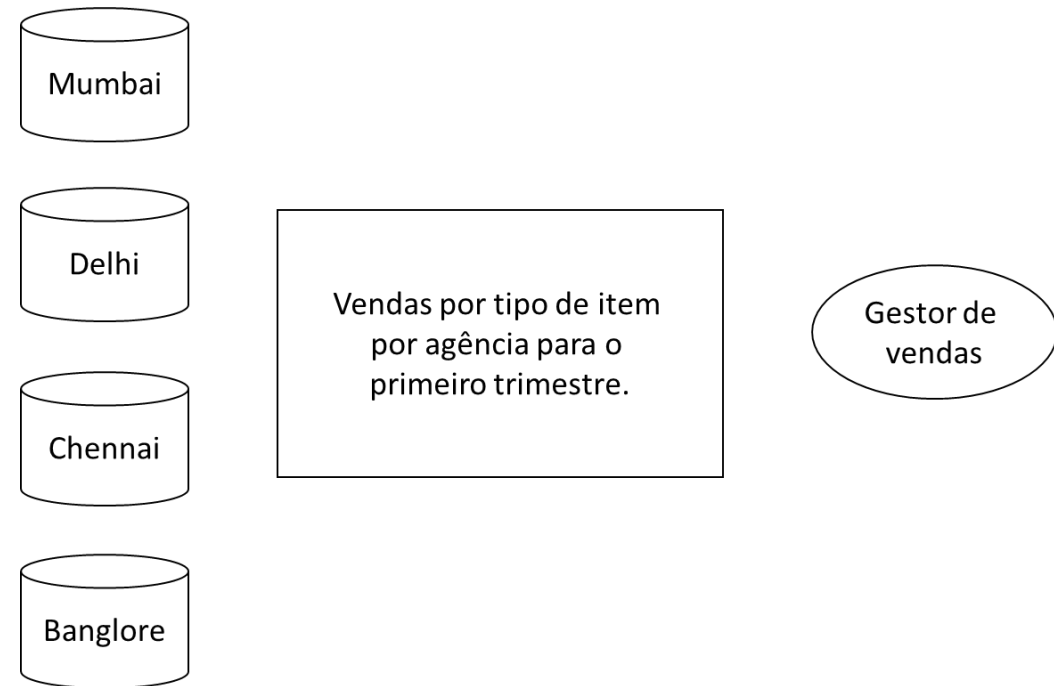


Bases de Dados

DATA WAREHOUSING

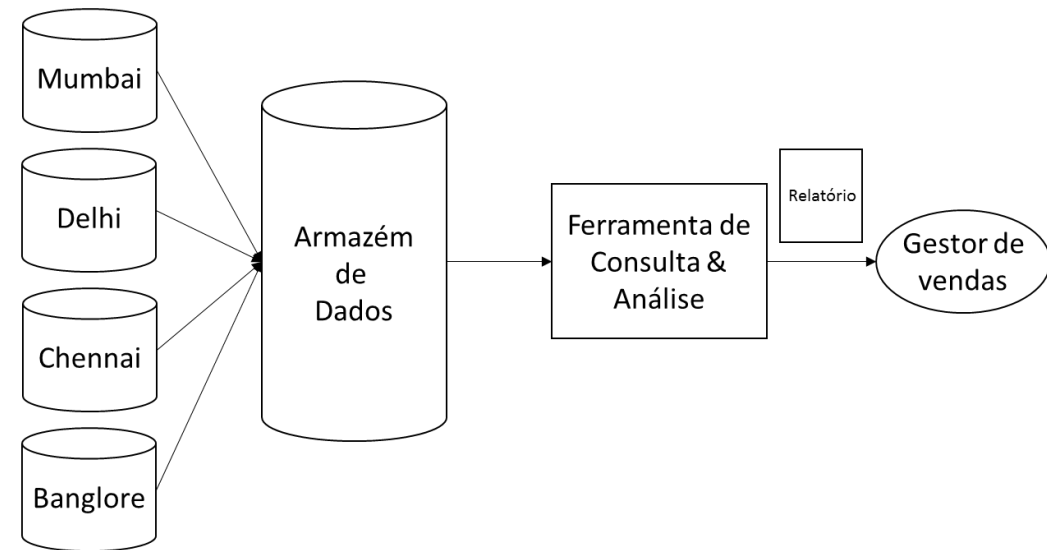
Exemplo 1

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



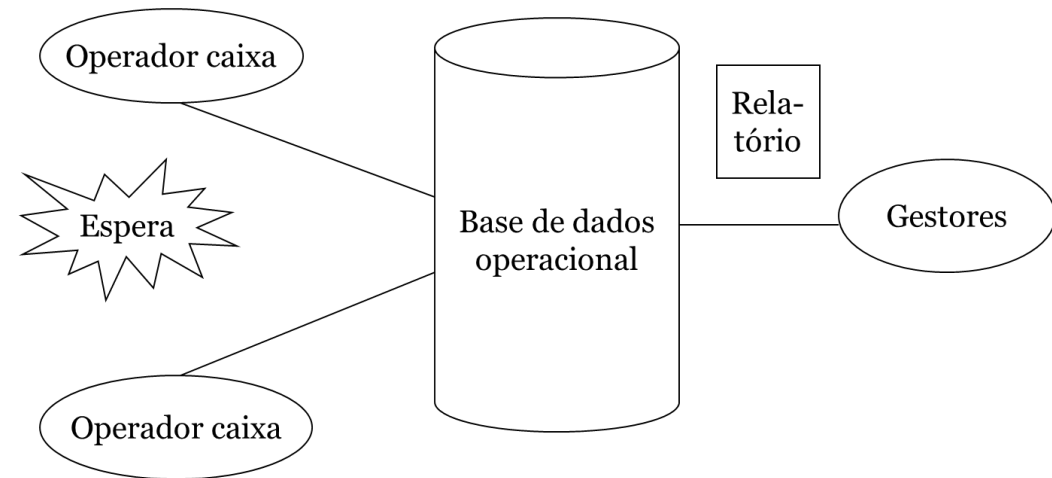
Exemplo 1 – solução

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



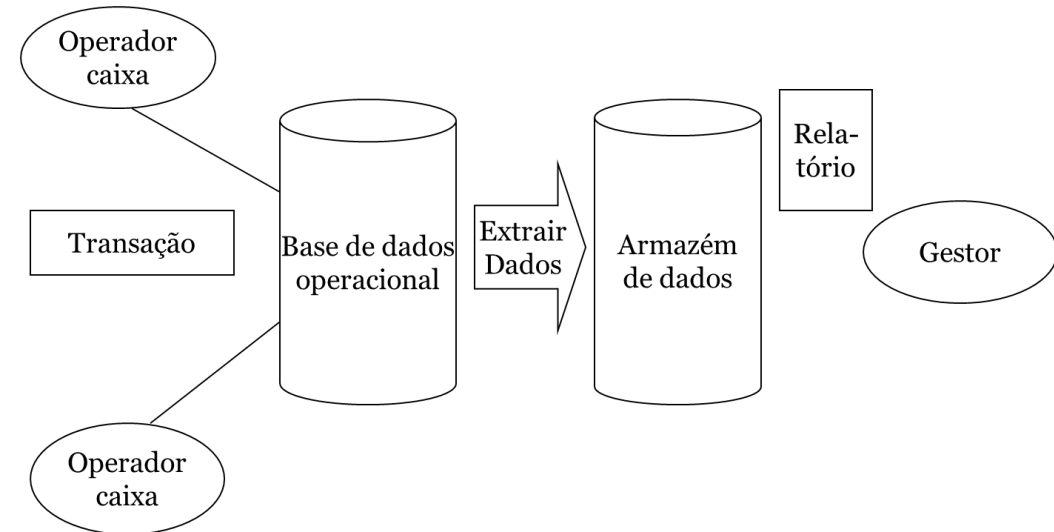
Exemplo 2

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



Exemplo 2 - solução

- 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 atualizados para análise.
- O armazém de dados irá conter dados numa perspetiva histórica.

