

**ITnerante**

**TIMASTERS**

Este curso lida com um conjunto de tecnologias que dão **suporte as decisões gerenciais** por meio do **informações** internas e externas às organizações. Essas tecnologias tem um profundo impacto na **estratégia corporativa**, na **performance** e na **competitividade**. Esse conjunto de tecnologias são coletivamente conhecidas como

**BUSINESS INTELLIGENCE**



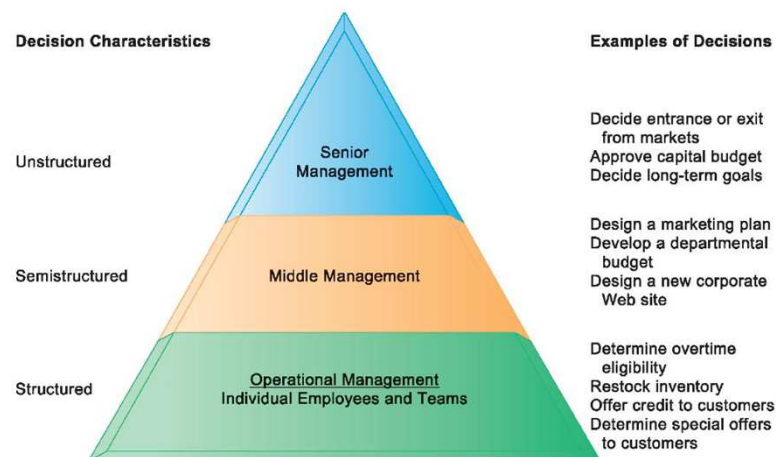
# Banco de dados

## Suporte a decisão – Business Intelligence

---

**Curso Preparatório - ITnerante**

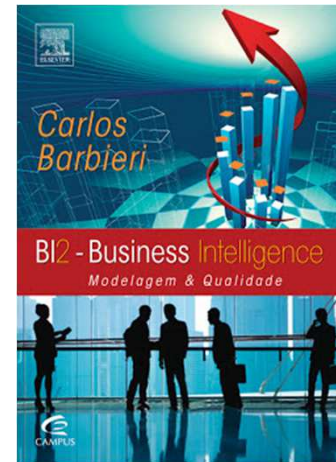
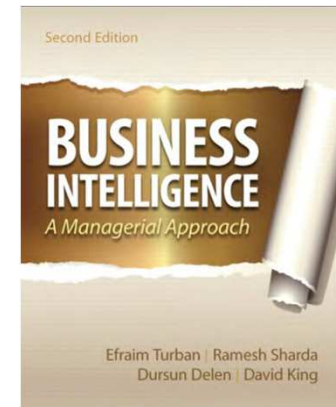
*Prof. Thiago Cavalcanti*



# Ementa

---

- Módulo 01 – Soluções de suporte à decisão:
  - Inteligência de negócios
    - Business intelligence (BI)
  - Data Warehouses e Data Marts.
  - Processo de Data Warehousing.
  - Modelagem multidimensional.
  - Extração, transformação e carga (ETL).



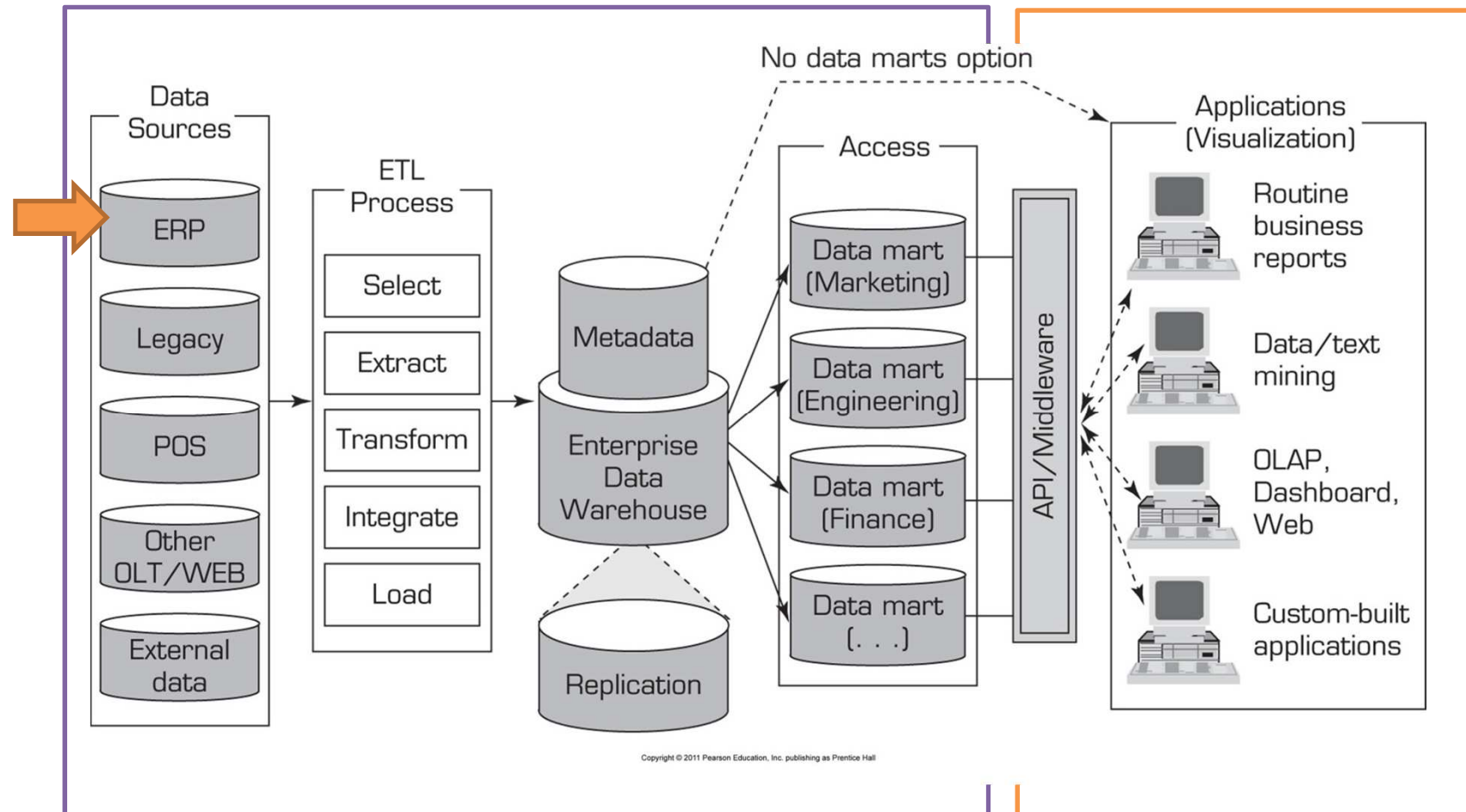
# Ementa

---

- Módulo 02 – Soluções de suporte à decisão:
  - Recuperação e visualização de dados
    - OLAP
    - Painéis e dashboards
    - Data Mining
  - ERP
  - Qualidade de dados.
  - Gestão de conteúdo (ECM).



