



## A NECESSIDADE É A MÃE DA INVENÇÃO

### Armazéns de Dados: porquê?

Observação: baseado em slides desenvolvidos por Ms Swapnil Shrivastava

João Mendes Moreira

FEUP

## Cenário 1

ABC Pvt Ltd é uma empresa com agências em Mumbai, Delhi, Chennai and Bangalore. O gestor de vendas quer relatórios trimestrais das vendas. Cada agência tem o seu próprio sistema operacional.

## Cenário 1 : ABC Pvt Ltd.

3

**Mumbai**

**Delhi**

**Chennai**

**Banglore**

**Vendas por tipo de item  
por agência para o  
primeiro trimestre.**

**Gestor de  
vendas**

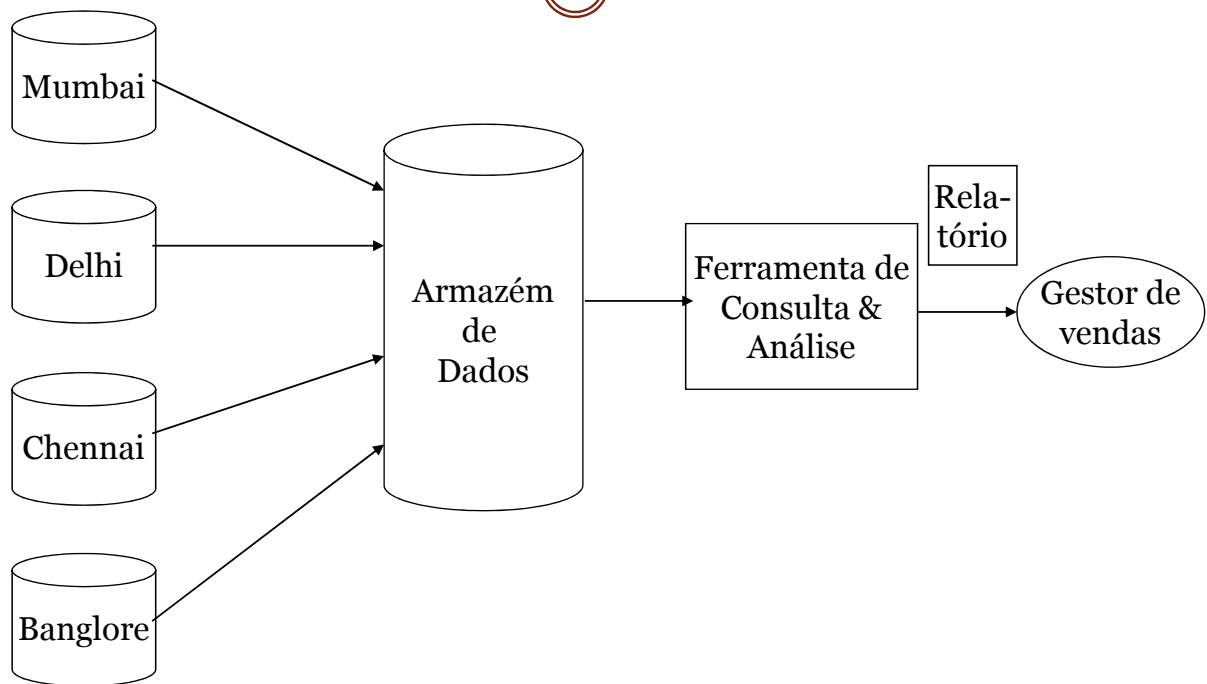
## Solução 1: ABC Pvt Ltd.

4

- Extrair informação de vendas em cada uma das quatro bases de dados.
- Armazenar a informação num repositório comum num sítio único.

## Solução 1: ABC Pvt Ltd.

5



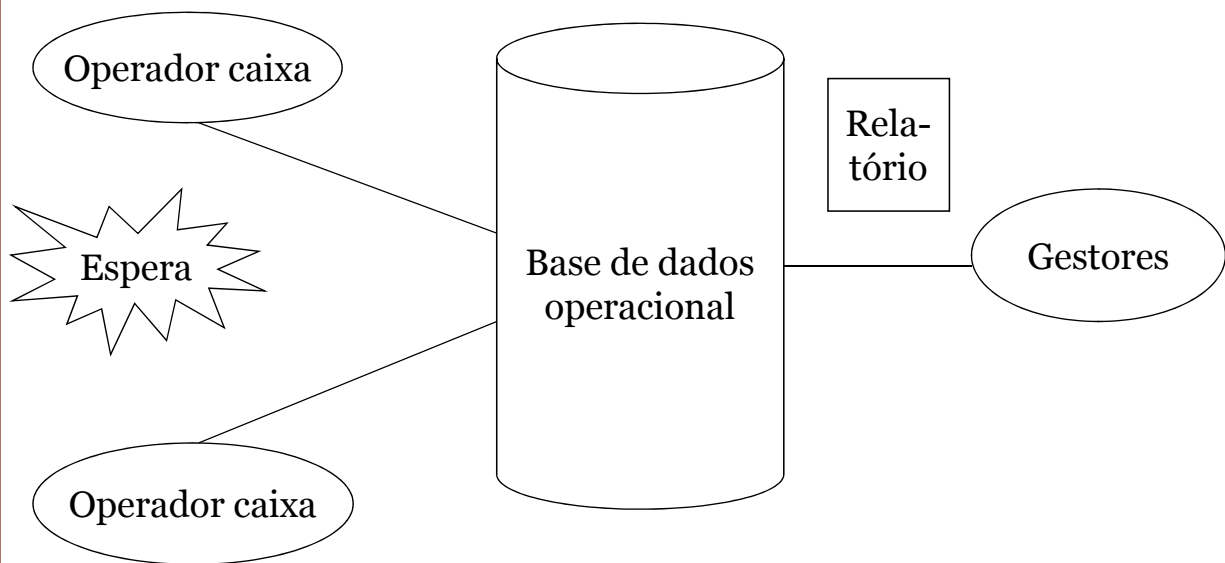
## Cenário 2

6

Um hiper-mercado tem uma base de dados operacional enorme. Sempre que os gestores pedem algum relatório, o sistema OLTP (On Line Transactional Processing) fica lento afectando o trabalho das caixas.

