

MBA em Gestão da Informação e *Business Intelligence*

Arquitetura de Data Warehouse

Ricardo Holanda, MSc.
Ago/2014



Conteúdo Programático

UNIDADE IV: PROJETO DE DADOS EM UM DW

- Princípios de modelagem dimensional e multidimensional;
- Esquemas Star e Snowflakes;
- Extração de dados, transformação e carga(ETL);
- Qualidade dos dados armazenados;
- Ferramentas e Operações OLAP;
- Modelo OLAP e variações: ROLAP e MOLAP.



Business School Brasil
O futuro é agora. Junte-se a nós.



Visão Multidimensional dos Dados

- **Modelo Empresarial:** Deve ser desenvolvido a partir da análise do modelo de dados normalizado do negócio. Nesta fase o importante é ter o foco na estrutura da informação
- **Modelo Dimensional:** Algumas perguntas complexas normalmente requerem uma perspectiva diferente na visão dos dados. O modelo dimensional tentar tornar real essa perspectiva peculiar
- **Modelo Físico:** Possui o propósito de obter desempenho. Como esse problema existe, a elaboração de um modelo físico bom e um não tão bom pode ser a diferença entre o projeto do DW dar certo ou não.



Business School Brasil
O Estudo É Sua Propriedade



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Visão Multidimensional dos Dados

Modelagem Multidimensional.

- Requisitos diferentes das aplicações do ambiente transacional:
 - ✓ Flexibilidade quanto às análises a suportar;
 - ✓ Medidas a analisar precisam ser vistas sob diferentes perspectivas (dimensões);
- Enfoque diferente da modelagem no ambiente operacional;
- Abordagem utilizada:
 - ❖ Modelagem Multidimensional.

Visão Multidimensional dos Dados

Modelagem Multidimensional

- Modela dados para o processo decisório;
- É o coração do DW;
- Viabiliza a evolução harmoniosa do sistema de informações gerenciais;
- Surgiu na década de 80, em vários lugares, simultaneamente;
- Organiza grandes volumes de dados.



Visão Multidimensional dos Dados

Modelagem DW para OLAP:

- Facilita o entendimento e visualização de problemas típicos de suporte à decisão;
- Mais intuitiva para o processamento analítico;
- Utilizada pelas ferramentas OLAP.

A visão lógica é multidimensional, embora a estrutura física possa ter a mesma visão tabular do modelo relacional.



Business School Brasil
O Estudo Língua Língua



Visão Multidimensional dos Dados

- Foco no cruzamento das informações:
 - ✓ Facilita o entendimento e visualização de problemas típicos de suporte à decisão;
 - ✓ Mais intuitiva para o processamento analítico;
 - ✓ Utilizada pelas ferramentas OLAP.

Qual a diferença da visão multidimensional para a visão “tabular” do modelo relacional?



Visão Multidimensional dos Dados

Estrutura Relacional:

MODEL	CITY	SALES VOLUME
MINI VAN	NEW YORK	6
MINI VAN	LOS ANGELES	5
MINI VAN	MADISON	4
SPORTS COUPE	NEW YORK	3
SPORTS COUPE	LOS ANGELES	5
SPORTS COUPE	MADISON	5
SEDAN	NEW YORK	4
SEDAN	LOS ANGELES	3
SEDAN	MADISON	2

Visão Multidimensional dos Dados

Visão Matricial ou Multidimensional

The diagram shows a 3x3 matrix representing data for three car models (Mini Van, Coupe, Sedan) across three cities (NY, LA, Madison). The matrix is labeled with 'MODEL' vertically on the left and 'CITY' horizontally at the bottom.

6	5	4
3	5	5
4	3	2

Dimensions:

- Vertical dimension (left): MODEL
- Horizontal dimension (bottom): CITY
- Depth dimension (rows): NY, LA, Madison

- ✓ Um array multidimensional tem um número fixo de dimensões e os valores são armazenados nas células;
- ✓ Cada dimensão consiste de um número de elementos.



Visão Multidimensional dos Dados

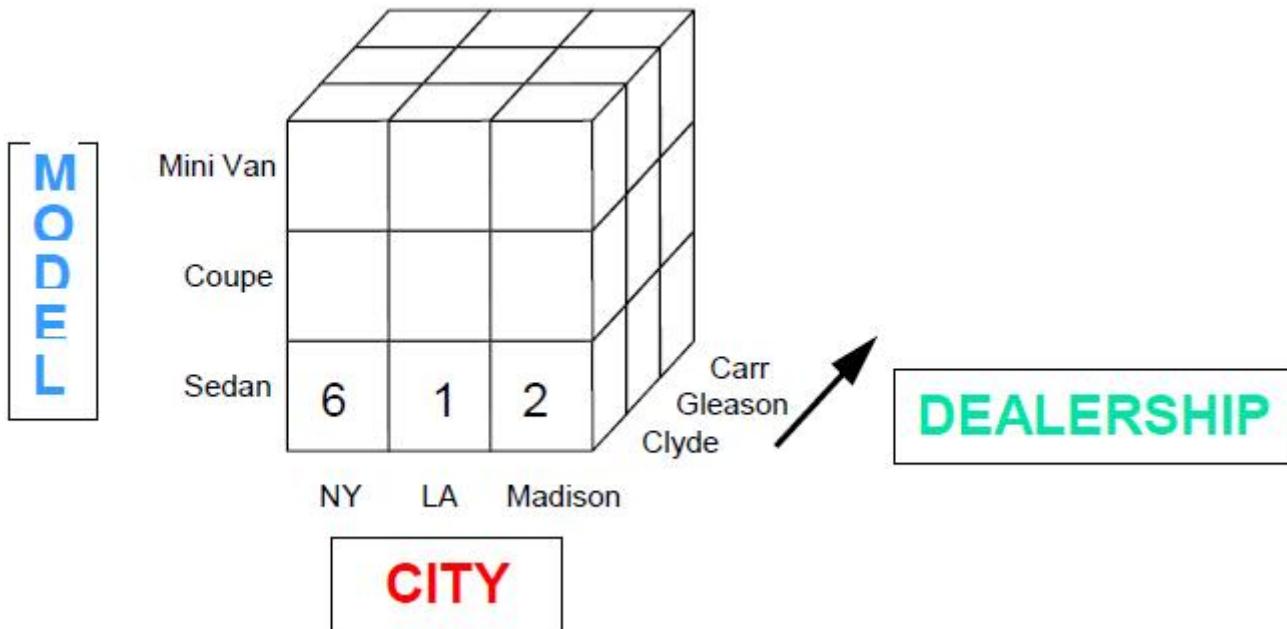
E se acrescentarmos mais uma coluna?

Volume de venda
por concessionária.

MODEL	CITY	DEALERSHIP	VOLUME
MINI VAN	NEW YORK	CLYDE	6
MINI VAN	NEW YORK	GLEASON	6
MINI VAN	NEW YORK	CARR	2
MINI VAN	LOS ANGELES	CLYDE	3
MINI VAN	LOS ANGELES	GLEASON	5
MINI VAN	LOS ANGELES	CARR	5
MINI VAN	MADISON	CLYDE	2
MINI VAN	MADISON	GLEASON	4
MINI VAN	MADISON	CARR	3
SPORTS COUPE	NEW YORK	CLYDE	2
SPORTS COUPE	NEW YORK	GLEASON	3
SPORTS COUPE	NEW YORK	CARR	2
SPORTS COUPE	LOS ANGELES	CLYDE	7
SPORTS COUPE	LOS ANGELES	GLEASON	5
SPORTS COUPE	LOS ANGELES	CARR	2
SPORTS COUPE	MADISON	CLYDE	4
SPORTS COUPE	MADISON	GLEASON	5
SPORTS COUPE	MADISON	CARR	1
SEDAN	NEW YORK	CLYDE	6
SEDAN	NEW YORK	GLEASON	4
SEDAN	NEW YORK	CARR	2
SEDAN	LOS ANGELES	CLYDE	1
SEDAN	LOS ANGELES	GLEASON	3
SEDAN	LOS ANGELES	CARR	4
SEDAN	MADISON	CLYDE	2
SEDAN	MADISON	GLEASON	2
SEDAN	MADISON	CARR	3



Visão Multidimensional dos Dados



- ✓ O cubo é, de fato, apenas uma metáfora visual.
- ✓ É uma representação intuitiva do fato porque todas as dimensões coexistem para todo ponto no cubo e são independentes umas das outras.



Business School Brasil
O futuro é agora, juntos!



PUC
GOIÁS

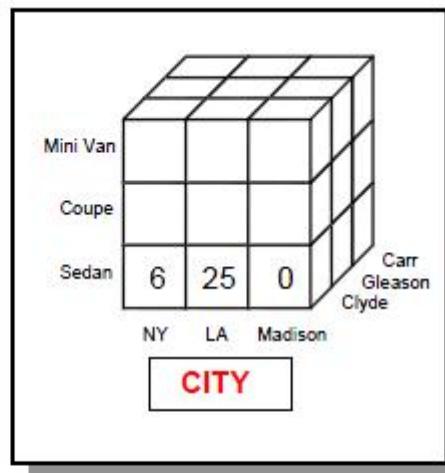
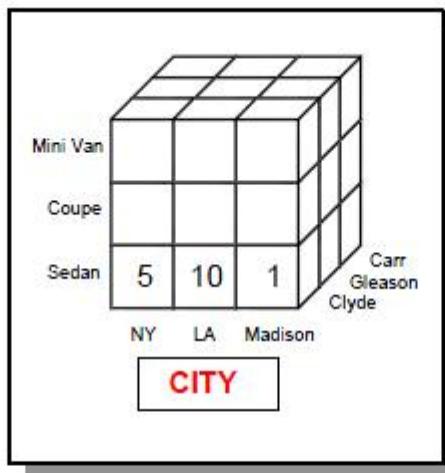
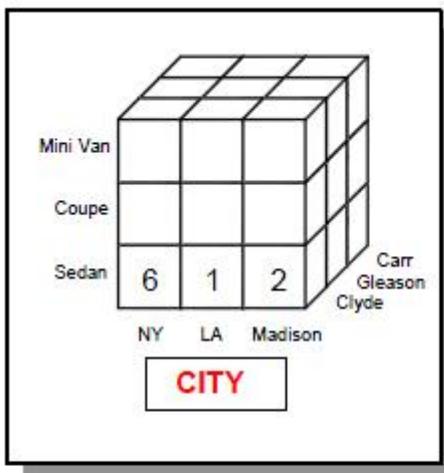


UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Visão Multidimensional dos Dados

Como faremos para visualizar mais de 3D?

MODEL



DEALERSHIP

JANUARY

FEBRUARY

MARCH

Hipercubo



Business School B
O futuro é agora, juntos!



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Visão Multidimensional dos Dados

Modelagem Multidimensional

- Utilização dos conceitos do modelo multidimensional a fim de representar, de forma clara, eficiente e flexível, a visão multidimensional dos dados.
- Conceitos
 - ❖ Fatos
 - ❖ Dimensões
 - Hierarquias e Agregações;



Visão Multidimensional dos Dados

Conceitos Iniciais

- Fatos
 - ❖ Geralmente os fatos são informações numéricas do negócio. Os fatos são o foco da curiosidade do cliente. Esse fatos estão dispostos num determinado grão, que é o menor nível de informação disponível.
- Dimensões
 - ❖ São as informações pelas quais se é permitido analisar um fato. Na ferramenta OLAP ela são caracterizadas pelos Atributos. A quantidade de dimensões associadas a um fato definem seu grão.



Business School Brasil
O futuro é agora. Junte-se a!



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Visão Multidimensional dos Dados

Conceitos Iniciais

- Métricas
 - ❖ São fórmulas matemáticas que atuam sobre os fatos, de forma a permitir a agregação de muitos registros em um único registro.
- Hierarquias
 - ❖ É um agrupamento de atributos relacionados. Em nível de negócio estabelecem uma relação de pai para filho (não necessariamente em nível físico). Ex: Ano → Mês → Dia



Visão Multidimensional dos Dados

A modelagem dimensional é a única etapa de um ciclo de vida de DW?



Business School Brasil
O futuro começa aqui!



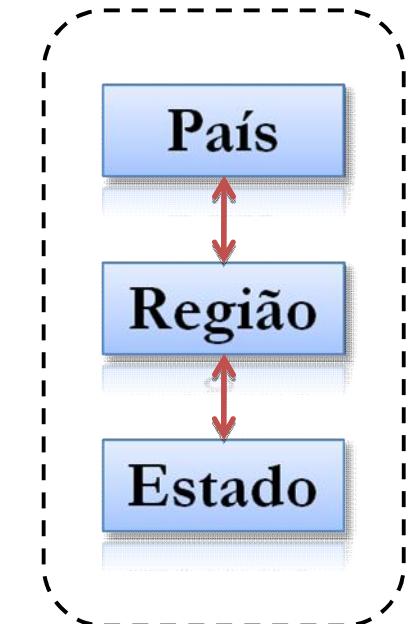
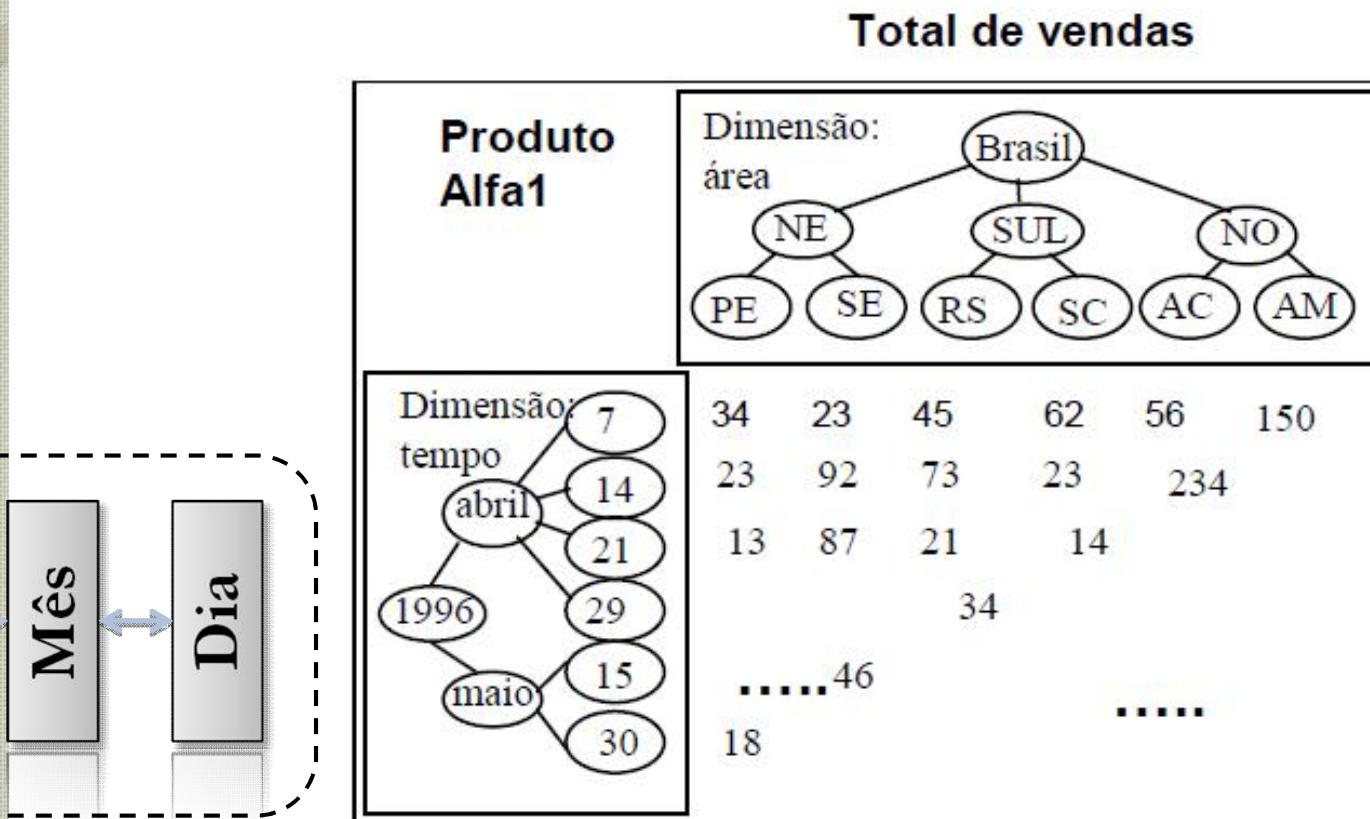
PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Agregados e hierarquias de dimensões

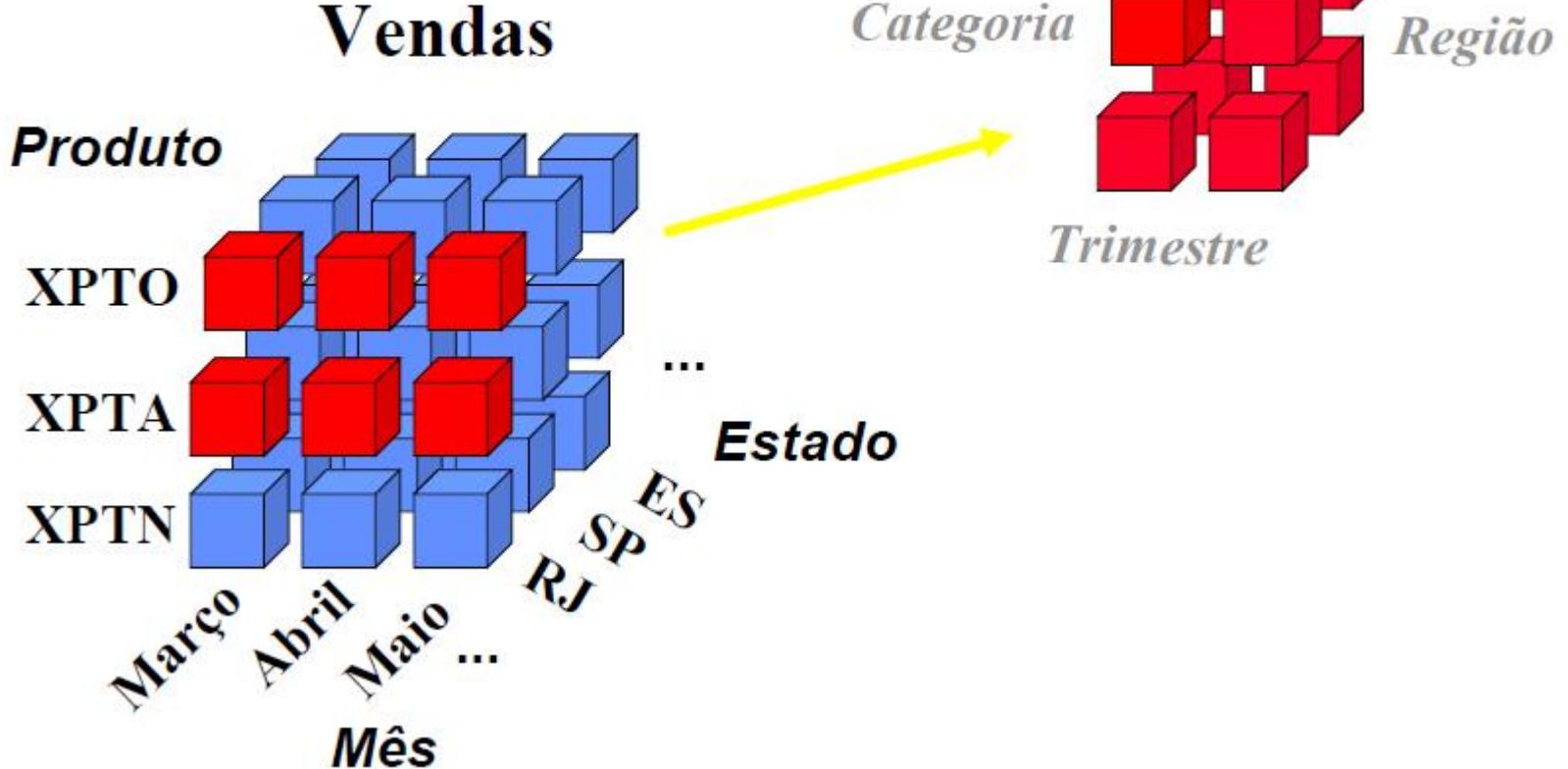
Níveis nas dimensões ou Hierarquias:



- Hierarquias são a base das agregações

Agregados e hierarquias de dimensões

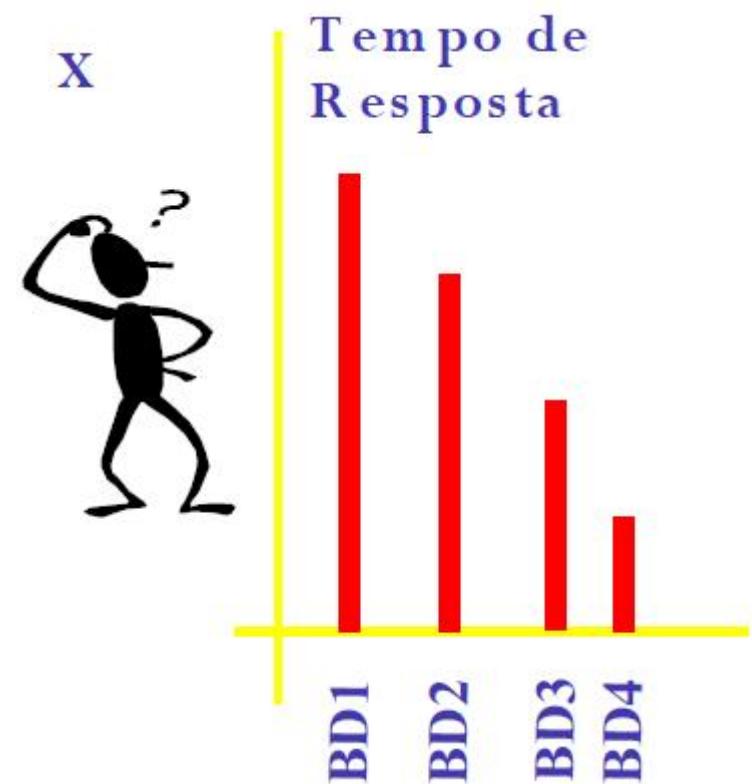
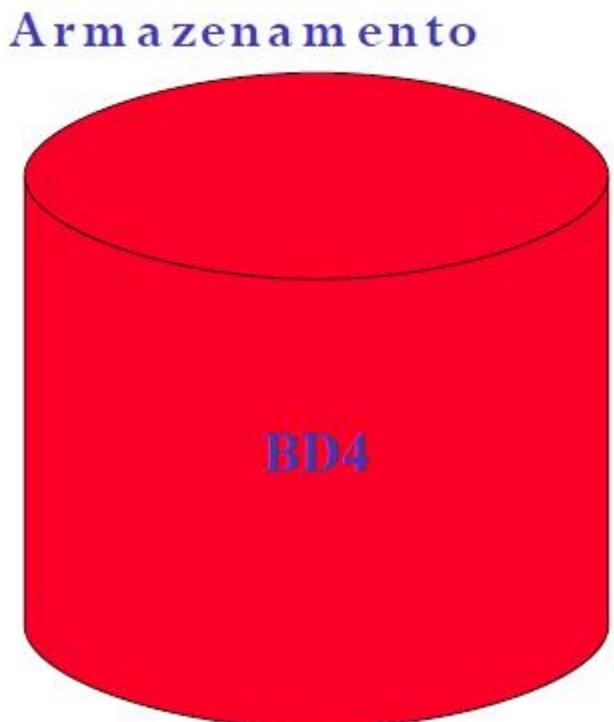
Agregados:



Agregados e hierarquias de dimensões

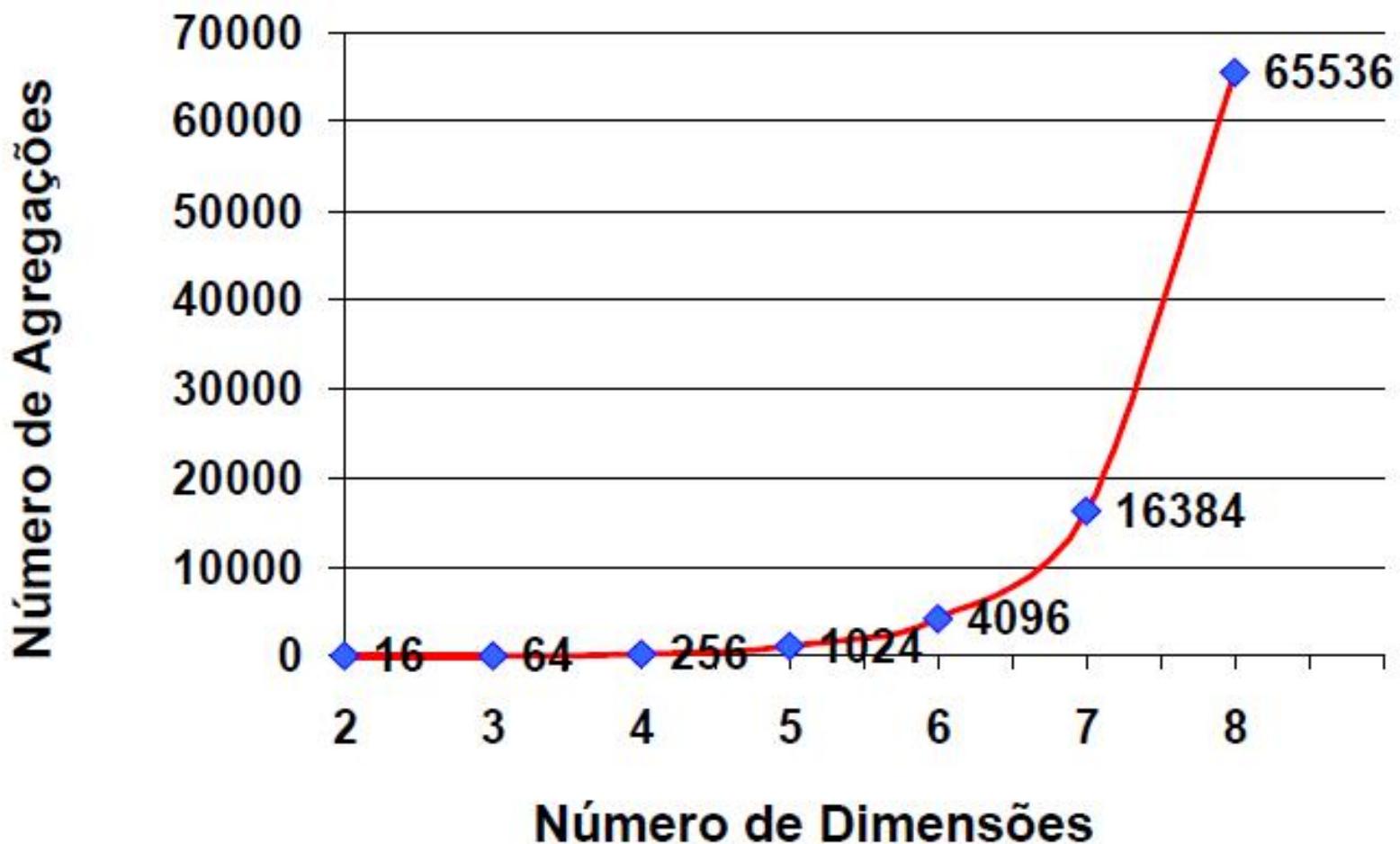
Problemas:

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



Agregados e hierarquias de dimensões

Explosão de Dados:



Visão Multidimensional dos Dados

Agregados

- As hierarquias permitem que o usuário possa ter acesso a dados com maior ou menor detalhe;
- Os valores apresentados quando o analista consulta dados em níveis hierárquicos mais altos são valores **agregados**.



