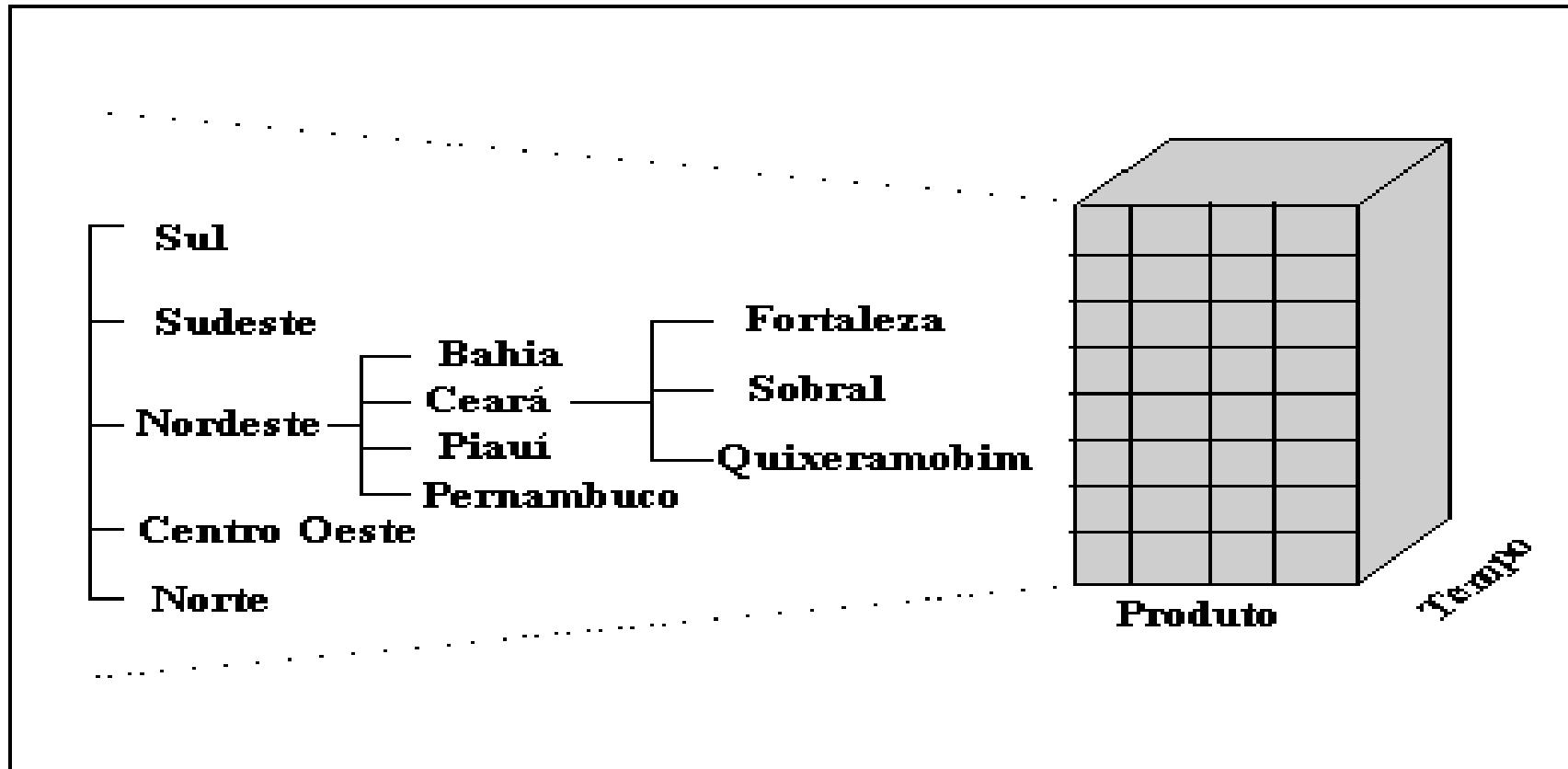
