

Motivação e Conceitos Básicos

Processamento Analítico de Dados
(em Larga Escala)

Profa. Dra. Cristina Dutra de Aguiar

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)**

Acesso às Informações

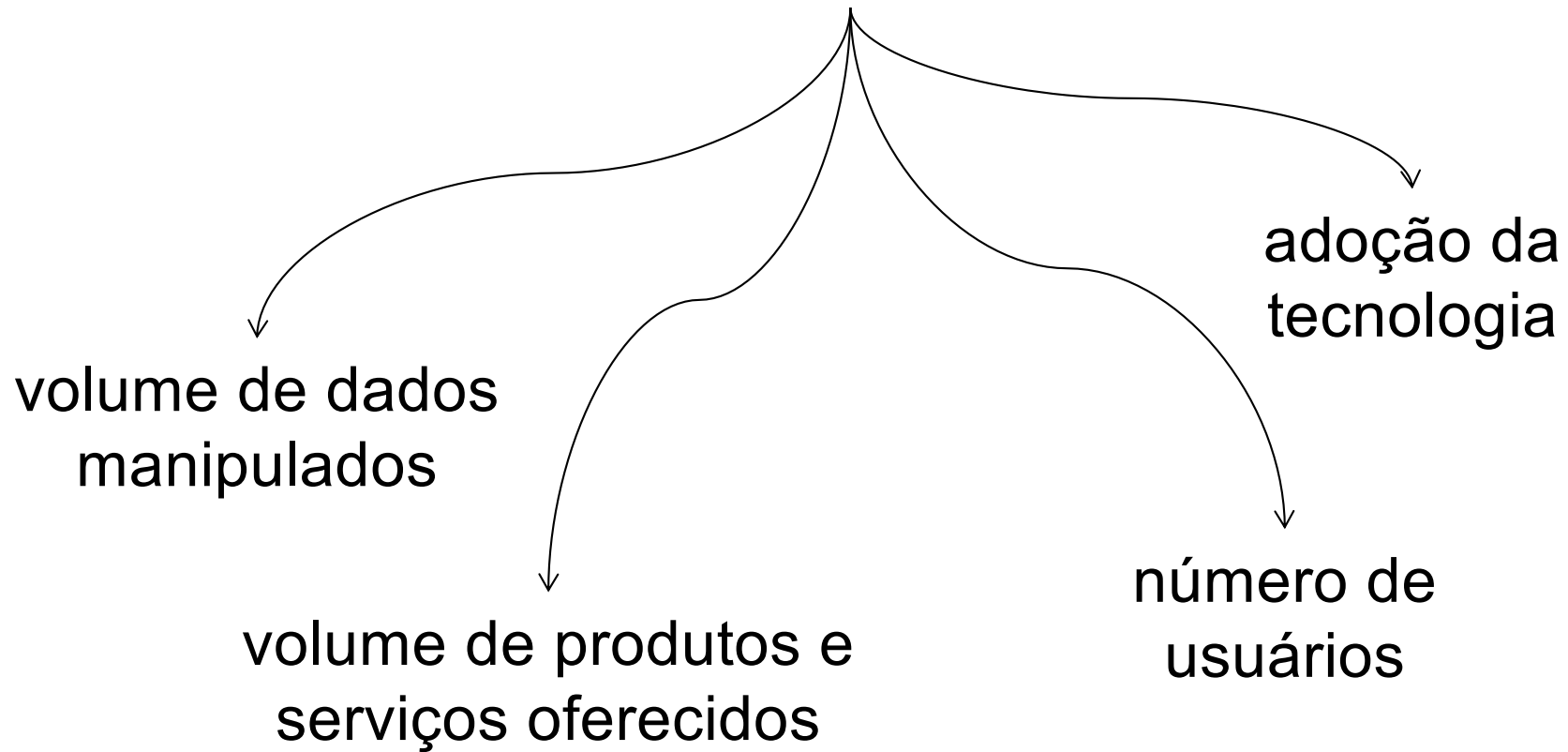
- Duas etapas
 - a informação de cada provedor é extraída previamente, devendo ser traduzida, filtrada, integrada à informação relevante de outros provedores e finalmente armazenada no **DW**
 - as consultas, quando realizadas, são executadas diretamente no **DW**, sem acessar os provedores de informação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

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 data marca nos últimos 3 anos, de acordo com as promoções de Natal?*

Visão do Mercado

Crescimento explosivo do uso da tecnologia de data warehousing



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.

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 (consultas analíticas)

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

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

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

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

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