# Ementa - Entendendo os módulos



# Tire suas dúvidas

---

- [rcthiago@gmail.com](mailto:rcthiago@gmail.com)
- [www.itnerante.com.br](http://www.itnerante.com.br)
- Lista: [timasters@yahoogroups.com.br](mailto:timasters@yahoogroups.com.br)



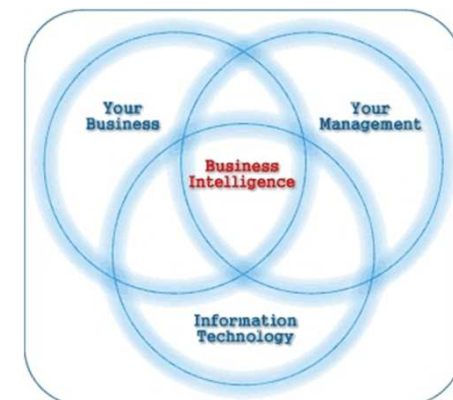


**ITnerante**

**TIMASTERS**

---

# BUSINESS INTELLIGENCE



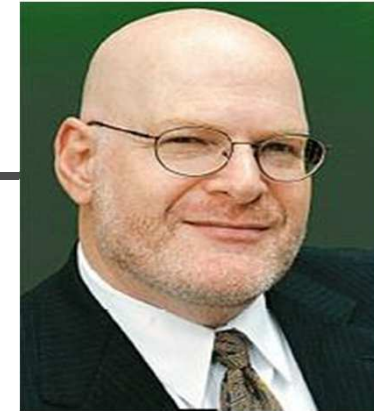
Prof. Thiago Cavalcanti

**ITnerante**

## Definição: Howard Dresner

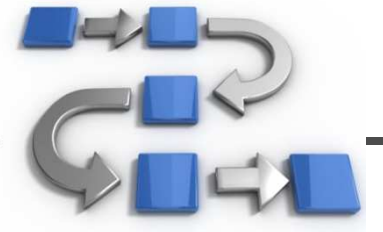
---

- “Um conjunto de conceitos, métodos e recursos tecnológicos que habilitam a obtenção e distribuição de informações geradas a partir de dados operacionais, históricos e externos, visando proporcionar subsídios para a tomada de decisões gerenciais e estratégicas.”
  - Gartner Group, 1989



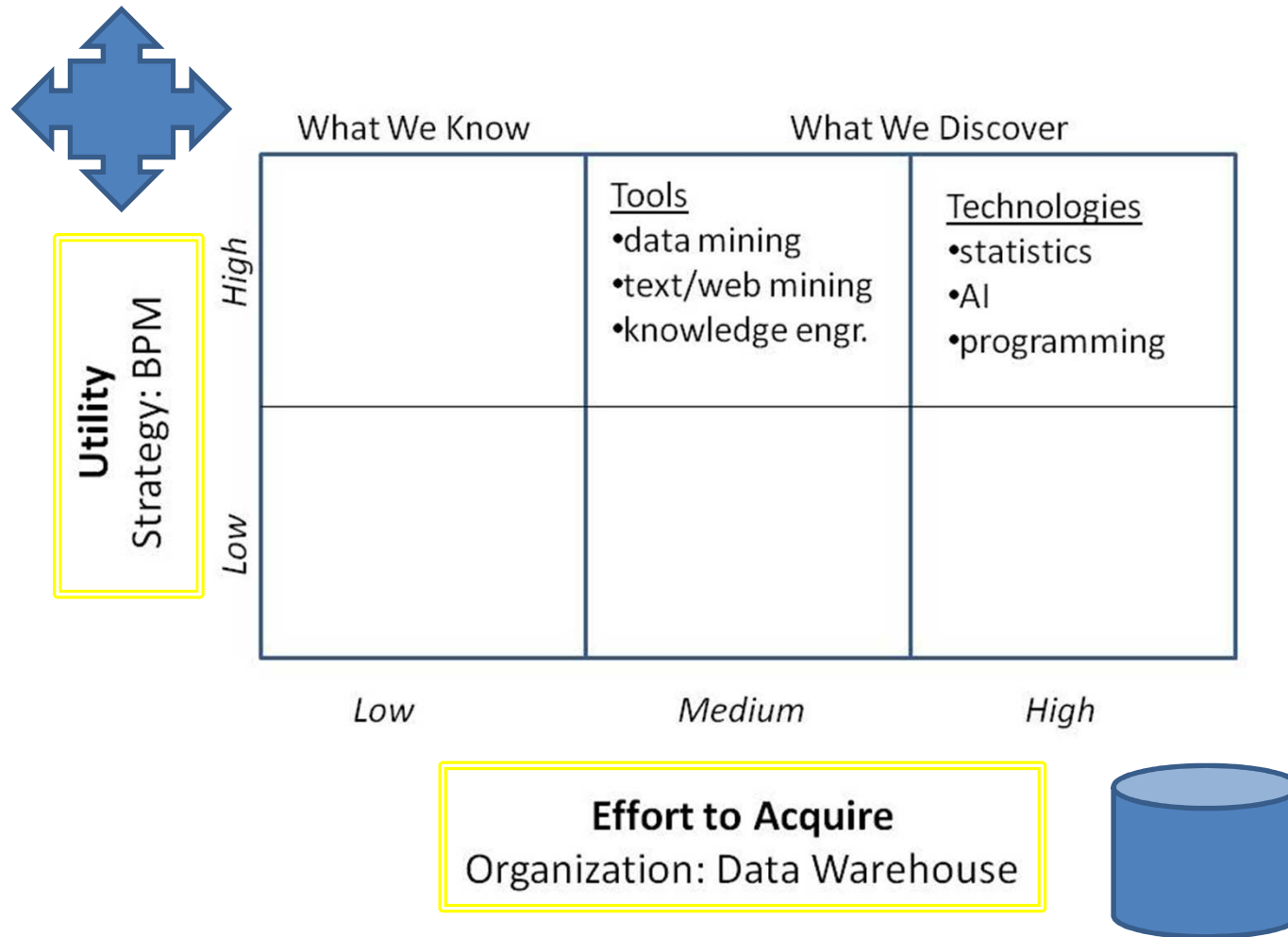
# Motivação

---



- Operacional x Analítico
  - Diferentes necessidades, estruturas e clientes
- Objetivos fundamentais de BI/DW (Kimball)
  - Fazer **a informação acessível** mais facilmente
  - Apresentar **a informação consistente** (credibilidade)
  - Adaptado a mudanças
  - Apresentar a **informação de forma temporal**
  - Um **bastião de segurança** que protege os ativos de informação
  - Servir como base de autoridade e de confiança para uma melhor tomada de decisão
  - Ser aceito pela comunidade organizacional para considerar bem-sucedido

# Informação



## Conceitos (Efrain Turban)

---

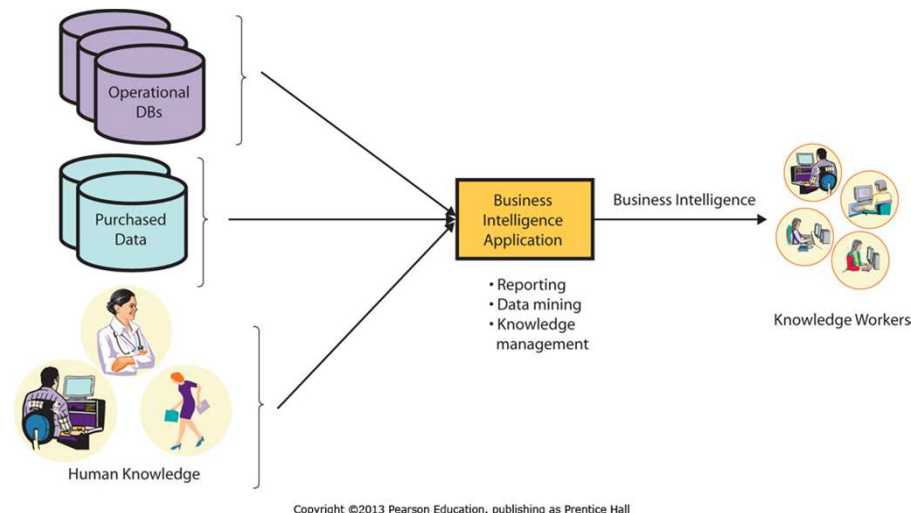
- *Business Intelligence* (BI) é um termo abrangente que combina arquiteturas, ferramentas, bancos de dados, ferramentas de análise, aplicações e metodologias
- Principais objetivos do BI
  - Permitir o acesso interativo (por vezes em tempo real) aos dados
  - Permitir a manipulação de dados
  - Dar aos gestores e analistas a capacidade de realizar análise adequada
- É baseado na transformação dos dados em informação, em seguida, informações em decisões e, finalmente, em ações



