

# Arquitetura de *Data Warehouse* Aula 01 - Fundamentos

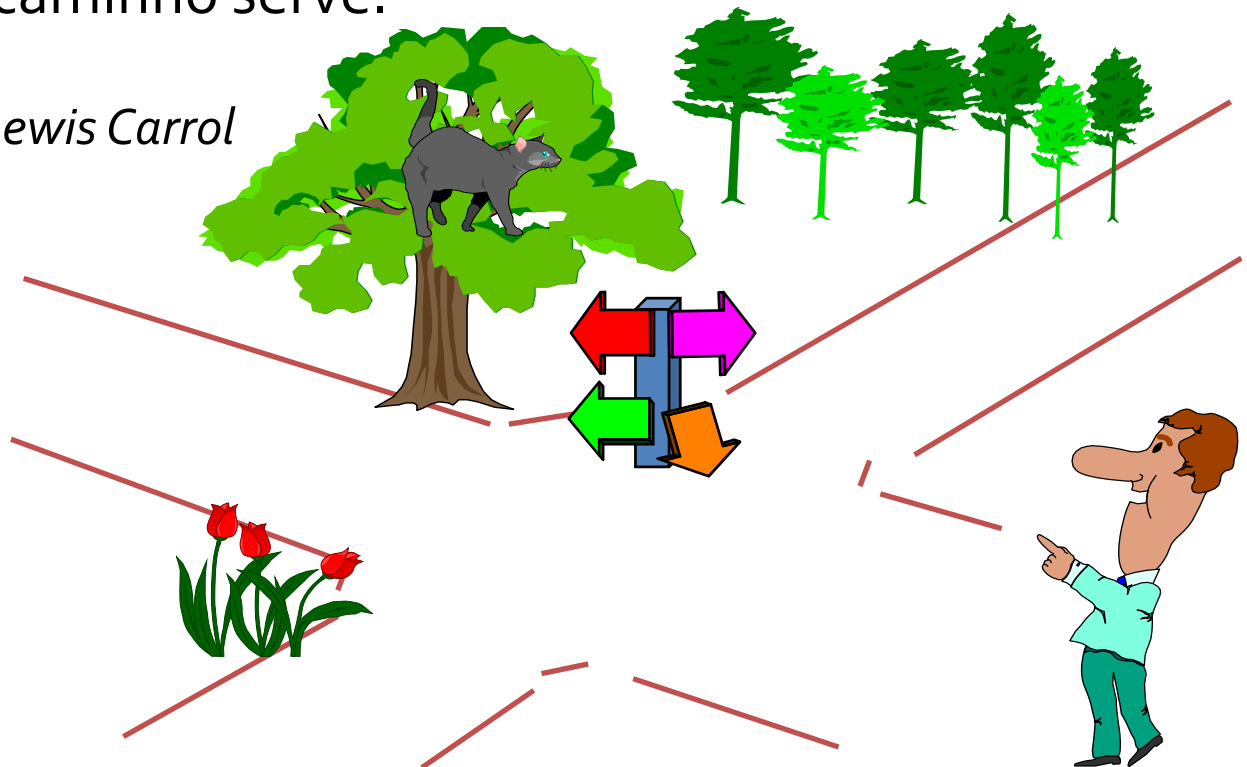
# Sistema de Informação



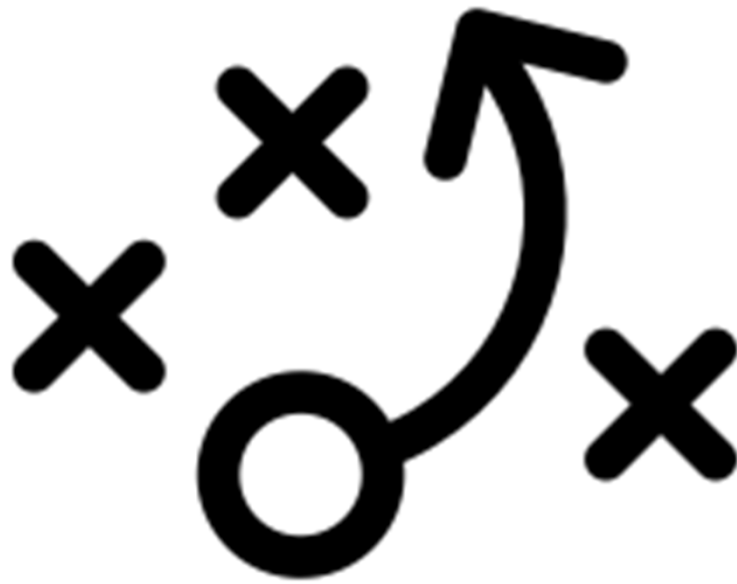
# Futuro...

- **Pode dizer-me que caminho devo tomar?**
- Isto depende do lugar para onde você quer ir.  
*(Respondeu com muito propósito o gato)*
- **Não tenho destino certo.**
- Neste caso qualquer caminho serve.

*Alice no País da Maravilhas, Lewis Carrol*



# Introdução



O que é  
estratégia ?

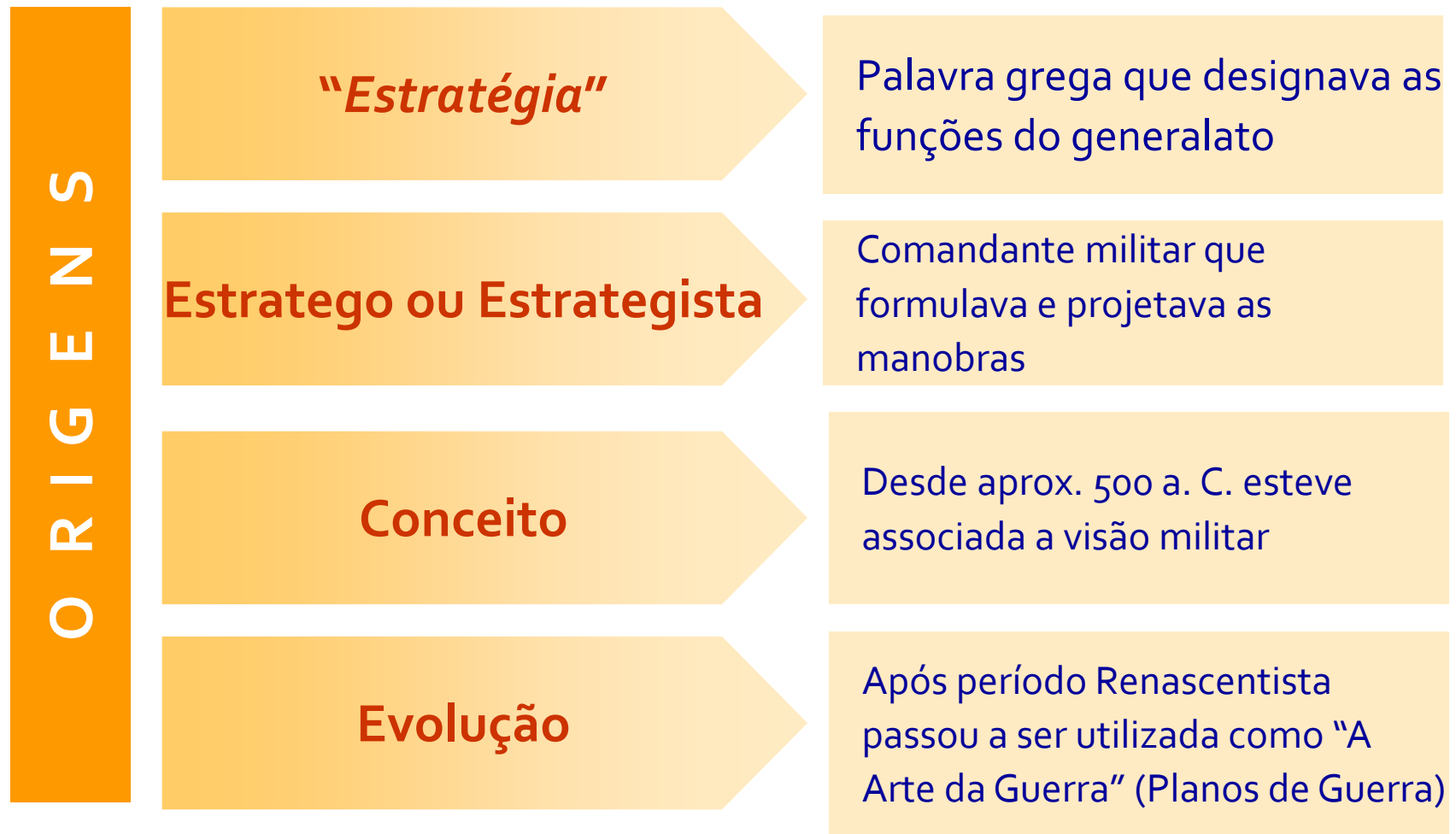
# Definição de Estratégia

Historicamente:

***STRATEGOS*** = **General**

**ESTRATÉGIA** = **“Arte do General”**

# Definição de Estratégia



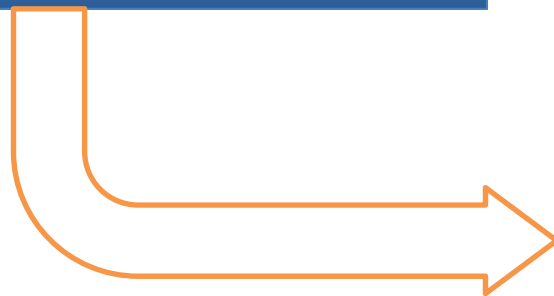
# Definição de Estratégia

**Estratégia Militar**



**Premissa de oposição  
inteligente pelos  
adversários**

**Estratégia Empresarial**



**Premissa de alto grau  
de incertezas, proveniente  
de um ambiente em  
mutação, veloz e ambíguo,  
tudo sem dependência da  
vontade e da ação  
de opositores conhecidos**

# Definição de Estratégia

*Estratégia é a direção e o escopo de uma organização o longo prazo, que obtém vantagem em um ambiente de mudança através de sua configuração de recursos e competências com o objetivo de atender às expectativas dos stakeholders.*



# Estratégia x Tática

*"O melhor plano de batalha é vencê-la de antemão através da estratégia."*

Sun Tzu

*"Todos os homens podem ver as táticas pelas quais conquisto, mas ninguém consegue antever a estratégia em que se baseia a minha vitória."*

Sun Tzu

# Estratégia x Tática

- Do ponto de vista militar:
  - Estratégia é a decisão referente ao lugar, tempo e condições da batalha.
  - Tática é a disposição e movimento das forças durante a batalha.

*Estratégia diz respeito a onde combater o inimigo,  
a tática a como o fazer.*

# Estratégia x Tática

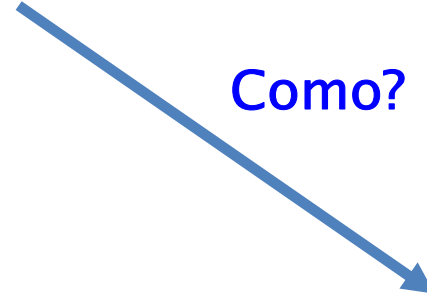
- Estratégia  $\neq$  Tática

Onde?



Plano global para desenvolver recursos que garantem um posicionamento favorável

Como?



Esquema para uma ação específica

# Estratégia x Tática

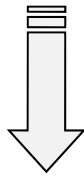
Decisões	Estratégia	Tática
Muito Importantes	Entrada em nova indústria	Substituição de publicidade na TV pela rádio ou jornal
Pouco Importantes	Saída de um segmento e entrada em outro semelhante na mesma indústria	Mudança de publicidade do programa X para Y
Longo Prazo	Entrada em nova área geográfica	Construção de uma nova fábrica
Curto Prazo	Saída de um segmento	Campanha de <i>Merchandising</i>

# Estratégia x Tática

**Estratégia**

**≠**

**Planejamento**



**Pensamento Estratégico**

**+**

**Planejamento Estratégico**

**+**

**Implementação**



**Processo formal de preparação para incertezas, através da formulação dos futuros planos de ação.**



# Dilúvio de Dados



# O que é *Business Intelligence*?

- A necessidade de cruzar informações para realizar uma gestão empresarial eficiente é hoje uma realidade tão verdadeira quanto no passado o foi descobrir se a alta da maré iria propiciar uma pescaria mais abundante.
- O interesse pelo *Business Intelligence* vem crescendo na medida em que seu emprego possibilita às corporações realizar uma série de análises e projeções, de forma a agilizar os processos relacionados às tomadas de decisão.



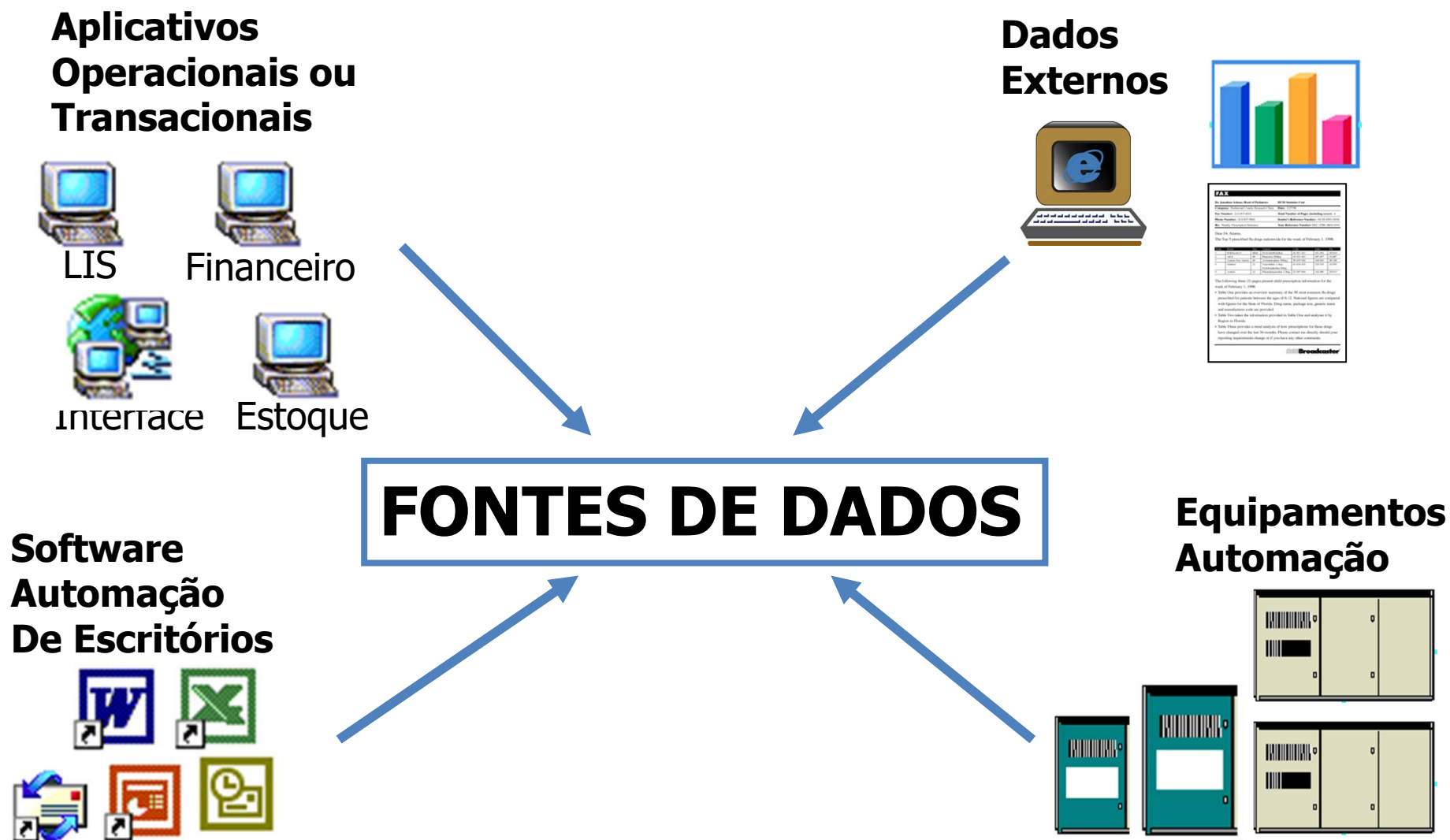
# O que é *Business Intelligence*?

- *Business Intelligence*, ou BI, é um conjunto de soluções tecnológicas que envolve um processo de coleta, transformação, análise e distribuição de dados para a tomada de decisões.
- Também conhecido como Inteligência Empresarial, define a habilidade das empresas em acessar dados e colher informações contidas, por exemplo, em um sistema de *Data Warehouse* (DW), analisando estas informações para dar suporte às tomadas de decisões nos negócios.