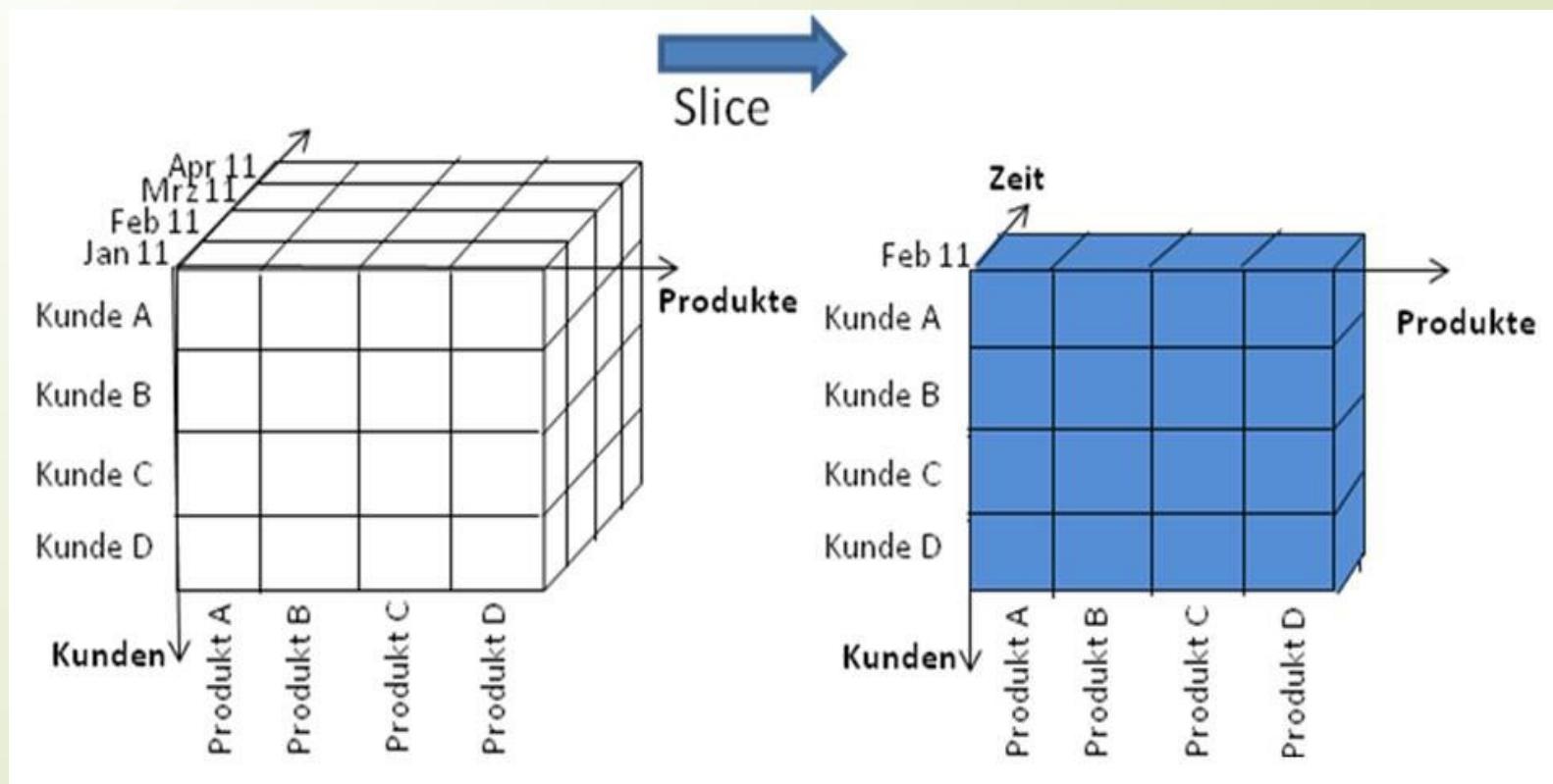


