



Unidade 24 – Data Warehouse



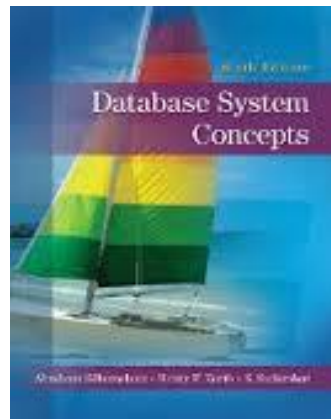
Prof. Aparecido V. de Freitas
Doutor em Engenharia
da Computação pela EPUSP



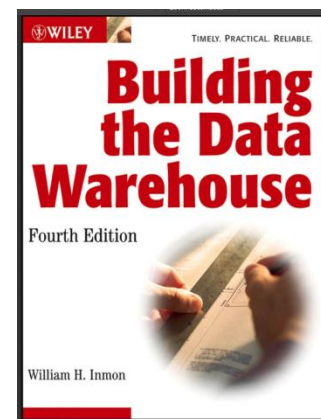
Bibliografia



Sistemas de Banco de Dados
Elmasri / Navathe 6ª edição



Sistema de Banco de Dados
Korth, Silberschatz – Sixth Edition



Building the Data Warehouse –
William H. Inmon – Fourth Edition



Projetando Sistemas de apoio
à decisão baseados em Data
Warehouse – Methanias C. Jr.
Axcel Books



Sistemas de Apoio a Decisão



✓ Informação

- Melhor recurso do qual uma empresa pode dispor para tomar decisões estratégicas;
- Obtida analisando dados históricos sobre vendas, produção, clientes, etc.

✓ Análise dos dados

- Fornecem informações vitais para as organizações;
- Podem aumentar a competitividade de uma organização;
- Era feita intuitivamente pelo grupo de staff de uma organização.





Sistemas de Apoio a Decisão



✓ Dificuldades para obter informação

- Quantidade de dados a serem analisados cresce com a expansão do negócio e com o passar dos anos;
- Dados conflitantes vindos de fontes diferentes podem gerar informações desencontradas;
- Extrema dificuldade para se manter e analisar dados de forma manual;
- Informação não é mais mantida apenas pelo grupo de staff da organização.





Sistemas de Apoio a Decisão (DSS)

- DSS – Decision Support Systems
- Também conhecidos por EIS – Executive Information Systems;
- Usam dados históricos mantidos em um banco de dados convencional;
- Dados históricos são analisados por meio de técnicas de mineração de dados para se obter informações usadas na tomada de decisões;
- Estatísticas de venda, produção, clientes, etc. podem ser levantadas e consideradas para tomar **decisões estratégicas de negócio**.





Sistemas de Apoio à Decisão (DSS)

✓ Benefícios

- Determinar o mercado-alvo de um produto;
- Definir o preço de um produto, criar promoções e condições especiais de compra;
- Verificar a eficácia de campanhas de marketing;
- Otimizar a quantidade de produtos no estoque;
- Responder rapidamente a mudanças no mercado e determinar novas tendências;

... **ou seja, ganhar eficiência e lucratividade.**





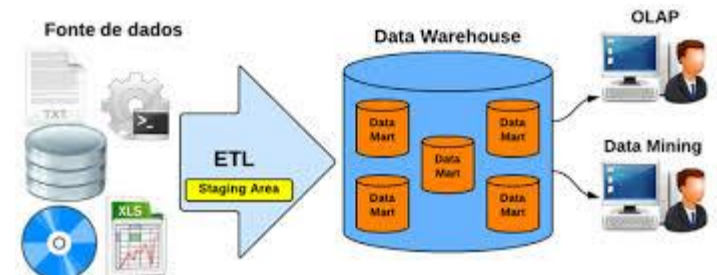
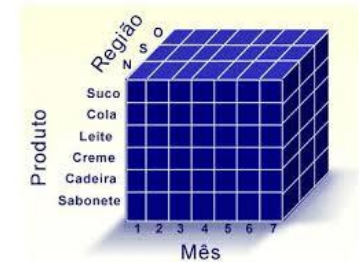
Sistemas de Apoio a Decisão (DSS)

✓ Problema

- Dados históricos não são mantidos nos BDs transacionais;
- Volume de dados seria muito grande;
- Desempenho seria insatisfatório;

✓ Solução

- criar um BD exclusivamente para manter os dados históricos;
- Especializado para realizar consultas em grandes volume de dados;
- Surge o Data Warehouse (DW).





Data Warehouse (DW)

✓ Histórico

- Criado pela **IBM** na década de 60 com o nome Information Warehouse;
- Relançado diversas vezes sem grande sucesso;
- O nome **Data Warehouse** foi dado por **William H. Inmon**, considerado o pai desta tecnologia;
- Tornou-se viável com o surgimento de novas tecnologias para armazenar e processar uma grande quantidade de dados.





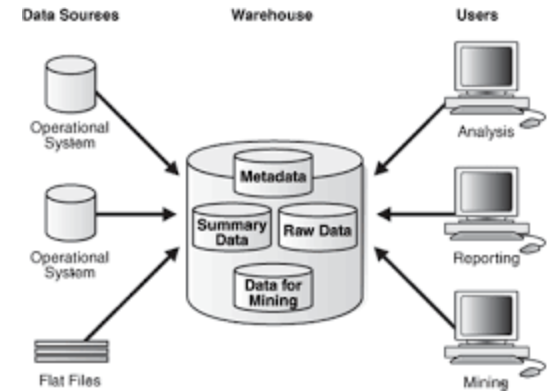
Introdução

- ✓ Um Data Warehouse corresponde a um armazém de dados;
- ✓ Oferecem armazenamento, funcionalidade e responsividade às consultas além das capacidades dos bancos de dados orientados à transação;
- ✓ Bancos de dados tradicionais são transacionais (relacionais, hierárquicos, redes);
- ✓ Data Warehouses têm a característica principal de servir para aplicações de apoio à decisão, usados pelo staff das organizações.
- ✓ Data Warehouses oferecem acesso aos dados para análise complexa, descoberta de conhecimento e tomada de decisão (Inmon, 1992);
- ✓ Aplicações que se utilizam de Data Warehouses são OLAP, DSS e mineração de dados.





