

# Business Intelligence (BI) No Olhar da Representação do Conhecimento Organizacional

Palestrante: Eduardo Manuel de Freitas Jorge  
[emjorge1974@gmail.com](mailto:emjorge1974@gmail.com)

# Agenda

---

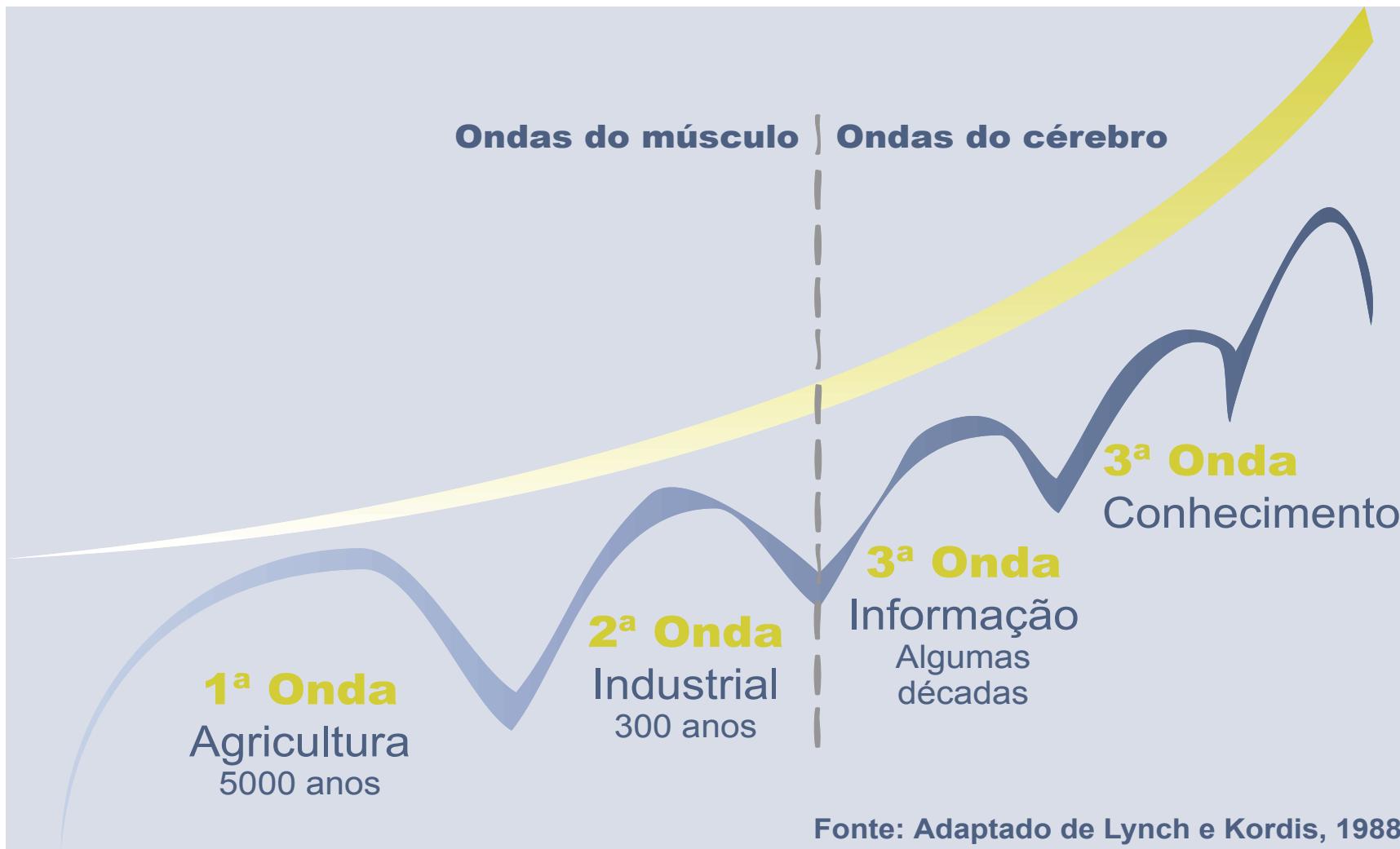
- ▶ O que é Conhecimento?
- ▶ O que envolve a Gestão do Conhecimento Organizacional?
- ▶ Qual a motivação para construção de Business Intelligence (BI)?
- ▶ Fatores de Sucesso ou Insucesso no BI
- ▶ Framework tecnológico do Sistema de Gestão do Conhecimento: Data Warehouse
  - ▶ BI: Na Perspectiva do Projeto do Data Warehouse
  - ▶ Framework tecnológico do Sistema de Gestão do Conhecimento: Mineração de Dados
- ▶ Quanto custa uma solução de BI?





# O que é Conhecimento?

# O que é Conhecimento?



# O que é Conhecimento?

## Dados, Informação e Conhecimento

Dado	Informação	Conhecimento
Simples observações sobre o estado do mundo	Dados dotados de relevância e propósito	Informação valiosa da mente humana inclui reflexão, síntese, contexto
<ul style="list-style-type: none"><li>* Facilmente estruturado</li><li>* Facilmente obtido por máquinas</li><li>* Freqüentemente quantificado</li><li>* Facilmente transferível</li></ul>	<ul style="list-style-type: none"><li>* Requer unidade de análise</li><li>* Exige consenso em relação ao significado</li><li>* Exige necessariamente a mediação humana</li></ul>	<ul style="list-style-type: none"><li>* De difícil estruturação</li><li>* De difícil captura em máquinas</li><li>* Freqüentemente tácito</li><li>* De difícil transferência</li></ul>

FONTE: Davenport, Prusak - Ecologia da informação -1998 - p.18



Thomas H. Davenport

# O que é Conhecimento?

---

- ▶ Gestor deve trabalhar com Inferências
- ▶ Inferências
  - ▶ Método baseado na Semiótica
- ▶ Inferências “adoção controlada de uma crença como consequência de um outro conhecimento” (PEIRCE, 1975)
- ▶ Indução, Dedução e Abdução
  - ▶ Dedução - Parte de uma premissa maior para uma menor.
  - ▶ Indução - Parte de uma premissa menor para uma maior.
  - ▶ A abdução ou hipótese é a adoção probatória da hipótese
  - ▶ [www.unimar.br/pos/semitotica/raciocinio.doc](http://www.unimar.br/pos/semitotica/raciocinio.doc)



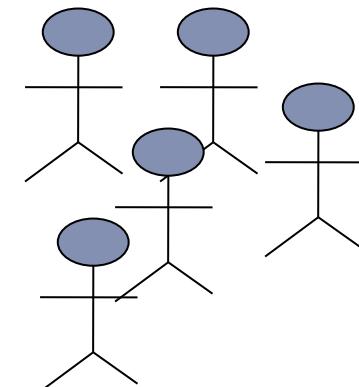
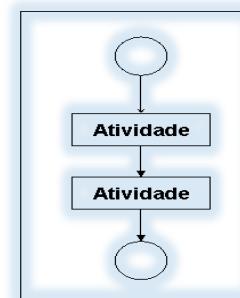
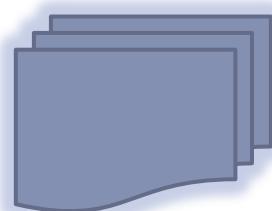
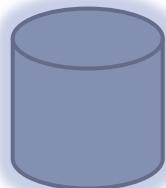


# O que envolve a Gestão do Conhecimento Organizacional?

# Definição Gestão do Conhecimento

“É uma disciplina que promove, com visão integrada, o **gerenciamento** e o **compartilhamento** de todo o ativo de informação possuído pela empresa. Esta informação pode estar em um banco de dados, documentos, procedimentos, bem como em pessoas, através de suas experiências e habilidades”.

Gartner Group



# Gestão do Conhecimento

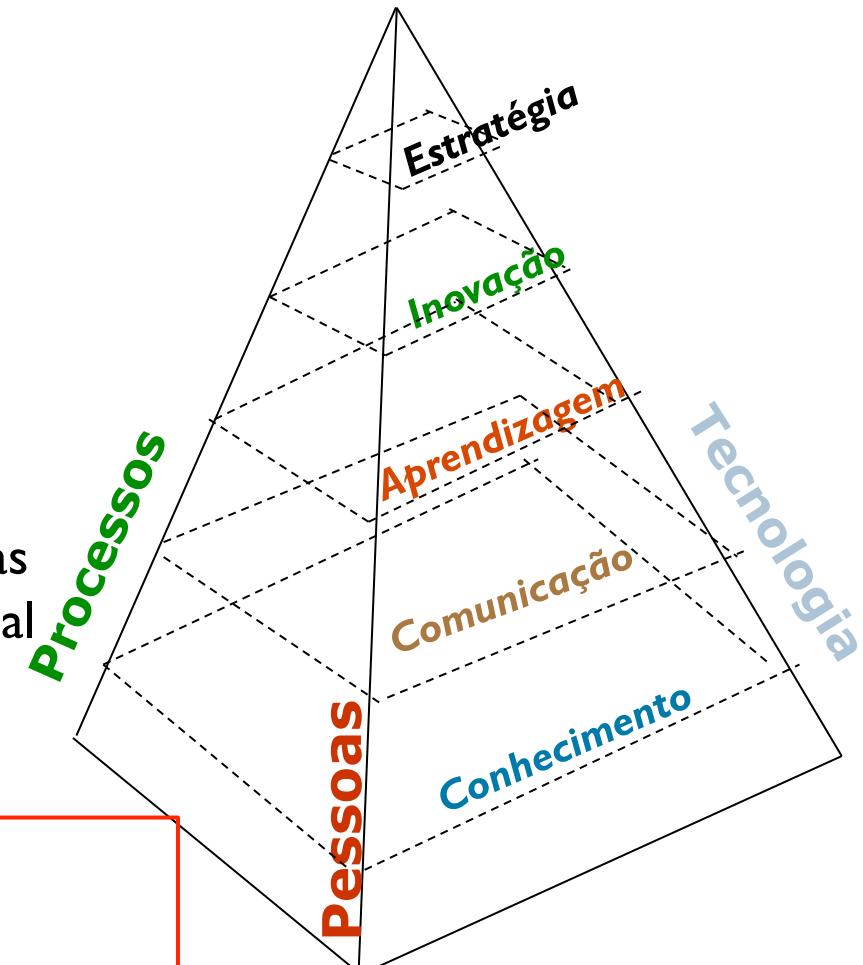
## ▶ Dimensões Estruturantes

- ▶ A GC organizacional deve contemplar três dimensões:

- ▶ **Processos**: organizam tarefas e atividades da organização;

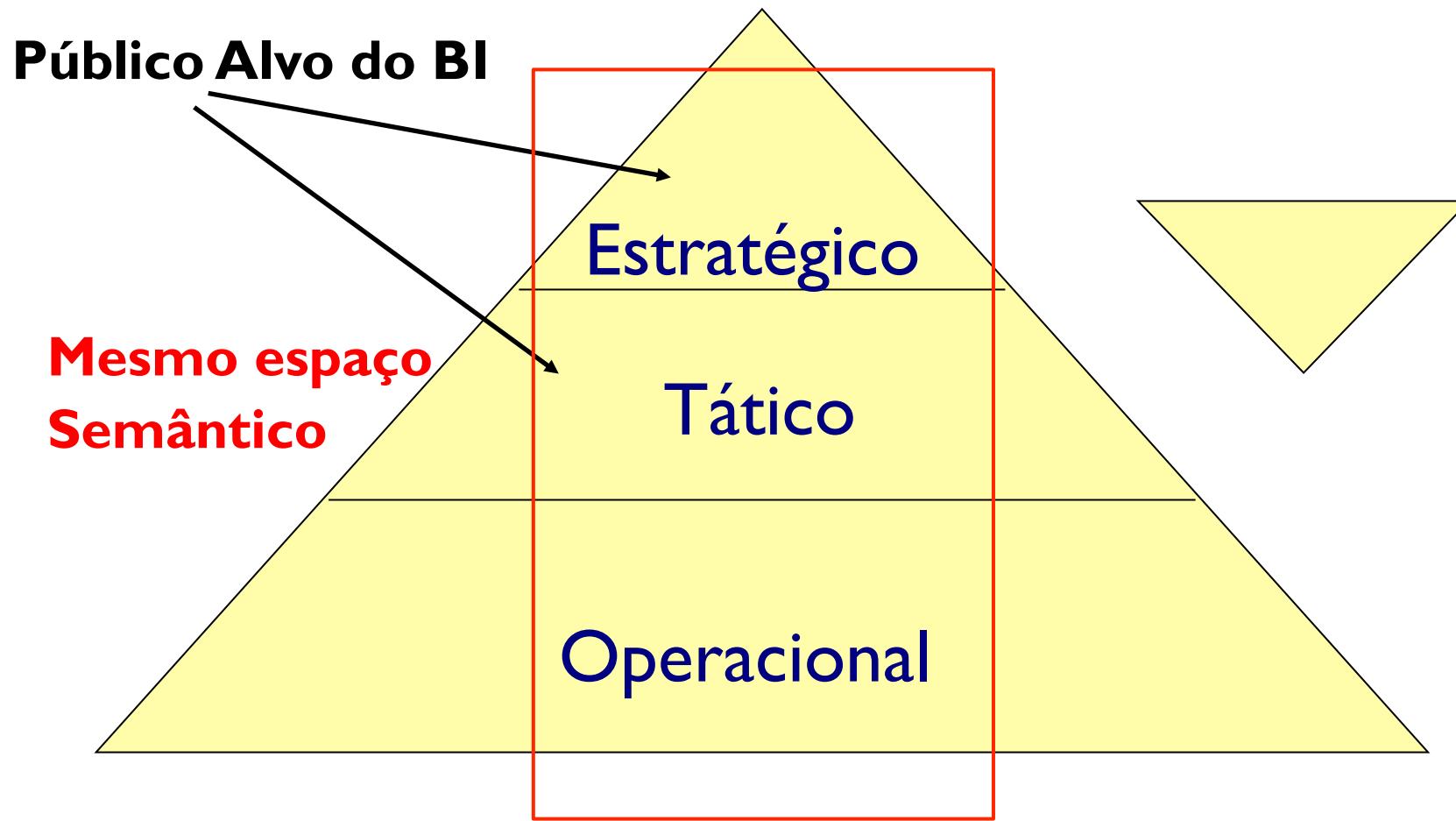
- ▶ **Pessoas**: congregam as competências (incluindo o conhecimento - principal fator de geração de valor organizacional);