Agregados e hierarquias de dimensões

Hierarquia e Agregados:

Produto

Tempo

Geografia

Consultas

Marca

Ano

País

Vendas por
Produto, Ano e
Região

Categ.

Mês

Região

Vendas por
Marca, Mês e
Estado

Produto

Dia

Estado



Business School Brasil
O maior projeto educativo do Brasil



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Visão Multidimensional dos Dados

Agregados

- Materializar (armazenar) ou não?
 - ✓ Vide síndrome da explosão do volume de dados.
- Critérios para definição de agregados
 - ✓ Passam pela análise dos principais tipos de informação necessárias e pela dificuldade de se obtê-las diretamente das tabelas granulares.
 - ✓ Exemplo:
 - TDLoja (chave-loja, nome-loja, endereco-loja, cidade, estado, **regiao**)
 - TDProduto (chave-produto, descricao, marca, **categoria, tipo-embalagem, departamento**)
 - TDDia (chave-dia, data-completa, dia, **mês, ano, período-fiscal, estação**)
 - TFVendas (chave-loja, chave-produto, chave-dia, valor-vendido-real, custo-real, lucro, qtdvendida)

Visão Multidimensional dos Dados

Agregados

- Hierarquias de dimensões

- ✓ Região → Loja
- ✓ Categoria → Produto
- ✓ Ano → Mês → Dia

- Combinações possíveis (35 combinações)

- ❖ Ternárias:

- Loja x Produto x Dia → $2 \times 2 \times 3 = 12$ combinações

- ❖ Binárias:

- Loja x Produto + Loja x Dia + Produto x Dia = 16 combinações

- ❖ Unárias:

- Loja + Produto + Dia = 7 combinações



Business School Brasil
O futuro é agora. Junte-se a!



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Visão Multidimensional dos Dados

Cuidados na definição dos agregados

- **Valores aditivos**

- ✓ Nem todas as métricas armazenadas nas tabelas granulares são aditivas em todas as dimensões (fatos semi-aditivos ou não aditivos). Isto significa que os atributos das tabelas fatos de agregados poderão ser diferentes das tabelas fatos granulares.

- **Precisão**

- ✓ Deve-se definir criteriosamente a precisão dos valores aditivos de agregados, que deverão ser maiores do que os usados nos respectivos valores das tabelas granulares (para evitar overflow na adição)
 - ✓ Fatos e dimensões agregados devem estar em tabelas fisicamente diferentes das tabelas granulares, mesmo que o número de tabelas cresça muito. Ferramentas de análise (OLAP, por exemplo) possuem mecanismo de navegação de agregados que escondem a complexidade da estrutura.



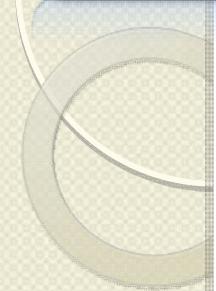
Visão Multidimensional dos Dados

- **Cuidados Operacionais**

- ✓ Modelos separados (agregados e granulares) para evitar contenções mútuas no momento de carga ou atualização;
- ✓ Carga total versus Atualização incremental: Tempo de processamento versus Complexidade de programas;
- ✓ Carga/atualização pode requerer processamento paralelo, para otimização.

- **Utilização de Agregados**

- ✓ Navegador de agregados: camada de interface entre a ferramenta OLAP e o servidor de DW. O navegador realiza transparentemente a conversão de comandos SQL granulares nos equivalentes que trabalham informações agregadas.



Ferramenta OLAP

Noções Iniciais

- Ferramentas OLAP fornecem suporte para funções de análise de dados, típicas de aplicações avançadas de planilhas eletrônicas;
- Nessa ferramenta que o usuário realiza as consultas e pode exportar os resultados para programas como Excel e formatos como .pdf;
- A consulta OLTP assemelha-se a uma câmera fotográfica, onde o usuário pode tirar fotografias do estado do sistema;
- A consulta OLAP assemelha-se a uma máquina de filmar, em que o usuário pode se deslocar livremente no meio dos dados com recursos poderosos de corte e zoom.

Ferramenta OLAP

Operações OLAP

- Operações dimensionais de ferramentas OLAP:
 - ✓ Slice and Dice (Ponto, Plano, Cubo);
 - ✓ Rotation (Rotação ou Pivotamento);
 - ✓ Drilling:
 - ▶ Drill Down
 - ▶ Drill Across e Drill Through
 - ▶ Drill Up (Roll Up)
 - ✓ Ranking (Classificação por uma coluna).



Business School Brasil
O futuro é agora. Junte-se aí.



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Ferramenta OLAP

Drill

- Com o drill (zoom) é possível aumentar ou diminuir o detalhamento de um relatório.
 - Ex: País → Estado → Município



Ferramenta OLAP

Drilling - Operações

Operação	Descrição	Exemplo
Drill Down	Aumento do nível de detalhe da informação e consequente diminuição do nível de granularidade.	Uma análise de vendas por estado é alterada para uma análise de vendas das cidades de um determinado estado.
Drill Up	Diminuição no nível de detalhe e consequente aumento do nível de granularidade.	Uma análise de vendas é alterada de uma cidade para seu estado correspondente
Drill Across	O nível de análise dentro de uma mesma dimensão é alterado, ou seja, o usuário avança um nível intermediário dentro de uma mesma dimensão	O nível da análise é alterado direto de ano para mês dentro da dimensão tempo, quando esta é composta por ano, semestre e mês.

Ferramenta OLAP

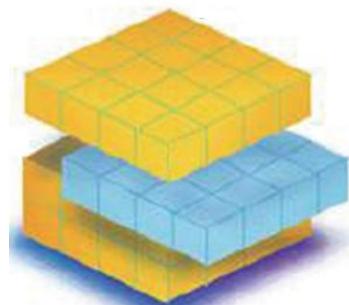
Drilling - Operações

Operação	Descrição	Exemplo
Drill Through	Ocorre quando o usuário passa de uma informação contida em uma dimensão para uma outra.	O usuário está realizando uma análise na dimensão tempo e no próximo passo analisa a informação por região.

Ferramenta OLAP

Slice - Operações

- O slice (filtro) permite que o usuário restrinja os resultados exibidos no relatório, utilizando um dos objetos existentes.



Operação	Descrição	Exemplo
Slice	Corta o cubo, mas mantém a mesma perspectiva de visualização dos dados. Funciona como um filtro que restringe uma dimensão à apenas um ou alguns de seus valores.	Em uma dimensão tempo de um modelo é selecionado somente o ano de 2000.

Ferramenta OLAP

Dice - Operações

- O dice (pivotagem) permite ao usuário trocar as informações de posição (de colunas para linha, por exemplo), auxiliando-o à analisar as informações sob diferentes perspectivas.



Operação	Descrição	Exemplo
Dice	Mudança de perspectiva da visão multidimensional, como se o cubo fosse girado. Permite descobrir comportamentos e tendências entre os valores das medidas analisadas em diversas perspectivas.	A análise é alterada de região (linha) por ano (coluna) para ano (linha) por região (coluna).

Ferramenta OLAP

Sort e Ranking - Operações

- Permite o ordenamento do resultados exibidos baseado em fatos (Rank) ou dimensões (Sort).

Operação	Descrição	Exemplo
Sort	Tem a função de ordenar a informação, podendo ser aplicada a qualquer tipo de informação, não somente a valores numéricos.	Ordenar as instituições em ordem alfabética.
Ranking	Permite agrupar resultados por ordem de tamanho, baseado em valores numéricos, refletindo somente na apresentação do resultado e não no resultado em si.	Ordenar a relação de filiais de acordo com os maiores volumes de vendas.

Ferramenta OLAP

Tipos de ferramentas OLAP

- **OLAP Multidimensional (MOLAP)**
 - ✓ Utilizam estrutura de dados multidimensional e permitem a navegação pelos níveis de detalhamento em tempo real;
 - ✓ O BD e o SGBD são multidimensionais;
 - ✓ Em produtos MOLAP, os cubos são alimentados por meio da cópia de um instantâneo dos dados residentes na fonte dos dados;
 - ✓ Maior velocidade. Maior espaço necessário. Menor escalabilidade.



Ferramenta OLAP

Tipos de ferramentas OLAP

- **OLAP Relacional (ROLAP)**

- ✓ Decorrência do uso consagrado de SGBDs relacionais nos BDs operacionais (transacionais), com as vantagens da tecnologia aberta e padronizada (SQL);
- ✓ Utiliza os metadados no apoio à descrição do modelo de dados e na construção de consultas. Através de uma camada semântica acima do esquema relacional, os dados são apresentados ao usuário com visão multidimensional.
- ✓ Trabalham diretamente com a fonte de dados, sem a cópias dos mesmos;
- ✓ Maior escalabilidade. Menor velocidade quando há uma grande volume de dados.

Ferramenta OLAP

Tipos de ferramentas OLAP

- **OLAP Híbrido (HOLAP)**

- ✓ Tendência dos modernos SGBDs relacionais de adicionar uma arquitetura multidimensional para prover facilidades a ambientes de suporte a decisão;
- ✓ Proporciona o desempenho e flexibilidade de um BD multidimensional e mantém a gerenciabilidade, escalabilidade, confiabilidade e acessibilidade conquistadas pelos BDs relacionais.



Ferramenta OLAP

12 regras de Codd para OLAP

- 1. Visão Conceitual Multidimensional:** Os analistas verão a empresa como multidimensional por natureza. Os dados são modelados em diversas dimensões podendo haver cruzamento de todos os tipos de informações;
- 2. Transparência:** O usuário não deve ser exposto as fontes de dados que alimentam a ferramenta OLAP, podendo a mesmas serem homogêneas ou heterogêneas. OLAP deve atender a todas as solicitações do analista, não importando de onde os dados virão. Todas as implicações devem ser transparentes para os usuários finais;
- 3. Acessibilidade:** As ferramentas OLAP devem permitir conexão com todas as bases de dados legadas, bem como realizar as conversões necessárias de forma a apresentar ao usuário uma visão coerente.



Ferramenta OLAP

12 regras de Codd para OLAP

4. **Relatórios de desempenho consistentes:** O desempenho da ferramenta OLAP não deve sofrer significativamente quando há o aumento do número de dimensões. Ela deve possuir conhecimento sobre todas as informações armazenadas que possa disponibilizar ao usuário;
5. **Arquitetura Cliente/Servidor:** OLAP deve ser construída em arquitetura C/S para que possa atender a qualquer usuário em qualquer ambiente operacional. O servidor deve ser suficientemente inteligente para conectar vários clientes com o mínimo de esforço, além de mapear e consolidar informações de bancos de dados distintos;
6. **Dimensionalidade Genérica:** Todos os dimensionais devem ser equivalentes em sua estrutura e capacidades operacionais. Deve ser capaz de tratar informações em qualquer quantidade de dimensões.

Ferramenta OLAP

12 regras de Codd para OLAP

7. **Manipulação de dados dinâmicos esparsos:** Devido ao grande volume de informações armazenadas nas diversas dimensões de um modelo multidimensional, é comum a esparsidade dos dados, e então essas células nulas devem ser tratadas para evitar custos com memória.
8. **Suporte multiusuário:** Nas grandes organizações, é comum vários analistas trabalharem com a mesma massa de dados, devendo a ferramenta OLAP fornecer recuperação concorrente, acessos de atualização, integridade e segurança.
9. **Operações de cruzamento dimensional sem restrição:** As ferramentas OLAP devem ser capazes de navegar nas diversas dimensões existentes, permitindo a manipulação e cálculos de dados em qualquer número de dimensões.



Ferramenta OLAP

12 regras de Codd para OLAP

10. **Manipulação intuitiva de dados:** O usuários devem ser capazes de manipular os dados livremente, devendo realizar operações de *drilling down* ou *drilling up*, por exemplo, sem necessitar de qualquer tipo de ajuda ou de procedimentos complexos ou que requeiram um grande número de etapas na interface;
11. **Relatórios flexíveis:** O usuário deve ter a flexibilidade para efetuar qualquer tipo de consulta, devendo a ferramenta disponibilizar facilidades na construção de relatórios.
12. **Ilimitadas dimensões e níveis de agregação:** Devido às várias dimensões existentes, deve haver vários níveis de agregação dos dados. Cada dimensão genérica deve permitir um número ilimitado de níveis de agregação em qualquer trajetória de consolidação;



Ferramenta OLAP

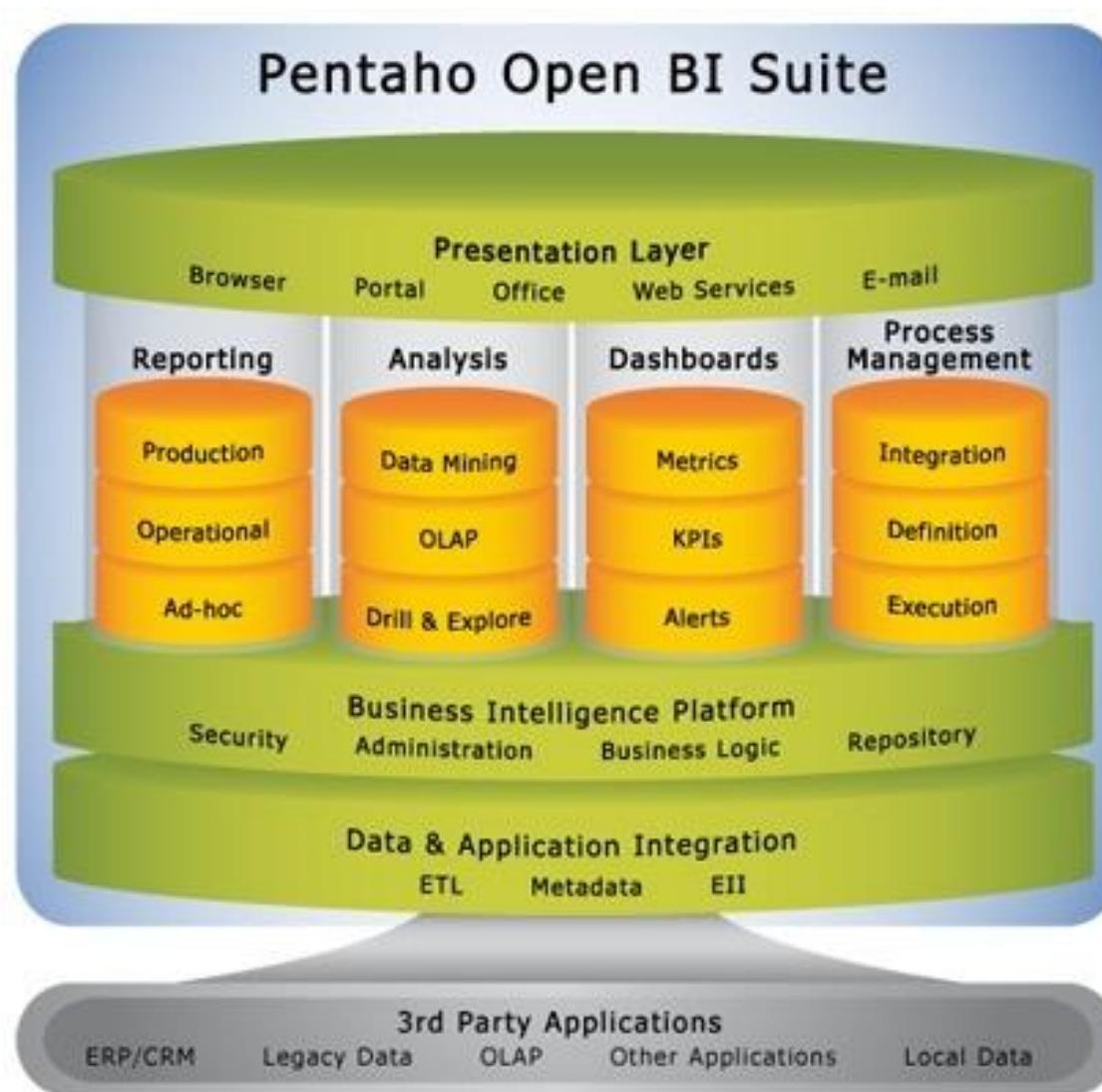
- Cognos
 - Execplan
 - Ascential
 - Microstrategy
 - SAS Institute
 - IBM Brasil
 - Business Object
 - Hyperion
 - Extend Software
 - Microsoft
 - Platinum
 - Hummingbird
-
- Sistemas oriundos de ERP ou outras soluções:
 - SAP
 - PeopleSoft
 - Oracle



Pentaho

- É uma plataforma de BI centrada em processos e orientada a soluções;
- É composta por componentes de BI que permitem as organizações desenvolverem uma solução completa e integrada para problemas de BI;
- O elemento central na arquitetura é a Pentaho Open BI Suíte;
- É centrada em processos cujo controle é realizado através de um mecanismo de workflow.

Ferramenta OLAP



Business School Brasil
O futuro começa aqui!



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Pentaho

- A Pentaho Open BI Suite consiste de uma plataforma de BI e o Pentaho Design Studio;
- O Pentaho Design Studio é um conjunto de ferramentas para o projeto e gestão que são integradas ao ambiente Eclipse;
- Estas ferramentas permitem ao analista de negócios tanto desenvolver como criar relatórios, *dashboards*, modelos de análise, regras de negócios e processos de BI.

Ferramenta OLAP

Pentaho

- A plataforma é construída em termos de processos e definições de processos;
- Os processos são definidos em uma linguagem padronizada de definição de processos;
- É externamente acessada, editada e customizada;
- Toda a lógica dos negócios é revelada, e nada é escondido.



Ferramenta OLAP

Pentaho

- A plataforma é construída em termos de processos e definições de processos;
- Os processos são definidos em uma linguagem padronizada de definição de processos;
- É externamente acessada, editada e customizada;
- Toda a lógica dos negócios é revelada, e nada é escondido.



Ferramenta OLAP

Pentaho



Business School Brasil
O futuro começa aqui!

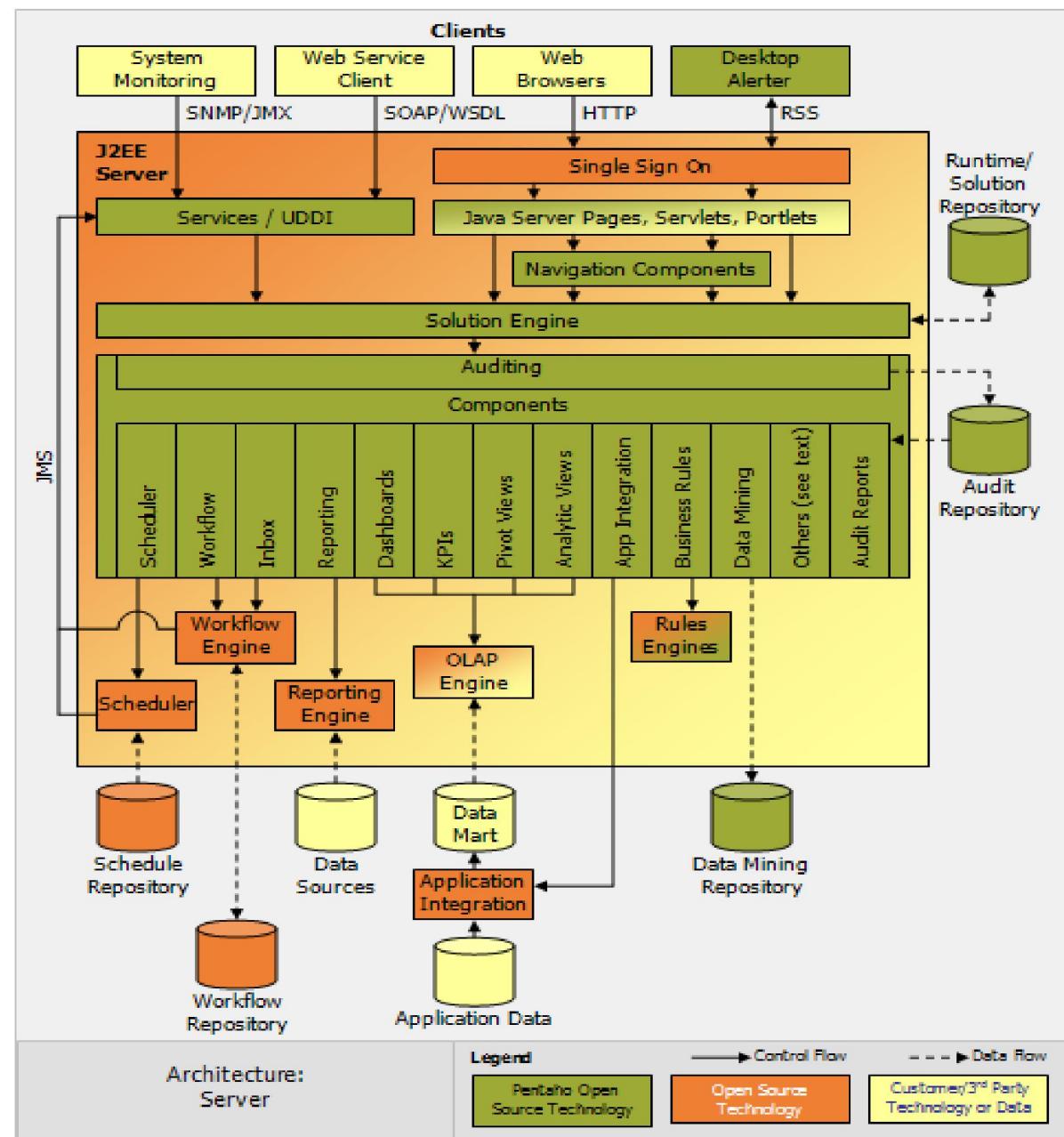


PUC
GOIÁS



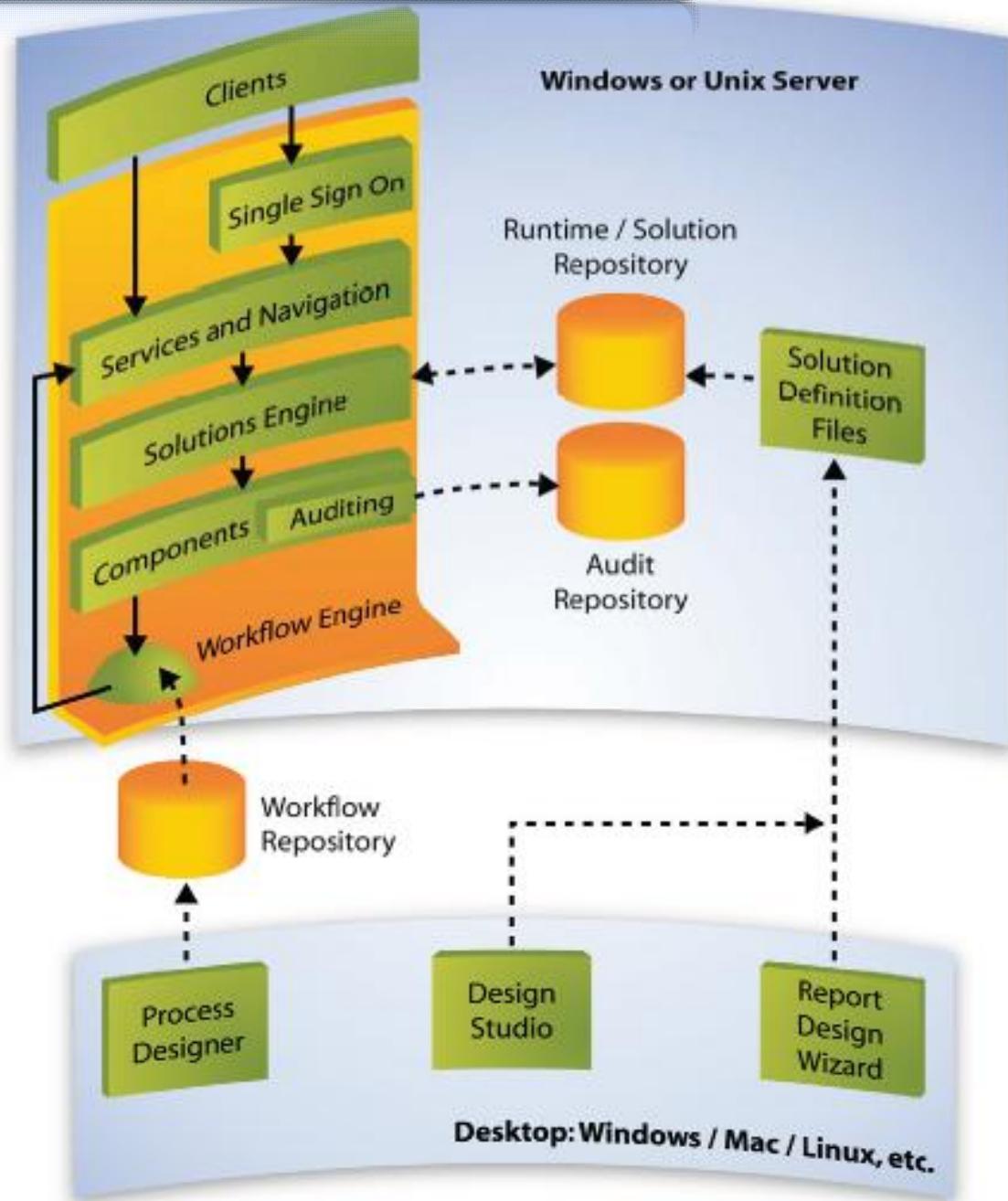
UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Architecture: Server