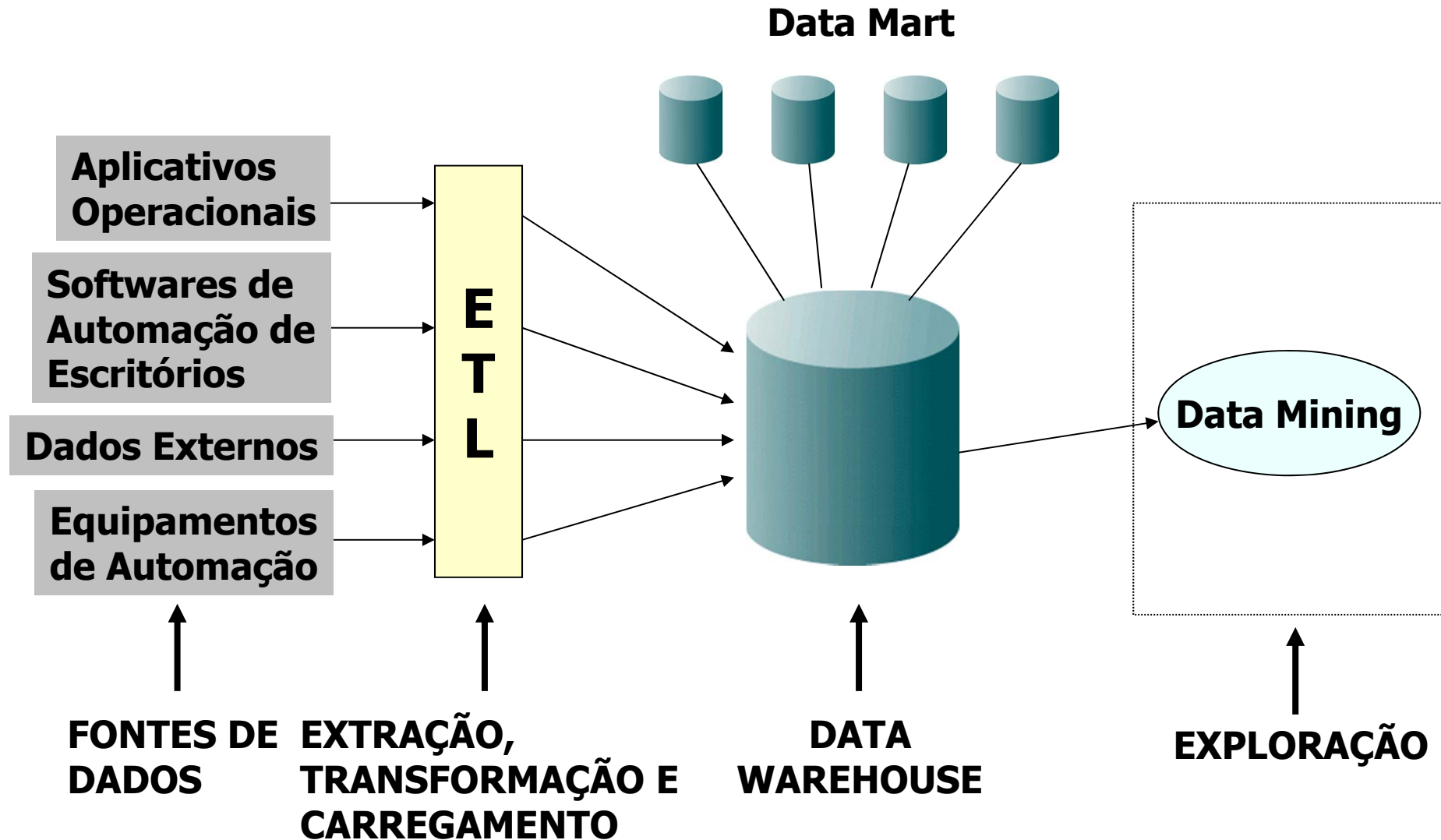
# Aplicativos Operacionais x BI

<b>Aplicativos Operacionais</b>	<b>Business Intelligence</b>
Visão do atual e do real	Visão histórica e de tendência
Solução para requisitos conhecidos	Permitir a identificação de fatos desconhecidos
Abrangência restrita	Abrangência ampla
Informação produzida por profissionais de informática	Informação produzida pelo próprio usuário
Custo e tempo para obtenção da informação altos	Informação obtida com baixo custo e em tempo real
Informação disponível a poucos usuários	Informação democratizada

# BI - Fontes de dados



# Ferramentas e Técnicas de BI

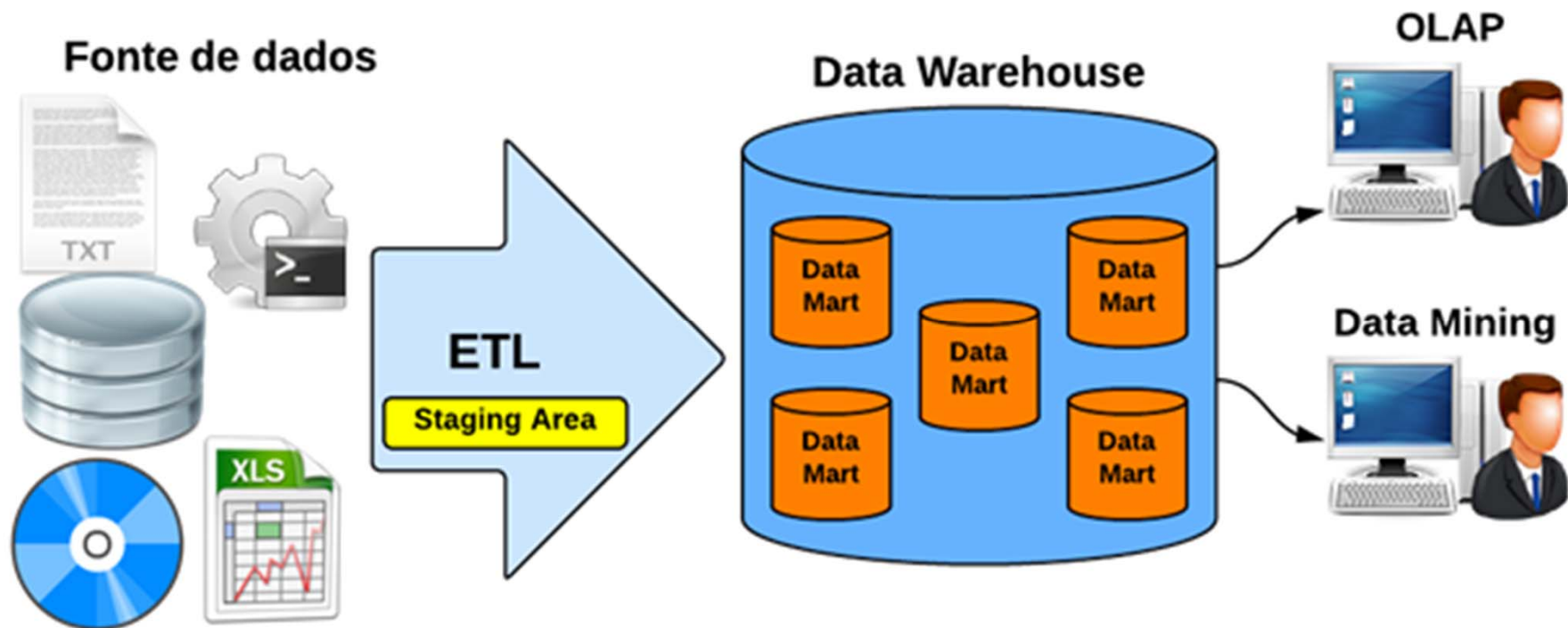


# Benefícios do BI

- Alinhar projetos de tecnologia com as metas estabelecidas pelas empresas na busca do máximo retorno do investimento;
- Ampliar a compreensão das tendências dos negócios, propiciando melhor consistência no momento de decisão de estratégias e ações;
- Facilitar a identificação de riscos e gerar segurança para migração de estratégias, criando maior efetividade nas implementações dos projetos;
- Permitir um planejamento corporativo mais amplo, substituindo soluções de menor alcance por resultados integrados pela informação consistente;

# Benefícios do BI

- Gerar, facilitar o acesso e distribuir informação de modo mais abrangente para obter envolvimento de todos os níveis da empresa e todos aqueles que podem valor usando-a;
- Ligar e consolidar dados de diferentes sistemas de modo a oferecer um visão comum do desempenho da empresa;
- Automatizar tarefas eliminando os erros ao colocar as pessoas no fim dos processos;
- Oferecer dados estratégicos para análise com um mínimo de atraso em relação a uma transação ou evento dentro da empresa;