- ▶ **Tecnologia**: deve servir aos processos e às pessoas

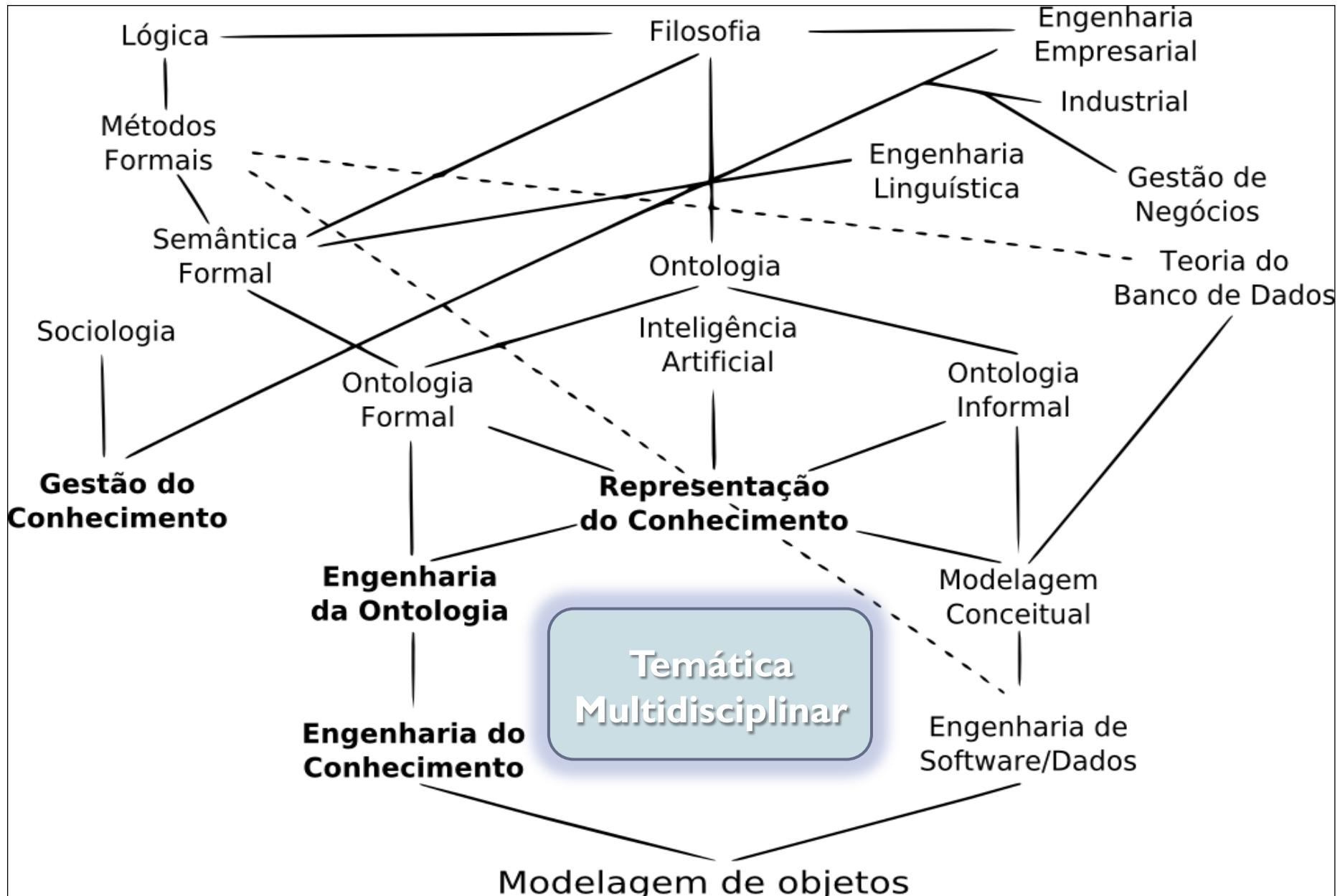


Pacheco, 2008

# Gestão do Conhecimento



► **Pirâmide Organizacional**



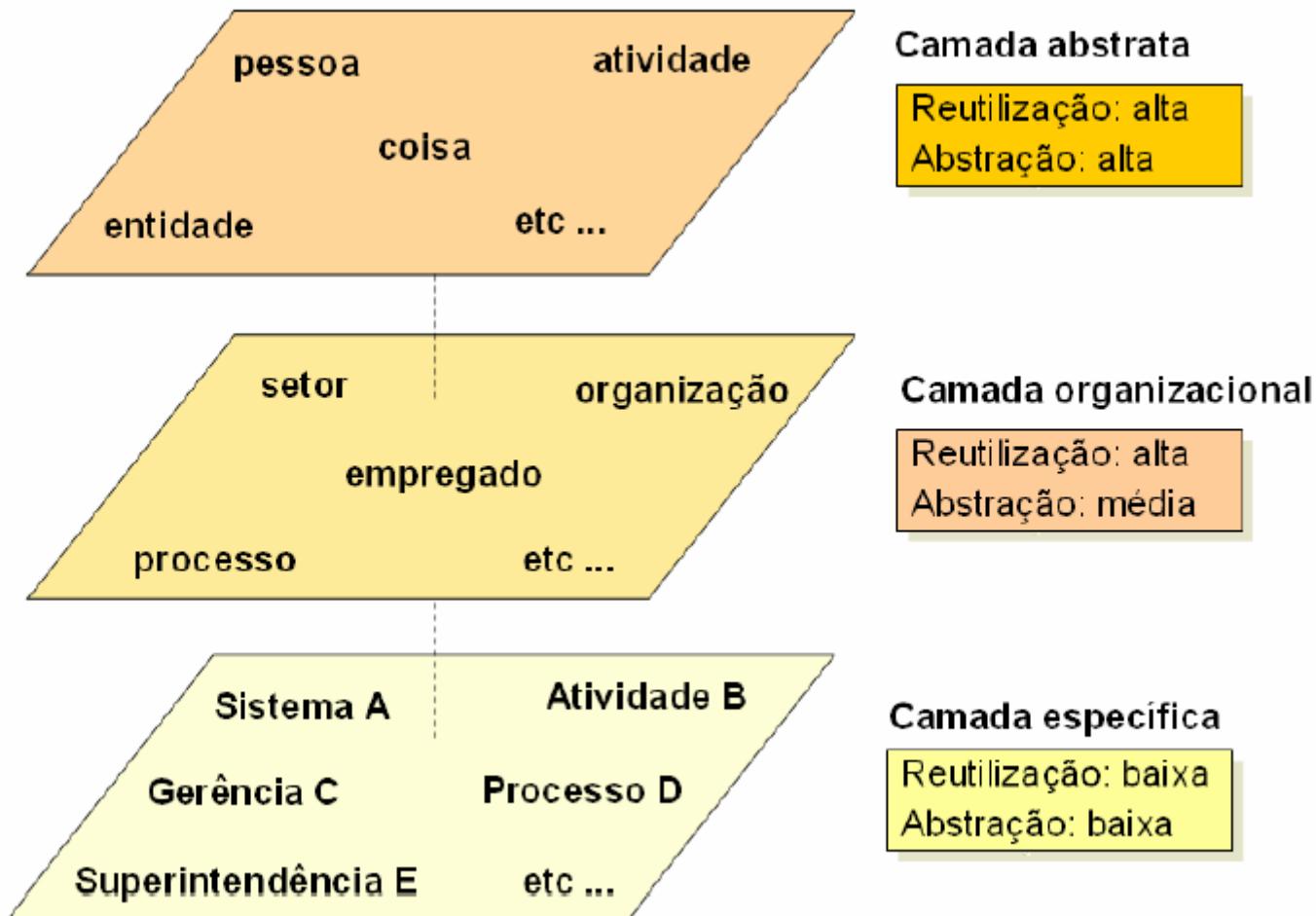
# Definição Ontologia

---

- ▶ Ontologia pode ser definida como “especificação explícita e formal de um conceito compartilhado” (GRUBER, 1993),
- ▶ É o estado da arte para a interoperabilidade de conceitos, **retirando dos sistemas** a propriedade das abstrações e garantindo uma visão mais epistemológica do domínio a ser representado.



# Ontologia Organizacional

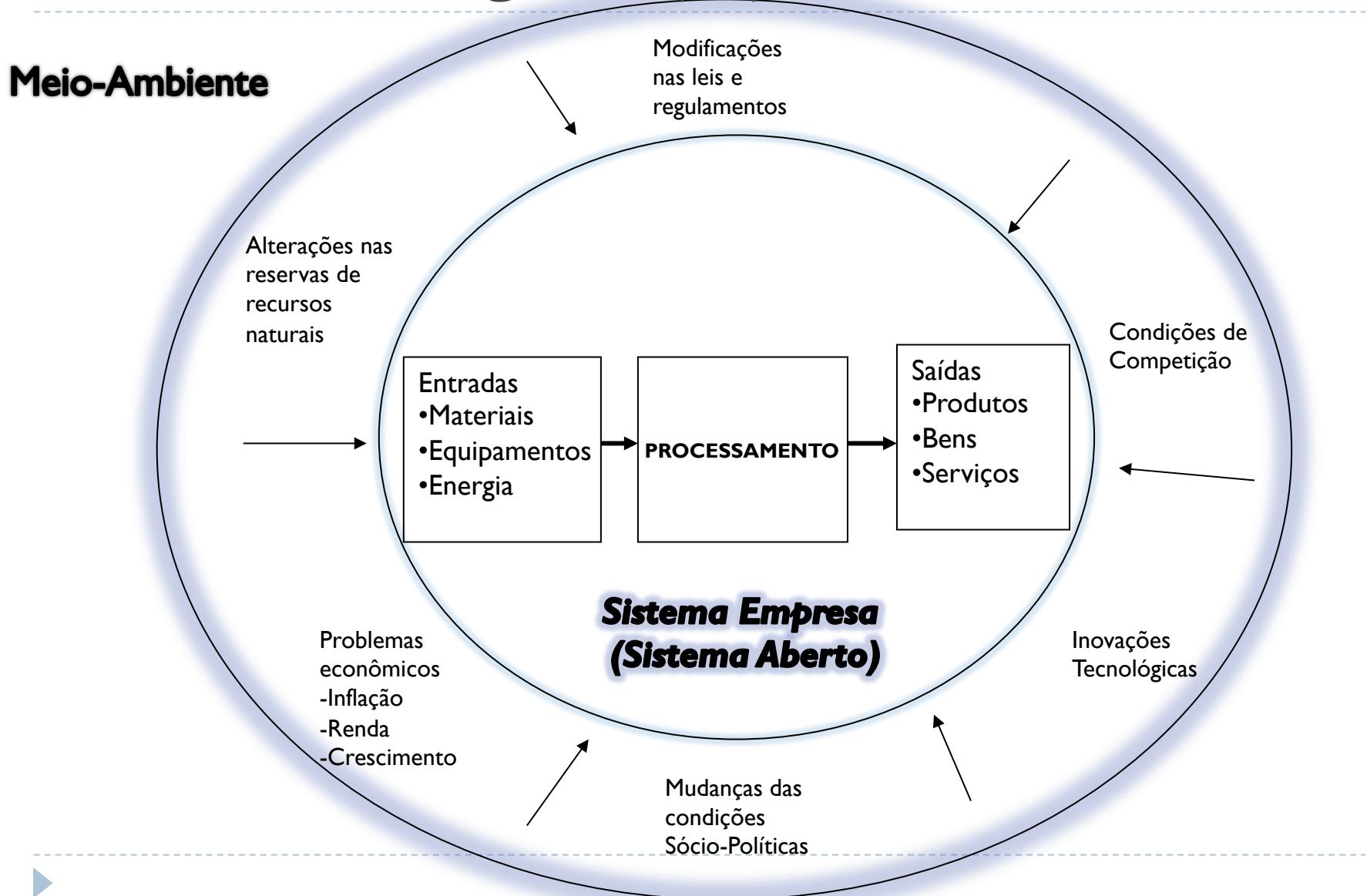


Ref(?)



Qual a motivação para construção  
de Business Intelligence (BI)?

# Qual a motivação para construção de Business Intelligence (BI)?



# Qual a motivação para construção de Business Intelligence (BI)?

---



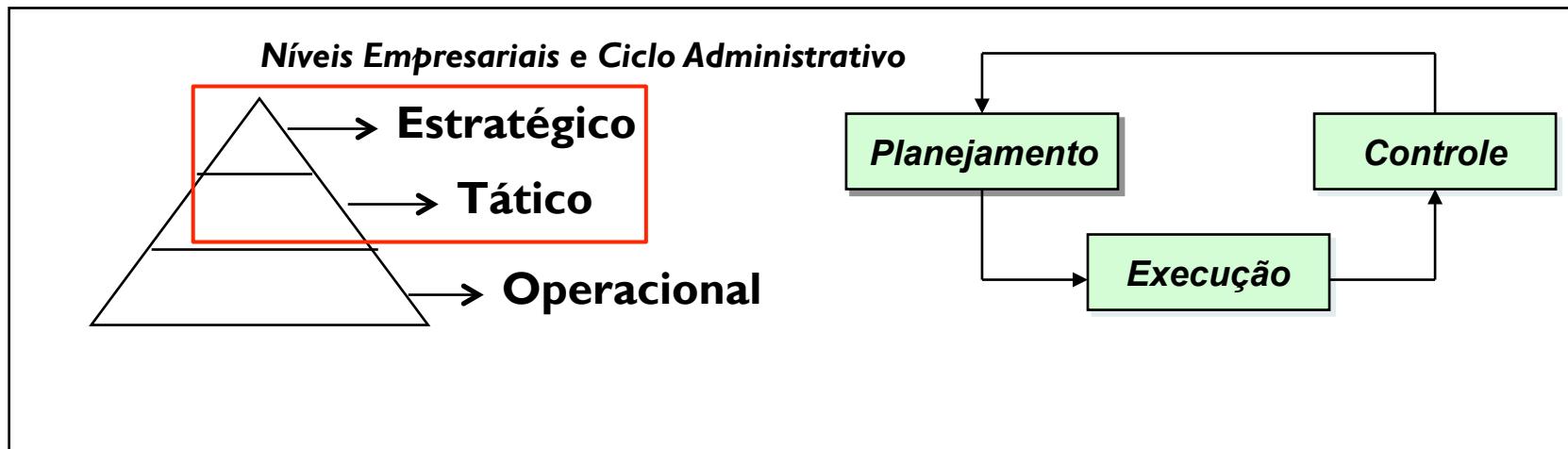
[...]Cenário não é uma realidade futura, mas um meio de representá-la, com o objetivo de nortear a ação presente, à luz dos futuros possíveis e desejados.

---



# Níveis Empresariais e Ciclo Administrativo

- ▶ Os Sistemas de Informações devem produzir informações que:
  - ▶ Possam atender aos níveis empresariais ( estratégico, tático, operacional)
  - ▶ Ciclo administrativo (planejamento, execução e controle).



