

Centro de Informática
UFPE

Banco de Dados

Sistemas de DW (Data Warehouse)

Valéria Cesário Times
vct@cin.ufpe.br

Cin/UFPE - DW 1

Centro de Informática
UFPE

Tópicos

- Introdução
 - Sistemas de Suporte à Decisão
 - Sistemas de DW
 - Arquitetura de SDW
 - Sistemas OLAP
- Modelo Multidimensional
- Operações OLAP

Cin/UFPE - DW 2

Centro de Informática
UFPE

Tópicos

- Modelo MultiDim
 - Notação gráfica
 - Hierarquias
 - Granularidade
 - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
- Data Marts
- Data Warehousing

Cin/UFPE - DW 3

Centro de Informática
UFPE

Sistemas de Suporte à Decisão

- O que são Sistemas de Suporte a Decisão(SSD)?
 - Permitem armazenar e analisar grandes BD para extrair informações que auxiliam a compreensão do comportamento dos dados sobre o negócio
 - Exemplo: Sistemas de Data Warehouse (SDW)
 - Fornecem apoio ao suporte estratégico de decisão
 - Vantagem competitiva e aumento da produtividade na tomada de decisão
- Consistem em sistemas computacionais de suporte à decisão que integram dados oriundos de diversas fontes de dados
- São usados por gerentes de negócio e gestores

Cin/UFPE - DW 4

Centro de Informática
UFPE

Sistemas de DW

- Por quê surgiram Sistemas de DW?
 - Necessidade de gerenciamento das atividades de negócio das organizações
 - Organizações acumularam uma grande quantidade de dados ao longo do tempo
 - Necessidade de se fazer o cruzamento dos dados
 - Surgimento de novos meios de armazenamento
 - Organizações sentiram a necessidade de avaliar tendências e o comportamento dos dados
 - Necessidade de manter não apenas dados atuais, mas também dados históricos
 - Necessidade de uma ferramenta de BD para SD

Cin/UFPE - DW 5

Centro de Informática
UFPE

Sistemas de DW

- Provêem diferentes níveis de análise
 - São chamados de sistemas analíticos
- Permitem que usuários naveguem nos diferentes níveis de detalhes sobre os dados
 - Dados são organizados por meio de modelos multidimensionais
 - Resultados de consultas são interpretados em uma variedade de visões multidimensionais
 - Consultas são providas pelas ferramentas OLAP (*On-Line Analytical Processing*)

Cin/UFPE - DW 6

Arquitetura de SDW

• Arquitetura em camadas de um Sistema de DW:

Fontes de dados: Fontes Internas, Fontes Externas, Dados Operacionais, ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP (Geração de Relatórios, Análises Estatísticas, Mineração de Dados).

Preparação de dados: ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

Data Warehouse: Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

OLAP: Servidor OLAP, Ferramentas OLAP.

Aplicação do Usuário: Ferramentas OLAP.

Cln/UFPE - DW © 7

Arquitetura de SDW

• Arquitetura em camadas de um Sistema de DW:

Fontes de dados: Fontes Internas, Fontes Externas, Dados Operacionais, ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP (Geração de Relatórios, Análises Estatísticas, Mineração de Dados).

Preparação de dados: ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

Data Warehouse: Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

OLAP: Servidor OLAP, Ferramentas OLAP.

Aplicação do Usuário: Ferramentas OLAP.

Cln/UFPE - DW © 8

Arquitetura de SDW

• Arquitetura em camadas de um Sistema de DW:

Fontes de dados: Fontes Internas, Fontes Externas, Dados Operacionais, ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP (Geração de Relatórios, Análises Estatísticas, Mineração de Dados).

Preparação de dados: ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

Data Warehouse: Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

OLAP: Servidor OLAP, Ferramentas OLAP.

Aplicação do Usuário: Ferramentas OLAP.

Cln/UFPE - DW © 9

Arquitetura de SDW

• Arquitetura em camadas de um Sistema de DW:

Fontes de dados: Fontes Internas, Fontes Externas, Dados Operacionais, ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP (Geração de Relatórios, Análises Estatísticas, Mineração de Dados).

Preparação de dados: ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

Data Warehouse: Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

OLAP: Servidor OLAP, Ferramentas OLAP.

Aplicação do Usuário: Ferramentas OLAP.

Cln/UFPE - DW © 10

Arquitetura de SDW

• Arquitetura em camadas de um Sistema de DW:

Fontes de dados: Fontes Internas, Fontes Externas, Dados Operacionais, ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP (Geração de Relatórios, Análises Estatísticas, Mineração de Dados).

Preparação de dados: ODS, Processos de ETL, Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

Data Warehouse: Metadados, DW, Data Marts, Servidor OLAP, Ferramentas OLAP.

OLAP: Servidor OLAP, Ferramentas OLAP.

Aplicação do Usuário: Ferramentas OLAP.

Cln/UFPE - DW © 11

Arquitetura de SDW

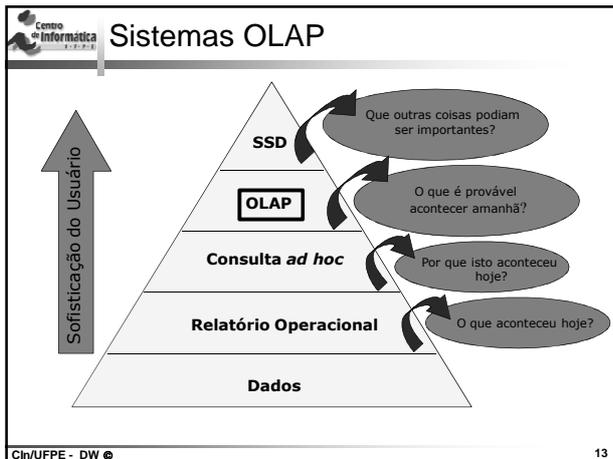
• Ferramentas de ETL

- Responsáveis pela conversão dos dados do ambiente operacional para o de suporte à decisão
- Realizam Acesso, Extração, Transformação, Validação e Carga dos dados
- Frequência da carga do DW e de sua integração com ODS depende da aplicação

• Operational Datastore (ODS)

- Repositório de dados operacionais integrados
- Benefícios
 - Otimiza a criação do DW
 - Possibilita a realização de consultas relacionais sobre dados históricos

Cln/UFPE - DW © 12



Sistemas OLAP

- Importância da análise de dados tem crescido desde os anos 90
 - Em todos os setores, empresas precisam melhorar seus processos de tomada de decisão → competitivas
 - Sistemas de BD tradicionais não satisfazem os requisitos de análise de dados
 - São projetados para registrar as operações do dia-a-dia
 - Garantem o acesso concorrente e rápido aos dados

processamento de transações, controle de concorrência e recuperação após falhas

- Mantêm a consistência dos dados por meio do gerenciamento de transações
- São chamados de BD operacionais ou sistemas OLTP

CIn/UFPE - DW © 14

Sistemas OLAP

- Sistema OLTP (*Online Transaction Processing*)
 - Gerencia a execução de milhares de transações e por isso, seu projeto deve prevenir anomalias de atualização
 - BD de sistemas OLTP são altamente normalizados
 - Exibe baixo desempenho na execução de consultas complexas que requerem
 - a junção de várias tabelas relacionais ou
 - a agregação de um grande volume de dados
 - Suas BD contêm dados detalhados e não possuem dados históricos

Mostra a necessidade de um novo paradigma: Sistemas OLAP!

CIn/UFPE - DW © 15

Sistemas OLAP

- O que é OLAP (*On-Line Analytical Processing*)?
 - Conjunto de tecnologias projetadas para analisar e acessar dados típicos de suporte a decisão que estão no DW
 - Fornece dados em alto nível (totais, médias, min...)
 - Acessa vários registros
 - Tem alto desempenho e consultas fáceis e interativas
 - Lida com dados históricos (dimensão temporal)
 - Oferece visões multidimensionais (perspectivas)

CIn/UFPE - DW © 16

Sistemas OLAP

- Sistemas OLAP
 - Projetados para análise gerencial
 - Priorizam facilidade de consultas e de navegação na estrutura dimensional
 - Dados são resumidos, históricos e baseados em assuntos
 - Consultas OLAP são geralmente *ad-hoc* e acessam grandes conjuntos de dados
 - Atualizações só em batch e por isso não há controle de concorrência
 - Dados são não normalizados já que o foco do OLAP está na recuperação e visualização e não no armazenamento

CIn/UFPE - DW © 17

Sistemas OLAP

- Exemplos de consultas típicas de OLAP:
 - Quais os produtos mais vendidos no mês passado?
 - Quais os 10 piores vendedores dos departamentos da filial Recife?
 - Qual a média salarial dos funcionários de Informática na região sul nos últimos 5 anos?
- OLAP implica em
 - Ordenar e classificar valores
 - Comparar valores entre períodos de tempo
 - Obter percentual de variação
 - Obter médias, somas, valores acumulativos, funções estatísticas e financeiras.

CIn/UFPE - DW © 18

Sistemas OLAP

- Exemplo de Relatório de Sistemas OLAP

	Leste	Oeste	Sul
Produto1	50	80	100
Produto2	40	70	80
Produto3	90	120	140
Produto4	20	10	30

CIn/UFPE - DW © 19

Sistemas OLAP

- Exemplo de Relatório de Sistemas OLAP

Produto	Região	Vendas no Mês	Comparação com o Mês Anterior
Pasta Colgate	Sul	110	**12%
Pasta Colgate	Sudeste	179	-3%
Pasta Colgate	Nordeste	55	5%
Total		344	**6%

CIn/UFPE - DW © 20

Sistemas OLAP

- Exemplo de Relatório de Sistemas OLAP

Produto	Região	Tamanho	Vendas no Mês	Comp. Com o Mês Anterior
Colgate	Sul	A	34	**10%
Colgate	Sul	B	36	**13%
Colgate	Sul	C	40	**11%
Colgate	Total		110	**12%
Colgate	Sudeste	A	63	-2.8%
Colgate	Sudeste	B	60	-3.1%
Colgate	Sudeste	C	56	-2.9%
Colgate	Total		179	-3%
Colgate	Nordeste	A	19	5%
Colgate	Nordeste	B	17	4%
Colgate	Nordeste	C	19	6%
Colgate	Total		55	5%
Total			344	6%

CIn/UFPE - DW © 21

Sistemas OLAP

- Exemplo de Relatório de Sistemas OLAP

Produto	Região	Equipe de Vendas	Vendas no Mês	Comp. Com o Mês Anterior
Colgate	Sul	P. Alegre	52	**21%
Colgate	Sul	Curitiba	28	5%
Colgate	Sul	Florianópolis	30	6%
Colgate	Total		110	**12%
Colgate	Sudeste	São Paulo	93	4%
Colgate	Sudeste	Rio	75	5%
Colgate	Sudeste	Belo Horiz.	11	-1.5%
Colgate	Total		179	-3%
Colgate	Nordeste	Salvador	21	5%
Colgate	Nordeste	Fortaleza	18	4%
Colgate	Nordeste	Recife	16	6%
Colgate	Total		55	5%
Total			344	6%

CIn/UFPE - DW © 22

Sistemas OLAP

- Exemplo de Relatório de Sistemas OLAP

Relatório mensal de vendas do nordeste – Maio/1999

Estado	Cidade	Un. Vendidas	Vendas \$
PE	Olinda	2.500	\$12.850
PE	Recife	2.750	\$14.135
PE Totais			
CE	Fortaleza	3.200	\$16.800
CE	Crato	1.725	\$ 9.143
CE Totais			
AL	Maceió	1.900	\$ 9.595
AL Totais			
Região Nordeste Total		12.075	\$62.523

CIn/UFPE - DW © 23

Sistemas OLAP

- Voltados para execução de consultas analíticas
 - BD de sistemas OLAP deve permitir uma carga elevada de consultas
- Exemplos típicos de consultas OLAP incluem:
 - Total da quantidade vendida por produto e por cliente
 - Produtos mais vendidos por cliente

Tais consultas envolvem agregação!

Em geral, requerem a varredura de todos os registros de uma tabela do BD!

CIn/UFPE - DW © 24

Sistemas OLAP

- Novas técnicas de otimização de consultas e indexação são necessárias para OLAP
- Normalização prejudica a execução de consultas OLAP já que ela particiona o BD em muitas tabelas
 - Reconstrução dos dados requer um alto número de junções entre milhares de tuplas

↑
Surgiu a necessidade de um novo modelo de BD:
modelo multidimensional!

- Isso levou ao surgimento de DW que se baseia no modelo multidimensional para permitir OLAP

CIn/UFPE - DW © 25

Tópicos

- Introdução
 - Sistemas de Suporte à Decisão
 - Sistemas de DW
 - Arquitetura de SDW
 - Sistemas OLAP
- Modelo Multidimensional
- Operações OLAP

CIn/UFPE - DW © 26

Modelo Multidimensional

- O que é um Data Warehouse (DW)?

“Coleção de dados orientada a assunto, integrada, não-volátil e variante no tempo, utilizada para tomada de decisões”.

- W. H. Inmon

“uma cópia dos dados de transação especificamente estruturados para consultas e análises”

R. Kimball

CIn/UFPE - DW © 27

Modelo Multidimensional

- O que é um DW?

“Repositório estruturado e corporativo de dados orientados ao assunto, variantes no tempo e históricos, usados para recuperação de informações e suporte à decisão. O DW armazena dados atômicos e sumarizados”.

- Definição de DW da Oracle

CIn/UFPE - DW © 28

Modelo Multidimensional

- Definição de DW
 - É uma base de dados que facilita a execução de consultas de apoio à decisão
- Propriedades de um DW

CIn/UFPE - DW © 29

Modelo Multidimensional

- Propriedades de um DW: Orientada ao assunto
 - Permite tomar decisões sobre o *assunto* - tema dos dados
 - Dados são divididos e armazenados por áreas de negócio
 - Satisfaz as necessidades de análise de diferentes áreas de uma organização
 - Áreas dependem do tipo de atividades executadas pela empresa

CIn/UFPE - DW © 30

Modelo Multidimensional

- Propriedades de um DW: Integrada
 - Consolida dados de diferentes fontes heterogêneas
 - Dados de um determinado assunto são definidos e armazenados apenas uma vez

Aplicações OLTP: Poupança, Contas, Empréstimos. Data Warehouse: Cliente.

CIn/UFPE - DW © 31

Modelo Multidimensional

- Propriedades de um DW: Variante no tempo
 - Mantém um histórico dos dados, permitindo comparações ao longo do tempo
 - Diferentes valores para a mesma informação são mantidas e registra-se o tempo de atualização destes valores

Data Warehouse

CIn/UFPE - DW © 32

Modelo Multidimensional

- Propriedades de um DW: Não volátil
 - Informações já presentes no BD são raramente modificadas (sobrescritas)
 - Novos dados são carregados no sistema, integrando-se com informações já armazenadas previamente
 - Durabilidade é garantida porque dados não são atualizados nem removidos

Operacional: Inserção, Atualização, Remoção e/ou Leitura. Warehouse: Leitura.

CIn/UFPE - DW © 33

Modelo Multidimensional

- Carga de Dados em um DW
 - Primeira Carga
 - Carga Incremental
 - Carga Incremental
 - Carga Incremental
 - Eliminar ou Arquivar

Bancos de Dados Operacionais, Data Warehouse.

CIn/UFPE - DW © 34

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tipo de Usuário	Operadores e técnicos administrativos	Gerentes e executivos
Uso do Sistema	Pré-definido e repetitivo	Ad hoc e não estruturado
Conteúdo dos dados	Atual e dados detalhados (menos volumosos)	Histórico e dados resumidos (mais volumosos)
Organização dos dados	Necessidades do ambiente operacional	Necessidades de análise

CIn/UFPE - DW © 35

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tipo de Usuário	Operadores e técnicos administrativos	Gerentes e executivos
Uso do Sistema	Pré-definido e repetitivo	Ad hoc e não estruturado
Conteúdo dos dados	Atual e dados detalhados (menos volumosos)	Histórico e dados resumidos (mais volumosos)
Organização dos dados	Necessidades do ambiente operacional	Necessidades de análise

CIn/UFPE - DW © 36

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tipo de Usuário	Operadores e técnicos administrativos	Gerentes e executivos
Uso do Sistema	Pré-definido e repetitivo	Ad hoc e não estruturado
Conteúdo dos dados	Atual e dados detalhados (menos volumosos)	Histórico e dados resumidos (mais volumosos)
Organização dos dados	Necessidades do ambiente operacional	Necessidades de análise

CIn/UFPE - DW © 37

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tipo de Usuário	Operadores e técnicos administrativos	Gerentes e executivos
Uso do Sistema	Pré-definido e repetitivo	Ad hoc e não estruturado
Conteúdo dos dados	Atual e dados detalhados (menos volumosos)	Histórico e dados resumidos (mais volumosos)
Organização dos dados	Necessidades do ambiente operacional	Necessidades de análise

CIn/UFPE - DW © 38

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Estruturas de dados	Otimizadas para pequenas transações	Otimizadas para consultas complexas
Frequência de acesso	Alta	De média à baixa
Tipo de Acesso	Leitura, inserção, atualização e remoção	Leitura e carga incremental
Número de Registros / acesso	Poucos (Dezenas)	Muitos (Milhares)

CIn/UFPE - DW © 39

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Estruturas de dados	Otimizadas para pequenas transações	Otimizadas para consultas complexas
Frequência de acesso	Alta	De média à baixa
Tipo de Acesso	Leitura, inserção, atualização e remoção	Leitura e carga incremental
Número de Registros / acesso	Poucos (Dezenas)	Muitos (Milhares)

CIn/UFPE - DW © 40

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Estruturas de dados	Otimizadas para pequenas transações	Otimizadas para consultas complexas
Frequência de acesso	Alta	De média à baixa
Tipo de Acesso	Leitura, inserção, atualização e remoção	Leitura e carga incremental
Número de Registros / acesso	Poucos (Dezenas)	Muitos (Milhares)

CIn/UFPE - DW © 41

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Estruturas de dados	Otimizadas para pequenas transações	Otimizadas para consultas complexas
Frequência de acesso	Alta	De média à baixa
Tipo de Acesso	Leitura, inserção, atualização e remoção	Leitura e carga incremental
Número de Registros / acesso	Poucos (Dezenas)	Muitos (Milhares)

CIn/UFPE - DW © 42

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tempo de resposta	Baixo	Pode ser maior
Nível de concorrência	Alto	Baixo
Uso de bloqueio	Necessário para evitar anomalias de atualização	Não necessário
Frequência de atualização	Alta	Nenhuma

CIn/UFPE - DW © 43

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tempo de resposta	Baixo	Pode ser maior
Nível de concorrência	Alto	Baixo
Uso de bloqueio	Necessário para evitar anomalias de atualização	Não necessário
Frequência de atualização	Alta	Nenhuma