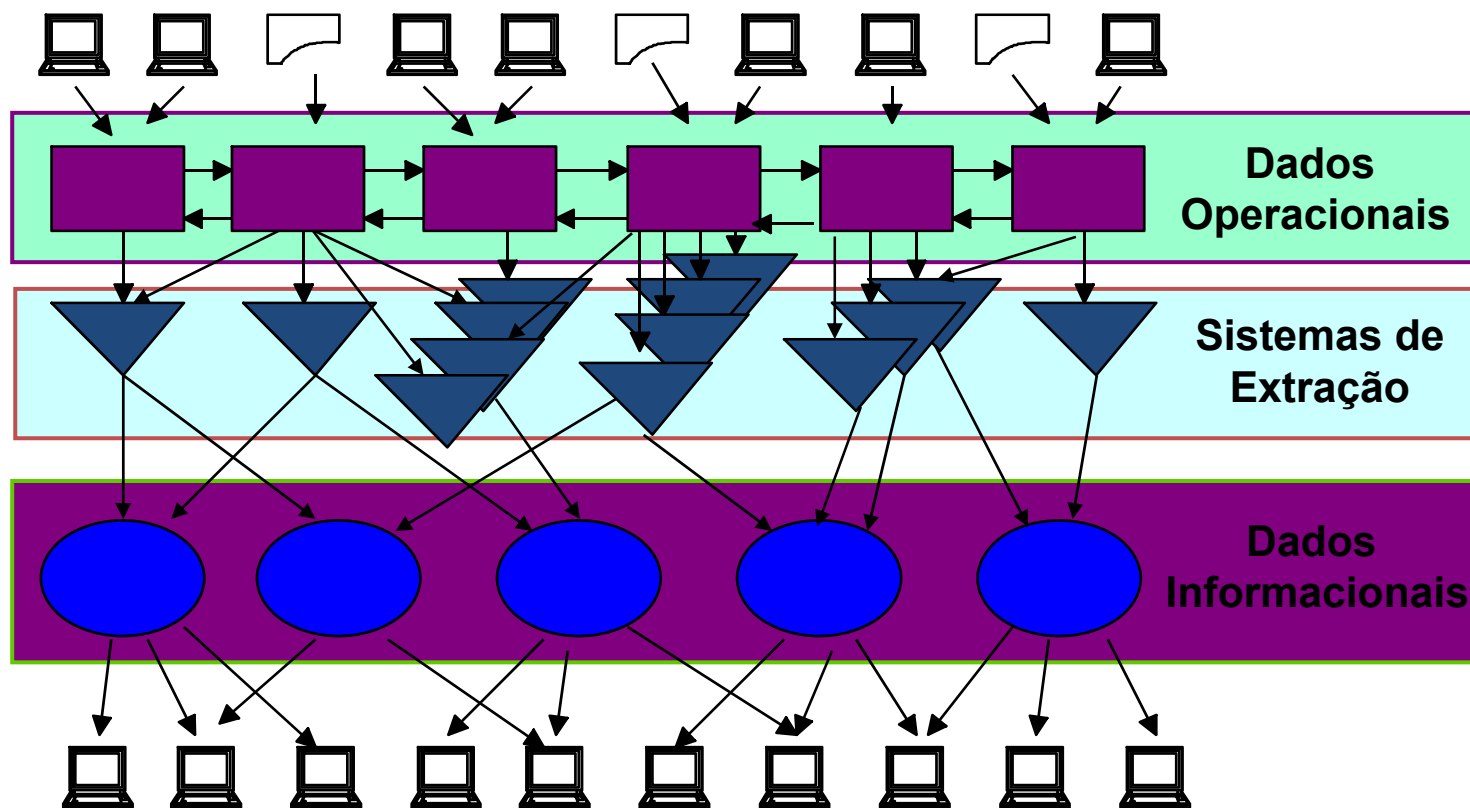
# *Data Warehouse*

- Um armazém de dados o qual integra e gerencia o fluxo de informações a partir dos bancos de dados corporativos.
- Banco de dados para apoio à decisão
- Dados em estruturas lógicas dimensionais
- Dados arrumados e etiquetados em prateleiras de fácil acesso
- *Data Marts* - depósito de dados para áreas específicas da empresa



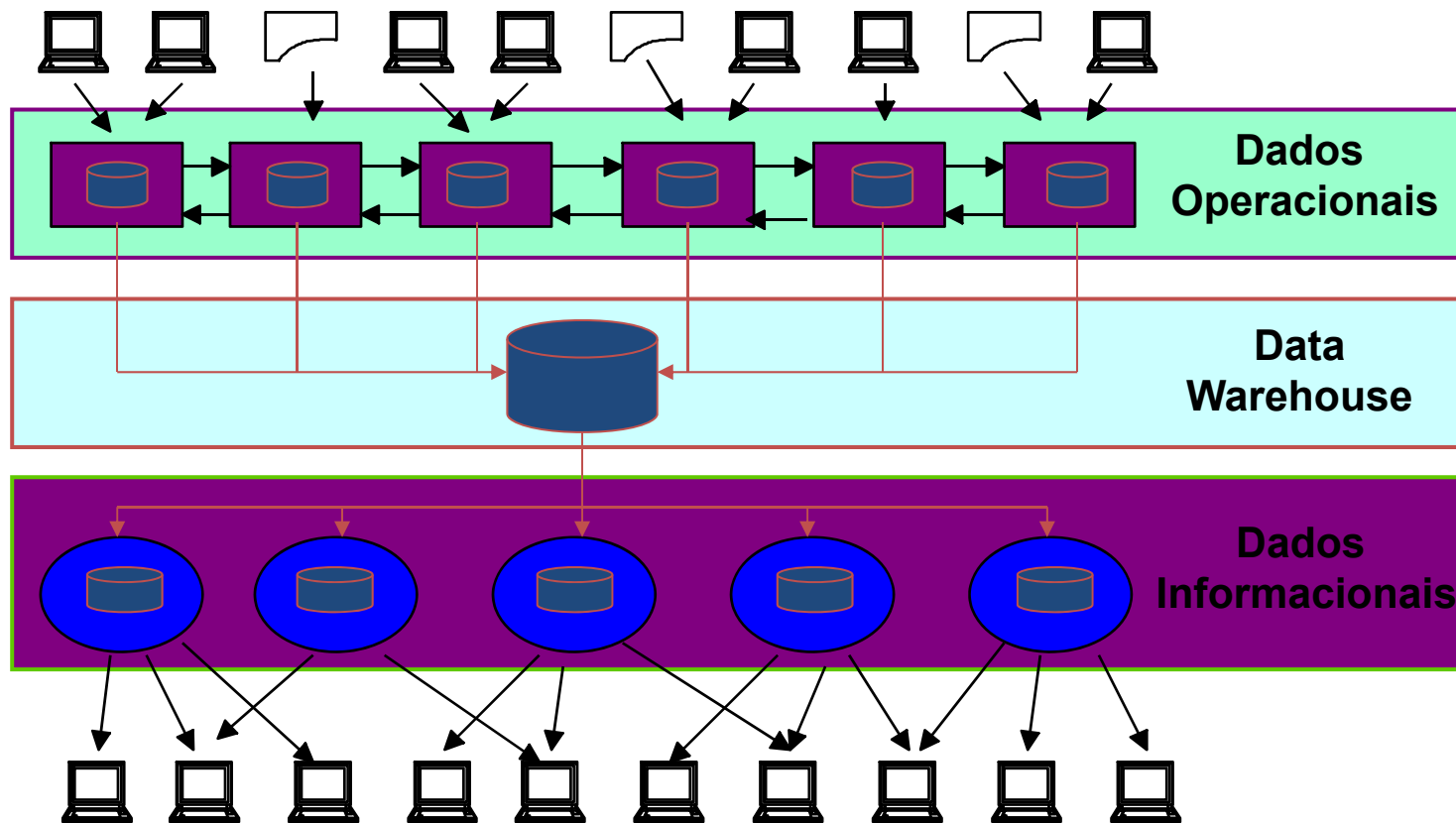
# Arquitetura do *Data Warehouse*

- Sistemas de Extração Tradicionais [Orr]



# Arquitetura do *Data Warehouse*

- Sistemas baseados em Data Warehouse [Orr]