Data Warehouse (DW)

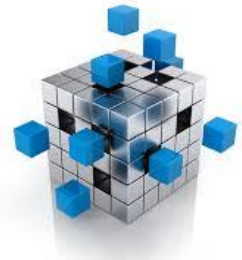


✓ O que é?

- Sistema que armazena **dados históricos** usados no processo de tomada de decisão;
- Integra os dados corporativos de uma empresa em um único **repositório**;

✓ Para que serve?

- Para criar uma visão única e centralizada dos dados que estavam dispersos em diversos Bancos de Dados;
- Permite que usuários finais executem **consultas**, gerem **relatórios** e façam **análises**.





Data Warehouse (DW)

- ✓ Bancos de dados usados nas aplicações transacionais são chamados Bancos de Dados operacionais ou Banco de Dados Transacionais;
- ✓ Bancos de dados tradicionais oferecem suporte para o processamento on-line de transações (**OLTP - On-line Transaction Processing**);
- ✓ **DW** é um BD informacional alimentado com dados dos Bancos de Dados operacionais da empresa ;
- ✓ Disponibiliza dados históricos e são separados lógica e fisicamente do ambiente de produção da organização, concebido para armazenar dados extraídos desse ambiente. ;
- ✓ Dados podem ser sumarizados (condensados) para que sejam processados em aplicações de análise de dados;
- ✓ Contém também metadados (dados que descrevem os dados armazenados em um DW).





Então o Data Warehouse é apenas um BD que contém também dados históricos?





Data Warehouse (DW)

- ✓ Para que seja considerado um Data Warehouse, um banco de dados deve:
 - Coletar dados de várias fontes;
 - Dados coletados devem ser transformados para que haja uma visão única dos dados;
 - Dados devem ser usados por aplicativos focados em apoio à decisão.





Data Warehouse (DW)

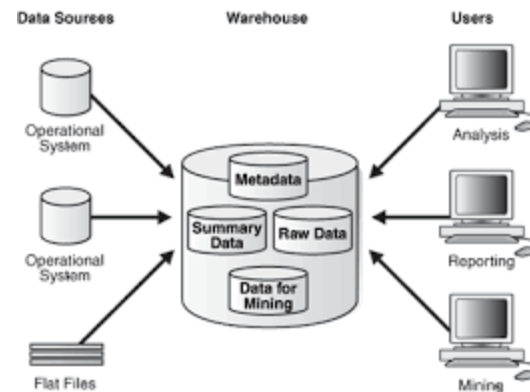
	BD Operacional	Data Warehouse
Usuários	Funcionários	Alta administração
Utilização	Tarefas cotidianas	Decisões estratégicas
Padrão de uso	Previsível	Difícil de prever
Princípio de funcionamento	Com base em transações	Com base em análise de dados
Valores dos dados	Valores atuais e voláteis	Valores históricos e imutáveis
Detalhamento	Alto	Sumarizado
Organização dos dados	Orientado a aplicações	Orientado a assunto



Data Warehouse – Principais Características

✓ De acordo com a definição dada por **Inmon**, um Data Warehouse deve ser:

- **Orientado a assunto;**
- **Integrado;**
- **Não-volátil;**
- **Variável com o tempo.**

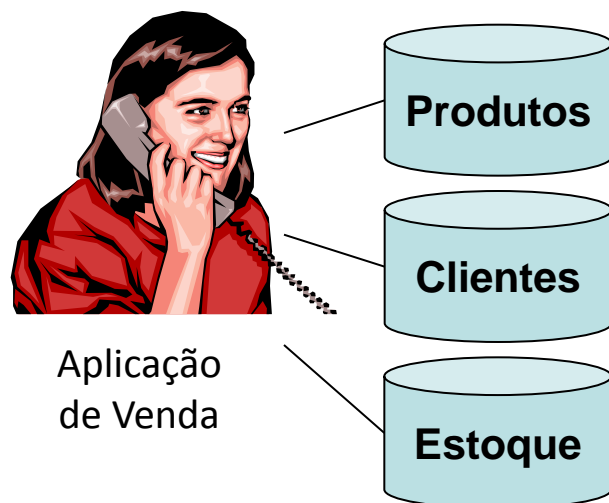




Data Warehouse – Principais Características

✓ Orientação a assunto

- Os dados em um DW são organizados de modo a facilitar a análise dos dados;
- Dados são organizados por assunto e não por aplicação, como em BDs operacionais.





Data Warehouse – Principais Características

✓ Orientação a assunto

- Para cada assunto (por exemplo, cliente) pode haver várias tabelas;
- Por exemplo, tabela com informações gerais (nome, endereço, email, telefone), outra com os clientes que tiveram contas superiores a R\$ 10.000,00;
- Além disso, podem haver tabelas cumulativas com os clientes que mais consumiram entre 2012 e 2016;
- Assim, para o mesmo assunto, pode haver vários níveis de detalhamento;