CIn/UFPE - DW © 44

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tempo de resposta	Baixo	Pode ser maior
Nível de concorrência	Alto	Baixo
Uso de bloqueio	Necessário para evitar anomalias de atualização	Não necessário
Frequência de atualização	Alta	Nenhuma

CIn/UFPE - DW © 45

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Tempo de resposta	Baixo	Pode ser maior
Nível de concorrência	Alto	Baixo
Uso de bloqueio	Necessário para evitar anomalias de atualização	Não necessário
Frequência de atualização	Alta	Nenhuma

CIn/UFPE - DW © 46

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Redundância de dados	Baixa (tabelas normalizadas)	Alta (tabelas desnormalizadas)
Modelagem de dados	Modelo UML ou ER	Modelagem multidimensional
Número de usuários	Milhares	Dezenas
Objetivo	Processamento de transações	Consultas Analíticas

CIn/UFPE - DW © 47

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Redundância de dados	Baixa (tabelas normalizadas)	Alta (tabelas desnormalizadas)
Modelagem de dados	Modelo UML ou ER	Modelagem multidimensional
Número de usuários	Milhares	Dezenas
Objetivo	Processamento de transações	Consultas Analíticas

CIn/UFPE - DW © 48

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Redundância de dados	Baixa (tabelas normalizadas)	Alta (tabelas desnormalizadas)
Modelagem de dados	Modelo UML ou ER	Modelagem multidimensional
Número de usuários	Milhares	Dezenas
Objetivo	Processamento de transações	Consultas Analíticas

CIn/UFPE - DW © 49

Modelo Multidimensional

- Comparação entre BD operacionais e DW :

	BD Operacional	Data Warehouse
Redundância de dados	Baixa (tabelas normalizadas)	Alta (tabelas desnormalizadas)
Modelagem de dados	Modelo UML ou ER	Modelagem multidimensional
Número de usuários	Milhares	Dezenas
Objetivo	Processamento de transações	Consultas Analíticas

CIn/UFPE - DW © 50

Modelo Multidimensional

- DW diferem de BD operacionais (ou transacionais)
 - Operacional → orientado por transação
 - Ex: o produto vendido, o artigo publicado, o imposto predial pago
 - Muitas das consultas feitas são previstas a priori
 - Decisão → orientado por assunto
 - Ex: Vendas, Publicações de Artigos, Gerenciamento Urbano
 - Consultas mudam continuamente
- Dividir os dados em dois ambientes ortogonais
 - Operacional (normalmente já existente)
 - Estratégico (construir um DW)

CIn/UFPE - DW © 51

Modelo Multidimensional

- Data Warehouse (DW):
 - É um repositório volumoso de dados que consolida dados oriundos de diversas fontes
 - Dados internos e externos à organização
 - É geralmente atualizado *off-line*
 - Provê suporte a vários tipos de análises:
 - Mineração de dados
 - Análise estatística
 - Consultas OLAP
 - Adota o modelo de dados multidimensional
 - Permite a visualização dos dados em um espaço n-dimensional chamado de cubo de dados ou hipercubo

CIn/UFPE - DW © 52

Modelo Multidimensional

- Cubo de dados

Produto	Loja	Vendas
COPO	F01	50
COPO	F02	60
COPO	F03	100
BOLA	F01	40
BOLA	F02	70
BOLA	F03	80
JOGO	F01	90
JOGO	F02	120
JOGO	F03	140
VELA	F01	20
VELA	F02	10
VELA	F03	30

Produto	Loja	Tempo	Vendas
COPO	F01	1999	50
COPO	F02	1999	60
COPO	F03	1999	100
BOLA	F01	1999	40
BOLA	F02	1999	70
BOLA	F03	1999	80
JOGO	F01	1999	90
JOGO	F02	1999	120
JOGO	F03	1999	140
VELA	F01	1999	20
VELA	F02	1999	10
VELA	F03	1999	30
COPO	F01	2000	50
COPO	F02	2000	60
COPO	F03	2000	100
BOLA	F01	2000	40
BOLA	F02	2000	70
BOLA	F03	2000	80
JOGO	F01	2000	90
JOGO	F02	2000	120
JOGO	F03	2000	140
VELA	F01	2000	20
VELA	F02	2000	10
VELA	F03	2000	30

CIn/UFPE - DW © 53

Modelo Multidimensional

- Cubo de Dados

Tempo (trimestre)	Produto (categoria)	Clientes (cidade)	Valores de Medida
Q1	Beverages	Lyon	21
		Paris	10
		Berlin	19
Q2	Seafood	Lyon	14
		Paris	11
		Berlin	30
Q3	Condiments	Lyon	26
		Paris	12
		Berlin	35
Q4	Produce	Lyon	14
		Paris	20
		Berlin	47

- Cubo de Dados é definido por dimensões e fatos
 - Dimensões são perspectivas usadas para analisar os dados
 - Exemplo de três dimensões: *Produto*, *Tempo* e *Cliente*.
- Nível de dimensão define a granularidade ou o nível de detalhe em que medidas são representadas para cada dimensão do cubo
 - Quantidade vendida é agregada nos níveis: *categoria*, *trimestre* e *cidade*

CIn/UFPE - DW © 54

Modelo Multidimensional

- Cubo de Dados
 - Instâncias de uma dimensão são membros
 - Condiments* e *Seafood* são exemplos de membros da dimensão *Produto* no nível *Categoria*
 - Dimensão pode também ter atributos que a descrevem
 - A dimensão *Produto* pode ter como atributos (propriedades):
 - Número do produto e Preço Unitário
 - Células do cubo são chamadas de fatos
 - São instâncias de atributos chamados de medidas
 - Medidas são usadas para avaliar quantitativamente vários aspectos de interesse numa análise
 - Quantidade: indica o número de itens vendidos (em milhares) por categoria, trimestre e cidade de cliente
 - Um cubo de dados geralmente possui várias medidas
 - Outro exemplo é o *Total_Vendido*: indica o valor total das vendas

CIn/UFPE - DW 55

Modelo Multidimensional

Níveis		Tempo		Propriedades	
Dia 1	Dia 2	Trimestre 1	Trimestre 2	Início	Fim
Jan. 1998	Fev. 1998	Trimestre 1	Trimestre 2	1 de Julho	30 de Setembro
				1 de Outubro	31 de Dezembro
				1 de Janeiro	31 de Março
				1 de Abril	30 de Junho

CIn/UFPE - DW 56

Modelo Multidimensional

- Exemplo de Cubo de Dados

CIn/UFPE - DW 57

Modelo Multidimensional

- Cubo de dados (Abstração do DW)
 - Abordagem multidimensional para visualização e organização dos dados
 - Várias dimensões podem ser consultadas simultaneamente
 - Dados são manipulados mais rapidamente e facilmente (agregação em níveis)

CIn/UFPE - DW 58

Modelo Multidimensional

- Cubo de Dados
 - Pode ser denso ou esparso
 - Depende de se ele tem valores de medida para cada combinação de valores de dimensão (combinação de membros)
 - Por exemplo:
 - Denso: se todos os produtos forem comprados por todos os clientes durante o período de tempo considerado
 - Esparso: se nem todos os clientes compraram produtos de todas as categorias durante todos os trimestres do ano
 - Em aplicações reais, cubos são geralmente esparsos
 - Tem uma granularidade que é determinada pela combinação de níveis de cada eixo do cubo
 - Para extrair conhecimento estratégico de um cubo, é necessário visualizar seus dados em vários níveis de detalhe
 - Nível de menor granularidade: quantidade vendida por mês
 - Nível de maior granularidade: quantidade vendida por ano

CIn/UFPE - DW 59

Modelo Multidimensional

- Hierarquias
 - Permite a definição de uma sequência de mapeamentos entre níveis menores ou de menor granularidade e níveis maiores ou de maior granularidade
 - Níveis menores representam conceitos mais detalhados
 - Níveis maiores representam conceitos mais genéricos
 - Dados dois níveis de uma hierarquia:
 - Menor nível é chamado de filho
 - Maior nível é chamado de pai

CIn/UFPE - DW 60

Modelo Multidimensional

- Hierarquias
 - Estrutura hierárquica de uma dimensão é chamada de esquema da dimensão
 - Conjunto de membros de todos os níveis de uma dimensão é chamado de instância da dimensão
 - Membros da hierarquia: Produto → Categoria

- Nível All possui um único membro, chamado all, que é usado para obter a agregação de medidas para toda a hierarquia
 - Por exemplo: para obter a quantidade vendida de todos os produtos

CIn/UFPE - DW 61

Modelo Multidimensional

- Agregação em Níveis de Hierarquias

CIn/UFPE - DW 62

Modelo Multidimensional

- Medidas
 - Cada medida do cubo está associada a uma função de agregação que mapeia vários valores de medida em um único valor
 - Agregação de medidas ocorre quando o usuário muda o nível de detalhe pelo qual os dados do cubo são visualizados
 - Navegação nas hierarquias das dimensões
 - Mudando de *Cidade* para *País*: as quantidades vendidas para todos os clientes de um mesmo país serão agregadas (e.g. SUM)
 - Obtendo o total de vendas: um cubo com uma única célula mostrará a soma das vendas de todos os produtos

Isto corresponde a visualizar o cubo no nível All de todas as hierarquias de dimensão!

CIn/UFPE - DW 63

Modelo Multidimensional

- Medidas
 - Sumarização: agregação correta de medidas do cubo ao longo das hierarquias de dimensão para obter resultados de agregação consistentes
 - Para garantir a sumarização, as seguintes condições devem ser satisfeitas
 - Disjunção de instâncias
 - Compleitude
 - Corretude

CIn/UFPE - DW 64

Modelo Multidimensional

- Medidas
 - Disjunção de instâncias: agrupamento de instâncias de um nível com relação aos seus pais do nível acima deve resultar em subconjuntos disjuntos de instâncias
 - Por exemplo: Um produto não pode pertencer a duas categorias
 - Caso contrário, as quantidades vendidas do produto seriam contadas duas vezes, uma para cada categoria
 - Compleitude: Todas as instâncias devem ser incluídas na hierarquia e cada instância deve ter um só pai do nível acima
 - Por exemplo: *Tempo* deve conter todos os dias do período de interesse e cada dia deve pertencer a um só mês
 - Caso contrário, haveria datas cujas quantidades vendidas não seriam contadas

CIn/UFPE - DW 65

Modelo Multidimensional

- Medidas
 - Corretude: refere-se ao uso correto de funções de agregação
 - Medidas podem ser de vários tipos, e isto define o tipo de função de agregação que pode ser aplicado sobre elas
 - Para definir uma medida é necessário determinar as funções de agregação que serão aplicadas sobre ela nas dimensões
 - Isto é particularmente importante no caso de medidas semi-aditivas e não-aditivas.
 - De acordo com o modo pelo qual elas podem ser agregadas, medidas podem ser:
 - Aditivas
 - Semi-aditivas
 - Não-aditivas

CIn/UFPE - DW 66

Modelo Multidimensional

- Medidas
 - Aditivas: representam valores aditivos
 - Quantidade de ocorrências criminais por região e mês
 - Número de transplantes por ano
 - Quantidade Vendida: *Produto*, *Tempo* e *Cliente*
 - Podem ser sumarizadas ao longo de todas as dimensões por meio da adição
 - Referem-se a fatos sobre os quais podem ser aplicadas operações de soma ao longo de qualquer dimensão

CIn/UFPE - DW © 67

Modelo Multidimensional

- Medidas
 - Semi-aditivas: podem ser sumarizadas usando a adição por meio de *algumas*, mas não todas as dimensões
 - Exemplo: quantidade estocada: não pode ser agregada na dimensão *Tempo*
 - Não faz sentido somar a quantidade em estoque de dois semestres consecutivos
 - É possível obter a média (ou valores min e max) da quantidade estocada ao longo da dimensão *Tempo* e computar a soma ao longo de outras dimensões

CIn/UFPE - DW © 68

Modelo Multidimensional

- Medidas Semi-Aditivas
 - Referem-se a fatos que podem ser adicionados por meio de algumas dimensões apenas
 - Referem-se a fatos que envolvem a contagem dupla
 - Representam valores semi-aditivos ao longo do tempo:
 - Quantidade de produtos estocados
 - Saldo bancário

↑

Não faz sentido adicioná-los ao longo da dimensão temporal!

Pois a soma dos totais (em estoque ou no banco) em mais de um período contabilizaria mais de uma vez o fato (item estocado ou dinheiro)!

Mas faz sentido somá-los ao longo de outra dimensão (e.g. loja, cliente,....)?

CIn/UFPE - DW © 69

Modelo Multidimensional

- Medidas
 - Não-aditivas: não podem ser sumarizadas pela adição aplicada ao longo de qualquer dimensão
 - Exemplo: preço de produtos, nota escolar, custo por unidade e taxas de câmbio
 - Dependendo da semântica da aplicação, outras funções podem ser usadas para agregar valores de medidas não-aditivas:
 - COUNT
 - MIN
 - MAX

CIn/UFPE - DW © 70

Modelo Multidimensional

- Medidas Não-Aditivas:
 - Dependendo do contexto da aplicação e das necessidades do usuário
 - A soma das medidas pode fazer menos sentido do que a aplicação de uma outra função de agregação
 - Média
 - Mínimo ou Máximo
 - Desvio Padrão
 - Exemplos
 - Temperatura
 - Preços Unitários
 - Taxas

CIn/UFPE - DW © 71

Modelo Multidimensional

- Medidas Não-Aditivas:
 - Temperatura
 - » Não faz sentido dizer que a temperatura de 2008 foi 200 graus
 - » Faz sentido obter a temperatura média do ano
 - Preço Unitário

Preço de Venda do Milho	R\$
Janeiro	10,00
Fevereiro	11,00
Março	12,00
Abril	10,00
Mai	8,00
Junho	7,00

← Não faz sentido dizer que o preço de venda da mão de milho no primeiro semestre foi R\$ 58,00

CIn/UFPE - DW © 72

Modelo Multidimensional

- Medidas Não-Aditivas:
 - Temperatura
 - » Não faz sentido dizer que a temperatura de 2008 foi 200 graus
 - » Faz sentido obter a temperatura média do ano
 - Preço Unitário

Preço de Venda do Milho	R\$
Janeiro	10,00
Fevereiro	11,00
Março	12,00
Abril	10,00
Mai	8,00
Junho	7,00

Mas faz sentido:

- Preço médio do semestre = R\$ 9,66 ou
- Preço mínimo = R\$ 7,00

CIn/UFPE - DW 73

Modelo Multidimensional

- Medidas
 - Para permitir usuários explorar o cubo de dados eficientemente em diferentes granularidades, técnicas de otimização são usadas.
 - » São baseadas na precomputação de agregados
 - » Evita a computação de todos valores agregados cada vez que o DW é consultado
 - » Ferramentas OLAP implementam mecanismos de agregação incremental
 - » Contudo, dependendo do tipo de função de agregação usada, a computação da agregação incremental não é possível.
 - Exemplo: RANK

Isso levou ao surgimento de uma nova classificação de medidas!

CIn/UFPE - DW 74

Modelo Multidimensional

- Medidas
 - Distributivas: são definidas por uma função de agregação que pode ser computada de forma distribuída
 - » Dados são particionados em n conjuntos
 - » Função de agregação é aplicada a cada conjunto, resultando em n valores agregados
 - aplicação da função ao conjunto inteiro de dados produz o mesmo resultado de sua aplicação aos valores agregados
 - Exemplo: COUNT, SUM, MIN, MAX
 - DISTINCT COUNT não é distributiva

DISTINCT COUNT → {3, 3, 4, 5, 8, 4, 7, 3, 8} ⇒ 5

DISTINCT COUNT → {3, 3, 4} {5, 8, 4} {7, 3, 8} ⇒ 8

CIn/UFPE - DW 75

Modelo Multidimensional

- Medidas
 - Algébricas: função de agregação que pode ser expressa por meio de uma função escalar definida sobre funções distributivas
 - Exemplo: $AVG = SUM / COUNT$
 - Holísticas: não podem ser computadas a partir de subconjuntos de valores agregados
 - Exemplo: RANK, MEDIAN
 - São computacionalmente custosas

CIn/UFPE - DW 76

Modelo Multidimensional

- Medidas podem ser classificadas pelo tipo da função de agregação:
 - Acumulativas: o valor de medida de uma célula é obtido pela computação dos valores de outras células
 - Exemplos: SUM, COUNT, AVG
 - Baseadas em Filtro: filtram os membros de uma dimensão que são mostrados como resultado.
 - Exemplos: MIN, MAX
 - Diferem das demais funções de agregação porque computam não apenas o valor agregado mas determinam também os membros que possuem este valor
 - Para determinar o melhor funcionário, deve-se obter a quantidade vendida máxima mas também identificar quem foi o responsável por esse valor máximo

CIn/UFPE - DW 77

Tópicos

- Introdução
 - Sistemas de Suporte à Decisão
 - Sistemas de DW
 - Arquitetura de SDW
 - Sistemas OLAP
- Modelo Multidimensional
- Operações OLAP

CIn/UFPE - DW 78

Operações OLAP

- Modelo multidimensional permite a visualização de dados a partir de múltiplas perspectivas e vários níveis de detalhe
 - Provendo um ambiente de análise de dados interativo
- A materialização de perspectivas de análise e níveis de detalhe permite a navegação em dimensões e hierarquias
- Cubo com quantidades de vendas trimestrais por categoria de produto e cidade de cliente e para o ano de 2012

Cin/UFPE - DW 79

Operações OLAP

- Cubo com quantidades de vendas trimestrais por categoria de produto e cidade de cliente e para o ano de 2012
- Roll-up: agrega medidas ao longo da hierarquia de dimensão para obter valores de um nível mais genérico
 - Sintaxe da operação roll-up:
 - ROLLUP (Nome_Cubo, Dimensão → Nível, Função(Medida))
- roll-up para o nível país da dimensão Cliente:
 - ROLLUP (Vendas, Cliente → País, SUM(Quantidade))

Cin/UFPE - DW 80

Operações OLAP

- Drill-down:
 - Navegação de um nível da hierarquia mais genérico para um nível mais detalhado (oposto de roll-up)
 - Sintaxe da operação drill-down:
 - DRILLDOWN (Nome_Cubo, Dimensão → Nível)
 - drill-down para o nível mês da dimensão Tempo:
 - DRILLDOWN (Vendas, Tempo → Mês)

Cin/UFPE - DW 81

Operações OLAP

- Drill Down e Roll Up (ou Drill Up)
 - Possibilita caminhar pela estrutura multidimensional (hierarquias), permitindo ter diferentes visões de dados
 - Navegação entre os níveis de detalhamento dos dados
 - Modifica o nível de granularidade da consulta