# Arquitetura do *Data Warehouse*

- 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

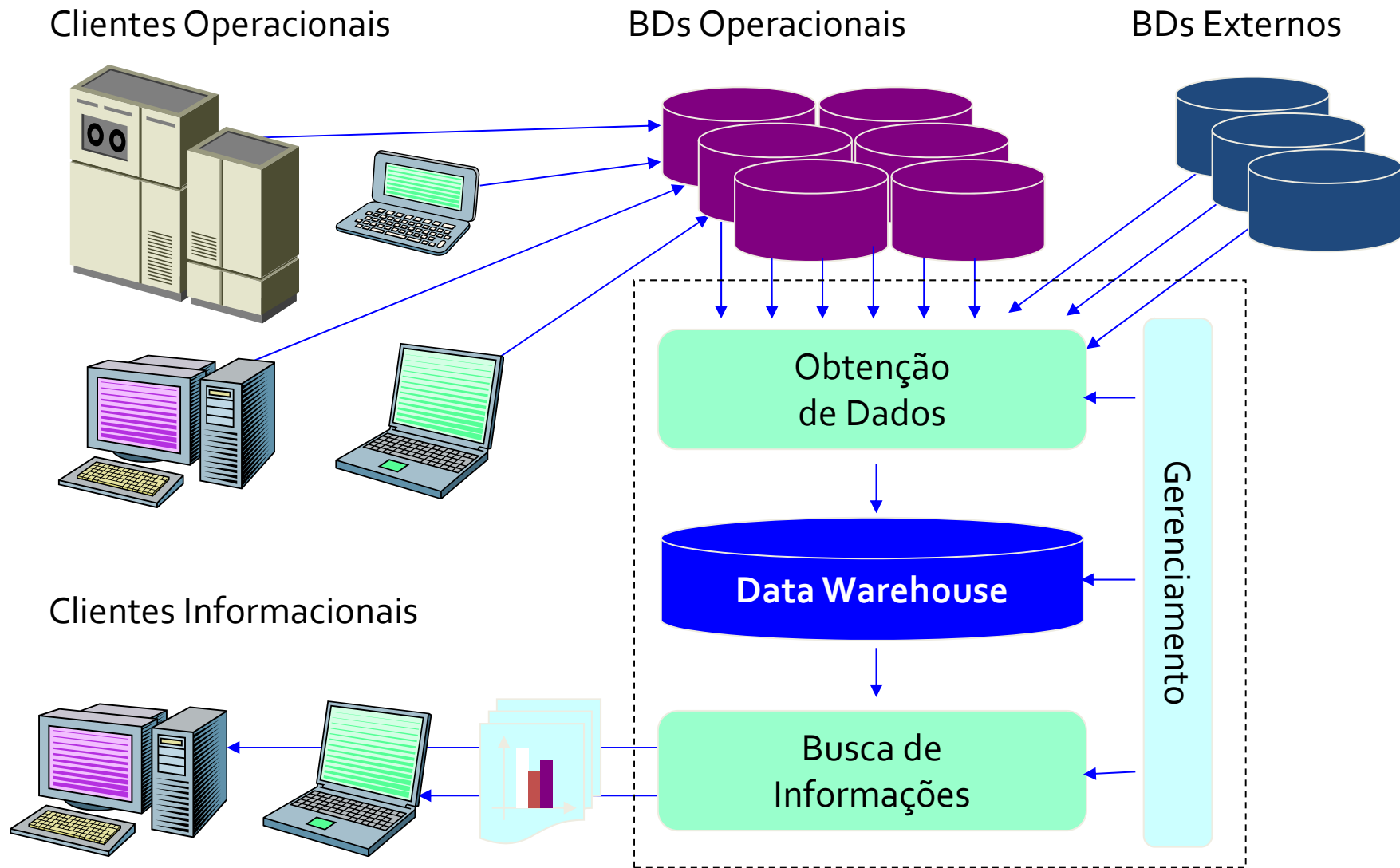
# Estrutura Interna

- 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

# Estrutura Interna

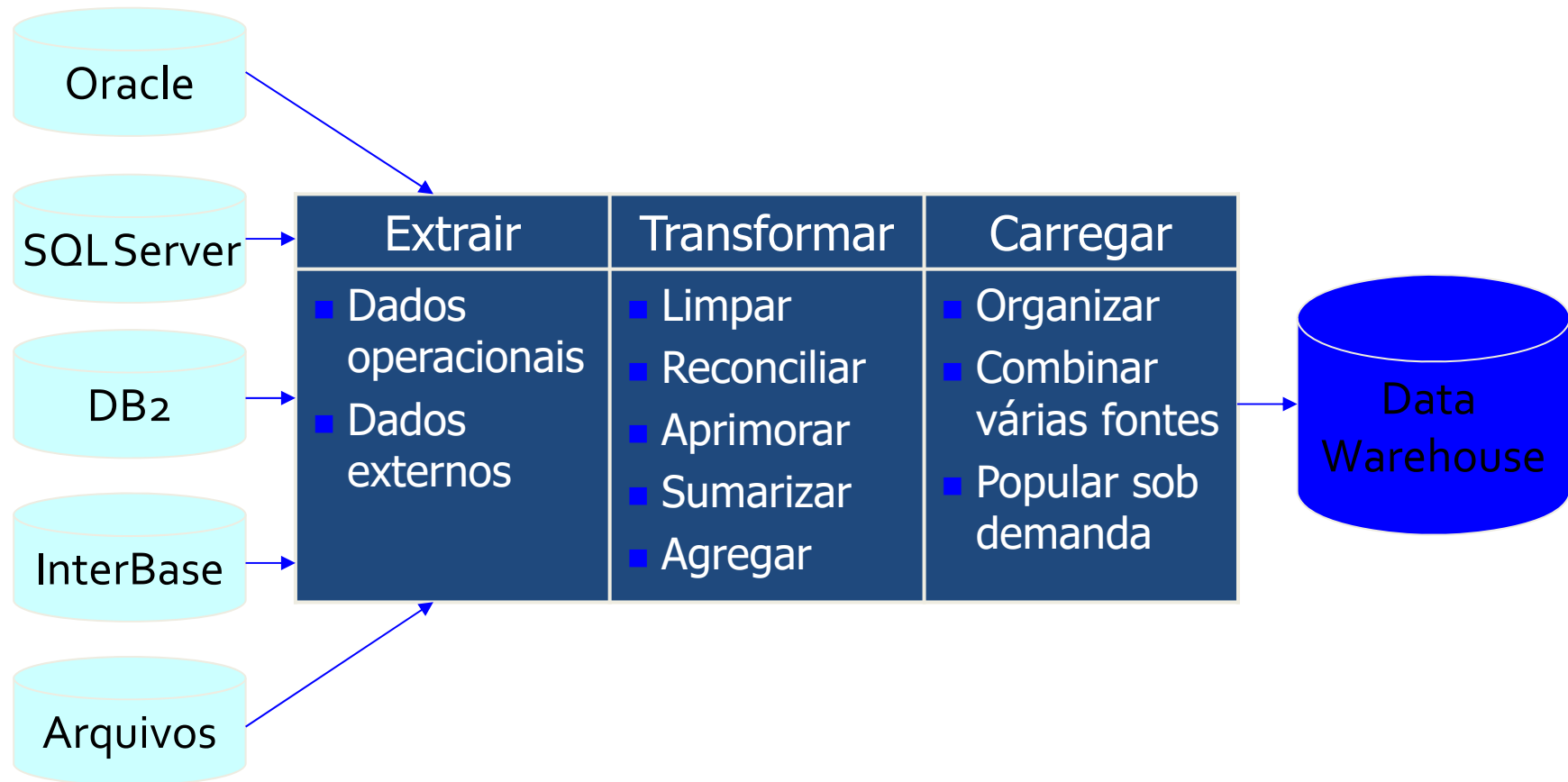
- Em geral são usados BDs relacionais para armazenar os dados do DW
  - Capazes de manter e processar grandes volumes de dados
  - Otimizados para lidar com dados imutáveis
- As ferramentas de análise empregam:
  - Técnicas de mineração de dados
  - Inteligência artificial: redes neurais, *fuzzy*, etc.
  - A Internet: *Web mining*, agentes móveis, etc.