Análise
de Vendas

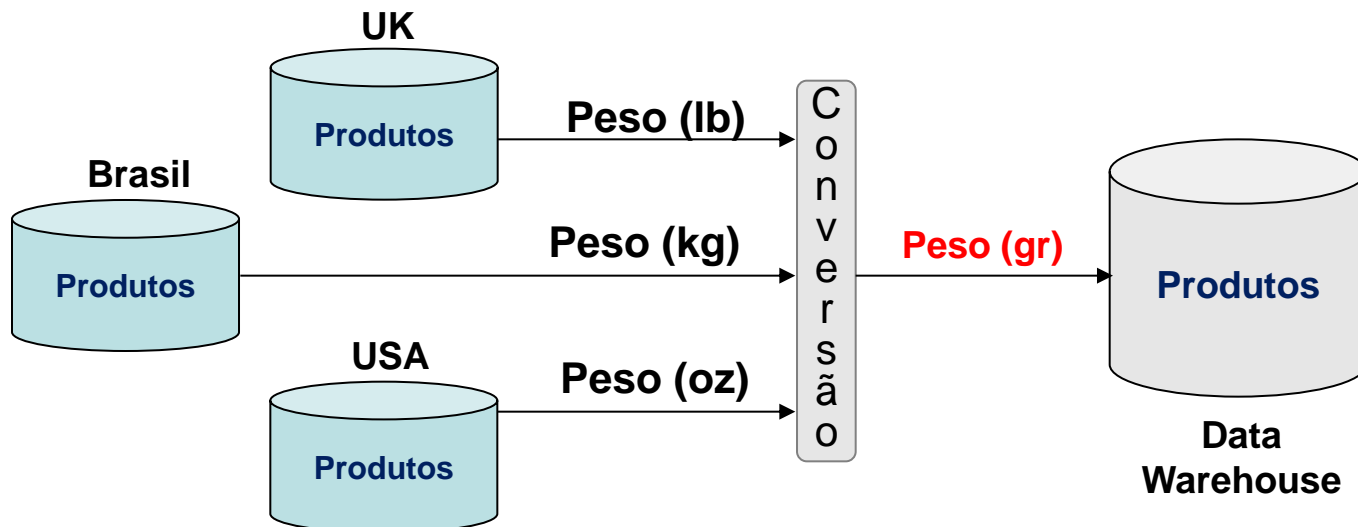




Data Warehouse – Principais Características

✓ Integração

- Dados de um **DW** provém de diversas fontes;
- Dados podem ser sumarizados ou eliminados;
- Formato dos dados deve ser padronizado para uniformizar nomes, unidades de medida, etc.





Data Warehouse – Principais Características

✓ Não-Volátil

- Dados não são mais alterados depois de incluídos no **DW** (carga);
- Novos dados são absorvidos, integrando-se aos dados já existentes;
- **DW** é otimizado para inclusão de dados (carga) e consultas;
- Operações no **DW**:
 - ✓ Em um **BD** operacional é possível incluir, alterar e eliminar dados;
 - ✓ Já no **DW** é possível apenas incluir dados;
- Garante que consultas subsequentes a um dado produzirão o mesmo resultado.

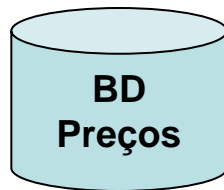




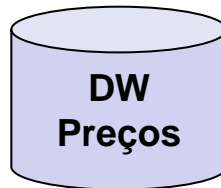
Data Warehouse – Principais Características

✓ Variável com o Tempo

- Os dados no **DW** são relativos a um determinado instante de tempo.



Produto	Preço
Caneta Azul	0,50
Lápis Preto	0,30
...	...

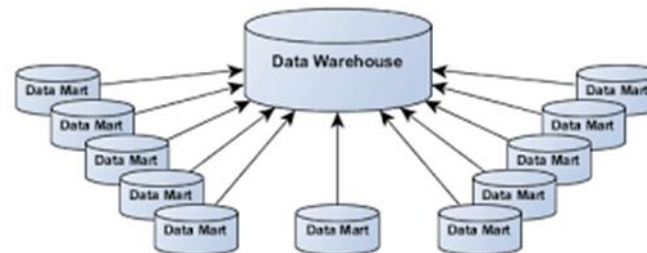
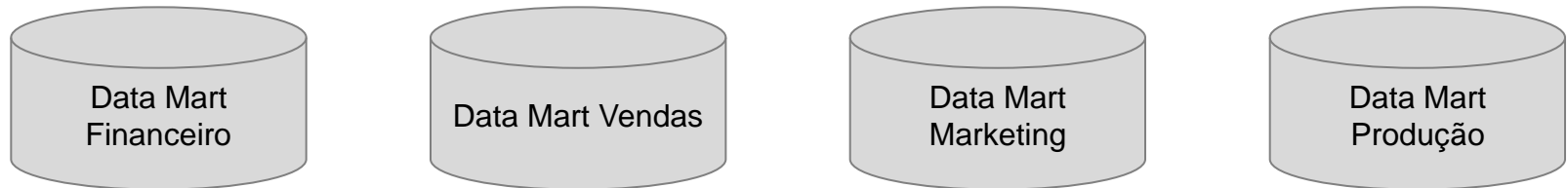


Produto	Jan/03	Fev/03	Mar/03
Caneta Azul	0,40	0,45	0,50
Lápis Preto	0,25	0,28	0,30
...



Data Marts

- ✓ Dados separados por assunto em subconjuntos de acordo com:
 - A estrutura interna da empresa;
 - O processo de tomada de decisão;
- ✓ Estes subconjuntos dos dados são chamados de **Data Marts.**



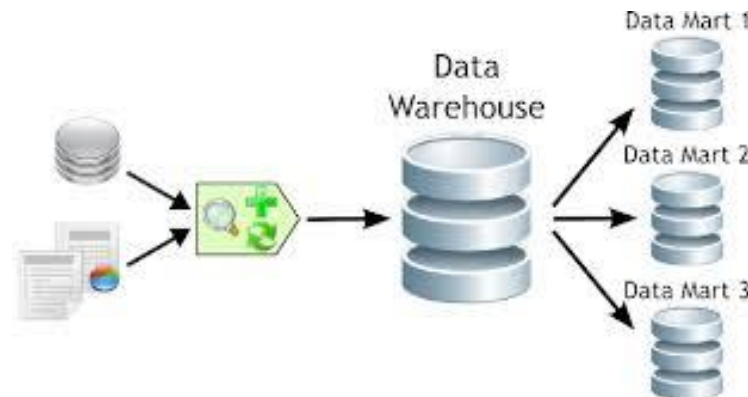


Qual a utilidade de um Data Mart?