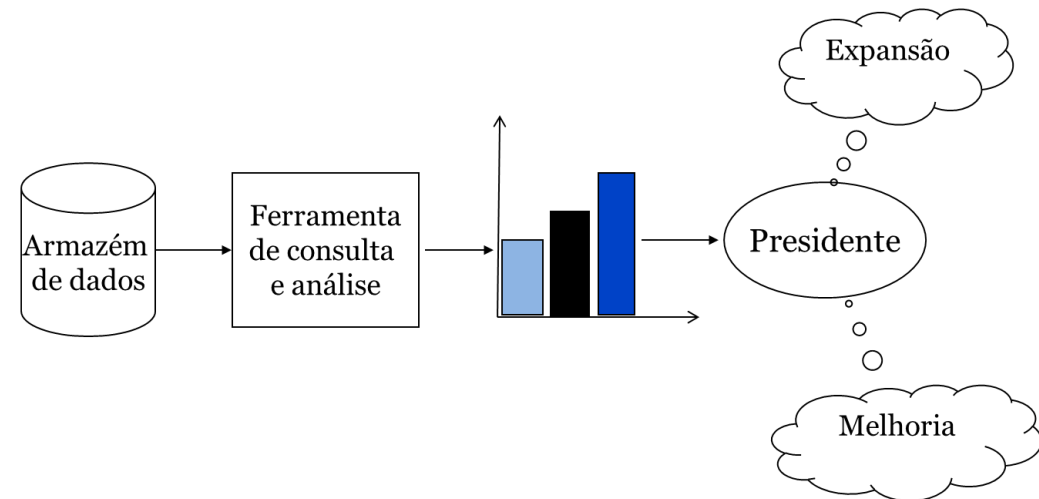


Exemplo 3

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

Exemplo 3

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

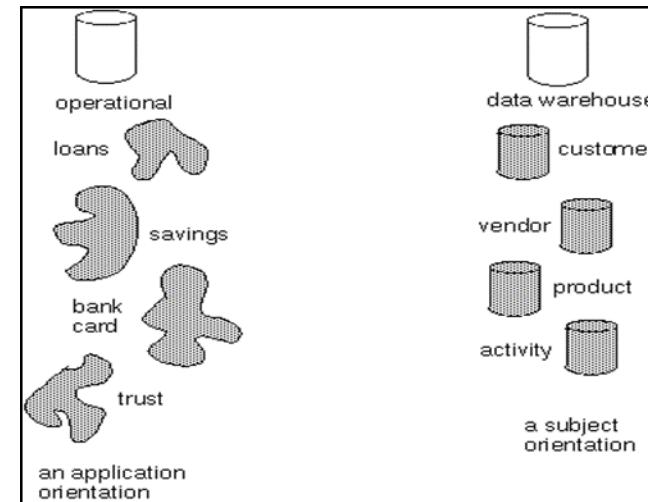


Data warehouse

- Um armazém de dados é
 - orientado aos assuntos,
 - integrado,
 - varia com o tempo,
 - não volátil
- Coleção de dados para apoio ao processo de tomada de decisão por parte dos gestores.

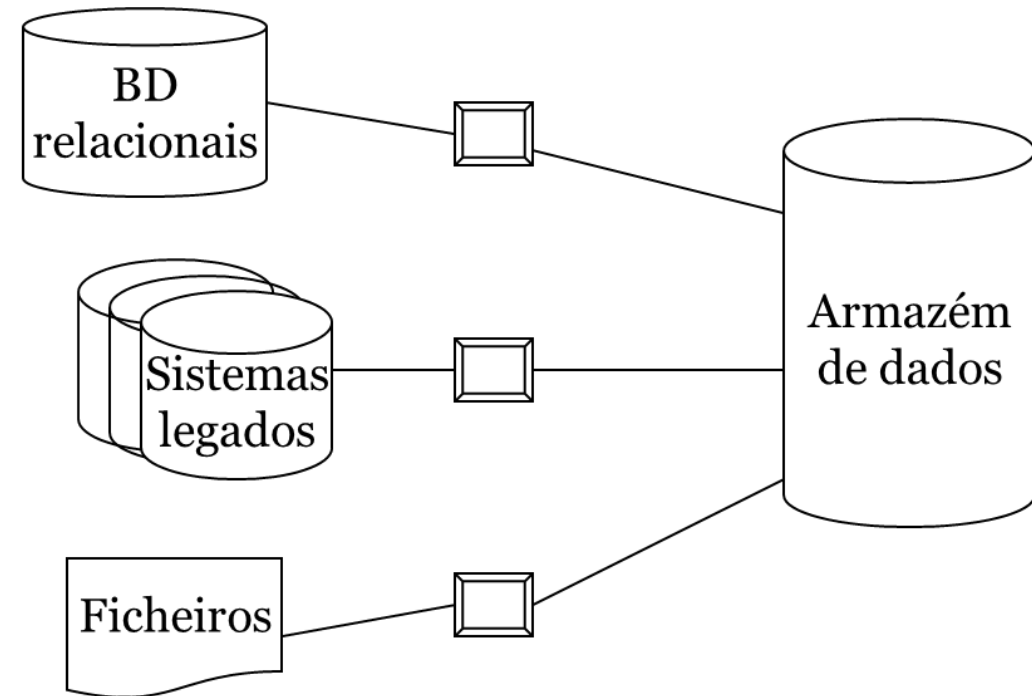
DW – orientado aos assuntos

- 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.