# Estrutura Interna



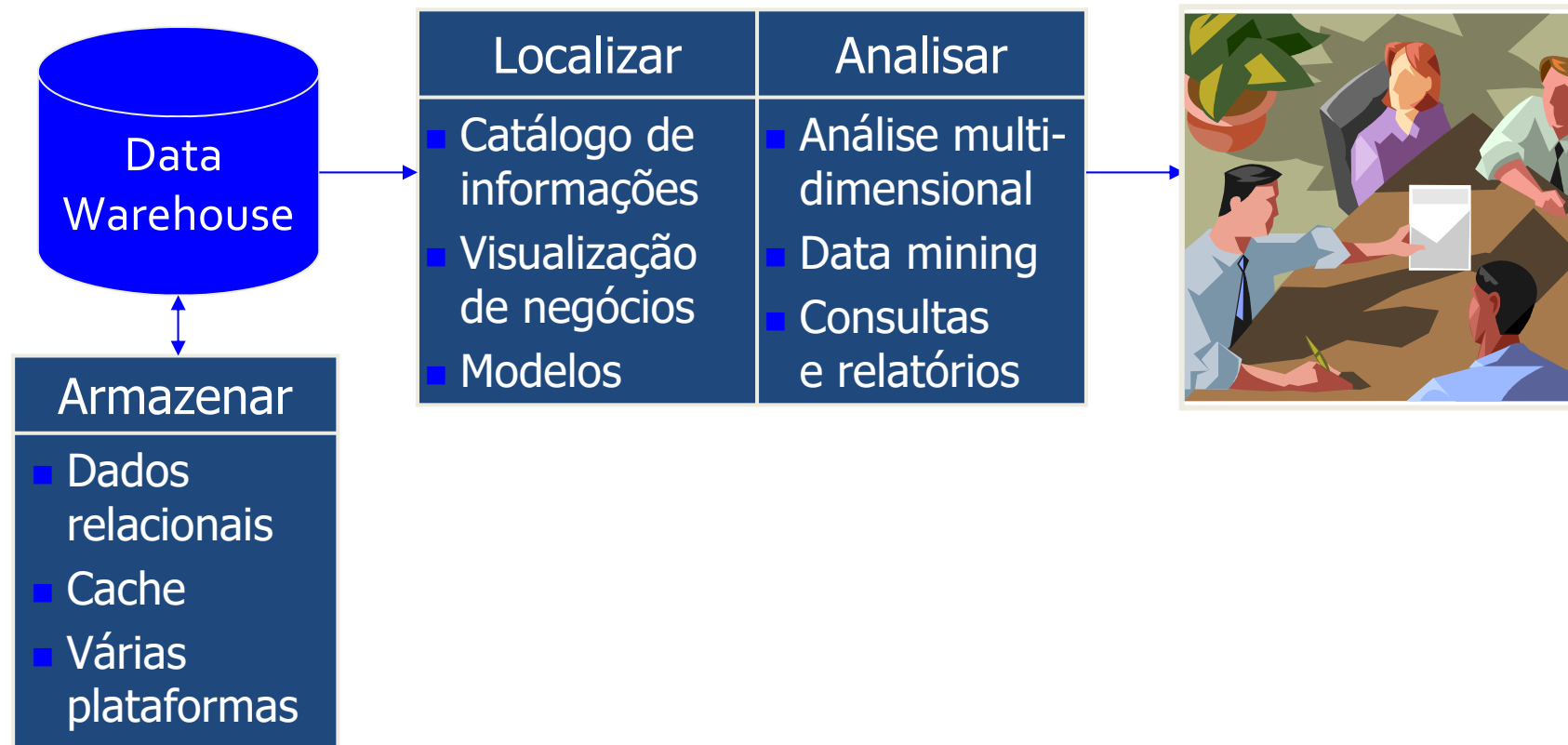
# Estrutura Interna

- Obtenção de Dados



# Estrutura Interna

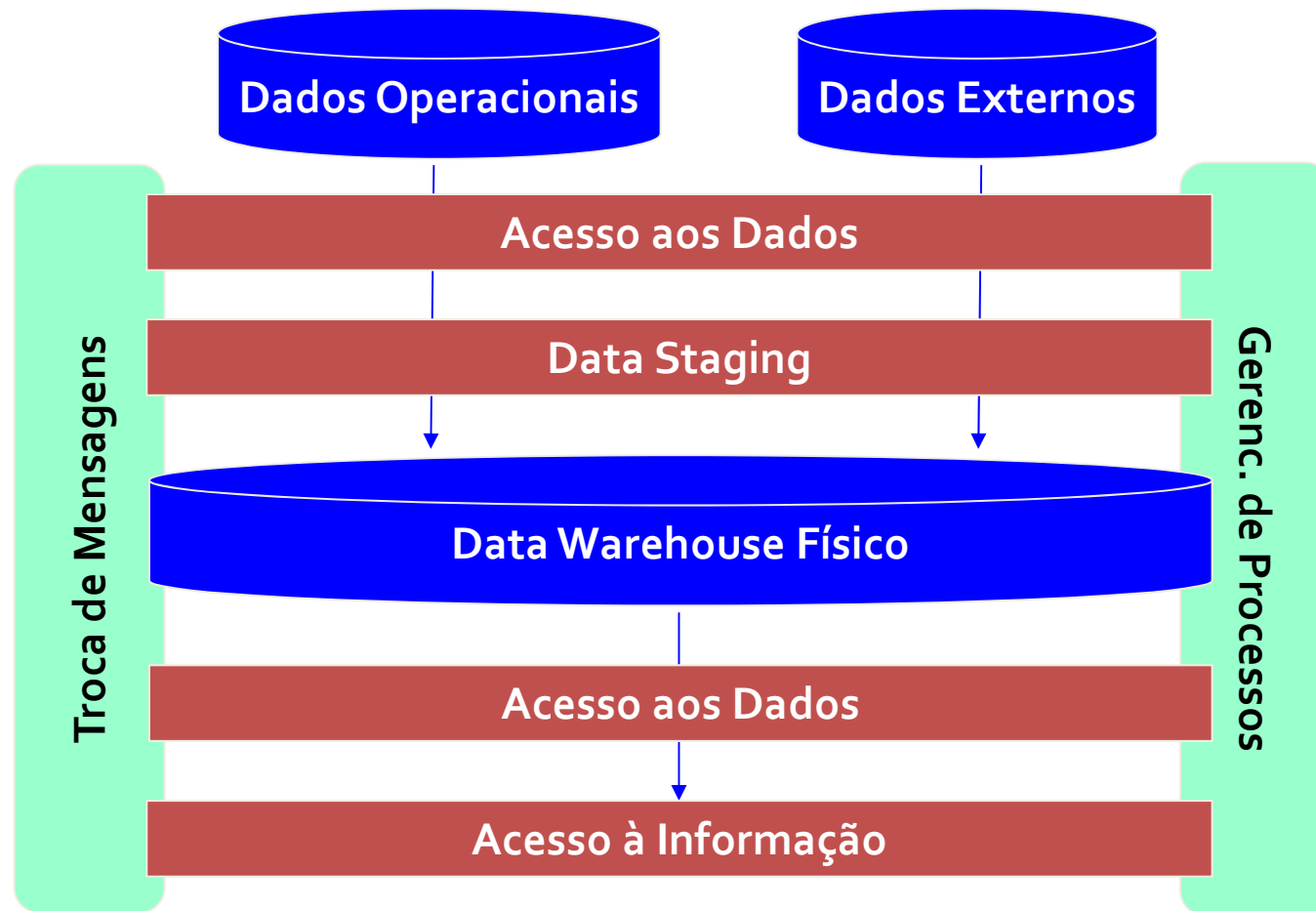
- Busca de Informações





# Estrutura Interna

- Modelo de Camadas



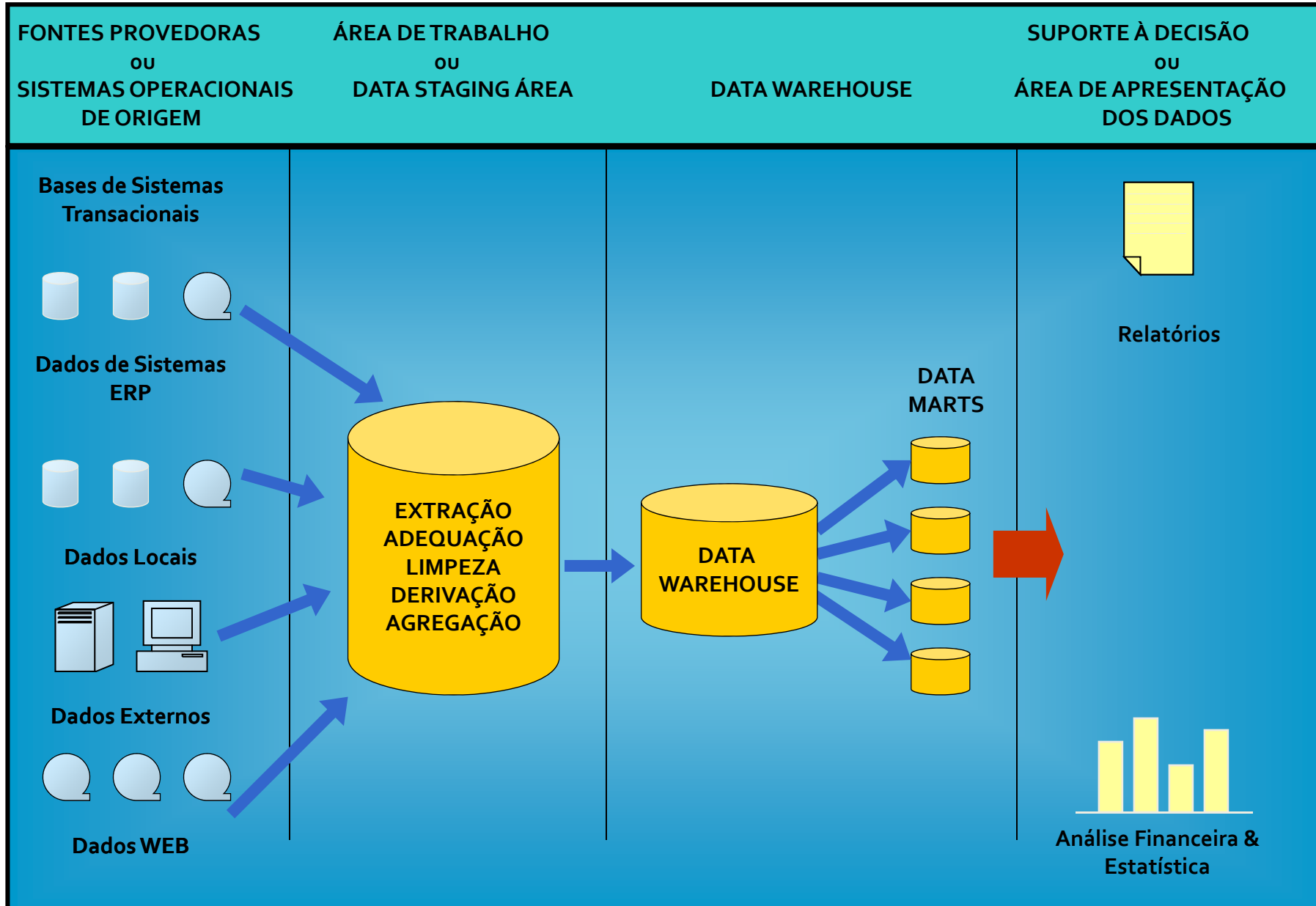
# Estrutura Interna

- 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
  - Gerenc. de Processos: controlar atividades

