



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

MODELAGEM DIMENSIONAL E INDICADORES DE DESEMPENHO/DASHBOARDS

Prof. Marcondes Alexandre
marcondesalexandre@gmail.com



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Conceitos Básicos

Processamento Analítico de Dados

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Data Warehousing

Engloba **arquiteturas**, **algoritmos** e **ferramentas** que possibilitam que dados selecionados de **provedores de informação** autônomos, heterogêneos e distribuídos sejam **integrados** em uma única base de dados, conhecida como **data warehouse (DW)**

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Acesso às Informações

- Duas etapas
 - a informação de cada provedor é extraída previamente, devendo ser traduzida, filtrada, integrada à informação relevante de outros provedores e finalmente armazenada no **DW**
 - as consultas, quando realizadas, são executadas diretamente no **DW**, sem acessar os provedores de informação originais

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Exemplos de Análises

- Análises de tendências simples
 - *Quais as vendas mensais de um certo produto no ano de 1998?*
- Análises comparativas
 - *Quais as vendas mensais dos produtos de uma dada marca nos últimos 3 anos?*
- Análises de tendência múltiplas
 - *Quais as vendas mensais dos produtos de uma dada marca nos últimos 3 anos, de acordo com as promoções de Natal?*

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Vantagens

- Análises podem ser realizadas mais rapidamente
 - DW armazena informações integradas, cujas diferenças semânticas e de modelo já foram eliminadas
- Existe maior disponibilidade dos dados
 - consultas são executadas diretamente no DW sem acessar os provedores de informação originais

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Vantagens

- Garante a autonomia dos provedores de informação originais
 - processamento local nos provedores de informação originais não é afetado por causa da participação destes no ambiente de data warehousing
- ...

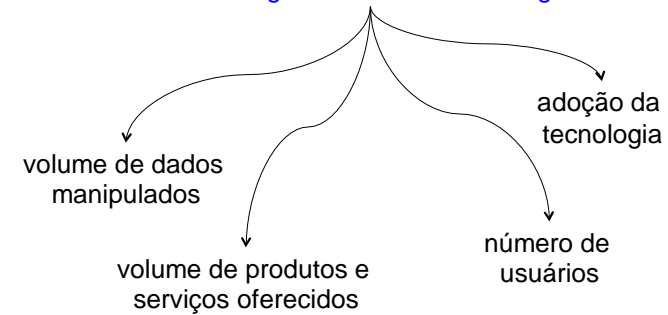
OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Visão do Mercado

Crescimento explosivo do uso da tecnologia de data warehousing



OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Pensamento Motivacional

A obtenção de **informações estratégicas**, relativas ao contexto de **tomada de decisão**, é de suma importância para o sucesso de uma empresa. Tais informações permitem à empresa um **planejamento rápido** frente às mudanças nas condições do negócio, essencial na atual conjuntura de um mercado globalizado.

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Principal Característica	voltado ao processamento de transações OLTP	voltado ao processamento de consultas OLAP
Tipos de Operação mais Frequentes	atualização remoção inserção	leitura (consulta)

o termo OLAP (on-line analytical processing) foi introduzido em 1993 por Codd *et al.* para definir a categoria de processamento analítico sobre um banco de dados histórico voltado para os processos de gerência e tomada de decisão

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Volume de Transações	relativamente alto	relativamente baixo
Características das Transações	pequenas e simples, acessam poucos registros por vez	longas e complexas, acessam muitos registros por vez e realizam várias varreduras e junções de tabelas

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Tipos de Usuários	administradores do sistema, projetistas, usuários de entrada de dados	usuários de SSD por exemplo: executivos, analistas, gerentes, administradores
Número de Usuários Concorrentes	grande (geralmente milhares)	relativamente pequeno (geralmente centenas)
Interações com os Usuários	pré-determinadas estáticas	<i>ad-hoc</i> dinâmicas

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Volume de Dados	<i>megabytes a gigabytes</i>	<i>gigabytes a terabytes</i>
Projeto do Banco de Dados	normalizado para suporte às propriedades ACID	multidimensional, refletindo as necessidades de análise dos usuários de SSD
Granularidade dos Dados	detalhado	detalhado e agregado

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Principal Questão de Desempenho	produtividade da transação	produtividade da consulta
Tempo de Resposta	geralmente poucos segundos	de minutos a horas
Exemplos de aplicações	transações bancárias, empréstimos de livros, contas a pagar	planejamento de <i>marketing</i> , análise financeira