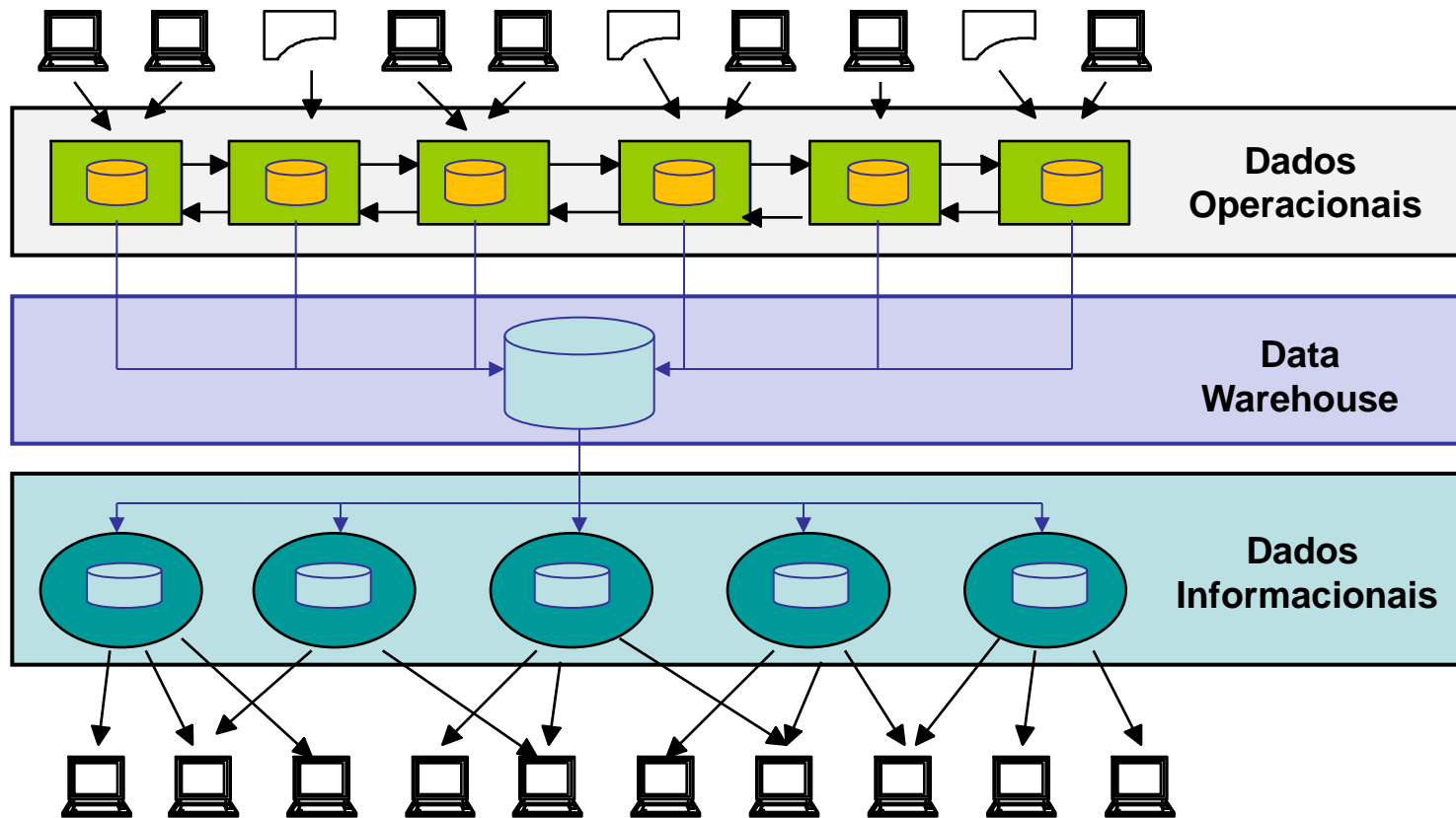
Data Marts

- ✓ Dados devem ser segregados para melhorar o **desempenho** do ponto de vista de usuário;
- ✓ Deve existir uma cópia dos dados onde só usuários autorizados devem ter **privilégios** de acesso;
- ✓ Diferentes usuários são **responsáveis** por diferentes Data Marts.



Data Warehouse – Arquitetura

✓ Sistemas baseados em Data Warehouse [Orr]





Data Warehouse – Arquitetura

✓ Principais tarefas efetuadas pelo DW

- Obter dados dos **BDs** operacionais e externos;
- Armazenar os dados;
- Fornecer informações para tomada de decisão;
- Administrar o sistema e os dados;

✓ Principais componentes do DW

- Mecanismos para acessar e transformar dados;
- Mecanismo para armazenamento de dados;
- Ferramentas para análise de dados;
- Ferramentas de gerência.