# Objetivos fundamentais de BI

---

- Acesso a dados confiáveis
- Aumento da transparência e compreensão do negócio
- Suporte para a tomada de decisão

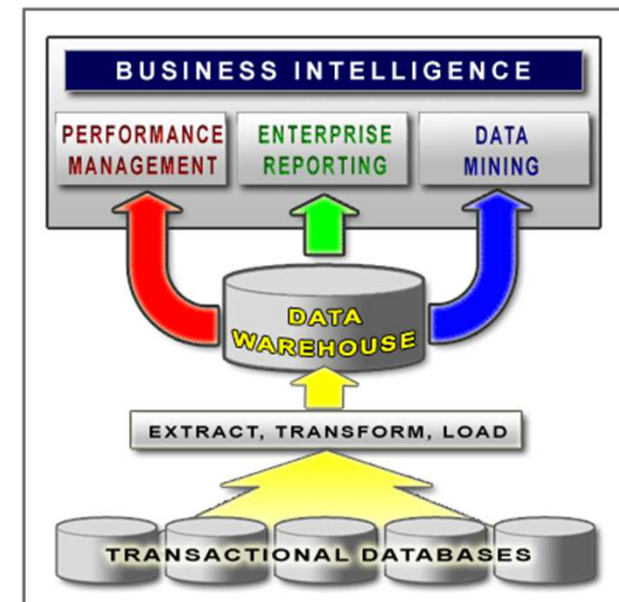




# Conceito

- “BI representa a habilidade de se estruturar, acessar e explorar informações, normalmente guardadas em um DW/DM (Data Warehouse/Data Mart), com o objetivo de desenvolver percepções, entendimentos, conhecimento, os quais podem produzir um melhor processo de tomada de decisão.”

- Carlos Barbieri



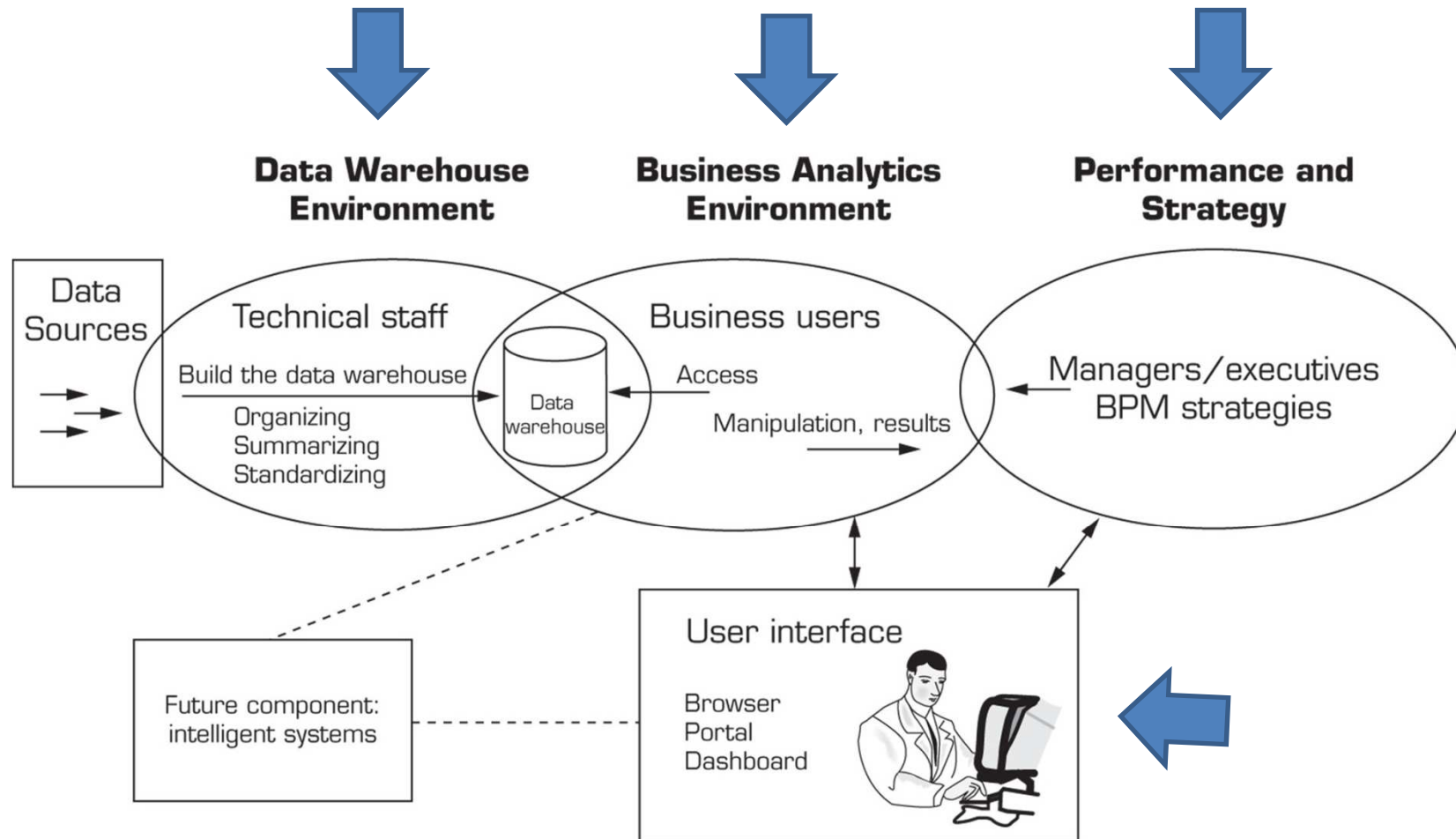
## Questão 01 – STN 2013 - ESAF

---

Q. 28- O processo de *Business Intelligence (BI)* baseia-se em:

- (a) transformação de dados em informações, depois em conhecimento e finalmente em decisões.
- (b) obtenção de dados a partir de informações, transformação em decisões e finalmente transformação em produtos.
- (c) transformação de dados em relações, depois em interfaces e finalmente em ações.
- (d) tomada de decisões, geração de informações e avaliação da relevância dos dados.
- (e) transformação de dados em informações, depois em decisões e finalmente em ações.

# Componentes de Sistemas de BI



Copyright © 2011 Pearson Education, Inc. publishing as Prentice Hall

# Data Warehouse

---

- O **armazém de dados** e suas variantes são a **pedra fundamental** (pedra angular) de qualquer sistema de BI de médio a grande porte
- Originalmente, incluiu apenas **dados históricos** que foram **organizados e resumidos**, para que os usuários finais pudessem facilmente ver ou manipular dados e informações
- Hoje, incluem também dados atuais para que eles possam fornecer apoio à decisão em tempo real

# Business Analytics (Análise de negócios)

---

- Os usuários finais podem trabalhar com os dados e informações em um armazém de dados usando uma variedade de ferramentas e técnicas
- Estas se enquadram em duas categorias principais:
  - 1. Relatórios e consultas**
    - Análise de negócios incluem relatórios estáticos e dinâmicos, todos os tipos de consultas, a descoberta de informações, visão multidimensional, *drill down*, ...
    - Esses relatórios também estão relacionados com BPM (introduzido no próximo slide)
  - 2. Dados, texto e mineração Web e outras ferramentas matemáticas e estatísticas sofisticadas**
    - A mineração de dados é um processo de busca de relações desconhecidas ou informações em grandes bases de dados ou armazéns de dados, utilizando ferramentas inteligentes como a computação neural, técnicas de análise preditiva, ou métodos estatísticos avançados