OLAP: Conceitos Básicos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Arquitetura de um Ambiente de *Data Warehousing*

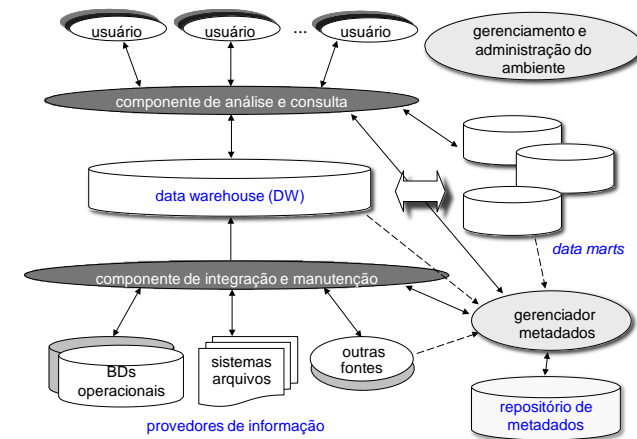


OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Arquitetura Típica



OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente: DW

- Coração do ambiente de data warehousing
- Banco de dados
 - voltado para o suporte aos processos de gerência e tomada de decisão
 - tem como principais objetivos prover eficiência e flexibilidade na obtenção de informações estratégicas e manter os dados sobre o negócio com alta qualidade

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Características dos Dados

- Orientados a assunto
 - relativos aos temas de negócio de maior interesse da corporação
 - *exemplos*: clientes, produtos, promoções, contas e vendas
- Integrados
 - dados obtidos dos provedores de informação corrigidos para eliminar possíveis inconsistências

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Características dos Dados

- Não-voláteis
 - o conteúdo do DW permanece estável por longos períodos de tempo
- Históricos
 - relevantes a algum período de tempo
 - *exemplo*: usualmente dados relativos a um grande espectro de tempo (5 a 10 anos) encontram-se disponíveis

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Características dos Dados

- Organizados em diferentes níveis de agregação
 - nível inferior: dados primitivos coletados do ambiente operacional
 - níveis intermediários: dados com graus de agregação crescente
 - nível superior: dados altamente resumidos (agregados)

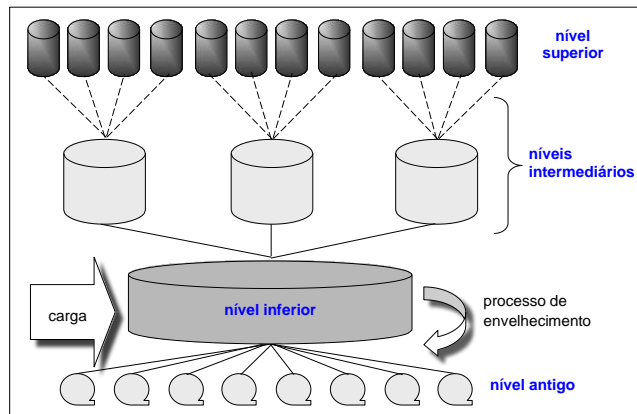
devido ao volume de dados armazenados no DW, esses dados podem ser transferidos periodicamente para o nível antigo

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Níveis de Agregação

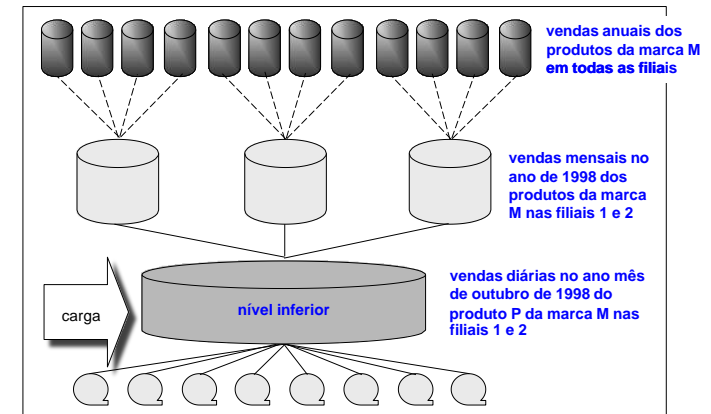


OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Níveis de Agregação



OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Granularidade

- Grau de detalhamento em que os dados são armazenados em um nível
- Questão de projeto muito importante
 - impacta no volume de dados armazenado
 - afeta as consultas que podem ser respondidas

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Granularidade

grão muito pequeno



tamanho do data
warehouse é muito
grande

+

praticamente qualquer
consulta pode ser
respondida

grão muito grande



tamanho do data
warehouse é menor

+

número de consultas que
podem ser respondidas
é menor

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente: Provedores de Informação

- Fontes de dados
 - autônomas
 - heterogêneas
 - distribuídas
- Contêm **dados operacionais**
- Exemplos
 - SGBD relacionais, objeto-relacionais, ...
 - documentos HTML, SGML, ...