# Definição BI

---

- ▶ **Business Intelligence (BI)** engloba as **tecnologias, processos, e melhores práticas** que coloquem na ponta dos dedos os **indicadores** chave de desempenho e tornam a **gerência proativa** do seu negócio uma realidade.
- ▶ [http://www.atsolutions.com.br/solucoes/  
businessintelligence.htm](http://www.atsolutions.com.br/solucoes/businessintelligence.htm)



# Definição BI

---

- ▶ O termo **Business Intelligence (BI)**, pode ser traduzido como **Inteligência de negócios**, refere-se ao processo de coleta, organização, **análise, compartilhamento e monitoramento** de informações que oferecem **suporte a gestão de negócios**.
- ▶ [http://pt.wikipedia.org/wiki/Business\\_intelligence](http://pt.wikipedia.org/wiki/Business_intelligence)



# Definição BI

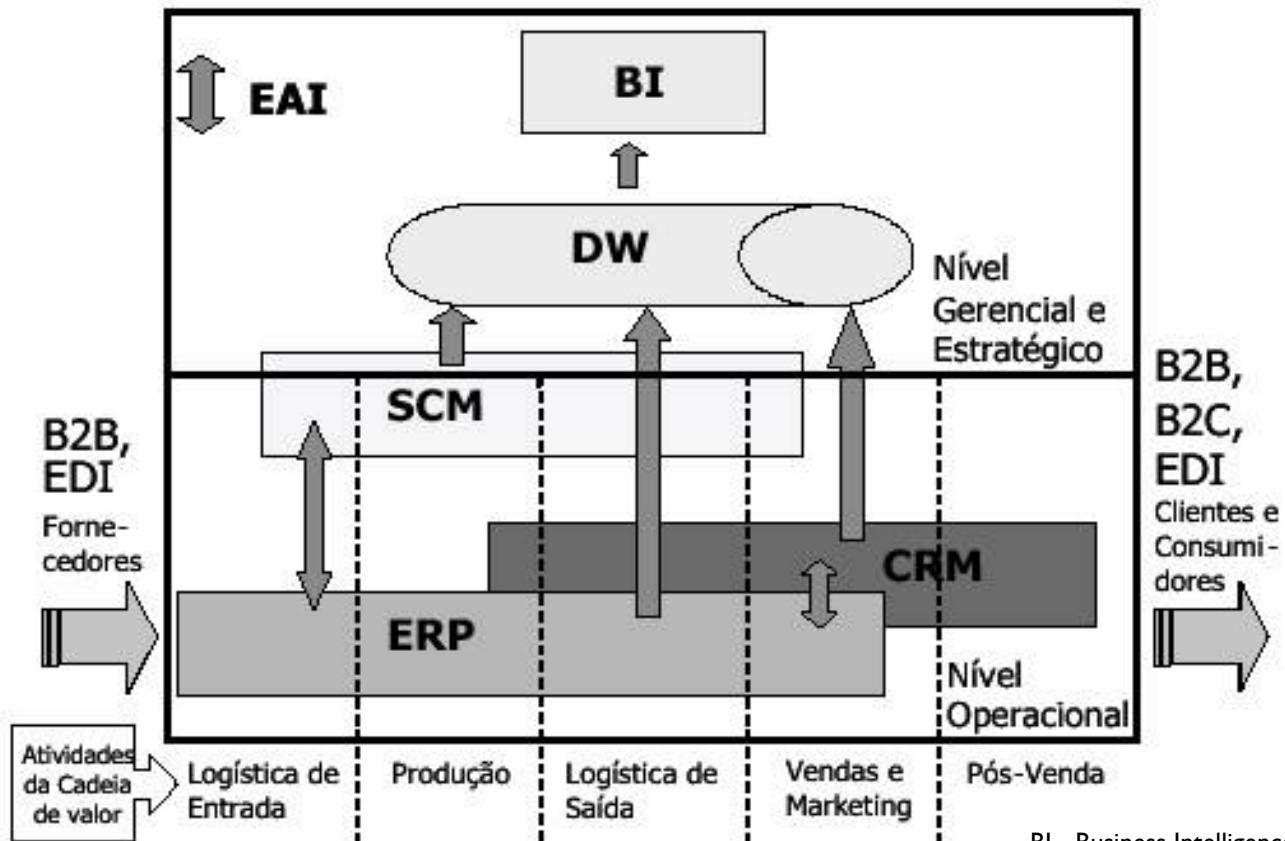
---

- ▶ “O conceito de **Business Analytics**, além das ferramentas tradicionais de BI, traz consigo inovações como características de **notícias on-line** e de **sites comunitários dispostos em relatórios com gráficos** e com **conteúdo desestruturado**, além do suporte à **colaboração** e **workflow** e, o mais importante, **foco no suporte à decisão**.”

- ▶ “<http://computerworld.uol.com.br/company-zone/IDC/quais-as-recomendacoes-para-implementar-bi-e-ba>”



# Sopa de Siglas do BI



<http://www.cyta.com.ar/biblioteca/bddoc/040401/v4n4a1.htm>

BI - Business Intelligence (BI)  
EAI - Enterprise Application Integration  
ERP –Enterprise Resource Planning  
SCM – Supply Chain Management  
CRM –Customer Relationship Management  
B2C – Business-to-Consumer  
B2B - Business-to-business  
EDI -Electronic Data Interchange

## Fatores Críticos de Sucesso

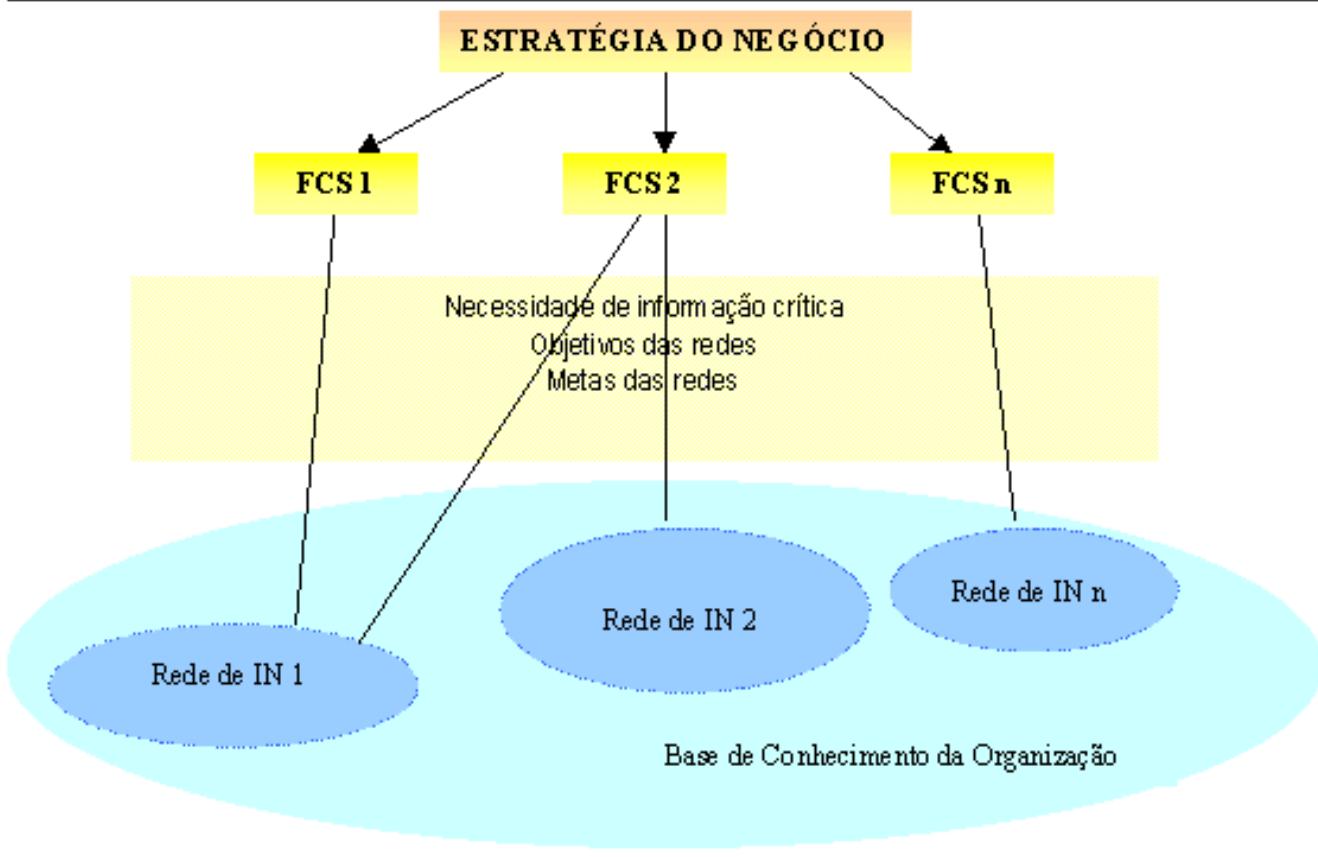
---

- ▶ São poucos (menos que dez; em geral de três a seis)
- ▶ Têm importância vital para a organização
- ▶ São diferentes entre as organizações
- ▶ Têm grande influência sobre as relações da empresa com ambiente, principalmente com os mercados atingidos ou pretendidos.



# Fatores Críticos de Sucesso

FIGURA 3  
FCS's e Redes de Inteligência



# Indicadores

---

- ▶ **O indicador é uma relação matemática entre duas ou mais medidas de desempenho**, com o objetivo de mensurar o **comportamento de um processo**, comparando-o com o padrão pré-definido.
- ▶ Exemplos:
  - ▶ **Percentual de atendimentos com atraso:** Este indicador é encontrado a partir das medidas: nº. de atendimentos com atraso e Total de atendimentos.
  - ▶
  - ▶ **Devem ser mesuráveis e observáveis**



## Fatores de Sucesso ou Insucesso no BI

# Fatores para o Sucesso do BI

---

- ▶ **Apoio e aprovação da alta gerência.**
- ▶ **Equipe composta por especialistas na tecnologia de Business Analytics e por especialistas no negócio da organização que tenham poder de decisão.**
- ▶ “  
<http://computerworld.uol.com.br/company-zone/IDC/quais-as-recomendacoes-para-implementar-bi-e-ba> ”
- ▶

# Fatores para o Sucesso do BI

---

- ▶ Foco na **integração contínua**, na qualidade e nos esforços magistrais para a gerência de informações.
- ▶ **Única fonte de conhecimento corporativa** para o processo de tomada de decisão.
- ▶ Políticas e procedimentos de governança estabelecidos.
- ▶ **Indicadores** de medida de performance **definidos** de acordo com as **necessidades de negócio** da organização.

“<http://computerworld.uol.com.br/company-zone/IDC/quais-as-recomendacoes-para-implementar-bi-e-ba>”



# Fatores para o Insucesso do BI

---

- ▶ Uma adoção **liderada apenas pela área de tecnologia**, sem o apoio da alta gerência e sem o devido envolvimento das outras áreas internas de interesse.
- ▶ Ecologia da Informação
  - ▶ Falta de conhecimento sobre os impactos que podem ser gerados pela nova solução, nos negócios da organização. Por exemplo, considerar **disponibilidade de 99,5%** como um indicador, sem prévia análise.
  - ▶ O quanto custaria para a empresa a indisponibilidade de 0.05%? Esta sim deveria ser a questão a ser respondida.
- ▶ “  
<http://computerworld.uol.com.br/company-zone/IDC/quais-sao-recomendacoes-para-implementar-bi-e-ba>”