Data Warehouse – Estrutura Interna

✓ Requisitos do DW

- **Eficiente**
 - Grande volume de dados imutáveis;
 - Processamento paralelo e/ou distribuído;
- **Confiável**
 - Funcionamento do sistema;
 - Resultado das análises;
- **Expansível**
 - Crescente volume de dados;
 - Maior número de fontes de dados;

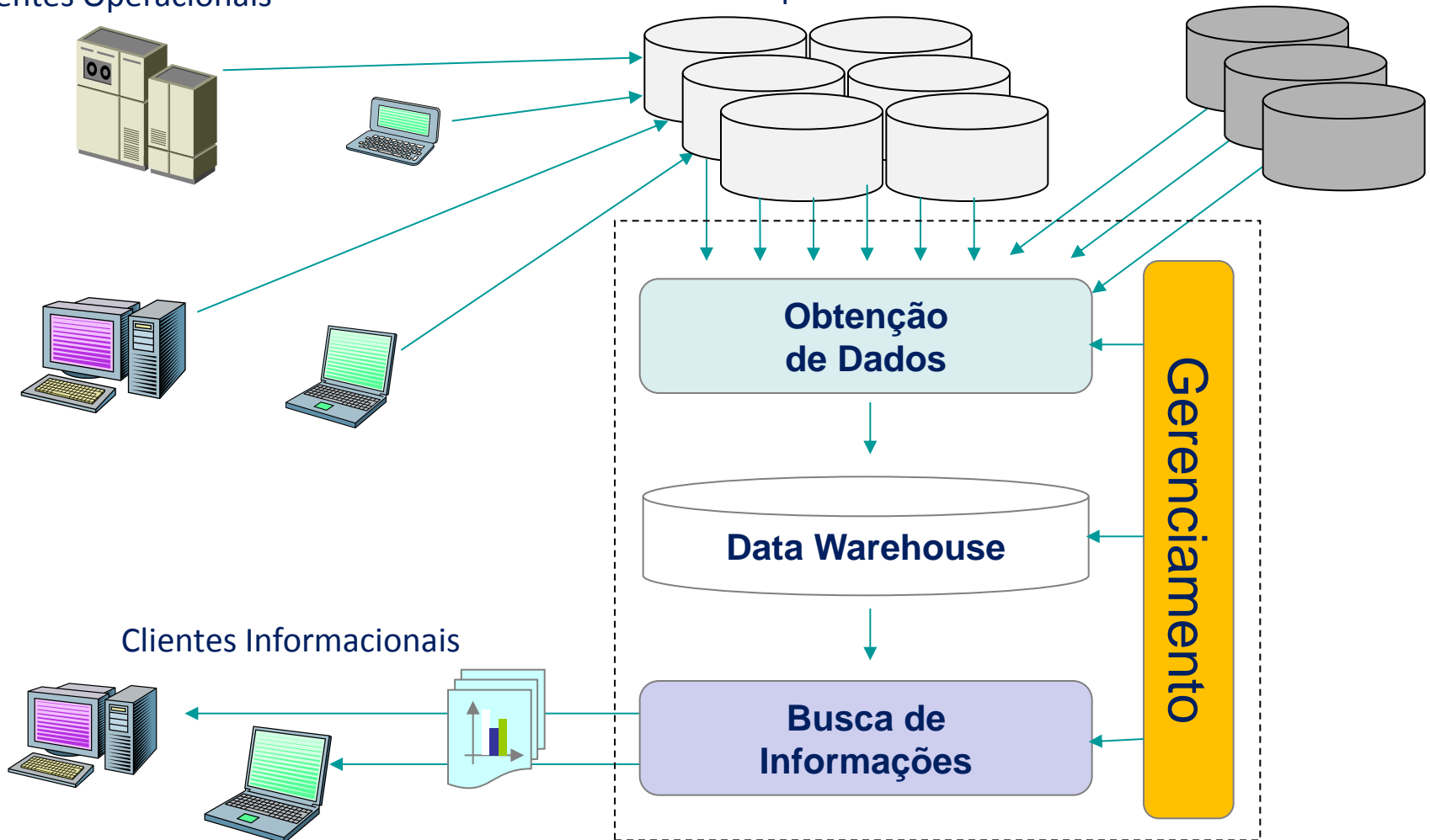


Data Warehouse – Estrutura Interna

Clientes Operacionais

BDs Operacionais

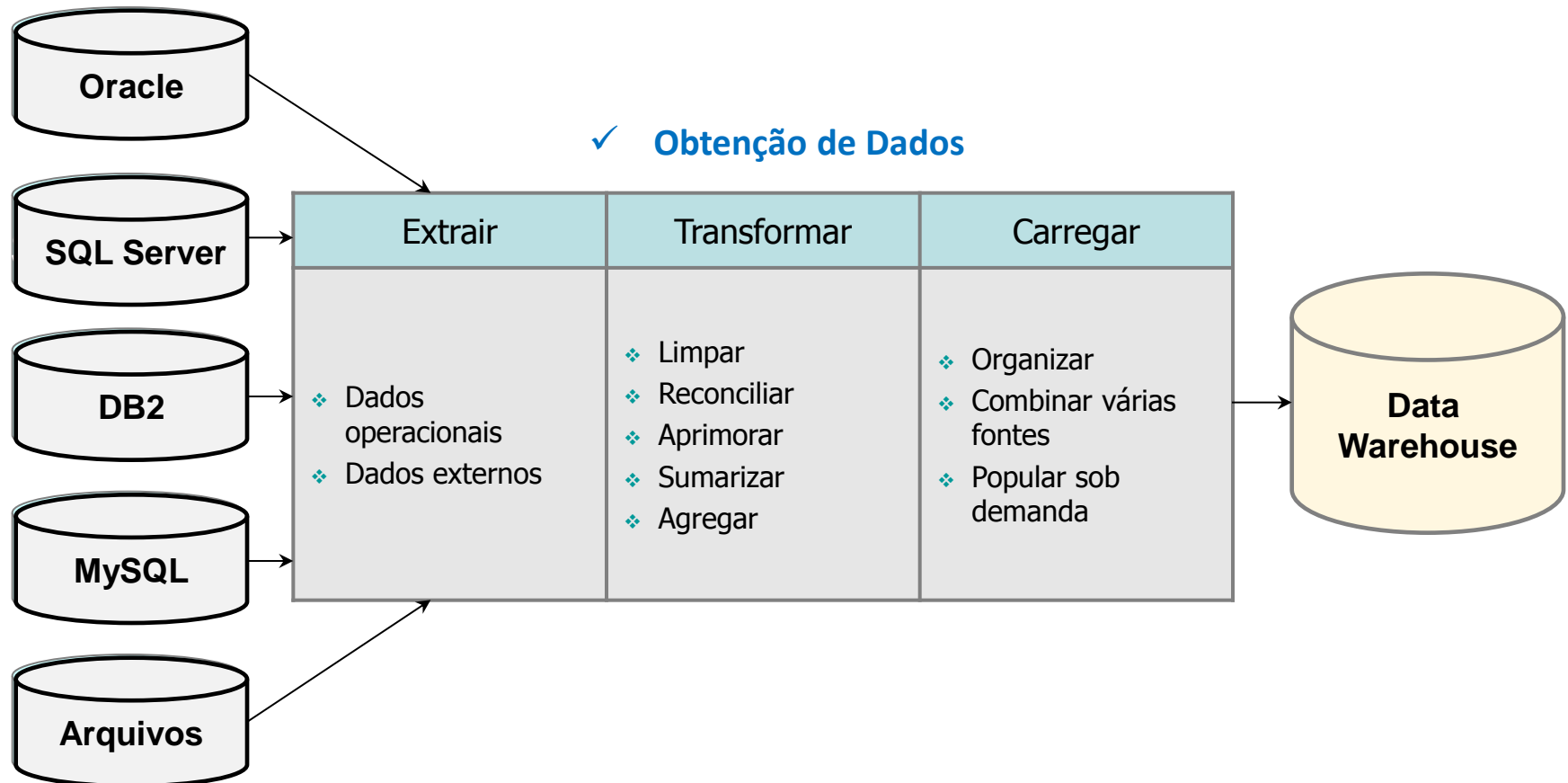
Fontes Externas





Data Warehouse – ETL – Extraction, Transform and Load

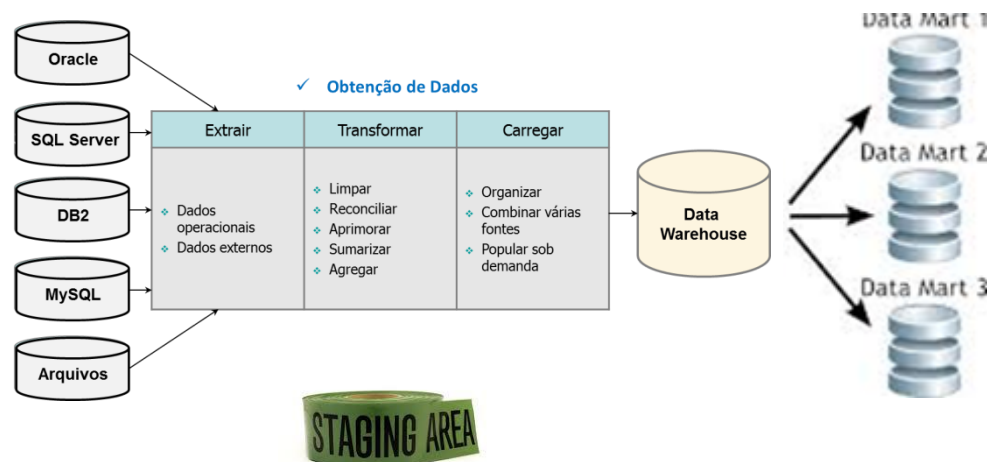