DW - Integrado

- 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 efetuado de forma a garantir a consistência desses mesmos dados.



DW - Integrado

- Em termos de dados.

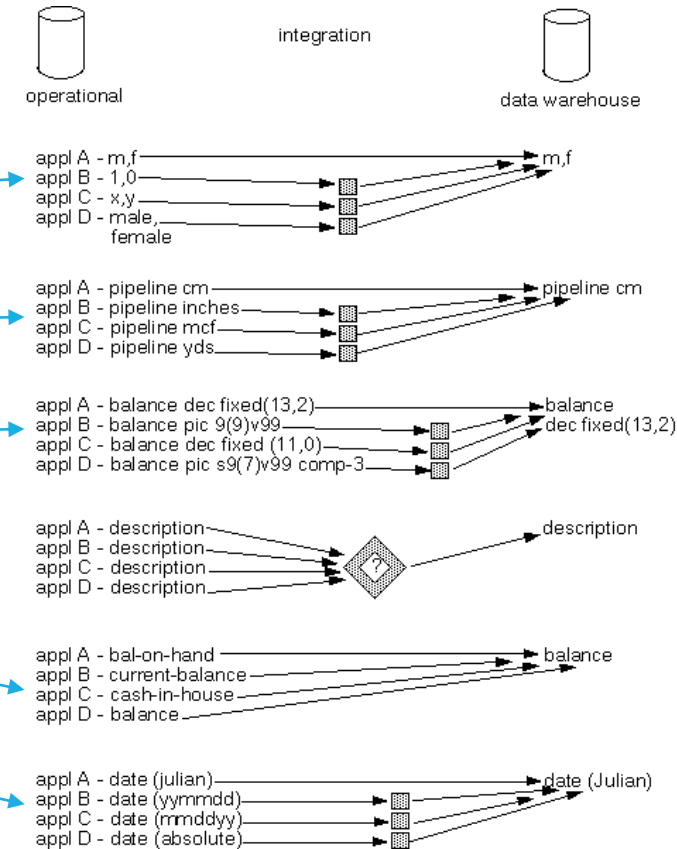
- estruturas codificadas.

- medidas de atributos.

- atributo físico de dados.

- Convenção de nomes.

- Formato do tipo de dados.

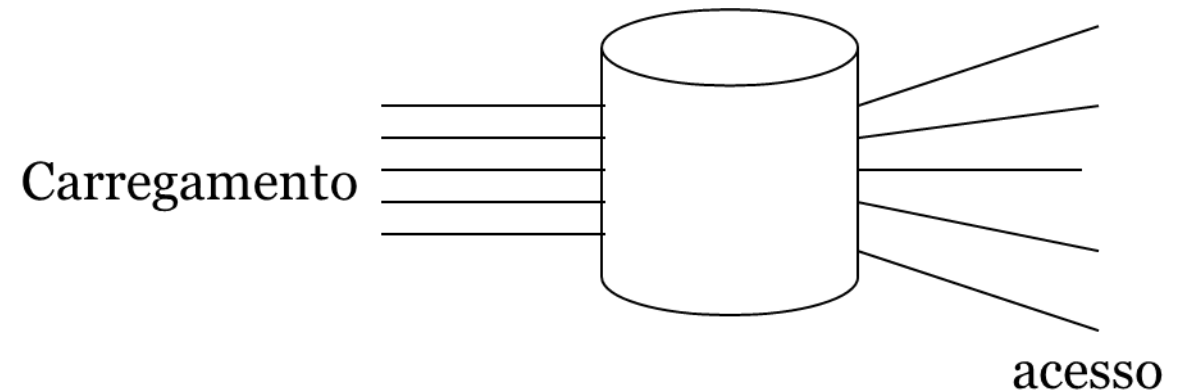


Varia com o tempo

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

Não volátil

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

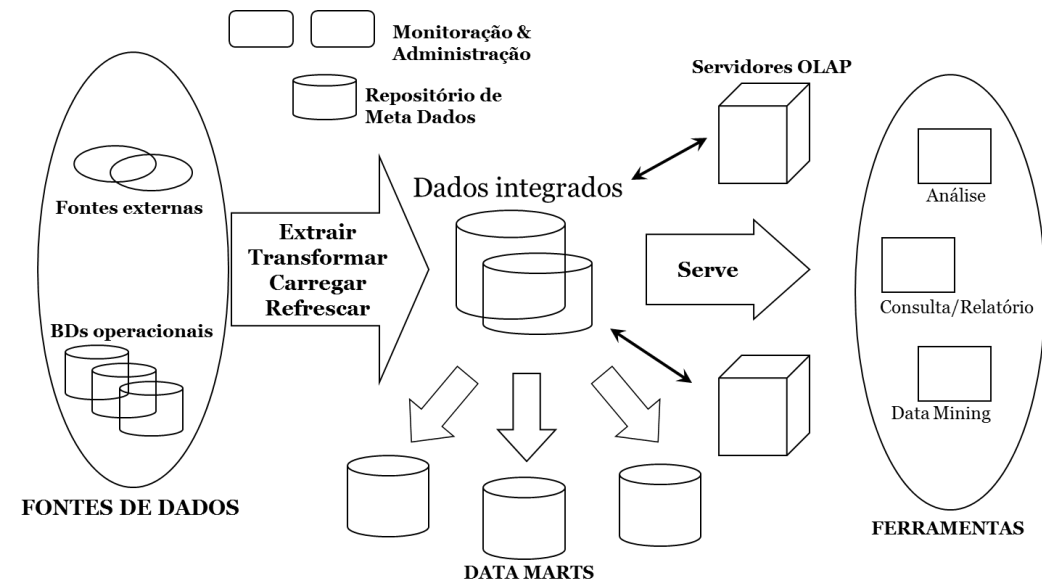


Sistemas operacionais vs. informativos

	Operacional	Informativo
Características	Processamento operacional	Processamento informativo
Orientação	Transação	Análise
Utilizador	Balconista, administradores de BDs, profissionais de BDs	Business Intelligence
Função	Operações diárias	Apoio à decisão
Dados	Atuais	Históricos
Vista	Detalhada, formato tabela	Agregada, multidimensional
Desenho da BD	Orientado às aplicações	Orientado aos assuntos
Unidade de trabalho	Curto, transação simples	Consulta complexa
Acesso	Ler/escrever	Maioritariamente de leitura
Foco	Dados de entrada	Informação pretendida
Nº de registos acedidos	Dezenas	Milhões
Nº de utilizadores	Milhares	Centenas
tamanho 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