# Framework tecnológico do Sistema de Gestão do Conhecimento:

Data Warehouse

Foco no BI

# Framework tecnológico dos Sistema de Gestão do Conhecimento

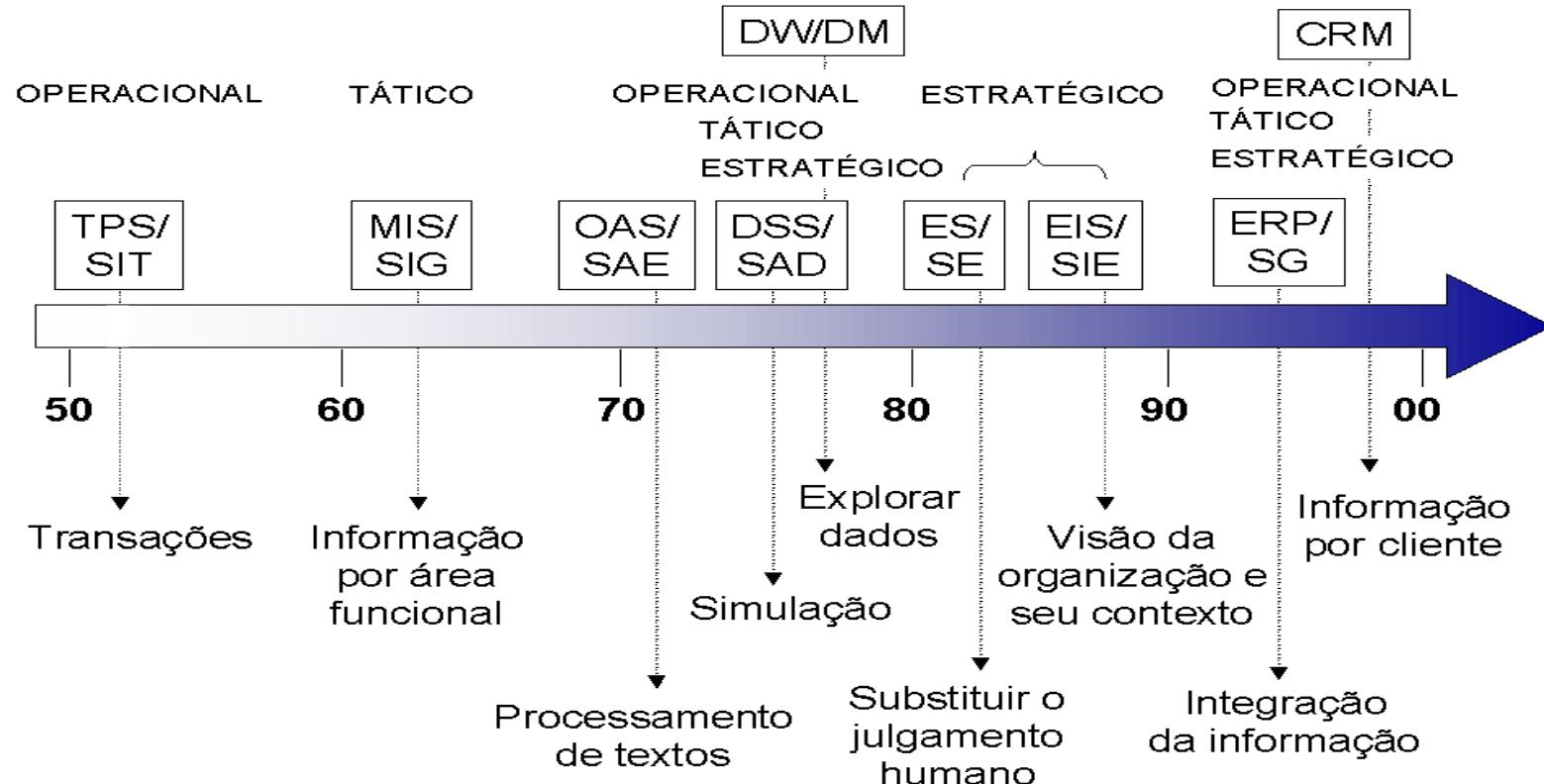
---

## ***FRAMEWORK TECNOLÓGICO***

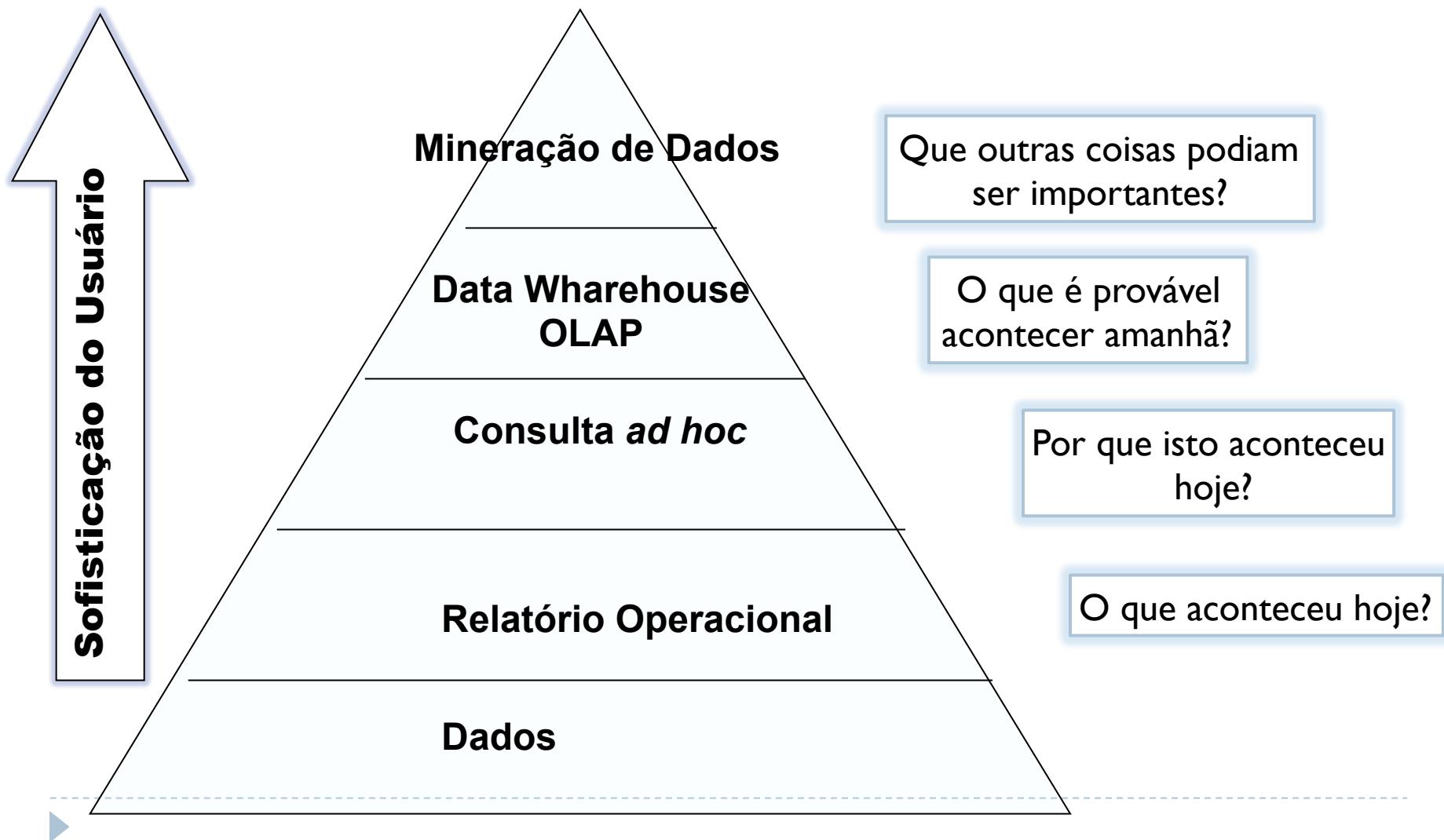
- ▶ Internet/Intranet/Extranet;
- ▶ Workflow;
- ▶ Mapas do conhecimento;
- ▶ Groupware;
- ▶ Gerenciamento eletrônico de documentos;
- ▶ Soluções de BI: Data Warehouse e Data Mining.



# Histórico



# Uma Hierarquia de Necessidades de Informação



# Típica consulta em um Data Wharehouse

- ▶ **Passo I:** A tabela apresenta informações do total do mês e a comparação percentual com o mês anterior das vendas de um produto.
- ▶ Este relatório é uma summarização dos dados do DW Vendas — mudança de granularidade dia para mês (“drill up”).

Produto	Filial	Vendas no Mês	Compraçāo com o Mês Anterior
Papel A4	Salvador	20.000	12%
<i>Papel A4</i>	<i>C a m p i n a Grande</i>	<i>3.000</i>	<i>-5%</i>
Total		23.000	7%

O que isto sugere?

# Típica consulta em um Data Wharehouse

- Neste contexto, o executivo deseja saber a razão para os 5% negativos . Assim, ele executa uma operação de “Drill down“, acrescentando o atributo ‘Tamanho da embalagem’ do produto. Esta operação em que o tipo do relatório é alterado é também chamada de “Slice and dice”.

Produto	Filial	Tamanho	Vendas no Mês	Comparação Com o Mês Anterior
Papel A4	Salvador	A	10.000	11%
Papel A4	Salvador	B	10.000	12%
Papel A4	Total		20.000	12%
Papel A4	Campina Grande	A	1.000	-4
Papel A4	Campina Grande	B	1000	-5
Papel A4	Campina Grande	C	1.000	-4
Papel A4	Total		3.000	-5 %
<i>Total</i>			23.000	7%

# Típica consulta em um Data Wharehouse

- ▶ O executivo com o resultado finalmente descobre uma notável inconsistência em suas equipes de venda, podendo agora tomar providências realmente efetivas.

Produto	Filial	Equipe de Vendas	Vendas no Mês	Comparação Com o Mês Anterior
Papel A4	Salvador	X1	2.000	5%
Papel A4	Salvador	Y1	18.000	19%
Papel A4	Total		20.000	12%
Papel A4	Campina Grande	X2	2.000	4%
Papel A4	Campina Grande	Y2	800	1%
<b>Papel A4</b>	<b>Campina Grande</b>	<b>Z2</b>	<b>200</b>	<b>-20%</b>
Papel A4	Total		3.000	-5 %
<i>Total</i>			23.000	7%



# BI: Na Perspectiva do Projeto do Data Warehouse

# BI: Na Perspectiva do Projeto do DW

- ▶ No cenário atual de globalização, é de suma importância para as organizações sistemas de informação consistentes, rápidos e acessíveis.
- ▶ É comum a informação estar dispersa em diversos bancos de dados ou outras fontes de informação
  - ▶ O que dificulta o seu acesso integrado de forma veloz e consistente.
- ▶ Esses ambientes de sistemas heterogêneos caracterizam-se por ser verdadeiras **“Ilhas de informação”**.



# BI: Na Perspectiva do Projeto do DW

---

- ▶ O problema das “ilhas de informação” tem um impacto menor em relação a **Sistemas Rotineiros de Produção**.
  - ▶ Também chamados de Sistemas Operacionais
- ▶ A **ausência de dados integrados** é **menos crítica** em Sistemas Operacionais;

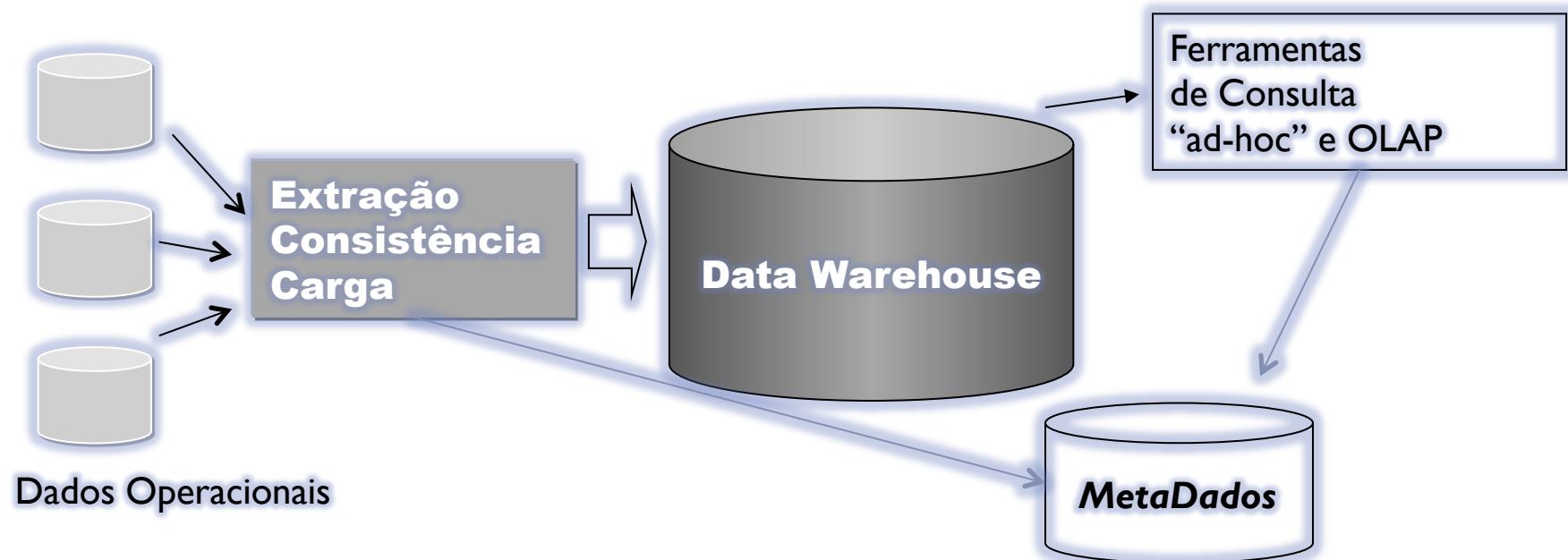


# BI: Na Perspectiva do Projeto do DW

- ▶ **Sistemas gerenciais** necessitam de informação menos detalhada e mais estratégica, implicando em dados integrados e históricos.
- ▶ \* Executar consultas gerenciais sobre sistemas operacionais é absolutamente contraproducente, pois o desempenho seria insuportavelmente baixo.



# Arquitetura de um DWing



# Fontes de Dados Operacionais

---

- ▶ Montagem de um BI é uma oportunidade para fazer uma revisão na qualidade dos Sistemas Operacionais.
- ▶ Dados em Banco de Dados Legado
- ▶ Banco de Dados dos Sistemas Operacionais
- ▶ Planilhas
- ▶ Documentos semi-estruturados e não-estruturados

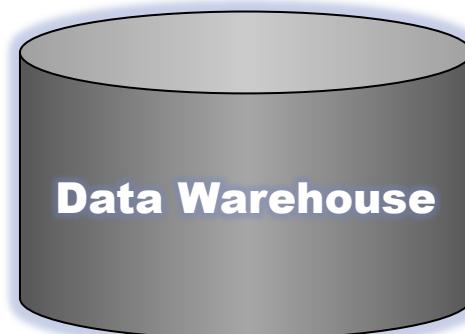


► Dados Operacionais

# Questões Importantes para construção de um DW

---

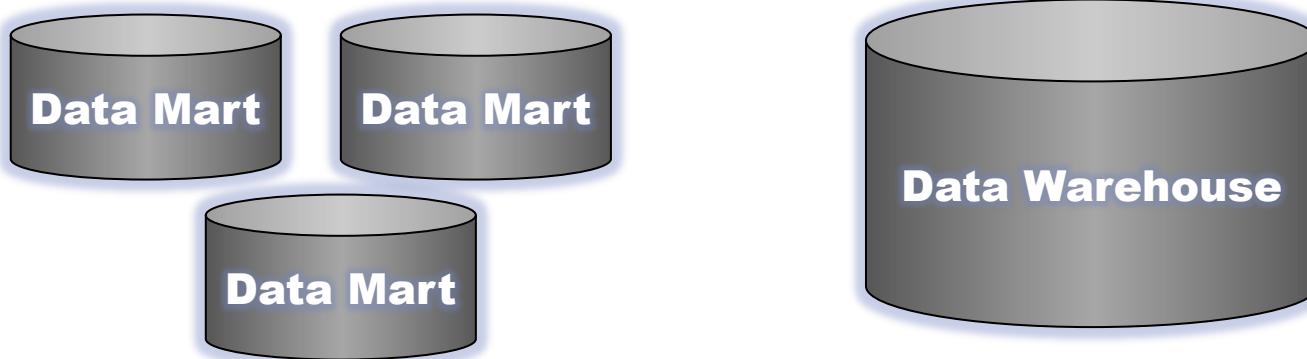
- ▶ Infra-estrutura para armazenamento de um grande volume de dados (dados históricos);
- ▶ Dados organizados de forma a facilitar as consultas pelos usuários responsáveis pela tomada de decisões – **Modelagem Dimensional;**
- ▶
- ▶



# Questões Importantes para construção de um DW

---

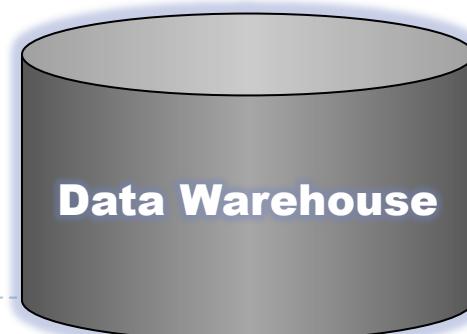
- ▶ Dados são coletados do ambiente transacional;
- ▶ Séries de dados no tempo;
- ▶ Abordagem de organização por assunto
  - ▶ Data Mart;



# Arquitetura de um DWing

---

- ▶ Definição do Nível de Granularidade do DW
- ▶ “**A definição do nível de granularidade é a mais importante questão de projeto do ambiente de data warehouse” (W. H. Inmon)**
- ▶ Erro comum : Querer obter informações no DW de nível operacional.
  - ▶ Amplia o volume de dados e por consequência a complexidade do projeto



# Arquitetura de um DWing

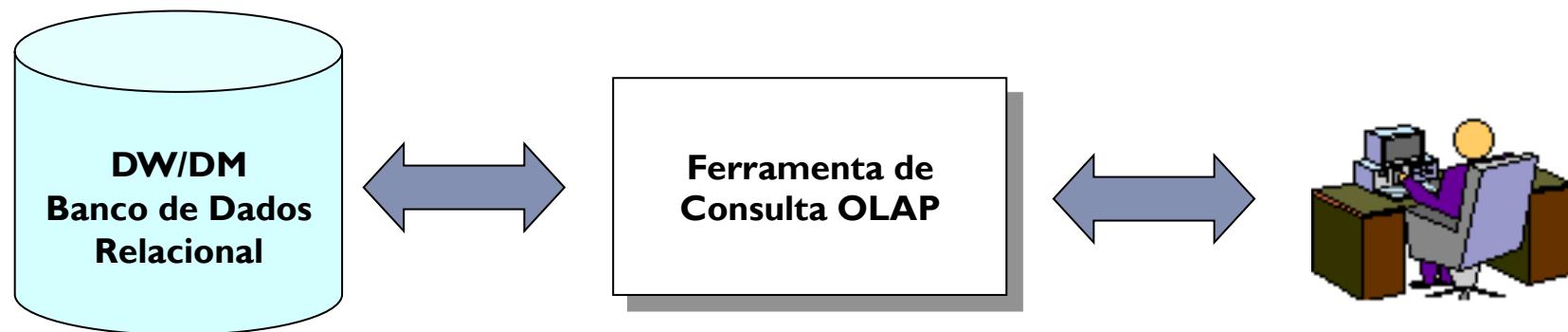
---

- ▶ ROLAP (“Relacional OLAP”) - suporte a consultas OLAP por meio de Sistemas de Gerência de Bancos de Dados Relacionais — SGBDR;
- ▶ MOLAP (“Multidimensional OLAP”) - suporte a consultas OLAP baseadas em arrays multidimensionais.
- ▶ HOLAP = ROLAP + MOLAP