## Cenário 2 : Hiper-mercado

7



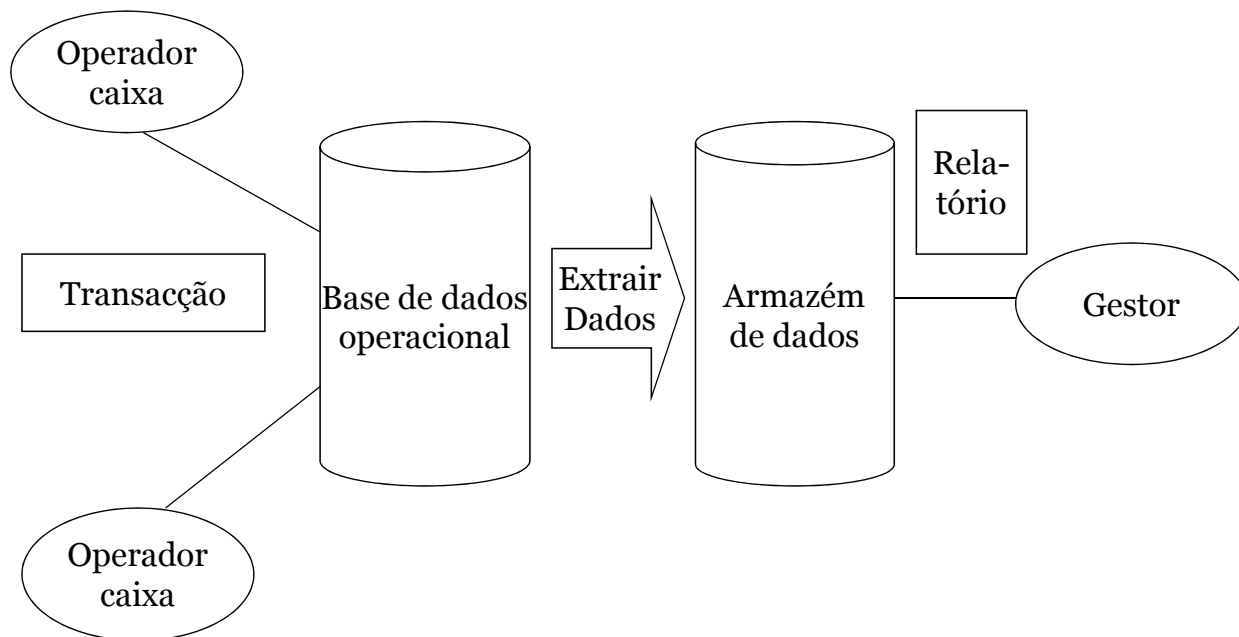
## Solução 2

8

- Extrair os dados necessários para análise da base de dados operacional.
- Armazená-los num armazém de dados.
- Refrescar o armazém de dados a intervalos de tempo regulares de forma a ter sempre dados actualizados para análise.
- O armazém de dados irá conter dados numa perspectiva histórica.

## Solução 2

9



## Cenário 3

10

Cakes & Cookies é uma pequena companhia nova. O presidente da companhia quer que ela cresça. Para isso precisa de informação para poder tomar as melhores decisões.

