

# **Data Warehouse**

## **Conteúdo Programático**

1. Introdução ao Data Warehouse
2. Arquitetura do Data Warehouse
3. OLAP
4. Projeto do Data Warehouse

## Unidade 1

# **Introdução ao Data Warehouse**

- Sistemas de Apoio à Decisão
- Conceituação de Data Warehouse
- Principais Características

## **Sistemas de Apoio à Decisão**

- Informação
  - Melhor recurso do qual uma empresa pode dispor para tomar decisões estratégicas
  - Obtida analisando dados históricos sobre vendas, produção, clientes, etc.
- Análise dos dados
  - Fornece informações vitais para a empresa
  - Pode aumentar a competitividade da empresa
  - Era feita intuitivamente pelos gerentes

## **Sistemas de Apoio à Decisão**

- Dificuldades para obter informação
  - Quantidade de dados a serem analisados cresce com a expansão do negócio e com o passar dos anos
  - Dados conflitantes vindos de fontes diferentes podem gerar informações desencontradas
  - Impossível para um ser humano manter e analisar todos os dados
  - Informação não é mais mantida por gerentes devido à mobilidade no mercado de trabalho

## **Sistemas de Apoio à Decisão**

- Sistemas de Apoio à Decisão (SAD)
  - Usam dados históricos mantidos em um banco de dados convencional
  - Dados históricos são analisados usando técnicas de mineração de dados para obter informações usadas na tomada de decisões
  - Estatísticas de venda, produção, clientes, etc. podem ser levantadas e consideradas para tomar decisões estratégicas de negócio

## **Sistemas de Apoio à Decisão**

- Benefícios dos Sistemas de Apoio à Decisão
  - Determinar o mercado-alvo de um produto
  - Definir o preço de um produto, criar promoções e condições especiais de compra
  - Verificar a eficácia de campanhas de marketing
  - Otimizar a quantidade de produtos no estoque
  - Responder rapidamente a mudanças no mercado e determinar novas tendências
- ... ou seja, ganhar eficiência e lucratividade

## **Sistemas de Apoio à Decisão**

- Problema: dados históricos não são mantidos nos BDs da empresa
  - Volume de dados seria muito grande
  - Desempenho seria insatisfatório
- Solução: criar um BD exclusivamente para manter os dados históricos
  - Especializado para realizar poucas consultas sobre um grande volume de dados
  - Surge o Data Warehouse (DW)

## **Data Warehouse**

### ■ Histórico

- Criado pela IBM na década de 60 com o nome *Information Warehouse*
- Relançado diversas vezes sem grande sucesso
- O nome Data Warehouse foi dado por William H. Inmon, considerado o pai desta tecnologia
- Tornou-se viável com o surgimento de novas tecnologias para armazenar e processar uma grande quantidade de dados

## **Data Warehouse**

### ■ O que é?

- Sistema que armazena dados históricos usados no processo de tomada de decisão
- Integra os dados corporativos de uma empresa em um único repositório

### ■ Para que serve?

- Para criar uma visão única e centralizada dos dados que estavam dispersos em diversos BDs
- Permite que usuários finais executem consultas, gerem relatórios e façam análises

## **Data Warehouse**

- BDs usados nas aplicações de negócio são chamados BDs operacionais
- DW é um BD informacional alimentado com dados dos BDs operacionais da empresa
  - Disponibiliza dados atuais e a dados históricos
  - Dados podem ser sumarizados (condensados) para que sejam analisados
  - Contém também metadados, que são dados sobre os dados armazenados no DW

## **Data Warehouse**

- Então o Data Warehouse é apenas um BD que contém também dados históricos?
- Para que seja considerado um Data Warehouse, um banco de dados deve:
  - Coletar dados de várias fontes
  - Dados coletados devem ser transformados para que haja uma visão única dos dados
  - Dados devem ser usados por aplicativos para obter informações que dêem apoio à decisão