# Business Performance Management (BPM)

---

- Uma carteira emergente de aplicativos e metodologias que contém evolução da arquitetura e ferramentas de BI em seu núcleo
  - BPM amplia o monitoramento, medição e comparação das vendas, lucro, custos, rentabilidade e outros indicadores de desempenho, introduzindo o conceito de gestão e feedback
  - Abrange processos tais como planejamento e previsão de princípios fundamentais de uma estratégia de negócios
  - Em contraste com as tradicionais DSS, EIS, e BI, que suportam a extração bottom-up de informação a partir dos dados, BPM proporciona uma aplicação top-down da estratégia corporativa
- Normalmente combinada com a metodologia *Balanced Scorecard* (BSC) e *dashboards*
- Também conhecido como gerenciamento de desempenho corporativo (*CPM - Corporate*)

## A interface com o usuário

---

- *Dashboards* (painéis) fornecem uma visão abrangente e visual das medidas de desempenho corporativo (também conhecido como indicadores chave de desempenho - KPI), tendências e exceções
- Integram informações de várias áreas de negócio
- Apresentam gráficos que mostram o desempenho real em comparação com as métricas desejadas, assim, um painel apresenta uma visão geral da saúde da organização

## A interface com o usuário

---

- Além dos painéis, outras ferramentas que transmitem informações são portais corporativos, cockpits digitais e outras ferramentas de visualização
- Muitas ferramentas de visualização, que vão desde apresentação do cubo multidimensional a realidade virtual, são parte integrante dos sistemas de BI
- BI surgiu de EIS, portanto recursos visuais para os executivos foram transformados em software de BI.
- Tecnologias como sistemas de informações geográficas (SIG) desempenham um papel crescente no apoio à decisão



## Questão 02 - STN 2013 - ESAF

---

29- São componentes do *Business Intelligence (BI)*:

- (a) *Data warehouse (DW)*. Análise de estratégias. Ferramentas para manipular e analisar os dados. *Business Performance Control (BPC)*.
- (b) *Data mining (DM)*. Análise de negócios. *Scorecards* para manipular e analisar os dados. *Business Performance Management (BPM)*.
- (c) *Data warehouse (DW)*. Análise de negócios. Ferramentas para manipular e analisar os dados. *Business Performance Management (BPM)*.
- (d) *Data warehouse (DW)*. Avaliação estratégica e tática de negócios. Ferramentas para manipular e analisar os dados. *Business Corporate Management (BCM)*.
- (e) *Data storeware (DSW)*. Análise de negócios. *Scorecards* para manipular e analisar os dados. *Business Performance Management (BPM)*.

# Fornecedores

---

TERADATA

**MicroStrategy**  
Best In Business Intelligence™

 Microsoft

 **COGNOS**  
AN IBM® COMPANY



**SPSS**

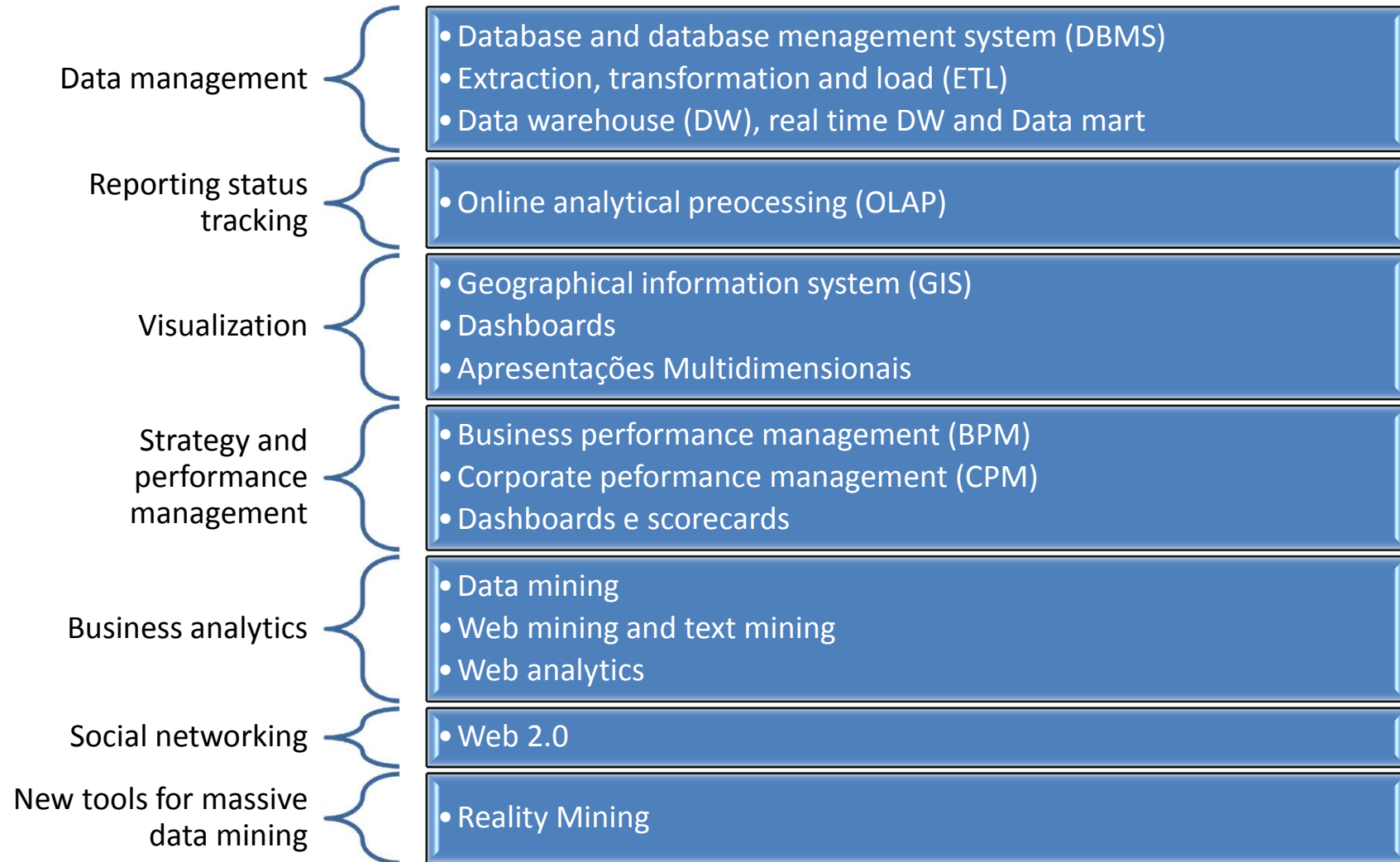
  
**SAP BusinessObjects**

**ORACLE**  
 **Hyperion**  
The future in sight

 **sas**  
TM

- Grandes empresas de software adquirir outros para construir um portfólio completo de ofertas.
  - A SAP adquiriu a Business Objects, a IBM adquiriu a Cognos em 2008 e SPSS em 2009, a Oracle adquiriu a Hyperion.
  - Novas empresas estão surgindo agora no texto, Web, e análise de dados.
  - Além disso, as empresas estão colaborando umas com as outras empresas para a construção de parcerias. Por exemplo, SAS e Teradata firmaram um parceria para oferecer armazenamento de dados e capacidades analíticas de previsão em conjunto.

# Ferramentas computadorizadas para SAD



## Questão 03 - FCC - SEFAZ-SP - Agente Fiscal de Rendas - Tecnologia da Informação – 2009 – Prova 3

---

A área de BI - *Business Intelligence* está diretamente envolvida com os projetos de implementação das aplicações de

- (a) B2B, B2C e BSC.
- (b) EAI, B2B e B2C.
- (c) EAI, CRM e ERP.
- (d) CI, KMS e BSC.
- (e) CRM, PRM e ERP.