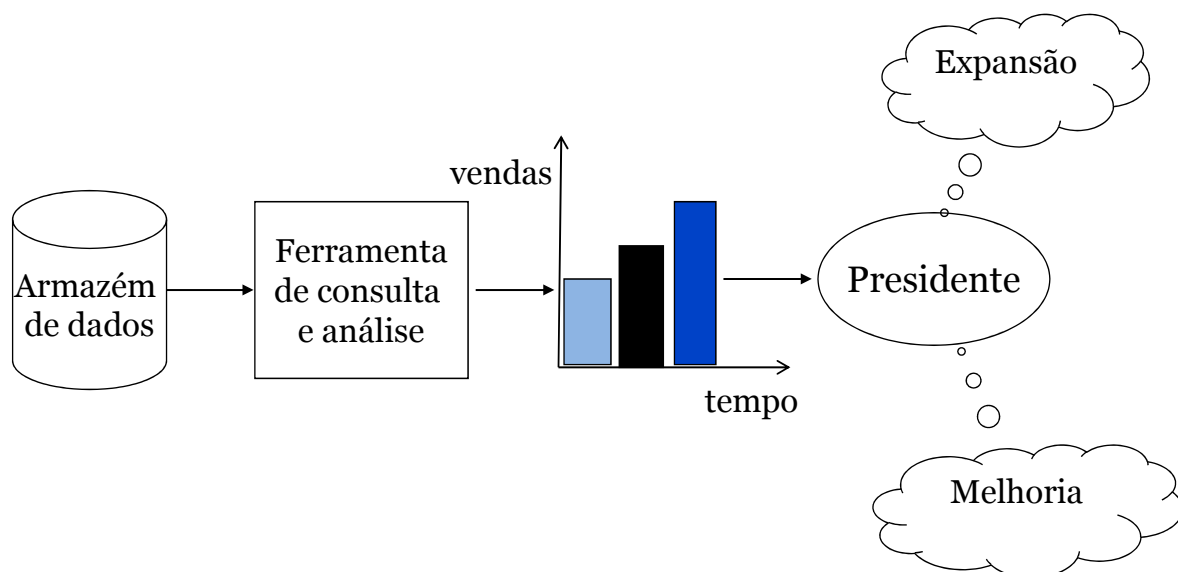
## Solução 3

11

- Melhorar a qualidade dos dados antes de os carregar para o armazém de dados.
- Efectuar limpeza e transformação dos dados antes de os carregar para o armazém de dados.
- Utilizar ferramentas para consulta de dados que permitam a realização de consultas ad-hoc.

## Solução 3

12



# O que é um armazém de dados?

## Definição de Bill Inmon

Um armazém de dados é

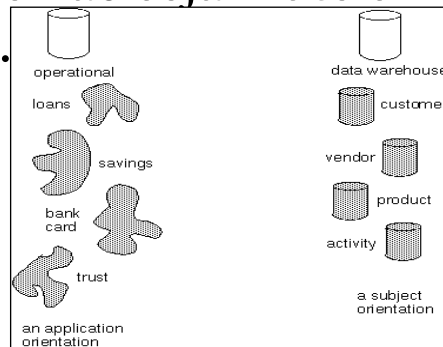
- orientado aos assuntos,
- integrado,
- varia com o tempo,
- não volátil

Colecção de dados para apoio ao processo de tomada de decisão por parte dos gestores.

# Orientado aos assuntos

15

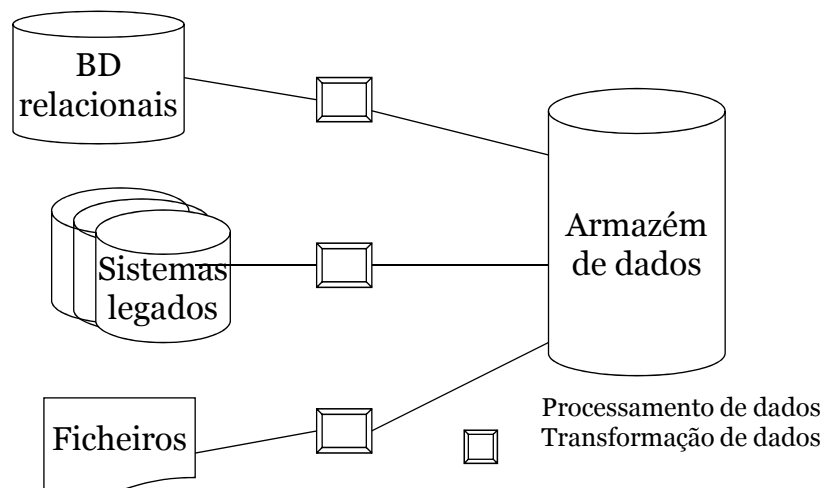
- O armazém de dados está organizado por assuntos tais como vendas, produtos, clientes, etc.
- Está orientado para a modelação e análise de dados para a tomada de decisão.
- Exclui os dados que não sejam úteis no processo de tomada de decisão.



# Integrado

16

- Um armazém de dados é construído pela integração de dados provenientes de mais do que uma fonte, sendo essas fontes tipicamente heterogêneas.
- O processamento dos dados é efectuado de forma a garantir a consistência desses mesmos dados.





# Integrado

17

- Em termos de dados.

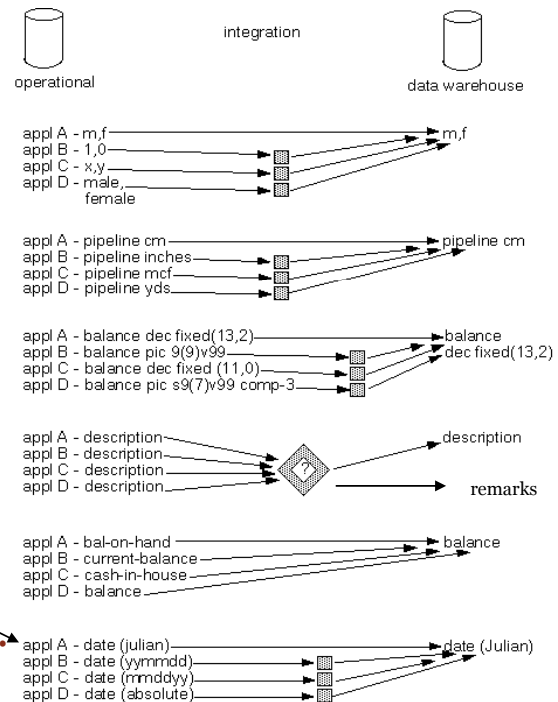
- estruturas codificadas.

- medidas de atributos.

- atributo físico de dados.

- Convenção de nomes.

- Formato do tipo de dados.