Arquitetura dos datawarehouses

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



Esquemas para Datawarehouses

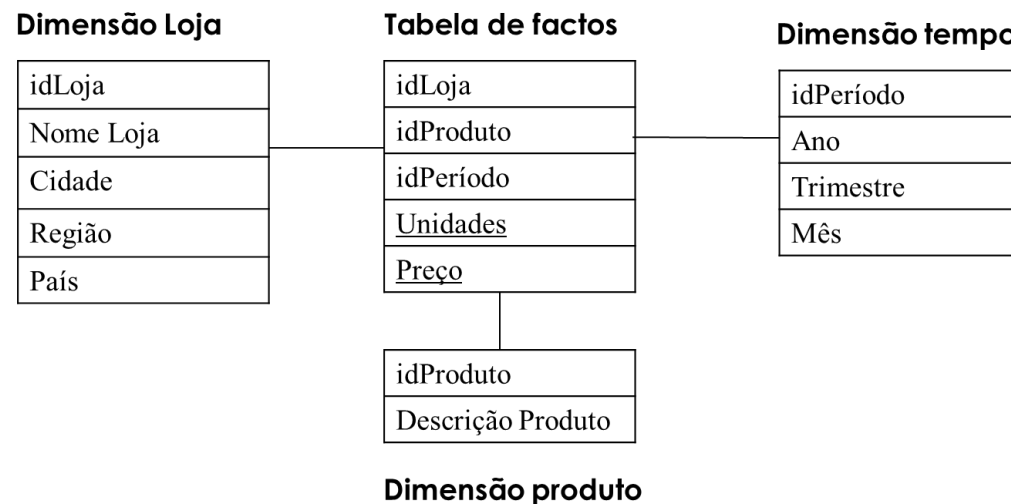
- Esquema em estrela
- Esquema de constelação de factos
- Esquema em floco de neve

Esquema em estrela

- 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 direta.

Benefícios:

- fácil de entender,
- fácil de definir hierarquias,
- número reduzido de junções necessárias.

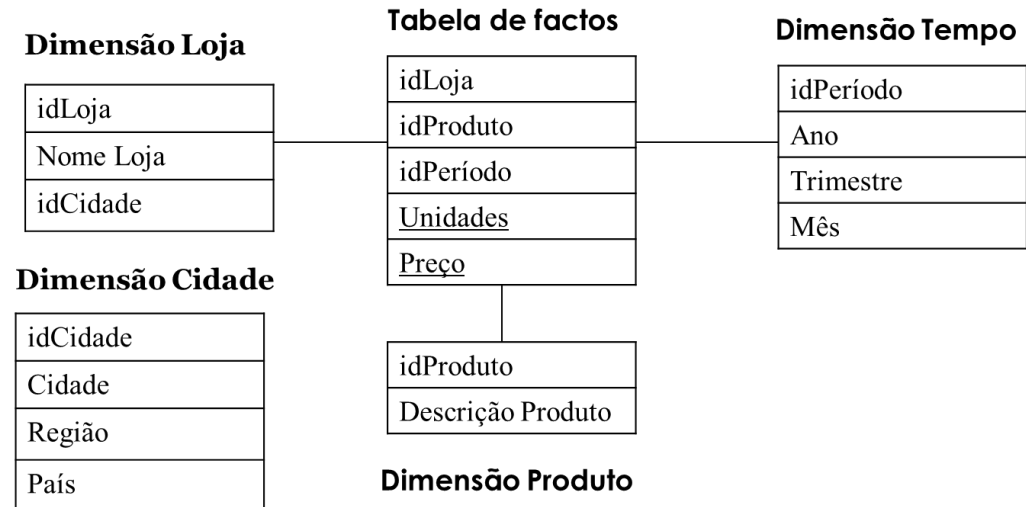


Esquema em floco de neve

- 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).

Desvantagens:

- Junções dispendiosas em termos de tempo.
- Geração de relatórios lenta.



Esquema de constelação de factos

- 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.

Tabela de factos Vendas

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

Dimensão Produto

idProduto
DescriçãoProduto

Dimensão Loja

idLoja
Nome Loja
Cidade
Região
País

Tabela de factos Transporte

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

Construção do Datawarehouse

- Seleção 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.

Exemplo

- **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

Relatório: Número de unidades vendidas.

113

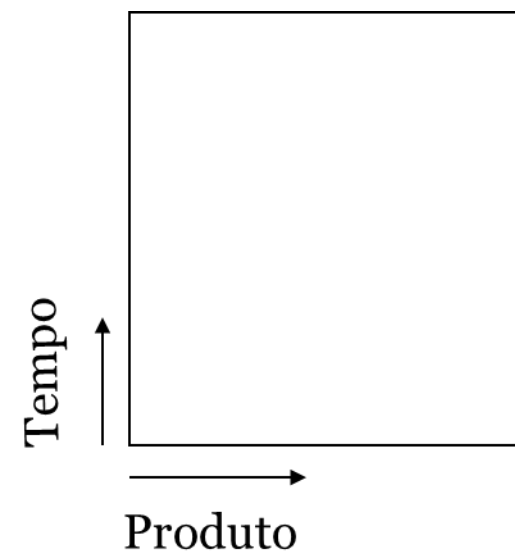
Relatório: Número de unidades vendidas ao longo do tempo

Janeiro	Fevereiro	Março	Abril
14	41	33	25

Informação das vendas

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

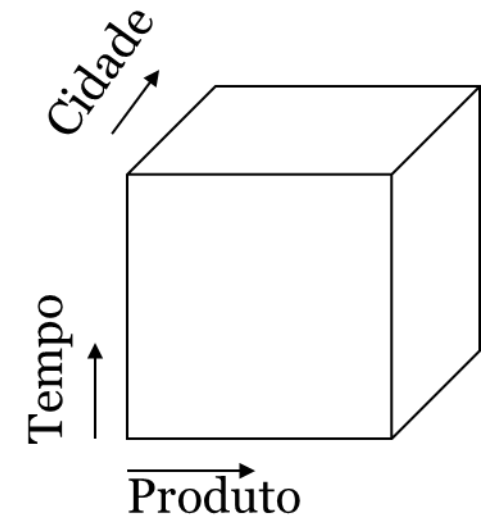
	Jan	Fev	Mar	Abr
Pão de trigo			6	17
Queijo	6	16	6	8
Alheira	8	25	21	



Informação das vendas

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

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

- Medida:
 - Unidades vendidas
 - Valor (rupias).
- Dimensões:
 - Produto,
 - Tempo,
 - Região.