## Data Warehouse

|                            | BD Operacional            | Data Warehouse                 |
|----------------------------|---------------------------|--------------------------------|
| Usuários                   | Funcionários              | Alta administração             |
| Utilização                 | Tarefas cotidianas        | Decisões estratégicas          |
| Padrão de uso              | Previsível                | Difícil de prever              |
| Princípio de funcionamento | Com base em transações    | Com base em análise de dados   |
| Valores dos dados          | Valores atuais e voláteis | Valores históricos e imutáveis |
| Detalhamento               | Alto                      | Sumarizado                     |
| Organização dos dados      | Orientado a aplicações    | Orientado a assunto            |

## Principais Características

- De acordo com a definição dada por Inmon, um Data Warehouse deve ser:
  - Orientado a assunto
  - Integrado
  - Não-volátil
  - Variável com o tempo

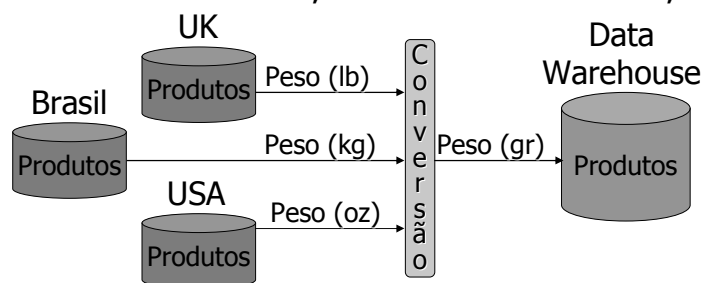
## Principais Características

- Orientação a assunto
  - Os dados em um DW são organizados de modo a facilitar a análise dos dados
  - Dados são organizados por assunto e não por aplicação, como em BDs operacionais



## Principais Características

- Integração
  - Dados de um DW provém de diversas fontes
  - Dados podem ser sumarizados ou eliminados
  - Formato dos dados deve ser padronizado para uniformizar nomes, unidades de medida, etc.



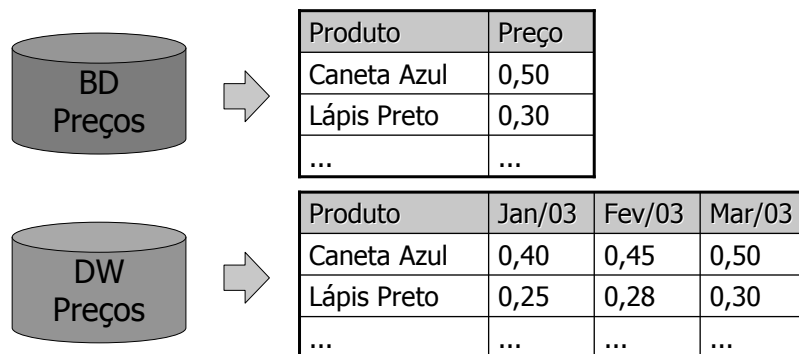


## Principais Características

- Não-Volátil
  - Dados não são mais alterados depois de incluídos no DW
- Operações no DW
  - Em um BD operacional é possível incluir, alterar e eliminar dados
  - Já no DW é possível apenas incluir dados
- Garante que consultas subseqüentes a um dado produzirão o mesmo resultado

## Principais Características

- Variável com o Tempo
  - Os dados no DW são relativos a um determinado instante de tempo