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente de Integração e Manutenção

- Carregamento dos dados
 - atividade mais complexa, cara e demorada
 - essencial ao bom funcionamento do ambiente de data warehousing
 - processos
 - extração – integração
 - tradução – armazenamento
 - limpeza – recuperação de falhas

fluxo de informação: provedores de informação → DW

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Carregamento dos Dados

- Extração
 - **quais** dados são extraídos de quais provedores
 - **como** esses dados são extraídos
- Tradução
 - **conversão** dos dados do formato nativo dos provedores de informação para o formato utilizado pelo ambiente de data warehousing
 - manutenção da **temporalidade** dos dados

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Carregamento dos Dados

- Limpeza
 - garante a **corretude** e a **qualidade** dos dados, de forma que esses dados atendam às restrições de integridade impostas pelas regras de negócio
- Integração
 - geração de um **dado único** a partir de várias cópias do mesmo dado extraídas de diferentes provedores

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Integração dos Dados

- Problema: dados armazenados nos provedores
 - são heterogêneos
 - seguem diferentes modelos de dados
 - são representados por conceitos diferentes
 - possuem diferentes formatos
 - etc
 - são redundantes, inconsistentes e até mesmo complementares
- Dois níveis: **esquema** e **instância**

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Integração: Nível de Esquema

- Conflitos de nome
 - refere-se aos nomes que representam os diferentes elementos a serem integrados
 - problema dos sinônimos: diferentes nomes são aplicados ao mesmo elemento
 - exemplo: **cliente** representa, em um esquema, todos os clientes atendidos por uma loja, enquanto que **comprador** é usado em outro esquema para representar a mesma situação
 - problema dos homônimos: mesmo nome é aplicado a diferentes elementos

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Integração: Nível de Esquema

- Conflitos semânticos
 - surgem quando o mesmo elemento é modelado em diferentes esquemas, porém representando conjuntos que se sobrepõem
 - exemplo: **produto** representa, em um esquema, todos os produtos de um supermercado, enquanto que **produto** é usado em outro esquema para representar apenas os produtos da seção de cosméticos

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Integração: Nível de Esquema

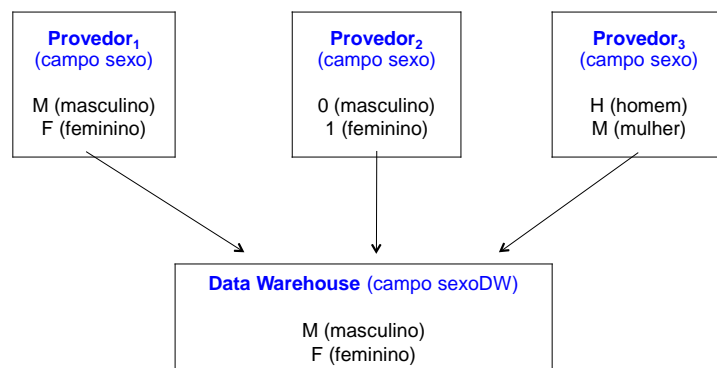
- Conflitos estruturais
 - surgem sempre que diferentes construtores estruturais são utilizados para modelar o mesmo conceito representado em diferentes aplicações
 - exemplo: considerando-se o modelo entidade-relacionamento, o mesmo conjunto de objetos do mundo real pode ser representado como um **tipo-entidade** em um esquema e como um **atributo** de um tipo-entidade em outro esquema

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Integração: Nível de Instância



OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Carregamento dos Dados

- Armazenamento
 - realização de **processamentos adicionais**, como verificação de restrições de integridade, geração de agregações, construção de índices, etc
- Recuperação de Falhas
 - **evita** que tanto leituras desnecessárias aos dados dos provedores de informação quanto computações cujos resultados já foram armazenados no DW **sejam realizadas novamente**

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente de Integração e Manutenção