Cin/UFPE - DW 82

Operações OLAP

- Drill Down e Roll Up (ou Drill Up)

Loja	Tipo	Vendas
F01	Luxo	50
	Popular	40
	Padrão	90
F02	Franquia	20
	Luxo	60
	Popular	70
F03	Padrão	120
	Franquia	10
	Luxo	100
	Popular	80
	Padrão	140
	Franquia	30

Cin/UFPE - DW 83

Operações OLAP

- Sort: retorna um cubo onde os membros de uma dimensão foram ordenados
 - Sintaxe da operação sort:
 - SORT (Nome_Cubo, Dimensão, Expressão, [TIPO])
 - TIPO pode ser ASC | DESC | BASC | BDESC
 - Membros de Dimensão são ordenados pelo valor de Expressão
 - Ordenação de produtos pelo nome:
 - SORT (Vendas, Produto, Nome, ASC)

Cin/UFPE - DW 84

Operações OLAP

- Pivot (ou Rotar): rotaciona os eixos de um cubo para prover uma apresentação alternativa dos dados
 - Sintaxe da operação pivot:
 - PIVOT (Nome_Cubo, Dimensão1 → Eixo1,..., DimensãoN → EixoN)
 - Rotação do cubo Vendas:
 - PIVOT (Vendas, Tempo → X, Cliente → Y, Produto → Z)

Cln/UFPE - DW © 85

Operações OLAP

- Pivot
 - Operação que gira o cubo, permitindo ter diferentes visões dos dados
 - Permite a troca de perspectiva da visão dos dados

Cln/UFPE - DW © 86

Operações OLAP

- Slice: remove uma dimensão do cubo
 - Sintaxe da operação slice:
 - SLICE (Nome_Cubo, Dimensão, Level = Value)
 - Dimensão será eliminada
 - Assume que a granularidade do cubo está em um nível específico definido por Level de Dimensão
 - Remoção da Dimensão Cliente:
 - SLICE (Vendas, Cliente, Cidade = 'Paris')

Cln/UFPE - DW © 87

Operações OLAP

- Dice: retorna células do cubo que satisfazem uma condição
 - Sintaxe da operação dice:
 - DICE (Nome_Cubo, Condição)
 - Condição é definida sobre níveis de dimensão, atributos e medidas
 - Assemelha-se à operação de seleção da álgebra relacional
 - Geração de um cubo Vendas para os 2 primeiros trimestres e para as cidades da França:
 - DICE (Vendas, (Cliente, Cidade = 'Paris' OR Cliente, Cidade = 'Lyon') AND (Tempo.Trimestre = 'Q1' OR Tempo.Trimestre = 'Q2'))

Cln/UFPE - DW © 88

Operações OLAP

- Dice
 - Operação que fatia o cubo, permitindo restringir a análise aos dados, sem inversão de eixos
 - Filtra os dados

Cln/UFPE - DW © 89

Operações OLAP

- Drill-across: combina células de 2 cubos que possuem os mesmos esquemas e as mesmas instâncias
 - Sintaxe da operação drill-across:
 - DRILLACROSS (Nome_Cubo1, Nome_Cubo2, [Condição])
 - Assemelha-se ao full outer join da álgebra relacional
 - Comparação de vendas de 2012 e 2011:
 - DRILLACROSS (Vendas, Vendas2011)

Cln/UFPE - DW © 90

Operações OLAP

- Union: mescla 2 cubos de dados que possuem os mesmos esquemas mas instâncias disjuntas
 - Sintaxe da operação union:
 - UNION (Nome_Cubo1, Nome_Cubo2)
 - União das vendas feitas aos clientes da França, Alemanha e Espanha
 - UNION (Vendas, VendasEspanha)

Cln/UFPE - DW 91

Operações OLAP

- Difference: remove células de um cubo que existem em outro cubo (ambos cubos possuindo o mesmo esquema)
 - Sintaxe da operação difference:
 - DIFFERENCE (Nome_Cubo1, Nome_Cubo2)
 - Diferença entre vendas de 2012 e duas maiores quantidades vendidas de 2012
 - DIFFERENCE (Vendas, MaioresVendas)

Cln/UFPE - DW 92

Operações OLAP

- Resumo

Operação	Objetivo
Dice	Retorna as células do cubo que satisfaz uma condição
Difference	Remove as células de um cubo que estão em outro cubo
Drill-across	Mescla 2 cubos com mesmo esquema e conjunto de instâncias
Drill-down	Desagrega medidas ao longo da hierarquia para obter dados mais detalhados. É o oposto do Roll-up.
Drill-through	Recupera dados do sistema operacional usados na construção do cubo
Pivot	Rotaciona os eixos de um cubo
Roll-up	Agrega medidas ao longo de uma hierarquia para obter dados mais genéricos. É o oposto do Drill-down.
Slice	Remove uma dimensão de um cubo
Sort	Ordena os membros de uma dimensão usando uma expressão
Union	Combina células de 2 cubos que têm o mesmo esquema mas possuem instâncias disjuntas

Cln/UFPE - DW 93

Tópicos

- Modelo MultiDim
 - Notação gráfica
 - Hierarquias
 - Granularidade
 - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
- Data Marts
- Data Warehousing

Cln/UFPE - DW 94

Modelo MultiDim

- Esquema Conceitual: descrição concisa dos requisitos de dados e análise dos usuários sem detalhes de implementação
- Vantagens
 - Podem ser facilmente convertidos para o esquema lógico usando um conjunto de regras de mapeamento
 - Facilita a comunicação entre projetistas de BD e usuários sobre as necessidades da aplicação
 - É mais estável do que esquemas orientados à implementação que precisam ser alterados quando há mudanças na plataforma alvo
- Contudo, não existe ainda um modelo conceitual para dados multidimensionais que seja universalmente adotado
- Apesar disso, existe o modelo MultiDim que permite a modelagem conceitual de todos os elementos de um DW

Cln/UFPE - DW 95

Modelo MultiDim

- Notação Gráfica

Cln/UFPE - DW 96

Modelo MultiDim

- Esquema: composto de um conjunto de dimensões e um conjunto de fatos
- Dimensão: composta de um nível ou de uma ou mais hierarquias
 - Uma hierarquia é composta por um conjunto de níveis
 - Não existe notação gráfica para representar uma dimensão
 - É modelada por meio da representação de seus componentes
 - O nome do nível folha define o nome da dimensão
 - Exceto quando o nível folha possui várias associações com fato → o papel define o nome da dimensão

Cln/UFPE - DW 97

Modelo MultiDim

- Membros: instâncias de um nível
- Nível: descreve um conjunto de objetos ou eventos com características similares
 - Conceito semelhante ao de tipo de entidade do modelo ER
 - Possui um conjunto de atributos que caracterizam seus membros
 - Possui um ou mais atributos identificadores que unicamente determinam seus membros
 - Cada identificador pode ser composto por um ou mais atributos
 - Cada atributo de um nível possui um tipo mas não possui notação gráfica

Cln/UFPE - DW 98

Modelo MultiDim

- Fato: associa vários níveis
 - Um mesmo nível pode participar várias vezes de um fato por meio da definição de diferentes papéis
 - Cada papel é identificado por um nome e é representado por uma ligação entre o nível correspondente e o fato

Cln/UFPE - DW 99

Modelo MultiDim

- Membros do Fato: são instâncias de um fato
- Cardinalidade do relacionamento entre níveis e fatos indica o número mínimo e máximo de membros do fato que podem se relacionar com os membros do nível
 - (1,n) entre o fato Vendas e o nível Produto → cada venda diz respeito a um produto o qual pode participar de várias vendas
 - (1,1) entre o fato Vendas e o nível Pedido → cada venda possui um único número de pedido o qual representa uma única venda
- Hierarquia é composta de vários níveis que se relacionam por um relacionamento pai-filho.
 - Mesmas cardinalidades dos relacionamentos entre fatos e níveis
 - (1,n) entre o nível filho Produto e o nível pai Categoria → todo produto pertence a uma categoria a qual possui vários produtos

Cln/UFPE - DW 100

Modelo MultiDim

- Atributos identificadores dos níveis envolvidos em um fato indicam a granularidade de medidas
- Uma dimensão pode ter várias hierarquias, cada uma delas expressando um critério de análise
 - Para diferenciá-las cada hierarquia possui um nome
- Um fato contém atributos chamados de Medidas

Cln/UFPE - DW 101

Modelo MultiDim

- Esquema conceitual de uma aplicação de Vendas

Cln/UFPE - DW 102

Tópicos

- Modelo MultiDim
 - Notação gráfica
 - Hierarquias
 - Granularidade
 - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
- Data Marts
- Data Warehousing

CIn/UFPE - DW 103

Hierarquias no MultiDim

- Balanceda: possui um só caminho à nível de esquema onde todos os níveis são obrigatórios
 - À nível de instância os membros formam uma árvore e todos caminhos têm o mesmo tamanho

All

Categoria

Produto

CIn/UFPE - DW 104

Hierarquias no MultiDim

- Desbalanceada: possui um só caminho à nível de esquema onde pelo menos um nível não é obrigatório
 - À nível de instância podem existir membros do nível pai sem membros do nível filho (membros formam árvore desbalanceada)

Banco

Filial

Agência

ATM

CIn/UFPE - DW 105

Hierarquias no MultiDim

- Recursiva (ou pai-filho): é um tipo especial da hierarquia desbalanceada, onde:
 - Membros de um mesmo nível são associados por meio de 2 papéis de um relacionamento pai-filho
 - Exemplo: relacionamento empregado-supervisor
 - Papeis do relacionamento pai-filho são associados aos membros do nível Empregado
 - É desbalanceada porque empregados sem subordinados não terão descendentes na árvore

Empregado

Empregado

Empregado

CIn/UFPE - DW 106

Hierarquias no MultiDim

- Generalizada:
 - Representa membros de um mesmo nível que são de diferentes tipos
 - Exemplo: clientes podem ser empresas ou pessoas
 - Medidas relativas a clientes devem ser agregadas diferentemente de acordo com o tipo do cliente:
 - Para empresas: Cliente → Ramo → Setor
 - Para pessoas: Cliente → Profissão → Setor
 - À nível de esquema, uma hierarquia generalizada possui múltiplos caminhos exclusivos, e pelo menos, compartilha o nível folha

Cliente

Ramo

Setor

Profissão

CIn/UFPE - DW 107

Hierarquias no MultiDim

- Generalizada:
 - Representa membros de um mesmo nível que são de diferentes tipos
 - Exemplo: clientes podem ser empresas ou pessoas
 - Medidas relativas a clientes devem ser agregadas diferentemente de acordo com o tipo do cliente:
 - Para empresas: Cliente → Ramo → Setor
 - Para pessoas: Cliente → Profissão → Setor
 - À nível de esquema, uma hierarquia generalizada possui múltiplos caminhos exclusivos, e pelo menos, compartilha o nível folha

Cliente

Ramo

Setor

Profissão

CIn/UFPE - DW 108

Hierarquias no MultiDim

- Generalizada:
 - Representa membros de um mesmo nível que são de diferentes tipos
 - Exemplo: clientes podem ser empresas ou pessoas
 - Medidas relativas a clientes devem ser agregadas diferentemente de acordo com o tipo do cliente:
 - Para empresas: Cliente → Ramo → Setor
 - Para pessoas: Cliente → Profissão → Setor
 - À nível de esquema, uma hierarquia generalizada possui múltiplos caminhos exclusivos, que compartilham pelo menos, o nível folha

Cln/UFPE - Div 109

Hierarquias no MultiDim

- Generalizada:
 - À nível de instância, cada membro da hierarquia pertence a apenas um caminho
 - Símbolo de relacionamento exclusivo indica que caminhos são exclusivos para cada membro

Existem 2 caminhos de agregação, um para cada tipo de cliente e ambos pertencem à mesma hierarquia

Cln/UFPE - DW 110

Hierarquias no MultiDim

- Generalizada: Não é necessário que níveis de divisão sejam unidos
 - Análise de publicações internacionais, onde 3 tipos de publicações são consideradas: revistas, livros e anais
 - Uma conferência produz vários anais de eventos satélites
 - Anais podem ser agregados ao nível de conferência

Cln/UFPE - DW 111

Hierarquias no MultiDim

- Irregular: é um tipo especial da hierarquia generalizada onde: caminhos alternativos são obtidos pela omissão de um ou mais níveis
 - Exemplo: Cidade → Estado → Região → País → Continente
 - Alguns países (e.g. Bélgica) são divididos em regiões, enquanto outros (e.g. Alemanha) não são
 - Países pequenos (e.g. Vaticano) não possuem regiões nem estados
 - À nível de instância:
 - cada membro filho tem apenas um pai
 - tamanho do caminho das folhas até um mesmo nível pai pode ser diferente para diferentes membros

Cln/UFPE - DW 112

Hierarquias no MultiDim

- Irregular:
 - Exemplo: Cidade → Estado → Região → País → Continente

Cln/UFPE - DW 113

Hierarquias no MultiDim

- Alternativa:
 - Representa a situação onde à nível de esquema, existem várias hierarquias não exclusivas que compartilham ao menos, o nível folha
 - Exemplo: Dimensão *Tempo* com 2 hierarquias para permitir agrupamentos diferentes de meses em: anos fiscais e anos de calendário

Cln/UFPE - DW 114

Hierarquias no MultiDim

- Alternativa: Exemplo de Instâncias
 - Anos fiscais começam em fevereiro
 - Hierarquias formam um grafo
 - Um membro filho possui mais de um membro pai e tais membros pai pertencem a diferentes níveis
 - Hierarquias alternativas são necessárias quando deseja-se analisar medidas a partir de uma única perspectiva (e.g. Tempo) usando agregações alternativas

Cln/UFPE - DW 115

Hierarquias no MultiDim

- Alternativa: Exemplo de Instâncias
 - Anos fiscais começam em fevereiro
 - Hierarquias formam um grafo
 - Um membro filho possui mais de um membro pai e tais membros pai pertencem a diferentes níveis
 - Hierarquias alternativas são necessárias quando deseja-se analisar medidas a partir de uma única perspectiva (e.g. Tempo) usando agregações alternativas

Cln/UFPE - DW 115

Hierarquias no MultiDim

- Diferença entre Hierarquias Alternativas e Generalizadas:
 - Ambos tipos de hierarquias compartilham alguns níveis mas representam situações distintas
 - Na Hierarquia Generalizada:
 - Um membro filho está associado a apenas um dos caminhos possíveis
 - Representação de múltiplos caminhos exclusivos
 - Exemplo: Cliente → Ramo → Setor
 - Cliente → Profissao → Setor
 - Na Hierarquia Alternativa:
 - Um membro filho está associado a todos os caminhos possíveis, e o usuário deve escolher um deles para análise
 - Representação de múltiplos caminhos não-exclusivos
 - Exemplo: Mes → TrimestreCalendario → AnoCalendario
 - Mes → TrimestreFiscal → AnoFiscal

Cln/UFPE - DW 117

Hierarquias no MultiDim

- Diferença entre Hierarquias Alternativas e Generalizadas:
 - Ambos tipos de hierarquias compartilham alguns níveis mas representam situações distintas
 - Na Hierarquia Generalizada:
 - Um membro filho está associado a apenas um dos caminhos possíveis
 - Representação de múltiplos caminhos exclusivos
 - Exemplo: Cliente → Ramo → Setor
 - Cliente → Profissao → Setor
 - Na Hierarquia Alternativa:
 - Um membro filho está associado a todos os caminhos possíveis, e o usuário deve escolher um deles para análise
 - Representação de múltiplos caminhos não-exclusivos
 - Exemplo: Mes → TrimestreCalendario → AnoCalendario
 - Mes → TrimestreFiscal → AnoFiscal

Cln/UFPE - DW 118

Hierarquias no MultiDim

- Esta é uma Hierarquia Alternativa ou Generalizada?

Hierarquia Generalizada com nível de junção!

Cln/UFPE - DW 119

Hierarquias no MultiDim

- Esta é uma Hierarquia Alternativa ou Generalizada?

Hierarquia Generalizada sem nível de junção!

Cln/UFPE - DW 120

Hierarquias no MultiDim

- Esta é uma Hierarquia Alternativa ou Generalizada?

Hierarquia Alternativa!

CIn/UFPE - DW 121

Hierarquias no MultiDim

- Paralela:
 - Existe quando uma dimensão possui várias hierarquias associadas a ela para modelar diferentes critérios de análise
 - As hierarquias componentes podem ser de diferentes tipos
 - Pode ser de dois tipos: dependente ou independente
 - Hierarquias componentes compartilham níveis → dependente
 - Hierarquias componentes não compartilham níveis → independente
 - Exemplo de uma dimensão com 2 hierarquias paralelas independentes:

CIn/UFPE - DW 122

Hierarquias no MultiDim

- Paralela:
 - Existe quando uma dimensão possui várias hierarquias associadas a ela para modelar diferentes critérios de análise
 - As hierarquias componentes podem ser de diferentes tipos
 - Pode ser de dois tipos: dependente ou independente
 - Hierarquias componentes compartilham níveis → dependente
 - Hierarquias componentes não compartilham níveis → independente
 - Exemplo de uma dimensão com 2 hierarquias paralelas independentes:

CIn/UFPE - DW 123

Hierarquias no MultiDim

- Paralela:
 - Existe quando uma dimensão possui várias hierarquias associadas a ela para modelar diferentes critérios de análise
 - As hierarquias componentes podem ser de diferentes tipos
 - Pode ser de dois tipos: dependente ou independente
 - Hierarquias componentes compartilham níveis → dependente
 - Hierarquias componentes não compartilham níveis → independente
 - Exemplo de uma dimensão com 2 hierarquias paralelas independentes:

CIn/UFPE - DW 124

Hierarquias no MultiDim

- Paralela:
 - Exemplo de uma dimensão com 2 hierarquias paralelas dependentes:
 - Análise de vendas para lojas localizadas em vários países
 - LocalFisico representa a divisão geográfica dos endereços de lojas
 - LocalVendas representa a divisão organizacional das empresas
 - Ambas hierarquias compartilham o nível Estado que possui diferentes papéis de acordo com a hierarquia escolhida para análise

CIn/UFPE - DW 125

Hierarquias no MultiDim

- Diferença entre Hierarquias Alternativas e Paralelas:
 - Ambos tipos de hierarquias compartilham alguns níveis e podem ser compostas de várias hierarquias mas representam situações distintas
 - Hierarquia Alternativa:
 - Possui apenas um único nome de hierarquia
 - Não é possível combinar níveis de diferentes hierarquias componentes
 - Exemplo: Qual o total de vendas por trimestre fiscal que pertencem ao 2º semestre do calendário regular de 2012? (Não faz sentido!)
 - Hierarquia Paralela:
 - Possui vários nomes de hierarquias que representam vários critérios de análise
 - Permite combinar níveis de diferentes hierarquias componentes
 - Exemplo: Qual o total de vendas das lojas da cidade A que pertencem ao distrito de vendas B? (Faz sentido!)