- ✓ Em projetos DW, processos **ETL** consomem mais de 70% do tempo de desenvolvimento;
- ✓ Processo desenvolvido especificamente para cada empresa.





Data Warehouse – Processo ETL

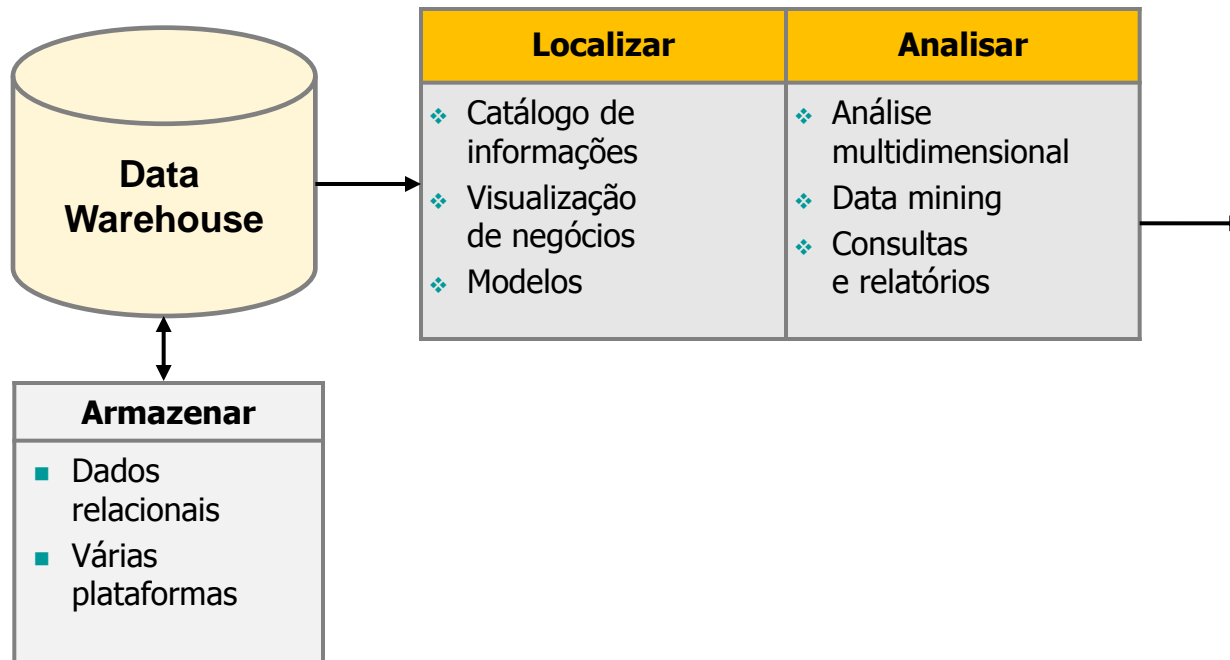
- ✓ Fluxo de dados se inicia nas aplicações fontes e passa por uma área intermediária de armazenamento chamada **Staging Area** (área de estágio);
- ✓ Nessa área os dados sofrem limpeza, integração e depois são exportados para o **DW**;
- ✓ Estas ações constituem a tarefa mais crítica na geração de um DW;
- ✓ Em seguida à carga do **DW**, os dados são transmitidos para os **Data Marts**.