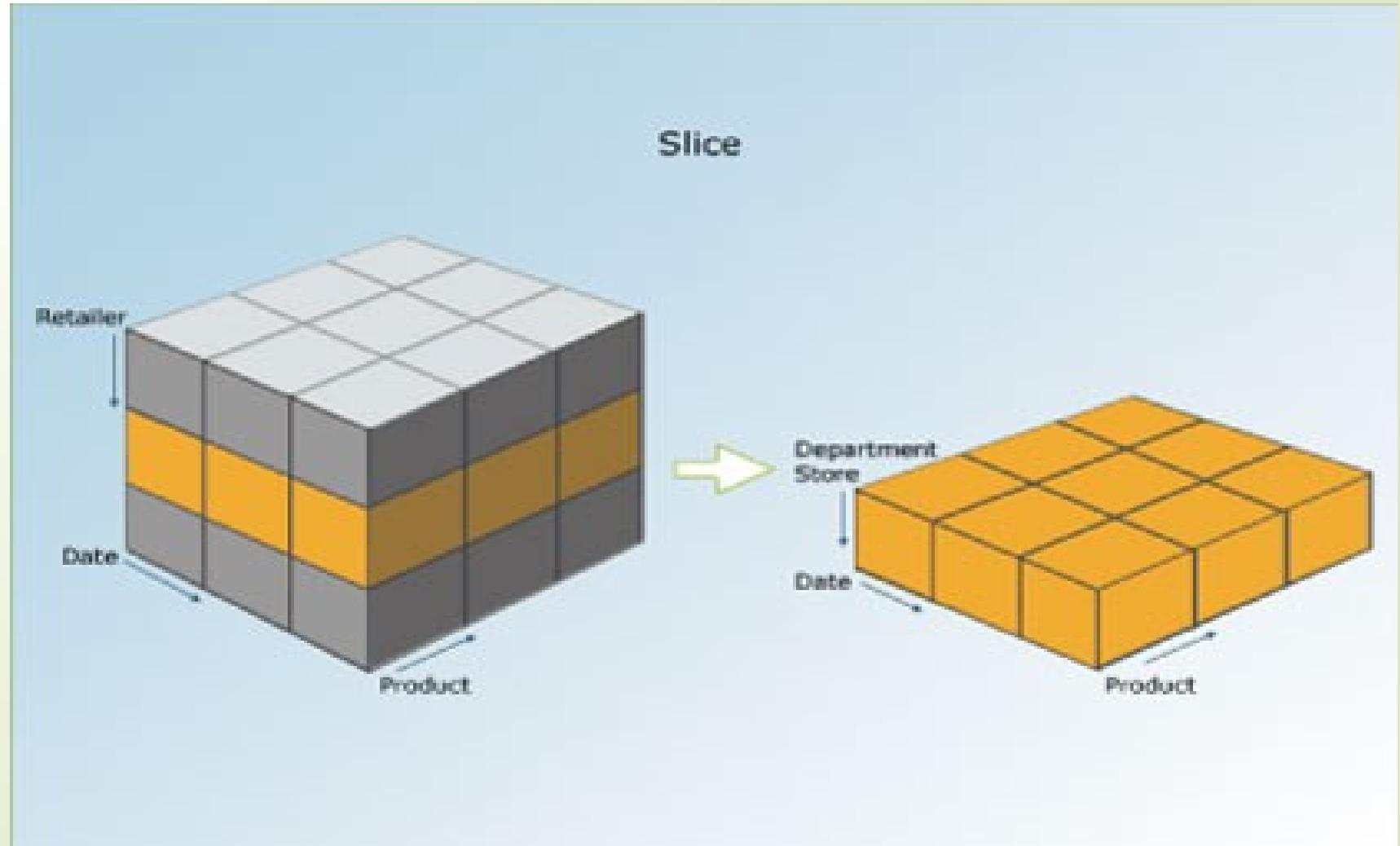
Figura 10 - Dimensões com níveis hierárquicos