Armazém de Dados: Vendas

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

Cidade

idCidade	idProduto	Mês	Unidades	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

Tabelas: Dimensão Produto

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

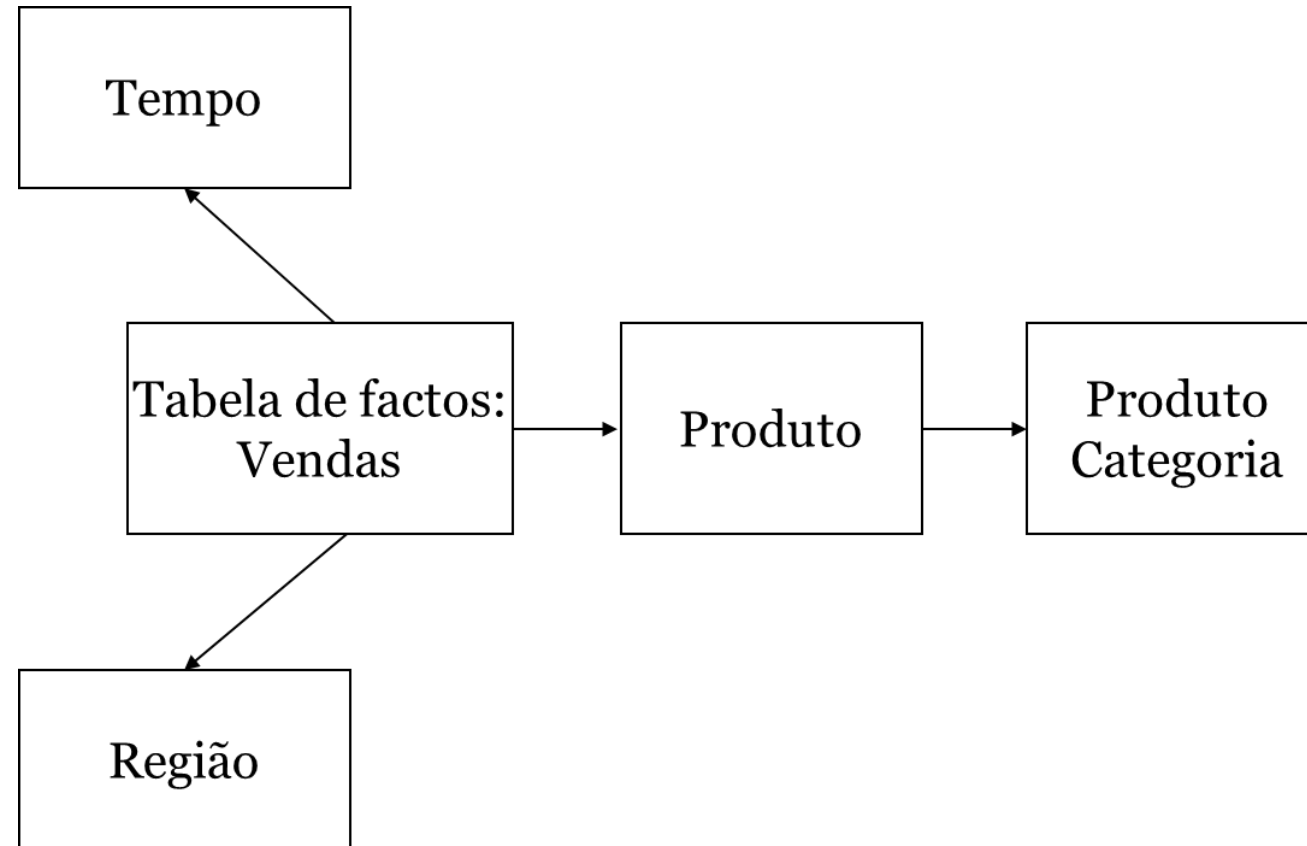
idProdutoCategoria	ProdutoCategoria
1	Pão
2	Bolos

Armazém de Dados: Vendas

Tabela: Dimensão Região

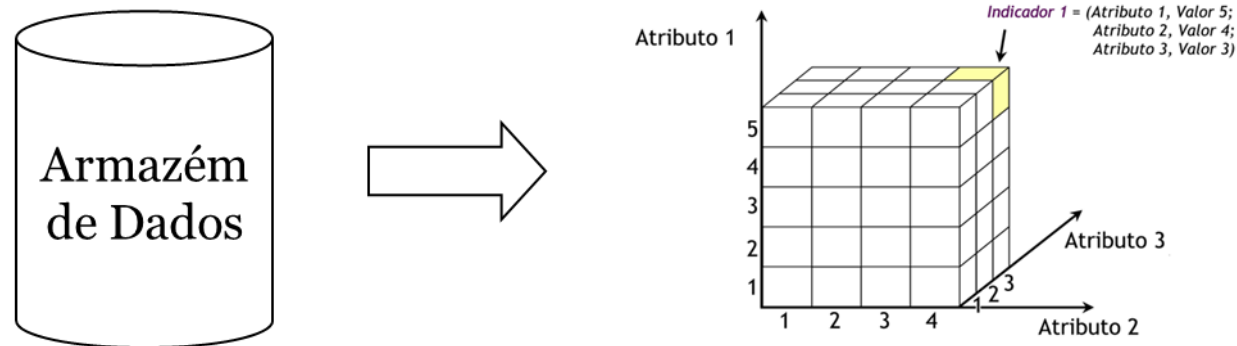
idCidade	Cidade	Região	País
1	Mumbai	Oeste	Índia
2	Pune	Noroeste	Índia