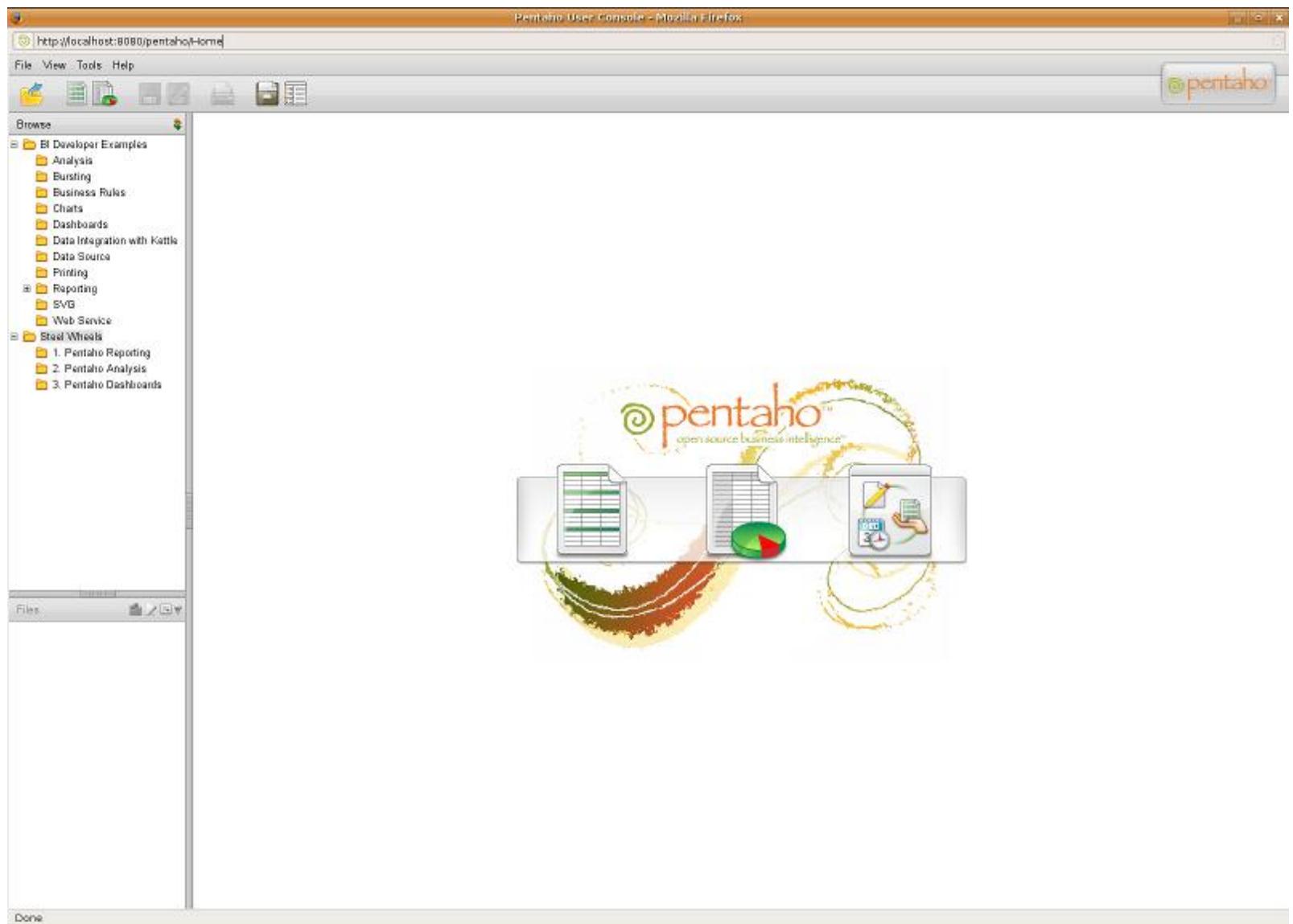
Ferramenta OLAP

Pentaho



Ferramenta OLAP

Pentaho



Ferramenta OLAP

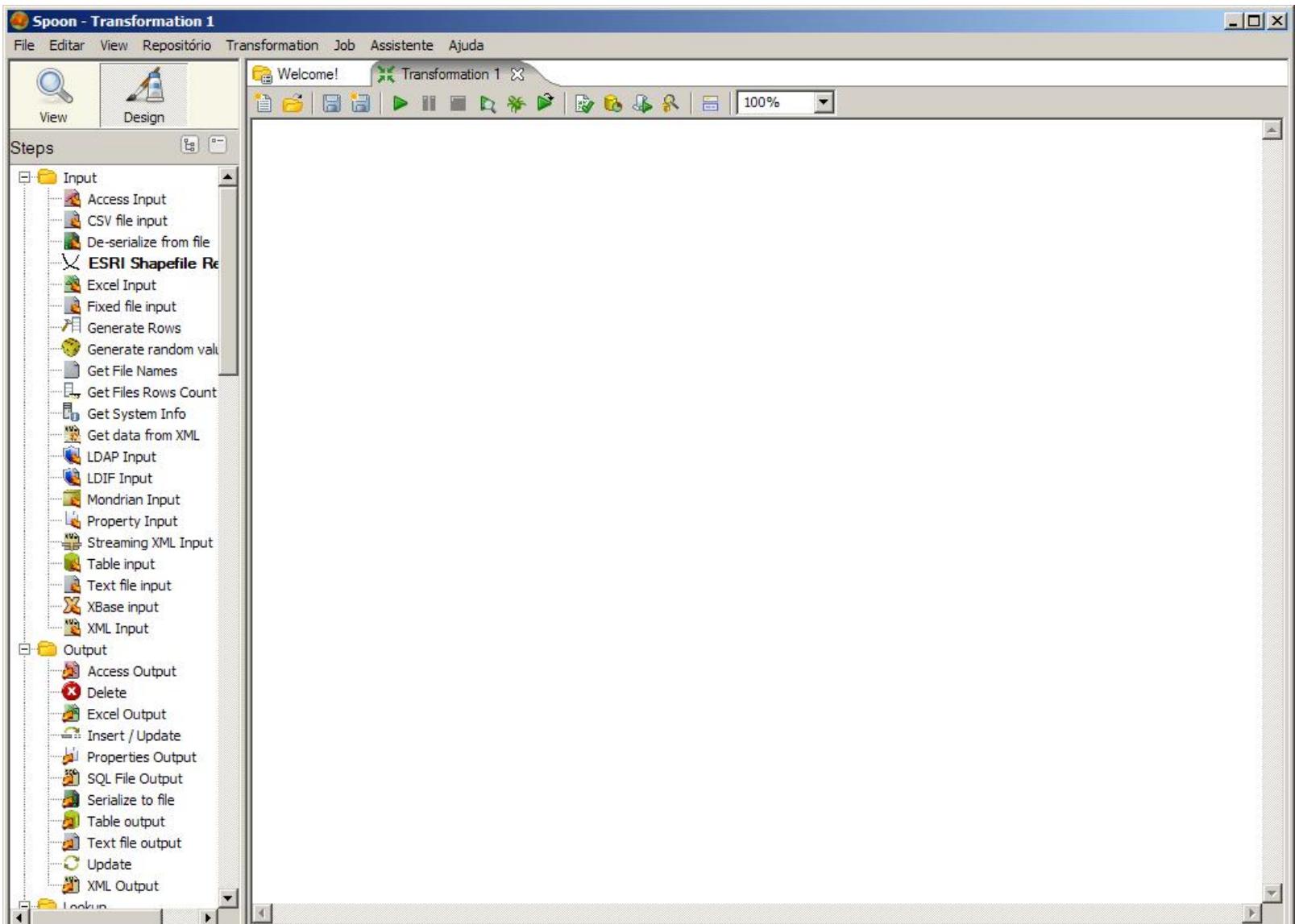
Pentaho

- Kettle ETL e ETL (PDI)
- JFree Report (*Report Designer*)
- Schema Workbench (cubos OLAP)
- Pentaho Design Studio (Edição de *action sequences*)
- MetaData Editor (Metamodelos)



Ferramenta OLAP

Pentaho - PDI



Business School Brasil
O futuro é agora, juntos!



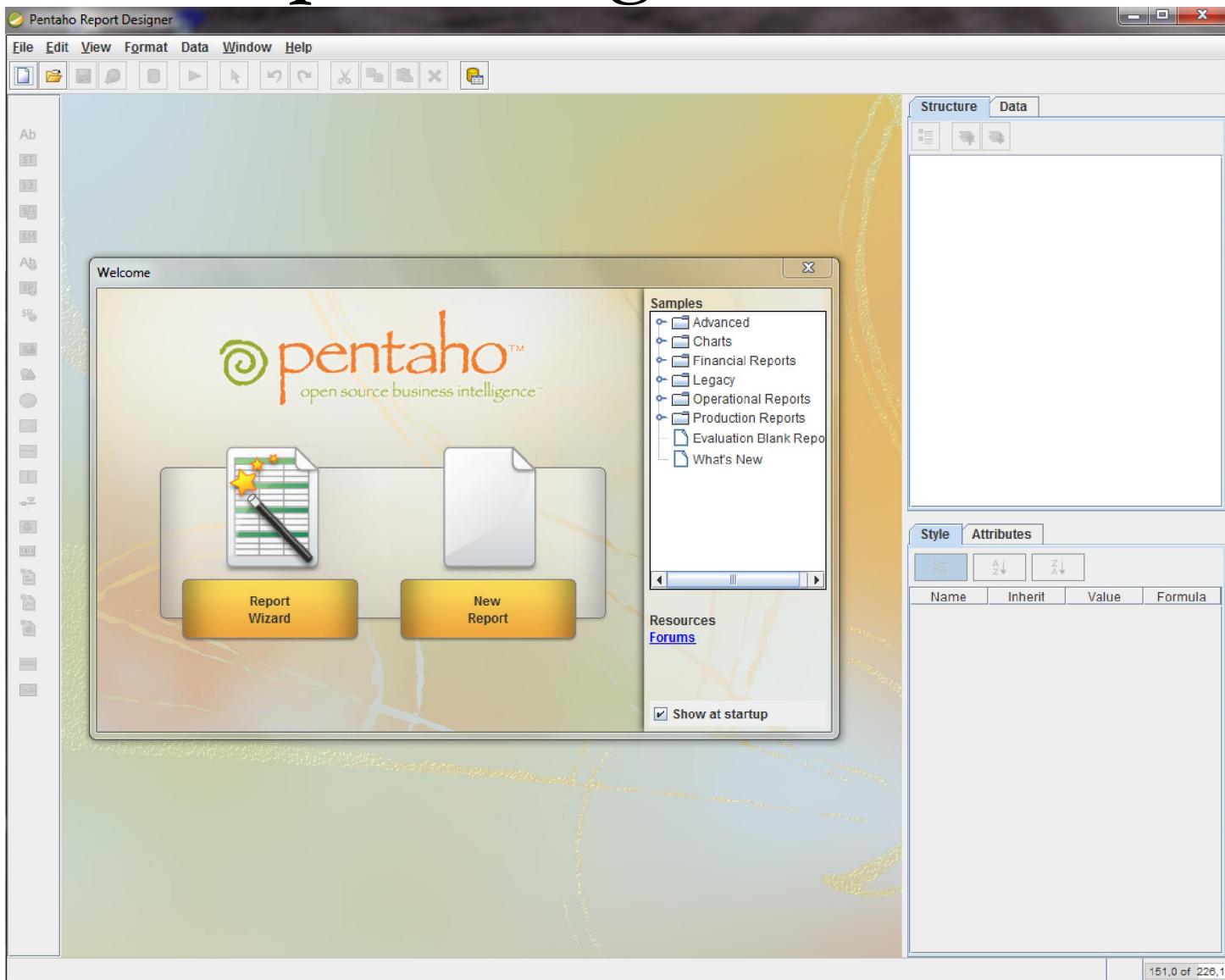
PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Ferramenta OLAP

Pentaho – Report Designer



Business School Brasil
O futuro é agora, juntos!



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Ferramenta OLAP

Pentaho - Workbeach

The screenshot shows the Pentaho Schema Workbench application window. The title bar reads "Schema Workbench". The menu bar includes "File", "Edit", "View", "Tools", "Windows", and "Help". The toolbar contains various icons for file operations like Open, Save, and Delete, along with other tools. The main area has a tab titled "Schema - SugarCRM Sales Performance (SugarCRM-Sales-Performance.xml)". On the left, there's a tree view of the schema structure under "SalesAnalysis", including tables like "sales_fact" and dimensions like "Account Categorization" and "Start Period". On the right, a table titled "Shared Dimension" is displayed with the following data:

Attribute	Value
name	End Period
foreignKey	
type	TimeDimension
usagePrefix	
caption	

A message at the bottom states "Hierarchy 'New Hierarchy 0' is invalid." The database connection at the bottom is set to "Database - sugarcrm (MySQL)".



Business School Brasil
O futuro começa aqui!



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Ferramenta OLAP

Pentaho – Designer Studio

The screenshot shows the Eclipse IDE interface for the Pentaho Designer Studio. The title bar indicates the project is 'Java - BurstActionSequence.xaction - Eclipse SDK'. The menu bar includes File, Edit, Navigate, Search, Project, Run, BI Platform, Window, and Help. The toolbar contains various icons for file operations like Open, Save, and Print.

The central workspace displays the 'BurstActionSequence.xaction' editor. On the left is the Package Explorer showing files under the 'bursting' folder, including 'burst.png', 'BurstActionSequence.xaction', 'BurstingExamples.xpj', 'BurstShark_cn.prop', 'BurstShark_de.prop', 'BurstShark_es.prop', 'BurstShark_fr.prop', 'BurstShark_it.prop', 'BurstShark_nl.prop', 'BurstShark_pt.prop', 'BurstShark_zh_CN.prop', 'BurstShark.png', 'BurstShark.properties', 'BurstShark.xaction', 'file.png', 'folder.png', 'index_cn.properties', 'index_de.properties', 'index_es.properties', and 'index_fr.properties'.

The main editor area is divided into sections:

- Process Inputs:** Contains 'inputs' and 'resources' folders.
- Process Actions:** Contains a 'Query For Manager Information' step, which is an 'Action Loop On' step. It includes sub-steps: 'Format The E', 'Format The E', 'Format The E', 'Print Paramet', 'Query For Re', and 'Create The R'. There are also 'Run Query Immediately' and 'Run Query From Pentaho Subreport' options.
- Query:** A SQL query:

```
select distinct REGION, MANAGER_NAME, EMAIL from DEPARTMENT_MANAGERS
```
- Result Set Type:** Set to 'In-Memory Result Set'.
- Result Set Name:** Set to 'burst-data'.
- Result Set Columns:** A table with three columns:

Name	Type
REGION	string
EMAIL	string
MANAGER_NAME	string
<<Add>>	

The right side of the interface shows an 'Outline' view which displays the message 'An outline is not available.'



Business School Brasil
O futuro é agora, construído



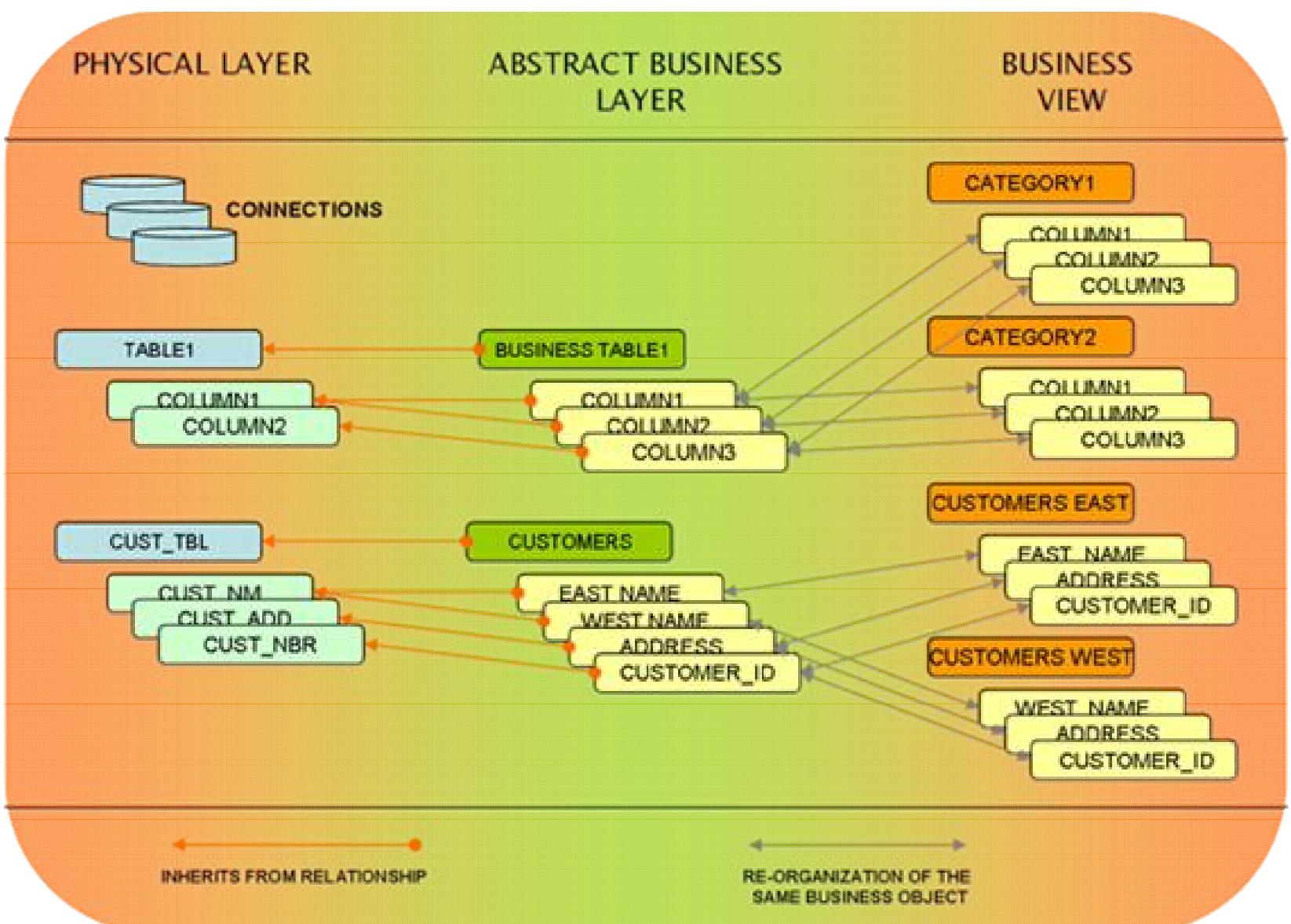
PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

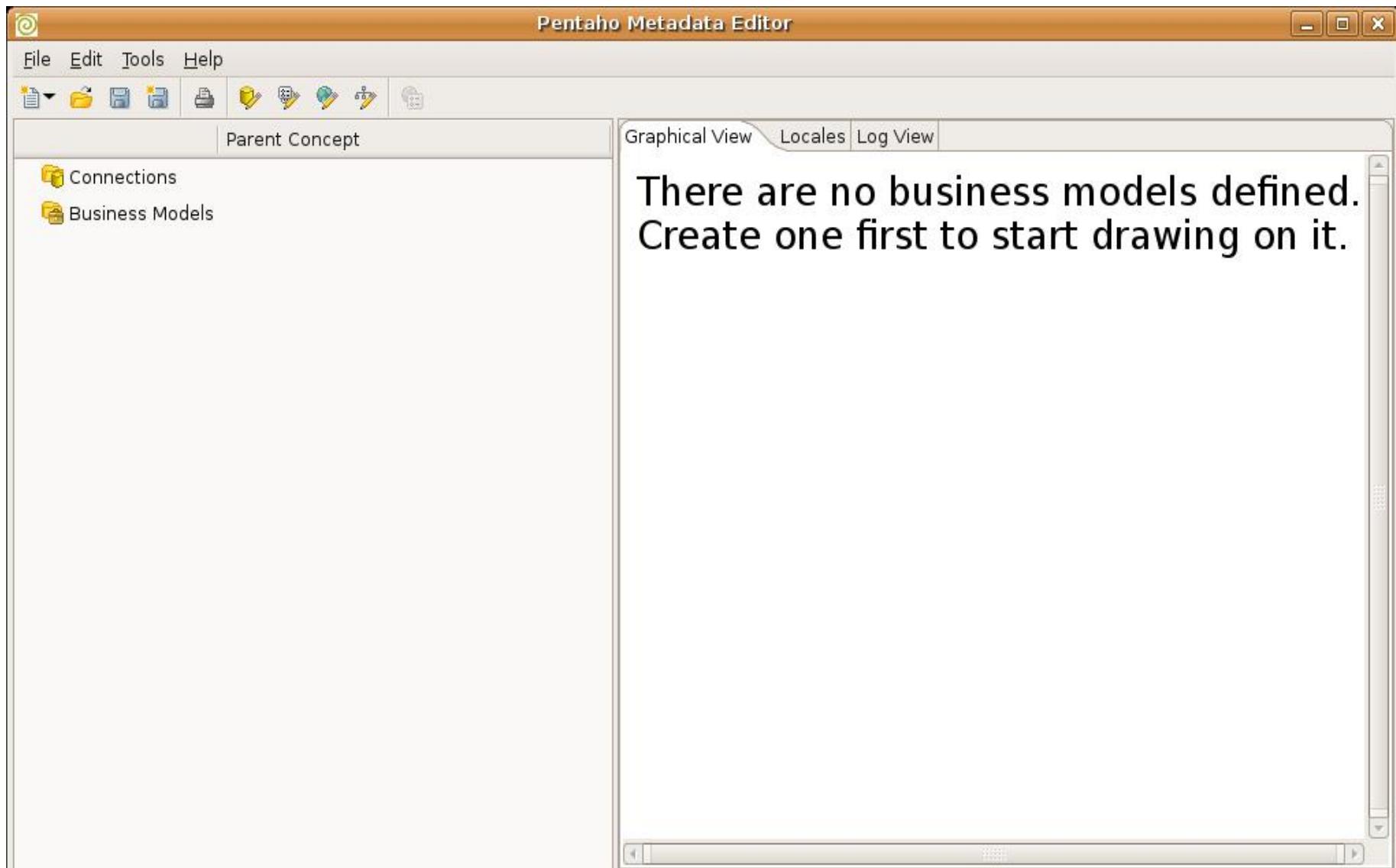
Ferramenta OLAP

Pentaho – Metadata Editor



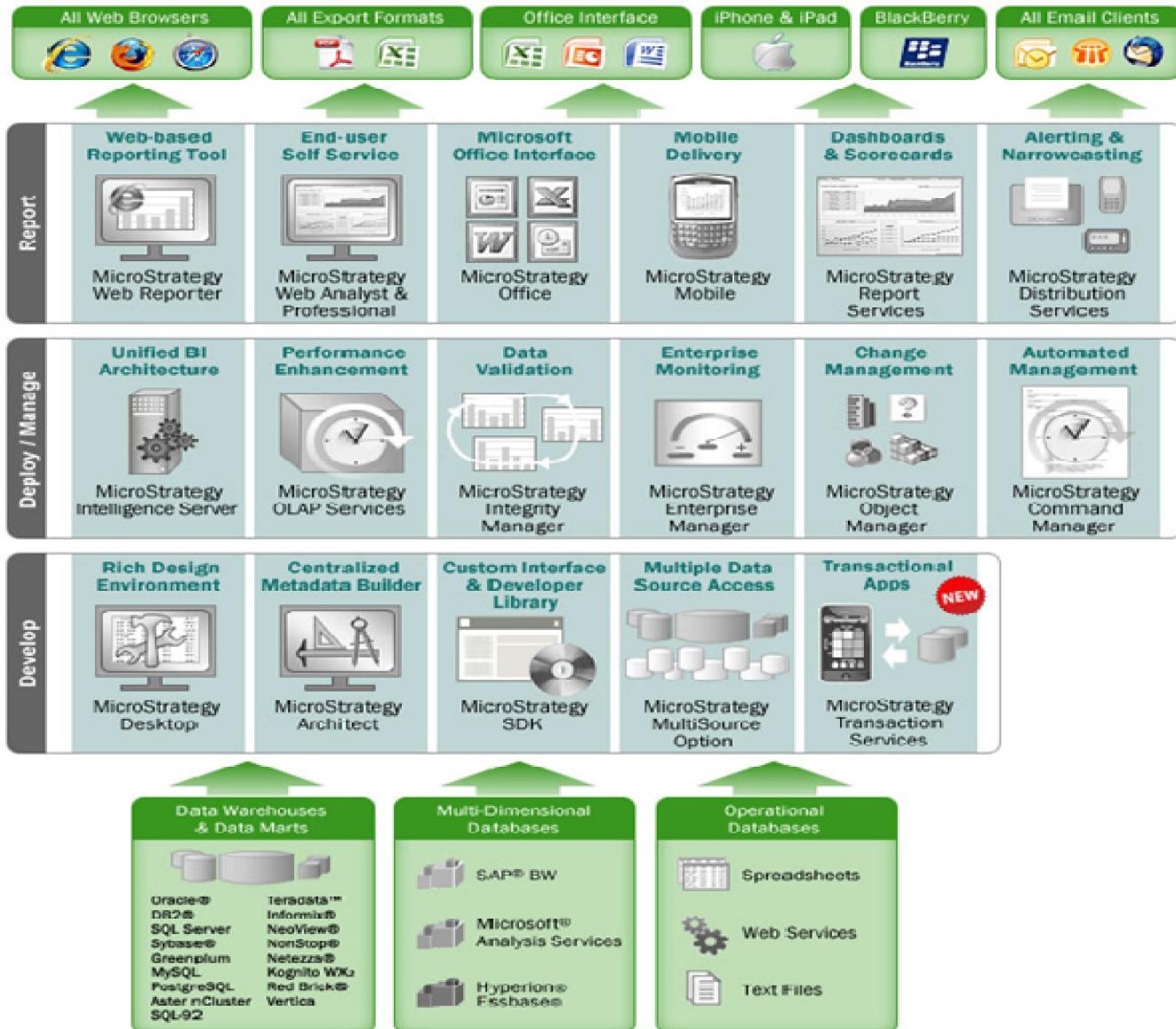
Ferramenta OLAP

Pentaho – Metadata Editor



Ferramenta OLAP

MicroStrategy



Conceitos de Modelagem Multidimensional

- Proposto por Ralph Kimball para projeto de DW
 - ✓ Dimensional Modeling Manifesto, 1997;
 - ✓ O próprio Kimball atribui a origem a um projeto conduzido por uma empresa (General Mills) e uma universidade (Dartmouth) nos anos 1960.
- Dominante no projeto de DW
 - ✓ Para Kimball, em todo o DW;
 - ✓ Para Inmon, nos data marts.