# Modelagem Multidimensional

- ***SLICE*** (“fatiar” uma seção do cubo).

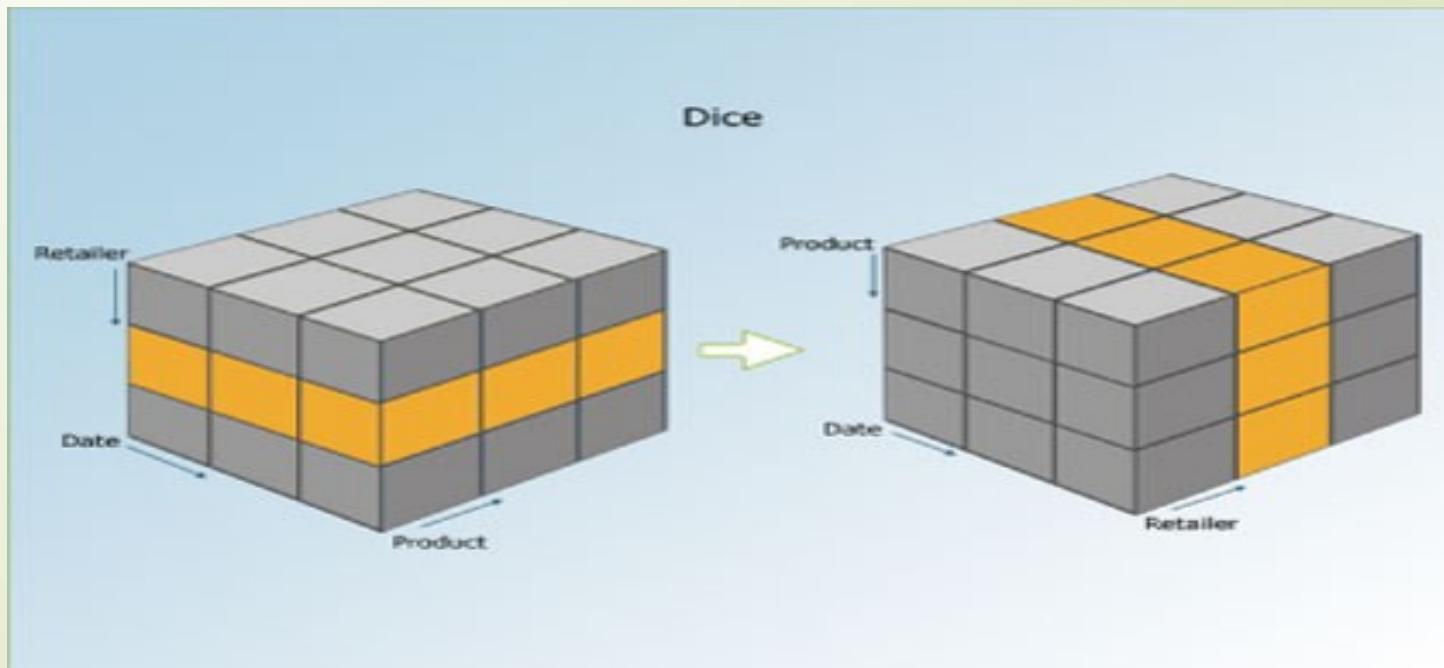


# Modelagem Multidimensional



# Modelagem Multidimensional

- **DICE** (“Dado” – lançamento de um dado: rotação do cubo).



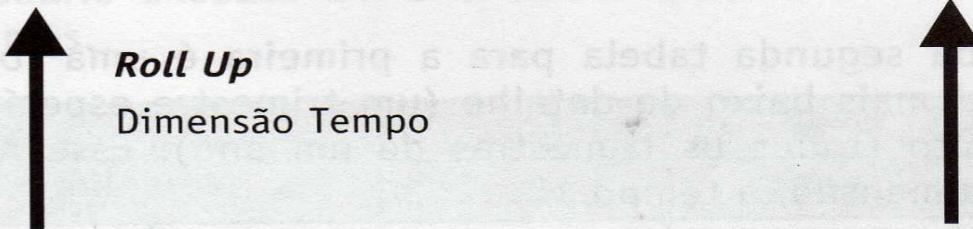
# Modelagem Multidimensional

- Operações básicas do **OLAP**:
  - Slice;
  - Dice;
  - Drill down;
  - Roll up(ou Drill up).

# Modelagem Multidimensional

- **Roll Up:** o usuário **diminui o nível de detalhe** da informação, aumentando o nível de granularidade.
- A diminuição do detalhamento da dimensão tempo (de trimestre para ano).

Volume de Produção (em milhares)		1999			
		Trim. 1	Trim. 2	Trim. 3	Trim. 4
Região Sul	RS	78	67	22	56
	SC	90	67	88	99

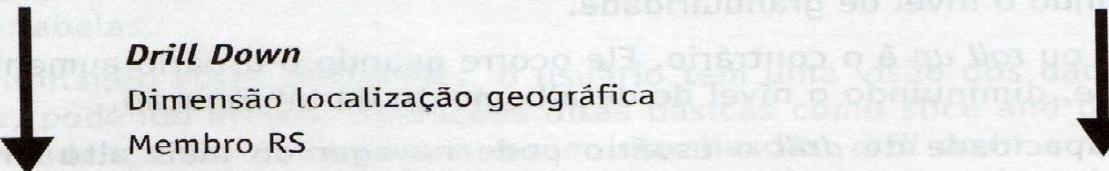


Volume de Produção (em milhares)		Trimestre 1		
		Janeiro	Fevereiro	Março
Região Sul	RS	30	26	22
	SC	28	30	32