## Unidade 2

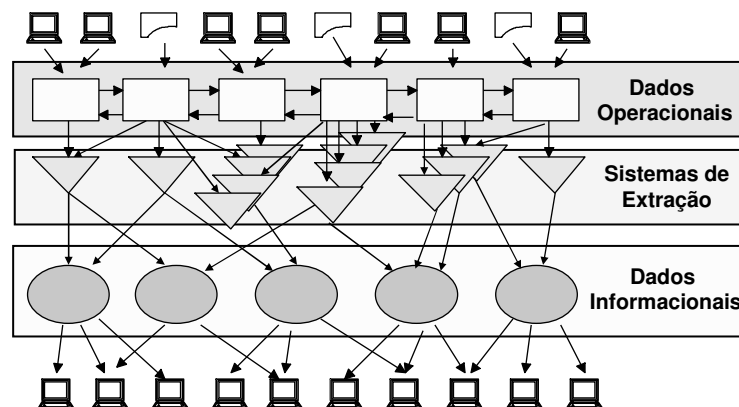
# Arquitetura do Data Warehouse

- Estrutura Interna
- Granularidade
- Data Marts
- Metadados
- Acesso aos Dados
- Tipos de DW

19

## Arquitetura do Data Warehouse

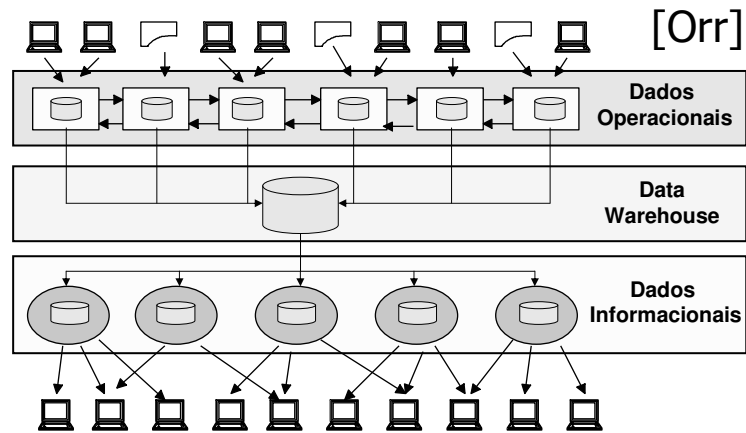
- Sistemas de Extração Tradicionais [Orr]



20

## Arquitetura do Data Warehouse

### ■ Sistemas baseados em Data Warehouse



21

## Arquitetura do Data Warehouse

- Principais tarefas efetuadas pelo DW
  - Obter dados dos BDs operacionais e externos
  - Armazenar os dados
  - Fornecer informações para tomada de decisão
  - Administrar o sistema e os dados
- Principais componentes do DW
  - Mecanismos para acessar e transformar dados
  - Mecanismo para armazenamento de dados
  - Ferramentas para análise de dados
  - Ferramentas de gerência

22

## **Estrutura Interna**

- Requisitos do DW
  - Eficiente
    - Grande volume de dados imutáveis
    - Processamento paralelo e/ou distribuído
  - Confiável
    - Funcionamento do sistema
    - Resultado das análises
  - Expansível
    - Crescente volume de dados
    - Maior número de fontes de dados