- Atualização dos dados
 - periodicidade
 - necessidades dos usuários de SSD
 - nível de consistência desejado
 - manutenção dos dados
 - **recomputação**: conteúdo do DW é descartado e os dados são carregados novamente a partir dos provedores de informação operacionais
 - **atualização incremental**: apenas as alterações nos dados dos provedores são refletidas no DW

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente de Integração e Manutenção

- Expiração dos dados
 - remoção de dados do DW visando diminuir o volume de dados armazenado
 - pode ocorrer quando
 - dados atingem o limite de tempo no qual tornam-se inválidos
 - dados não são mais relevantes ou necessários ao ambiente de data warehousing
 - espaço de armazenamento é insuficiente

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente de Análise e Consulta

- Permite a interação do usuário com o ambiente de data warehousing por meio de **ferramentas** dedicadas à análise e consulta dos dados
- Ferramentas
 - oferecem facilidades de navegação e de visualização
 - possuem diferentes classificações, com base nas funcionalidades oferecidas

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ferramentas

- De consulta gerenciáveis e geradores de relatório
 - tipos mais simples de ferramentas
 - têm como objetivo produzir relatórios periódicos
 - permitem que os usuários realizem consultas independentemente da estrutura do banco de dados e/ou da linguagem de consulta

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ferramentas

- Para sistemas de informações executivas
 - oferecem visualização gráfica simplificada, por exemplo representando exceções a atividades normais de negócio ou a regras por meio de diferentes cores
 - oferecem capacidades analíticas limitadas

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ferramentas

- OLAP
 - oferecem capacidades analíticas sofisticadas, permitindo que os dados sejam analisados usando visões multidimensionais complexas e elaboradas
 - oferecem navegação facilitada nessas visões
 - exemplo: usuários de SSD podem analisar os dados sob diferentes perspectivas e determinar tendências por meio da navegação entre diferentes níveis de hierarquias de agregação

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Ferramentas

- De mineração de dados
 - permitem que informações, padrões e tendências de negócio “escondidas” nos dados sejam descobertas

IMPORTANTE: Independentemente da ferramenta utilizada, um fator primordial refere-se à **visualização dos resultados obtidos**. Técnicas de visualização dos dados devem determinar a melhor forma de se exibir relacionamentos e padrões complexos em um monitor bidimensional, de modo que o problema inteiro e/ou a solução sejam claramente visíveis usuários de SSD

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente: Data Mart

- DW que possui escopo limitado
- Armazena dados que compartilham as mesmas características dos dados do DW
- Enfoques
 - subconjunto dos dados do DW
 - política no projeto de construção de um DW corporativo

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Componente: Repositório de Metadados

- Dados de nível mais alto que descrevem dados de nível mais baixo
- Características
 - permite que os usuários de SSD conheçam a **estrutura** e o **significado** dos dados
 - representa o principal recurso para a administração dos dados no ambiente de data warehousing

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Exemplos de Metadados

Metadados Administrativos	contêm informações relacionadas à construção e à utilização do data warehousing, tais como os esquemas dos provedores de informação e do DW, além dos mapeamentos existentes entre os diversos esquemas; regras de extração, de tradução, de limpeza e de atualização dos dados, em adição às regras de mapeamento utilizadas para a solução de problemas de heterogeneidade existentes entre os dados dos diversos provedores de informação que participam do ambiente; especificações sobre grupos de usuários e privilégios a eles associados, incluindo políticas de controle de acesso, autorização e perfis; ferramentas de integração e manutenção, e regras associadas aos processos envolvidos; ferramentas de análise e consulta; consultas, agregações e relatórios pré-definidos
----------------------------------	--

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Exemplos de Metadados

Metadados Específicos da Aplicação	incluem um conjunto de terminologias específicas ao domínio da aplicação, além de restrições da aplicação e outras políticas
Metadados de Auditoria	mantêm informações relacionadas à linhagem dos dados, à geração de relatórios de erros, às ferramentas de auditoria empregadas e às estatísticas de utilização do ambiente de data warehousing, incluindo dados sobre a frequência das consultas, os custos para se processar uma determinada consulta, o tipo de acesso aos dados e o desempenho do sistema

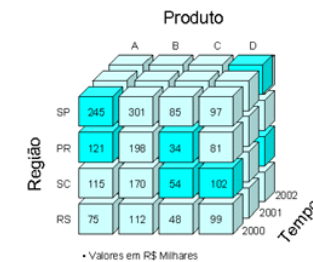
classificação baseada em Wu, M.-C., Buchmann, A.P. Research Issues in Data Warehousing. In *Proceedings of The German Database Conference*, pages 61-82, Ulm, Germany, March 1997.