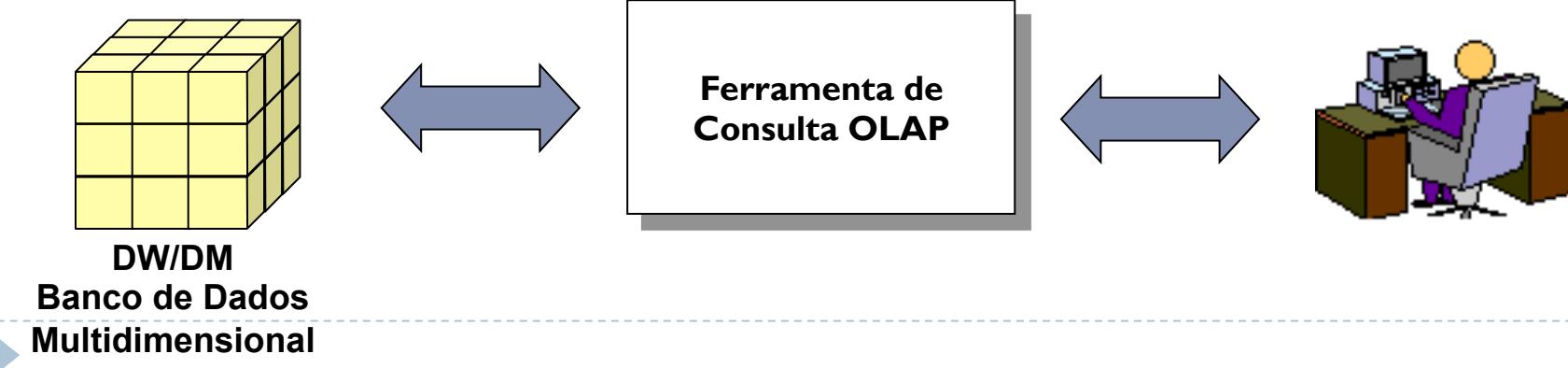
# ROLAP

Vantagens	Desvantagens
<ul style="list-style-type: none"><li>✓ Utilização de uma tecnologia consolidada e robusta;</li><li>✓ Arquitetura mais facilmente escalável;</li><li>✓ Tecnologia e cultura já disponíveis nas organizações.</li></ul>	<ul style="list-style-type: none"><li>✓ Tecnologia relacional (SQL) não é natural para consultas analíticas (OLAP);</li><li>✓ Performance na manipulação de registros em caso de grandes volumes de dados;</li></ul>



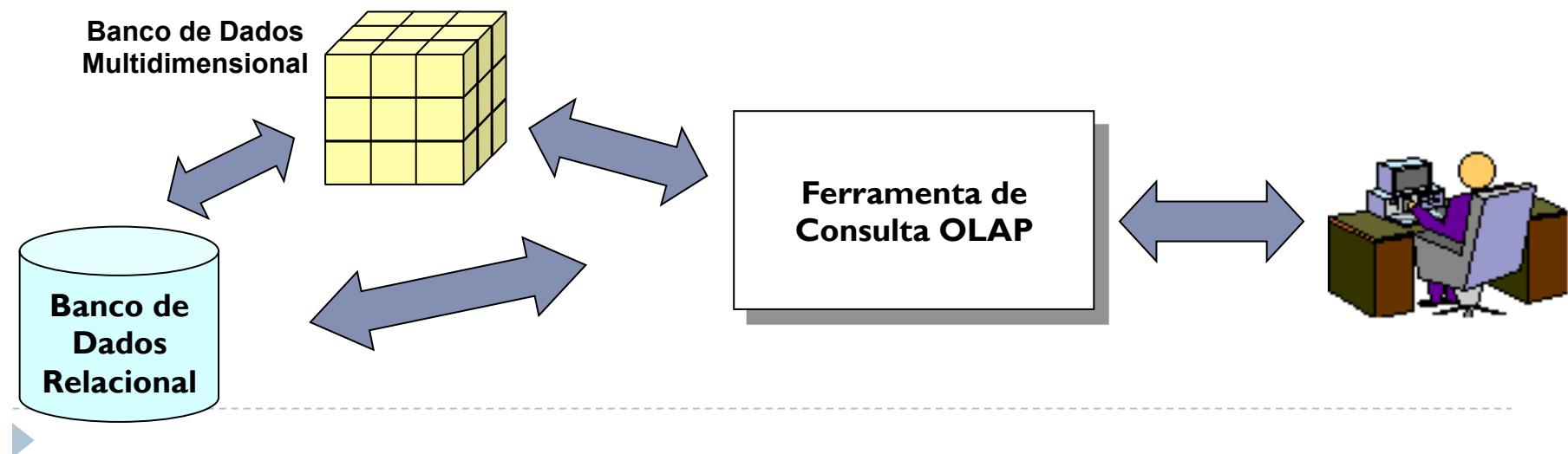
# MOLAP

Vantagens	Desvantagens
<ul style="list-style-type: none"><li>✓ Utilização de tecnologia mais próxima da consulta analítica (OLAP)</li><li>✓ Performance.</li></ul>	<ul style="list-style-type: none"><li>✓ Tecnologia e cultura não muito difundidas e consolidadas nas organizações;</li><li>✓ Limitações para armazenamento de grande volume de dados.</li></ul>

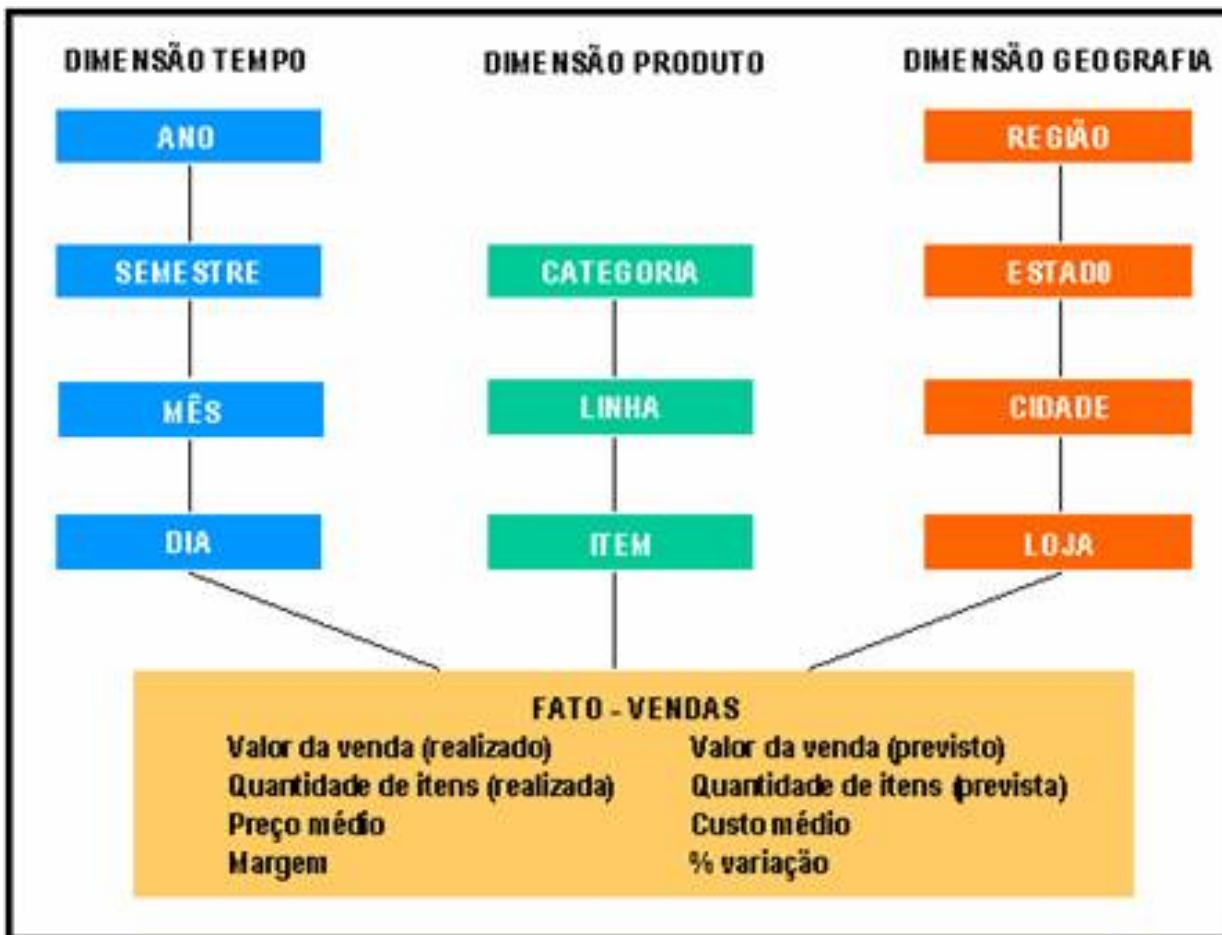


# HOLAP

Vantagens	Desvantagens
<ul style="list-style-type: none"><li>✓ Combina as vantagens das arquiteturas ROLAP e MOLAP;</li><li>✓ Permite o armazenamento de informações transacionais detalhadas sem causar impacto na performance das consultas.</li></ul>	<ul style="list-style-type: none"><li>✓ Ambiente híbrido;</li><li>✓ Necessidade de ferramentas para armazenar dados em duas tecnologias diferentes.</li></ul>

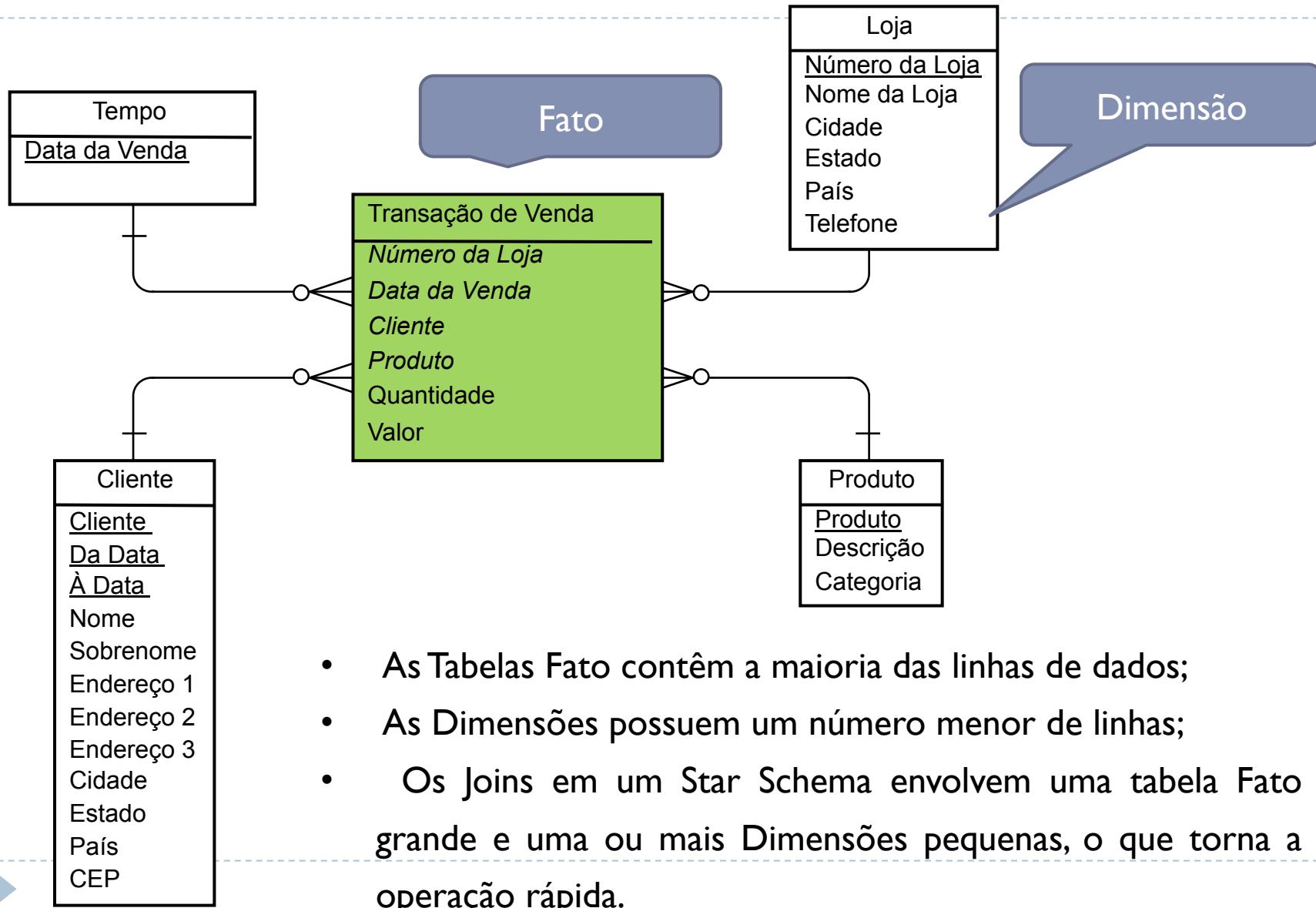


# Modelagem Multidimensional



<http://www.devmedia.com.br/articles/viewcomp.asp?comp=5686>

# Performance no Modelo Star Schema



# Prevendo o tamanho dos fatos

---

- **Vendas diárias:**

- Dimensão Tempo: 730 dias (2 anos)
- Dimensão Loja: 300 lojas
- Dimensão Produto: 20.000 produtos (média de 10 por compra)
- Dimensão Cliente: 100.000 clientes (média de 500 por dia)
- Número de registros:  $730 \times 300 \times 10 \times 500 = 1.095.000.000$
- Tamanho médio do registro = 24 bytes
- **Tamanho da tabela fatos = 1,095 milhões x 24 bytes = 26 GB**



# Metadados

---

- ▶ A gerência de metadados é um problema mais complicado quando se trata de ambientes de DWing.
  - ▶ Até mesmo ‘primitivos’ sistemas de arquivos (“legacy systems”) podem estar envolvidos.
  - ▶ Um gerente de metadados precisa descrever globalmente o sistema de DWing.
  - ▶ Não existe um padrão.
- ▶ Metadados podem ser classificados como: (Segundo Kimball)
  - ▶ Metadados Técnicos (“Back room metadata” )
  - ▶ Metadados Semânticos (“Front room metadata” )



# Metadado

---

- ▶ **Metadados Técnico** cobrem informação sobre:
  - ▶ Esquemas dos sistemas operacionais e do DW;
  - ▶ Esquemas lógico e físico do DW (p.e. ROLAP, MOLAP);
  - ▶ Dependências e mapeamentos entre as fontes operacionais e o DW, nos níveis lógico e físico;
  - ▶ Históricos dos dados e das ações dos diferentes tipos de usuário — administradores, desenvolvedores e usuários OLAP.