# BI2 → Barbieri → Índice :O

---

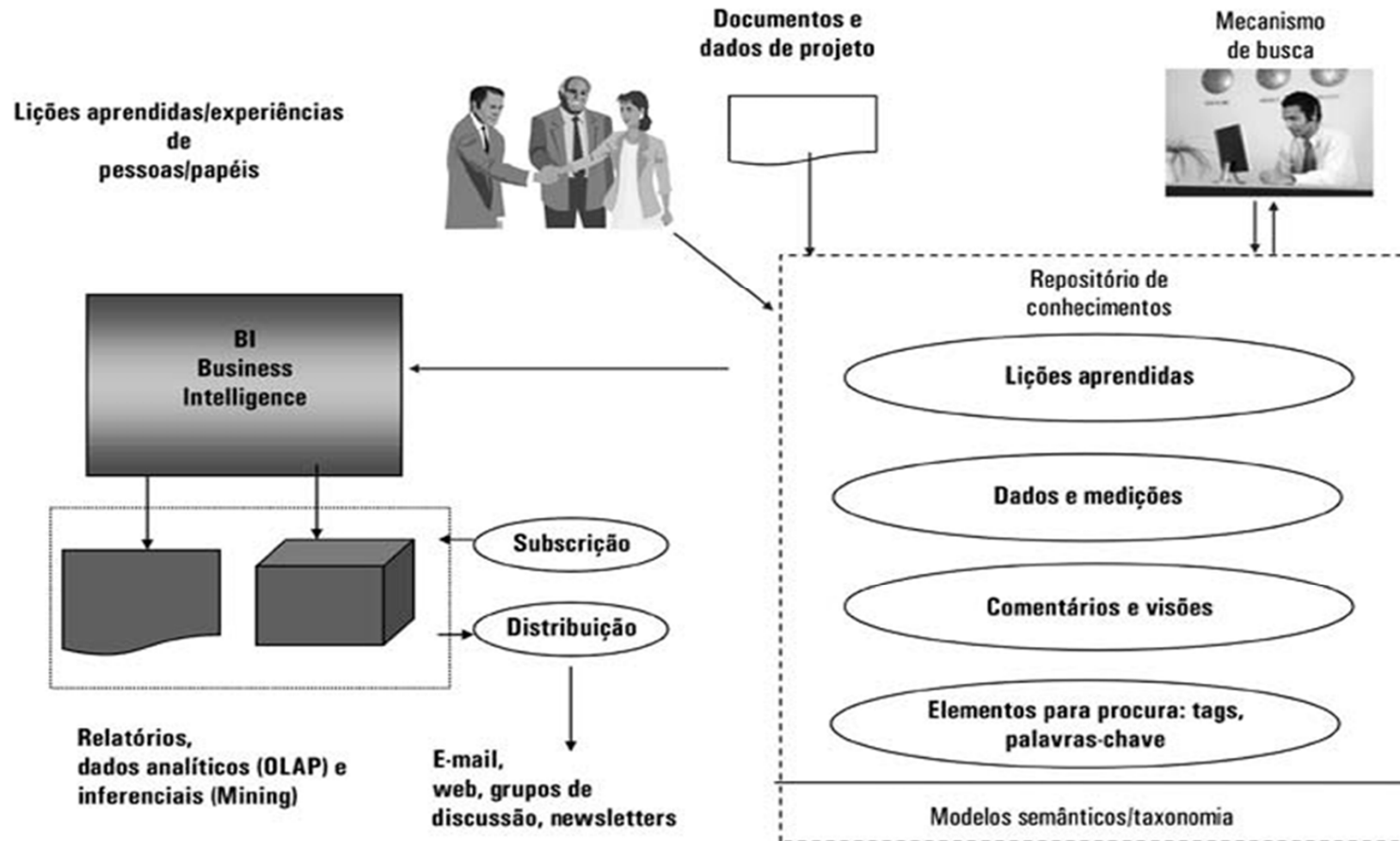
Capítulo 6 – <i>Conceitos correlatos de BI</i> .....	121
Gerência de conhecimento .....	121
Inteligência competitiva .....	123
BSC: Balanced ScoreCard.....	127
Resumo .....	130

## Gerência de Conhecimento (KMS)

---

- Objetiva estabelecer uma aproximação integrada e colaborativa para capturar, criar, organizar e usar todos os ativos de informação de uma empresa.
- Enquanto *Business Intelligence (BI)* é mais *compartimentada, objetiva e focada em estruturas definidas*, a **KMS** trabalha o ativo de informações, independentemente da sua *forma, estrutura e domínio*.

# Visão geral de um KMS



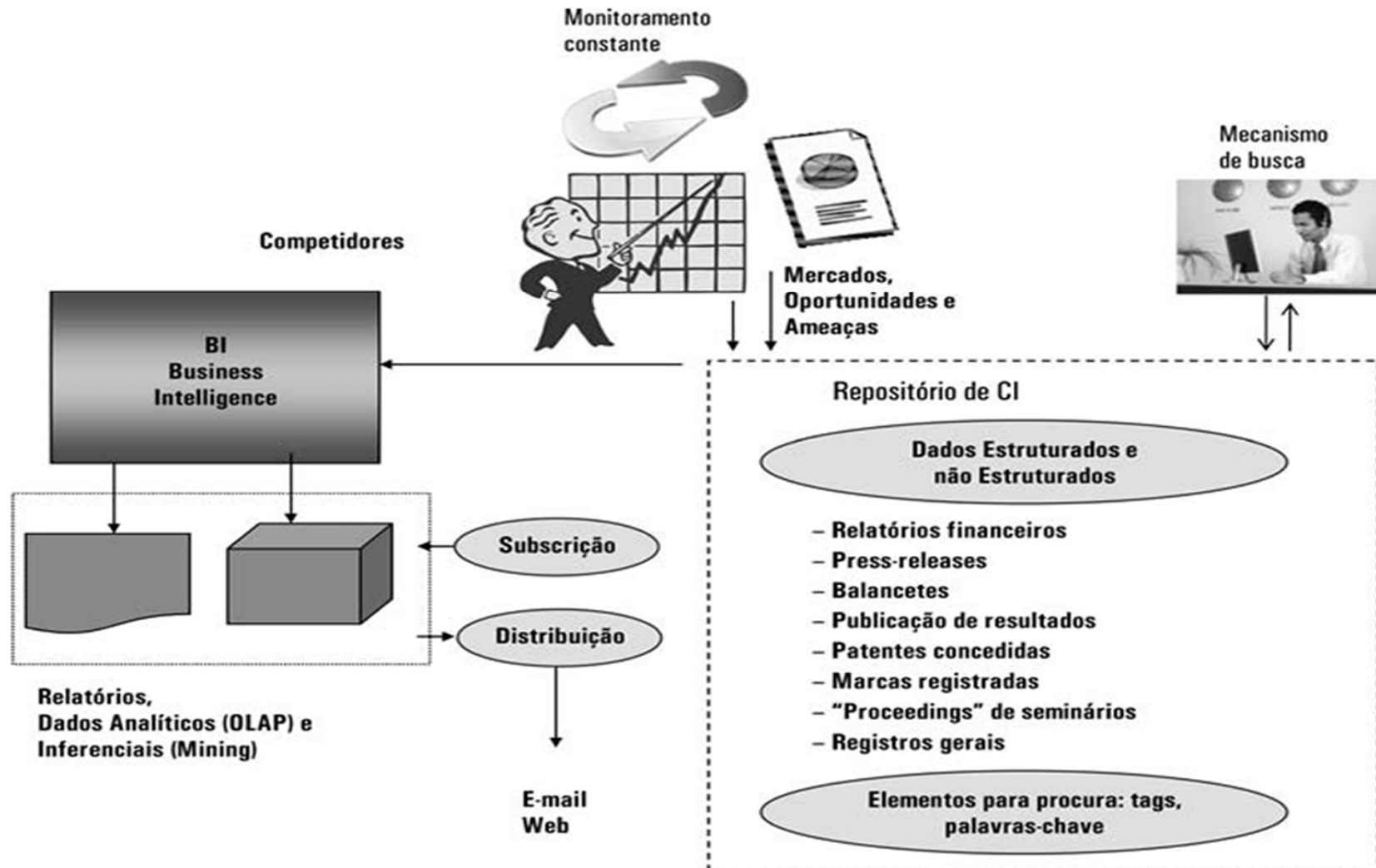
# Inteligência Competitiva

---

- Os sistemas de inteligência competitiva:
  - Têm por objetivo **prover** as organizações de um processo sistemático de coleta e análise de informações sobre as atividades correlatas e tendências de mercado.
  - Visam **absorver** o maior número de informações que possibilitem melhoria no processo de tomada de decisão e de desenvolvimento empresarial.
    - Romani et al