Conceitos de Modelagem Multidimensional

- **Características:**

- ✓ Distingue melhor as dimensões dos fatos medidos;
- ✓ Simplifica a visualização dimensional (essencial em consultas OLAP);
- ✓ Na verdade é uma mistura de modelagem conceitual com modelagem lógica, pois já é bastante voltada para a abordagem relacional (a literatura fala sempre em tabelas).

- **Fatos e dimensões podem ser dispostos segundo diferentes configurações**

- Esquema Estrela;
- Esquema Flocos de Neve;
- Esquema Constelação de Fatos.



Conceitos de Modelagem Multidimensional

Fatos:

- **Medidas numéricas do negócio**
 - ✓ Volume de vendas (número de itens, total em reais), quantidade de itens em estoque, volume de transações de cartão de crédito;
- **Representados em uma Tabela de Fatos**
 - ✓ Valores das medidas (numéricas e aditivas) e Referências para as dimensões (granularidade);
 - ✓ Não armazena informação redundante! (textos, valores zerados);
 - ✓ Quantidade de registros é normalmente muito grande em um DW.

Fato Vendas

pk_tempo
pk_produto
pk_loja
preco_venda
unidades_venda
preco_custo



Conceitos de Modelagem Multidimensional

Dimensões:

- Pontos de vista ou perspectivas do negócio sobre os quais uma organização deseja guardar registros
 - ✓ Loja, Produto, Fornecedor, Tempo.

Dimensão Produto

pk_produto
descricao
categoria
marca

- Representadas em Tabelas de Dimensão

- ✓ Descrição completa da dimensão
 - ▶ Atributos textuais e de domínio discreto preferencialmente.
- ✓ Essenciais para tornar o DW usável e legível
 - ▶ “Tipo de armazenamento = V” ou “Tipo de armazenamento = Vácuo” ?

Dimensão Loja

pk_loja
nome_loja
endereco
cidade
estado

Agregados e hierarquias de dimensões

Dimensões:

Dimensão
Tempo

pk_tempo
data
mes
ano
...

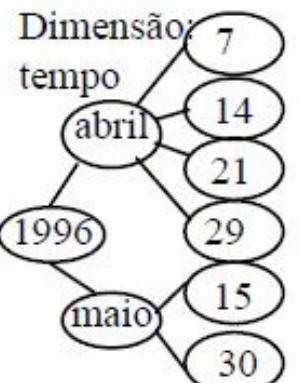


Ano

Mês

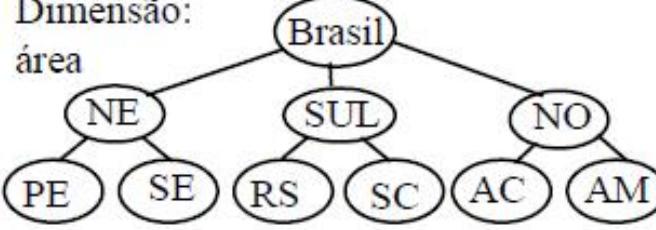
Dia

Produto
Alfa1



Total de vendas

Dimensão:
área



34	23	45	62	56	150
23	92	73	23	234	
13	87	21	14		
		34			
	46		
		18			

País

Região

Estado

Dimensão Area

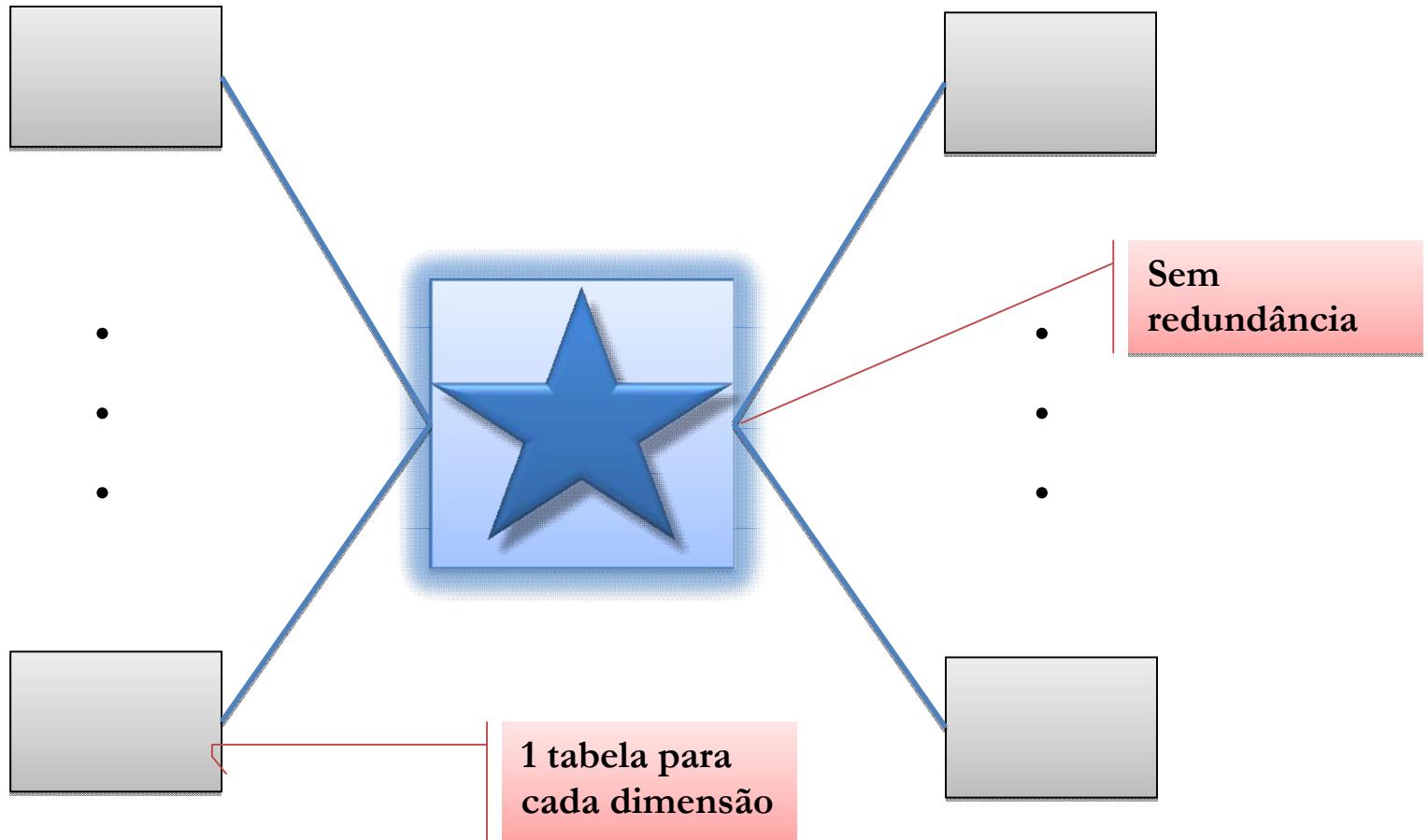
pk_area
estado
regiao
pais
...

bsis

Conceitos de Modelagem Multidimensional

Esquema Estrela

Uma tabela de fato cercada de tabelas de dimensões



Conceitos de Modelagem Multidimensional

Esquema Estrela:

Dominante em projetos de Data Warehouse (KIMBALL e ROSS, 2002).

- **Características:**

- ✓ Distingue bem as dimensões dos fatos medidos;
- ✓ Simplifica a visualização dimensional;
- ✓ Simetria;
- ✓ Eficiente para a realização de consultas;
- ✓ Acomodam mudanças mais facilmente.

- Hierarquias são representadas pelos atributos da dimensão;



Conceitos de Modelagem Multidimensional

Esquema Estrela: Exemplo

Dimensão Produto	
pk_produto	
descricao	
categoria	
marca	

Fato Vendas	
pk_tempo	
pk_produto	
pk_loja	
preco_venda	
unidades_venda	
preco_custo	

Dimensão Tempo	
pk_tempo	
data	
mes	
quadrimestre	
ano	
Flag_feriado	

Dimensão Loja	
pk_loja	
nome_loja	
endereco	
cidade	
estado	

Conceitos de Modelagem Multidimensional

Esquema Estrela: Exemplos de Consultas

- ✓ "Vendas por categoria de produto sobre os últimos seis meses".
- ✓ "Vendas por marca entre 1990 e 1995".

Dimensão
Tempo

pk_tempo
data

Dimensão Produto

Dimensão Loja

pk_loja
nome_loja
endereço
cidade
estado

Colunas da chave composta
ligando a tabela de fatos às
tabelas de dimensão

Medidas Numéricas

pk_tempo	pk_produto	pk_loja	preco_venda	Unidades_venda	Preco_cust o



Conceitos de Modelagem Multidimensional

- **Esquema Estrela é simétrico**
 - ✓ – Comparado a esquemas ER típicos.
- **Tabela de Fatos**
 - ✓ Expressa relacionamento M:N entre dimensões;
 - ✓ Tabela dominante:
 - ▶ Usualmente com grande volume de dados; ocupam 90% do espaço em um DW típico;
 - ▶ Tendem a ter muitas linhas e poucas colunas.



Business School Brasil
O futuro começa aqui



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Conceitos de Modelagem Multidimensional

- **Tabelas de Dimensões**

- ✓ Tabelas que “qualificam” os fatos, com muitos campos descritivos (é comum ter dimensões com dezenas de colunas);
- ✓ Dimensões apresentam-se em consultas qualificadas como “por dimensão” (vendas “por semana” “por marca” “por loja”) e são as bases para agregações e agrupamentos;
- ✓ Uma junção liga cada tabela de dimensão à tabela de fatos;
- ✓ Volume bem menor que as tabelas de fatos;
- ✓ O poder de um DW é diretamente proporcional à qualidade e profundidade dos atributos das dimensões.



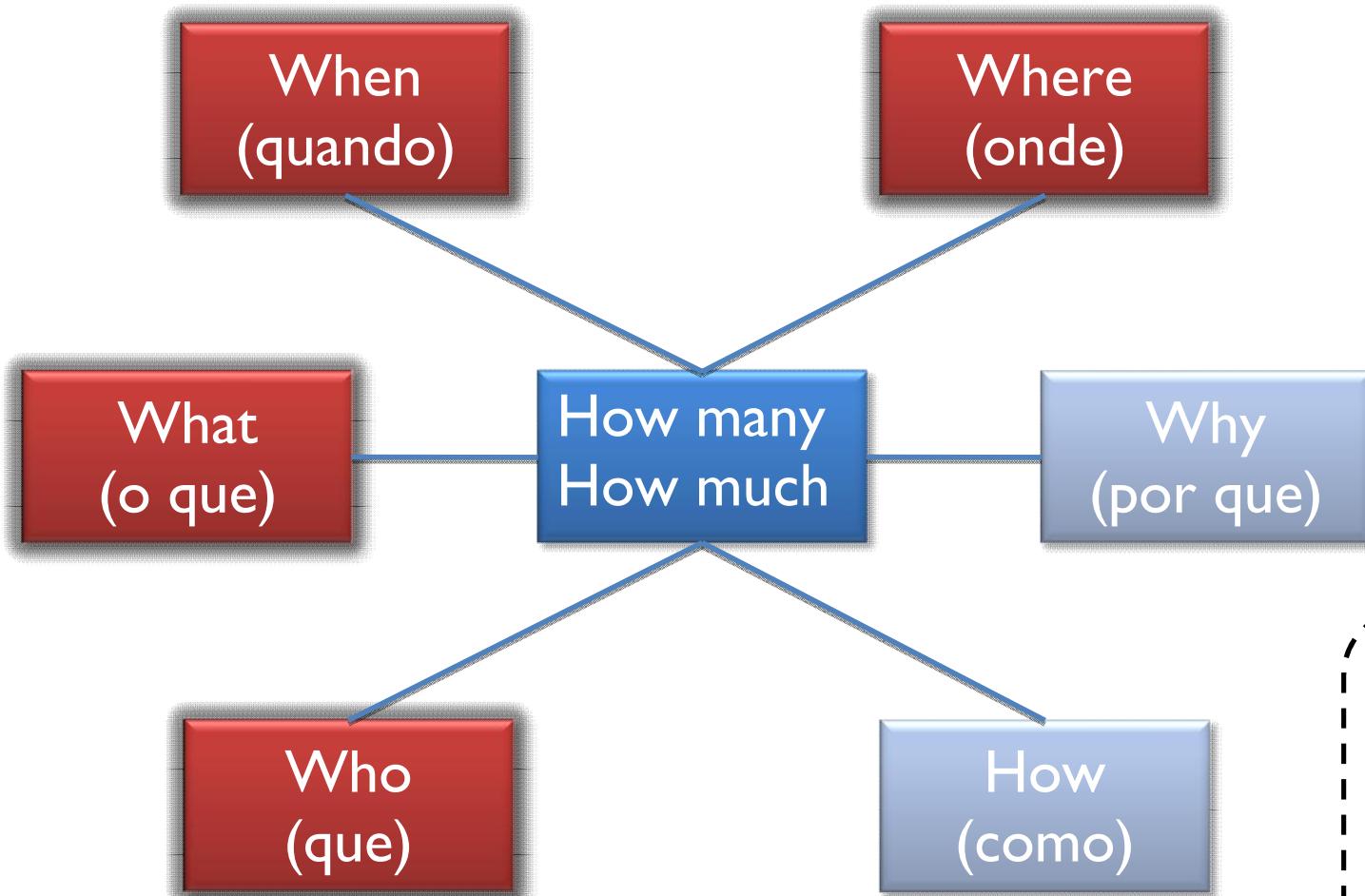
Conceitos de Modelagem Multidimensional

Modelagem ER x Multidimensional

ER	Multidimensional
1 diagrama (vários processos de negócio)	Vários diagramas dimensionais (1 para cada processo de negócio)
Usuários acham difícil entender e navegar pelo modelo	Usuários reconhecem “o seu negócio”
Muitas junções para responder a consultas	Poucas junções
Dados atômicos	Dados atômicos e agregados
Planos de consultas extremamente distintos e específicos para as consultas previstas	Planos de consultas “genéricos” (simetria do modelo)

Conceitos de Modelagem Multidimensional

Esquema Estrela: 5W3H



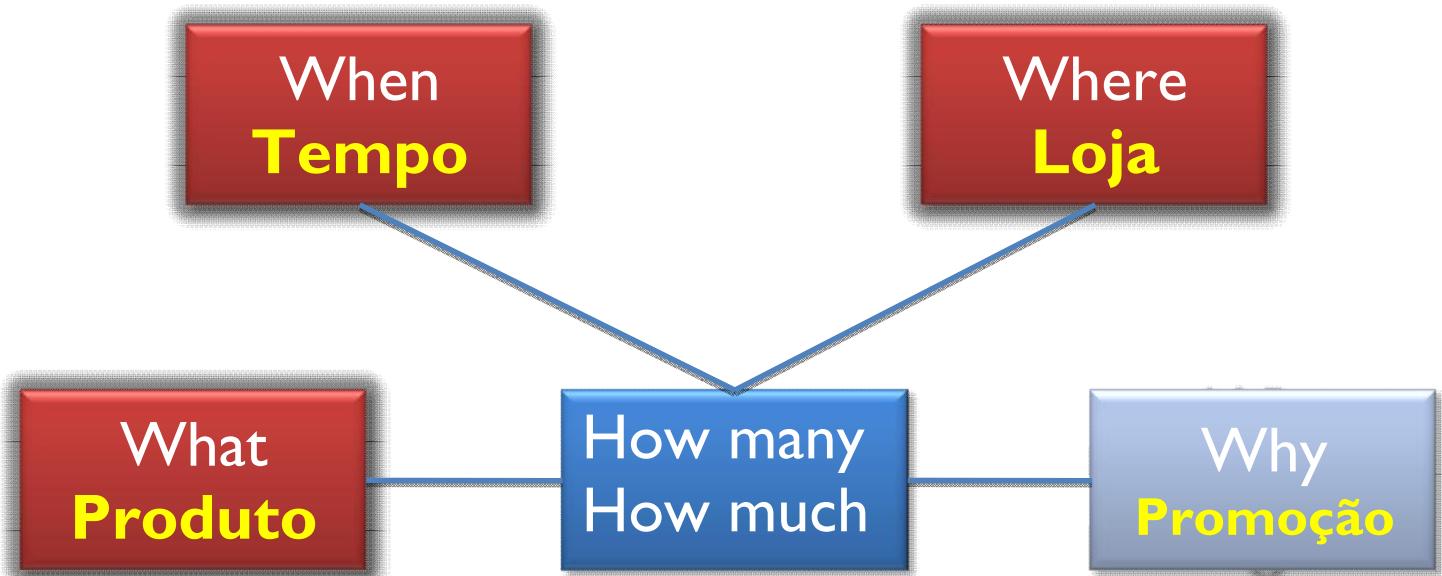
Tipos de dimensão mais comuns



Conceitos de Modelagem Multidimensional

Esquema Estrela: 5W3H - Exemplo

Exemplo de uma empresa com várias loja de vendas de produtos, que podem estar em promoção.



Conceitos de Modelagem Multidimensional

Dimensão Tempo (When)

- A dimensão Tempo (Data) é muito poderosa e importante em todo DW. Como tal deve ser tratada de forma diferenciada em relação às outras dimensões. Usualmente está presente em todo Data Mart, pois o DW é histórico.
- Costuma ser complexa no mundo real:
 - ❖ Dia, Mês, Trimestre, Semestre, Ano;
 - ❖ Dia Acumulado no Mês, no Ano;
 - ❖ Período Fiscal, Semana de Cinco Dias;
 - ❖ Feriados, Fim de semana.



Conceitos de Modelagem Multidimensional

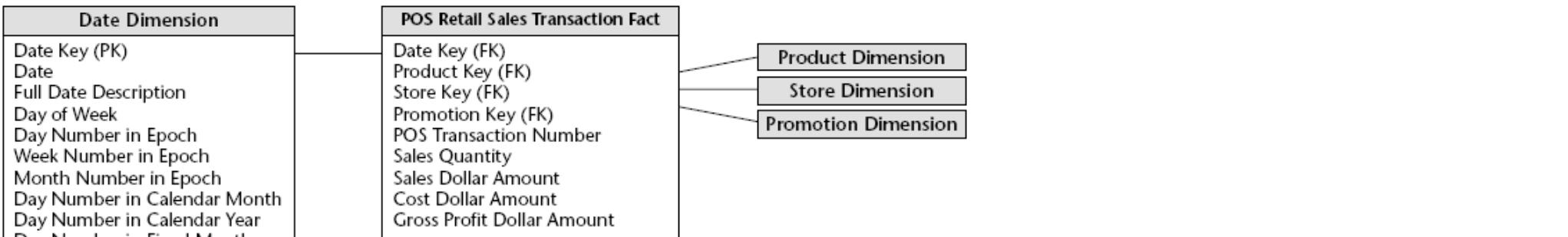
Dimensão Tempo (When)

Qual a granularidade ideal?

- É claro, depende do projeto
 - ❖ Com granularidade diária, podemos organizar os dados por dias, meses, anos, por períodos fiscais (artificiais) da empresa, etc. Essa modelagem é mais flexível a mudanças nos requisitos do negócio.
- Diferente das outras dimensões, a tabela Data pode ser carregada antecipadamente, de uma só vez e não requer fonte de dados
 - ❖ – Exemplo: 5 anos passados + 5 anos futuros = 10 anos = 3.650 dias (linhas na tabela)

Conceitos de Modelagem Multidimensional

Dimensão Tempo (When)



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

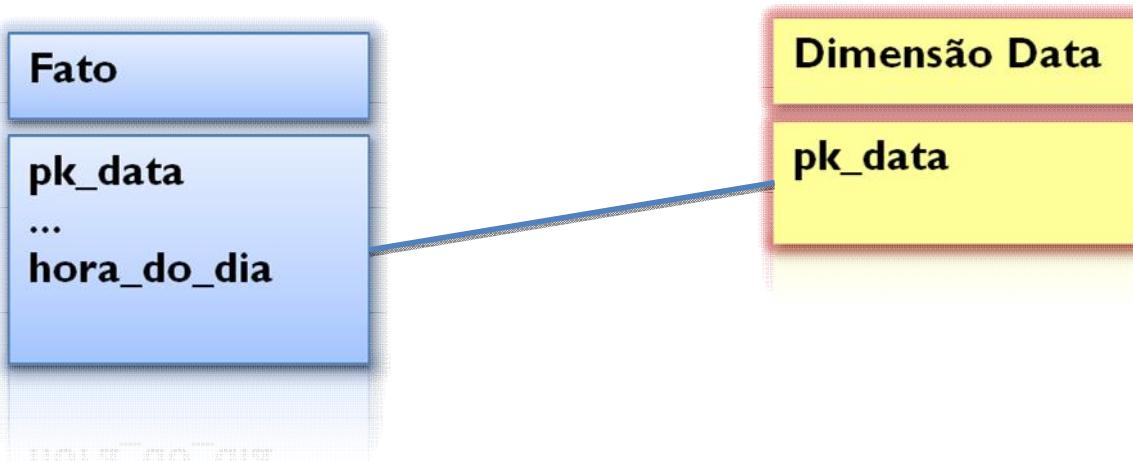
Tipo de dados SQL (Date, Time) não suportam essa riqueza de descrições, daí a necessidade de uma dimensão Data explícita.

Conceitos de Modelagem Multidimensional

Dimensão Tempo: Horas, minutos, segundos

- Várias soluções são possíveis, graças à extensibilidade do modelo dimensional.

1^a Alternativa: Colocar a "hora do dia" na Tabela de Fatos.



- ❖ Pode ser usado quando não há descrições adicionais sobre a hora do dia.
- ❖ Pode sobrecarregar a tabela de fatos (tipo Timestamp requer 8 bytes) 8 bytes x bilhões de linhas na tabela de fatos ...

Conceitos de Modelagem Multidimensional

Dimensão Tempo: Horas, minutos, segundos

2^a Alternativa: Criar uma dimensão Hora do dia (24h x 60min = 1440 valores.

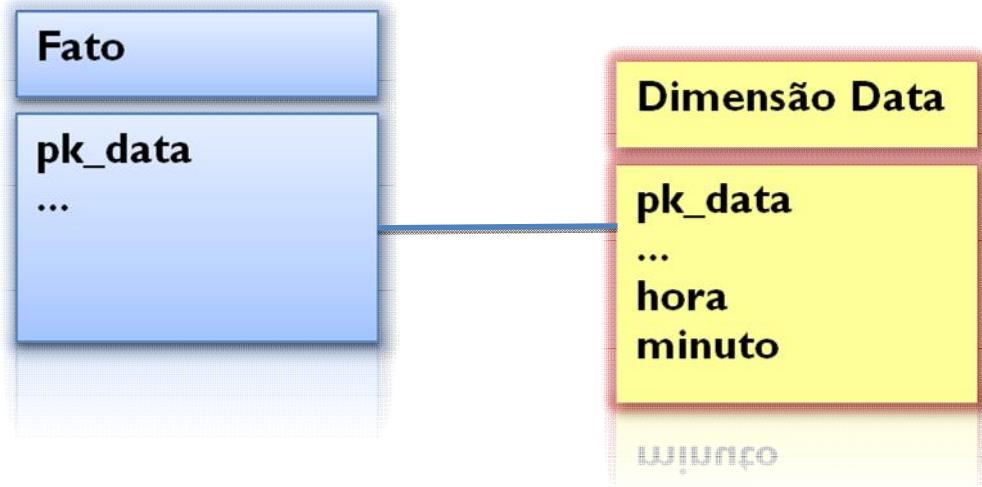


- ❖ Permite descrições e agregações úteis: (nomes de horas, nomes de turnos, etc);
- ❖ A chave `pk_hora_do_dia` pode ter 2 bytes (suficiente para $24 \times 60 = 1.440$ minutos) ou 4 bytes (suficiente para $1.440 \times 60 = 86.400$ segundos).

Conceitos de Modelagem Multidimensional

Dimensão Tempo: Horas, minutos, segundos

3^a Alternativa: hora e minutos na mesma tabela de dimensão que as datas.



- ❖ Tabela de dimensão muito grande;
- ❖ $10 \text{ anos} = 3.650 \times 1.440 \text{ minutos} = 5.256.000 \text{ linhas}$ (526.000 linhas para cada ano adicional).

Conceitos de Modelagem Multidimensional

Dimensão Produto (What)

Product Dimension
Product Key (PK)
Product Description
SKU Number (Natural Key)
Brand Description
Category Description
Department Description
Package Type Description
Package Size
Fat Content
Diet Type
Weight
Weight Units of Measure
Storage Type
Shelf Life Type
Shelf Width
Shelf Height
Shelf Depth
... and more

POS Retail Sales Transaction Fact
Date Key (FK)
Product Key (FK)
Store Key (FK)
Promotion Key (FK)
POS Transaction Number
Sales Quantity
Sales Dollar Amount
Cost Dollar Amount
Gross Profit Dollar Amount



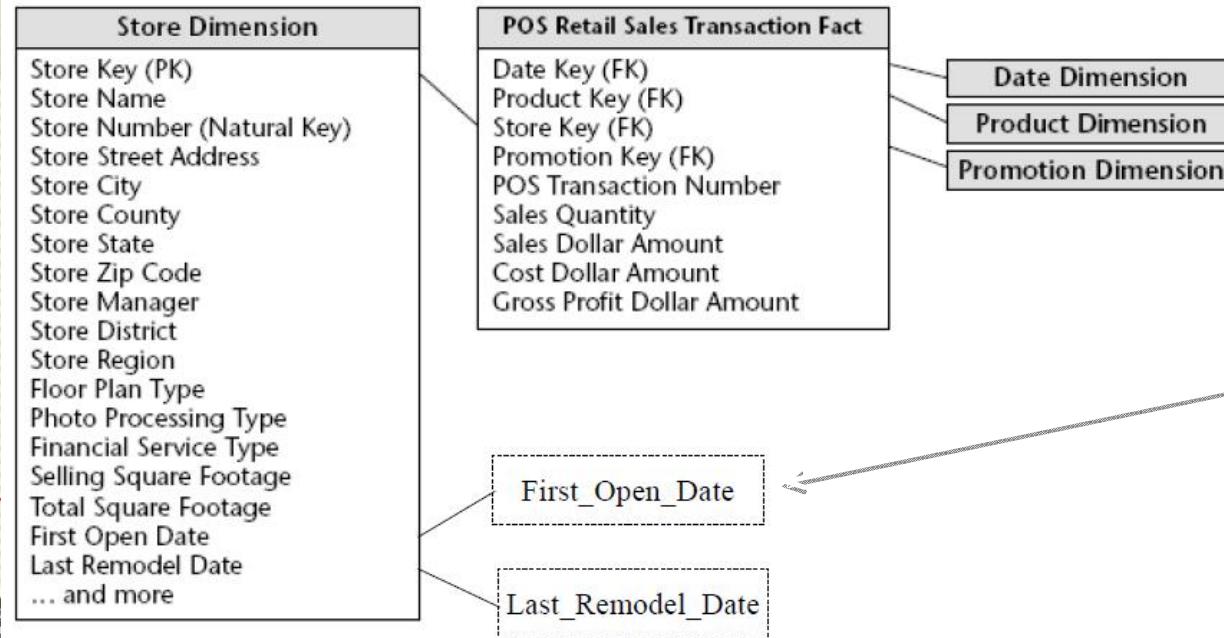
Product Key	Product Description	Brand Description	Category Description	Department Description	Fat Content
1	Baked Well Light Sourdough Fresh Bread	Baked Well	Bread	Bakery	Reduced Fat
2	Fluffy Sliced Whole Wheat	Fluffy	Bread	Bakery	Regular Fat
3	Fluffy Light Sliced Whole Wheat	Fluffy	Bread	Bakery	Reduced 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

Redundância à custa da 3FN vale a pena, pois as tabelas de dimensões são pequenas em relação às tabelas de fatos.