# Varia com o tempo

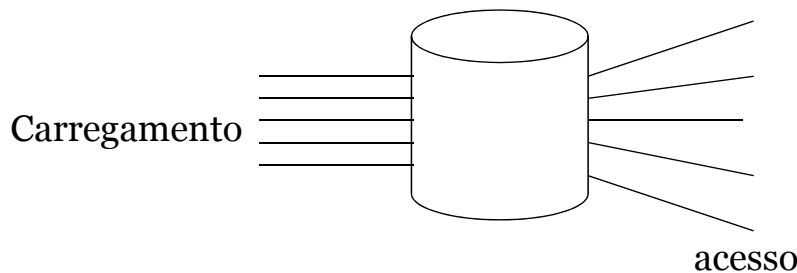
18

- Dispõe de informação com perspectiva histórica, por ex., últimos 5-10 anos
- Todas as estruturas chave contêm de forma implícita ou explícita um elemento de tempo

# Não volátil

19

- Os dados uma vez inseridos no armazém de dados não podem ser actualizados.
- O armazém de dados permite efectuar dois tipos de operações sobre os dados:
  - Carregamento inicial dos dados
  - Acesso aos dados



Relem  
brar

## O que é uma Base de Dados?

20

### Tecnologias de Informação

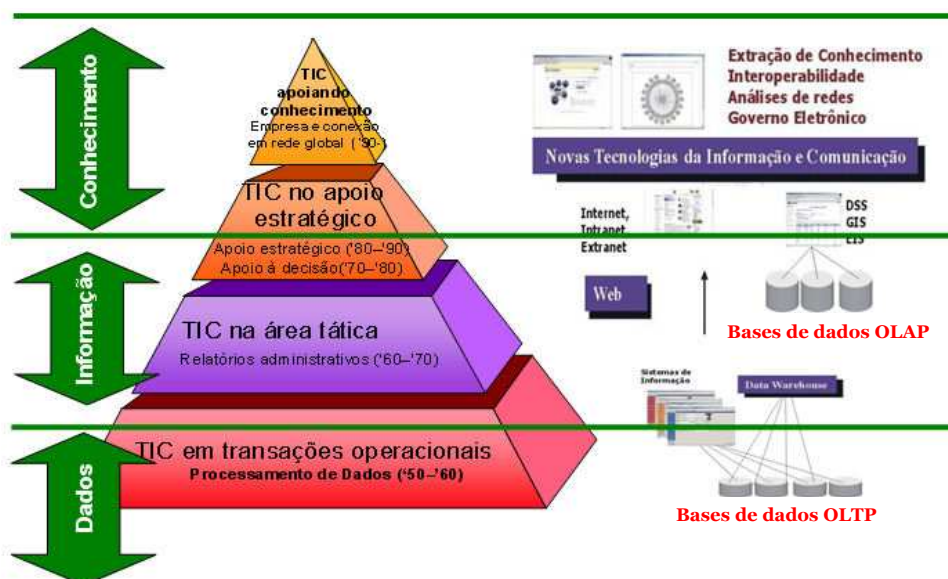


Imagem retirada de <http://revista.ibict.br/index.php/ciinf/article/viewFile/795/644/2174> em 15 de Fevereiro de 2010

## Sistemas operacionais vs. informativos

21

	Operacional	Informativo
Características	Processamento operacional	Processamento informativo
Orientação	Transacção	Análise
Utilizador	Balconista, administrador es de BDs, profissionais de BDs	Trabalhadores de conhecimento
Função	Operações diárias	Apoio à decisão
Dados	Actuais	Históricos
Vista	Detalhada, formato tabela	Agregada, multidimensional
Desenho da BD	Orientado às aplicações	Orientado aos assuntos
Unidade de trabalho	Curto, transacção simples	Consulta complexa
Acesso	Ler/escrever	Maioritariamente de leitura

João Mendes Moreira

FEUP

## Sistemas operacionais vs. informativos

22

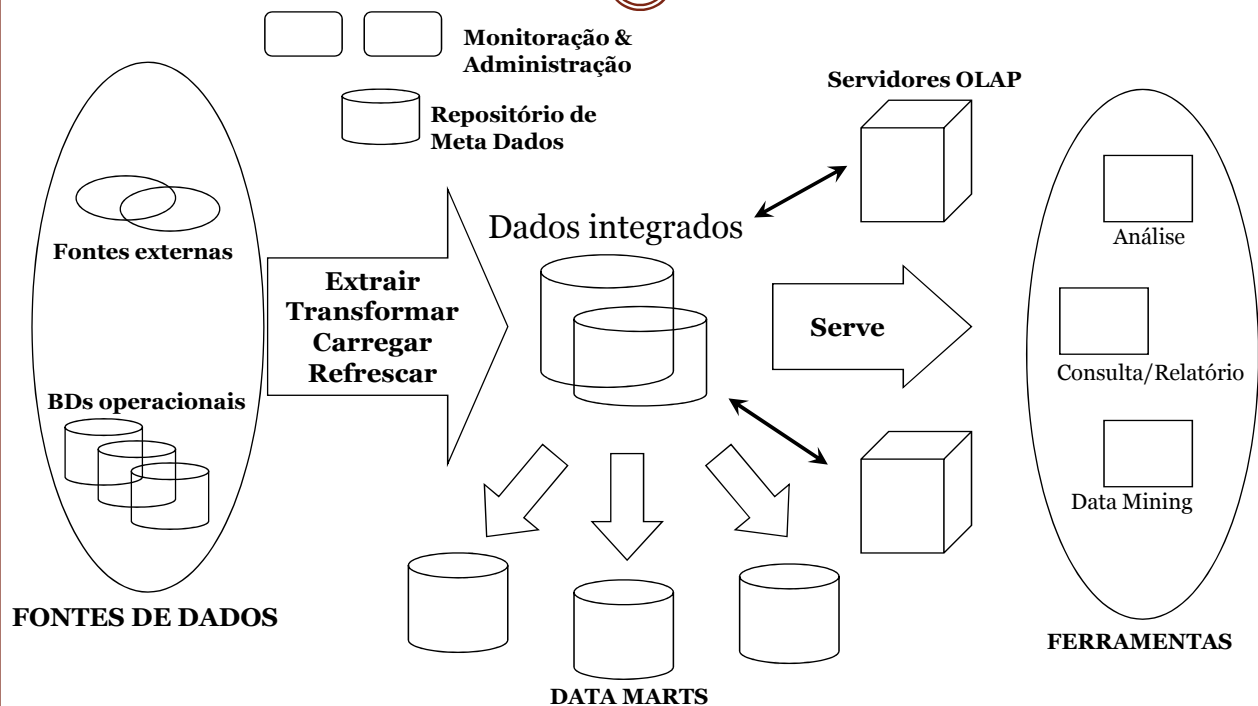
	Operacional	Informativo
Foco	Dados de entrada	Informação pretendida
Nº de registos acedidos	Dezenas	Milhões
Nº de utilizadores	Milhares	Centenas
tamnhoo da BD	De 100MB a GB	De 100 GB a TB
Prioridade	Desempenho elevado, Disponibilidade elevada	Flexibilidade elevada, Autonomia do utilizador final
Medida	Transacção	Consulta