CIn/UFPE - DW 126

Hierarquias no MultiDim

- Diferença entre Hierarquias Alternativas e Paralelas:
 - Em hierarquias dependentes paralelas, um membro folha pode estar associado a vários diferentes membros de um nível compartilhado
 - Exemplo: local de moradia e região de vendas de empregados
 - Navegação em ambas hierarquias do nível *Empregado* para o nível *Estado* produz resultados distintos mas ambos corretos
 - Isto ocorre quando empregados residem em um estado e trabalham em outro

Cln/UFPE - DW © 127

Hierarquias no MultiDim

- Diferença entre Hierarquias Alternativas e Paralelas:
 - Valores de medida agregados de níveis compartilhados de hierarquias alternativas podem ser reusados
 - Exemplo: valores de vendas por mês (nível compartilhado) podem ser usados para computar valores de vendas por ano fiscal e ano de calendário
 - Isto não é o caso de hierarquias paralelas dependentes
 - Exemplo: Vendas geradas por Empregados que residem em PE

Cln/UFPE - DW © 128

Hierarquias no MultiDim

- Não-estrita: possui pelo menos um relacionamento n:m
 - N:M relacionamentos entre níveis pai e filho são comuns em aplicações reais:
 - Um diagnóstico pode pertencer a vários grupos de diagnósticos
 - Uma semana pode pertencer a dois meses
 - Um produto pode ser classificado em várias categorias
 - O fato de uma hierarquia ser estrita ou não independe de seu tipo
 - Hierarquias previamente discutidas podem ser estritas ou não
 - Exemplo de instâncias de uma hierarquia não-estrita:
 - Estado: Ceará, Pernambuco
 - Cidade: Fortaleza, Recife, Olinda
 - Empregado: Paulo

Cln/UFPE - DW © 129

Hierarquias no MultiDim

- Problemas com hierarquias não-estritas:
 - Contagem dupla de valores de medida em operações de drill-down
 - Exemplo: Vendas de empregados agregadas por cidade e estado

Cln/UFPE - DW © 130

Hierarquias no MultiDim

- Soluções para a contagem dupla de valores:
 - Transformar uma hierarquia não-estrita em uma hierarquia estrita:
 - Criar um novo membro pai para cada grupo de membros pai associados a um único membro filho de um relacionamento n:m

Cln/UFPE - DW © 131

Hierarquias no MultiDim

- Soluções para a contagem dupla de valores:
 - Transformar uma hierarquia não-estrita em uma hierarquia estrita:
 - Escolher um membro pai como o principal e ignorar a existência dos outros membros pai
 - Exemplo: Fortaleza pode ser considerado o membro principal

Cln/UFPE - DW © 132

Hierarquias no MultiDim

- Soluções para a contagem dupla de valores:
 - Transformar uma hierarquia não-estrita em uma hierarquia estrita:
 - Mapear a hierarquia não-estrita em 2 dimensões independentes
 - Esta solução geralmente muda o foco da análise, produzindo um esquema conceitual diferente

- Só pode ser usada quando a distribuição exata de medidas é conhecida
 - Exemplo: quando se sabe os valores das vendas obtidos pelos funcionários ao trabalhar em diferentes cidades
- Embora medidas sejam agregadas corretamente em roll-ups, o problema de contagem dupla de um mesmo funcionário ocorrerá

Cln/UFPE - DW 133

Hierarquias no MultiDim

- Soluções para a contagem dupla de valores:
 - Considere o uso do esquema anterior para contar o número de empregados por cidade ou por divisão
 - Este valor pode ser calculado contando as instâncias de Empregado na tabela de fatos

Contar o número de empregados distintos produz valor certo!

Valores agregados de cada cidade não podem ser reusados para calcular o número de empregados de cada divisão!

Cln/UFPE - DW 134

Hierarquias no MultiDim

- Soluções para a contagem dupla de valores:
 - Uma abordagem alternativa consiste no uso de um atributo de distribuição
 - Indica como medidas são distribuídas entre os vários membros pai de relacionamentos n:m (e.g. Percentagem de tempo)

- Se um funcionário possui uma posição mais elevada, então embora ele trabalhe menos em uma cidade, ele pode receber mais
- Atributo de distribuição pode ser definido pelo número total de membros pais com os quais o membro filho se relaciona.

Cln/UFPE - DW 135

Tópicos

- Modelo MultiDim
 - Notação gráfica
 - Hierarquias
 - Granularidade
 - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
- Data Marts
- Data Warehousing

Cln/UFPE - DW 136

Granularidade no MultiDim

- Granularidade de Fatos
 - Refere-se ao nível de sumarização e de detalhe dos dados
 - Quanto mais detalhe existir, mais baixo é o nível de granularidade
 - Quanto menos detalhe existir, mais alto é o nível de granularidade

Cln/UFPE - DW 137

Granularidade no MultiDim

- Medidas podem ser coletadas em várias granularidades

Maior Nível de Detalhes Menor Nível de Granularidade	Menor Nível de Detalhes Maior Nível de Granularidade
<p>Exemplo: Detalhe de cada transação de venda de um vendedor realizada durante um mês</p> <p>Data Hora Vendedor Valor → 50 registros por mês</p>	<p>Exemplo: Sumário das transações de venda de um vendedor realizadas durante um mês</p> <p>Mês Vendedor Valor → 01 registro por mês</p>

Cln/UFPE - DW 138

Granularidade no MultiDim

- Granularidade de fatos é importante?
 - Afeta o volume de dados mantidos no DW
 - Quanto maior for o volume de dados:
 - Menor é o desempenho do sistema
 - Maior é o espaço em disco necessário
 - Maior é quantidade de índices mantidos pelo sistema
 - Impacta a disponibilização de consultas e relatórios
 - Impossibilidade de uso dos dados para atender a consultas detalhadas
 - Impossibilita a satisfação dos requisitos de análise
 - Seria interessante saber a evolução das vendas no dia 5 de cada mês nos últimos 5 anos? Não!
 - Mas poderia ser interessante saber o comportamento das vendas no mês de Dezembro nos últimos 5 anos!
 - Poderia também ser interessante agrupar dias em períodos festivos

Granularidade no MultiDim

- Não se pode analisar informações de DW diferentes se apresentadas em granularidades diferentes
 - Vendas diárias realizadas por vendedores, independente de produto
 - Vendas de produto no mês, independente de vendedor
- Exceto se os dados do DW de menor granularidade forem sumarizados para o mesmo nível de detalhe do DW de maior granularidade
 - Vendas de produtos por mês ← Isto não se aplica apenas à dimensão temporal
 - Vendas diárias de produtos
- MultiDim possui uma notação para representação de fatos com múltiplas granularidades
 - Uso de relacionamentos exclusivos entre os vários níveis de granularidade

Granularidade no MultiDim

- Fatos com Múltiplas Granularidades:
 - Exemplo: vendas podem ser reportadas por estado ou por cidade, dependendo do local da loja

Tópicos

- Modelo MultiDim
 - Notação gráfica
 - Hierárquias
 - Granularidade
 - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
- Data Marts
- Data Warehousing

Dimensão no MultiDim

- Dimensão n:m: vários membros desta dimensão participam de um mesmo fato
 - Exemplo: uma conta bancária pode pertencer a vários clientes
- Problema de contagem dupla: obtenção do saldo de uma conta por cliente computará o saldo desta conta por tantos clientes quantos a conta tiver

Tempo	Conta	Cliente	Saldo
T1	A1	C1	100
T1	A1	C2	100
T1	A1	C3	100
T1	A2	C1	500
T1	A2	C2	500

 - Em T1, existem 2 contas: A1 e A2 com 100 e 500, respectivamente
 - A1 é uma conta conjunta de C1, C2 e C3, enquanto A2 é de C1 e C2
 - Balanco total das 2 contas é igual a 600 mas a agregação por tempo ou cliente produz um valor igual a 1.300

Dimensão no MultiDim

- Dimensão n:m: exemplo de um esquema conceitual

- Forma Normal Multidimensional (FNM): cada medida é unicamente identificada pelo conjunto de níveis folha associados.
- Como Conta e Tempo determinam o saldo, i.e. o saldo depende apenas do número da conta e do tempo considerado → não satisfaz a FNM
 - Numero da conta e Tempo → Saldo

Dimensão no MultiDim

- Dimensão m:n: Duas abordagens para decompor os fatos
 - 1) Criando dois fatos

- Se os correntistas não mudam ao longo do tempo então clientes são determinados apenas pelas Contas
 - A ligação entre o nível *Tempo* e os fatos *Correntistas* pode sair

CIn/UFPE - DW © 145

Dimensão no MultiDim

- Dimensão m:n: Duas abordagens para decompor os fatos
 - 2) Incluindo uma hierarquia não-estruta

CIn/UFPE - DW © 146

Tópicos

- Modelo MultiDim
 - Notação gráfica
 - Hierarquias
 - Granularidade
 - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
- Data Marts
- Data Warehousing

CIn/UFPE - DW © 147

Mapeamento ER-MultiDim

- Construção do Modelo Multidimensional
 - Análise dos dados do ambiente operacional
 - Levantamento de requisitos das atividades de negócio dos usuários do DW
- A partir do esquema ER, vários modelos dimensionais podem ser gerados
 - Identificar as atividades de negócio e modelá-las separadamente
 - Relacionamentos N:M com propriedades numéricas e aditivas são mapeados em tabelas fato
 - Desnormalizar todas as entidades restantes e mapeá-las em tabelas de dimensão

CIn/UFPE - DW © 148

Mapeamento ER-MultiDim

- Metodologia
 - Especificar os fatos de negócio
 - Definir as dimensões
 - Identificar os atributos das dimensões
 - Identificar as medidas
 - Escolher a granularidade

CIn/UFPE - DW © 149

Mapeamento ER-MultiDim

- Metodologia
 - Especificar os fatos de negócio
 - Fatos normalmente aparecem em relacionamentos N:M do ER
 - Por exemplo:
 - » Uma Loja vende vários Produtos
 - » Cada Produto é vendido em várias Lojas
 - » Um Crime ocorre em várias Cidades
 - » Cada Cidade é o local de ocorrência de vários Crimes
 - » Um Operário produz várias Peças
 - » Cada peça é manufaturada por vários Operários
 - » Secretarias de saúde monitoram casos de Dengue
 - » Cada caso é monitorado por várias secretarias

CIn/UFPE - DW © 150

Mapeamento ER-MultiDim

- Metodologia
 2. Definir as dimensões
 - Dimensões são definidas conforme a necessidade de visualização dos fatos por parte do usuário
 - Para facilitar, pode-se adotar o uso de pontos cardeais principais:
 - ONDE o fato ocorre?
 - O QUE acontece no fato?
 - QUANDO o fato ocorre?
 - QUEM participa (ou realiza o) do fato?
 - Dimensão temporal sempre deve existir (i.e. QUANDO)
 - Quebrar o tempo em atributos individuais
 - Representar o tempo como dias úteis, finais de semana, feriados, ou períodos fiscais
 - Uma dimensão é derivada de uma entidade do E-R ou pode ser obtida indiretamente por meio do relacionamento entre entidades

CIn/UFPE - DW © 151

Mapeamento ER-MultiDim

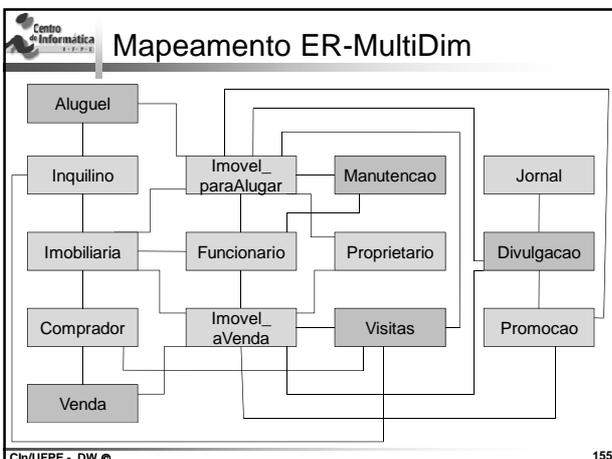
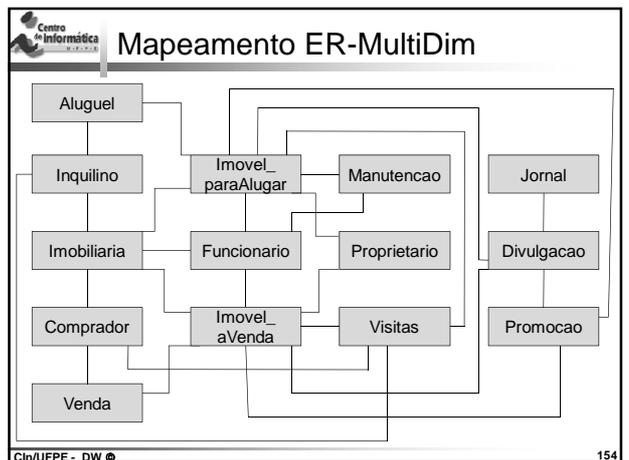
3. Identificar os atributos das dimensões
 - Atributos hierárquicos são obtidos a partir de uma seqüência de entidades interligadas por relacionamentos 1:N
 - Múltiplas hierarquias são normalmente derivadas por entidades com auto-relacionamento no diagrama E-R
 - Atributos não hierárquicos são identificados junto aos atributos de uma entidade do diagrama E-R.
 - Considerações de projeto de atributos de dimensões
 - Evitar o uso de códigos e abreviações
 - » SCMSB3340 ⇒ sapato de couro de altura média e solado de borracha com numeração de 33 a 40
 - Remover variações textuais que deveriam ser iguais
 - » Por exemplo: *separado* e *separado judicialmente*
 - Criar colunas úteis para níveis de agregação
 - Evitar valores nulos ou inexistentes
 - Incluir colunas altamente correlacionadas e descritivas

CIn/UFPE - DW © 152

Mapeamento ER-MultiDim

4. Identificar as medidas
 - Medidas normalmente correspondem aos valores numéricos e aditivos dos relacionamentos N:M que derivaram os fatos;
 - Para medidas semi-aditivas ou não aditivas, deve-se verificar a necessidade de armazenamento destas na tabela de fatos
5. Escolher a granularidade
 - Avaliar os requisitos dos dados
 - Escolher o nível mais baixo de detalhe
 - Requer mais espaço em disco
 - Envolve maior tempo de processamento
 - Permite análise de dados detalhada
 - Adaptar medidas à granularidade estabelecida

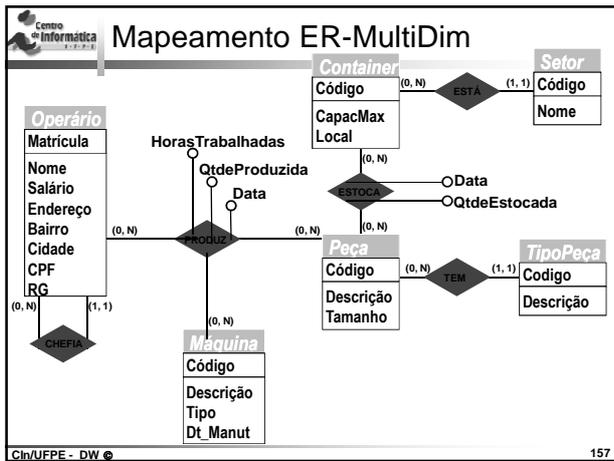
CIn/UFPE - DW © 153



Mapeamento ER-MultiDim

- Entidades da atividade de negócio *Vendas de Imóvel*
- Escolher a granularidade
 - Vendas de Imóveis Individuais
- Escolher as dimensões
 - Entidades restantes + dimensão temporal

CIn/UFPE - DW © 156



Mapeamento ER-MultiDim

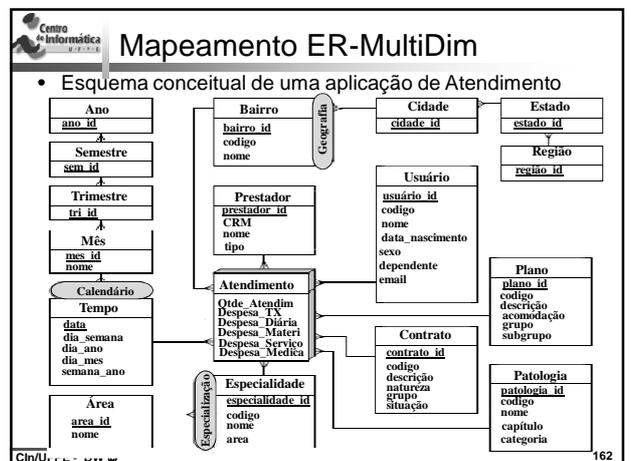
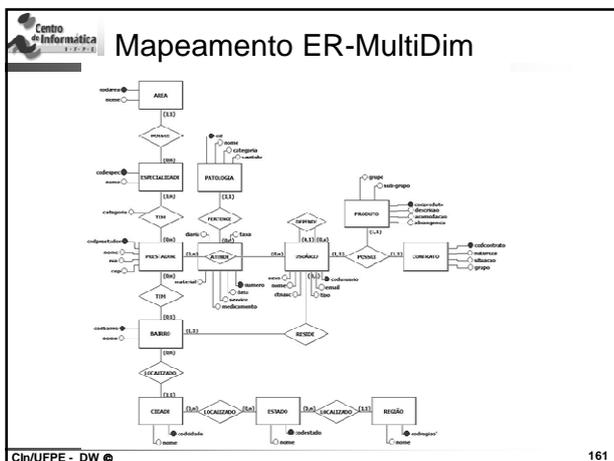
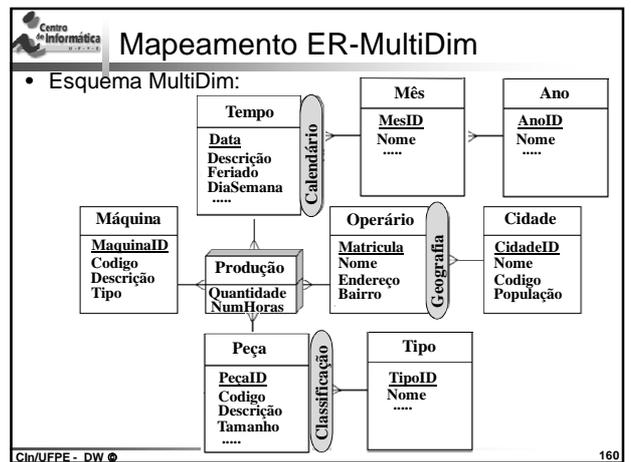
- O gerente de fabricação de uma indústria de peças automotivas necessita de um DW que o auxilie na alocação de recursos da produção
 - Avaliar a produção diária por operário e por tipos de peças produzidas
 - Descobrir quais são as máquinas mais alocadas na fabricação peças de motor, carroceria e engrenagens
 - Avaliar a evolução da quantidade de horas trabalhadas na produção de peças por tipos de peça, por operários e por semana
- Mapeamento ER-MultiDim
 - Especificar os fatos de negócio
 - Relacionamentos M-N: produz e estoca
 - Fatos de interesse do usuário:
 - Produção de peças
 - Tempo gasto na produção de peças