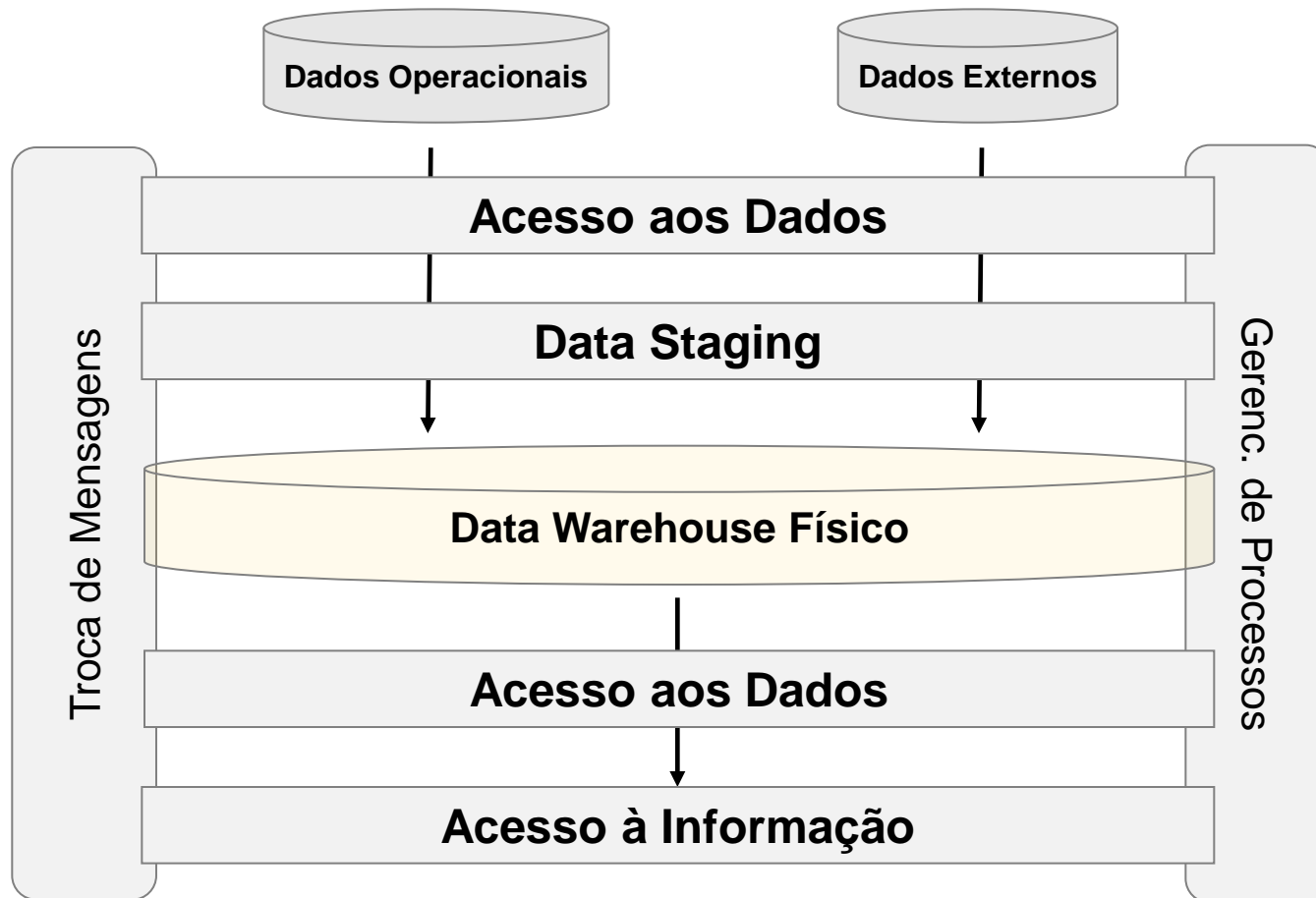
Data Warehouse – Consultas

✓ Busca de Informações





Data Warehouse – Estrutura Interna



Data Staging: Área intermediária para tratamento de dados



Data Warehouse – Estrutura Interna

✓ Funções das Camadas do DW

- Dados Operacionais/Externos: **Fontes de Dados**;
- Acesso aos Dados: **extrair** dados dos BDs;
- **Data Staging**: transformar e carregar dados;
- Data Warehouse Físico: **armazenar** dados;
- Acesso aos Dados: **localizar** dados para análise;
- Acesso à Informação: **analisar** dados;
- Troca de Mensagens: **transportar** dados;
- Gerenciamento de Processos: **controlar** atividades.