Armazém de Dados: Vendas



OLAP: On-Line Analytical Processing

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

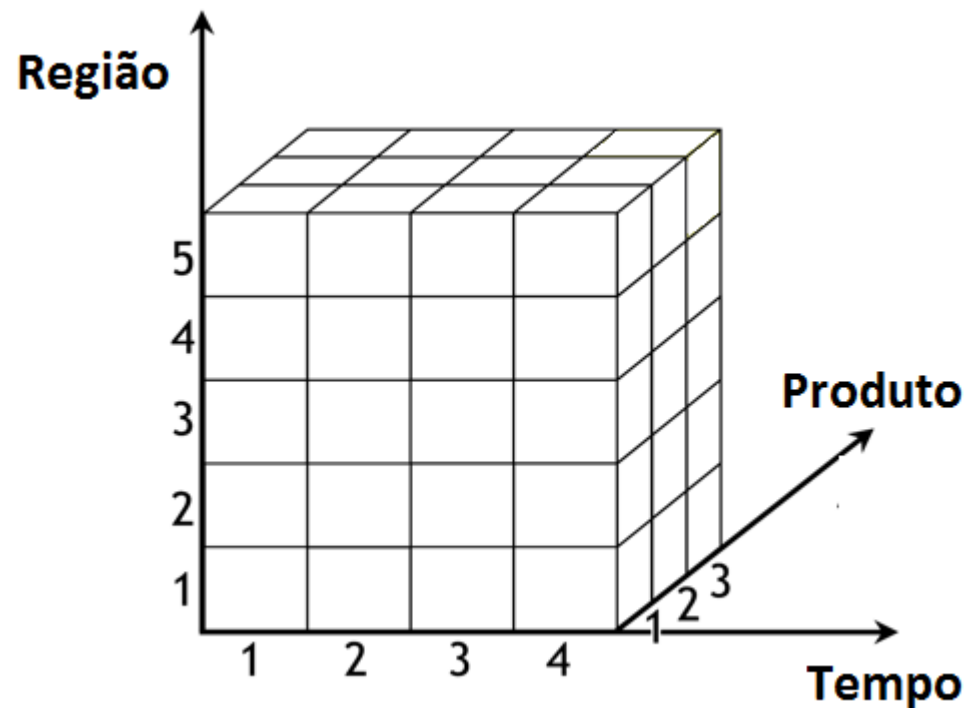


Cubo OLAP

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

Detalhar



Categoria, por ex., aplicação eléctrica



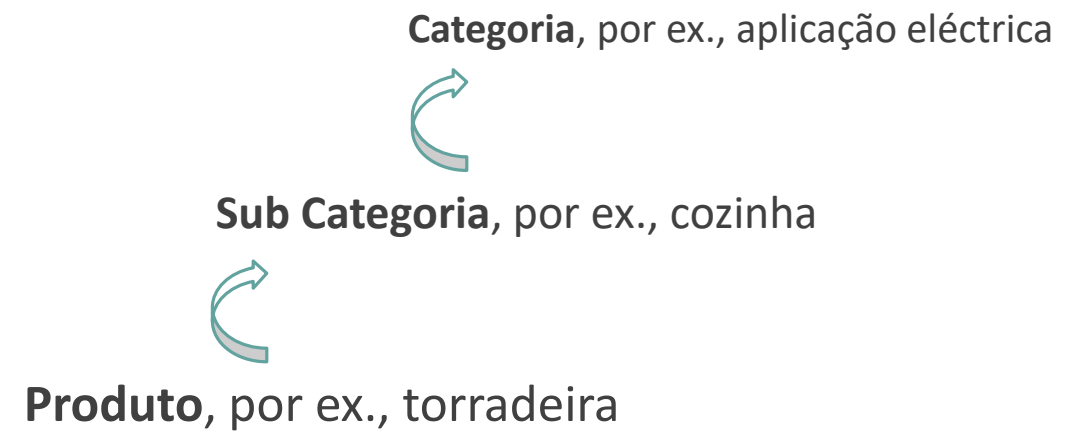
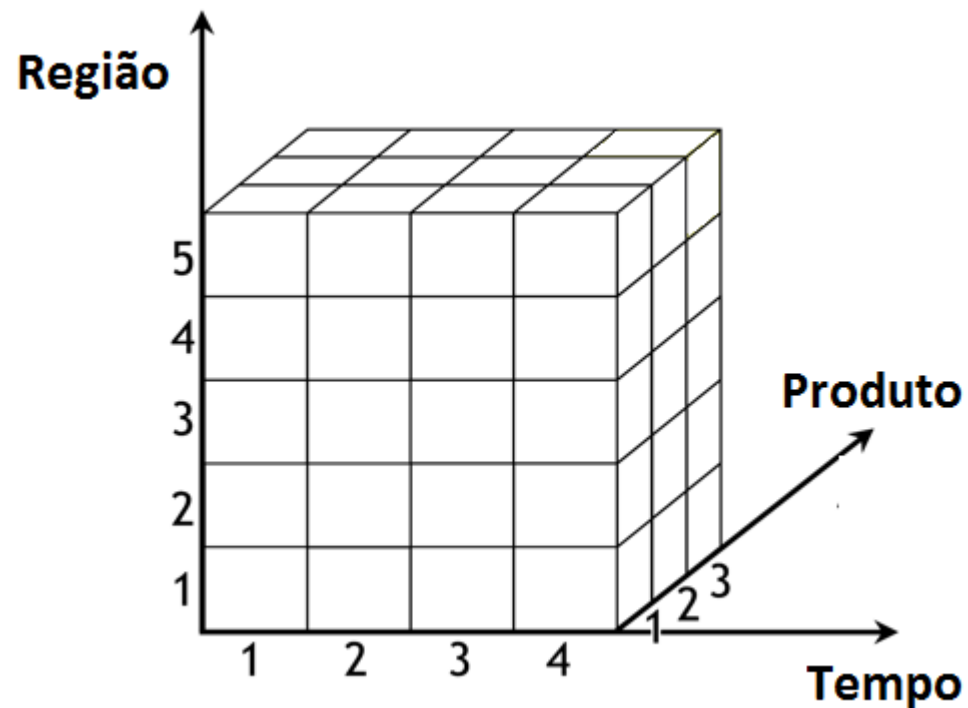
Sub Categoria, por ex., cozinha