OLAP: Arquitetura



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Modelagem Multidimensional



OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Modelagem Multidimensional

- Análises dos usuários de SSD
 - representam requisições multidimensionais aos dados do DW
 - permitem a identificação de problemas e de tendências
- Principais enfoques
 - modelo de dados multidimensional
 - abordagens para a representação lógica

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Modelo de Dados Multidimensional

aspectos **estáticos**



modelagem dos dados



dimensões (atributos)
medidas numéricas

aspectos **dinâmicos**



operações analíticas



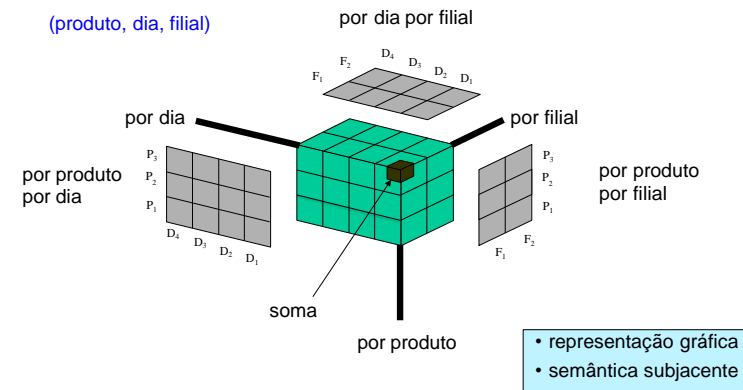
drill-down/roll-up
slice and dice
pivot
drill-across ...

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Cubo de Dados Multidimensional



OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Dimensão

- Representa uma perspectiva de análise dos usuários de SSD
- Composta por atributos
- *Exemplo:* dimensão **filial**
 - *atributos:* cidade, estado, região, país
 - *semântica:* a filial “Primeira Filial” está localizada na cidade de “São Carlos”, estado de “São Paulo”, região “Sudeste” do país “Brasil”

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Hierarquia de Atributos

- Definição
 - permite que atributos de uma dimensão relacionem-se com outros atributos da mesma dimensão
 - especifica níveis de agregação e, portanto, granularidade dos itens de dados
- *Exemplo*: dimensão **filial**
 - cidade → estado → região → país
 - *hierarquia de nível quatro na dimensão filial*

Definição formal: **grafo de derivação**

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Medida Numérica

- Objeto de análise relevante ao negócio
- Definida como uma função de suas dimensões correspondentes

Classificação	Definição	Exemplo
aditiva	somada através de todas as suas dimensões	unidades-vendidas
semi-aditivas	somadas somente através de algumas de suas dimensões	número-clientes
não aditivas	não podem ser somadas	preço

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Operações Analíticas

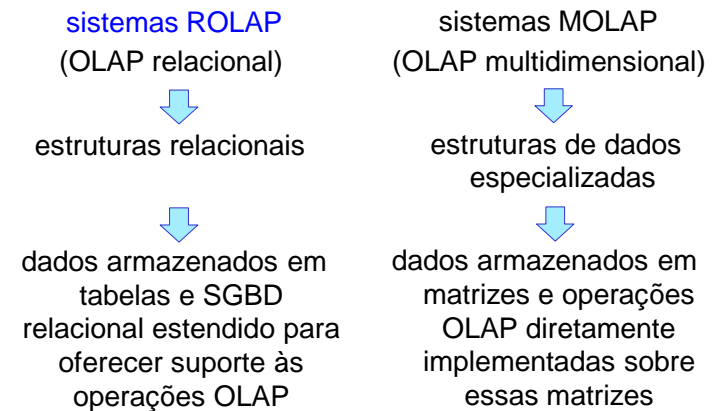
Operação	Definição
<i>drill-down</i>	analisa os dados em níveis de agregação progressivamente mais detalhados, ou de menor granularidade
<i>roll-up</i>	analisa os dados em níveis de agregação progressivamente menos detalhados, ou de maior granularidade
<i>slice and dice</i>	restringe os dados sendo analisados a um subconjunto destes dados <i>slice</i> : corte para um valor fixo <i>dice</i> : seleção de faixas de valores
<i>pivot</i>	reorienta a visão multidimensional dos dados, oferecendo diferentes perspectivas dos mesmos dados
<i>drill-across</i>	compara medidas numéricas distintas que são relacionadas entre si através de pelo menos uma dimensão em comum