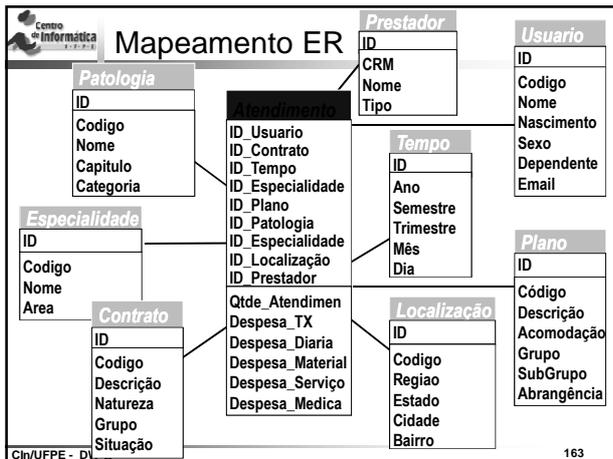
Cln/UFPE - DW 158

Mapeamento ER-MultiDim

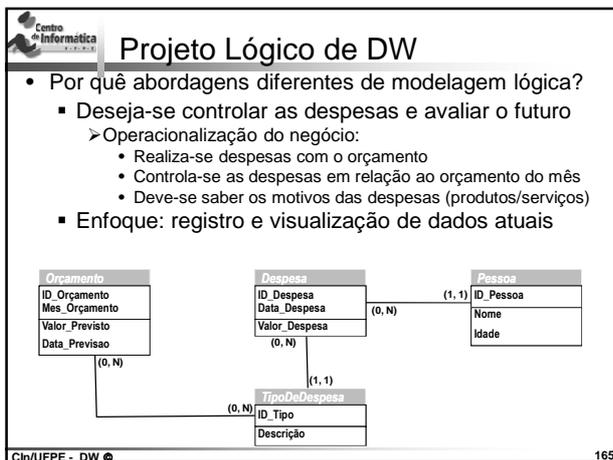
- Mapeamento ER-MultiDim (Cont.)
 - Definir as dimensões
 - O que são produzidos? Peças
 - Quando são produzidas? Diariamente (Dimensão Tempo)
 - Quem produziu? Operários
 - Onde são produzidas? Máquinas
 - Identificar os atributos das dimensões
 - Para a dimensão Peça:
 - Código, Descrição e Tamanho: obtidos da entidade Peça
 - Descrição do tipo da peça: obtido da entidade TipoPeça
 - Para a dimensão Tempo:
 - Dia, Mês, Trimestre, Semestre e Ano: obtidos do atributo do relacionamento Produz
 - Para a dimensão Operário: Matrícula, Nome, Rua, Bairro e Cidade.
 - Para a dimensão Máquina: Código, Descrição e Tipo.

Cln/UFPE - DW 159

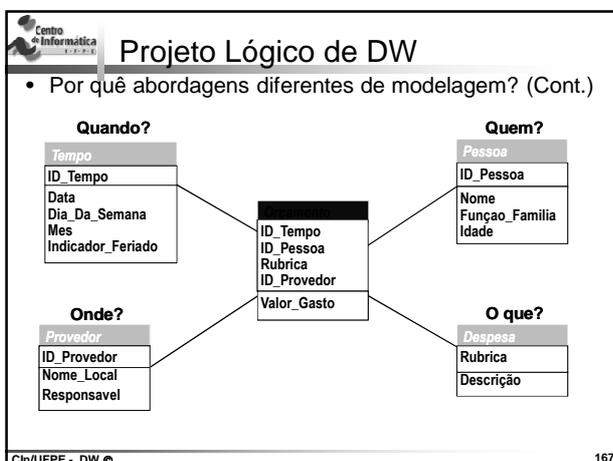




- ### Tópicos
- Modelo MultiDim
 - Notação gráfica
 - Hierárquias
 - Granularidade
 - Dimensão
 - Mapeamento ER-MultiDim
 - Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
 - Data Marts
 - Data Warehousing
- Cln/UFPE - DW 164



- ### Projeto Lógico de DW
- Por que abordagens diferentes de modelagem? (Cont.)
 - Deseja-se controlar as despesas e avaliar o futuro
 - Gestão do negócio:
 - Como está a evolução das despesas durante o ano?
 - Como está o percentual de cada tipo de gasto em relação ao orçamento mensal?
 - Quais são as despesas que possuem evolução histórica de aumento de valores em relação a variações do orçamento?
 - Como a redução de um gasto afeta o orçamento?
 - Enfoque:
 - Desempenho de determinados assuntos ao longo do tempo
 - Simulações táticas (cenários) para análise estratégica de decisões
- Cln/UFPE - DW 166



- ### Projeto Lógico de DW
- Comparando Modelos de Dados
- | Operacional: OLTP | Analítico: Data Warehouse |
|---|--|
| <ul style="list-style-type: none"> ▪ Entidades normalizadas ▪ Segue terceira forma normal ou maior ▪ Produz um design complexo de BD ▪ Armazena dados no nível transacional mais baixo ▪ Aumenta o nível de JOIN de tabelas em consultas ▪ Estrutura tipicamente estática | <ul style="list-style-type: none"> ▪ Entidades desnormalizadas ▪ Produz um único design de BD mais facilmente compreensível pelos usuários ▪ Armazena dados <ul style="list-style-type: none"> • Nível de transação • Nível de sumário ▪ Diminui o número de JOIN de tabelas em consultas ▪ Estrutura dinâmica |
- Cln/UFPE - DW 168

Projeto Lógico de DW

- Objetivos de um Projeto Lógico Multidimensional
 - Fornecer uma imagem global da realidade do negócio
 - Exibir informações em níveis apropriados de detalhes (resumido ou detalhado)
 - Otimizar o processamento de consultas complexas (Modelo Estrela ou Flocos de Neve)
 - Integrar dados de diversas fontes em uma única BD para facilitar a geração de relatórios

CIn/UFPE - DW © 169

Projeto Lógico de DW

- Abordagens para implementar o modelo multidimensional dependem de como o cubo de dados é armazenado:
 - Relacional OLAP (ROLAP):
 - Armazena os dados em BD relacionais
 - Dados pré-computados são mantidos em visões materializadas
 - Provê o suporte à extensões de SQL
 - Usa métodos de acesso especiais (e.g., índice bitmap de junção)
 - Multidimensional OLAP (MOLAP):
 - Armazena dados em estruturas de dados especiais (arrays)
 - Implementa operações OLAP sobre estas estruturas
 - Provê menos capacidade de armazenamento do que ROLAP
 - Provê melhor desempenho quando dados multidimensionais são consultados ou agrupados e quando comparado com ROLAP

CIn/UFPE - DW © 170

Projeto Lógico de DW

- Abordagens para implementar o modelo multidimensional dependem de como o cubo de dados é armazenado:
 - Híbrido OLAP (HOLAP):
 - Combina ambas as abordagens anteriores
 - Dados detalhados são mantidos em BD relacionais, enquanto dados agregados são armazenados em estruturas multidimensionais (arrays)
 - Beneficia-se da capacidade de armazenamento de ROLAP e da capacidade de processamento de MOLAP

CIn/UFPE - DW © 171

Projeto Lógico de DW

Calcular os agregados no momento da recuperação ou armazená-los?

Tempo de Resposta

BD1, BD2, BD3, BD4

Quanto maior o número de agregados, melhor o desempenho e mais dados para serem armazenados e gerenciados!

CIn/UFPE - DW © 172

Projeto Lógico de DW

- Resumo da comparação entre ROLAP, MOLAP e HOLAP

	MOLAP	HOLAP	ROLAP
Dados de base	---	Tabela Relacional	Tabela Relacional
Agregações	Arrays	Arrays	Tabela Relacional
Desempenho de Consultas	Mais Rápido	Rápido	Menos Rápido
Capacidade de Armazenamento	Baixa	Média	Alta
Memória	Principal	Principal e Secundária	Secundária

- Maioria dos servidores OLAP provêem todas as opções acima
- Mas a abordagem mais usada é a ROLAP

CIn/UFPE - DW © 173

Tópicos

- Modelo MultiDim
 - Notação gráfica
 - Hierarquias
 - Granularidade
 - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
 - Projeto de DW Relacional
 - Mudanças em Dimensões
- Data Marts
- Data Warehousing

CIn/UFPE - DW © 174

Projeto de DW Relacional

- Em sistemas ROLAP, tabelas são organizadas por meio de estruturas especializadas → esquema lógico
- Esquema lógico: um conjunto de medidas que caracterizam o desempenho de um dado negócio e de um conjunto de perspectivas de análise que são usadas para sumarizar e reestruturar os dados
- Uma possível representação relacional do modelo multidimensional é o esquema estrela o qual possui:
 - uma tabela central, chamada de fatos, contendo as medidas
 - Medidas: atributos geralmente numéricos cujos valores indicam o desempenho de uma atividade de negócio
 - um conjunto de tabelas de dimensão, uma para cada perspectiva de análise
 - possuem normalmente atributos descritivos (não-numéricos) pois servem para caracterizar os fatos observados na análise

CIn/UFPE - DW 175

Projeto de DW Relacional

- Exemplo de Esquema Estrela

CIn/UFPE - DW 176

Projeto de DW Relacional

- Componentes do Esquema Estrela

CIn/UFPE - DW 177

Projeto de DW Relacional

- Tabela de Fatos
 - Principal tabela do esquema estrela que contém medidas numéricas sobre o desempenho de uma atividade de negócio
 - Fato representa o valor de uma medida de negócio
 - Uma linha da tabela de fato corresponde ao valor de uma ou mais medidas
 - Tabela central com grande volume de dados
 - Armazena as medidas numéricas do negócio e chaves das dimensões (ID das dimensões)
 - Possui grande quantidade de linhas mas poucas colunas
 - Cada fato é a interseção entre todas dimensões
 - Existem três tipos: transacional, periódica e acumulativa

CIn/UFPE - DW 178

Projeto de DW Relacional

- Tabela de Fatos
 - Pode ser esparsa ou densa mas é única por esquema
 - Fatos possuem valores evolutivos (i.e mudam com o tempo)
 - Pode conter fatos pré-calculados que melhoram o desempenho mas aumentam o tamanho da tabela
 - Expressam relacionamentos N:M entre dimensões
 - Idealmente medidas são numéricas e aditivas
 - Vendas(R\$), Lucro(R\$), Despesas(R\$), Quantid_Vendida
 - Na teoria, um fato pode ser textual
 - O valor desta medida textual é a descrição de algo e é obtido de uma lista discreta de valores
 - Exemplo de uma medida não-numérica: Temperatura
 - Por exemplo: 20 – 40C ⇒ quente

CIn/UFPE - DW 179

Projeto de DW Relacional

- Tabela de Fatos
 - Quando um dado numérico é *Medida* ou *Atributo*?
 - Medida ⇒ varia continuamente a cada amostragem
 - Quantidade vendida de um produto
 - Atributo ⇒ praticamente constante
 - Peso de um produto
 - Em geral, as medidas da tabela de fatos possuem a mesma granularidade
 - Requer a determinação do nível de detalhe
 - Maioria dos fatos são numéricos e aditivos. Mas existem fatos:
 - Aditivos, não-aditivos e semi-aditivos

CIn/UFPE - DW 180

Projeto de DW Relacional

- Tabela de Fatos
 - Exemplo

Sua PK é formada por um subconjunto das FK

Vendas Diárias
 Data (FK)
 Produto (FK)
 Loja (FK)
 Quantidade Vendida
 Valor da Venda

Todas as chaves apontam para PK das dimensões correspondentes

Se não existirem vendas em um dado dia, em uma dada loja e para um dado produto?

CIn/UFPE - DW © 181

Projeto de DW Relacional

- Excepcionalmente, alguns campos não aditivos podem ser mantidos na tabela de fatos
 - São usados para calcular totais obtidos por fatores multiplicativos que não são explicitamente mantidos na fato
 - São valores unitários (não aditivos) relativos a estes fatores
 - Usuários têm uma visão estendida desta tabela mas não a realidade armazenada

Loja	Produto	Qtde_Danificada	Preço_Venda	Custo
A	10	10	200,00	100,0
A	20	22	40,00	20,00
B	30	53	80,00	50,00
C	10	33	200,00	100,0
C	50	21	60,00	40,00
C	80	10	140,00	70,00

CIn/UFPE - DW © 182

Projeto de DW Relacional

- Tabelas de Dimensões
 - Exemplo

Dimensão Produto
 Produto (PK)
 Descrição
 Código (Natural Key)
 Marca
 Categoria Departamento
 Tipo da Embalagem
 Tamanho da Embalagem
 Peso

Possui PK para garantir integridade referencial com qualquer tabela de fato

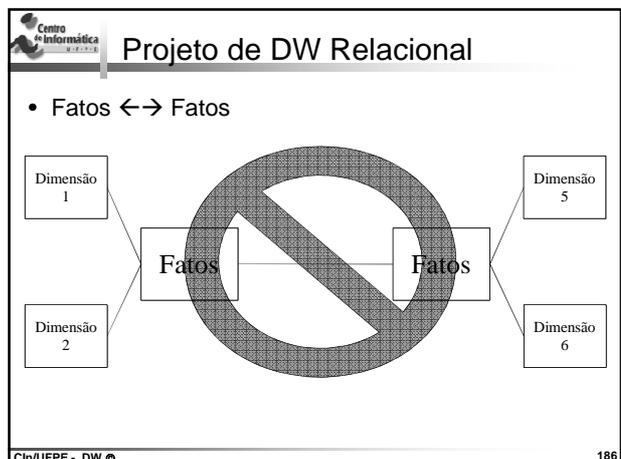
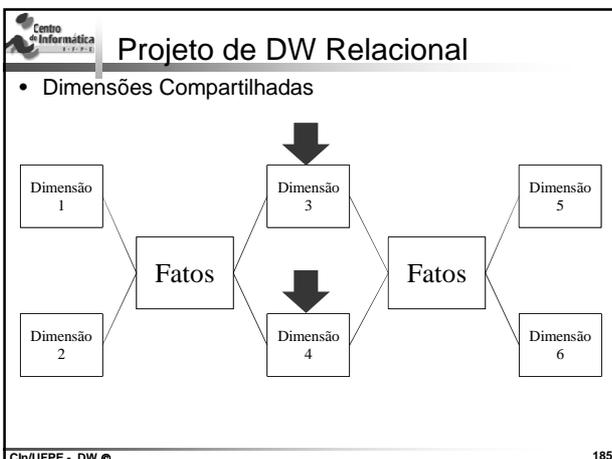
- Embora **tamanho** e **peso** sejam numéricos, eles são atributos descritivos (não são medidas!)
 - Eles são descritores constantes e discretos de um dado produto

CIn/UFPE - DW © 183

Projeto de DW Relacional

- Tabelas de Dimensões
 - Tabelas periféricas com pouco volume de dados
 - Armazenam as descrições do negócio
 - São usadas como filtros, agrupamentos e rótulos
 - Não devem ser abreviadas para aumentar legibilidade
 - Possuem poucas linhas mas muitos atributos
 - São normalmente desnormalizadas (esquema estrela)
 - Atributos ⇒ organizados em hierarquias
 - Produto (Categoria → Marca → Descrição)
 - Loja (Tipo → Endereço → Nome_Loja)
 - Tempo (Ano → Mês → Dia_Do_Mês)
 - Podem ser compartilhadas ou privadas

CIn/UFPE - DW © 184



Projeto de DW Relacional

- Tabelas de Dimensões
 - Geralmente representam relacionamentos hierárquicos (1:N) entre seus atributos
 - Informação descritiva e hierárquica é armazenada redundantemente
 - Por exemplo, para cada linha da dimensão produto, as descrições de **marca** e **categoria** são repetidas para cada uma de suas **subcategorias**
 - Elas são pequenas e representam menos de 10% do custo total de armazenamento do BD dimensional
 - Desnormalização favorece o desempenho de consultas e facilita o uso do sistema

Dimensões Desnormalizadas ⇒ modelo estrela
Dimensões Normalizadas ⇒ modelo snowflake

CIn/UFPE - DW 187

Projeto de DW Relacional

- Tabelas de Dimensões
 - Códigos operacionais são quebrados em atributos separados das tabelas de dimensões
 - Permite que eles possam ser diretamente filtrados, agrupados e usados em relatórios facilmente
 - Possuem uma ou mais hierarquias juntamente com outros atributos que não fazem parte das hierarquias

CIn/UFPE - DW 188

Projeto de DW Relacional

- Exemplo da Tabela de Dimensão Produto

Product Dimension	Department Description	Brand Description	Sales Dollar Amount	Sales Quantity
Product Key (PK)				
Product Description				
SKU Number (Natural Key)	Bakery	Baked Well	\$3,009	1,138
Brand Description	Bakery	Fluffy	\$3,024	1,476
Category Description	Bakery	Light	\$6,298	2,474
Department Description	Frozen Foods	Coldpack	\$5,521	2,640
Package Type Description	Frozen Foods	Freshlike	\$10,476	5,234
Package Size	Frozen Foods	Frigid	\$7,328	3,192
Fat Content	Frozen Foods	Icy	\$2,184	1,437
Diet Type	Frozen Foods	QuickFreeze	\$6,467	3,162
Weight				
Weight Units of Measure				
Storage Type				
Shelf Life Type				

Product Key	Product Description	Brand Description	Category Description	Department Description	Fat Content
1	Baked Well Light Sourdough Fresh Bread	Baked Well	Bread	Bakery	Reduced Fat
2	Fluffy Sliced Whole Wheat	Fluffy	Bread	Bakery	Regular Fat
3	Fluffy Light Sliced Whole Wheat	Fluffy	Bread	Bakery	Reduced Fat
4	Fat Free Mini Cinnamon Rolls	Light	Sweeten Bread	Bakery	Non-Fat
5	Diet Lovers Vanilla 2 Gallon	Coldpack	Frozen Desserts	Frozen Foods	Non-Fat
6	Light and Creamy Butter Pecan 1 Pint	Freshlike	Frozen Desserts	Frozen Foods	Reduced Fat
7	Chocolate Lovers 1/2 Gallon	Frigid	Frozen Desserts	Frozen Foods	Regular Fat
8	Strawberry Ice Creamy 1 Pint	Icy	Frozen Desserts	Frozen Foods	Regular Fat
9	Icy Ice Cream Sandwiches	Icy	Frozen Desserts	Frozen Foods	Regular Fat