# Metadados

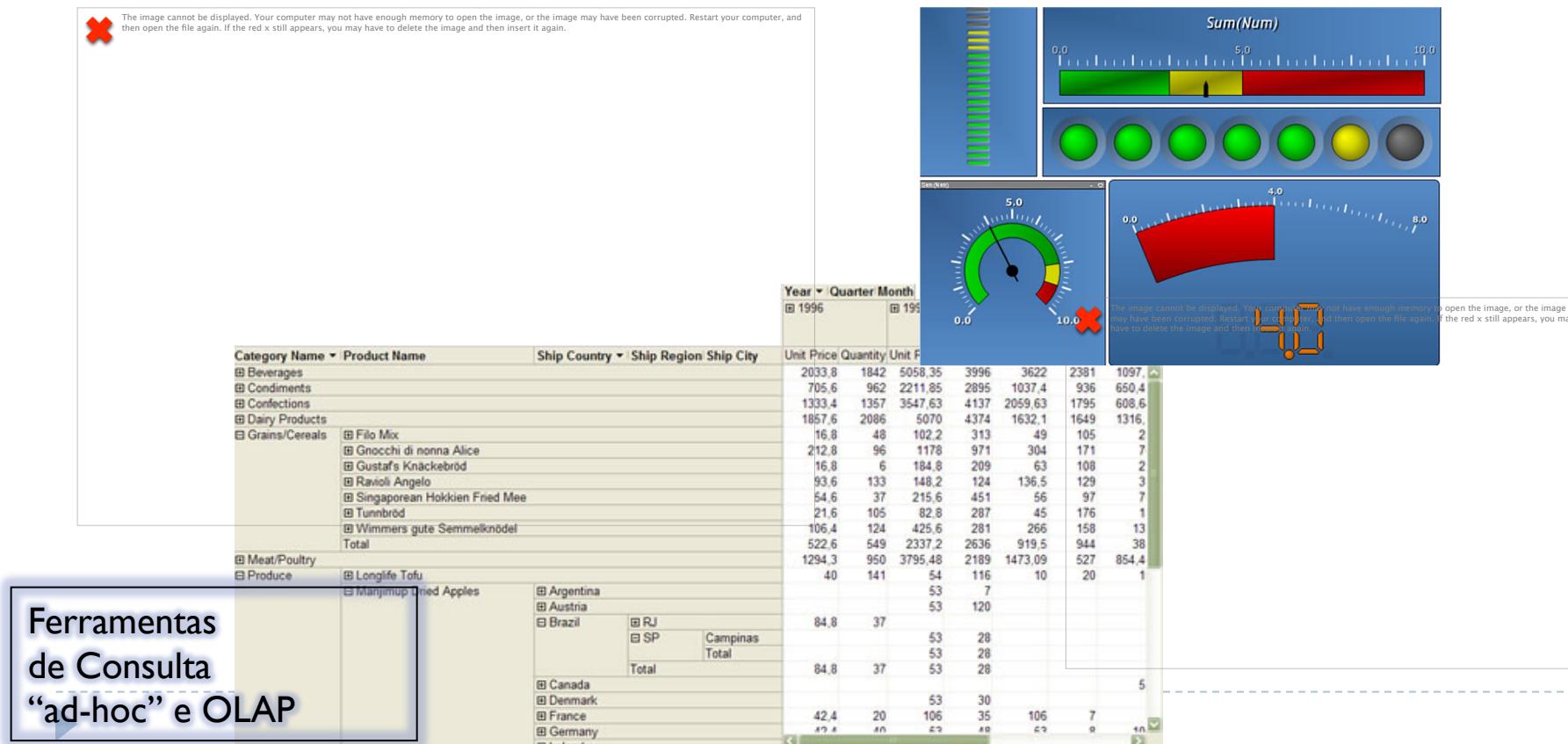
---

- ▶ **Metadados Semânticos** cobrem os seguintes tipos de metadados:
  - ▶ Esquema conceitual do DW: dimensões (critérios de agregação de dados) e atributos de informações agregadas;
  - ▶ Indicadores
  - ▶ Baseado na Ontologia Organizacional
  - ▶ Dependências entre o esquema conceitual e o esquema lógico do DW.
- ▶ Típicos usuários de metadados semânticos são usuários OLAP e desenvolvedores de aplicação.



# Ferramentas de Consultas OLAP

- ▶ Ferramentas de Consulta Dinâmicas (Possibilitar o Gestor Elaborar as suas próprias Consultas)
    - ▶ Mineração Visual / Metáforas de Interface
    - ▶ Dashboards, relatórios, gráficos, Mapas(geoprocessamento) e etc.

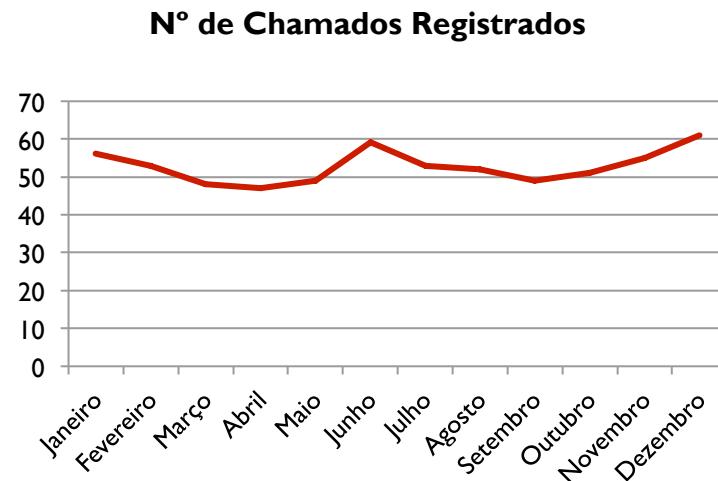


# Metáfora Visual

---

- ▶ Por que usar uma Metáfora Visual?

Período	Nº de Chamados Registrados
Janeiro	56
Fevereiro	53
Março	48
Abril	47
Maio	49
Junho	59
Julho	53
Agosto	52
Setembro	49
Outubro	51
Novembro	55
Dezembro	61



# Ferramentas de Consultas OLAP

---

- ▶ Integradas com o Porta da Organização
- ▶ Ferramenta Web
  - ▶ Cliente Universal e Com Baixo poder Computacional
  - ▶ Diminuição da Necessidade de Treinamento
  - ▶ Alertas por Email e SMS
  - ▶ Dispositivos Móveis
- ▶ Ferramentas de Consulta Dinâmicas (Possibilitar o Gestor Elaborar as suas próprias Consultas)

Ferramentas  
de Consulta  
“ad-hoc” e OLAP



# Extração, Transformação e Carga

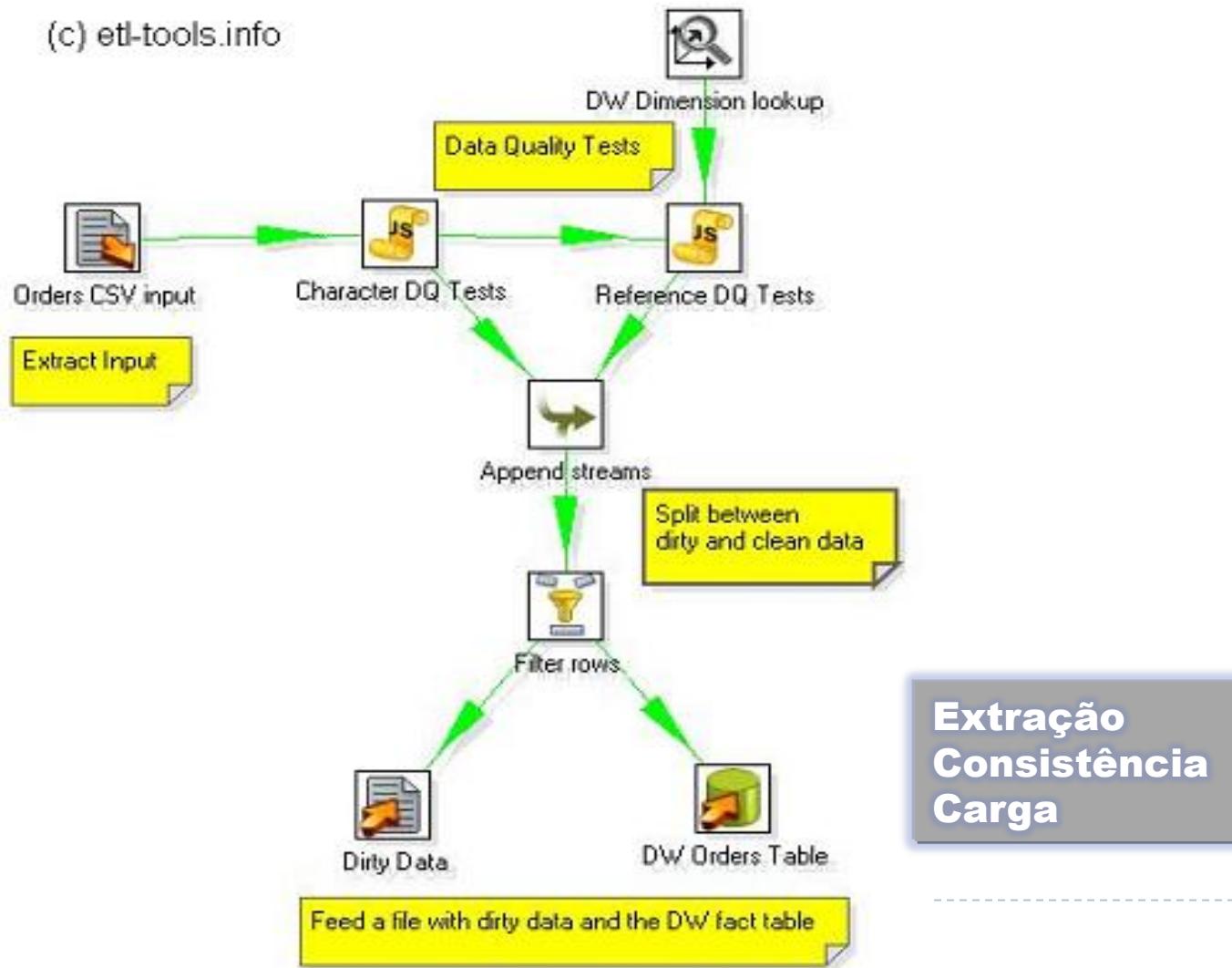
---

- ▶ ETL – Extract, Transformation and Load; (Extração, Transformação e Carga)
- ▶ ETL pode **consumir cerca de 80% do tempo** de desenvolvimento do projeto.
  - ▶ Necessidade de uma Ferramenta de ETL
- ▶ Processo de carga do DW a partir das fontes de dados (ODS);
- ▶ **Implementação de regras de transformação e garantia da qualidade dos dados.**

Extração  
Consistência  
Carga



# Ferramenta de ETL (Pentaho)



# Extração, Transformação e Carga

---

- ▶ Pode ajudar destruir alguns índices para agilizar a carga;
  - ▶ Cuidado com índices que estão relacionados com a Integridade Referencial.
  - ▶ Recriar os índices ao final do processo de carga.
- ▶ Carga em horários que os usuários não estão usando as fontes operacionais.



# Extração, Transformação e Carga

---

- ▶ Evitar atualizações no DW
  - ▶ Alterações, deleções de dados podem ser novos insert's. (na tabela de fato)
  - ▶ Pensar em como gerenciar transações longas no processo de carga.
    - ▶ Arquivo de Log menores
    - ▶ O processo de recuperação pode ser muito custoso



# Equipe de Projeto do Data Wharehouse

---

- ▶ Analista de Negócio
- ▶ Especialistas em Modelagem Multidimensional e Ferramentas OLAP
- ▶ Especialista em Interface Homem Computador
  - ▶ Análise de Interação
- ▶ Especialista em Banco de Dados
- ▶ Estatístico
- ▶ Analista de Testes



# Analisar BI sobre várias Perspectivas

---

- ▶ **Perspectiva do Gestor**
- ▶ **Nível Estratégico**
  - ▶ Ligação com o Planejamento Estratégico
  - ▶ Definição de Fatores Críticos de Sucesso
  - ▶ Futuros Indicadores
    - ▶ Mensuráveis e Observáveis
    - ▶ Análise de Cenários
- ▶ **Definição de Metadados Conceitual**
  - ▶ Dimensões que se deseja analisar o negócio
  - ▶ Fatos históricos usados nas futuras análises



# Analisar BI sobre várias Perspectivas

---

- ▶ **Equipe Técnica**
  - ▶ Qual Ferramenta OLAP adotar?
    - ▶ Interface Web, Dashbord,
  - ▶ Necessidade de um estatístico
  - ▶ Esforço empregado principalmente na atividade de ETL (80% do esforço do projeto)
  
- ▶ **Definição de Metadados Conceitual**
  - ▶ Dimensões metadados lógico e físico
  - ▶ Quais Bases de Dados Operacionais serão utilizadas?
  - ▶ Quais a regras de transformação serão aplicadas?



# Framework tecnológico do Sistema de Gestão do Conhecimento:

Mineração de Dados

# Definição: Mineração de Dados

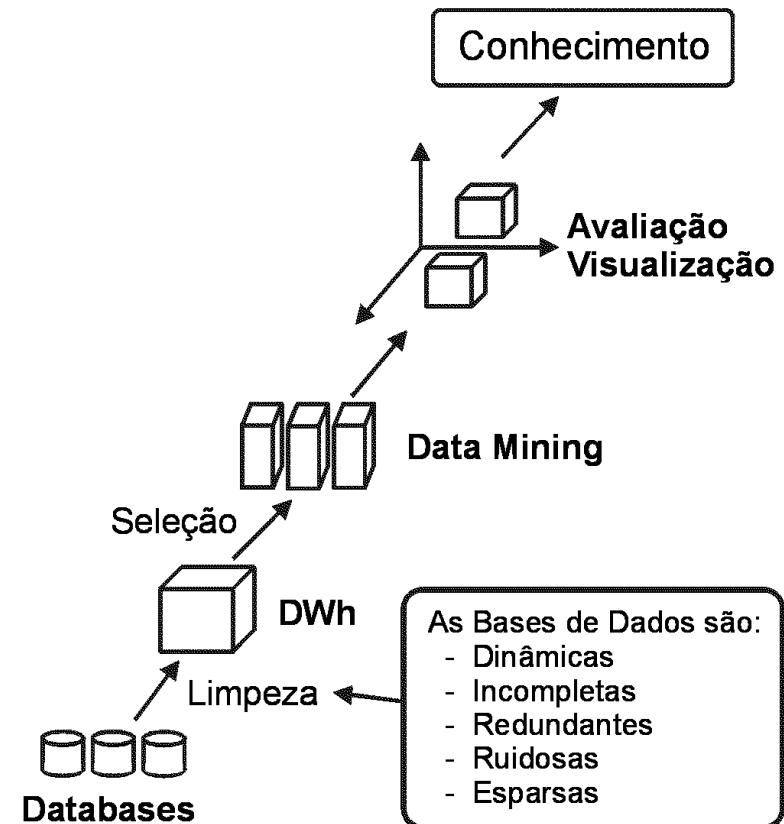
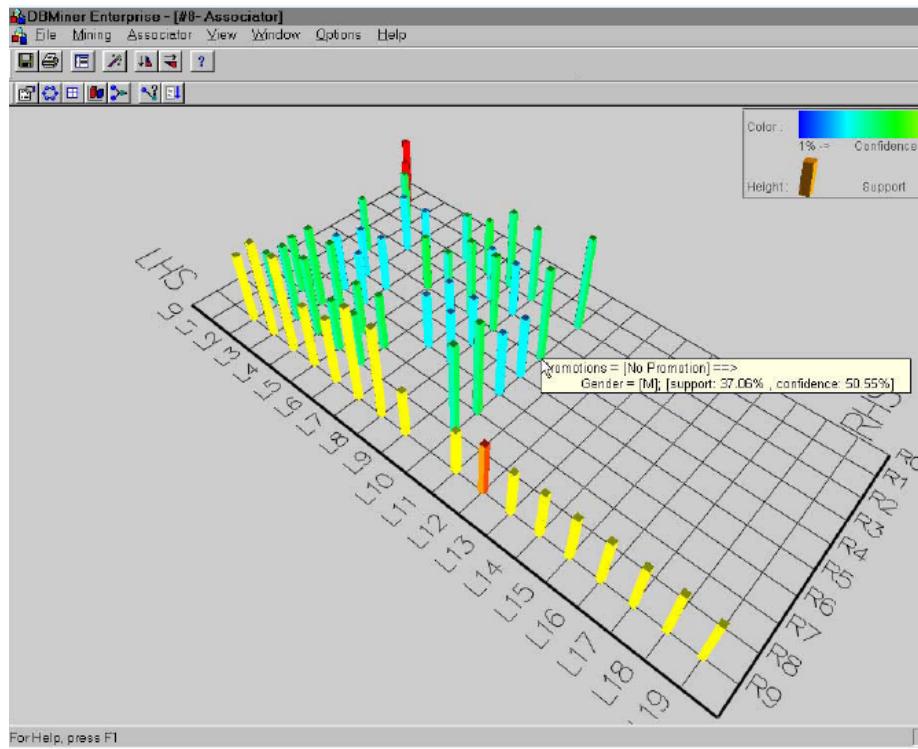
---

- ▶ “Extração de **conhecimento** de BD é o **processo** de identificação de **padrões válidos, novos, potencialmente úteis e comprehensíveis embutidos nos dados**” Fayyad
  
- ▶ Dados: fatos ou casos em um repositório de dados;
- ▶ Padrões: abstração de um subconjunto de dados em alguma linguagem descritiva de conceitos;
- ▶ Processo: extração de conhecimento de base de dados envolvendo diversas etapas;
- ▶ Válido: padrões descobertos devem **possuir algum grau de certeza**;
- ▶ Novos: um padrão encontrado deve fornecer novas informações.