Produto, por ex., torradeira

Operações OLAP

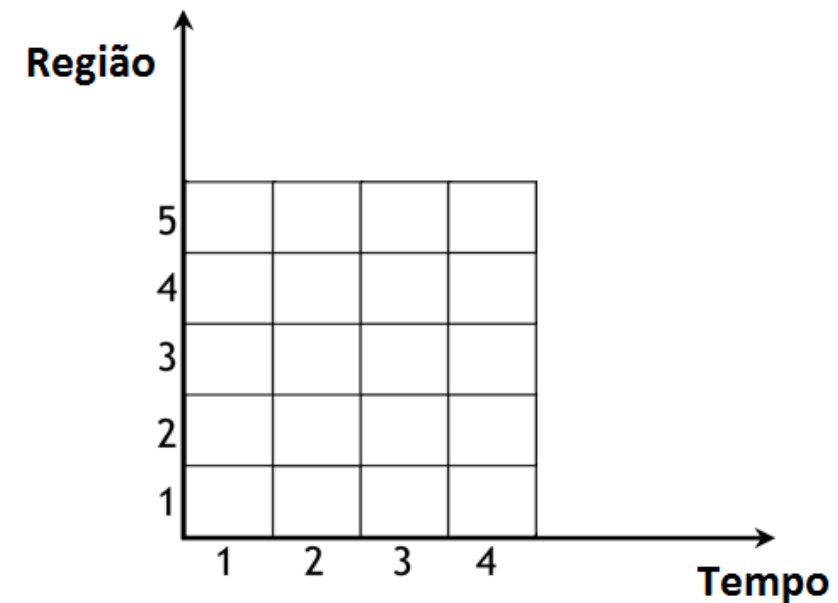
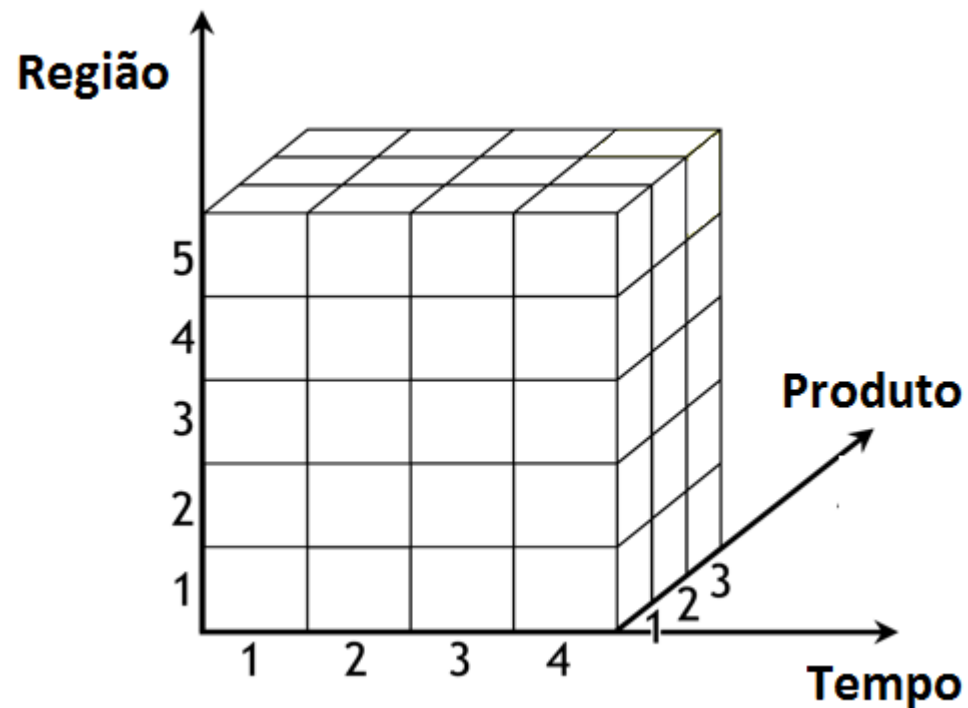
Generalizar



Operações OLAP

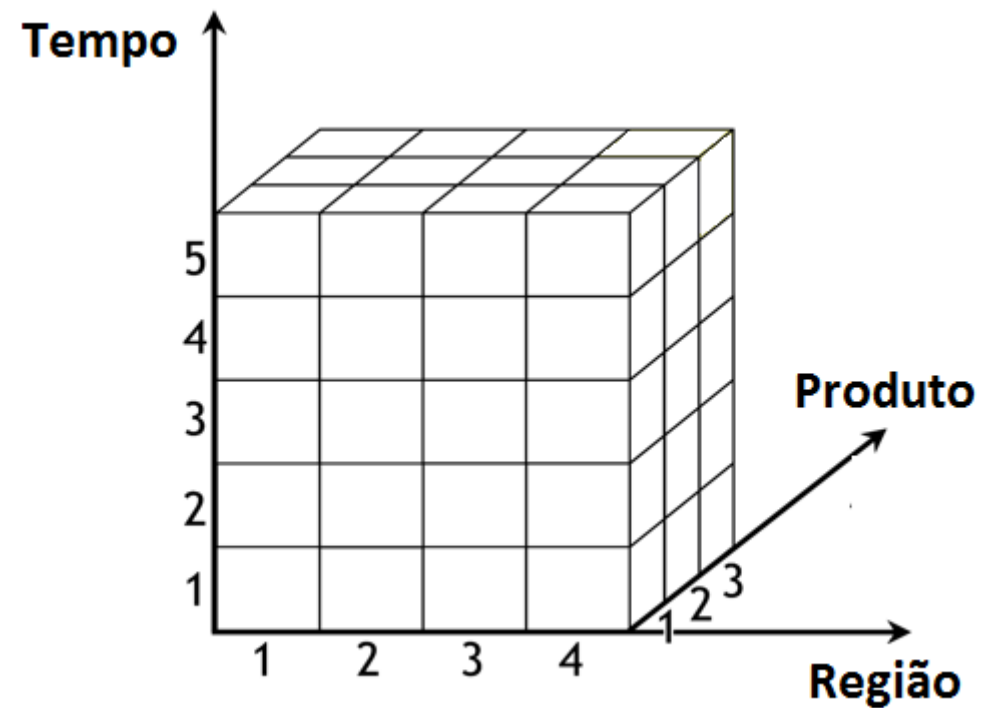
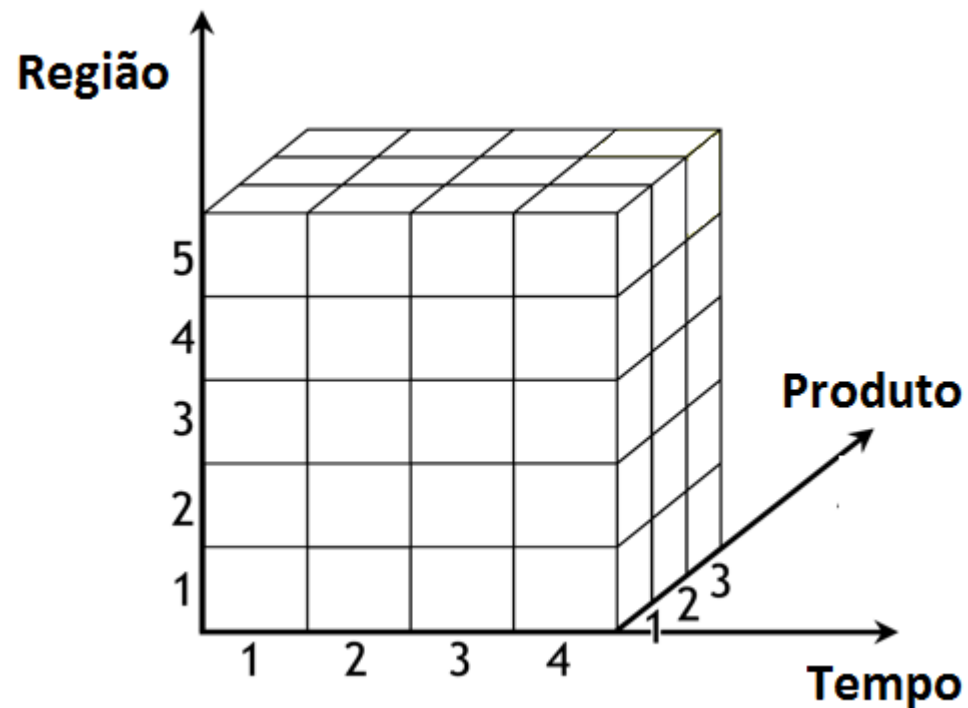
Slice and Dice