Conceitos de Modelagem Multidimensional

Dimensão Loja (Where)



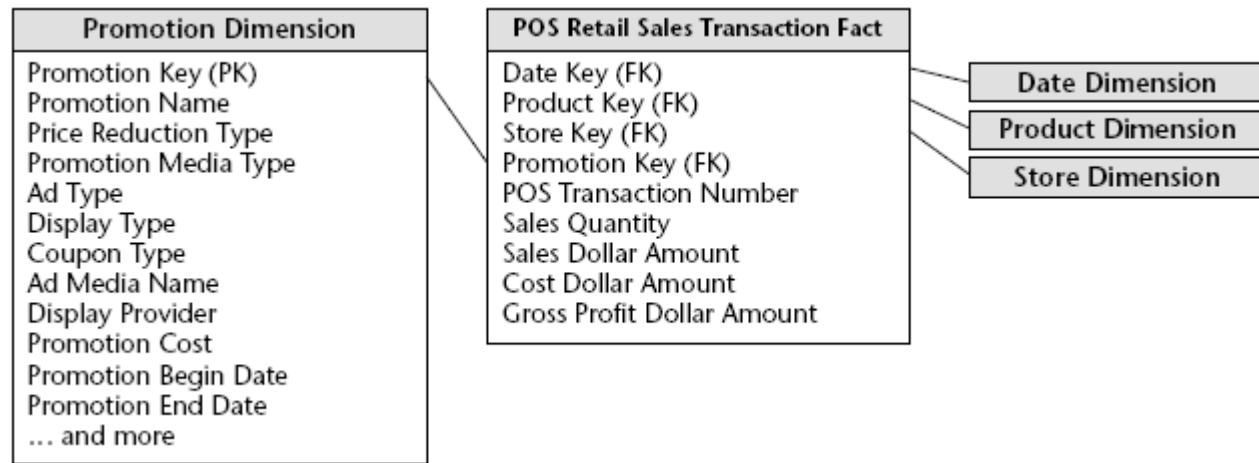
Esse tipo de tabela virtual para relacionar dimensões é denominado "outtrigger".

- ❖ Note os atributos First Open Date e Last Remodel Date, são DATAS. São chaves de junção com cópias da tabela de dimensão Date.



Conceitos de Modelagem Multidimensional

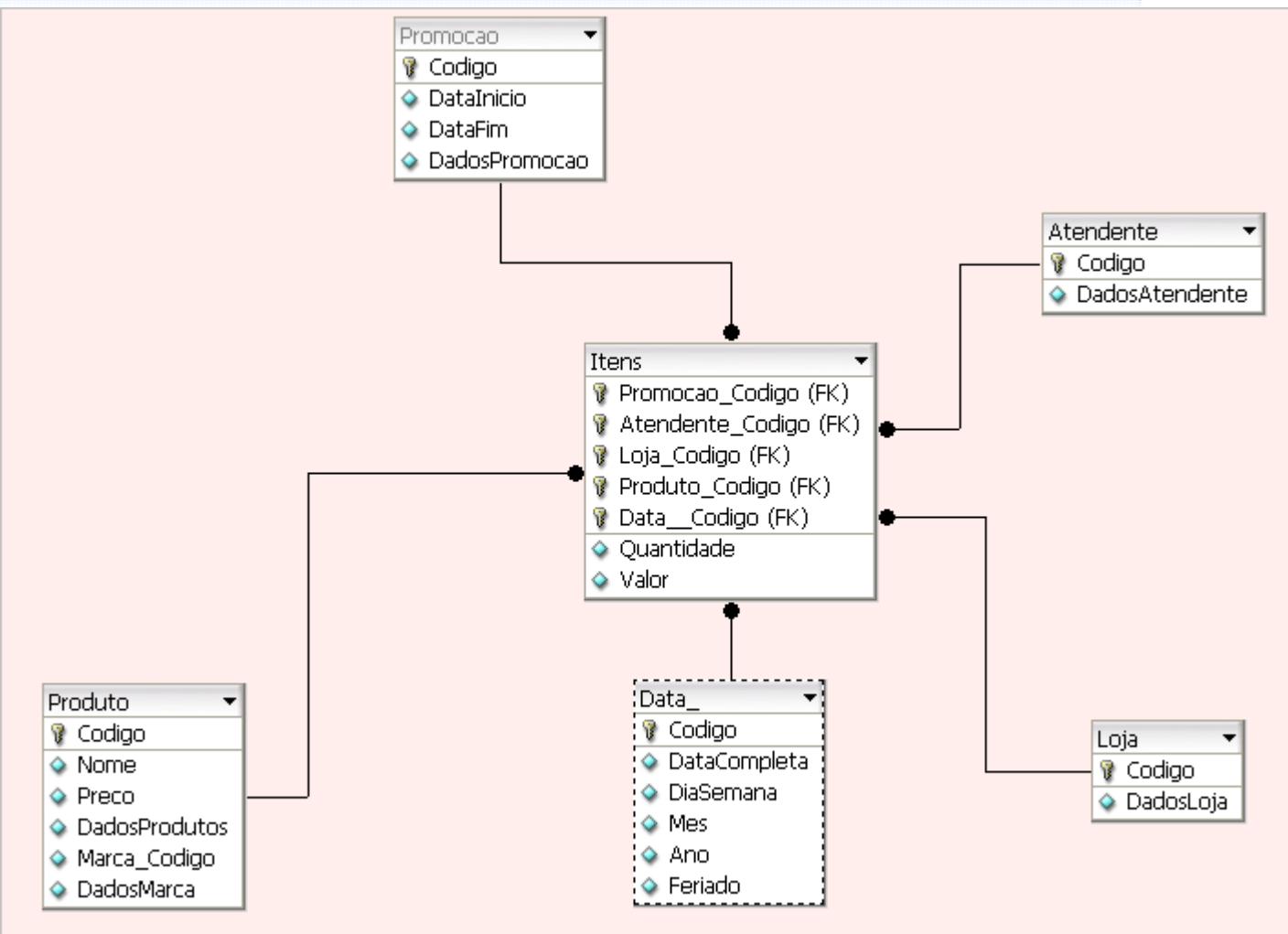
Dimensão Promoção (Why)



- ❖ A dimensão Promoção do exemplo é, de fato uma COMBINAÇÃO DE DIMENSÕES causais (price reduction, ads, display, coupon) que poderiam estar em quatro tabelas separadas, com o mesmo efeito;
- ❖ No caso, estão combinadas numa única tabela de dimensão porque são altamente correlatas;
- ❖ Dimensões combinadas economizam espaço da tabela de Fatos, embora separadas pudessem ser mais bem entendidas e mais facilmente administradas.



Tabelas de Fatos



Esquema estrela com tabela de fatos **Itens** e dimensões **Produto**, **Promoção**, **Atendente**, **Loja** e **Data**

Tabelas de Fatos

- **Chave primária composta por um subconjunto das chaves das dimensões (subconjunto que garanta unicidade – às vezes todas as chaves):**
 - ❖ Vide por exemplo a tabela Itens, se houvesse uma dimensão CupomFiscal (bastariam as chaves de CupomFiscal e de Produto como chave primária)
- **Por ser o DW histórico, a tabela de fatos tem muitas linhas (milhões, bilhões) e poucas colunas (chaves das dimensões e medidas dos fatos);**
- **Medidas do fatos são usualmente numéricas, mas podem ser não numéricas ou sem medida (tabelas sem fato).**
- **Fatos são tipicamente aditivos, mas podem ser:**
 - ❖ Semi-aditivos ou mesmo Não aditivos.

Tabelas de Fatos

Fatos Aditivos

- São numéricos e podem ser somados em relação às dimensões existentes;
 - ❖ Quantidade e valor podem ser somados ao longo de qualquer dimensão (Produto, Promoção, Atendente, Loja e Data);
 - ❖ Ex: Quantidade, Valor na Tabela Itens.
- Sempre que, em uma modelagem, um dado numérico for apresentado, então este será um bom indício de uma métrica que deverá ser colocada na FATO;
- Em geral, fatos aditivos representam medidas de atividade do negócio, ligadas ao seus indicadores de desempenho (**KPI – Key performance indicators**).



Tabelas de Fatos

Fatos Semi-aditivos

- Também são numéricos, mas não podem ser somados em relação a todas as dimensões existentes (a semântica não permite):
 - ❖ Ex.: Quantidade em estoque só pode ser somada ao longo da dimensão Produto. Nas dimensões Loja e Data, a soma não faria muito sentido (especialmente nesta última, nenhum sentido)
- Em geral, fatos semi-aditivos representam leituras medidas de intensidade do negócio:
 - ❖ São snapshots destas leituras que entram no DW.
 - ❖ O valor atual já leva em consideração valores passados.
- Fatos semi-aditivos típicos: Níveis de Estoque, Saldos, Fechamento diário/mensal de conta, etc...



Tabelas de Fatos

Fatos Não-aditivos

- Algumas observações não numéricas podem eventualmente ser fatos.
 - ❖ Ex: DW de registro de acidentes de trânsito
 - ▶ Atributos: carro1, carro2, motorista1, motorista2, descrição do acidente, descrição do tempo e descrição da pista.
- Informações textuais são fatos que só permitem contagem e estatísticas associadas a contagens. Alternativamente, poderiam ser modeladas como dimensões ligadas a uma tabela de fatos “sem fatos”, isto é, só para contagem.
 - ❖ Ex: DW de registro de inscrições em turmas por disciplina, por semestre, por curso, por aluno, por professor.



Business School Brasil
O futuro é agora. Junte-se a.

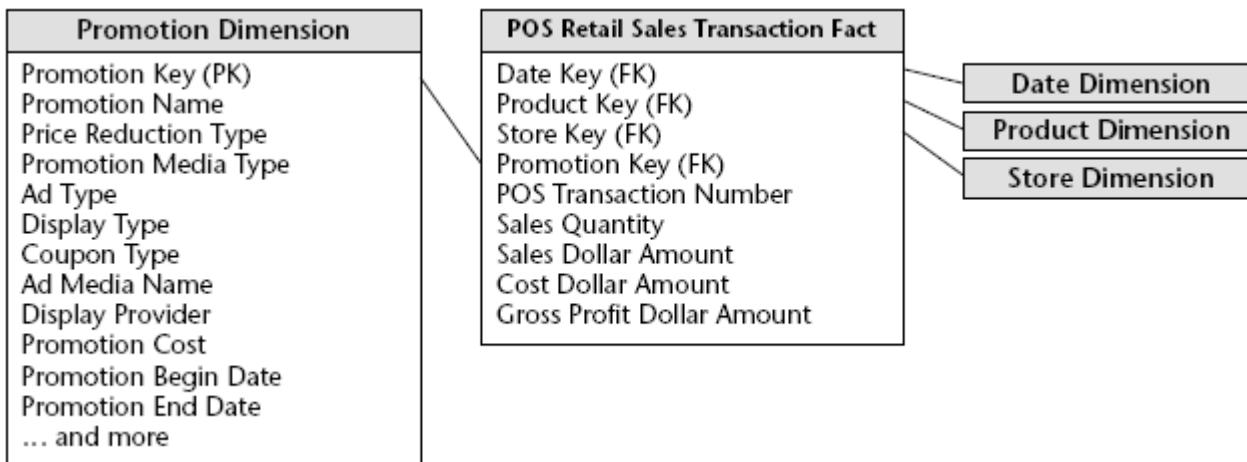


UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Tabelas de Fatos sem Fatos

Factless Fact Tables

- Uma tabela de fatos que não tem fatos mas captura alguns relacionamentos muitos-para-muitos entre chaves de dimensões. Mais frequentemente usada para representar eventos ou prover informação de cobertura que não aparece em outras tabelas de fatos.



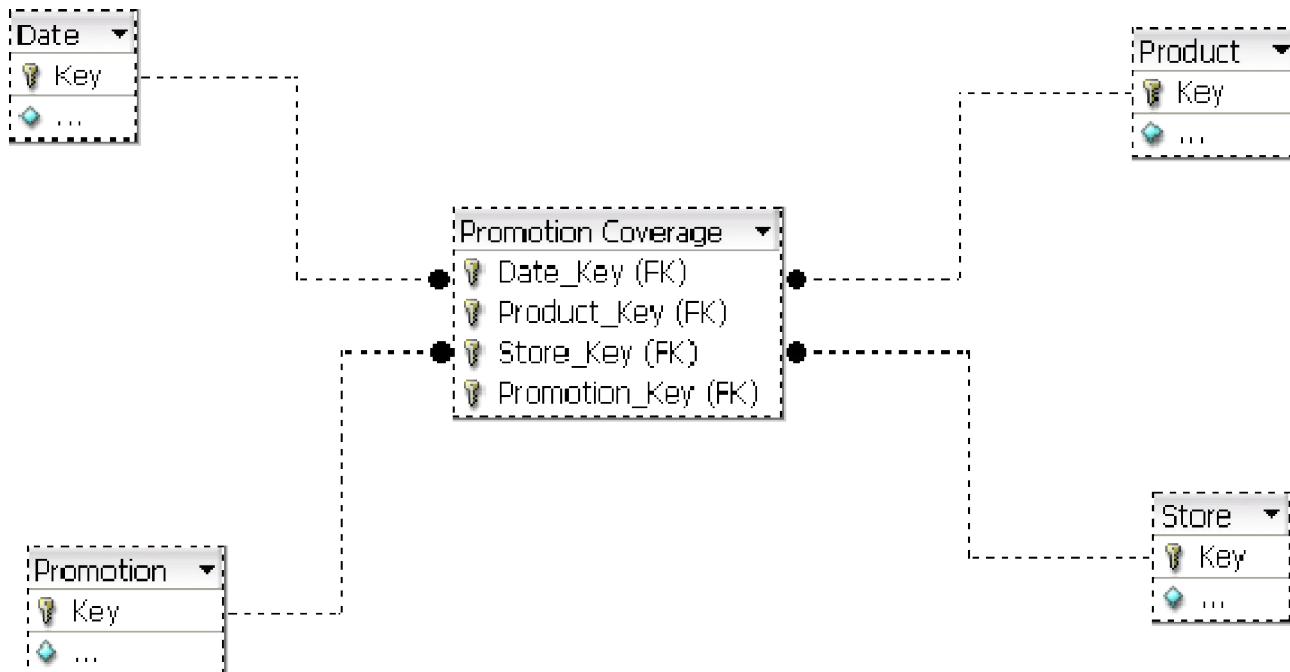
Quais produtos estavam em promoção mas não venderam?

Tabelas de Fatos sem Fatos

Factless Fact Tables

Por que não pode? Por que não deveria?

- A solução é criar uma Tabela de Cobertura de promoção com as mesmas dimensões da tabela de Vendas (Data, Produto, Loja, Promoção).



Tabelas de Fatos sem Fatos

Factless Fact Tables

- Os produtos em promoção que não venderam será o conjunto diferença entre a cobertura e as vendas;
- Uma tabela de fatos, tipicamente sem fatos, que registra todos os produtos que estão em promoção numa determinada loja, independentemente de ser vendidos ou não;
- Consulta: Quais produtos estavam em promoção mas não venderam?

```
SELECT Product_Key, ... FROM Promotion_Coverage, ...
```

```
WHERE ...
```

```
MINUS
```

```
SELECT Product_Key, ... FROM POS_Retail_Sales, ... WHERE
```

```
...
```



Tabela de Dimensões

- **Objetivo:**

- ❖ Contém descrições textuais do negócio (fato);
- ❖ Atributos de dimensões servem como cabeçalho das linhas e colunas das análises e filtro nas consultas e relatórios.

Dimensão Produto

pk_produto
descricao
categoria
marca

- **Características:**

- ❖ Chaves simples (em geral, artificiais: “surrogate keys”);
 - ▶ Números inteiros de 4 bytes.
- ❖ Muitas colunas (dezenas); poucas linhas (centenas ou milhares) se comparadas com tabelas de fatos;

Dimensão Loja

pk_loja
nome_loja
endereco
cidade
estado

Tabela de Dimensões

- **Características:**
 - ❖ Usualmente não dependente do tempo
 - ▶ » Tempo é outra dimensão (quase sempre presente)
 - ❖ Desnormalizada (em geral, na 2a forma normal);
 - ❖ Hierarquias implícitas (à custa da 3a forma normal).
- **Fonte principal das cláusulas das consultas, agrupamentos e títulos de relatórios:**
 - ✓ Volume de vendas “por produto”.

Dimensão Produto

pk_produto
descricao
categoria
marca

Dimensão Loja

pk_loja
nome_loja
endereço
cidade
estado



Business School Brasil
O futuro é agora. Junte-se a!



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Tabelas de Dimensões

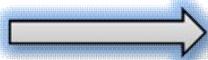
2^a Forma Normal

- Informalmente:

❖ Uma relação está em 2FN se todo atributo não-primo (isto é, que não seja membro de chave) for totalmente dependente de qualquer chave.

Tabela
Estoque_1
não está em
2FN

Estoque_1
Loja_Codigo
Produto_Codigo
NomeLoja
DadosLoja
NomeProduto
DadosProduto
Quantidade
Valor



Estoque_2
Loja_Codigo
Produto_Codigo
Quantidade
Valor

Tabela
Estoque_2
está em
2FN e em
3FN

Em geral, as tabelas de fatos são normalizadas em 3FN

Tabelas de Dimensões

3^a Forma Normal

- Informalmente:

❖ Uma relação está em 3FN se estiver em 2FN e nenhum atributo não-primo (isto é, que não seja membro de uma chave) for transitivamente dependente da chave.

Tabela
Produto
não está
em 3FN

Produto
Codigo
Nome
Preco
DadosProdutos
Marca_Codigo
DadosMarca
Categoria
DadosCategoria



Produto
Codigo
Nome
Preco
DadosProdutos
Marca_Codigo (FK)
Categoria_Codigo (FK)



Tabela
Produto
está em
3FN

Num esquema estrela, as tabelas de dimensões não são normalizadas em 3FN; estão apenas em 2FN



Tabelas de Dimensões

A tabela abaixo está na 3FN ou na 2FN?

Vencedores de Torneios

Torneio	Ano	Vencedor	Data de nascimento do vencedor
Indiana Invitational	1998	Al Fredrickson	21 de julho de 1975
Cleveland Open	1999	Bob Albertson	28 de setembro de 1968
Des Moines Masters	1999	Al Fredrickson	21 de julho de 1975
Indiana Invitational	1999	Chip Masterson	14/3/1977

- ❖ A tabela não está na terceira forma normal porque o atributo "data de nascimento do vencedor" é dependente transitivamente de {Torneio, Ano} via o atributo "Vencedor". O fato de a data de nascimento do vencedor não ser dependente do vencedor torna a tabela vulnerável a inconsistências lógicas, já que nada impediria de uma mesma pessoa aparecer com datas de nascimento diferentes em dois registros



Business School Brasil
O futuro começa aqui



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAU

Tabelas de Dimensões

Vencedores de Torneios			
<u>Torneio</u>	<u>Ano</u>	Vencedor	Data de nascimento do vencedor
Indiana Invitational	1998	Al Fredrickson	21 de julho de 1975
Cleveland Open	1999	Bob Albertson	28 de setembro de 1968
Des Moines Masters	1999	Al Fredrickson	21 de julho de 1975
Indiana Invitational	1999	Chip Masterson	14/3/1977



Vencedores de Torneios

<u>Torneio</u>	<u>Ano</u>	Vencedor
Indiana Invitational	1998	Al Fredrickson
Cleveland Open	1999	Bob Albertson
Des Moines Masters	1999	Al Fredrickson
Indiana Invitational	1999	Chip Masterson

Datas de nascimento de jogadores

<u>Jogador</u>	Data de nascimento
Chip Masterson	14/3/1977
Al Fredrickson	21 de julho de 1975
Bob Albertson	28 de setembro de 1968

Tabelas de Dimensões

Hierarquia de Dimensões

- **Informalmente:**

- ❖ Uma dimensão pode ter múltiplas hierarquias além de outros atributos descritivos

- **Exemplos:**

- ❖ Para a dimensão Loja:

- ▶ Geografia física: CEP, cidade, estado, região, país;
 - ▶ Geografia de vendas: território, região, zona;
 - ▶ Geografia de distribuição: área primária, região.

- ❖ Para a dimensão Produto:

- ▶ Hierarquia de Marcas;
 - ▶ Hierarquia de Categorias;
 - ▶ Hierarquia de Tipo de Armazenamento.

Tabelas de Dimensões

- Segundo KIMBALL, as tabelas de dimensão não devem ser normalizadas pois:

Não há atualização freqüente nas bases

O espaço em disco economizado é relativamente pequeno

Esse ganho de espaço não justifica a perda de performance na realização de consultas por conta das junções necessárias em caso de normalização.

Tabelas de Dimensões

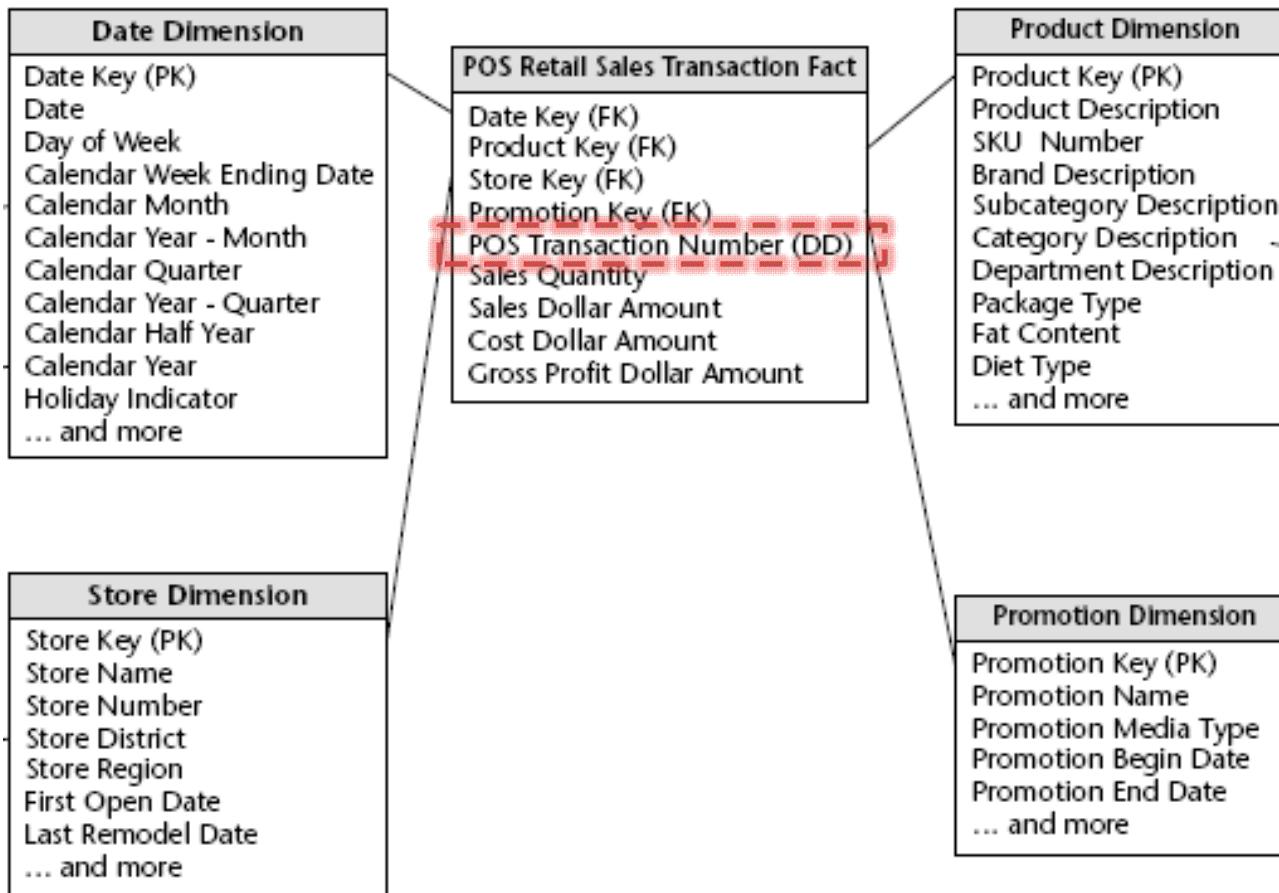
Dimensões Degeneradas

- Chaves de dimensão na tabela de fatos sem tabelas de dimensão correspondentes;
- Uma chave de dimensão, como o número de uma transação, número de fatura, tiquete, nota fiscal, pedido ou ordem de compra, que não tenha nenhum atributo portanto não se junta com uma tabela de dimensão;
- Esses documentos normalmente são compostos de itens, e se a granularidade da tabela de fatos for item, o número do documento estará na tabela fato apenas para permitir o agrupamento dos itens por documento.