João Mendes Moreira

FEUP

# Arquitetura dos armazéns de dados

23



João Mendes Moreira

FEUP

# Arquitetura dos armazéns de dados

24

- Servidor de Armazém de Dados
  - Quase sempre um SGBD relacional, raramente ficheiros tipo tabela
- Servidores OLAP
  - Para gerir e operar estruturas de dados multi-dimensionais
- Clientes
  - Ferramentas de consulta e geração de relatórios
  - Ferramentas de análise
  - Ferramentas para descoberta de conhecimento (*data mining*)

João Mendes Moreira

FEUP

# Esquemas para Armazéns de Dados

25

- Esquema em estrela
- Esquema de constelação de factos
- Esquema em floco de neve

## Esquema em estrela

26

- Uma só tabela de factos, grande, e uma tabela por cada dimensão.
- Cada registo da tabela de factos tem uma chave estrangeira para cada uma das dimensões às quais acrescem atributos adicionais.
- Não guarda as hierarquias existentes de forma directa.

## Esquema em estrela (continuação)

27

### Dimensão Loja

idLoja
Nome Loja
Cidade
Região
País

### Tabela de factos

idLoja
idProduto
idPeríodo
<u>Unidades</u>
<u>Preço</u>

### Dimensão tempo

idPeríodo
Ano
Trimestre
Mês

idProduto
Descrição Produto

### Dimensão produto

**Benefícios:** fácil de entender, fácil de definir hierarquias, número reduzido de junções necessárias.

## Esquema em floco de neve

28

- Variante do esquema em estrela.
- Uma só tabela de factos, grande, e uma ou mais tabelas por cada dimensão.
- As tabelas de dimensões são normalizadas, i.e., representa cada dimensão em várias tabelas em vez de o fazer numa tabela só (como o esquema em estrela).

# Esquema em floco de neve (continuação)

29

## Dimensão Loja

idLoja
Nome Loja
idCidade

## Dimensão Cidade

idCidade
Cidade
Região
País

## Tabela de factos

idLoja
idProduto
idPeríodo
<u>Unidades</u>
<u>Preço</u>

## Dimensão Produto

idProduto
Descrição Produto

## Dimensão Tempo

idPeríodo
Ano
Trimestre
Mês

Desvantagem: Junções dispendiosas em termos de tempo.  
Geração de relatórios lenta.

# Esquema de constelação de factos

30

- Várias tabelas de factos que partilham tabelas de dimensões.
- Este esquema pode ser visto como um conjunto de estrelas, e por isso, é designado por esquema em galáxia ou esquema de constelação de factos.
- Esquema adequado para aplicações mais complexas.

## Esquema em constelação de factos (continuação)

31

**Tabela de factos  
Vendas**

idLoja
idProduto
idPeríodo
<u>Unidades</u>
<u>Preço</u>

**Dimensão Produto**

idProduto
DescriçãoProduto

**Tabela de factos  
Transporte**

idTransporte
idLoja
idProduto
idPeríodo
<u>Unidades</u>
<u>Preço</u>

**Dimensão Loja**

idLoja
Nome Loja
Cidade
Região
País

## Construção do Armazém de Dados

32

- Selecção de dados
- Pré-processamento de dados
  - Preenchimento de valores em falta
  - Remoção de inconsistências
- Transformação & Integração de dados
- Carregamento de dados

Os dados de um Armazém de Dados são guardados na forma de tabelas de factos e tabelas de dimensões.



# Caso de estudo

33

- Afco Foods & Beverages é uma nova companhia que produz leite e produtos dele derivados, pão e produtos variados à base de carne. A unidade de produção situa-se em Baroda, Índia.
- Estes produtos são vendidos nas regiões Norte, Noroeste e Oeste da Índia.
- Existem pontos de venda em Mumbai, Pune, Ahemdabad, Delhi e Baroda.
- O presidente da companhia quer informação das vendas.