# Desenvolvimento de *DW*

- Identificação dos dados - modelagem de dados voltada para áreas específicas
- Catalogação das informações gerenciais, possibilitando a busca por assunto
- Extração, Transformação e Carga (ETC) dos dados oriundos de diversas fontes da empresa, com rotinas de consolidação, sumarização e agregação dos dados
- Disponibilização através de interfaces gráficas amigáveis

# Processo de *Data Warehousing*



# Descoberta de Conhecimento em Bases de Dados

- Processo não trivial de identificação de padrões válidos, inéditos, potencialmente úteis e essencialmente compreensíveis.
- Nova tendência para aplicações comerciais
- Busca de correlações escondidas em altos volumes de dados, não evidentes nos sistemas de informação do cotidiano da empresa

## *Data Mining*

- Aplicação de algoritmos de baixo nível projetados para analisar ou extrair padrões de dados. (Fayyad, 1996)
- Mais que a interpretação dos dados – descobrir possíveis fatos e correlações não explicitadas nos SGDDs ou DW

# *Data Mining* para Descoberta de Regras de Associação

- $X \rightarrow Y$ , onde:
  - X e Y são conjuntos de itens de dados
- Ex: “30% das transações de compra realizadas em um supermercado que contêm cerveja, também contêm fraldas descartáveis, sendo que 2% do total de compras contêm ambos os produtos”.

## Aplicações - Regras de Associação

- Planejamento de catálogos
- Promoções
- Disposição física dos produtos
- Estratégia de vendas
- Segmentação de clientes



# *Data Mining* para Classificação

- Tarefa de previsão
  - Vários atributos de entrada e um de classificação
- Aplicações
  - Determinar risco de empréstimo bancário
  - Detectar fraude em cartão de crédito ou seguro
  - Identificar compradores em potencial para um produto

# *Data Mining x Data Warehouse*

- Pergunte ao seu DW:
  - “ Qual o valor médio para o pagamento de seguros de vida para fumantes, na região metropolitana de Fortaleza em setembro de 2009?”
- Pergunte à sua ferramenta DM:
  - “Que perfil de cliente tem mais chances de participar de um acidente de automóvel?”

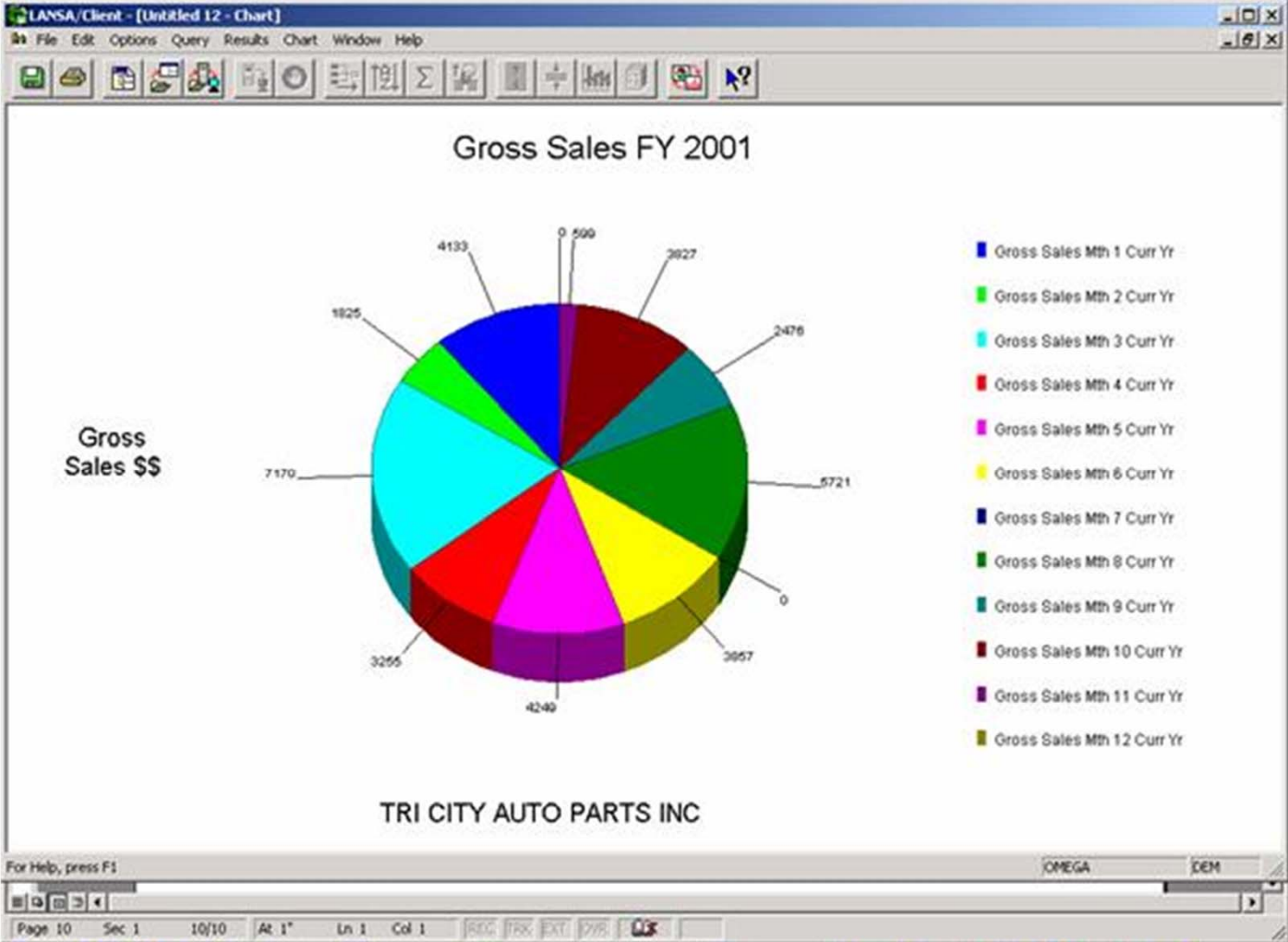
# *OLAP – Processamento Analítico On-line*

- É constituído de um conjunto de tecnologias especialmente projetadas para dar suporte ao processo decisório através de consultas, análises e cálculos sofisticados nos dados corporativos estando armazenados em um Data warehouse por partes dos seus usuários, geralmente analistas, gerentes e executivos.