Tabelas de Dimensões

Dimensões Degeneradas

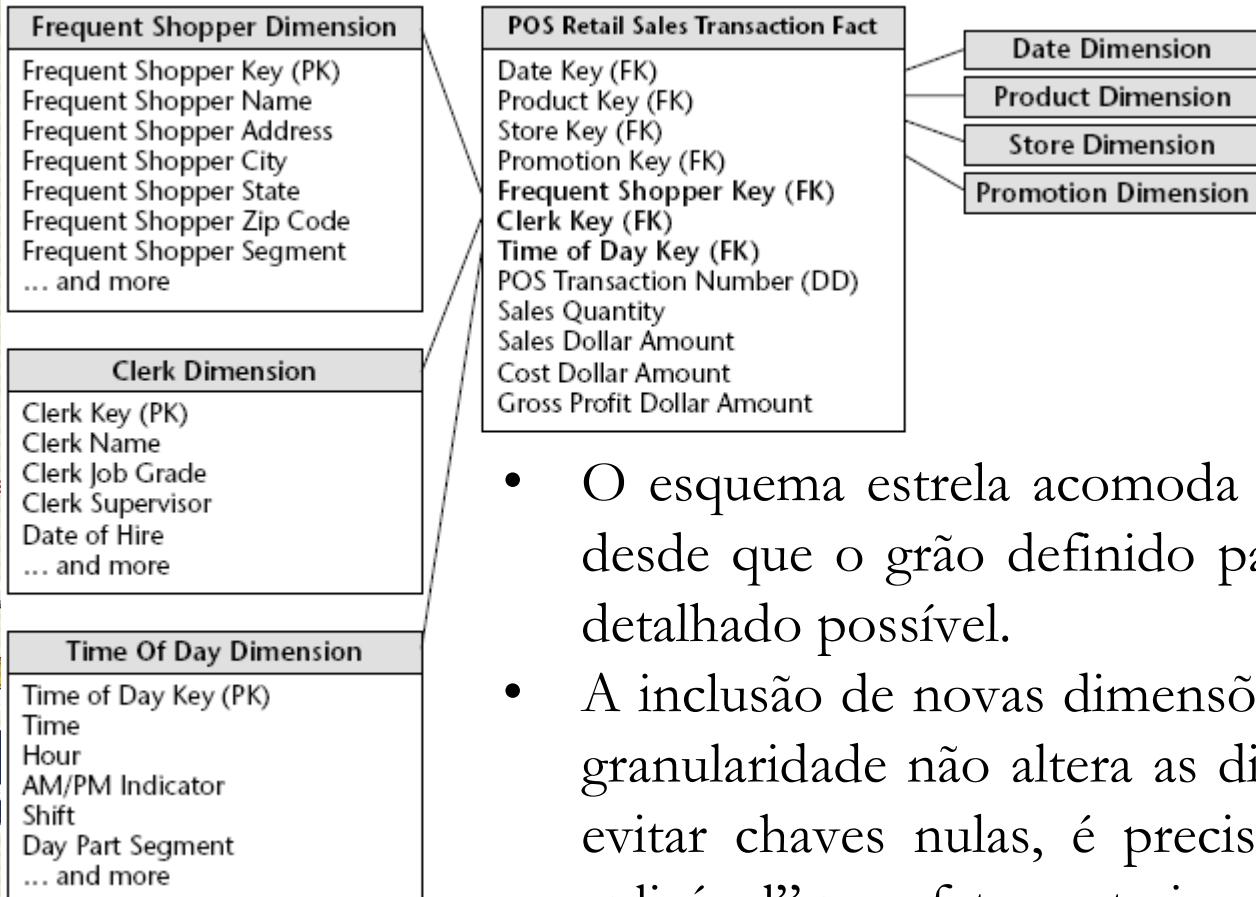


POS Transaction Number é uma Dimensão Degenerada (DD)



Tabelas de Dimensões

Extensibilidade do Esquema Estrela



- O esquema estrela acomoda sem problema novas dimensões desde que o grão definido para a tabela de fatos seja o mais detalhado possível.
- A inclusão de novas dimensões que se aplicam a esse nível de granularidade não altera as dimensões e fatos existentes. Para evitar chaves nulas, é preciso prever valores do tipo “Não aplicável” para fatos anteriores à inclusão da nova dimensão.
- Todas as consultas e aplicações existentes continuam a rodar sem nenhuma alteração.

Tabelas de Dimensões

Extensibilidade do Esquema Estrela

- Modificações absorvidas naturalmente pelo esquema estrela, devido a mudança nas fontes ou por decisão de modelagem, sem impacto nas aplicações existentes:
 - ❖ Novos atributos de dimensões;
 - ❖ Novas dimensões;
 - ❖ Novos fatos medidos (na mesma tabela de fatos ou em nova tabela);
 - ❖ Dimensões mais granulares;
 - ❖ Adição de uma fonte de dados nova envolvendo dimensões existentes assim como novas dimensões não previstas.
- A extensibilidade é possível graças à simetria do esquema estrela, contanto que o grão inicial escolhido seja o mais detalhado possível.



Tabelas de Dimensões

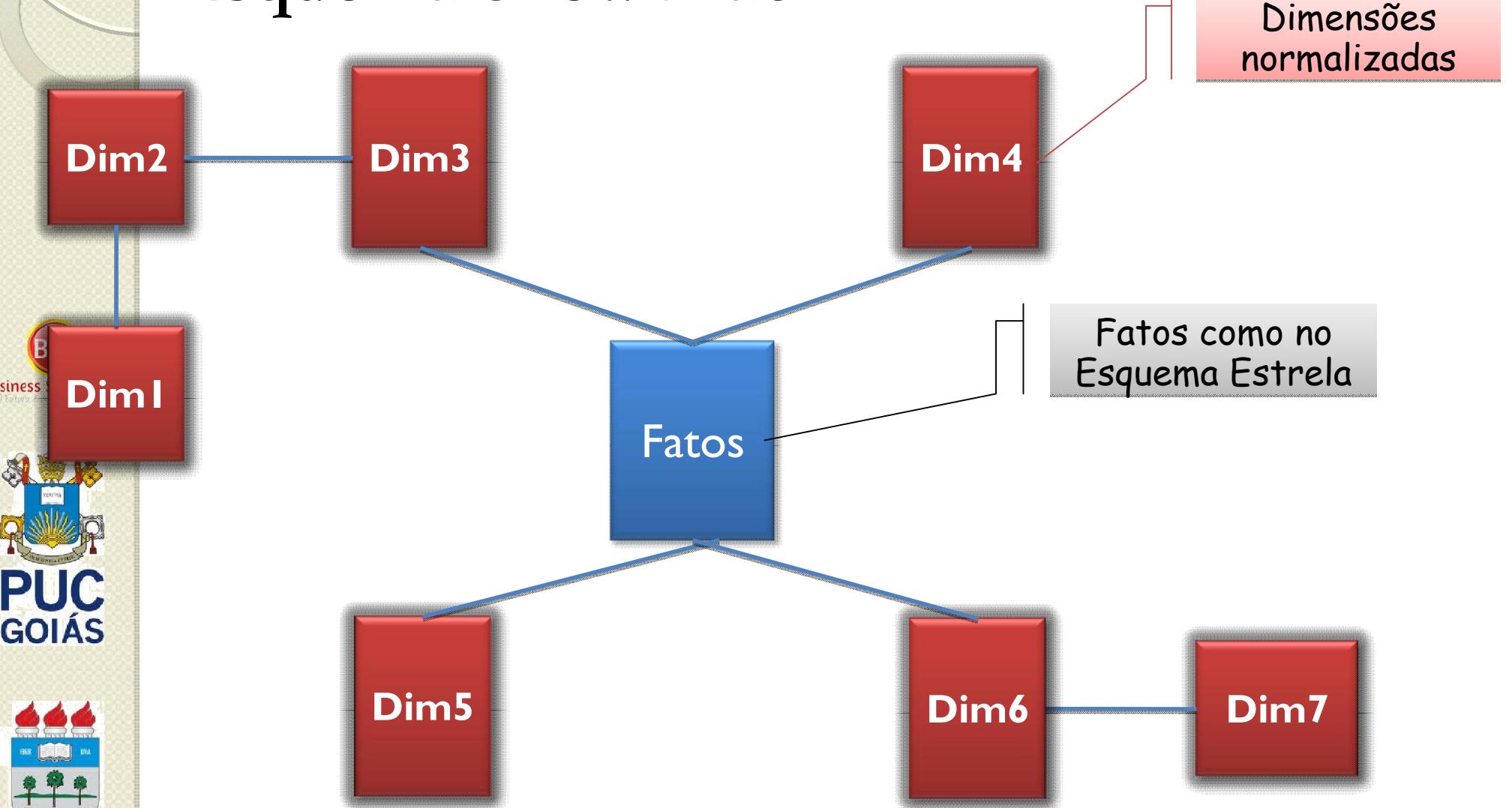
Esquema Dimensional Snow Flack

- O esquema floco de neve é uma variação do esquema estrela no qual todas as tabelas dimensão são normalizadas na terceira forma normal (3FN);
 - ❖ Tabela de fatos;
 - ✓ Sem redundância
 - ❖ K tabelas de dimensões
 - ✓ 1 dimensão pode ter várias tabelas
- Reduzem a redundância mas aumentam a complexidade do esquema e consequentemente a compreensão por parte dos usuários;
- Dificultam as implementações de ferramentas de visualização dos dados;
- Hierarquias representadas pelos relacionamentos entre as dimensões.



Tabelas de Dimensões

Esquema Snow Flack



Tabelas de Dimensões

Esquema Snow Flack

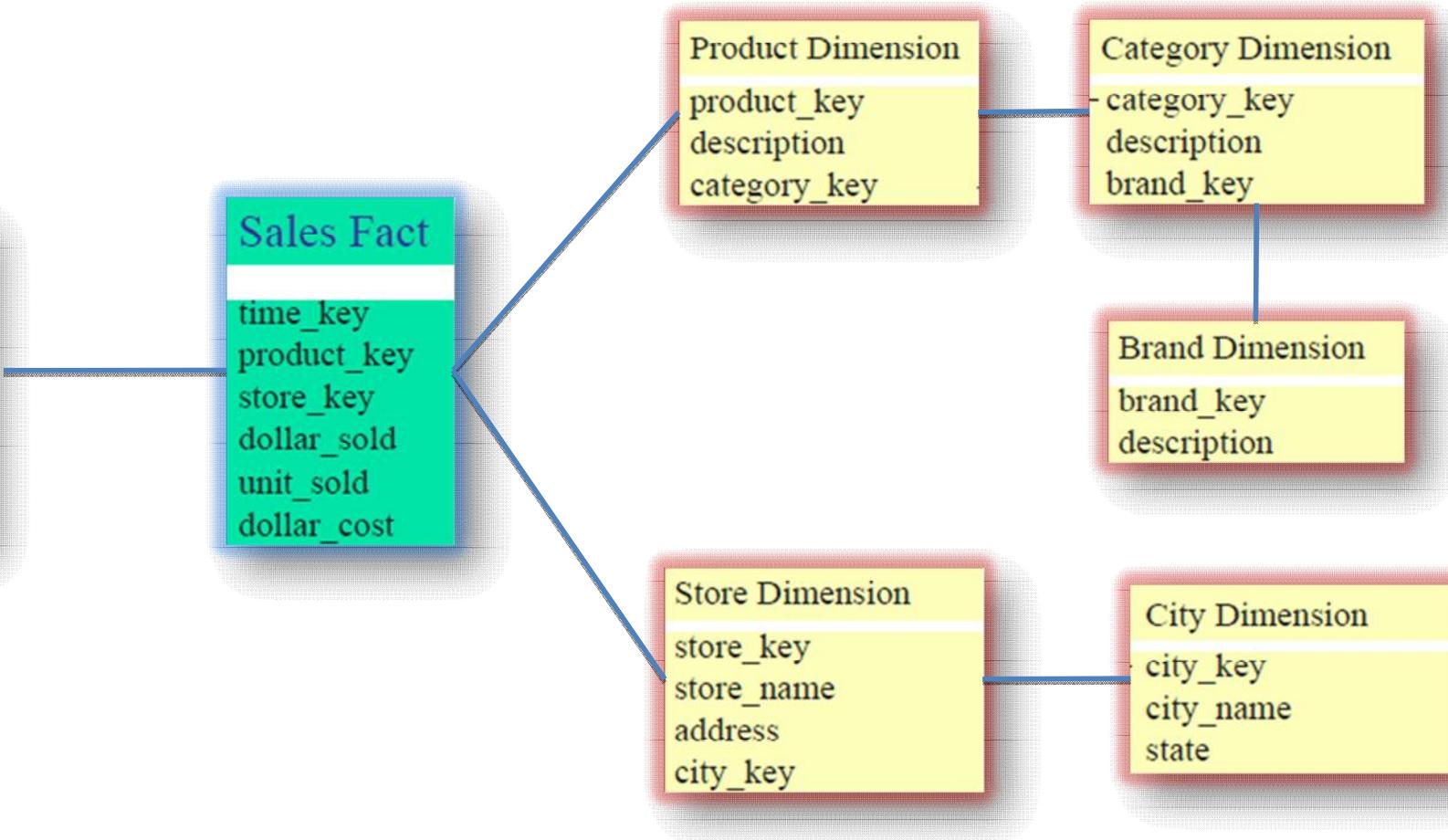
Business
Analytics



PUC
GOIÁS

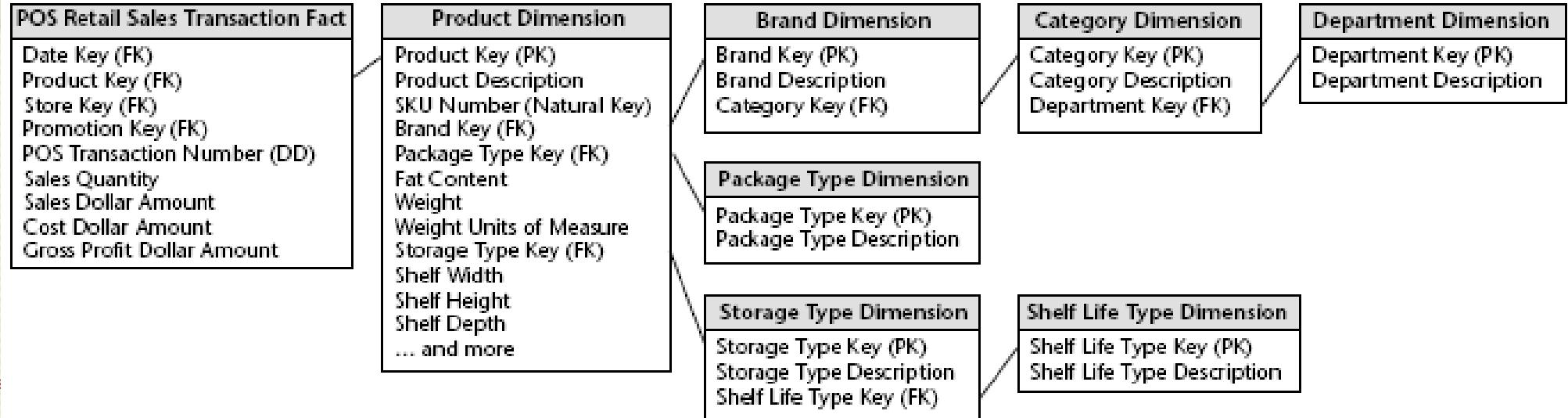


UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ



Tabelas de Dimensões

Esquema Dimensional Snow Flack



- Embora aceitável, a normalização de dimensões não é recomendável por razões de desempenho e facilidade de uso
 - ❖ A quantidade de tabelas torna a apresentação do modelo mais complexa;
 - ❖ Otimizadores do SGBD têm mais dificuldade com esquema complexo;
 - ❖ A economia de espaço em disco é insignificante em relação ao DW completo;



Tabelas de Dimensões

Esquema Dimensional Snow Flack

POS Retail Sales Transaction Fact
Date Key (FK)
Product Key (FK)
Store Key (FK)
Promotion Key (FK)
POS Transaction Number (DD)
Sales Quantity
Sales Dollar Amount
Cost Dollar Amount
Gross Profit Dollar Amount

Product Dimension
Product Key (PK)
Product Description
SKU Number (Natural Key)
Brand Key (FK)
Package Type Key (FK)
Fat Content
Weight
Weight Units of Measure
Storage Type Key (FK)
Shelf Width
Shelf Height
Shelf Depth
... and more

Brand Dimension
Brand Key (PK)
Brand Description
Category Key (FK)

Category Dimension
Category Key (PK)
Category Description
Department Key (FK)

Department Dimension
Department Key (PK)
Department Description

Package Type Dimension
Package Type Key (PK)
Package Type Description

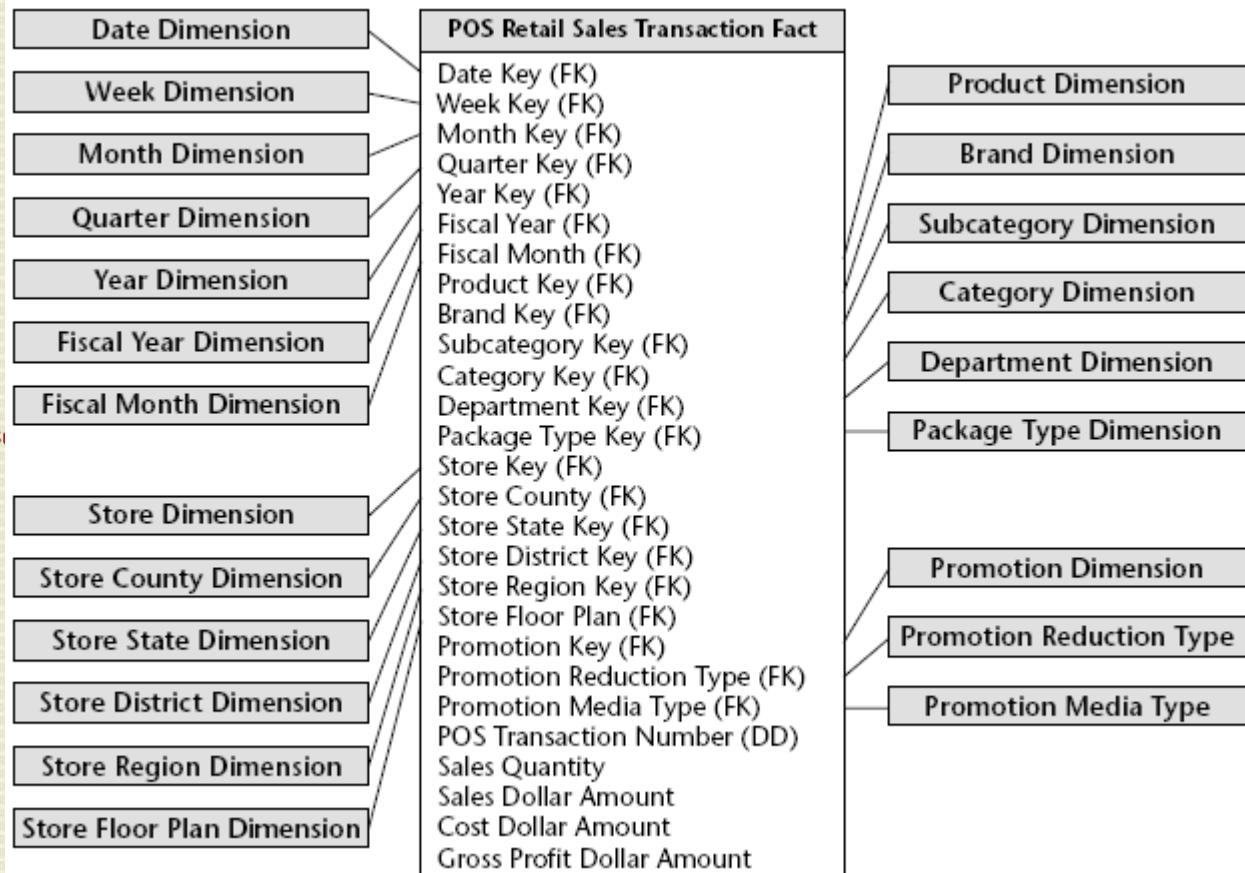
Storage Type Dimension
Storage Type Key (PK)
Storage Type Description
Shelf Life Type Key (FK)

Shelf Life Type Dimension
Shelf Life Type Key (PK)
Shelf Life Type Description

- ❖ Snowflaking diminui a habilidade de usuários de navegar na dimensão;
- ❖ Snowflaking impede o uso de índices tipo BitMap, que são usados por SGBD para indexar campos com baixa cardinalidade.

Tabelas de Dimensões

Esquema Centopeia



- ❖ Um número de dimensões muito grande (25+) é um sinal de que muitas dimensões não são completamente independentes e deveriam ser combinadas numa única. É um erro em modelagem dimensional representar elementos de uma hierarquia como dimensões separadas.



Tabelas de Dimensões

Campos Chaves de Tabelas de Dimensões

- Regra básica: uso de surrogates ou chaves artificiais.
 - ❖ Ajudam a manter a estabilidade, através da neutralidade;
 - ❖ Evitam manutenção custosa de tabelas, especialmente das tabelas fatos;
 - ❖ Chaves naturais podem ter problemas de unicidade, ausência, tamanhos exagerados;
 - ❖ Chaves artificiais podem ser especificadas como inteiros de 4 bytes, alcançando até 2^{32} , isto é, mais de 2 bilhões de ocorrências (inteiros positivos), o que é mais do que necessário para qualquer tabela dimensão;
 - ❖ Chaves artificiais ficam transparentes (invisíveis) para os usuários, servindo apenas como ligação entre dimensões e fatos;
 - ❖ Campos naturais não chave poderão ser indexados, tornando as consultas amistosas;



Tabelas de Dimensões

Campos Chaves de Tabelas de Dimensões

- Regra básica: uso de surrogates ou chaves artificiais.
 - ❖ Se produzidas automaticamente, deve-se ter cuidado no processo de preparação (ETL), especialmente nos reprocessamentos.
 - ❖ A única desvantagem das chaves artificiais é que não faz sentido a tabela fato ser consultada diretamente, pois os campos descritivos de filtro estarão armazenados nas dimensões.

Cada join entre dimensão e tabelas de fato no Data Warehouse deve basear-se em surrogates. Deve-se evitar usar os códigos naturais do sistema origem. Nenhuma das chaves deve ser inteligente.



Tabelas de Dimensões

Dinâmica das Dimensões

- Atualização das dimensões que mudam lentamente (**Slowly Changing Dimensions**)
 - ❖ Exemplos: Endereço de Cliente, Descrição de Produto.
 - ▶ Várias alternativas
 - ❖ Tipo 1: Atualizar por cima do valor antigo
 - ▶ É simples mas não preserva histórico.
 - ❖ Tipo 2: Adicionar uma nova linha com o novo valor do atributo atualizado, mantendo os demais.
 - ▶ A nova linha particiona o histórico na tabela fato.;
 - ▶ É a técnica predominante para dimensões que mudam lentamente (*slowly changing dimensions*).

Tabelas de Dimensões

Dinâmica das Dimensões

- ❖ **Tipo 3: Adicionar uma nova coluna, preservando o valor anterior e inserindo o novo valor na nova coluna.**
 - ▶ Permite a manutenção de duas visões simultâneas do histórico, mas dá margem a muitos valores nulos quando as mudanças são lentas.
- ❖ **Tipo 6: Soluções híbridas, com múltiplas versões (linhas) combinadas ou não com coluna de valor anterior.**
 - ▶ Mais flexíveis e completas, porém mais complexas.



Tabelas de Dimensões

Dinâmica das Dimensões

Product Key	Product Description	Department	SKU Number (Natural Key)
12345	IntelliKidz 1.0	Education	ABC922-Z

Tupla Original

Mudança: O produto Intellikidz 1.0 mudou de departamento



**PUC
GOIÁS**

Product Key	Product Description	Department	SKU Number (Natural Key)
12345	IntelliKidz 1.0	Strategy	ABC922-Z

Tipo 1



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Tabelas de Dimensões

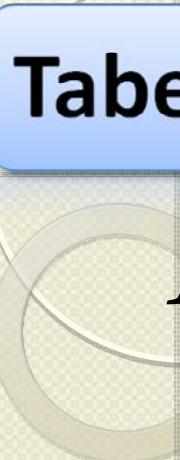
Dinâmica das Dimensões

Product Key	Product Description	Department	SKU Number (Natural Key)
12345	IntelliKidz 1.0	Education	ABC922-Z
25984	IntelliKidz 1.0	Strategy	ABC922-Z

Tipo 2

Product Key	Product Description	Department	Prior Department	SKU Number (Natural Key)
12345	IntelliKidz 1.0	Strategy	Education	ABC922-Z

Tipo 3



BSB
Business School Brasil
O futuro começa aqui!



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Tabelas de Dimensões

Dinâmica das Dimensões: Tipo 6

Product Key	Product Description	Current Department	Historical Department	SKU Number (Natural Key)
12345	IntelliKidz 1.0	Education	Education	ABC922-Z

Tupla Original

Requisito: Preservar histórico e ao mesmo tempo suportar consultas a dados históricos de acordo com valores atuais.