# Informação das vendas

34

**Relatório:** Número de unidades vendidas.

113

**Relatório:** Número de unidades vendidas ao longo do tempo

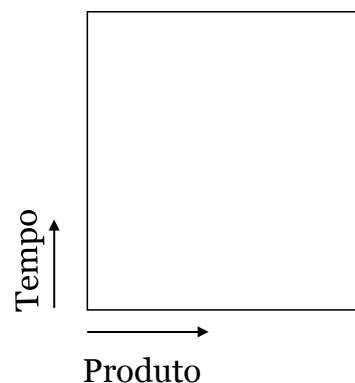
Janeiro	Fevereiro	Março	Abril
14	41	33	25

# Informação das vendas

35

**Relatório:** Número de unidades vendidas de cada produto ao longo do tempo

	Jan	Fev	Mar	Abr
Pão de trigo			6	17
Queijo	6	16	6	8
Alheira	8	25	21	

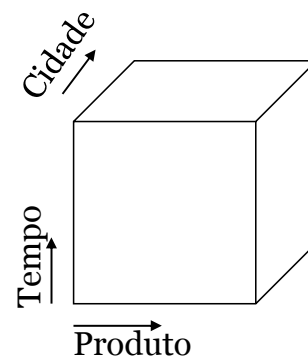


# Informação das vendas

36

**Relatório:** Número de unidades vendidas em cada cidade por produto, ao longo do tempo

		Jan	Fev	Mar	Abr
Mumbai	Pão de trigo			3	10
	Queijo	3	16	6	
	Alheira	4	16	6	
Pune	Pão de trigo			3	7
	Queijo	3			8
	Alheira	4	9	15	



# Informação das vendas

37

Relatório: Número de unidades vendidas e receita por região e produto ao longo do tempo.

		Jan		Fev		Mar		Abr	
		Rec.	U	Rec.	U	Rec.	U	Rec.	U
Mumbai	Pão de trigo					7.44	3	24.80	10
	Queijo	7.95	3	42.40	16	15.90	6		
	Alheira	7.32	4	29.98	16	10.98	6		
Pune	Pão de trigo					7.44	3	17.36	7
	Queijo	7.95	3					21.20	8
	Alheira	7.32	4	16.47	9	27.45	15		

# Medidas de venda & Dimensões

38

- Medida – Unidades vendidas, Quantidade.
- Dimensões – Produto, Tempo, Região.

# Armazém de Dados: Vendas

39

Tabela de factos

Cidade	Produto	Mês	Unidades	Rupias
Mumbai	Pão de trigo	Janeiro	3	7.95
Mumbai	Queijo	Janeiro	4	7.32
Pune	Pão de trigo	Janeiro	3	7.95
Pune	Queijo	Janeiro	4	7.32
Mumbai	Alheira	Fevereiro	16	42.40

# Armazém de Dados: Vendas

40

idCidade	idProduto	Mês	Unitdades	Rupias
1	589	1/1/1998	3	7.95
1	1218	1/1/1998	4	7.32
2	589	1/1/1998	3	7.95
2	1218	1/1/1998	4	7.32
1	589	2/1/1998	16	42.40

# Armazém de Dados: Vendas

41

Tabelas: Dimensão Produto

idProduto	Nome Produto	idProdutoCategoria
589	Pão de trigo	1
590	Pão de trigo	1
288	Bolos de côco	2