# Modelagem Multidimensional

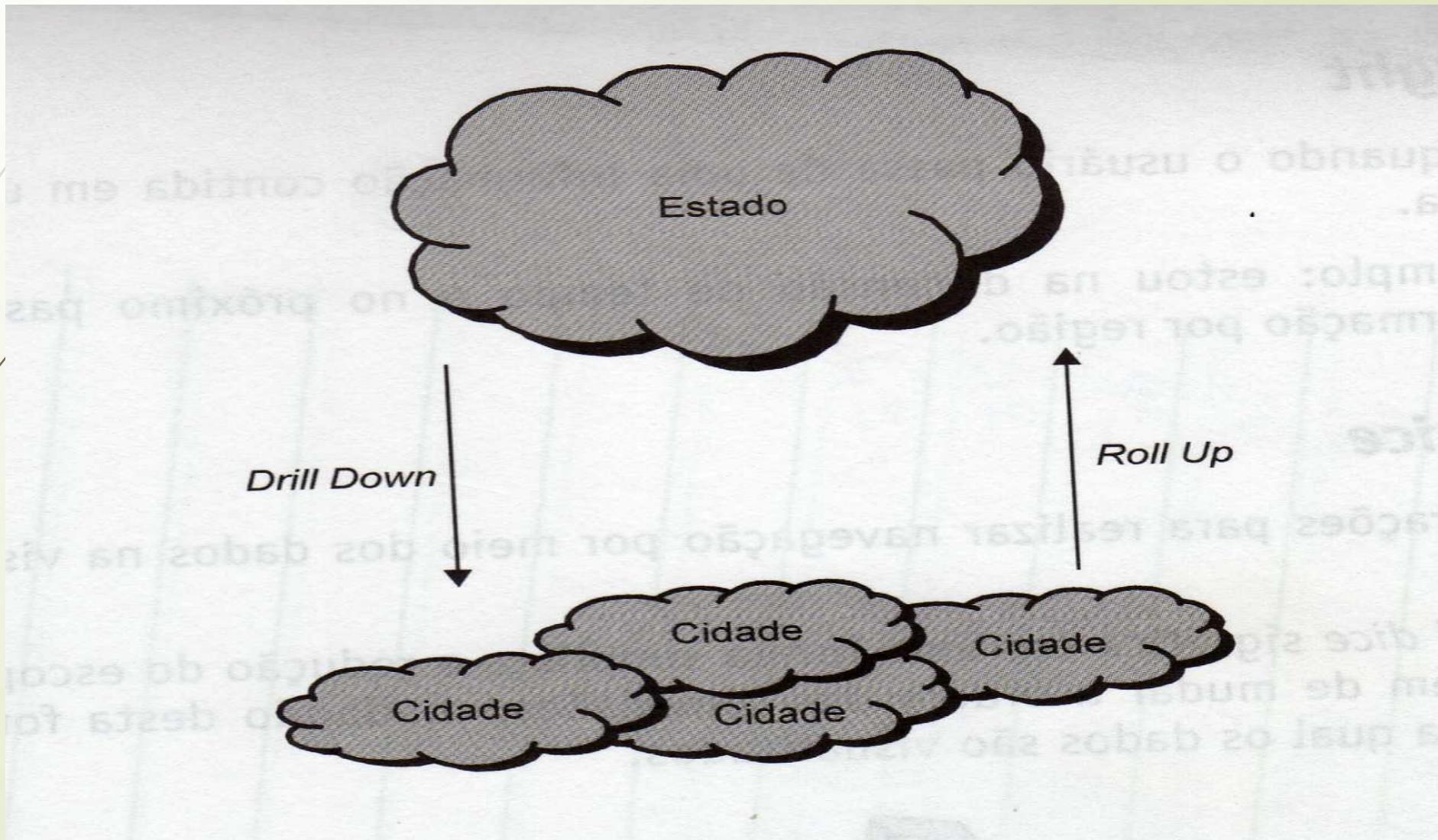
- **Drill Down:** o usuário **aumenta o nível de detalhamento** da informação (diminuindo o nível de granularidade).
- O aumento do detalhe na dimensão localização geográfica: dos estados da região sul para as cidades do Rio Grande do Sul.

Volume de Produção (em milhares)		Telefone Celular		Pagers	
Região Sul	RS	1001	1002	2001	2002
	SC	33	12	8	12
		45	34	20	23



Volume de Produção (em milhares)		Telefone Celular		Pagers	
RS	Canoas	1001	1002	2001	2002
	Porto Alegre	13	4	2	5
		20	8	6	7

# Modelagem Multidimensional



# Modelagem Multidimensional

## *Drill Across*

Ocorre quando o usuário pula um nível intermediário dentro de uma mesma dimensão. Por exemplo: a dimensão tempo é composta por ano, semestre, trimestre, mês e dia.

O usuário executa um *drill across* quando ele passar de ano direto para semestre ou mês.