Product Key	Product Description	Current Department	Historical Department	SKU Number (Natural Key)
12345	IntelliKidz 1.0	Strategy	Education	ABC922-Z
25984	IntelliKidz 1.0	Strategy	Strategy	ABC922-Z

1ª Mudança

Product Key	Product Description	Current Department	Historical Department	SKU Number (Natural Key)
12345	IntelliKidz 1.0	Critical Thinking	Education	ABC922-Z
25984	IntelliKidz 1.0	Critical Thinking	Strategy	ABC922-Z
31726	IntelliKidz 1.0	Critical Thinking	Critical Thinking	ABC922-Z

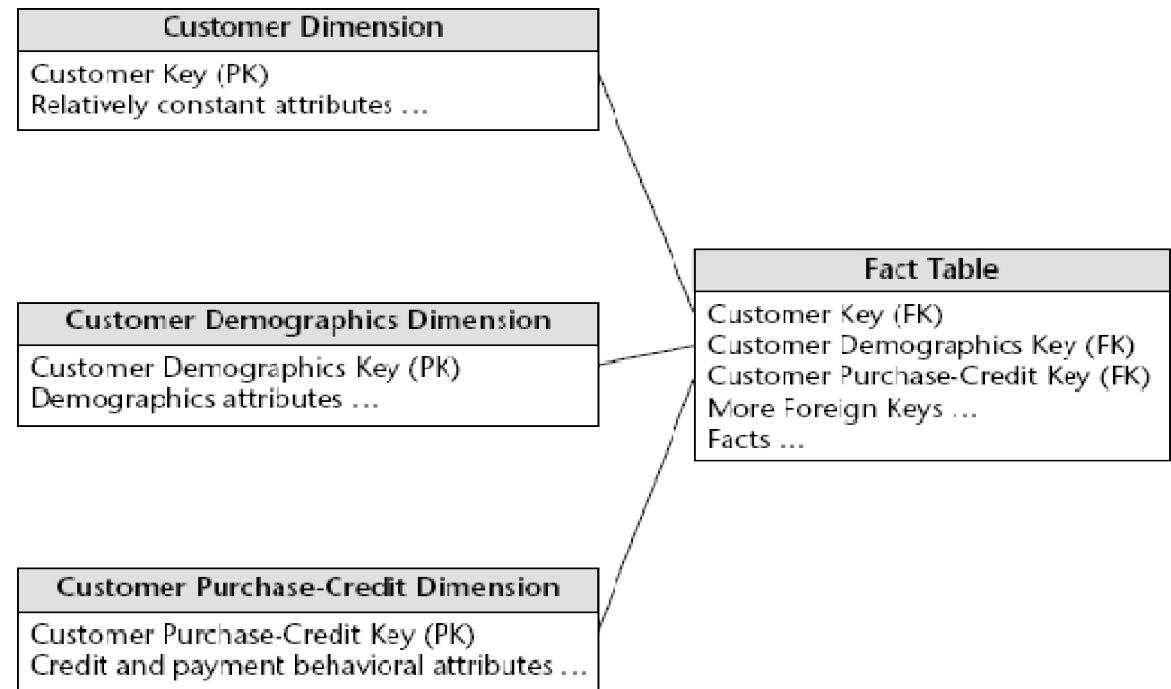
2ª Mudança

Tabelas de Dimensões

Dimensões: Grande volume e alta volatilidade

- Também chamadas de *Rapidly Changing Monster Dimensions*.
- Solução para dimensões grandes com mudanças frequentes (por exemplo, alguns atributos mudam mensalmente)
 - ❖ Particionamento da dimensão em tabelas diferentes, separando-se dados estáticos de dados voláteis.
 - ▶ Dimensões são relacionadas entre si e ambas relacionadas com a tabela fato.

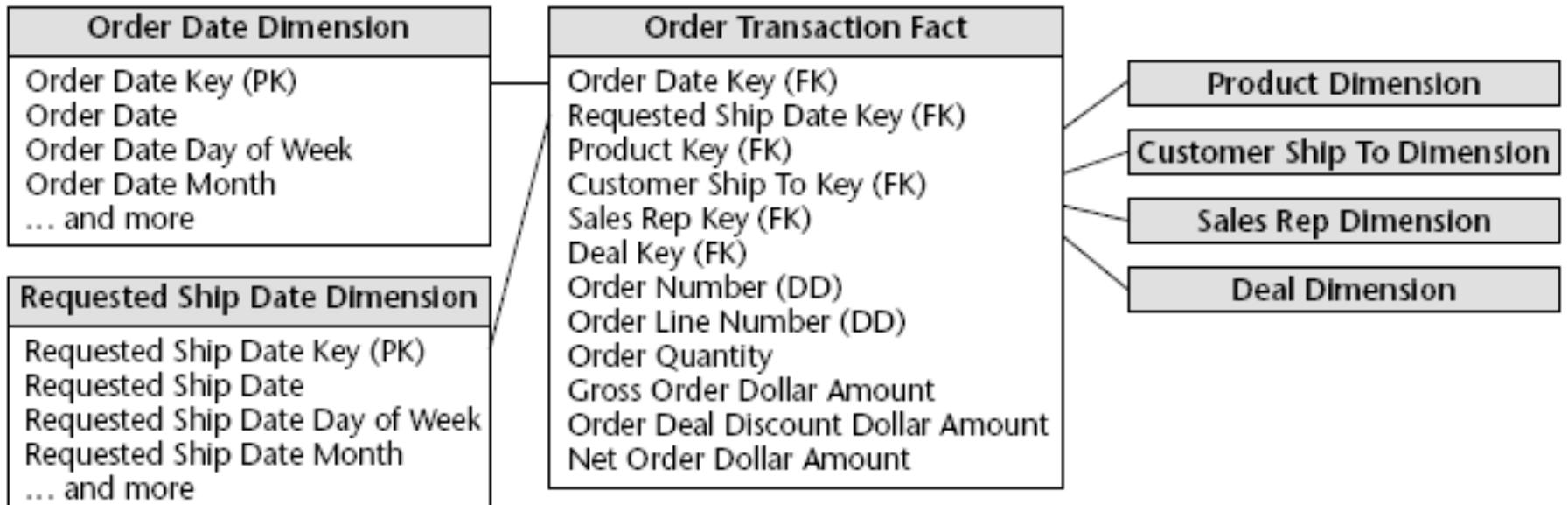
As várias tabelas quebradas de uma dimensão grande são também chamadas de minidimensões



Tabelas de Dimensões

Dimensões com vários papéis

- Também chamadas de *Role Playing Dimensions*;
- A situação onde uma mesma dimensão aparece várias vezes na mesma tabela de fatos. Cada um dos papéis da dimensão é representado por uma tabela lógica separada com nomes de coluna únicos através de visões.



Tabelas de Dimensões



Valores de Exceção



Tabelas de Dimensões

Valores de Exceção

- **Sem Informação** – *Informa que o campo correspondente do sistema-fonte está vazio;*
- **Código Inválido** – *Denuncia um código que não está tabelado no sistema-fonte;*
- **Sem Contexto** – *Informa que, no MD, não há significado associado ao cruzamento da métrica com a dimensão;*



Tabelas de Dimensões

Outros Tipos Especiais de Dimensões

- **Dimensão lixo ou sucata (*junk dimension*)**
 - ❖ Uma dimensão abstrata com a decodificação de um grupo de *flags* e indicadores de baixa cardinalidade, portanto removendo os *flags* da tabela de fatos.
- **Minidimensões**
 - ❖ Subconjuntos de uma dimensão grande, como Cliente, que são quebrados em dimensões artificiais menores para controlar o crescimento explosivo de uma dimensão grande, com mudança rápida;
 - ❖ Os atributos demográficos continuamente mutáveis de um cliente são freqüentemente modelados como uma minidimensão separada.

Tabelas de Dimensões

Outros Tipos Especiais de Dimensões

- **Dimensões com “Outrigger”**
 - ❖ Solução normalizada (*snowflake*) para conjuntos de atributos de baixa cardinalidade em dimensões grandes, como Cliente;
 - ❖ A economia de espaço vale a pena porque a dimensão é grande, e a carga de dados é separada do restante da dimensão porque os dados provêm de fontes externas diferentes.
- **Dimensões multivaloradas (*Bridge Table*)**
 - ❖ Normalmente, uma tabela de fatos possui conexões somente para dimensões representando um valor simples, como uma data ou produto;



Tabelas de Dimensões

Outros Tipos Especiais de Dimensões

- **Dimensões multivaloradas (tabela ponte)**

- ❖ Ocasionalmente, é válido conectar um registro de fato a uma dimensão representando um número aberto de valores, como o número de diagnósticos simultâneos que um paciente pode ter num momento de um mesmo tratamento;
- ❖ Neste caso, dizemos que a tabela de fatos tem uma dimensão multivalorada. Tipicamente manipulada por uma tabela ponte (*Bridge Table*) também chamada *Helper Table*, Tabela Associativa).



Tabelas de Dimensões

Dimensão lixo ou sucata (junk dimension)

- Relacionadas com tabelas tipo código-descrção com baixa cardinalidade: Sexo, Estado Civil, Tags diversos, Textos descritivos, etc. São campos tipo miscelânea que não trazem muita correlação com os outros campos da tabela fato, mas são usados como filtro, daí serem dimensões.
- Podem ser usadas de forma combinada.
 - ❖ Exemplo: três tags binários → $2^3 = 8$ combinações possíveis
- Usado como artifício para diminuir a tabela de fatos.
Exemplo:



Tabelas de Dimensões

Dimensão lixo ou sucata (junk dimension)

Order Indicator Key	Payment Type Description	Payment Type Group	Inbound/Outbound Order Indicator	Commission Credit Indicator	Order Type Indicator
1	Cash	Cash	Inbound	Commissionable	Regular
2	Cash	Cash	Inbound	Non-Commissionable	Display
3	Cash	Cash	Inbound	Non-Commissionable	Demonstration
4	Cash	Cash	Outbound	Commissionable	Regular
5	Cash	Cash	Outbound	Non-Commissionable	Display
6	Discover Card	Credit	Inbound	Commissionable	Regular
7	Discover Card	Credit	Inbound	Non-Commissionable	Display
8	Discover Card	Credit	Inbound	Non-Commissionable	Demonstration
9	Discover Card	Credit	Outbound	Commissionable	Regular
10	Discover Card	Credit	Outbound	Non-Commissionable	Display
11	MasterCard	Credit	Inbound	Commissionable	Regular
12	MasterCard	Credit	Inbound	Non-Commissionable	Display
13	MasterCard	Credit	Inbound	Non-Commissionable	Demonstration
14	MasterCard	Credit	Outbound	Commissionable	Regular

Tabelas de Dimensões

Minidimensões

- A melhor abordagem para tratar atributos em dimensões muito grandes é quebrar em uma ou mais minidimensões, cada uma contendo atributos que tenham um número limitado de valores. Exemplo: dimensão Cliente com milhões de ocorrências.

Customer Dimension
Customer Key (PK)
Customer ID (Natural Key)
Customer Name
Customer Address
Customer Date of Birth
Customer Date of 1st Order
...
Age
Gender
Annual Income
Number of Children
Marital Status



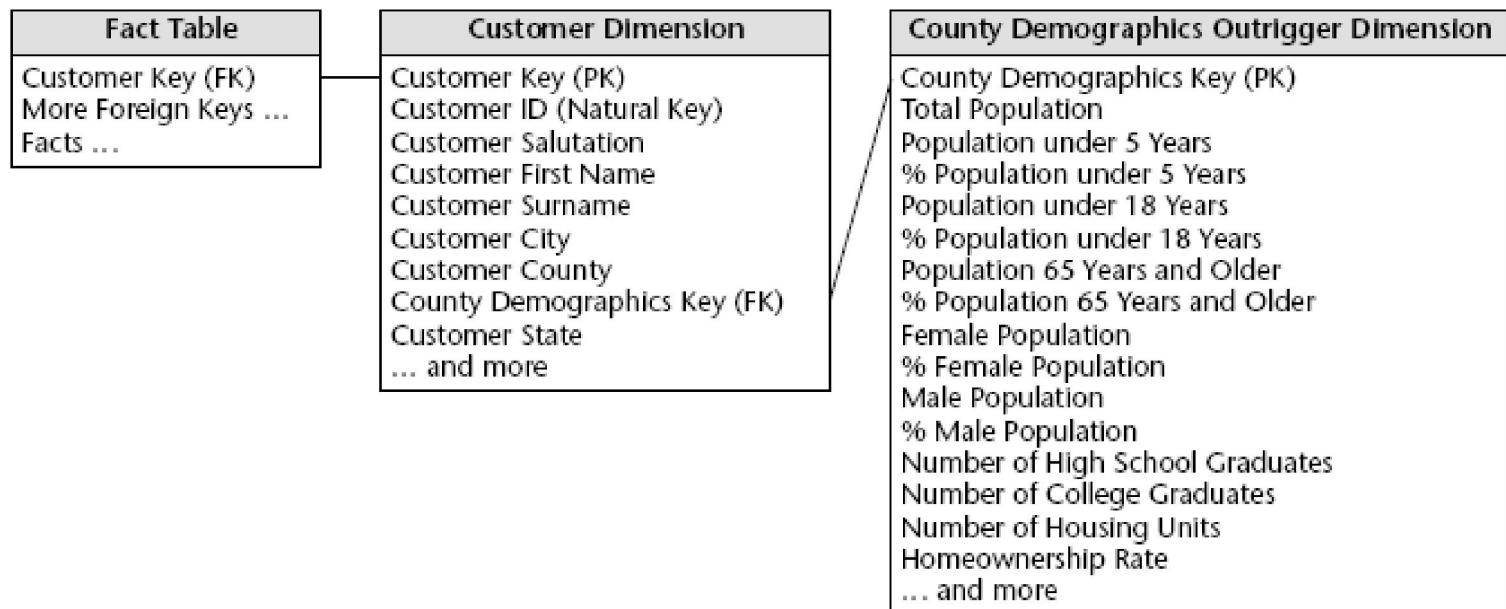
Customer Dimension	Fact Table
Customer Key (PK) Customer ID (Natural Key) Customer Name Customer Address Customer Date of Birth Customer Date of 1st Order ...	Customer Key (FK) Customer Demographics Key (FK) More Foreign Keys ... Facts ...
Customer Demographics Dimension	
Customer Demographics Key (PK) Customer Age Band Customer Gender Customer Income Band Customer Number of Children Band Customer Marital Status	

- Vide também o caso de dimensões com alta volatilidade (minidimensão com atributos que mudam rapidamente);
- Tamanho de cada minidimensão = Produto cartesiano da cardinalidade dos atributos da minidimensão.

Tabelas de Dimensões

Dimensões como “outrigger”

- No exemplo, o “outrigger” agrupa atributos de baixa cardinalidade, que são mantidos em tabela separada da dimensão principal (Customer) para economia de espaço, e também porque a carga dessa tabela é feita com frequência diferente e a partir de fonte externa.
- Note que se a solução fosse ligar o “outrigger” diretamente à tabela de Fatos, seria uma minidimensão. Seria possível?
Vantagens e desvantagens?



Tabelas de Dimensões

Dimensões Multivaloradas (Bridge Table)

- Uma tabela com chave composta capturando um relacionamento muitos-para-muitos que não possa ser acomodado pela granularidade natural de uma tabela de fatos ou tabela de dimensão. Serve como uma ponte entre a tabela de fatos e a tabela de dimensão de forma a permitir dimensões multivaloradas.



- Outros exemplos de dimensões multivaloradas: titulares de conta bancária, códigos de classificações, etc

Tabelas de Dimensões

Dimensões Multivaloradas (Bridge Table)

- Tabela ponte conta-cliente para associar múltiplos clientes com fatos de contas.

Month Account Snapshot Fact
Month End Date Key (FK)
Branch Key (FK)
Product Key (FK)
Account Key (FK)
Account Status Key (FK)
Household Key (FK)
Primary Month Ending Balance
Average Daily Balance
Number of Transactions
Interest Paid
Interest Charged
Fees Charged

Account Dimension
Account Key (PK)
Account Number (Natural Key)
Primary Account Holder Name
Secondary Account Holder Name
Account Address Attributes ...
Account Open Date
Account Type Description
... and more

Account-to-Customer Bridge
Account Key (FK)
Customer Key (FK)
Weighting Factor

Customer Dimension
Customer Key (PK)
Customer Name
Customer Date of Birth
... and more



Modelagem Dimensional e DW

Contexto de Análise

é a associação entre um conjunto de métricas e um conjunto de dimensões



Business School Brasilia
O futuro é agora. Construí-lo.



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Modelagem Dimensional e DW



*O nome do contexto de análise revela
a sua granularidade*



Modelagem Dimensional e DW

Fatos Conformados

- Estabelecer dimensões conformadas para amarrar os data marts representa 90% do esforço de arquitetura de projeto. O restante do esforço consiste em estabelecer definições de fatos conformados.
- Preços, custos, lucros, medidas de qualidade, medidas de satisfação do cliente e outros KPIs são fatos que devem ser conformados. Em geral, dados de fatos não são duplicados explicitamente em múltiplos data marts. Mas isso pode ocorrer em data marts de primeiro nível (originários de um sistema fonte primário de dados) e data marts consolidados (a partir de múltiplas fontes que podem referenciar mais de um processo de negócio).



Modelagem Dimensional e DW

Fatos Conformados

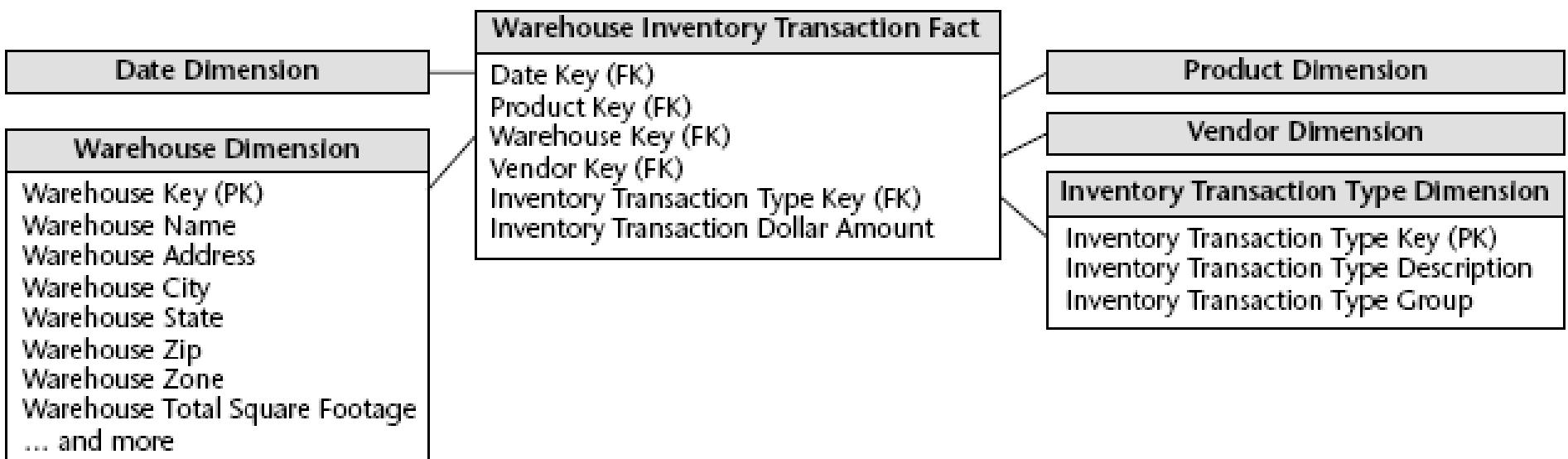
- Se os fatos forem rotulados identicamente, precisam ser definidos no mesmo contexto dimensional e com as mesmas unidades de medida de *data mart* para *data mart*;
- Algumas vezes, um fato tem uma unidade de medida natural em uma tabela de fatos e outra unidade de medida em outra tabela de fatos. Ao invés de prover um fator de conversão numa tabela de dimensão, a abordagem correta é levar o fato com as duas unidades de medida para facilitar os relatórios sem preocupação de conversão. Por exemplo, produtos medidos em caixas no depósito e em peças na loja.



Modelagem Dimensional e DW

Fatos de Transações

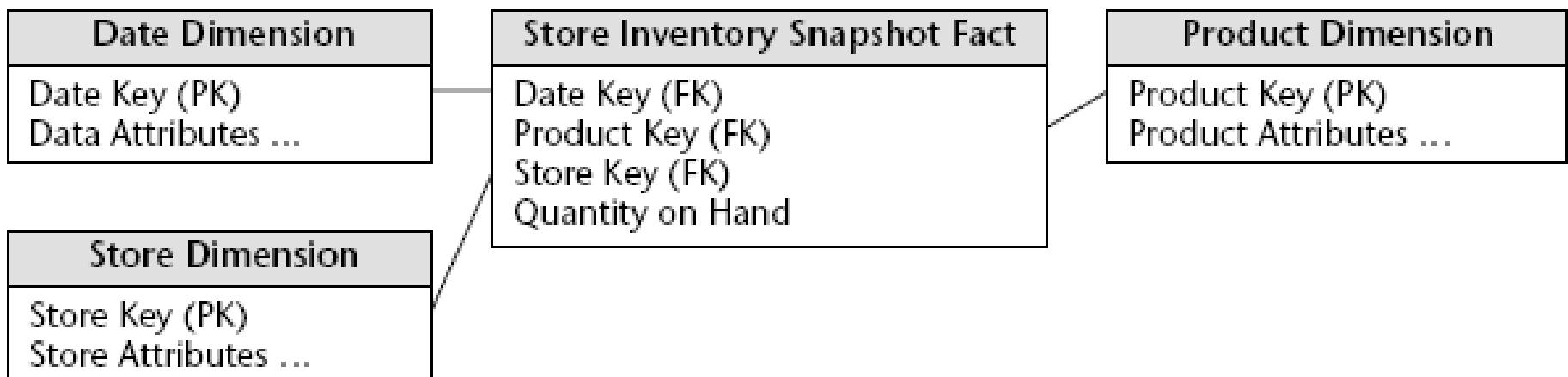
- O nível de transação individual representa a visão mais fundamental das operações do negócio. Essas tabelas de fatos representam um evento que ocorreu num ponto instantâneo do tempo.



Modelagem Dimensional e DW

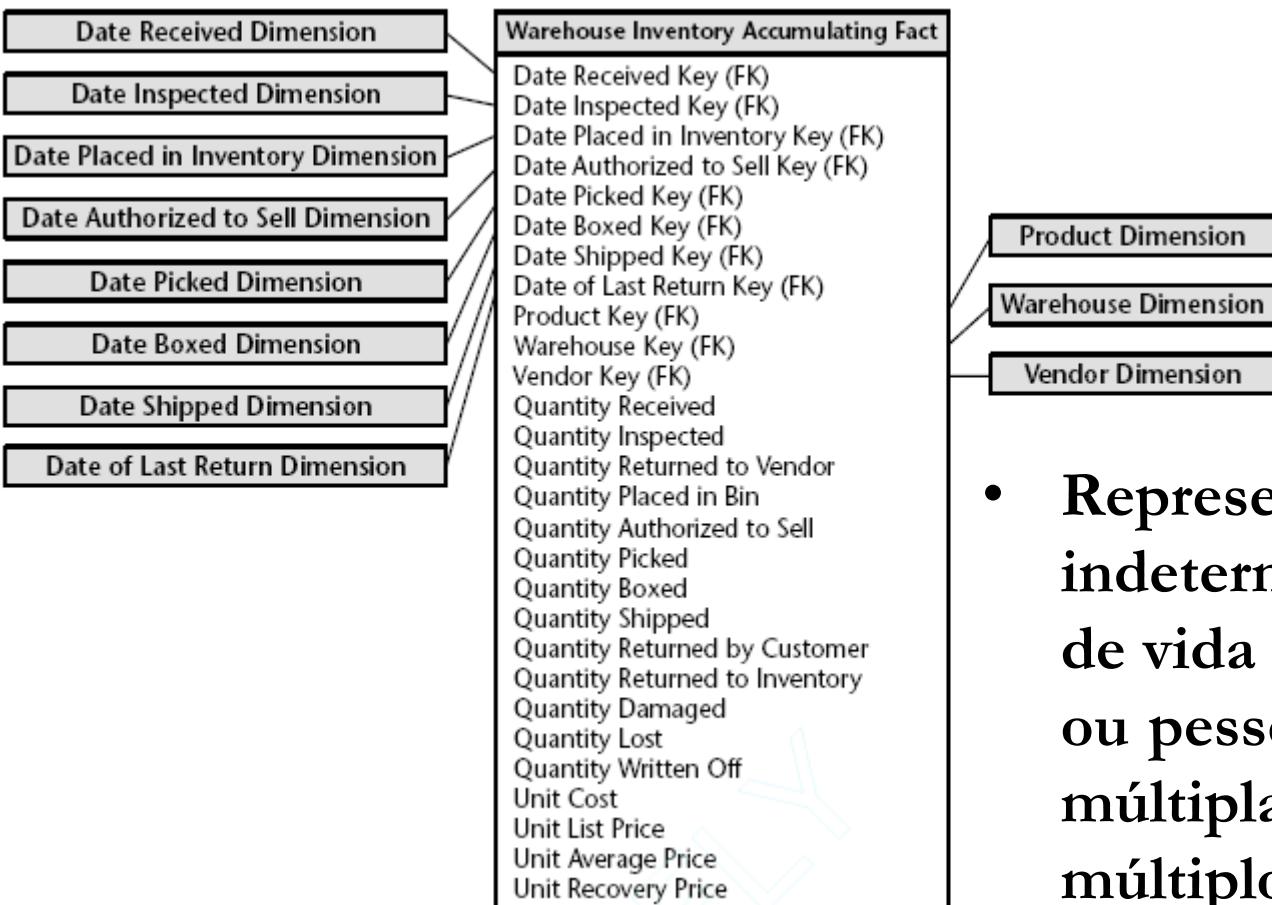
Fatos Instantâneos Periódicos

- São necessários para observar o desempenho cumulativo do negócio em intervalos de tempo regulares e previsíveis. Diferentemente do fato de transação, onde se carrega uma linha para cada ocorrência de evento, com o instantâneo periódico, tira-se uma fotografia da atividade no fim de um dia, uma semana ou um mês, e repetidamente ao fim de cada período.



Modelagem Dimensional e DW

Fatos Instantâneos Acumulados

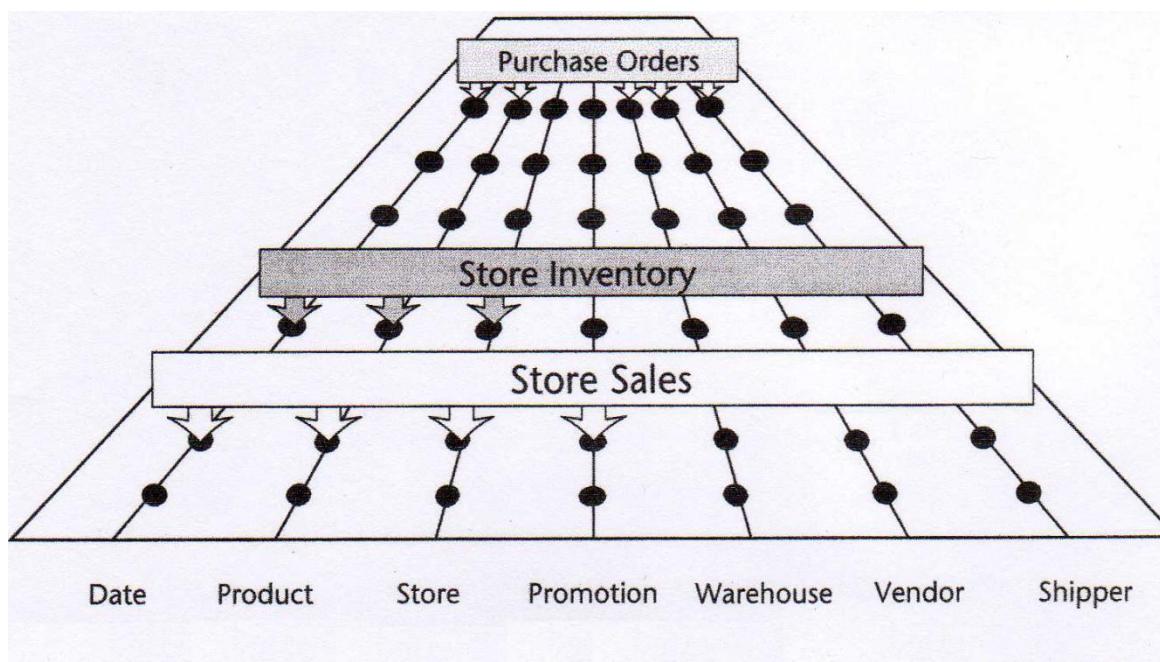


- Representam um tempo indeterminado, que cobre o ciclo de vida da transação ou do produto ou pessoa. Quase sempre possuem múltiplas datas, representando os múltiplos eventos ou fases que ocorrem durante o curso de um ciclo de vida.