CIn/UFPE - DW 189

Projeto de DW Relacional

- Comparativo entre Tabela de Fatos e de Dimensão

Fatos

- Atributos **quantitativos** sobre o desempenho do negócio em um determinado ramo
- Exemplo sobre o fato vendas: a quantidade vendida, o preço da venda, a margem de lucro, etc

Dimensões

- Atributos **qualitativos** sobre os ramos do negócio envolvidos na medida de desempenho de um determinado fato
- Exemplo sobre a dimensão produto: a descrição, o código, o preço, etc

CIn/UFPE - DW 190

Projeto de DW Relacional

- Exemplo de esquema estrela:

CIn/UFPE - DW 191

Projeto de DW Relacional

- Identificação de hierarquias nas dimensões
 - Definir quais classificações e agrupamentos em que se deseja analisar os fatos

Hierarquia Consolidada

Local da Loja	Continte	País	Região	Cidade	Loja

Hierarquia em Separado

CIn/UFPE - DW 192

Projeto de DW Relacional

- Esquema estrela:
 - Restrição de integridade referencial entre a tabela de fatos e as tabelas de dimensão
 - Tabelas de dimensão podem ser redundantes em função da presença de hierarquias
 - Tabelas de dimensão desnormalizadas e tabela de fatos normalizada
- Outros esquemas:
 - Flocos de neve
 - Constelação de fatos

Cln/UFPE - DW © 193

Projeto de DW Relacional

- Exemplo de Esquema Flocos de Neve
 - Dimensão Principal: **Dim_Produto** (Chave_Produto, Nome, Tamanho, CodigolMarca)
 - Tabelas-Dimensão Secundária: **MarcaProduto** (CodigolMarca, CodigolCategoria) e **Categoria** (CodigolCategoria, Nome)

- Dimensões são normalizadas
- Ganho de espaço de armazenamento é pouco relevante
 - Estrutura complexa
 - Custo de junções

Cln/UFPE - DW © 194

Projeto de DW Relacional

- Esquema flocos de neve: Evita a redundância de esquemas estrela por meio da normalização das tabelas de dimensão
 - Tabelas normalizadas otimizam o espaço de armazenamento mas degradam o desempenho de consultas analíticas

Cln/UFPE - DW © 195

Projeto de DW Relacional

- Esquema flocos de neve:
 - Proz tabelas normalizadas que são fáceis de manter e otimiza o espaço de armazenamento
 - Degrada o desempenho porque mais junções precisam ser executadas quando operações de navegação em hierarquias são processadas
 - Por exemplo: *Total de vendas por categoria*

```
SELECT Categoria, SUM(TotalValor)
FROM Produto P, Vendas V
WHERE P.ProdutoPK = V.ProdutoFK
GROUP BY Categoria
```
 - Esquema Estrela*
 - ```
SELECT C.Nome, SUM(TotalValor)
FROM Produto P, Vendas V, Categoria C
WHERE P.ProdutoPK = V.ProdutoFK AND
P.CategoriaKey = C.CategoriaPK
GROUP BY C.Nome
```
  - Esquema Flocos de Neve*

Cln/UFPE - DW © 196

### Projeto de DW Relacional

- Comparação entre Modelos Lógicos Multidimensionais

|                           | Estrela   | Flocos de Neve |
|---------------------------|-----------|----------------|
| Clareza                   | + fácil   | + difícil      |
| Número de tabelas         | menor     | maior          |
| Complexidade de consultas | + simples | + complexo     |
| Desempenho de consultas   | + rápido  | + lento        |

Economia de espaço de disco obtida pelo *Flocos de Neve* é geralmente menos de 1% do espaço de disco total para manter o BD multidimensional!

Cln/UFPE - DW © 197

### Projeto de DW Relacional

- Esquema constelação de fatos: possui múltiplas tabelas de fatos que compartilham tabelas de dimensão
  - Podem incluir tabelas de dimensão normalizadas ou desnormalizadas
  - Exemplo: Duas tabelas de fatos *Vendas* e *Compras* que compartilham as dimensões *Produto* e *Tempo*

Cln/UFPE - DW © 198

**Projeto de DW Relacional**

- Dimensão Causal:
  - Descreve fatos que causam mudanças em outros fatos

**Dimensão Promoção**

PromoçãoID (PK)  
 NomePromocao  
 TipoReducaoPreço  
 TipoPropaganda  
 TipoDisplay  
 TipoCupom  
 Custo  
 DataInicial  
 DataFinal

←

Valores de chaves nulos não são usados na tabela de fatos!

←

Uma tupla na dimensão deve indicar que ela não é aplicável a uma dada medida.

Cln/UFPE - DW © 199

**Projeto de DW Relacional**

- Esquema estrela da atividade *Vendas*

**Produto**

ProdutoID (PK)  
 Atributos....

**Vendas**

DataID (FK)  
 ProdutoID (FK)  
 LojaID (FK)  
 PromoçãoID (FK)  
 NumeroTransação  
 QtdeVendida  
 TotalVendas  
 CustoVendas

**Data**

DataID (PK)  
 Atributos...

**Loja**

LojaID (PK)  
 Atributos...

Quais produtos da promoção que não foram vendidos?

↓

Tabela de Fatos Sem Fatos

Cln/UFPE - DW © 200

**Projeto de DW Relacional**

- Esquema estrela da atividade *Promoção sem Vendas*

**Produto**

ProdutoID (PK)  
 Atributos....

**SemFatos**

DataID (FK)  
 ProdutoID (FK)  
 LojaID (FK)  
 PromoçãoID (FK)

**Data**

DataID (PK)  
 Atributos...

**Loja**

LojaID (PK)  
 Atributos...

- Para determinar quais produtos da promoção não venderam
  - Consulta à tabela SemFatos para determinar quais produtos da promoção em um dado dia
  - Consulta à tabela Vendas para determinar quais produtos foram vendidos no mesmo dia
  - Obtem a diferença entre as duas listas de produtos

Cln/UFPE - DW © 201

**Projeto de DW Relacional**

- Dimensão Degenerada:
  - Consiste em uma dimensão vazia
    - Contém apenas uma chave que é incluída na tabela de fatos
  - Números de controle do ambiente operacional são geralmente exemplos deste tipo de dimensão
    - Número de Pedido
    - Número de Fatura
  - Se atributos adicionais forem identificados no futuro e forem relacionados com a dimensão degenerada
    - Uma nova dimensão normal pode ser criada e a dimensão degenerada removida

Cln/UFPE - DW © 202

**Projeto de DW Relacional**

- Dimensão Degenerada:
  - É útil para agrupar todos os produtos vendidos em uma única transação
  - É útil para identificar cada tupla da tabela de fatos

Chave = NumeroTransação + ProdutoID

**Produto**

ProdutoID (PK)  
 Atributos....

**Vendas**

DataID (FK)  
 ProdutoID (FK)  
 LojaID (FK)  
 PromoçãoID (FK)  
 NumeroTransação  
 QtdeVendida  
 TotalVendas  
 CustoVendas

**Data**

DataID (PK)  
 Atributos...

**Loja**

LojaID (PK)  
 Atributos...

Cln/UFPE - DW © 203

**Projeto de DW Relacional**

- Tabela de Fatos Periódica:
  - Fatos são inseridos periodicamente e independem da ocorrência de transações
  - Exemplo: Minimizar a indisponibilidade de produtos para serem vendidos
    - Precisa saber a quantidade de produtos existentes em cada loja
  - Atividade de negócio a ser modelada consiste em analisar o inventário de produtos de uma cadeia de lojas
  - Granularidade desta atividade de negócio consiste em obter o inventário diário por produto e por loja
    - Dimensões são: *Data*, *Produto* e *Loja*
  - Tabela de fatos contém a quantidade disponível em estoque

Cln/UFPE - DW © 204

**Projeto de DW Relacional**

- Esquema estrela da atividade *Inventário Periódico*

- Tabela de fatos Inventario difere da Tabela de fatos Vendas
  - Tabela de Fatos Vendas é esparsa
    - Apenas um percentual dos produtos de cada loja é vendido em cada dia
    - Se o produto não é vendido na loja em um dado dia
      - Nenhuma tupla para esta combinação de chaves é inserida na tabela de fatos

CIn/UFPE - DW 205

**Projeto de DW Relacional**

- Tabela de fatos Inventario difere da Tabela de fatos Vendas
  - Tabela de Fatos Inventario é densa
    - Para evitar estoques zerados nos quais o produto não está disponível para venda
      - Existe uma tupla na tabela de fatos para cada produto de cada loja e em cada dia
      - Se existem 60.000 produtos estocados em 100 lojas, 6 milhões de tuplas são inseridas em cada dia na tabela de fatos

Reduzir o tamanho dos dados:

- Últimos 60 dias de inventário ⇒ Nível diário
- Dados restantes ⇒ Nível semanal

CIn/UFPE - DW 206

**Projeto de DW Relacional**

- Tabela de fatos Inventario difere da Tabela de fatos Vendas
  - Tabela de Fatos Inventario contém medidas semi-aditivas
    - Quantidade estocada pode ser sumarizada com base nas dimensões Loja e Produto
    - Mas não pode com relação à dimensão Data
      - Fatos representam balanços feitos em um dado instante de tempo
      - Não podem ser somados ao longo do tempo
      - Isto também ocorre com outros tipos de balanços de saldo (i.e. Contas bancárias)
      - Pode-se obter a média, o mínimo ou o máximo do balanço realizado

CIn/UFPE - DW 207

**Projeto de DW Relacional**

- Tabela de Fatos Inventario contém medidas semi-aditivas
  - Todas as medidas que registram um nível estático (i.e. valor medido em um dado instante de tempo)
    - Balanços de estoque
    - Balanços de contas financeiras
    - Medidas de intensidade (e.g. temperatura)

São inerentemente não-aditivas ao longo da dimensão temporal!

Nestes casos, elas podem ser agregadas pelo eixo do tempo, a partir da média dos valores mensurados em um dado intervalo de tempo

CIn/UFPE - DW 208

**Projeto de DW Relacional**

- Regras de mapeamento: MultiDim → Estrela ou Flocos de neve
  - REGRA 1: Um nível N (que não possui um relacionamento 1:1 com o fato F) é mapeado para uma tabela TN que contém todos os atributos do nível
    - Uma chave artificial pode ser adicionado à tabela, ou o identificador do nível será a chave da tabela

CIn/UFPE - DW 209

**Projeto de DW Relacional**

- Regras de mapeamento: MultiDim → Estrela ou Flocos de neve
  - REGRA 1: É preferível usar chaves artificiais como identificadores de cada dimensão do DW
    - Garante independência das chaves dos sistemas operacionais
      - Códigos de ambientes operacionais mudam com o tempo
      - Códigos de produtos se tornam obsoletos
      - Números de contas se tornam inativas
    - Permite diferenciar instâncias de um mesmo elemento do ambiente operacional
    - Permite representar situações inexistentes no ambiente operacional
    - Chaves artificiais são usualmente representadas por inteiros
      - Facilita as operações de junção-estrela
      - Identificadores compostos não devem ser usados
    - Mas, chaves de sistemas operacionais devem ser mantidas no DW para manter seus dados associados aos dados fontes
      - Elas possuem semântica embutida nelas

CIn/UFPE - DW 210

**Projeto de DW Relacional**

- Regras de mapeamento: MultiDim → Estrela ou Flocos de neve
  - REGRA 2: Um fato F é mapeado para uma tabela TF que possui como atributos todas as medidas do fato
    - Uma chave artificial pode ser adicionado à tabela

Cln/UFPE - DW © 211

**Projeto de DW Relacional**

- Regras de mapeamento: MultiDim → Estrela ou Flocos de neve
  - REGRAS 1 e 2:
    - Para cada nível associado ao fato por um relacionamento 1:n, é inserida a FK do nível na tabela de fatos
      - Isto é feito para cada papel exercido pelo nível
    - A chave da tabela de fatos é composta pelas FK de todos os níveis associados ao fato
    - Se um identificador artificial for adicionado à tabela de fatos, a composição das FK torna-se uma chave alternada

Cln/UFPE - DW © 212

**Projeto de DW Relacional**

- Regras de mapeamento: MultiDim → Estrela ou Flocos de neve
  - REGRA 3: Um relacionamento entre uma fato F e um nível de dimensão N, ou entre um nível de dimensão pai NP e um nível de dimensão filho NF pode ser mapeado de 3 formas:
    - REGRA 3a:
      - Se o relacionamento é 1:1, a tabela de fatos (TF) ou a tabela do nível filho (TNF) é estendida com todos os atributos do nível de dimensão ou do nível pai, respectivamente.

Cln/UFPE - DW © 213

**Projeto de DW Relacional**

- Regras de mapeamento: MultiDim → Estrela ou Flocos de neve
  - REGRA 3: Um relacionamento entre uma fato F e um nível de dimensão N, ou entre um nível de dimensão pai NP e um nível de dimensão filho NF pode ser mapeado de 3 formas:
    - REGRA 3b:
      - Se o relacionamento é 1:n, a tabela de fatos (TF) ou a tabela do nível filho (TNF) é estendida com a FK do nível de dimensão ou do nível pai, respectivamente.

Cln/UFPE - DW © 214

**Projeto de DW Relacional**

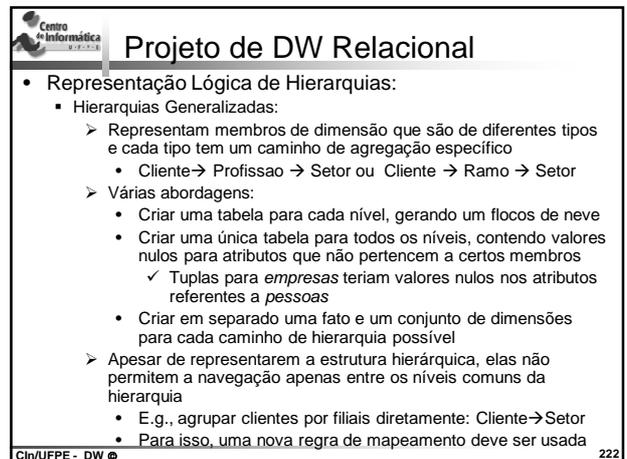
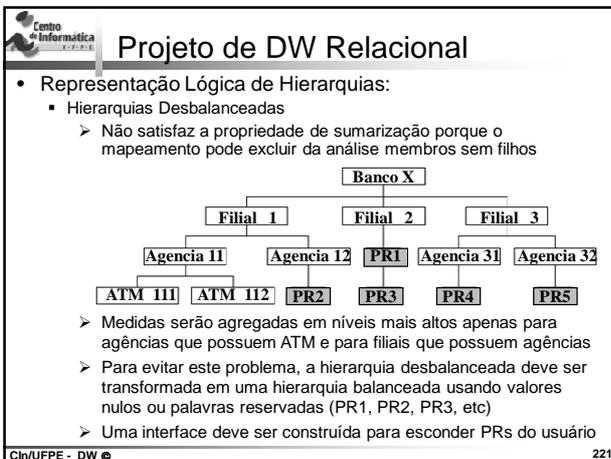
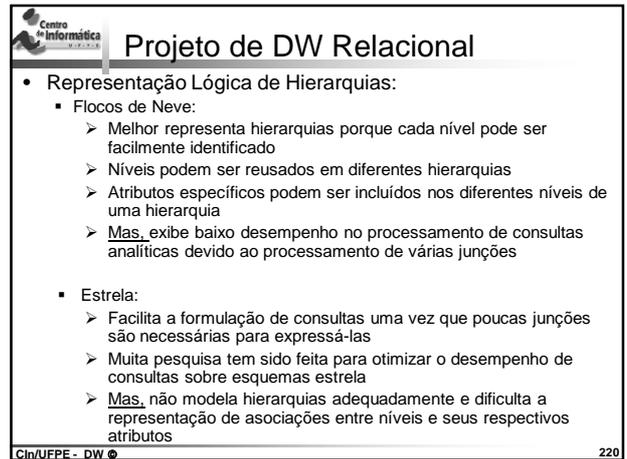
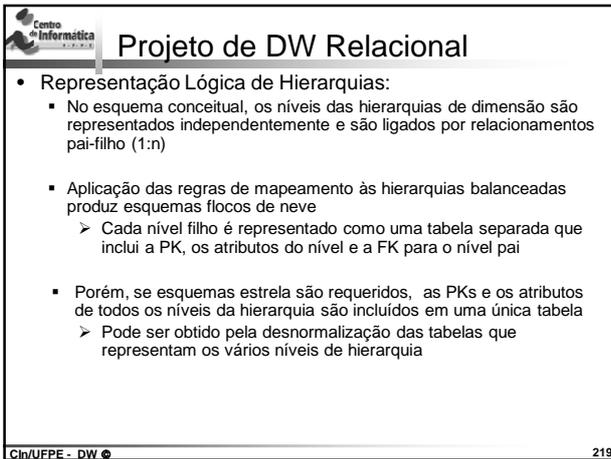
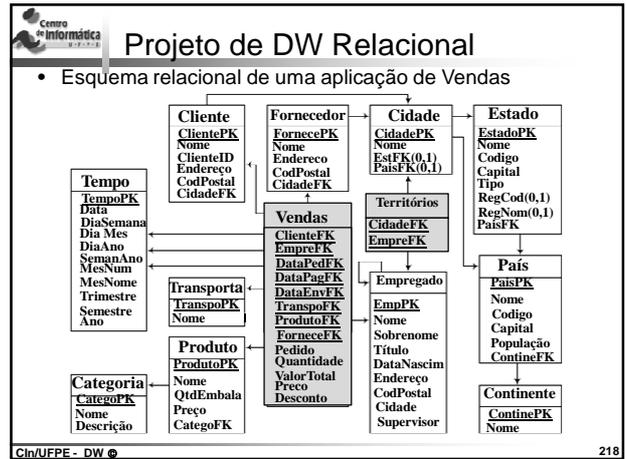
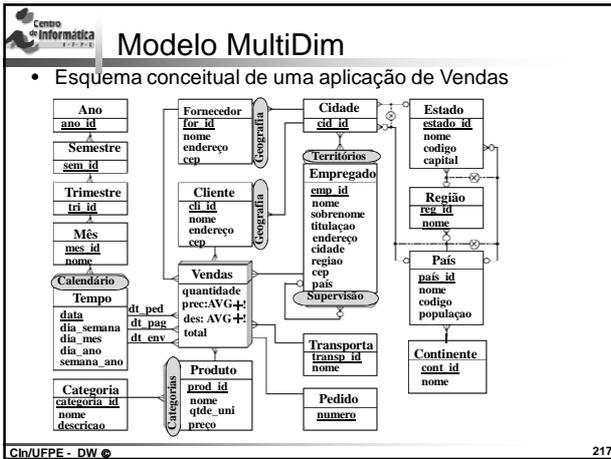
- Regras de mapeamento: MultiDim → Estrela ou Flocos de neve
  - REGRA 3: Um relacionamento entre uma fato F e um nível de dimensão N, ou entre um nível de dimensão pai NP e um nível de dimensão filho NF pode ser mapeado de 3 formas:
    - REGRA 3c:
      - Se o relacionamento é m:n, uma nova tabela T é criada cuja é composta pela: PK da tabela de fatos (TF) e do nível de dimensão ou PK da tabela do nível filho e PK da tabela do nível pai
      - Se o relacionamento possui um atributo de distribuição, um atributo adicional é adicionado à nova tabela criada para representá-lo
      - Exemplo: Um relacionamento n:m entre dois níveis *Cidade* e *Empregado* é mapeado para uma nova tabela cuja PK é composta e formada pela PK de *Cidade* e pela PK de *Empregado*

Cln/UFPE - DW © 215

**Projeto de DW Relacional**

- REGRA 3c: Exemplo
  - Diagram showing the transformation of a fact table 'Vendas' and dimension table 'Empregado' into a star schema. The fact table 'Vendas' has attributes: Quantidade, TotalVendas, Preço, Desconto. The dimension table 'Empregado' has attributes: EmpregadoID, Nome, Sobrenome, Título, DataNascimento, Endereço, CódigoPostal. The resulting star schema has a fact table 'Vendas' with primary key 'FatoPK' and foreign key 'EmpregadoFK', and a dimension table 'Empregado' with primary key 'EmpregadoPK' and foreign key 'EmpregadoFK'.