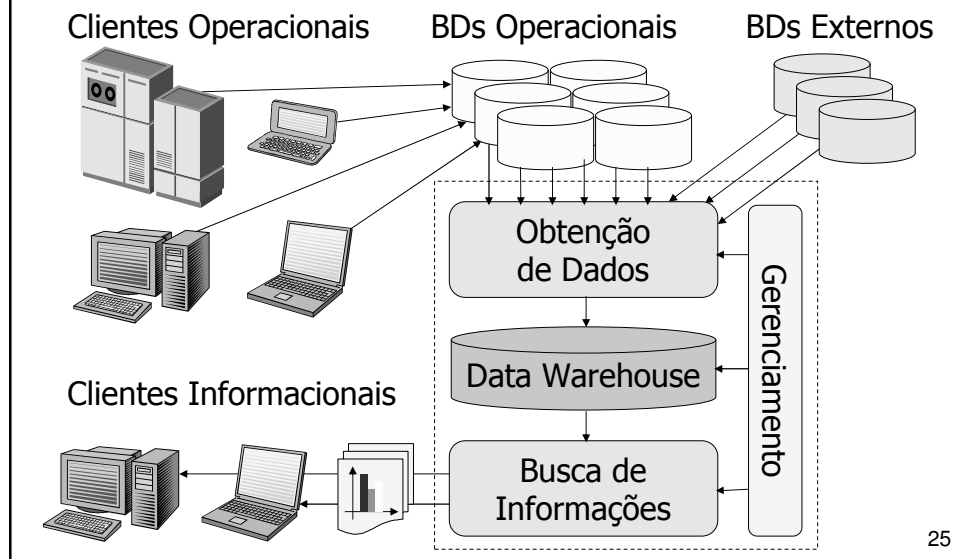
23

## **Estrutura Interna**

- Em geral são usados BDs relacionais para armazenar os dados do DW
  - Capazes de manter e processar grandes volumes de dados
  - Otimizados para lidar com dados imutáveis
- As ferramentas de análise empregam:
  - Técnicas de mineração de dados
  - Inteligência artificial: redes neurais, *fuzzy*, etc.
  - A Internet: *Web mining*, agentes móveis, etc.

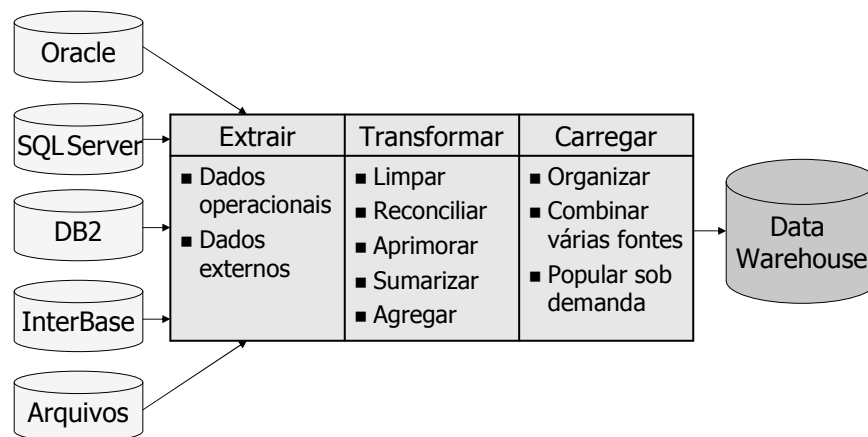
24

## Estrutura Interna



## Estrutura Interna

### ■ Obtenção de Dados



## Estrutura Interna

### ■ Busca de Informações



27

## Estrutura Interna

### ■ Modelo de Camadas



28

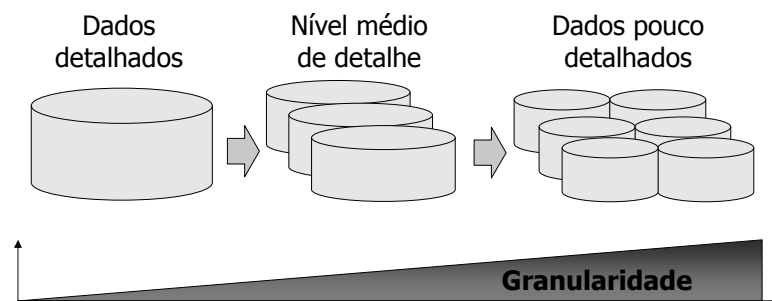
## Estrutura Interna

- Funções das Camadas do DW
  - Dados Operacionais/Externos: fontes de dados
  - Acesso aos Dados: extrair dados dos BDs
  - Data Staging: transformar e carregar dados
  - Data Warehouse Físico: armazenar dados
  - Acesso aos Dados: localizar dados para análise
  - Acesso à Informação: analisar dados
  - Troca de Mensagens: transportar dados
  - Gerenc. de Processos: controlar atividades

29

## Granularidade

- Granularidade
  - Nível de detalhe dos dados
  - De extrema importância no projeto do DW