Volume de Produção (em milhares)		Telefone Celular		Pagers	
Região Sul	RS	1001	1002	2001	2002
	SC	33	12	8	12
		45	34	20	23

*Drill across*

Dimensão localização geográfica

Membro RS

Volume de Produção (em milhares)		Telefone Celular		Pagers	
Região Sul	RS	Janeiro 2001	Janeiro 2002	Janeiro 2001	Janeiro 2002
	SC	2	4	1	4
		5	3	3	2

# Modelagem Multidimensional

## ***Drill Through***

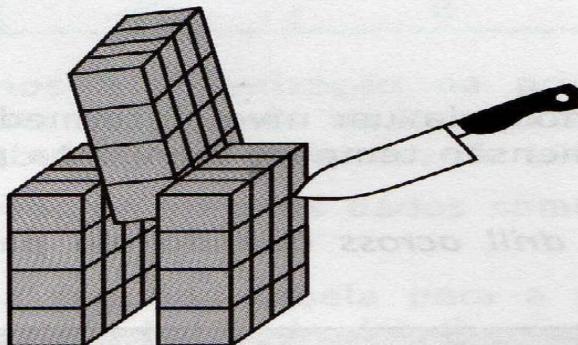
Ocorre quando o usuário passa de uma informação contida em uma dimensão para uma outra.

Por exemplo: estou na dimensão de tempo e no próximo passo começo a analisar a informação por região.

## ***Slice and Dice***

São operações para realizar navegação por meio dos dados na visualização de um cubo.

*Slice and dice* significa em uma forma simplista a redução do escopo dos dados em análise, além de mudar a ordem das dimensões, mudando desta forma a orientação segundo a qual os dados são visualizados.



Essa característica das ferramentas OLAP é de extrema importância, pois com ela nós podemos analisar as informações de diferentes prismas limitados somente pela nossa imaginação.

# Modelagem Multidimensional

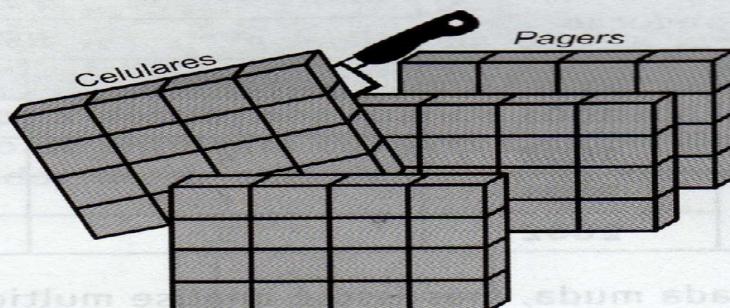
Volume de Produção (em milhares)		Celulares e Pagers		
Região Sul	RS	Janeiro	Fevereiro	Março
	SC	28	30	32

Fatiar ou realizar slice é a operação de visualizarmos somente a produção de um tipo de produto, por exemplo, celulares.

Volume de Produção (em milhares)		Celulares		
Região Sul	RS	Janeiro	Fevereiro	Março
	SC	22	18	18
		19	27	25

Esta tabela representa uma fatia dos dados.

Fatiamos um cubo.

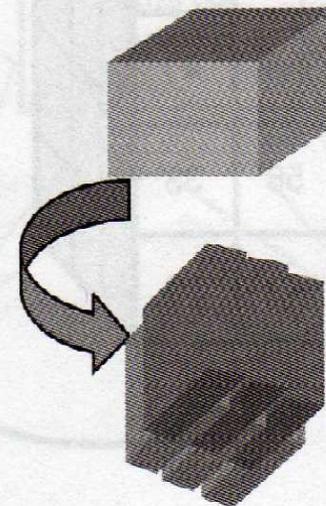


# Modelagem Multidimensional

*Dice* é a mudança de perspectiva da visão.

É a extração de um "subcubo" ou a interseção de vários slices.

Vamos nos utilizar de tabelas de dados para ilustrar melhor essa operação. É como se girássemos o cubo em nossas mãos.



# Modelagem Multidimensional

Volume de Produção (em milhares)		1999			
		Telefone Celular		Pagers	
		1001	1002	2001	2002
RS	Canoas	13	4	2	5
	Porto Alegre	20	8	6	7

Estávamos visualizando e analisando no sentido estado, cidade, ano, modelo de produto e produto.

Cada valor apresentado significa o volume produzido em um estado, em uma cidade, em um ano, de um modelo de produto e de um produto.

Mudamos nossa perspectiva para modelo de produto, produto, ano, estado e cidade.