idProdutoCategoria	ProdutoCategoria
1	Pão
2	Bolos

# Armazém de Dados: Vendas

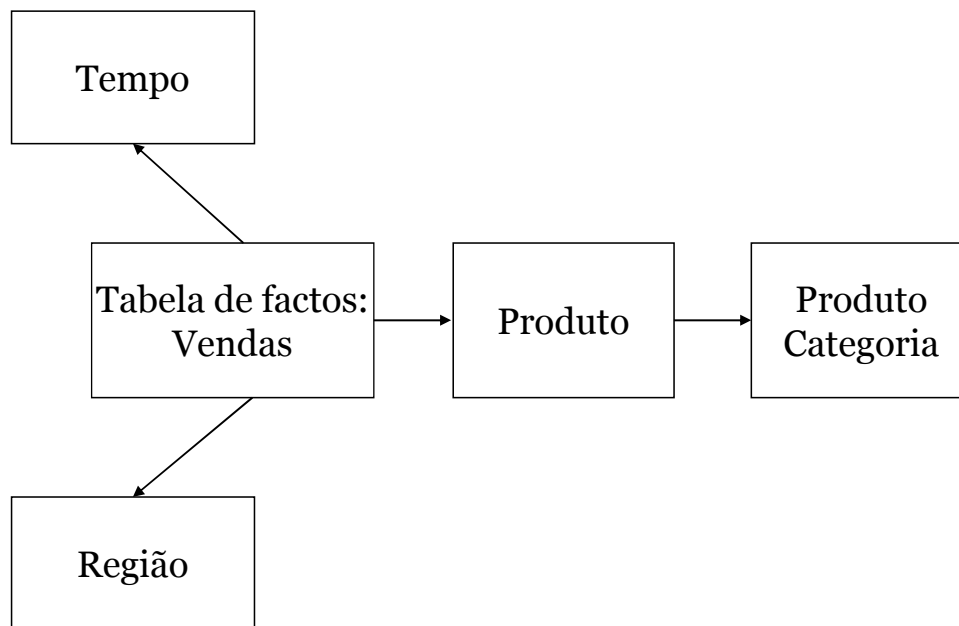
42

Tabela: Dimensão Região

idCidade	Cidade	Região	País
1	Mumbai	Oeste	Índia
2	Pune	Noroeste	Índia

# Armazém de Dados: Vendas

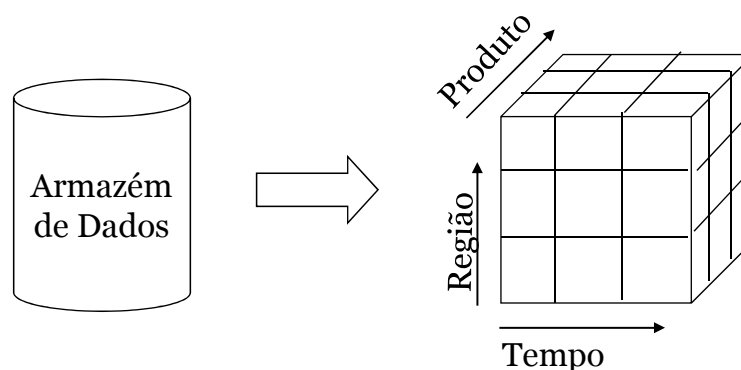
43



## OLAP: On-Line Analytical Processing

44

- Permite a analistas, gestores e executivos analisar dados através do acesso rápido, consistente e interactivo a uma grande variedade de diferentes vistas obtidas a partir de dados em formato tipo tabela de forma a transmitir a dimensionalidade real da empresa tal como é percebida pelo utilizador.



# Cubo OLAP

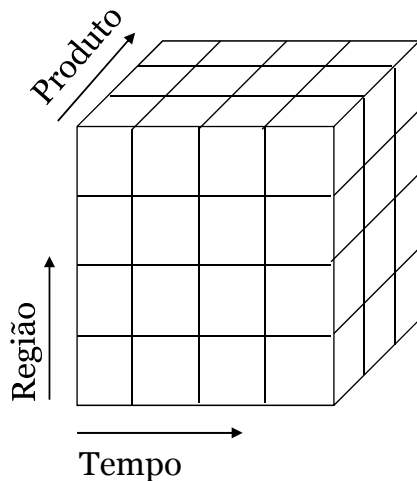
45

Cidade	Produto	Tempo	Unidades	Dólares
Todas	Todos	Todos	113	251.26
Mumbai	Todos	Todos	64	146.07
Mumbai	Pão branco	Todos	38	98.49
Mumbai	Pão de trigo	Todos	13	32.24
Mumbai	Pão de trigo	1ºTrimestre	3	7.44
Mumbai	Pão de trigo	Março	3	7.44

## Operações OLAP

46

Detalhar



Categoria, por ex., aplicação eléctrica



Sub Categoria, por ex., cozinha

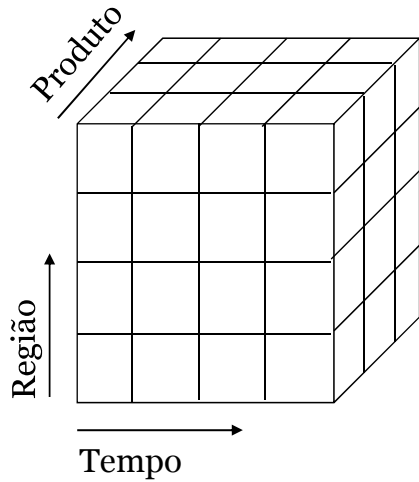


Produto, por ex., torradeira

# Operações OLAP

47

Generalizar



Categoria, por ex., aplicação eléctrica

Sub Categoria, por ex., cozinha



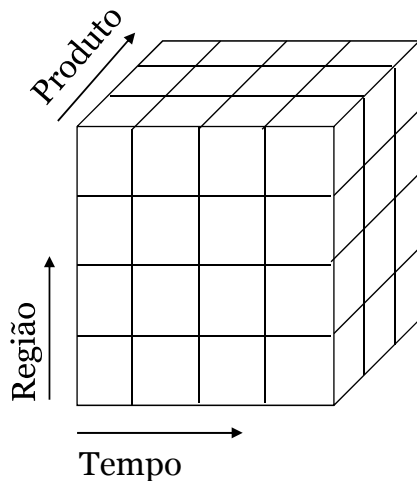
Produto, por ex., torradeira



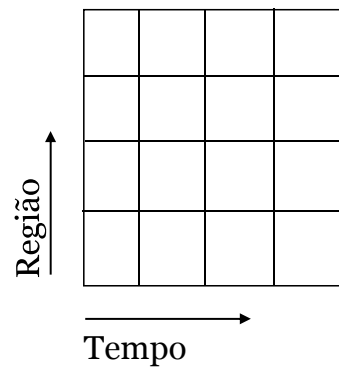
# Operações OLAP

48

Slice and Dice



Produto=Torradeira

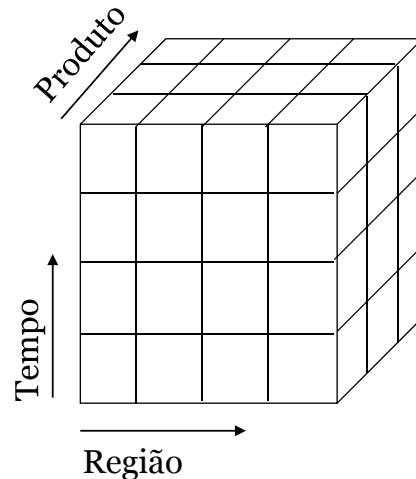
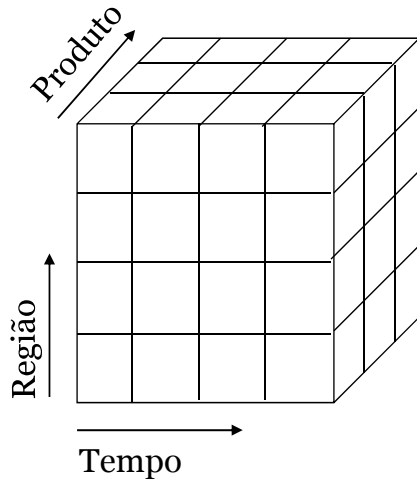