30

## **Granularidade**

- Definir a granularidade adequada é vital para que o DW atenda seus objetivos
  - Mais detalhes → Mais dados → Análise mais longa → Informação mais detalhada
  - Menos detalhes → Menos dados → Análise mais curta → Informação menos detalhada
- Para evitar que se perca informação são criados vários níveis de granularidade

31

## **Granularidade**

- Dados x Granularidade
  - Dados Atuais
    - Refletem acontecimentos recentes
    - Alto nível de detalhe (baixa granularidade)
  - Dados Sumarizados (1 ou + níveis)
    - Dados históricos condensados
    - Menor nível de detalhe (maior granularidade)
  - Dados Antigos
    - Dados históricos mantidos em fita, CD, etc
    - Alto nível de detalhe (baixa granularidade)

32

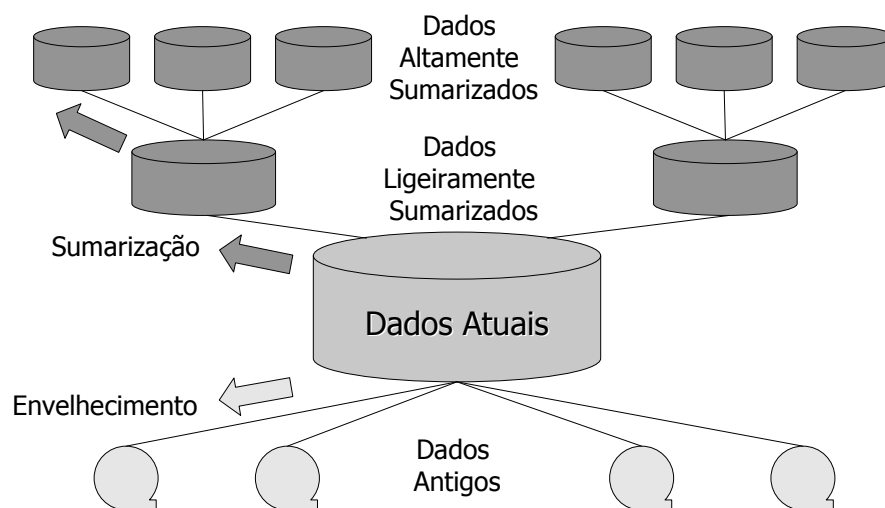


## Granularidade

- Processo de sumarização
  - Aplica um novo esquema de modo a condensar os dados
  - Ex.: armazenar totais, médias, etc.
- Processo de envelhecimento
  - Transfere os dados antigos do HD para fita, CD, etc.
  - Mantém o nível de detalhe para que nenhuma informação seja perdida