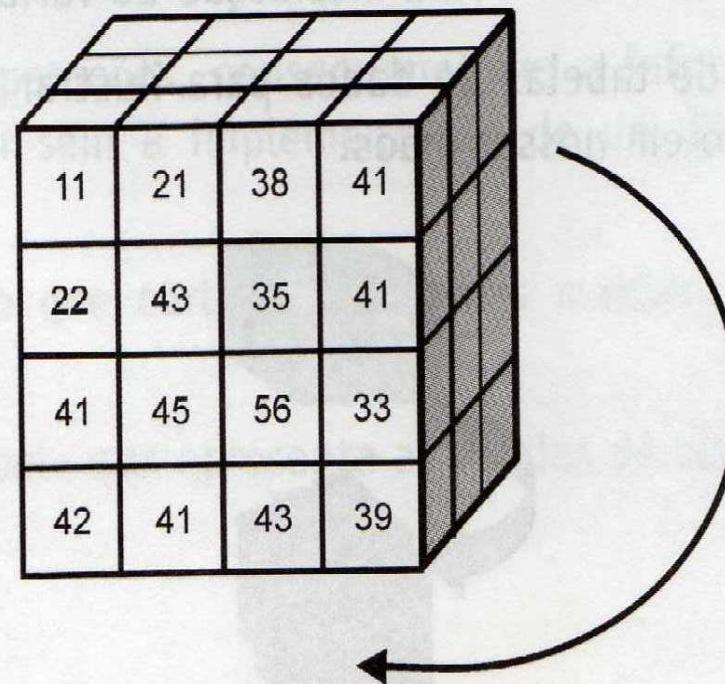
Agora cada valor representa a produção de um modelo de produto e um produto, em um ano, em um estado e em uma cidade.

Volume de Produção (em milhares)		1999	
		RS	
		Canos	Porto Alegre
Telefone Celular	1001	13	2
	1002	4	5
Pagers	2001	2	6
	2002	5	7

# Modelagem Multidimensional

Pode parecer que nada muda, mas isso é análise multidimensional.

Nesse tipo de análise busca-se descobrir comportamentos conforme a perspectiva de análise dos dados.



# Modelagem Multidimensional

## Modelo Estrela



# Modelo Estrela

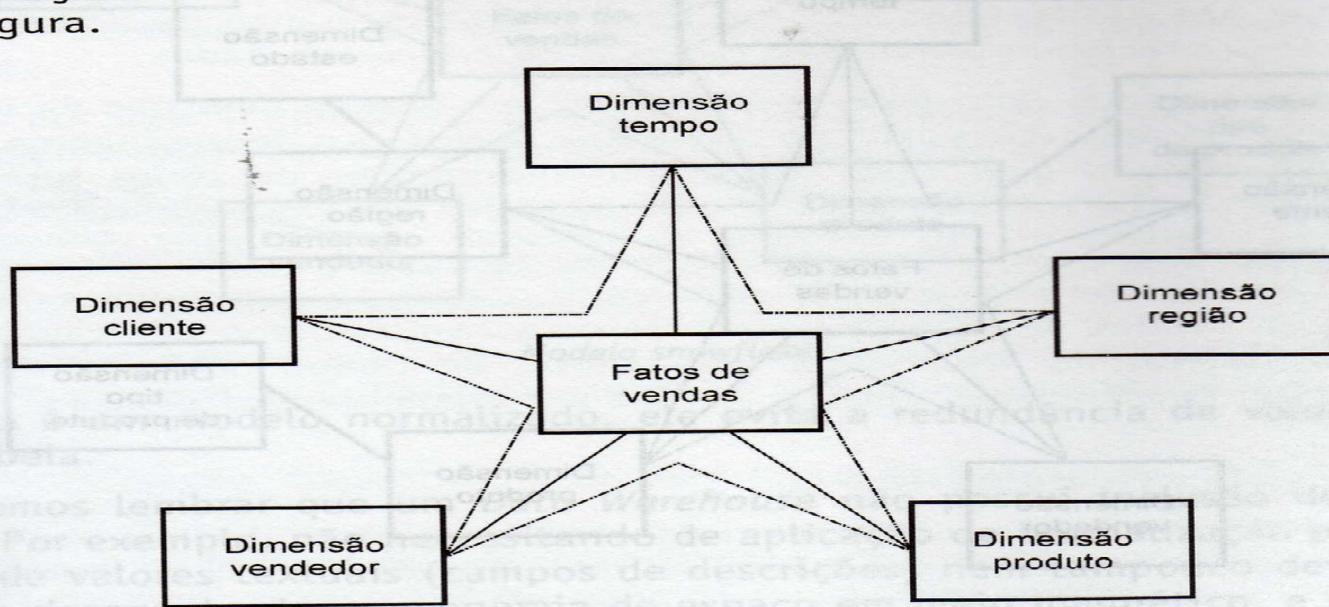
## Modelo *Star* ou Estrela

*Star schema* é o termo comum para a designação de modelos de dados multidimensionais.