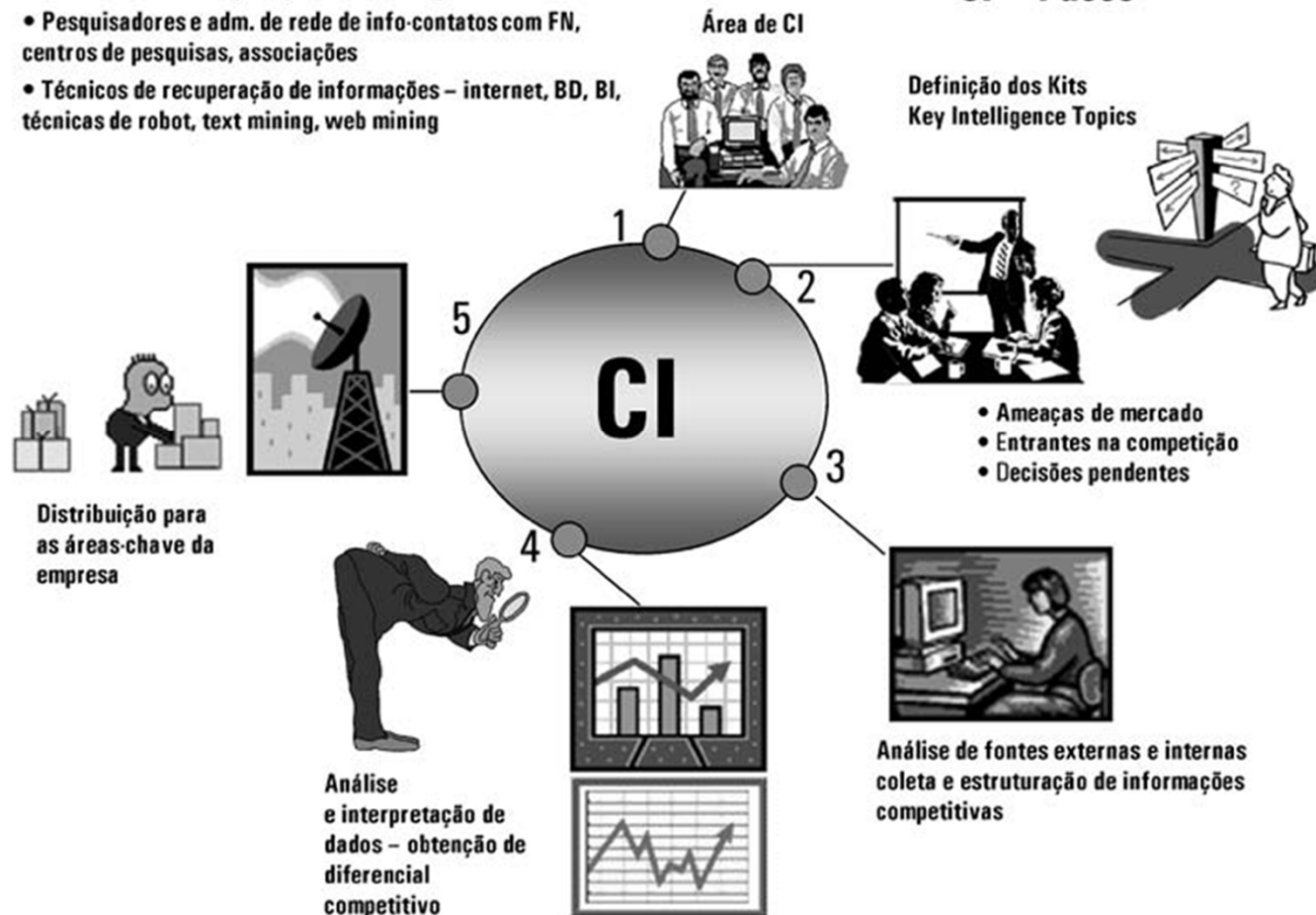


# Conceito de Inteligência Competitiva

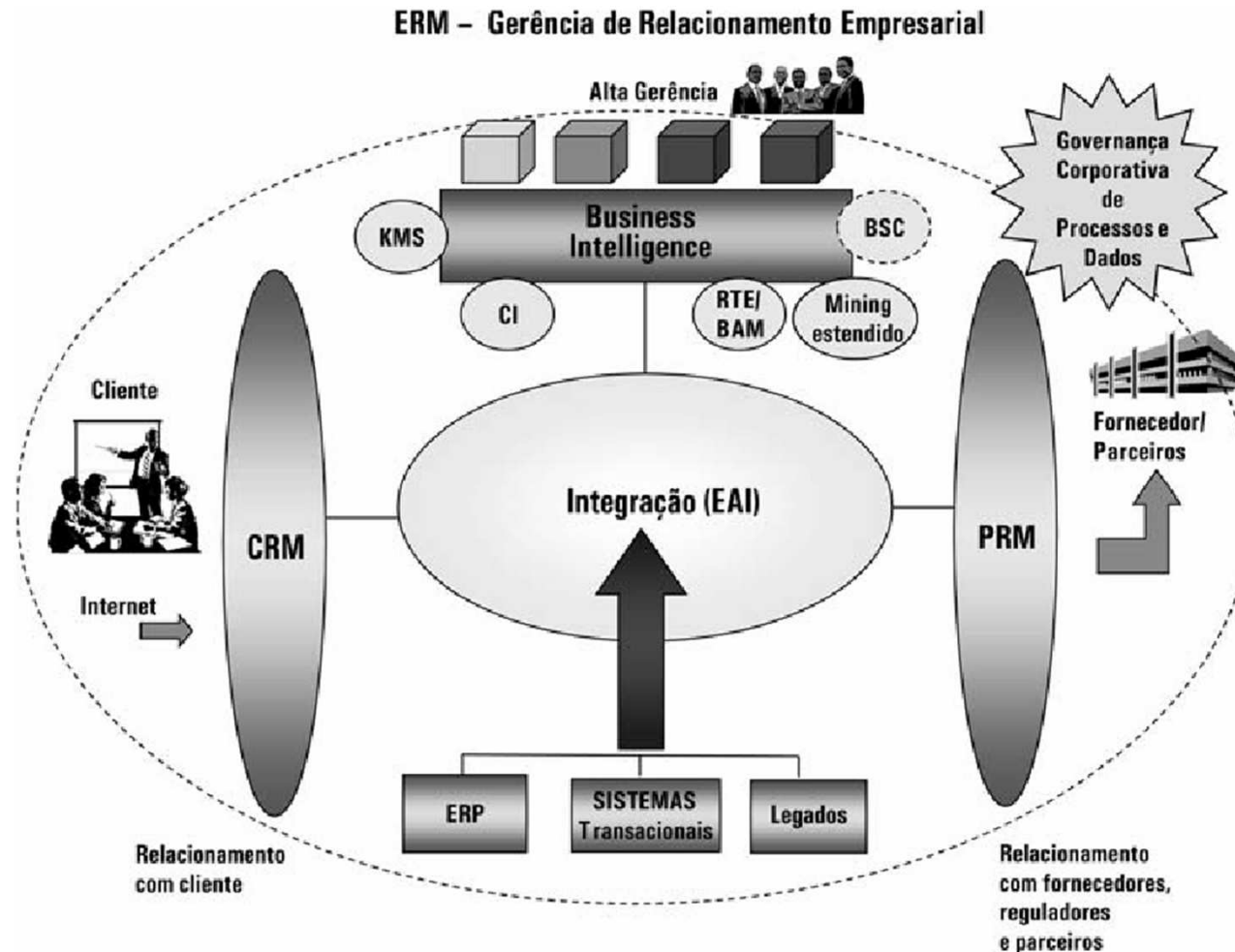


# Atividades de um processo de CI

- Coordenador de CI
- Analistas de informações, profissionais de pensamentos intuitivos
- Pesquisadores e adm. de rede de info-contatos com FN, centros de pesquisas, associações
- Técnicos de recuperação de informações – internet, BD, BI, técnicas de robot, text mining, web mining



# ERM – Enterprise Relationship Management

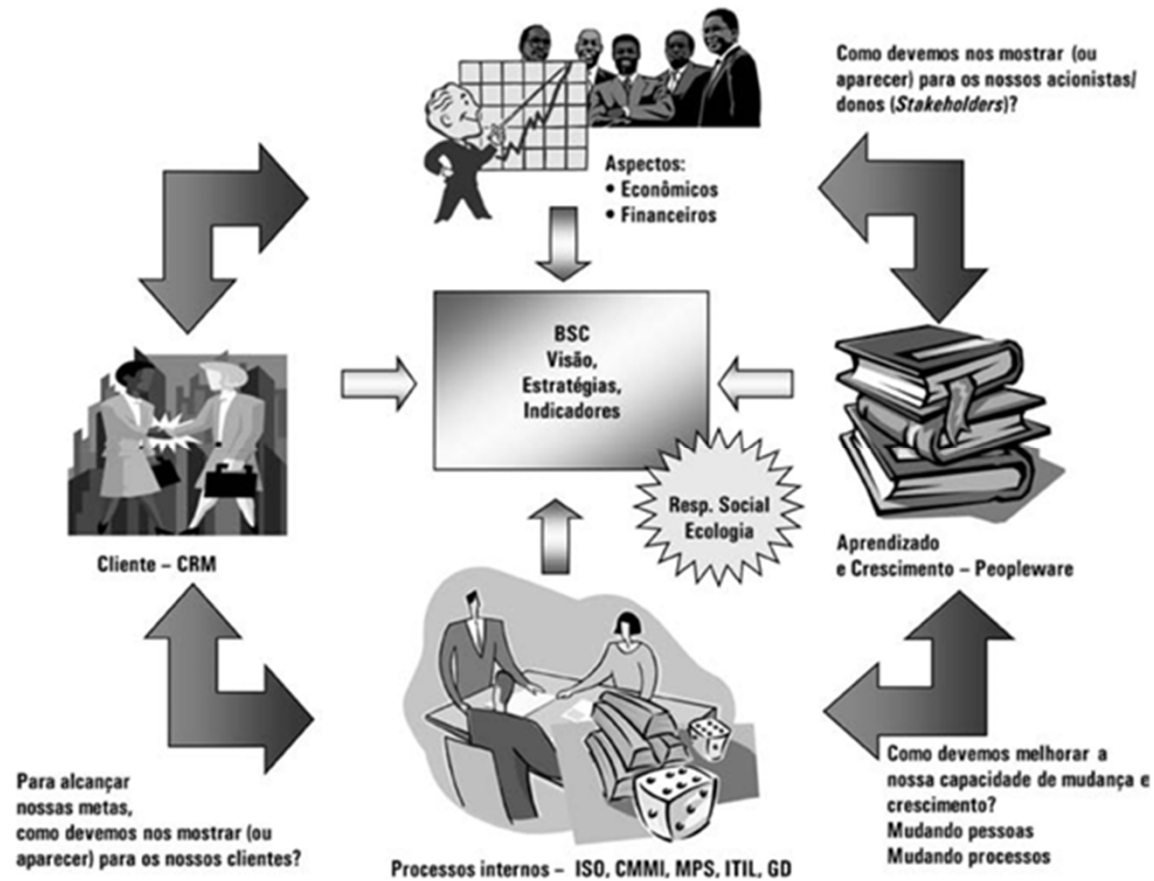


# Balanced Scorecard (BSC)

- Balanced Scorecard ou método de avaliação de desempenho empresarial foi criado por Robert Kaplan e David Norton nos anos 90.
- Uma vez definida a estratégia corporativa, o BSC a traduz em objetivos, indicadores, metas e planos de ação, constituindo a base de um processo de monitoramento e gerenciamento.