# Operações OLAP

49

Pivot



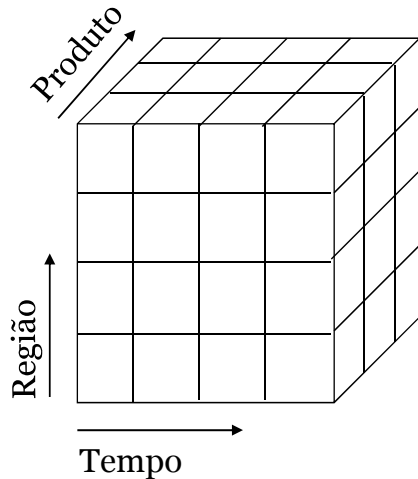
## Servidor OLAP

50

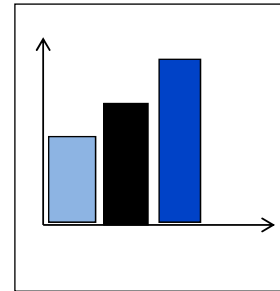
- Um Servidor OLAP é um engenho para manipulação de dados, de elevada capacidade e multi utilizador, desenhado especificamente para apoiar e operar sobre estruturas de dados multi-dimensionais.
- Existem os seguintes tipos de servidores OLAP:
  - Servidor MOLAP
  - Servidor ROLAP
  - Servidor HOLAP

# Apresentação

51



Ferramenta  
para geração  
de relatórios

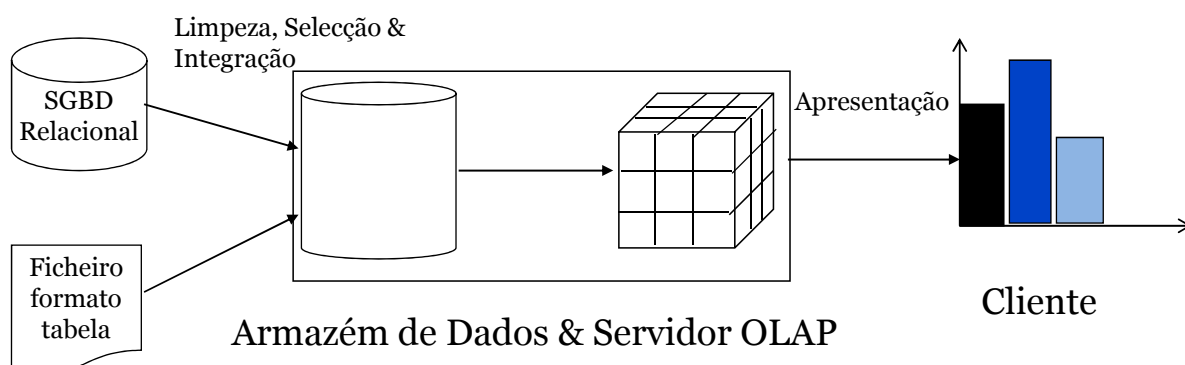


Relatório

## Armazenamento de Dados inclui

52

- Construção do Armazém de Dados
- Processamento analítico online (OLAP).
- Apresentação.



# Necessidade de Armazenamento de Dados

53

- As empresas têm grandes quantidades de dados operacionais
- Trabalhadores do conhecimento querem transformar esses dados em informação útil.
- Esta informação é útil para apoiar a tomada de decisões estratégicas.
- É uma plataforma de dados históricos consolidados para análise.
- Armazena dados de boa qualidade permitindo aos trabalhadores de conhecimento a tomada de decisões correctas.

# Necessidade de Armazenamento de Dados

54

- Numa perspectiva de negócio:
  - é uma armas do marketing
  - ajuda a manter clientes pela obtenção de mais conhecimento sobre as suas necessidades.
  - ferramenta valiosa num mundo competitivo e em rápida mudança.

# Ferramentas de Armazenamento de Dados

55

- Armazém de Dados
  - Microsoft SQL Server 2008 & Microsoft BI platform
  - Oracle 11g Warehouse Builder
- Ferramentas OLAP
  - SQL Server 2008 Analysis Services
  - Oracle OLAP 11g
  - OpenI (software livre)
- Reporting tools
  - DataWarehouse Explorer (para SQL Server)
  - OLAP Reporting Tool (para MS Excel)
  - IBM Cognos 8 Business Intelligence

## Referências

56

- The Data Warehouse lifecycle toolkit (2<sup>nd</sup> edition), Ralph Kimball, Margy Ross, Warren Thornthwaite, and Joy Mundy, 2008.
- Building the Data Warehouse, William H. Inmon, 2005
- <http://www.dwinfocenter.org>
- <http://datawarehouse.ittoolbox.com/>