Como o termo esquema nos leva-nos diretamente a uma idéia de modelo físico, o mais usual é denominarmos de modelo estrela.

O modelo estrela é a estrutura básica de um modelo de dados multidimensional.

Sua composição típica possui uma grande entidade central denominada fato (*fact table*) e um conjunto de entidades menores denominadas dimensões (*dimension tables*), arranjadas ao redor dessa entidade central, formando uma estrela, como mostra a figura.

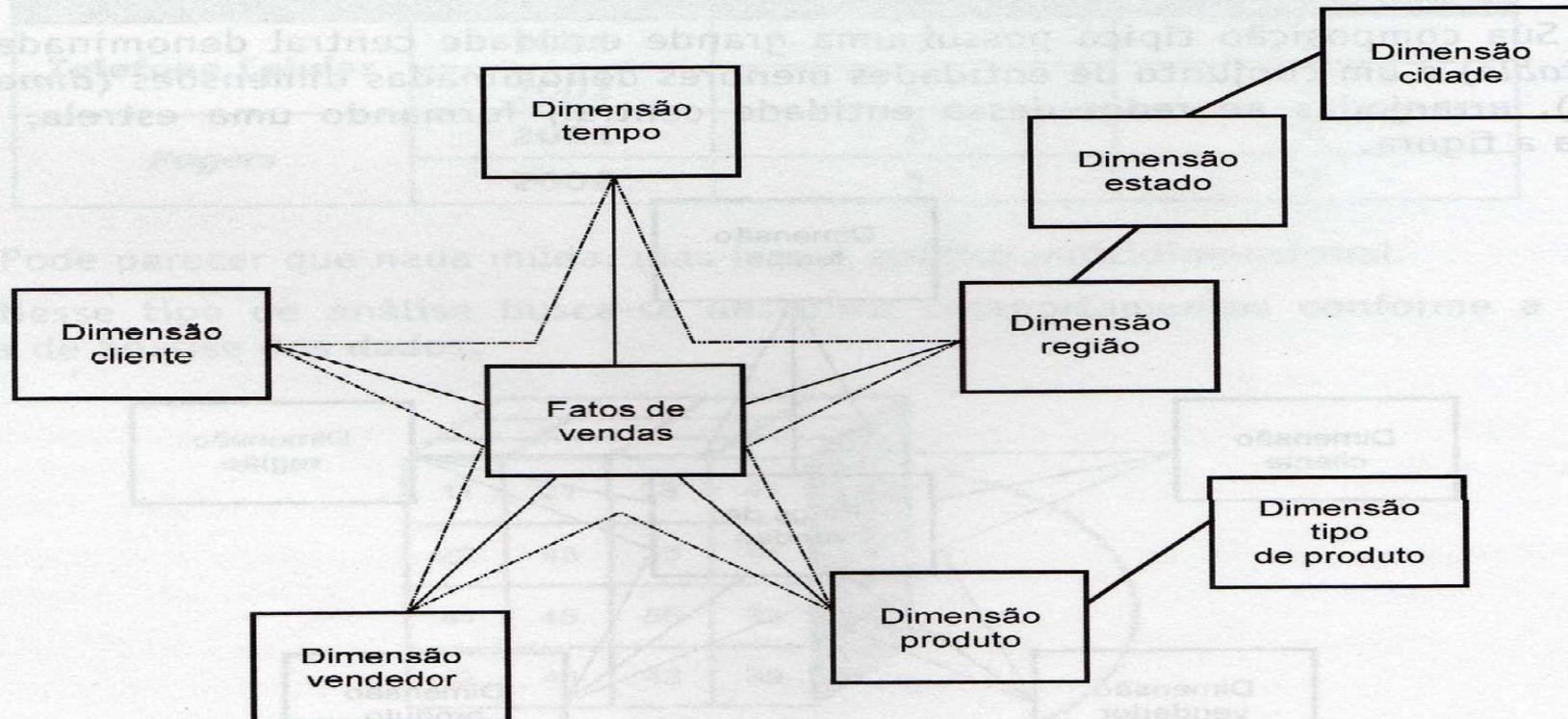


# Modelo Floco de Neve

## Modelo *Snowflake* (Floco de Neve)

A modelagem multidimensional normalmente inicia com a identificação de fatos e dimensões, depois de levantados os requisitos de negócio a serem satisfeitos.

O modelo de dados multidimensional é usualmente similar a uma estrela em seu *design*. Um fato está no centro da estrela e as dimensões ao seu redor, formando as pontas dessa estrela.



# Modelo Floco de Neve