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Representação Lógica



OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Sistemas ROLAP: Componentes

- Máquina relacional
 - SGBD adaptado para processamento analítico
 - armazenamento dos dados
 - esquema estrela
 - esquema floco de neve
 - funcionalidades adicionais
 - novos operadores SQL
 - estruturas de indexação especializadas
 - otimização de consultas complexas
 - execução paralela de consultas complexas

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Sistemas ROLAP: Componentes

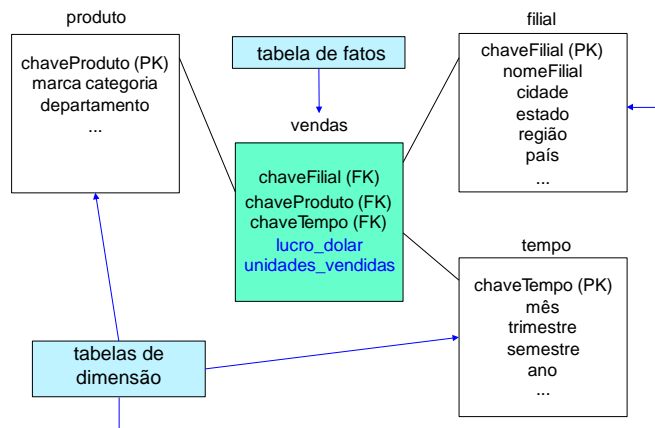
- Máquina ROLAP
 - camada de *software* que complementa a máquina relacional
 - objetivos
 - suprir as limitações da máquina relacional
 - apresentar os dados multidimensionalmente
 - funcionalidade adicional
 - materialização de visões

OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Esquema Estrela

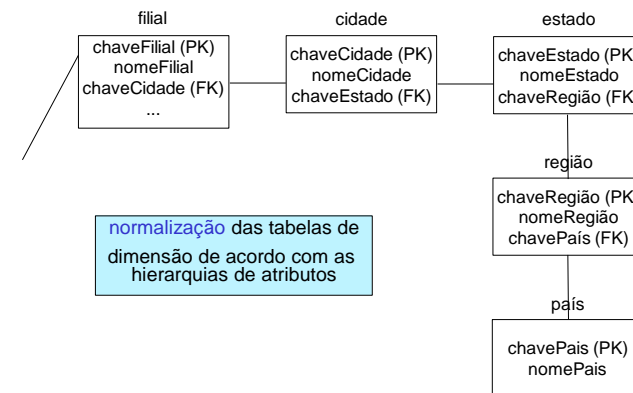


OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Esquema Floco de Neve

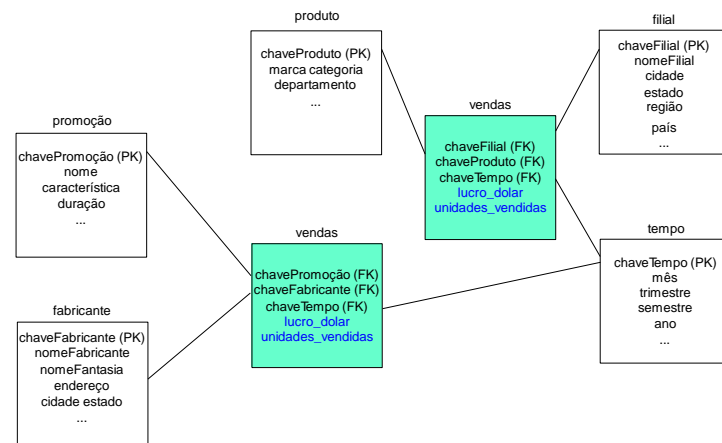


OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Constelação de Fatos



OLAP: Modelagem Multidimensional



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Mitos sobre a Modelagem Dimensional

- Modelos dimensionais e data marts são apenas para dados resumidos
- Modelos dimensionais e data marts são soluções departamentais e não corporativas
- Modelos dimensionais e data marts não são escalonáveis
- Modelos dimensionais e data marts são apropriados apenas quando existe um padrão de utilização previsível
- Modelos dimensionais e os data marts não podem ser integrados e, portanto, levam a soluções isoladas

60



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Processo de Criação Dimensional em 4 Etapas