# BSC – Visão geral

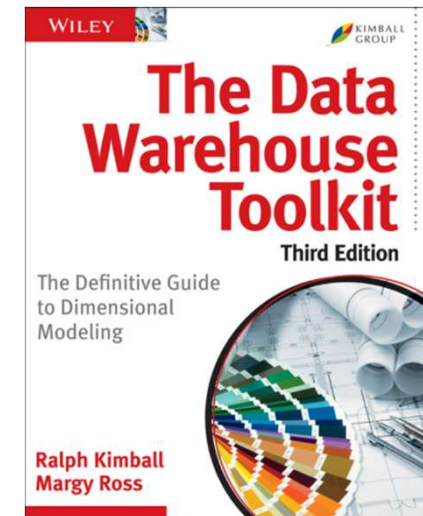
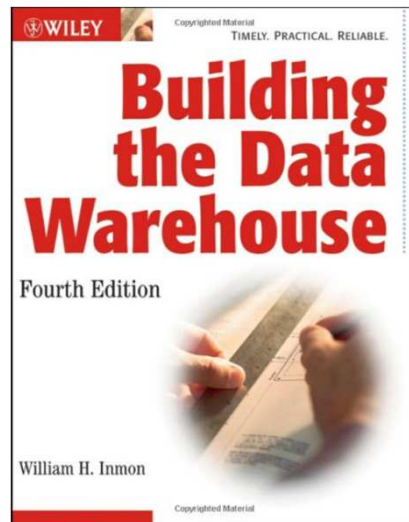




**ITnerante** 

**TIMASTERS** 

# Data Warehouse





# Conceito de Data Warehouse

---

- Inmon: É uma coleção de dados orientados por assunto, integrados, variáveis com o tempo e não voláteis, para dar suporte ao processo de tomada de decisão
- Kimball: É um conjunto de ferramentas e técnicas de projeto, que quando aplicadas às necessidades específicas dos usuários e aos bancos de dados específicos permitirá que planejem e construam um Data Warehouse
- Laudon&Laudon: Banco de dados, com ferramentas de consulta e relatório, que armazena dados atuais e históricos extraídos de vários sistemas operacionais e consolidados para fins de análises e relatórios administrativos