Produto = torradeira



Operações OLAP

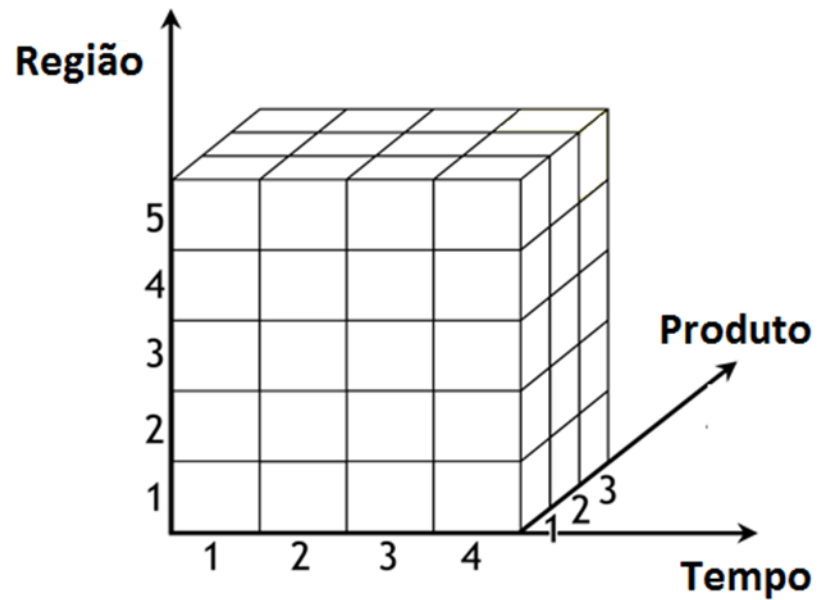
Pivot



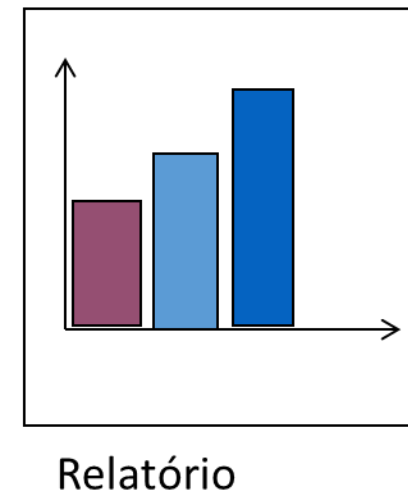
Servidor OLAP

- 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 multidimensionais.
- Existem os seguintes tipos de servidores OLAP:
 - Servidor MOLAP
 - Servidor ROLAP
 - Servidor HOLAP

Apresentação

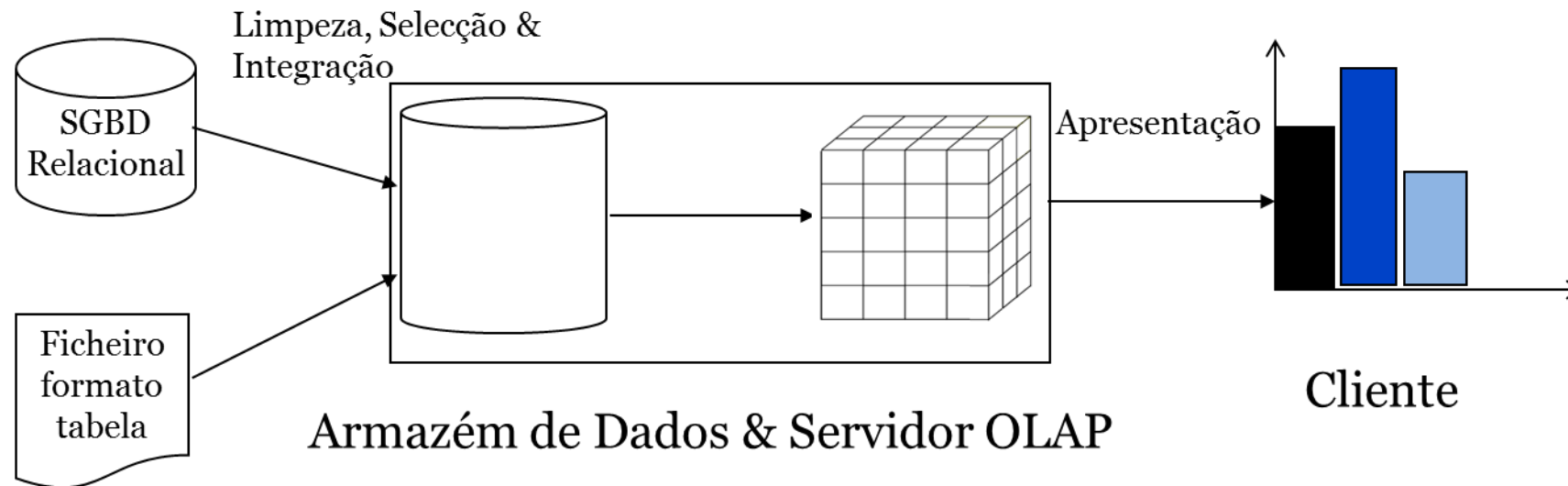


Ferramenta
para geração
de relatórios



Armazenamento de Dados inclui

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



Necessidade de Armazenamento de Dados

- 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 corretas.

Necessidade de Armazenamento de Dados

- Numa perspetiva de negócio:
 - é uma ferramenta de 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

- 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)
 - Pentaho (software livre)
- Reporting tools
 - DataWarehouse Explorer (para SQL Server)
 - OLAP Reporting Tool (para MS Excel)
 - IBM Cognos 8 Business Intelligence