1. Selecione o processo do negócio
2. Declare o grão do processo do negócio
3. Escolha as dimensões que aplicam a cada linha da tabela fato
4. Identifique os fatos numéricos que preencherão cada linha da tabela de fatos

61



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Estudo de Caso sobre o Varejo

- Empresa do ramo de alimentos
 - 100 supermercados em 5 estados
 - Todas as lojas têm os mesmos depts.
 - 60.000 produtos individuais, as SKU's
 - 55.000 SKU têm código de barras, os UPC's
 - 5.000 SKU são internas à rede (produtos a granel)
- Coleta de dados no POS e na entrada de mercadorias

62



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Modelando esse Caso

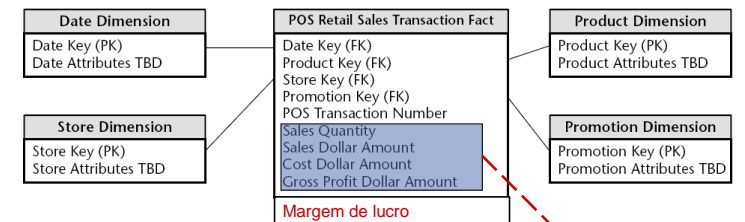
1. Selecione o processo do negócio
 - POS de vendas a varejo
2. Declare o grão
 - Item individual de venda no POS
3. Escolha as dimensões
 - Loja, Produto, Data, Promoção, nro. de transação POS
4. Identifique os fatos
 -

63



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Analizando os Fatos



Fato NAO-Aditivo

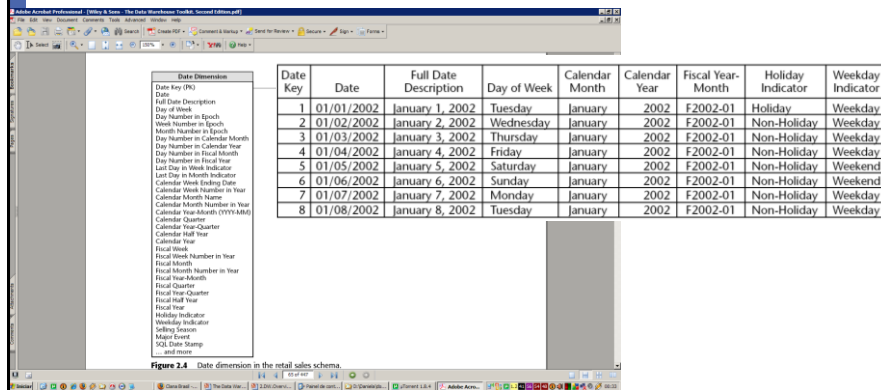
Fatos Aditivos

- Porcentagens e proporções são não-aditivas. É possível calculá-las armazenando numerador e denominador, e depois a proporção das somas

64



Dimensão Data



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

Figure 2.4 Date dimension in the retail sales schema.

65



Dimensão Produto



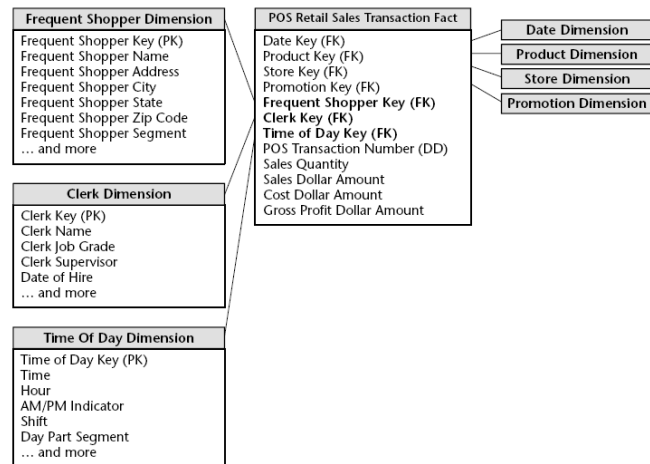
Department Description	Brand Description	Sales Dollar Amount	Sales Quantity
Bakery	Baked Well	\$3,009	1,138
Bakery	Fluffy	\$3,024	1,476
Bakery	Light	\$6,298	2,474
Frozen Foods	Coldpack	\$5,321	2,640
Frozen Foods	Freshlike	\$10,476	5,234
Frozen Foods	Frigid	\$7,328	3,092
Frozen Foods	Icy	\$2,184	1,437
Frozen Foods	QuickFreeze	\$6,467	3,162