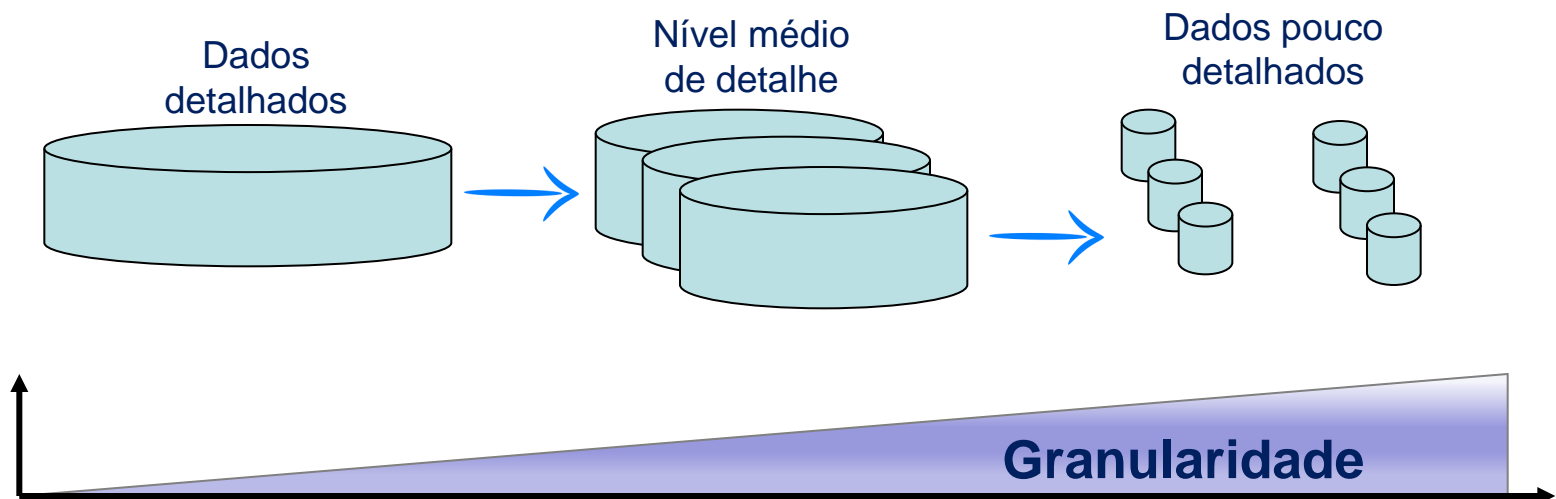




Data Warehouse – Granularidade

✓ Granularidade

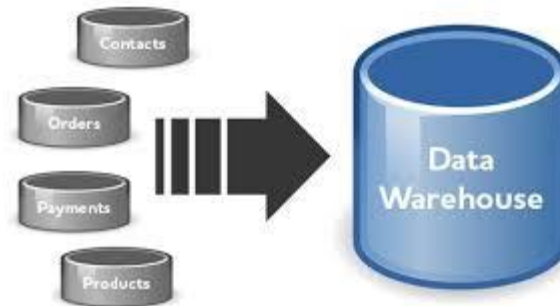
- Nível de detalhe dos dados;
- De extrema importância no projeto do DW.





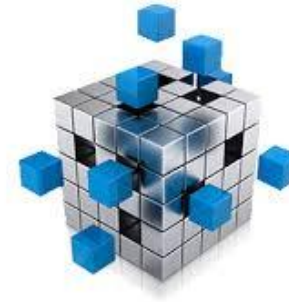
Data Warehouse – Granularidade

- ✓ Definir a granularidade adequada é vital para que o **DW** atenda seus objetivos;
 - **Mais detalhes** → Mais dados → Análise mais longa → Informação mais detalhada;
 - **Menos detalhes** → Menos dados → Análise mais curta → Informação menos detalhada;
- ✓ Para evitar que se perca informação são criados vários níveis de granularidade.





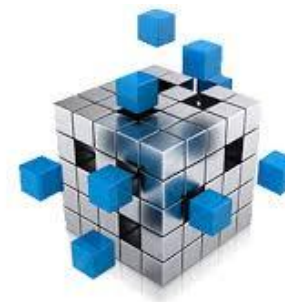
OLAP



- ✓ **OLAP** (On-line Analytical Processing) é um termo que se usa para descrever a análise de dados;
- ✓ Conjunto de técnicas utilizadas para se tratar a informação contida em um **DW**;
- ✓ Ferramentas **OLAP** utilizam capacidades de computação para análises que exigem mais armazenamento e poder de processamento .
 - ✓ Aplicações de **Finanças**: Orçamento, Análise de Balanço, Fluxo de Caixa, Contas a Receber;
 - ✓ Aplicações de **Vendas**: Análise de vendas, Previsões, Análise de Canais de Distribuição;
 - ✓ Aplicações de **Marketing**: Análise de Preço/Volume, Análise de Mercados;
 - ✓ Aplicações de **RH**: Análise de Benefícios, Projeção de Salários;
 - ✓ Aplicações de **Manufatura**: Gerência de Estoque, Cadeia de Fornecimento, Planejamento de Demanda, Análise de custos de matéria-prima.



OLAP x OLTP

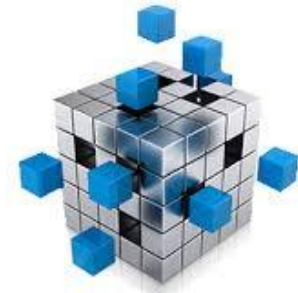


- ✓ Bancos de dados desenvolvidos para **OLTP**, em geral, são inapropriados para **OLAP**;
- ✓ Banco de dados **OLTP** não representam repositórios de fatos e dados históricos;

OLAP	OLTP
Relevância para dados históricos;	Mantém usualmente a situação corrente;
Necessidade de ver o dado sob diferentes perspectivas;	Voltado para automação de funções repetidas;
Atualizações quase inexistentes, apenas novas inserções;	Atualizações em grande número;
Baseado em dados históricos, consolidados e frequentemente totalizados;	Baseado em transações;



Características dos Sistemas OLAP



- ✓ A principal característica dos sistemas **OLAP** é permitir uma visão conceitual multidimensional dos dados de uma organização;
- ✓ Os dados são modelados numa estrutura chamada **CUBO**, no qual cada dimensão representa os assuntos (temas), como por exemplo, produto, cliente e tempo;
- ✓ Geralmente pode-se efetuar operações de filtros (**Ranking**), alteração da forma de visualização dos dados (**Rotation**), sumarização de informações (**Drill Up**), etc