33

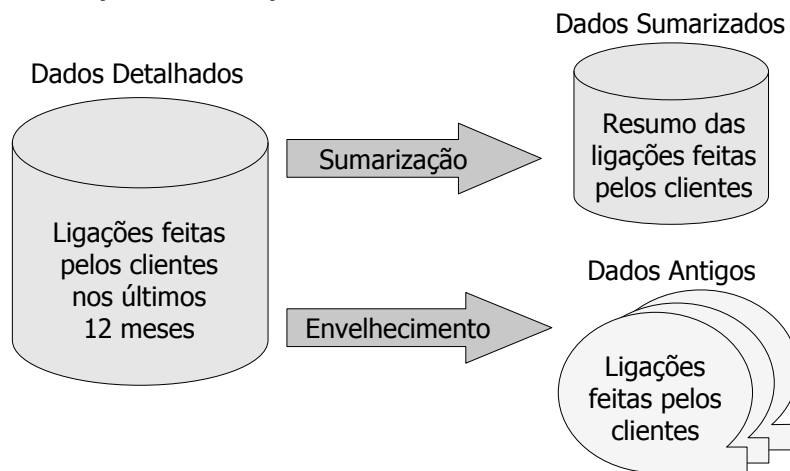
## Granularidade



34

# Granularidade

## ■ Exemplo: Companhia Telefônica [Inmon]



35

# Granularidade

## ■ Exemplo: Companhia Telefônica (cont.)

| Dados Detalhados                                                                                                                                                                        | Dados Sumarizados                                   | Dados Antigos                                          |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|--------------------------------------------------------|---------|--------|-----|--------|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|---------|-----|--------|-----------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|--------|---------|--------|-----|--------|--------|
| <table><tr><th>Ligações</th></tr><tr><td>Origem</td></tr><tr><td>Destino</td></tr><tr><td>Início</td></tr><tr><td>Fim</td></tr><tr><td>Tarifa</td></tr><tr><td>Status</td></tr></table> | Ligações                                            | Origem                                                 | Destino | Início | Fim | Tarifa | Status | <table><tr><th>Ligações</th></tr><tr><td>Cliente</td></tr><tr><td>Mês</td></tr><tr><td>Pulsos</td></tr><tr><td>LongaDist</td></tr><tr><td>ValConta</td></tr></table> | Ligações | Cliente | Mês | Pulsos | LongaDist | ValConta | <table><tr><th>Ligações</th></tr><tr><td>Origem</td></tr><tr><td>Destino</td></tr><tr><td>Início</td></tr><tr><td>Fim</td></tr><tr><td>Tarifa</td></tr><tr><td>Status</td></tr></table> | Ligações | Origem | Destino | Início | Fim | Tarifa | Status |
| Ligações                                                                                                                                                                                |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Origem                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Destino                                                                                                                                                                                 |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Início                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Fim                                                                                                                                                                                     |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Tarifa                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Status                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Ligações                                                                                                                                                                                |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Cliente                                                                                                                                                                                 |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Mês                                                                                                                                                                                     |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Pulsos                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| LongaDist                                                                                                                                                                               |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| ValConta                                                                                                                                                                                |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Ligações                                                                                                                                                                                |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Origem                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Destino                                                                                                                                                                                 |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Início                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Fim                                                                                                                                                                                     |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Tarifa                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Status                                                                                                                                                                                  |                                                     |                                                        |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |
| Nº de registros:<br>ligações nos<br>últimos 12 meses                                                                                                                                    | Nº de registros:<br>contas emitidas<br>pela empresa | Nº de registros:<br>ligações efetuadas<br>pela empresa |         |        |     |        |        |                                                                                                                                                                      |          |         |     |        |           |          |                                                                                                                                                                                         |          |        |         |        |     |        |        |

36

## Granularidade

- Quanto menor a granularidade, mais detalhada é a informação disponível
  - No exemplo anterior, poderíamos determinar se o cliente A ligou para B na semana passada
  - Também poderíamos verificar se A faz muitas chamadas de longa distância
- Durante o processo de sumarização, algumas informações podem ser perdidas
  - Não seria possível saber se A ligou para B
  - É possível verificar o padrão de consumo de A

37

## Data Marts

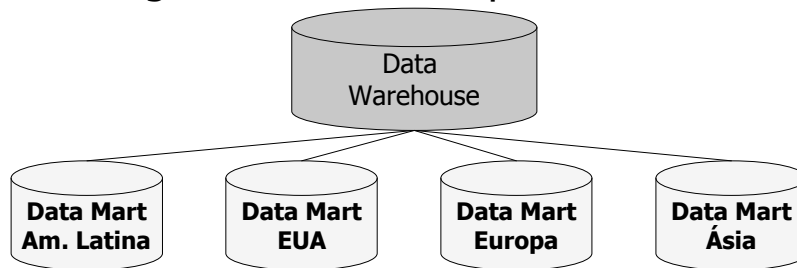
- Dados mantidos no DW são separados por assunto em subconjuntos de acordo com:
  - A estrutura interna da empresa
  - O processo de tomada de decisão
- Estes subconjuntos dos dados são chamados de Data Marts