Modelagem Dimensional e DW

Data Warehouse Bus Architecture

- Definindo um barramento padrão para o ambiente de DW, data marts separados podem ser implementados por grupos diferentes em tempos diferentes. Todos os processos da cadeia de valores da organização criarão uma família de modelos dimensionais que compartilham um conjunto completo de dimensões comuns e conformadas.



Modelagem Dimensional e DW

Data Warehouse Bus Matrix

- As linhas da matriz correspondem a data marts e as colunas a dimensões conformadas. A matriz é a ferramenta usada para criar, documentar, gerenciar e comunicar a arquitetura de barramento. Segundo Kimball, é o artefato de análise mais importante do desenvolvimento de um DW. É uma ferramenta híbrida, que serve para design técnico, para gerência de projeto e como forma de

BUSINESS PROCESSES	COMMON DIMENSIONS							
	Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
Store Sales	X	X	X	X				
Store Inventory	X	X	X					
Store Deliveries	X	X	X					
Warehouse Inventory	X	X			X	X		
Warehouse Deliveries	X	X			X	X		
Purchase Orders	X	X			X	X	X	X

Modelagem Dimensional e DW

Bus Matriz of Implementation

Business Process	Fact Table	Granularity	Facts	Date	Policyholder	Coverage	Covered Item	Employee	Policy	Claim	Claimant	3rd Party
Policy Transactions	Corporate Policy Transactions	1 row for every policy transaction	Policy Transaction Amount	X Trxn Bf	X	X	X	X	X			
	Auto Policy Transactions	1 row per auto policy transaction	Policy Transaction Amount	X Trxn Bf	X	X Auto	X Auto	X	X			
	Home Policy Transactions	1 row per home policy transaction	Policy Transaction Amount	X Trxn Bf	X	X Home	X Home	X	X			
Policy Premium Snapshot	Corporate Policy Premiums	1 row for every policy, covered item, and coverage each month	Written Premium Revenue Amount, Earned Premium Revenue Amount	X	X	X	X	X Agent	X			
	Auto Policy Premiums	1 row per auto policy, auto covered item, and auto coverage each month	Written Premium Revenue Amount, Earned Premium Revenue Amount	X	X	X Auto	X Auto	X Agent	X			
	Home Policy Premiums	1 row per home policy, home covered item, and home coverage each month	Written Premium Revenue Amount, Earned Premium Revenue Amount	X	X	X Home	X Home	X Agent	X			
Claims Transactions	Claim Transactions	1 row for every claim transaction	Claim Transaction Amount	X Trxn Bf	X	X	X	X	X	X	X	X
	Claim Accumulating Snapshot	1 row per covered item and coverage on a claim	Original Reserve Amount, Assessed Damage Amount, Reserve Adjustment Amount, Current Reserve Amount, Open Reserve Amount, Claim Amount Paid, Payments Received, Salvage Received, Number of Transactions	X	X	X	X	X Agent	X	X	X	
	Accident Event	1 row per loss party and affiliation in an auto claim	Implied Accident Count	X	X	X Auto	X Auto		X	X Auto	X	

Exemplo de Matriz de Barramento em Negócio de Seguradora

Modelagem Dimensional e DW

Mitos sobre Modelagem Dimensional

1. Modelos dimensionais e Data Marts são para dados summarizados somente;
2. Modelos dimensionais e Data Marts são soluções departamentais, não empresariais;
3. Modelos dimensionais e Data Marts não são escaláveis;
4. Modelos dimensionais e Data Marts são apropriados somente quando há um padrão de uso previsível;
5. Modelos dimensionais e Data Marts não podem ser integrados e levam a soluções isoladas.



Business School Brasil
O Melhor Lugar para Crescer



PUC GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Modelagem Dimensional e DW

Quatro Passos da Modelagem Dimensional

1. Selecionar o **processo de negócio** a modelar:

- Um processo é uma atividade de negócio natural da organização que tipicamente é suportada por um sistema fonte de coleções de dados;
- Departamental ou corporativo;
- Os de maior impacto para o usuário primeiro;
- Exemplos: vendas, compras de matéria prima, pedidos, expedições, faturamento, inventário, contas a pagar/receber, etc.



Business School Brasil
O Estudo Língua Língua



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Modelagem Dimensional e DW

Quatro Passos da Modelagem Dimensional

2. Escolher a **granularidade** do processo de negócio:

- Significa especificar exatamente o que uma linha da tabela fato representa;
- Nível mais detalhado das medidas da tabela de fatos;
- Transações individuais ou posições periódicas;
- Exemplos: uma linha de um cupom fiscal, um cartão de embarque individual, um nível diário de estoque de cada produto, um saldo mensal de cada conta bancária, etc.

Modelagem Dimensional e DW

Quatro Passos da Modelagem Dimensional

3. Escolher as **dimensões** que se aplicam a cada linha da tabela de fatos:

- Implica em responder à pergunta: “Como o pessoal do negócio descreve os dados que resultam do processo de negócio?”;
- Exemplos: data, produto, cliente, tipo de transação, status de pedido, etc.



Modelagem Dimensional e DW

Quatro Passos da Modelagem Dimensional

4. Escolher as **medidas** que irão popular cada linha da tabela de fatos:

- Implica em responder à pergunta: “O que nós estamos medindo?” Os fatos candidatos devem ser coerentes com o grão declarado no passo 2;
- Percentuais não devem ser armazenados, e sim seus numeradores e denominadores;
- Exemplos: quantidade, valor, etc.



Business School Brasil
O futuro é agora. Junte-se a!



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Modelagem Dimensional e DW

Quatro Passos da Modelagem Dimensional



Modelagem Dimensional e DW

Dicas sobre Modelagem Dimensional

- Resista à tentação de simplesmente examinar as fontes de dados somente: não há substituto para o input dos usuários do negócio;
- Caso exista, use um modelo de dados convencional E-R como ponto de partida para o trabalho de modelagem dimensional:
 - ❖ Observe os relacionamentos 1:N existentes. Eles podem sugerir dimensões;
 - ❖ Observe as entidades fortes. Elas também podem sugerir dimensões;
 - ❖ Observe as entidades que expressam documentos como Nota Fiscal, Pedido, Ordem de Compra, etc. Elas podem sugerir fatos;
 - ❖ Observe os relacionamentos M:N. Na sua interseção, pode haver valores numéricos. Isto sugere fatos;
 - ❖ Observe os atributos que estarão nas tabelas de dimensões. Analise a relação de hierarquias entre esses atributos de dimensão. Atente para os relacionamentos M:N entre eles. Isto pode definir Granularidade.



Modelagem Dimensional e DW

Dicas sobre Modelagem Dimensional

- As tabelas **FATOS**, tipicamente, armazenam dados, valores atômicos ou agregados obtidos a partir destes;
- As medidas das tabelas **FATOS** são normalmente aditivas em certas dimensões (ou em todas);
- As tabelas **FATOS** possuem chaves que as conectam às diferentes **DIMENSÕES** que as circundam. Essa conexão se dá num nível de granularidade compatível entre elas (**FATO** e **DIMENSÃO**);
- As tabelas **DIMENSÃO** armazenam os valores de filtro, acesso e textos que caracterizam os dados trabalhados;



Modelagem Dimensional e DW

Dicas sobre Modelagem Dimensional

- As tabelas FATOS são normalmente normalizadas (3^a forma normal);
- As tabelas DIMENSÕES são normalmente desnormalizadas (2^a forma normal - Esquema Estrela);
- A granularidade combinada da tabela FATO com a de suas tabelas DIMENSÕES determina o número de linhas das tabelas do projeto.



Modelagem Dimensional e DW

Questões Críticas para Modelagem Dimensional

- **Foco nos requisitos e objetivos do negócio;**
 - ❖ Não na tecnologia e nos dados
- **Envolvimento do patrocinador e usuários gerenciais é essencial para o sucesso;**
- **Adote uma abordagem incremental e iterativa para o desenvolvimento do DW;**
 - ❖ Não tente fazer tudo de uma vez;
- **Desempenho das consultas do usuário e facilidade de uso são os fatores mais críticos;**
 - ❖ Otimização de consultas OLAP;
- **Apresente os dados de forma simples, e com a semântica clara;**
- **Nível de detalhe deve chegar até os dados atômicos;**
- **Esteja preparado para mudanças no negócio e nos dados;**
- **Dê especial atenção à aceitação dos usuários.**



Modelagem Dimensional e DW

10 Armadilhas a evitar em projetos DW

1. Negligenciar o reconhecimento de que o sucesso do DW está amarrado à aceitação do usuário;
2. Presumir que o negócio, seus requisitos e análises, assim como os dados subjacentes e a tecnologia, são estáticos;
3. Carregar somente dados sumarizados nas estruturas dimensionais da área de apresentação;
4. Popular modelos dimensionais de forma isolada, sem levar em conta a arquitetura que os amarra juntos usando dimensões compartilhadas e conformadas;



PUC
GOIÁS



Modelagem Dimensional e DW

10 Armadilhas a evitar em projetos DW

5. Tornar os dados supostamente consultáveis na área de apresentação desnecessariamente complexos;
6. Prestar mais atenção no desempenho operacional e na facilidade de desenvolvimento do “back-room” do que no desempenho de consultas e facilidade de uso do “front-room”;
7. Alocar energia para construir uma estrutura de dados normalizada, mesmo estourando o orçamento, do que para construir um área de apresentação viável baseada no modelo dimensional;



Modelagem Dimensional e DW

10 Armadilhas a evitar em projetos DW

8. Atacar um projeto galático plurianual ao invés de perseguir esforços de desenvolvimento mais gerenciáveis, porém ainda desafiadores e iterativos;
9. Falhar em identificar e adotar uma gerência influente, acessível e razoavelmente visionária como patrocinador do negócio;
10. Tornar-se enamorado da tecnologia e dos dados ao invés de focar nos requisitos e objetivos do negócio.



Modelagem Dimensional e DW

Projeto de Data Warehouse

Requisito de Dados

Projeto Conceitual

Modelagem dos requisitos de dados através de diagramas de Entidades e Relacionamentos (DER) ou de Classes e Objetos (DCO)

Esquema Conceitual

Projeto Lógico

Mapeamento do esquema conceitual para o modelo de dados do SGBD escolhido, através de diagrama de estruturas de dados (DED)

Esquema Lógico

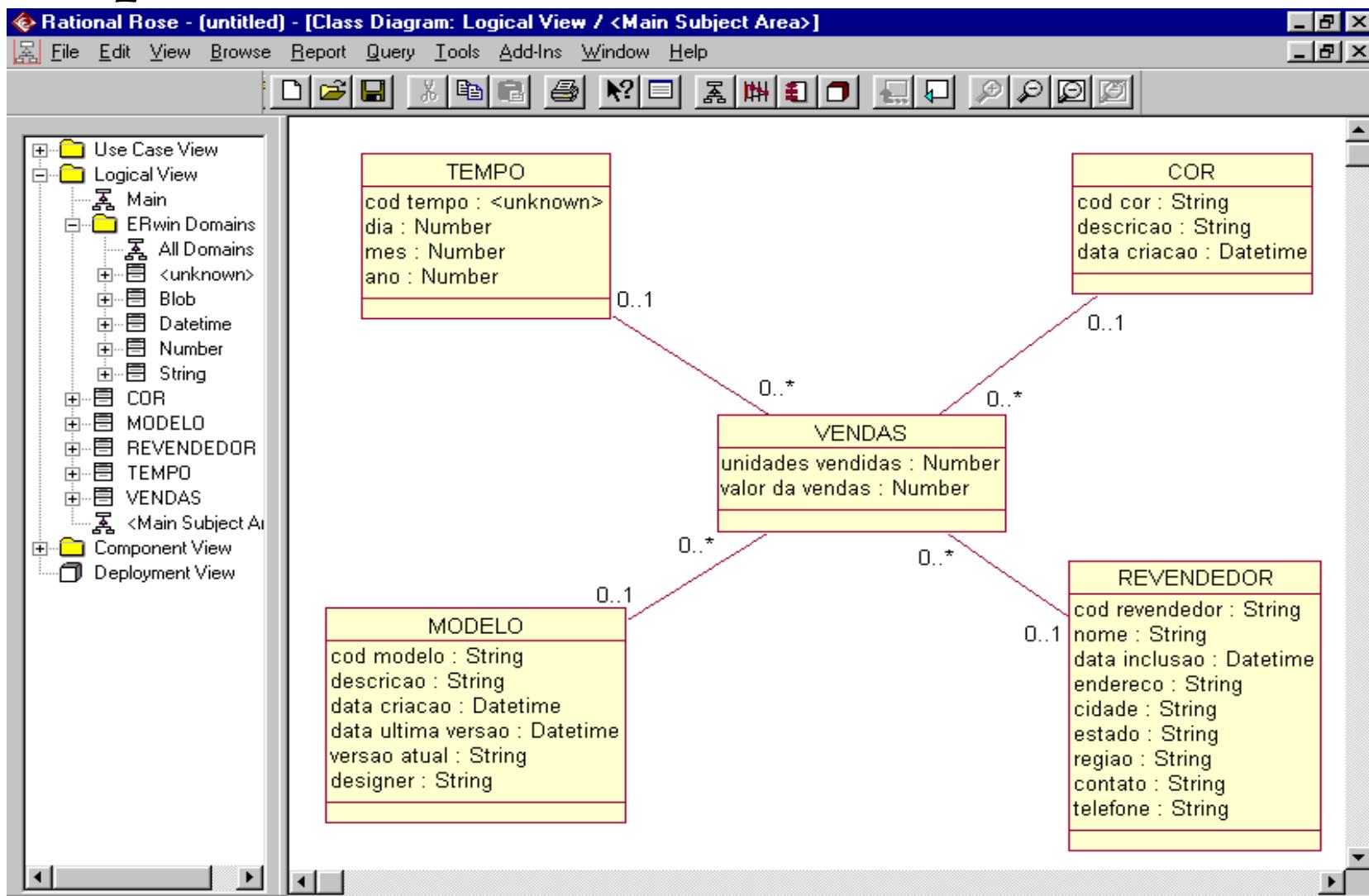
Projeto Físico

Mapeamento do esquema lógico para os tipos de dados e restrições de integridade do SGBD escolhido; criação de visões e índices.

Esquema Físico

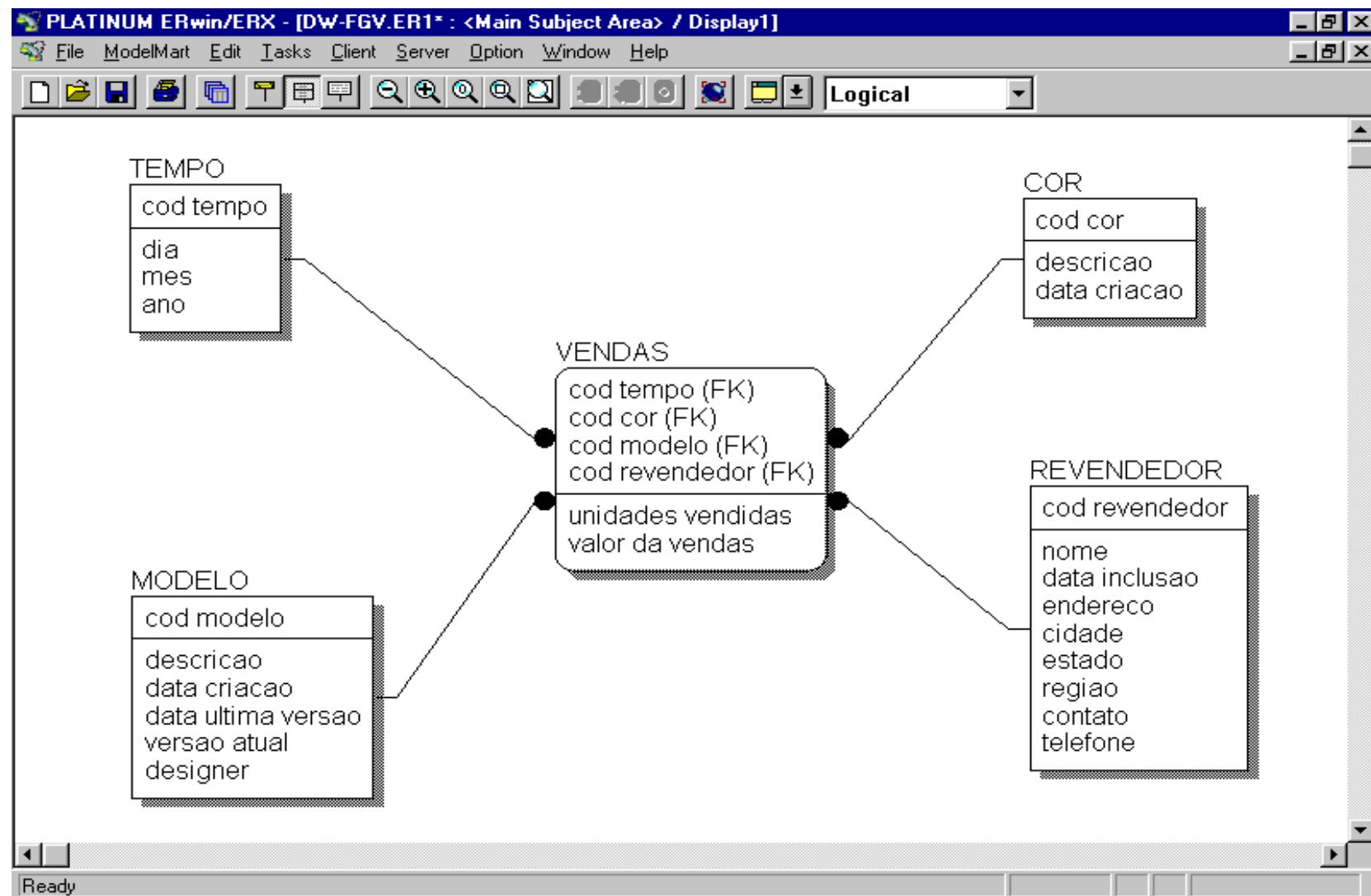
Modelagem Dimensional e DW

Esquema Estrela: Modelo Conceitual



Modelagem Dimensional e DW

Esquema Estrela: Modelo Lógico



Business School Brasil
O futuro começa aqui!



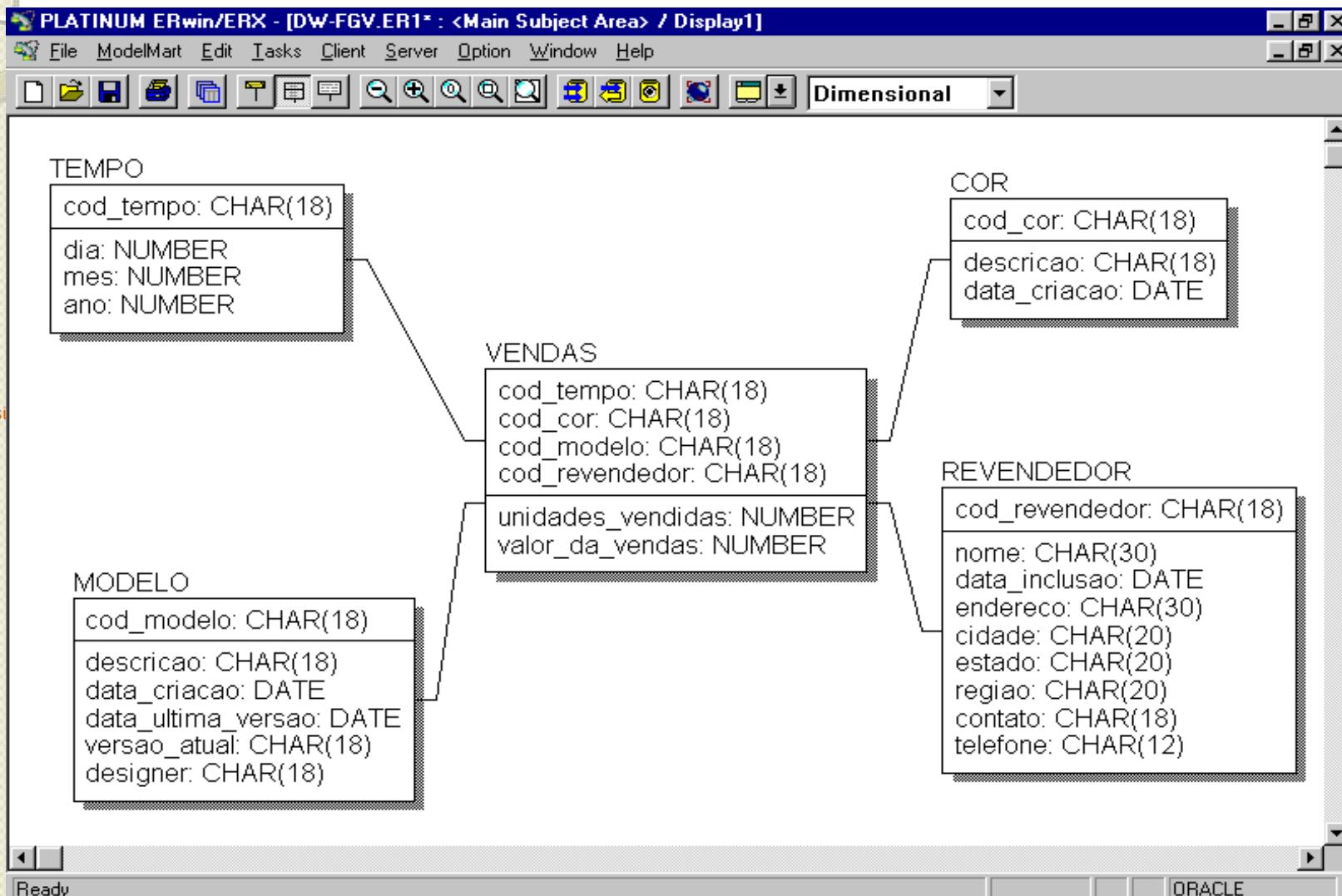
PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Modelagem Dimensional e DW

Esquema Estrela: Modelo Físico



Modelagem Dimensional e DW

Implementação do Modelo Dimensional

- **SGBDs multidimensionais:**
 - ❖ Implementam fisicamente o modelo dimensional;
 - ❖ Problemas de desempenho, segurança e confiabilidade;
 - ❖ Problema de esparsidade: células onde não há dados (nulos).
- **SGBDs relacionais:**
 - ❖ Maior aceitação (força do mercado de BD relacional);
 - ❖ Exige mapeamento (como qualquer projeto de BD relacional).



Modelagem Dimensional e DW

10 Erros comuns em Projetos de DW

1. Aceitar a premissa de que os responsáveis pelos sistemas fontes mais relevantes da organização são muito importantes e ocupados para gastar tempo com a equipe de DW;
2. Após a equipe de DW ter sido acordada, marcar uma reunião para discutir comunicações com os usuários de negócio, se o orçamento permitir;
3. Assegurar para o pessoal de suporte do DW escritórios agradáveis no prédio da TI, que fica próximo dos usuários de negócio, e providenciar um número de telefone de suporte de DW com várias opções de menu;



Modelagem Dimensional e DW

10 Erros comuns em Projetos de DW

4. Treinar cada usuário em cada característica da ferramenta de acesso a dados na primeira aula de treinamento, adiar o treinamento sobre conteúdo de dados porque a aula usa dados falsos (os dados reais não estarão prontos nos próximos dois meses) e declarar sucesso ao término da primeira aula de treinamento já que o DW foi disponibilizado para os usuários de negócio;
5. Assumir que os usuários de negócio vão naturalmente gravitar em direção a dados robustos e desenvolver suas próprias *killer applications* analíticas;



Modelagem Dimensional e DW

10 Erros comuns em Projetos de DW

6. Antes de implementar o DW, fazer uma análise completa descrevendo todos os possíveis ativos de dados da empresa e todos os usos desejados de informação, e evitar a ilusão sedutora de desenvolvimento iterativo, que é somente uma desculpa para não fazer certo da primeira vez;
7. Não aborrecer os executivos *seniors* de sua organização com o DW até que você tenha o implementado e possa apontar para um sucesso significativo;



Modelagem Dimensional e DW

10 Erros comuns em Projetos de DW

8. Encorajar os usuários de negócio a lhe dar feedback contínuo ao longo do ciclo de desenvolvimento sobre novas fontes de dados e métricas chaves de desempenho que eles gostariam de acessar, e assegurar a inclusão desses requisitos na release em desenvolvimento;
9. Concordar em entregar um *data mart* centrado em cliente de alto perfil, idealmente lucratividade de cliente ou satisfação de cliente, como seu primeiro produto;
10. Não conversar com os usuários de negócio; ao invés disso, confiar em consultores ou especialistas internos para lhe dar interpretação dos requisitos de usuários do DW.



Modelagem Dimensional e DW

Conclusão

- Modelagem dimensional é a abordagem utilizada para o projeto de DWs:
 - ❖ Visão dimensional dos dados
 - ▶ Intuitiva, flexível e eficiente para consultas;
 - ▶ Fatos, dimensões e hierarquias.
 - ❖ Esquemas
 - ▶ Estrela, flocos de neve e constelação de fatos
 - ❖ Fases do projeto
 - ▶ Escolha do: Processo de negócio, granularidade, dimensões e medidas de fatos (nesta ordem).



Business School Brasil
O futuro começa aqui



PUC
GOIÁS



UNIVERSIDADE ESTADUAL
VALE DO ACARAÚ

Dúvidas e Esclarecimentos



Ricardo Holanda
rhnobre@gmail.com.br