Cln/UFPE - DW © 216



### Projeto de DW Relacional

- Representação Lógica de Hierarquias:
  - Regra 4:
    - Uma tabela representando um nível de divisão em uma hierarquia generalizada deve ter um atributo discriminador que indica o caminho de agregação específico de cada membro
      - Cliente → Profissao → Ramo ou Cliente → Setor → Ramo
    - A tabela deve também conter um atributo adicional o qual é uma FK para o próximo nível de junção, caso este nível exista.
      - Cliente → Ramo

Cln/UFPE - DW © 223

### Projeto de DW Relacional

- Representação Lógica de Hierarquias:
  - Hierarquias Generalizadas:
    - A aplicação da Regra 4 promove caminhos diferentes de análise
      - Navegar em hierarquias com níveis específicos, tais como: Cliente → Profissao → Setor ou Cliente → Ramo → Setor
      - Acessar os níveis comuns a todos os membros: usando a hierarquia Cliente → Setor para analisar todos os clientes independente de seu tipo
    - Restrições de integridade devem ser definidas para garantir que apenas uma das FKs para níveis especializados possua um valor
    - Tal mapeamento também pode ser aplicado em hierarquias irregulares já que elas são um caso especial das hierarquias generalizadas

Cln/UFPE - DW © 224

### Projeto de DW Relacional

- Representação Lógica de Hierarquias:
  - Fatos com Múltiplas Granularidades:
    - Primeira Abordagem
      - Usar múltiplas FKs, uma para cada granularidade alternativa

- Ambos atributos CidadeFK e EstadoFK são opcionais
- Gatilhos devem ser definidos para garantir que apenas uma delas tenha um valor

Cln/UFPE - DW © 225

### Projeto de DW Relacional

- Representação Lógica de Hierarquias:
  - Fatos com Múltiplas Granularidades:
    - Segunda Abordagem
      - Remover variação de granularidade à nível de instância usando palavras reservadas

- Palavras reservadas (PR) são usadas para fatos que se referem aos níveis não-folha. Dois casos:
  - Um membro de fato aponta para um membro não-folha que tem filhos (i.e., PR1 representa todas as demais cidades)
  - Um membro de fato aponta para um membro não-folha sem filhos (i.e., PR2 representa todas as cidades do estado)

Cln/UFPE - DW © 226

### Projeto de DW Relacional

- Esquema estrela da atividade *Aluguel de Imóvel*

Cln/UFPE - DW © 227

### Projeto de DW Relacional

- Esquema estrela da atividade *Aluguel de Imóvel*

Cln/UFPE - DW © 228

**Projeto de DW Relacional**

- Muitas Dimensões?
  - Um grande número de dimensões no esquema é uma indicação que:
    - Várias dimensões não são completamente independentes
    - Elas devem ser combinadas em uma única dimensão
  - Consiste em um erro de modelagem representar elementos de uma hierarquia como dimensões separadas
  - Dimensões correlacionadas de um esquema (com várias dimensões) devem ser combinadas

CIn/UFPE - DW © 229

**Projeto de DW Relacional**

- Muitas Dimensões?
  - Diagrama de um esquema de fatos para "POS Retail Sales Transaction Fact" com 20 dimensões: Date, Week, Month, Quarter, Year, Fiscal Year, Fiscal Month, Store, Store County, Store State, Store District, Store Region, Store Floor Plan, Product, Brand, Subcategory, Category, Department, Package Type, Promotion, Promotion Reduction Type, Promotion Media Type. Um grande 'X' preto é sobreposto ao diagrama.

CIn/UFPE - DW © 230

**Projeto de DW Relacional**

- Várias tabelas de fatos?
  - Diagrama mostrando tabelas de fatos agrupadas por entidade: Produto, Fornecedor, Desconto, Funcionário, Pedido de Compra, Ordem de Compra, Nota de Transferencia, Recibo de Entrega, Ordem de pagamento, Data, Contrato, Condição Recebimento.

CIn/UFPE - DW © 231

**Projeto de DW Relacional**

- Várias tabelas de fatos?
  - Quais são os requisitos de análise dos usuários?
    - Eles usualmente analisam várias transações de uma vez ou cada uma em separado?
  - Existem várias atividades de um único negócio ou são várias atividades de negócio distintas?
  - Vários sistemas operacionais são usados para carga?
    - Compras, Estoque e Contas a Pagar
  - Qual é a dimensionalidade da tabela de fatos?
    - Todas as medidas são descritas pelas mesmas dimensões?

CIn/UFPE - DW © 232

**Projeto de DW Relacional**

- Dimensão Temporal:
  - Um DW é um BD histórico
    - A dimensão *Tempo* está presente em todo DW
  - Informação temporal é incluída como:
    - FK na tabela de fatos, indicando o tempo em que o fato ocorreu
    - Dimensão temporal que contém os níveis de agregação que indicam como os fatos podem ser agregados ao longo do tempo
  - A granularidade da dimensão temporal varia de acordo com a aplicação
    - Dados mensais → a granularidade da dimensão temporal = mês
    - DW com 5 anos teria 5 x 12 = 60 tuplas na dimensão temporal
    - Se a granularidade da dimensão temporal = segundos, então: 5 x 12 x 30 x 24 x 3600 = 155.520.000 tuplas
  - A dimensão temporal:
    - Geralmente, é povoada automaticamente
    - Podem ter mais de uma hierarquia (e.g. fiscal e de calendário)

CIn/UFPE - DW © 233

**Projeto de DW Relacional**

- Dimensão Temporal:
  - Exemplo: *Total das vendas durante fins de semana*

```
SELECT SUM (TotalValor)
FROM Tempo T, Vendas V
WHERE T.TempoPK = V.TempoFK AND
T.FimSemana
```
- É preciso satisfazer a condição de sumarização
  - Um dia pode ser agregado corretamente para o nível *Mes* ou *Ano*
  - Mas uma semana pode pertencer a diferentes meses. Por isso, membros do nível *Semana* não podem ser agregados para o nível *Mes* em uma hierarquia de dimensão temporal.

CIn/UFPE - DW © 234

## Projeto de DW Relacional

- Exemplo de uma Tabela de Dimensão Temporal

| Date Dimension               | Date Key | Date       | Full Date       | Day of Week | Calendar Month | Calendar Year | Fiscal Year-Month | Holiday Indicator | Weekday Indicator |
|------------------------------|----------|------------|-----------------|-------------|----------------|---------------|-------------------|-------------------|-------------------|
| Date Key (FK)                | 1        | 01/01/2002 | January 1, 2002 | Tuesday     | January        | 2002          | F2002-01          | Holiday           | Weekday           |
| Full Date Description        | 2        | 01/02/2002 | January 2, 2002 | Wednesday   | January        | 2002          | F2002-01          | Non-Holiday       | Weekday           |
| Day of Week                  | 3        | 01/03/2002 | January 3, 2002 | Thursday    | January        | 2002          | F2002-01          | Non-Holiday       | Weekday           |
| Week Number in Epoch         | 4        | 01/04/2002 | January 4, 2002 | Friday      | January        | 2002          | F2002-01          | Non-Holiday       | Weekday           |
| Month Number in Epoch        | 5        | 01/05/2002 | January 5, 2002 | Saturday    | January        | 2002          | F2002-01          | Non-Holiday       | Weekend           |
| Day Number in Calendar Month | 6        | 01/06/2002 | January 6, 2002 | Sunday      | January        | 2002          | F2002-01          | Non-Holiday       | Weekend           |
| Day Number in Fiscal Month   | 7        | 01/07/2002 | January 7, 2002 | Monday      | January        | 2002          | F2002-01          | Non-Holiday       | Weekday           |
| Day Number in Fiscal Year    | 8        | 01/08/2002 | January 8, 2002 | Tuesday     | January        | 2002          | F2002-01          | Non-Holiday       | Weekday           |

Cln/UFPE - DW 235

## Tópicos

- Modelo MultiDim
  - Notação gráfica
  - Hierarquias
  - Granularidade
  - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
  - Projeto de DW Relacional
  - Mudanças em Dimensões
- Data Marts
- Data Warehousing

Cln/UFPE - DW 236

## Mudanças em Dimensões

**Ambiente Operacional**  
Quando um dado é modificado, ele é sobrescrito

**Ambiente de DW**  
Dado não é modificado para registrar a evolução histórica e não distorcer os indicadores de análise

- Se um vendedor X que atuava na área 1 durante anos muda para área 3, então pode-se alterar este valor na dimensão vendedor?
  - Não, porque os valores que estavam historicamente associados à área 1 em função de X passam para 3 que não tinha estes valores antes
    - Inser-se uma nova ocorrência de X com a nova área

Cln/UFPE - DW 237

## Mudanças em Dimensões

- Suponha que:
  - Dados são acumulados de Janeiro a Maio das áreas 1 e 3
  - Vendedor X foi movido para área 3 em Maio

| Area | Vendedor | Janeiro | Fevereiro | Março | Abril | Maio |
|------|----------|---------|-----------|-------|-------|------|
| 1    | X        | 100     | 120       | 110   | 90    | 100  |
| 1    | Y        | 22      | 34        | 21    | 20    | 20   |
| 1    | Z        | 53      | 120       | 110   | 112   | 98   |
| 3    | W        | 33      | 44        | 55    | 66    | 77   |
| 3    | S        | 21      | 12        | 24    | 34    | 13   |
| 3    | D        | 45      | 43        | 46    | 56    | 34   |

Cln/UFPE - DW 238

## Mudanças em Dimensões

- Suponha que:
  - Dados são acumulados de Janeiro a Maio das áreas 1 e 3
  - Vendedor X foi movido para área 3 em Maio

| Area | Vendedor | Janeiro | Fevereiro | Março | Abril | Maio |
|------|----------|---------|-----------|-------|-------|------|
| 1    | X        | 100     | 120       | 110   | 90    | 100  |
| 1    | Y        | 22      | 34        | 21    | 20    | 20   |
| 1    | Z        | 53      | 120       | 110   | 112   | 98   |
| 3    | W        | 33      | 44        | 55    | 66    | 77   |
| 3    | S        | 21      | 12        | 24    | 34    | 13   |
| 3    | D        | 45      | 43        | 46    | 56    | 34   |

Cln/UFPE - DW 239

## Mudanças em Dimensões

- Se a área do Vendedor X for modificada de 1 para 3:
  - Total de vendas feitas por X não é modificado
  - Total de vendas por área sofreu mudança no histórico

| Area | Vendedor | Janeiro | Fevereiro | Março | Abril | Maio |
|------|----------|---------|-----------|-------|-------|------|
| 1    | Y        | 22      | 34        | 21    | 20    | 20   |
| 1    | Z        | 53      | 120       | 110   | 112   | 98   |
| 3    | W        | 33      | 44        | 55    | 66    | 77   |
| 3    | S        | 21      | 12        | 24    | 34    | 13   |
| 3    | D        | 45      | 43        | 46    | 56    | 34   |
| 3    | X        | 100     | 120       | 110   | 90    | 100  |

↑ ALTERAÇÃO ↑

Cln/UFPE - DW 240

## Mudanças em Dimensões

- Uma solução para o caso da mudança de área:
  - Facilidade de manutenção *versus* desempenho
  - Falta de necessidade de verificação de consistência de relacionamentos

CIn/UFPE - DW © 241

## Mudanças em Dimensões

- Pensava-se que dimensões eram independentes do tempo e estáveis e que apenas novos fatos eram inseridos no DW
  - Mudanças ocorrem no mundo real
    - Linhas de produto são reestruturadas, causando mudanças em hierarquias de produto
    - Clientes se mudam, causando mudanças em seus formatos de endereços
  - Não dá para registrar todas mudanças como fatos
  - Em muitas aplicações reais, dimensões podem mudar tanto a nível de estrutura quanto a nível de instância
  - Exemplo de mudanças estruturais:
    - Um atributo é removido das fontes de dados e portanto deve ser removido da tabela de dimensão

CIn/UFPE - DW © 242

## Mudanças em Dimensões

- Exemplos de mudanças à nível de instância:
  - Por causa de um erro, um novo dado deve substituir o antigo para fins de correção
  - Por causa de novas políticas administrativas ou comerciais, o conteúdo das dimensões deve mudar de acordo
- Assumindo que tais mudanças não são frequentes, dimensões que permitem tais mudanças são chamadas de slowly changing dimensions
- Existem várias abordagens:
  - Sobrescrita do valor
  - Adição de uma tupla e de atributos de validade
  - Adição de uma coluna para cada atributo propenso à mudança
  - Adição de uma mini-dimensão
  - .....
  - Temporal DW que permite registrar mudanças no esquema e nas instâncias das dimensões
- Verificar se o levantamento de requisitos foi bem realizado

CIn/UFPE - DW © 243

## Mudanças em Dimensões

- Exemplo: uma tabela de fatos *Vendas* associada às dimensões: *Tempo*, *Empregado*, *Cliente* e *Produto*

| TempoFK | EmpregadoFK | ClienteFK | ProdutoFK | ValorTotal |
|---------|-------------|-----------|-----------|------------|
| t1      | e1          | c1        | p1        | 100        |
| t2      | e2          | c2        | p1        | 100        |
| t3      | e1          | c3        | p3        | 100        |
| t4      | e2          | c4        | p4        | 100        |

| ProdutoPK | Nome  | Categoria | Descrição |
|-----------|-------|-----------|-----------|
| p1        | prod1 | cat1      | desc1     |
| p2        | prod2 | cat1      | desc1     |
| p3        | prod3 | cat2      | desc2     |
| p4        | prod4 | cat2      | desc2     |

- A categoria do produto p1 pode ser mudada para cat2

CIn/UFPE - DW © 244

## Mudanças em Dimensões

- Consulta: *Total de vendas por empregado e por categoria de produto*

| EmpregadoFK | Categoria | ValorTotal |
|-------------|-----------|------------|
| e1          | cat1      | 100        |
| e2          | cat1      | 100        |
| e1          | cat2      | 100        |
| e2          | cat2      | 100        |

- No instante  $t > t4$ , a categoria de p1 mudou para cat2
- Se a categoria for apenas sobrescrita, a mesma consulta retornaria:

| EmpregadoFK | Categoria | ValorTotal |
|-------------|-----------|------------|
| e1          | cat2      | 200        |
| e2          | cat2      | 200        |

- Se a nova categoria é o resultado de uma correção de carga (i.e., a categoria de p1 é de fato cat2, então o resultado da mudança está correto
- Senão, o resultado é incorreto: produtos afetados pela mudança de categoria já estavam associados às vendas

CIn/UFPE - DW © 245

## Mudanças em Dimensões

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 1: sobrescrever o valor antigo do atributo com o novo valor
    - Histórico da alteração de valores do atributo é perdida
    - Adequada apenas quando a alteração é feita para fins de correção
  - UPDATE Produto
    - SET Categoria = cat2
    - WHERE ProdutoPK = p1

| ProdutoPK | Nome  | Categoria | Descrição |
|-----------|-------|-----------|-----------|
| p1        | prod1 | cat2      | desc1     |
| p2        | prod2 | cat1      | desc1     |
| p3        | prod3 | cat2      | desc2     |
| p4        | prod4 | cat2      | desc2     |

CIn/UFPE - DW © 246

## Mudanças em Dimensões

- TIPO 1: Exemplo de Sobrescrita do Valor

| Chave Produto | Descrição | Departamento    | Código Produto |
|---------------|-----------|-----------------|----------------|
| 12345         | BrushKids | Higiene Pessoal | ABC922-Z       |

↓

| Chave Produto | Descrição | Departamento | Código Produto |
|---------------|-----------|--------------|----------------|
| 12345         | BrushKids | Infantil     | ABC922-Z       |

- Não há mudanças em chaves de qualquer tabela
- É a abordagem mais simples, tendo como vantagens a rapidez e facilidade de implementação
- Não mantém o histórico dos valores anteriores
- Adequada para realização de correções em carga de dados

Cln/UFPE - DW 247

## Mudanças em Dimensões

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 2: as tuplas da dimensão são versionadas e uma nova tupla é inserida cada vez que uma alteração ocorre
    - Garante uma associação correta entre as tuplas da fato e as tuplas da dimensão
    - Uma nova tupla é criada na dimensão *Produto* para o produto *prod1* da categoria *cat2*
    - A dimensão *Produto* é estendida com 2 atributos indicando o intervalo de validade da tupla

| ProdutoPK | Nome  | Categoria | Descrição |            |            |
|-----------|-------|-----------|-----------|------------|------------|
| p1        | prod1 | cat1      | desc1     | 01/01/2010 | 31/12/2011 |
| p11       | prod1 | cat2      | desc2     | 01/01/2012 | 31/12/9999 |
| p2        | prod2 | cat1      | desc1     | 01/01/2010 | 31/12/9999 |
| p3        | prod3 | cat2      | desc2     | 01/01/2010 | 31/12/9999 |
| p4        | prod4 | cat2      | desc2     | 01/01/2011 | 31/12/9999 |