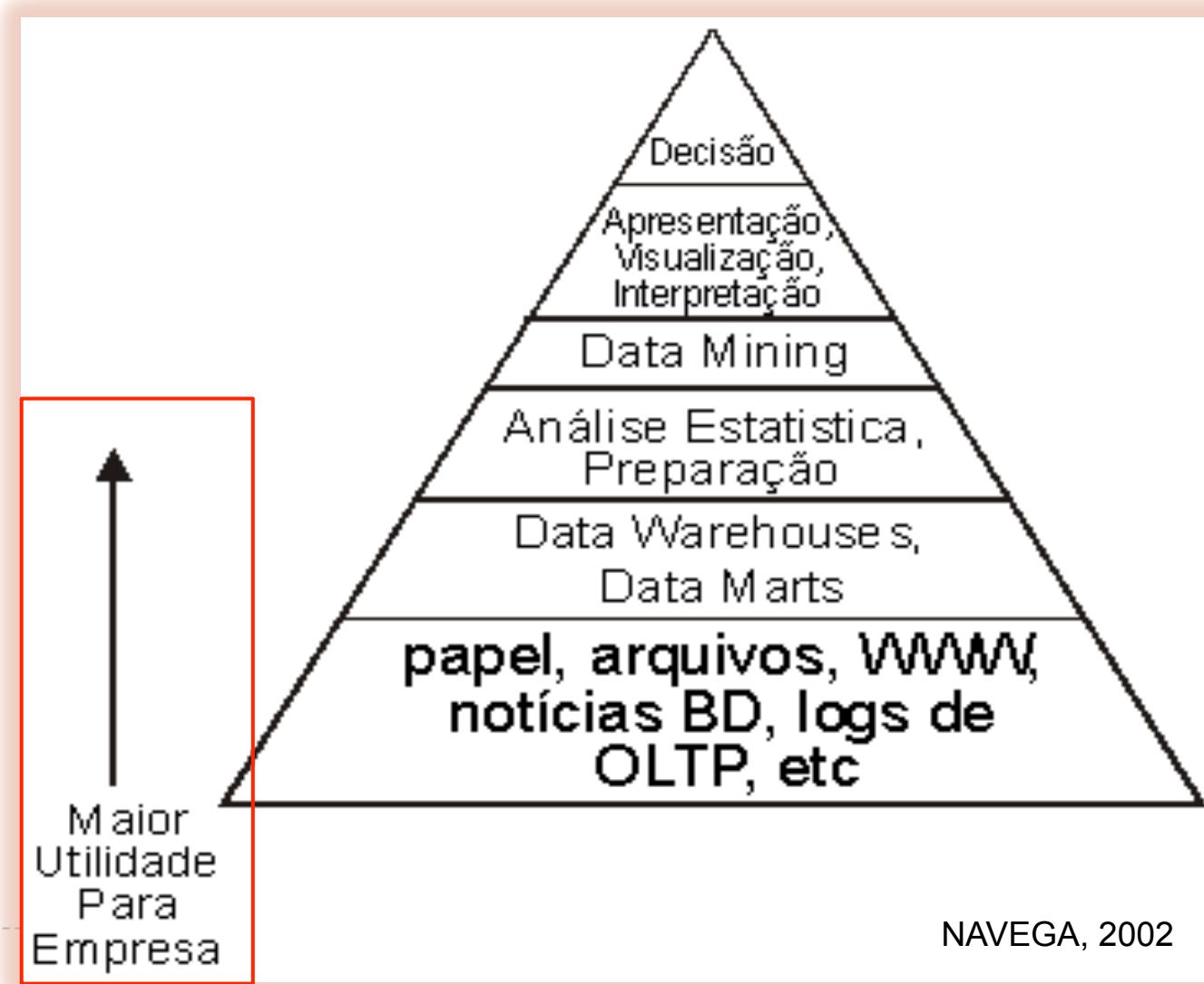


# Mineração de Dados

- ▶ Processo iterativo de criação, teste e refinamento de hipóteses;

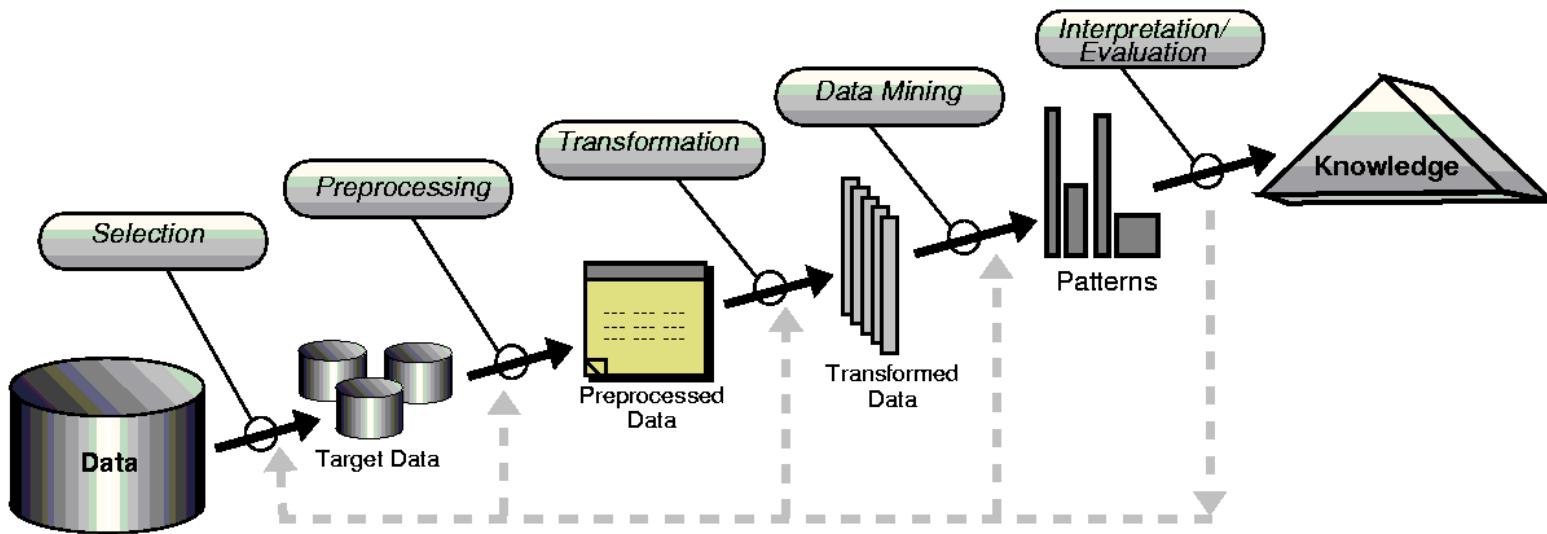


# Mineração de Dados



# Mineração de Dados

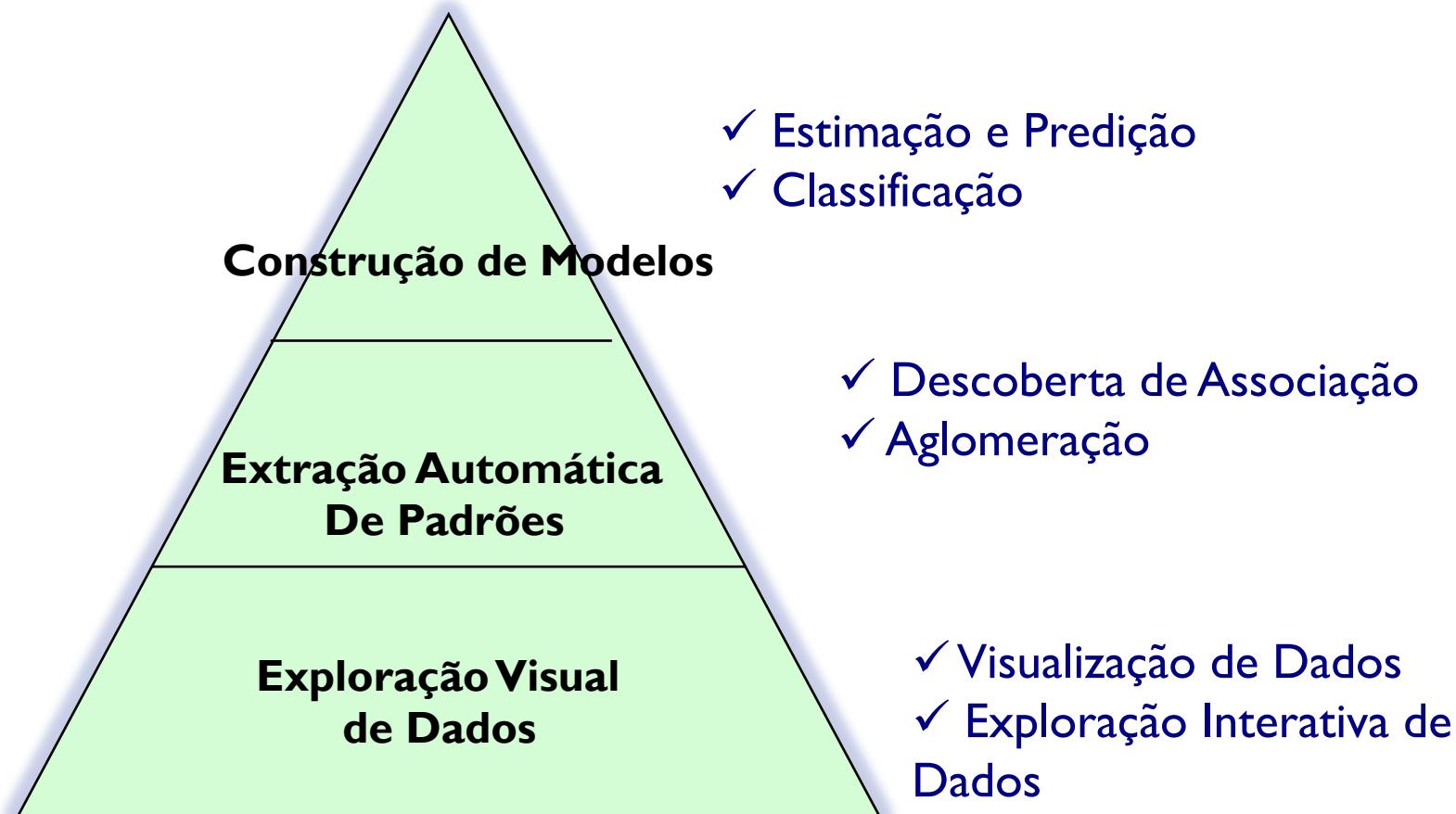
- ▶ Aproveitamento da informação permite ganho de competitividade: “**conhecimento é poder**”;



Ref(?)

# Mineração de Dados

---



# Exemplo: Regras de Associação

---

- ▶ O conhecimento é uma ou mais regras do tipo:
  - ▶ De todas as compras com fraldas, 30% também tiveram cerveja;
  - ▶ Se (Fraldas) então (Cerveja) (30%);
- ▶ Representa elementos/objetos com comportamento conjunto nos eventos descritos na base/negócio;
- ▶ Problema facilmente descrito, solução nem tanto.



# Regras de Associação

---

- ▶ Base de Dados:
- ▶ Cada registro é uma venda, com itens comprados assumindo valores binários (S ou N). TID (Id-Transação)

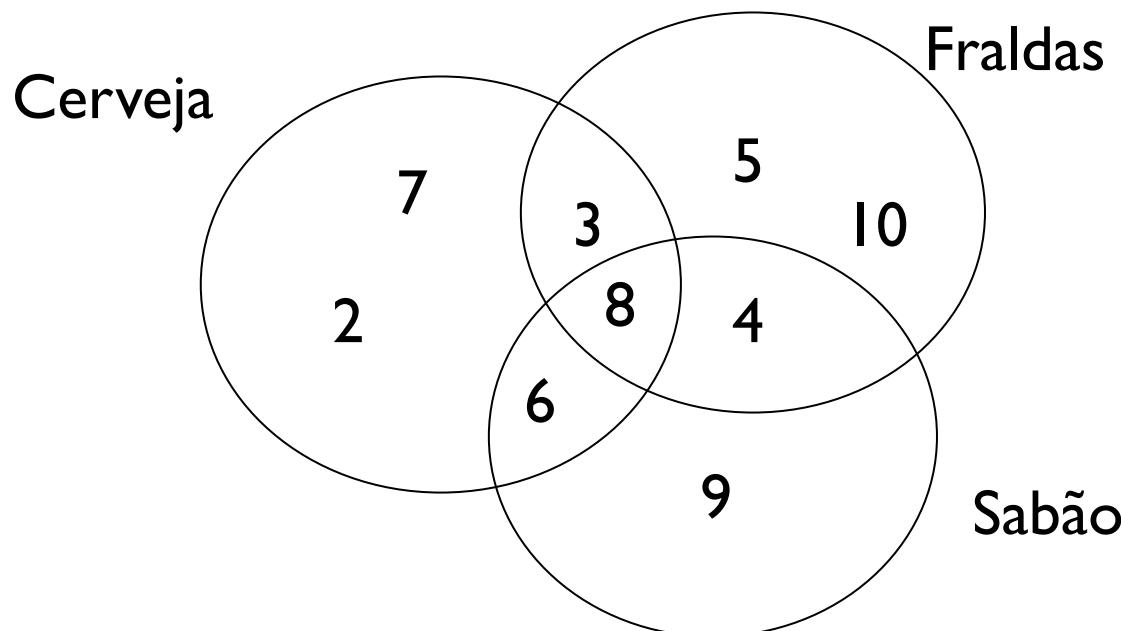
TID	Leite	Guaraná	Fraldas	Cerveja	Pão	Sabão	Arroz	Maça	Carne
1	N	S	N	N	S	N	N	S	S
2	N	N	N	S	N	N	S	N	S
3	N	N	S	S	S	N	N	S	N
4	N	N	S	N	S	S	S	N	N
5	N	N	S	N	S	N	N	S	N
6	S	N	N	S	N	S	S	N	N
7	N	N	N	S	S	N	N	N	S
8	S	N	S	S	N	S	S	N	N
9	N	N	N	N	S	S	S	S	N
10	N	N	S	N	S	N	N	S	S