# Exemplos de visualizações finais dos usuários

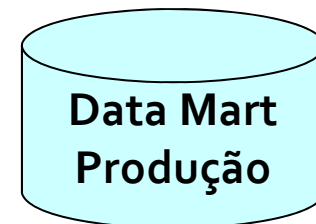
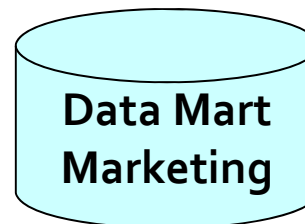
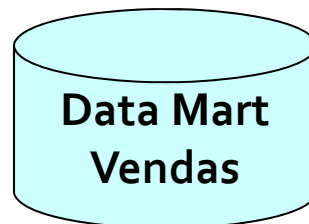
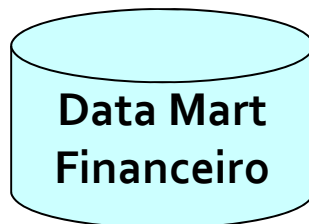


## *Exemplos de visualizações finais dos usuários*



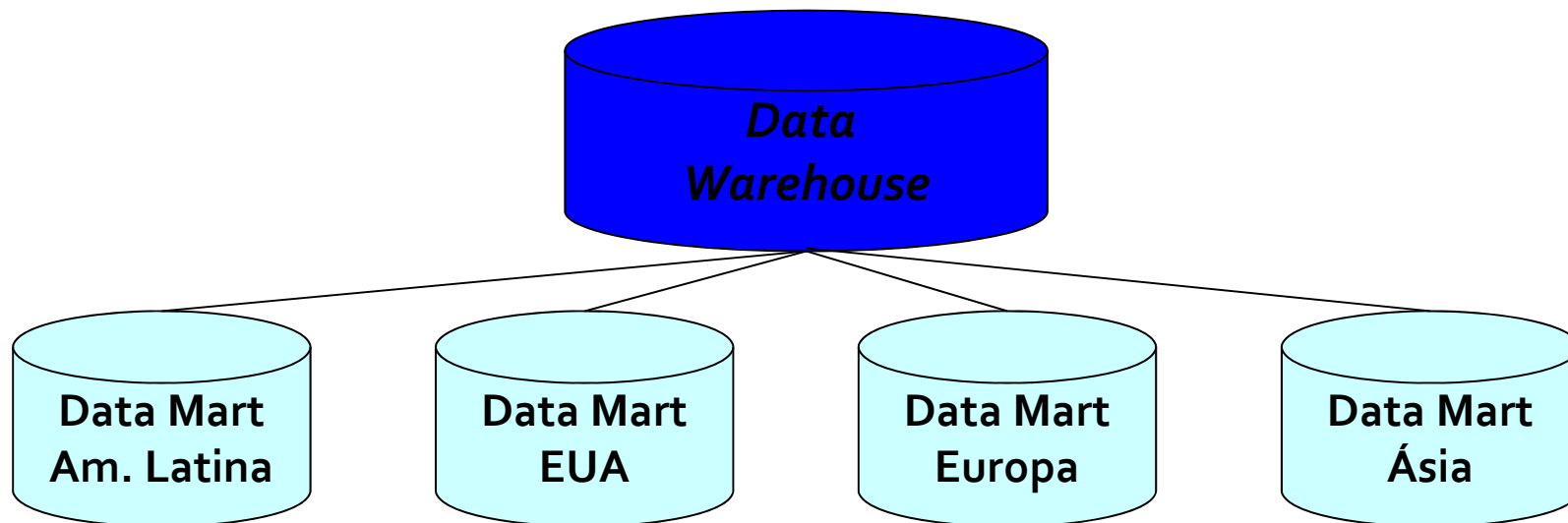
# *Data Marts*

- Dados mantidos no DW são 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



# *Data Marts*

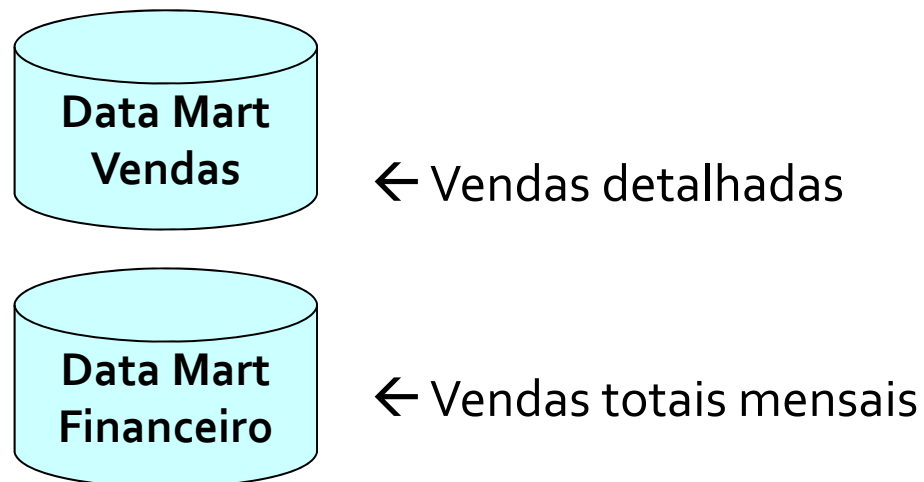
- Um Data Mart desempenha o papel de um DW departamental, regional ou funcional
- Uma empresa pode construir seus Data Marts gradativamente a partir do DW



# *Data Marts*

- Dados podem ser repetidos em dois ou mais Data Marts
- Os mesmos dados podem estar representados com granularidade diferente

Ex:





# Metadados

- Os Metadados são dados sobre os dados
  - Para cada atributo mantido no DW há uma entrada no dicionário de dados
  - Os dados são processados, atualizados e consultados partindo dos metadados
  - Usuários ficam conhecendo a estrutura e o significado dos dados
  - No BD operacional, a estrutura e o significado dos dados estão embutidos nas aplicações

# Metadados

- Camadas de Metadados
  - Metadados Operacionais
    - ✓ Definem a estrutura dos dados operacionais
  - Metadados do DW
    - ✓ Orientados por assunto
    - ✓ Informam como os dados do DW foram calculados e como devem ser interpretados
  - Metadados do Usuário
    - ✓ Organizam os metadados do DW com base em conceitos familiares ao usuário final

# Metadados

- Classificação em função dos dados descritos
  - Metadados de Mapeamento
    - ✓ Como BDs operacionais são mapeados no DW
  - Metadados de Sumarização
    - ✓ Como os dados foram sumarizados no DW
  - Metadados Históricos
    - ✓ Como a estrutura dos dados vem mudando
  - Metadados de Padrões de Acesso
    - ✓ Como os dados do DW vem sendo acessados
  - Metadados de Miscelânea

# Metadados

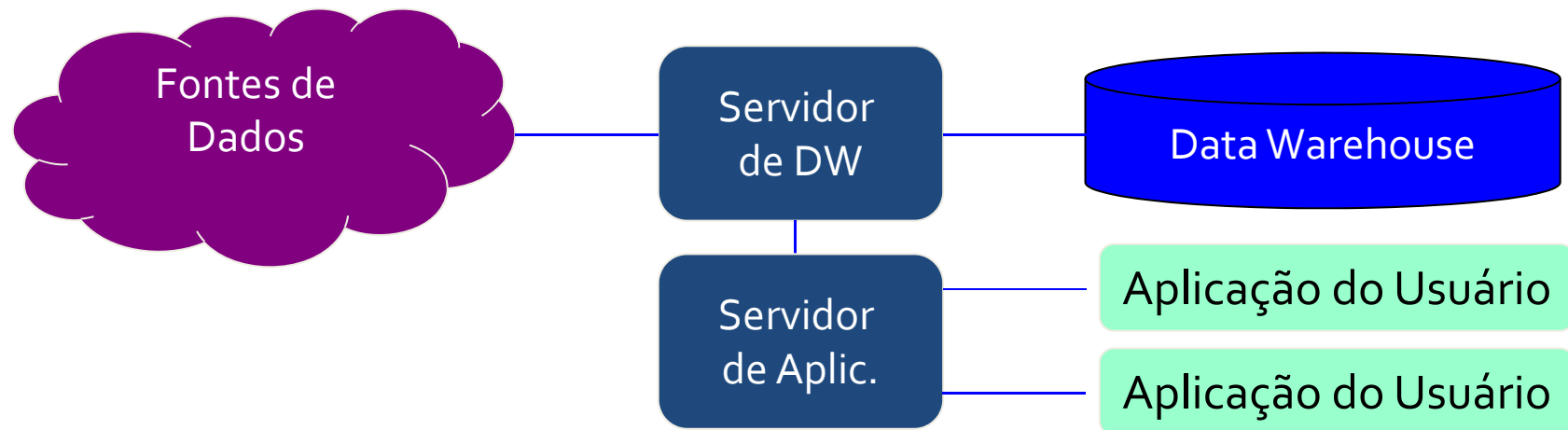
- Fontes de Metadados
  - Código fonte dos SBDs operacionais
  - Diagramas CASE de BDs operacionais e do DW
  - Documentação dos BDs operacionais e do DW
  - Entrevistas com usuários, administradores e programadores dos BDs e do DW
  - O ambiente de DW
    - ✓ Frequência de acesso aos dados, tempo de resposta, controle de usuários etc.

# Acesso aos Dados

- Acesso em Duas Camadas



- Acesso em Três Camadas



# Tipos de *Data Warehouse*

- DW baseado em Servidor
  - Mainframe ou servidor de rede local (LAN)
- DW Virtual
  - Reúne dados operacionais e dados históricos mantidos em BDs – não há um DW central
- DW Distribuído
  - DW global reúne dados de vários DWs locais
- DW baseado na Web
  - Dados provenientes da World Wide Web

Obrigado!

...e agora suas perguntas?



*ricardo.avila@outlook.com.br*



*@theavila*