38

## Data Marts

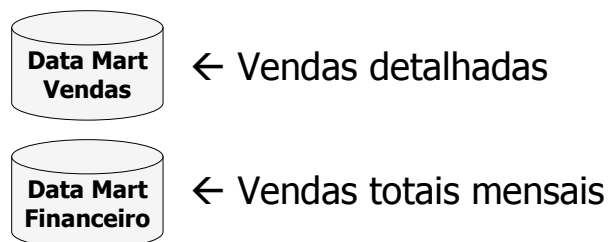
- Um Data Mart desempenha o papel de um DW departamental, regional ou funcional
- Uma empresa pode construir seus Data Marts gradativamente a partir do DW



39

## Data Marts

- Dados podem ser repetidos em dois ou mais Data Marts
- Os mesmos dados podem estar representados com granularidade diferente
- Ex:



40

## **Metadados**

- Os Metadados são dados sobre os dados
  - Para cada atributo mantido no DW há uma entrada no dicionário de dados
  - Os dados são processados, atualizados e consultados partindo dos metadados
  - Usuários ficam conhecendo a estrutura e o significado dos dados
  - No BD operacional, a estrutura e o significado dos dados estão embutidos nas aplicações

41

## **Metadados**

- Camadas de Metadados
  - Metadados Operacionais
    - Definem a estrutura dos dados operacionais
  - Metadados do DW
    - Orientados por assunto
    - Informam como os dados do DW foram calculados e como devem ser interpretados
  - Metadados do Usuário
    - Organizam os metadados do DW com base em conceitos familiares ao usuário final

42

## **Metadados**

- Classificação em função dos dados descritos
  - Metadados de Mapeamento
    - Como BDs operacionais são mapeados no DW
  - Metadados de Sumarização
    - Como os dados foram sumarizados no DW
  - Metadados Históricos
    - Como a estrutura dos dados vem mudando
  - Metadados de Padrões de Acesso
    - Como os dados do DW vem sendo acessados
  - Metadados de Miscelânea

43

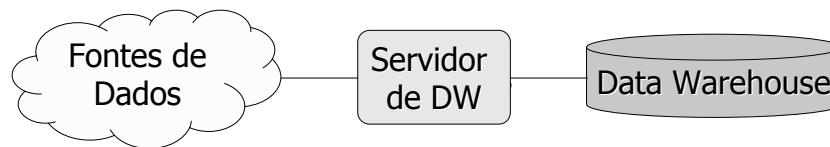
## **Metadados**

- Fontes de Metadados
  - Código fonte dos SBDs operacionais
  - Diagramas CASE de BDs operacionais e do DW
  - Documentação dos BDs operacionais e do DW
  - Entrevistas com usuários, administradores e programadores dos BDs e do DW
  - O ambiente de DW
    - Frequência de acesso aos dados, tempo de resposta, controle de usuários, etc.

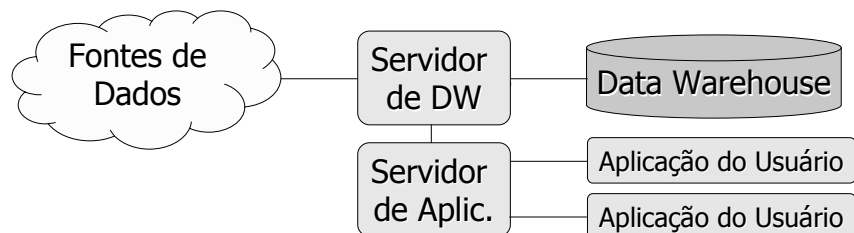
44

## Acesso aos Dados

### ■ Acesso em Duas Camadas



### ■ Acesso em Três Camadas



45

## Tipos de Data Warehouse

- DW baseado em Servidor
  - Mainframe ou servidor de rede local (LAN)
- DW Virtual
  - Reúne dados operacionais e dados históricos mantidos em BDs – não há um DW central
- DW Distribuído
  - DW global reúne dados de vários DWs locais
- DW baseado na Web
  - Dados provenientes da World Wide Web

46