# Regras de Associação

---

- ▶ Cada elemento tem um conjunto de TID;
- ▶ O objetivo é achar a intersecção entre estes conjuntos;



# Regras de Associação

- ▶ Generalizando...
- ▶ Se ( $X$ ) então ( $Y$ ), onde  $X$  e  $Y$  são conjunto de itens (itemsets)
- ▶ Exemplo:
- ▶  $X = \{\text{Fralda, Cerveja}\} = \{3, 8\}$
- ▶  $Y = \{\text{Sabão}\} = \{4, 6, 8, 9\}$
- ▶ Se (Fralda e Cerveja) então (Sabão)
- ▶ Em 10% do total de transações existentes (1 de 10)
- ▶ Em 50% das transações com Fraldas e Cervejas



# Regras de Associação

---

- ▶ Suporte (ou Freqüência) e Confiança servem para limitar a quantidade de regras extraídas e são definidos antes da extração;
- ▶ Confiança está relacionada com o grau de certeza;

$$\text{Suporte} = \frac{\text{Nº Registros com X e Y}}{\text{Total de Transações}}$$

$$\text{Confiança} = \frac{\text{Nº Transações com X e Y}}{\text{Nº Transações com X}}$$



# Tipos de Técnicas: Mineração de Dados

---

- ▶ Redes Neurais
- ▶ Algoritmos Genéticos
- ▶ Árvores de Decisão
- ▶ Software Weka (*Engloba vários algoritmos*)
- ▶ <http://sourceforge.net/projects/weka/>
  - ▶ ZeroR, OneR, NaiveBayes, J48 (C4.5)
  - ▶ Derivações do J48 - Podas



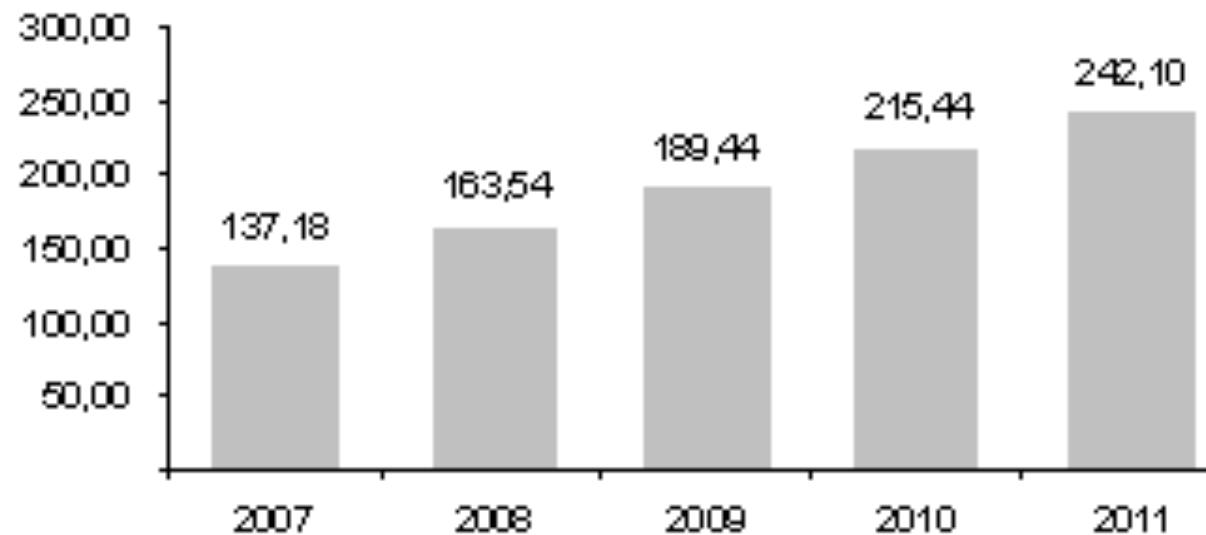


Quanto custa uma solução de BI?

# Custo de Uma Solução BI

---

Mercado Brasileiro de Business Analytics,  
2007-2011 (US\$ M)



Fonte: IDC, 2008



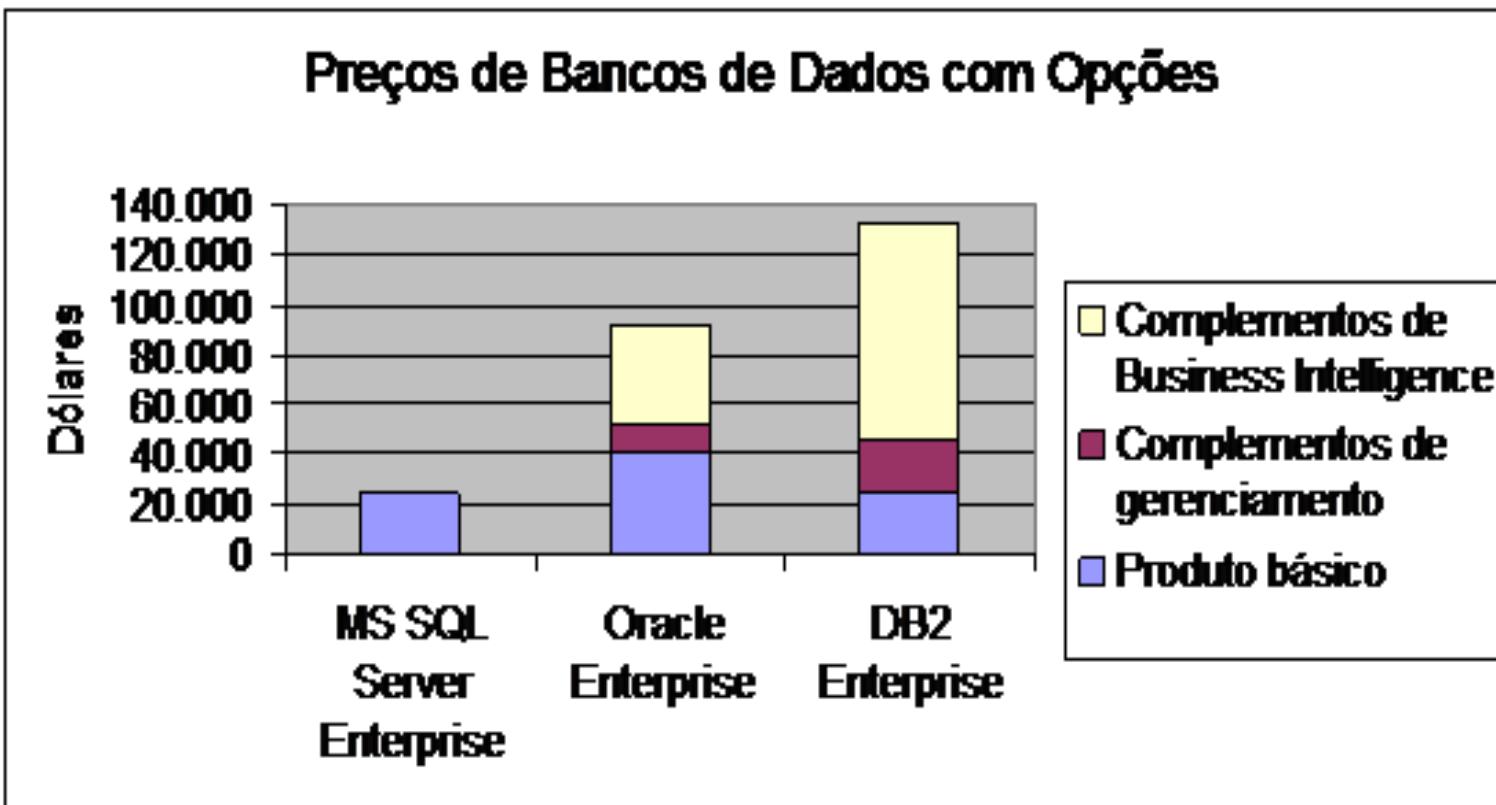
# Custo de Uma Solução BI

---

- ▶ “As licenças de software, por si só, representam uma parcela significativa, **entre 30% e 50%, do custo total de um projeto** de BI de médio porte, podendo alcançar até 80 % do valor final, dependo do projeto.”
- ▶ [http://sol7.com.br/doc/Apresentacao\\_Sol7-BI.pdf](http://sol7.com.br/doc/Apresentacao_Sol7-BI.pdf)



# Custo de Uma Solução BI



Publicado: Novembro de 2005 pela Microsoft

# Custo de Uma Solução BI

- ▶ **Nota:** Todos os preços são por processador e refletem os preços para compras dentro dos Estados Unidos e em Dólares Americanos. Preços baseados em informações disponíveis em sites dos fabricantes. No Brasil consulte um revendedor autorizado pelo fabricante.

<b>Microsoft SQL Server</b>		<b>Oracle</b>	<b>IBM</b>
Enterprise Edition	US\$24.999	US\$40.000	US\$25.000
Complemento de Ferramentas de Gerenciamento	Incluso	US\$3.000 (Diagnostics Pack) US\$3.000 (Tuning Pack) US\$3.000 (Configuration Management Pack) US\$3.000 (Change Management Pack) Total = US\$12.000	US\$9.910 (Recovery Expert) US\$10.000 (Performance Expert) Total = US\$19.910
Complemento de Servidor OLAP	Incluso	US\$20.000 (OLAP Opcional)	US\$28.000 (OLAP Server)
Complemento de Data Mining	Incluso	US\$20.000 (Data Mining Opcional)	US\$60.000 (Intelligent Miner)
<b>Custo Total</b>	<b>US\$24.999</b>	<b>US\$92.000</b>	<b>US\$132.910</b>

Publicado: Novembro de 2005 pela Microsoft



# Custo de Uma Solução BI

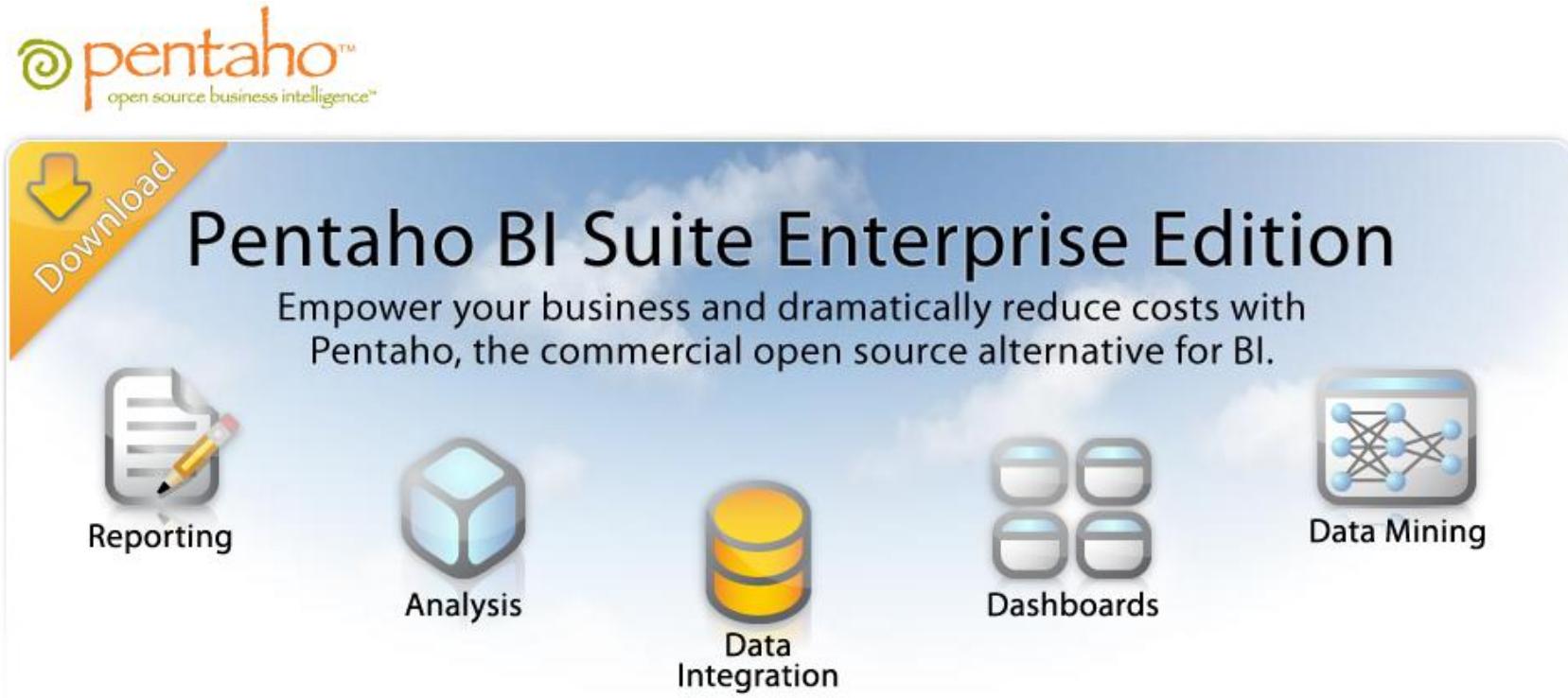
---

- ▶ Já há alguns anos que aplicações para Inteligência de Negócios vêm sendo desenvolvidas para plataformas de **código aberto**.
- ▶ Recentes pesquisas realizadas pela IDC mostram que o interesse por aplicações de BA(Business Analytics) em plataformas de **código aberto** ainda é pequeno.
  - ▶ **Ainda não há grandes perspectivas de mudanças para o próximo ano. Em razão do ceticismo do mercado, estas soluções terão que amadurecer e provar seu valor.**
- ▶ <http://computerworld.uol.com.br/company-zone/IDC/futuro-e-tendencias-de-bi-e-ba/>
- ▶ (informações de 2008 COMPUTERWORLD)



# Custo de Uma Solução BI

- ▶ Possibilidade Interessante



The image shows the landing page for Pentaho BI Suite Enterprise Edition. At the top left is the Pentaho logo with the tagline "open source business intelligence™". A yellow download button is visible on the left. The main title "Pentaho BI Suite Enterprise Edition" is prominently displayed in the center. Below it, a subtitle reads: "Empower your business and dramatically reduce costs with Pentaho, the commercial open source alternative for BI." Five icons represent different features: Reporting (document with pencil), Analysis (cube), Data Integration (stack of cylinders), Dashboards (grid of squares), and Data Mining (neural network). Each icon has its name written below it.

pentaho  
open source business intelligence™

Download

## Pentaho BI Suite Enterprise Edition

Empower your business and dramatically reduce costs with Pentaho, the commercial open source alternative for BI.

Reporting

Analysis

Data Integration

Dashboards

Data Mining



# Custo de Uma Solução BI

---

- Esforço empregado principalmente na atividade de ETL (80% do esforço do projeto)
- Depende do número de indicadores
- ????????





Fim: Obrigado  
Perguntas?