Cln/UFPE - DW 248

## Mudanças em Dimensões

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 2: um atributo adicional pode ser adicionado à dimensão para explicitamente indicar a tupla atual
    - Na dimensão *Produto*, o atributo *Situação* indica o valor atual de *prod1*

| ProdutoPK | Nome  | Categoria | Descrição | De         | Para       |         |
|-----------|-------|-----------|-----------|------------|------------|---------|
| p1        | prod1 | cat1      | desc1     | 01/01/2010 | 31/12/2011 | Vencida |
| p11       | prod1 | cat2      | desc2     | 01/01/2012 | 31/12/9999 | Atual   |
| p2        | prod2 | cat1      | desc1     | 01/01/2010 | 31/12/9999 | Atual   |
| p3        | prod3 | cat2      | desc2     | 01/01/2010 | 31/12/9999 | Atual   |
| p4        | prod4 | cat2      | desc2     | 01/01/2011 | 31/12/9999 | Atual   |

- Um mesmo produto aparece na tabela de fatos por meio de tantos identificadores artificiais quantos forem o número de mudanças
- Por exemplo:
  - prod1 → aparece na fato como p1 antes da mudança
  - prod1 → aparece na fato como p11 depois da mudança

Cln/UFPE - DW 249

## Mudanças em Dimensões

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 2:
    - É necessário rastrear todas as tuplas que pertencem ao mesmo atributo para obtenção de resultados corretos
    - Exemplo: Contar o número de diferentes produtos vendidos pela empresa em determinado período de tempo não pode ser feito pela contagem de vezes que um produto aparece na fato
    - Outra desvantagem da abordagem TIPO 2 é que a dimensão pode crescer consideravelmente pois um novo registro é inserido cada vez que um atributo tem seu valor alterado
      - Isto pode degradar o desempenho de operações de junção estrela (i.e., junção entre a dimensão e a tabela de fatos)

Cln/UFPE - DW 250

## Mudanças em Dimensões

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 2: para a representação do tipo floco de neve (normalizada)

| ProdutoPK | Nome  | CategoriaFK |
|-----------|-------|-------------|
| p1        | prod1 | c1          |
| p2        | prod2 | c1          |
| p3        | prod3 | c2          |
| p4        | prod4 | c2          |

| CategoriaPK | Nome | Descrição |
|-------------|------|-----------|
| c1          | cat1 | desc1     |
| c2          | cat1 | desc1     |
| c3          | cat2 | desc2     |
| c4          | cat2 | desc2     |

- Suponha que agora *prod1* muda sua categoria para *cat2* e dois atributos são adicionados à dimensão *Produto*

| ProdutoPK | Nome  | CategoriaFK | De         | Para       |
|-----------|-------|-------------|------------|------------|
| p1        | prod1 | c1          | 01/01/2010 | 31/12/2011 |
| p11       | prod1 | c2          | 01/01/2012 | 31/12/9999 |
| p2        | prod2 | c1          | 01/01/2010 | 31/12/9999 |
| p3        | prod3 | c2          | 01/01/2010 | 31/12/9999 |
| p4        | prod4 | c2          | 01/01/2011 | 31/12/9999 |

Cln/UFPE - DW 251

## Mudanças em Dimensões

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 2: Se a mudança acontece em um nível mais acima da hierarquia, i.e. se a *Descrição* da categoria for alterada)?

| CategoriaPK | Nome | Descrição | De         | Para       |
|-------------|------|-----------|------------|------------|
| c1          | cat1 | desc1     | 01/01/2010 | 31/12/2011 |
| c11         | cat1 | desc11    | 01/01/2012 | 31/12/9999 |
| c2          | cat1 | desc1     | 01/01/2010 | 31/12/9999 |
| c3          | cat2 | desc2     | 01/01/2010 | 31/12/9999 |
| c4          | cat2 | desc2     | 01/01/2011 | 31/12/9999 |

- Uma nova tupla é inserida na tabela *Categoria* para inclusão da nova descrição de *cat1*
- Atributos que denotam a validade da tupla são adicionados
- A alteração precisa ser propagada para o nível mais baixo da hierarquia, i.e., precisa ser propagada para a tabela *Produto*

Cln/UFPE - DW 252

**Mudanças em Dimensões**

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 2:
    - Todas as vendas antes da mudança devem se referir à versão anterior de cat1 (com PK = c1) enquanto as novas vendas devem apontar para a nova versão de cat1 (com PK = c11)
    - Para isso, a tabela *Produto* também é alterada

| ProdutoPK | Nome  | CategoriaFK | De         | Para       |
|-----------|-------|-------------|------------|------------|
| p1        | prod1 | c1          | 01/01/2010 | 31/12/2011 |
| p11       | prod1 | c11         | 01/01/2012 | 31/12/9999 |
| p2        | prod2 | c1          | 01/01/2010 | 31/12/9999 |
| p3        | prod3 | c2          | 01/01/2010 | 31/12/9999 |
| p4        | prod4 | c2          | 01/01/2011 | 31/12/9999 |

- É possível também substituir 31/12/9999 por um marcador "agora" que denota que a tupla é ainda válida até o momento atual

Cln/UFPE - DW 253

**Mudanças em Dimensões**

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 3: uma coluna adicional é acrescentada para cada atributo de dimensão que seja considerado propenso à mudança
    - Esta coluna adicional conterá o novo valor do atributo
    - Apenas as 2 versões mais recentes do atributo são representadas
    - Intervalo de validade de tuplas não é guardado
    - Na dimensão *Produto*, os atributos *Categoria* e *Descrição* mudaram quando *prod1* mudou da categoria *cat1* para *cat2*
    - Duas novas colunas são adicionadas à dimensão *Produto*

| ProdutoPK | Nome  | Categoria | Nova Categoria | Descrição | Nova Descrição |
|-----------|-------|-----------|----------------|-----------|----------------|
| p1        | prod1 | cat1      | cat2           | desc1     | desc2          |
| p2        | prod2 | cat1      | Null           | desc1     | Null           |
| p3        | prod3 | cat2      | Null           | desc2     | Null           |
| p4        | prod4 | cat2      | Null           | desc2     | Null           |

Cln/UFPE - DW 254

**Mudanças em Dimensões**

- TIPO 3: Exemplo de Adição de uma coluna

| Chave Produto | Descrição | Novo Depto | Antigo Depto    | Codigo   |
|---------------|-----------|------------|-----------------|----------|
| 12345         | BrushKids | Infantil   | Higiene Pessoal | ABC922-Z |

- Uma nova coluna na tabela de dimensão é inserida para refletir a mudança no valor do atributo.
- Fatos antigos e novos podem ser sumarizados em termos dos antigos e novos valores de atributo
- Não permite análise sobre o impacto das inúmeras mudanças feitas ao valor do atributo

Cln/UFPE - DW 255

**Mudanças em Dimensões**

- Tipo 2 e 3: Abordagem híbrida

| Chave | Descrição | Atual Depto     | Historico Depto | Codigo   |
|-------|-----------|-----------------|-----------------|----------|
| 12345 | BrushKids | Higiene Pessoal | Higiene Pessoal | ABC922-Z |

Cln/UFPE - DW 256

**Mudanças em Dimensões**

- Tipo 2 e 3: Abordagem híbrida

| Chave | Descrição | Atual Depto | Historico Depto | Codigo   |
|-------|-----------|-------------|-----------------|----------|
| 12345 | BrushKids | Infantil    | Higiene Pessoal | ABC922-Z |
| 25984 | BrushKids | Infantil    | Infantil        | ABC922-Z |

Cln/UFPE - DW 257

**Mudanças em Dimensões**

- Tipo 2 e 3: Abordagem híbrida

| Chave | Descrição | Atual Depto | Historico Depto | Codigo   |
|-------|-----------|-------------|-----------------|----------|
| 12345 | BrushKids | Educacional | Higiene Pessoal | ABC922-Z |
| 25984 | BrushKids | Educacional | Infantil        | ABC922-Z |
| 31726 | BrushKids | Educacional | Educacional     | ABC922-Z |

- Permite registrar o histórico das mudanças
- Permite sumarizar fatos de acordo com qualquer valor do atributo modificado
- As abordagens vistas anteriormente são apropriadas para *slowly changing dimensions* apenas!
  - Se há muitas alterações pode ser uma indicação de que o levantamento de requisitos não foi bem feito

Cln/UFPE - DW 258

**Mudanças em Dimensões**

- Quatro formas básicas de tratar slowly changing dimensions:
  - TIPO 4: projetada para dimensões volumosas que mudam frequentemente e para minimizar custos de junção estrela
    - Uma nova dimensão, chamada MiniDimensão, é criada para manter os atributos que mudam mais frequentemente
    - Em *Produto*, os atributos *IntervaloPreço* e *RanqueamentoPreço* mudam frequentemente para acompanhar as condições do mercado
      - Existirá uma tupla na minidimensão para cada combinação única de *IntervaloPreço* e *RanqueamentoPreço*
      - FK da minidimensão é adicionada à tabela de fatos

| MiniProdutoPK | IntervaloPreço | RanqueamentoPreço |
|---------------|----------------|-------------------|
| mp1           | 1 - 100        | 1                 |
| mp2           | 1 - 100        | 2                 |
| ...           | ...            | ...               |
| mp200         | 500 - 600      | 7                 |

CIn/UFPE - DW 259

**Tópicos**

- Modelo MultiDim
  - Notação gráfica
  - Hierarquias
  - Granularidade
  - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
  - Projeto de DW Relacional
  - Mudanças em Dimensões
- Data Marts
- Data Warehousing

CIn/UFPE - DW 260

**Data Marts**

- O que é um Data Mart (DM)?
  - DM é um subconjunto de um DW
    - Subconjunto do DW que satisfaz os requisitos de um certo tema ou atividade de negócio
    - Projetado para um dado grupo de usuários
  - Departamentos organizacionais querem apenas uma parte do DW corporativo → DM = DW departamental
    - Vendas → dados sobre vendas
    - RH → dados sobre funcionários e dados demográficos
  - Não é exclusivo de um departamento ou uma divisão
  - Tem o propósito de qualquer DW
    - Permitir consultas sobre dados históricos e navegação em diferentes hierarquias de dados

CIn/UFPE - DW 261

**Data Marts**

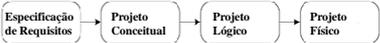
- Comparativo entre DW e DM

| Propriedade            | Data Warehouse | Data Mart        |
|------------------------|----------------|------------------|
| Escopo                 | Empresa        | Departamento     |
| Assuntos               | Vários         | Um único assunto |
| Fonte de Dados         | Várias         | Poucas           |
| Tempo de Implementação | Meses a anos   | Meses            |

- DM podem ser construídos antes ou depois do DW, o qual pode ser visto como uma coleção de DM

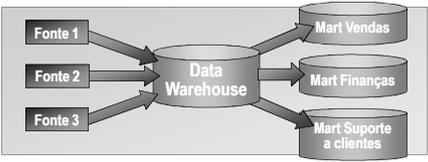
CIn/UFPE - DW 262

**Data Marts**

- Abordagens de Projeto de DW:
 
- Top-down:
  - Requisitos de dados coletados para diferentes níveis organizacionais são integrados
  - Um esquema para o DW corporativo é construído, a partir do qual os DM são obtidos
- Bottom-up:
  - Um esquema é construído para cada DM, de acordo com os requisitos dos usuários de cada área de negócio ou departamento
  - Esquemas de DM são integrados em um esquema de DW global

CIn/UFPE - DW 263

**Data Marts**

- Movendo dados de um DW para DM: Top Down
 
- Se DM são construídos depois do DW
  - Produz uma visão integrada dos dados
    - Campos compartilhados
  - Aumenta a complexidade de desenvolvimento
    - Mais custoso e demorado
  - Requer um maior planejamento prévio

CIn/UFPE - DW 264

### Centro de Informática

## Data Marts

- Movendo dados de DM para um DW: Bottom Up
 
- Se DM são construídos antes do DW
  - Pode-se criar DM incompatíveis
    - Fragmentos de dados ou replicação de dados
    - Falta de garantia de padrões únicos de metadados
  - Reduz a complexidade de desenvolvimento
    - Menos custoso e mais rápido

CIn/UFPE - DW © 265

### Centro de Informática

## Data Marts

- Estratégia Combinada: Top Down + Bottom Up
  - Efetua-se a modelagem de dados do DW de visão macro
  - Inicia um processo de implementação de partes deste modelo

A principal vantagem desta abordagem é a garantia da consistência dos dados!
- Características da Implementação Combinada
  - Planejamento Top Down
    - Atividades de Negócio
  - Desenvolvimento Bottom Up
    - Um DM de cada vez (incremental)
  - Gestão de metadados e de manutenção
    - Coerência entre os vários DM

CIn/UFPE - DW © 266

### Centro de Informática

## Data Marts

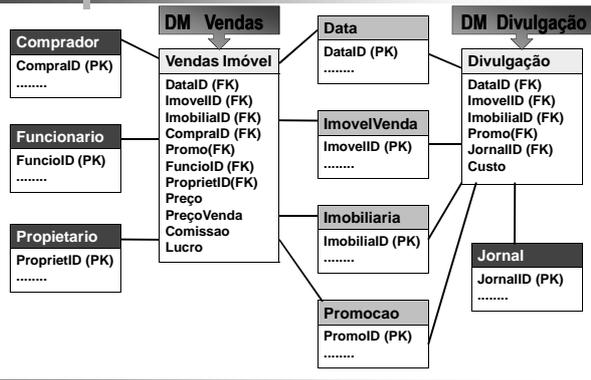
- DM Integrados
  - Dimensões definem o contexto das consultas sobre os fatos analisados
  - O projeto bem elaborado de um conjunto de dimensões facilita o entendimento e uso do DM
    - Um conjunto incompleto de dimensões limita o universo de análises que podem ser feitas
  - Se duas dimensões existem em dois DM, então:
    - Elas devem ser exatamente as mesmas dimensões ou
    - Uma deve ser um subconjunto da outra

Dois DM podem compartilhar 2 ou mais dimensões em uma mesma aplicação

CIn/UFPE - DW © 267

### Centro de Informática

## Data Marts



CIn/UFPE - DW © 268

### Centro de Informática

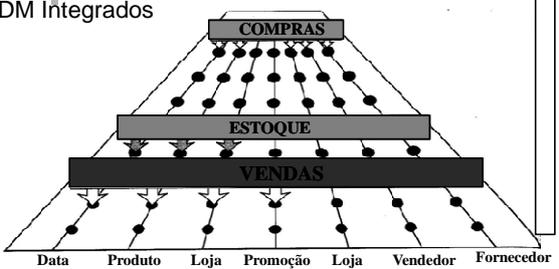
## Data Marts

- Compartilhamento de dados é essencial
  - Geração de relatórios consistentes a partir de diferentes DM
  - Aumento de desempenho no desenvolvimento do sistema a partir do reuso de componentes
- Tabelas de fatos mantidas em DM separados representam várias atividades de negócio
- Uso de dimensões compartilhadas permite a integração dos DM
  - Medidas de desempenho de várias atividades de negócio podem ser exibidas em um único relatório
  - Drill Across permite a combinação de resultados de consultas feitas à DM distintos por meio do atributo comum da dimensão compartilhada

CIn/UFPE - DW © 269

### Centro de Informática

## Data Marts

- DM Integrados
 
- Consiste em um conjunto de modelos dimensionais que compartilham um conjunto de dimensões
- Possui várias atividades de negócio acopladas

CIn/UFPE - DW © 270

**Data Marts**

- Matriz DM Integrados
  - Usada p/ criar, documentar e divulgar a arquitetura de DW

| Atividades de Negócio | Dimensões de DM Integrados |         |      |          |          |            |
|-----------------------|----------------------------|---------|------|----------|----------|------------|
|                       | Data                       | Produto | Loja | Promoção | Vendedor | Fornecedor |
| Vendas                | X                          | X       | X    | X        |          |            |
| Inventário            | X                          | X       | X    |          |          |            |
| Entrega               | X                          | X       | X    |          |          |            |
| Compras               | X                          | X       |      |          | X        | X          |

- Colunas são dimensões compartilhadas e linhas são DM
- Possibilita a visualização de quais dimensões merecem atenção especial por participarem de vários DM

CIn/UFPE - DW © 271

**Data Marts**

- Dimensões de DM integrados possuem:
  - Conteúdo de dados
  - Definição de dados
  - Formatos de exibição
 } Consistentes
- Dimensões de DM integrados:
  - São definidas no maior nível de detalhe possível
  - Podem pertencer a esquemas de atividades de negócio definidas em granularidades diferentes
    - Basta que uma dimensão seja um subconjunto da outra
    - Antes de seus esquemas serem comparados, uma delas é sumarizada para o mesmo nível de detalhe da outra

CIn/UFPE - DW © 272

**Data Marts**

- Exemplo:
  - Vendas ⇒ captura dados no nível de produto atômico
  - Previsão ⇒ gera dados no nível marca de produto

A tabela de dimensão Produto só pode ser compartilhada entre os esquemas destas 2 atividades de negócio se:

A tabela Produto do esquema Previsão for um subconjunto da tabela Produto do esquema Vendas!

Atributos comuns de ambas tabelas (i.e. marca e categoria) devem ser rotulados, definidos e instanciados identicamente!

CIn/UFPE - DW © 273

**Data Marts**

- Exemplo de dimensões de DM Integrados

| ProdutoVendas                                                                                                                                                     | ProdutoPrevisao                                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|
| ProdutoID (PK)<br>Descrição<br>Numero<br>Marca<br>Subcategoria<br>Categoria<br>Departamento<br>TipoEmbalagem<br>TamanhoEmbalagem<br>Peso<br>Demais Atributos..... | MarcaID (PK)<br>Marca<br>Subcategoria<br>Categoria<br>Departamento |

Integrada

- Operação *DrillAcross* só pode ser realizada por meio de:
  - atributos comuns das tabelas de dimensões integradas
  - tuplas comuns das tabelas para não violar a restrição de integridade referencial

CIn/UFPE - DW © 274

**Data Marts**

- Uso de dimensões de DM Integrados requer
  - Coordenação de implementação para que adição ou modificação de atributos seja revista
  - Disciplina na nomeação do dados
  - Estratégia de liberação da dimensão para área de apresentação
    - Dimensões idênticas devem ser replicadas de forma síncrona para todos DM associados
  - Não é sempre o caso que o uso de DM Integrados é aplicável
    - Pode não haver necessidade de relacionar dados de uma atividade de negócio com dados de outra
    - DW separados podem ser construídos

CIn/UFPE - DW © 275

**Tópicos**

- Modelo MultiDim
  - Notação gráfica
  - Hierarquias
  - Granularidade
  - Dimensão
- Mapeamento ER-MultiDim
- Projeto Lógico de DW
  - Projeto de DW Relacional
  - Mudanças em Dimensões
- Data Marts
- Data Warehousing

CIn/UFPE - DW © 276