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



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Extensão do Modelo



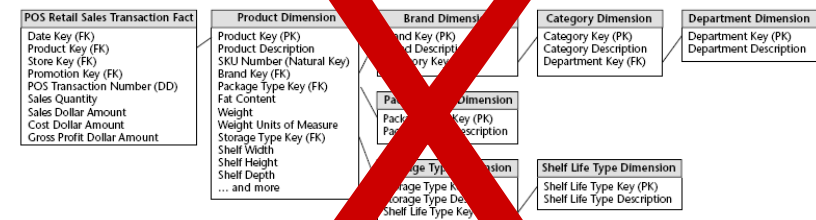
67



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Normalização das Dimensões

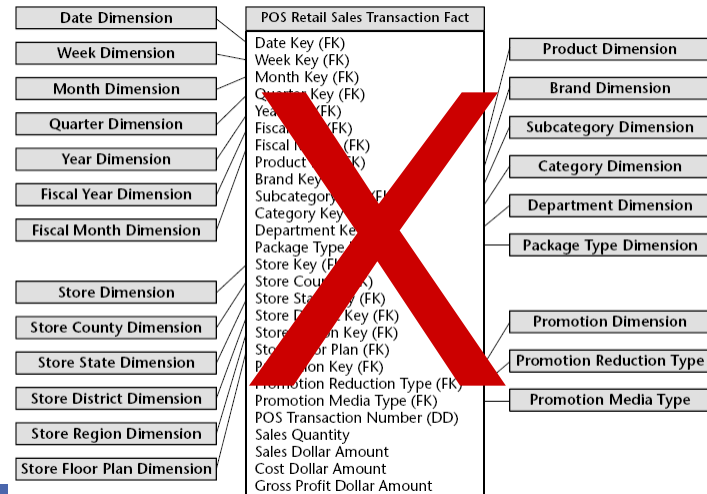
Snowflaking



68



Excesso de Dimensões



69



Outras Considerações

- Dimensão de Degeneração – nro. da transação no POS
- Chaves substitutas
 - Sem “inteligência”
 - Possivelmente menor que chave operacional
- Registram condições para dimensões em que não há valor operacional
- Seu uso evita complicações com as chaves operacionais reutilizadas
- Ao usá-las, evite a composição de chaves unindo chaves de uma dimensão

70



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Exercício de Fixação

O negócio: Rede de Supermercados

- São 100 lojas em cinco estados
- Cada loja tem departamentos de: mercearia, alimentos congelados, laticínios, carnes, produção, padaria, floral e auxiliares da saúde / beleza
- Cada loja tem aproximadamente 60.000 produtos nas suas prateleiras
- Os produtos individuais são chamados unidades em estoque (SKU – Stock Keeping Units)
- Cerca de 55.000 produtos vêm de fabricantes externos e têm códigos de barras impressos no pacote do produto chamados códigos de produto universais (UPC – Universal Product Codes)



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Exercício de Fixação

- As 5.000 unidades de estoque restantes vêm dos departamentos, tais como, carnes, produção, padaria ou floral. Esses produtos têm códigos locais individuais (SKU). Dados são coletados principalmente nos caixas da loja
- Os códigos de venda dos produtos são escaneados nos pontos de venda (POS – Point of Sales)
- Outro ponto de coleta de dados interessante é na entrada dos fundos das lojas onde os fornecedores entregam as encomendas
- Nas lojas, a administração está preocupada com a logística dos pedidos, armazenamento, e vendas dos produtos enquanto tenta maximizar o lucro



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Exercício de Fixação

- Lucro vem da venda do melhor preço possível para cada produto, da redução de custos na aquisição de produtos, e da atração do maior número de clientes possível através de preços competitivos
- As decisões mais importantes tem haver com preços e promoções
- Promoções incluem reduções temporárias de preços, propagandas em jornais, amostras em lojas e cupons.



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Exercício de Fixação

Agora é a sua vez! Vamos colocar o conhecimento em prática. Vamos preencher os 4 passos do processo da modelagem dimensional:

1. Processo de Negócio
2. Granularidade
3. Dimensões
4. Fatos



Especialização em Ciência de Dados com Big Data, BI e Data Analytics

Bibliografia

- Livro: The Data Warehouse Toolkit – guia completo para modelagem dimensional
 - Autor: Ralph Kimball / Margy Ross
 - 2ª Edição
 - Editora: Campus