# Características de um DW

---

Orientação  
por Assunto

Integração

Variação no  
Tempo

Não  
Volatilidade

Granularidade

Credibilidade  
dos dados



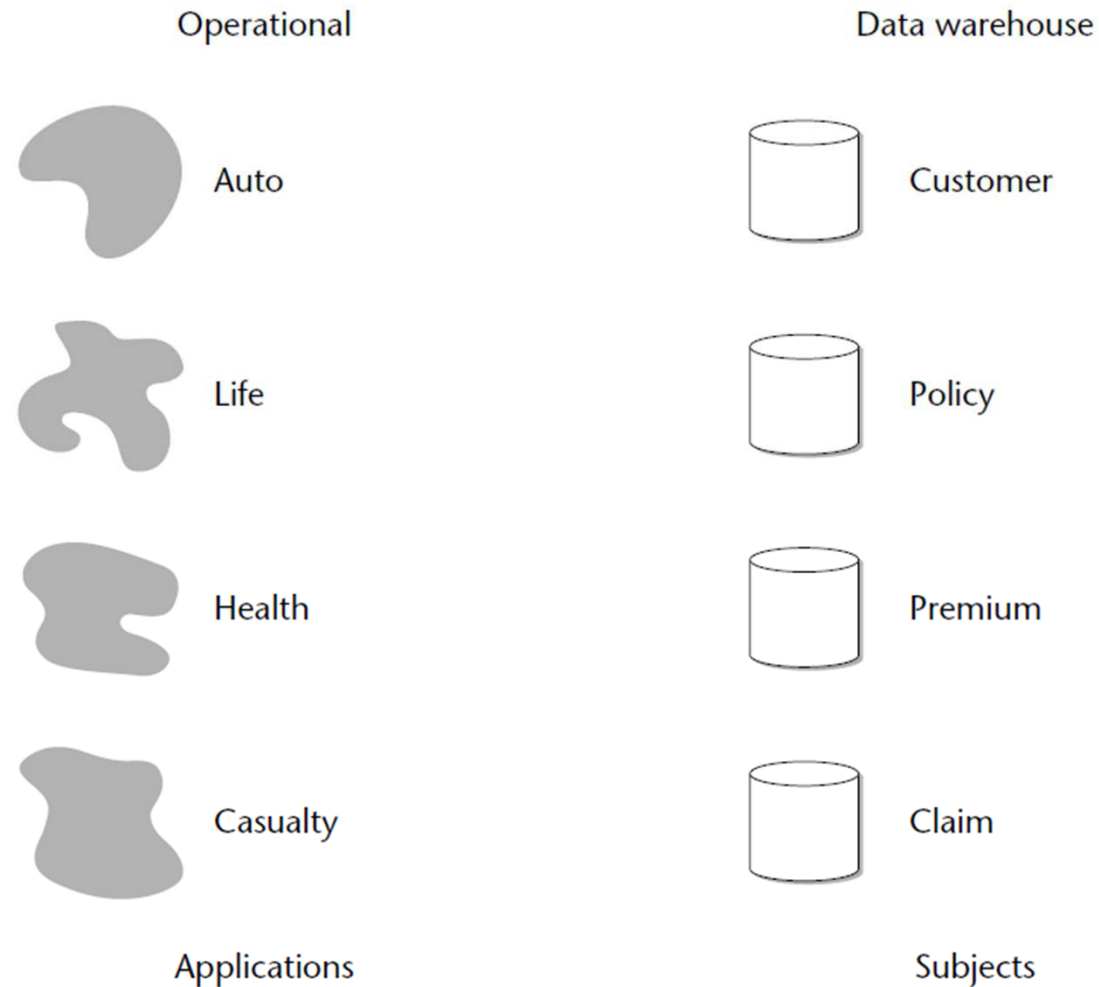
## Orientado por assunto

---

- Refere-se ao fato do Data Warehouse armazenar informações sobre **temas específicos** importantes para o **negócio** da empresa
  - Exemplos típicos de temas são: produtos, atividades, contas, clientes
- Em contrapartida, **o ambiente operacional** é organizado por **aplicações funcionais**
  - Por exemplo, em uma organização bancária, estas aplicações incluem empréstimos, investimentos e seguros

# Orientado por assunto

---

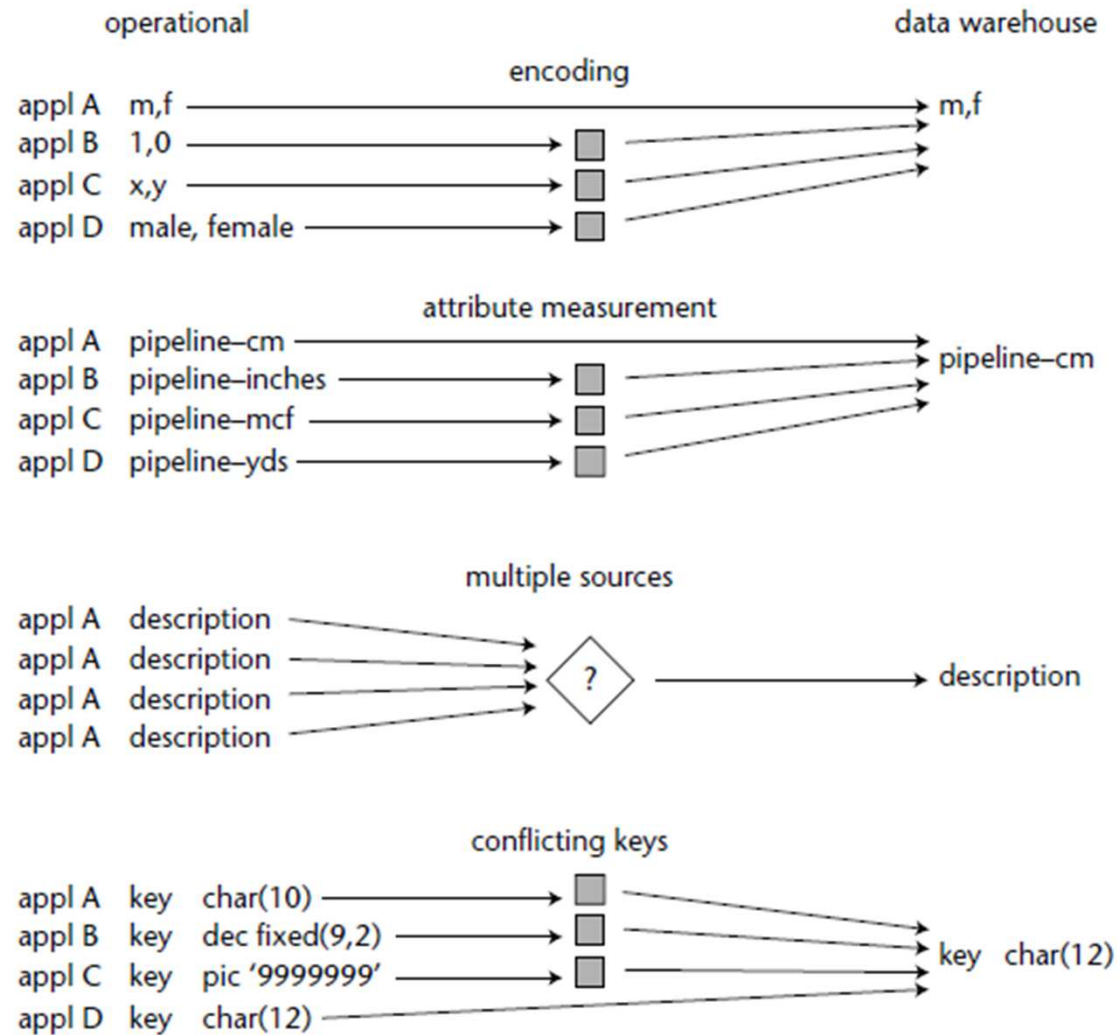


# Integrado

---

- Refere-se à consistência de nomes, das unidades, das variáveis, ... no sentido de que os dados foram transformados até **um estado uniforme**
  - Por exemplo, considere-se sexo como um elemento de dado. Uma aplicação pode codificar sexo como M/F, outra como 1/0 e uma terceira como H/M
- Conforme os dados são inseridos para o Data Warehouse, eles **são convertidos para um mesmo padrão**
  - Sexo é codificado apenas de uma forma
  - Da mesma maneira, se um elemento de dado é medido em centímetros em uma aplicação, em polegadas em outra, ele será convertido para **uma representação única** ao ser colocado no Data Warehouse

# Integrado



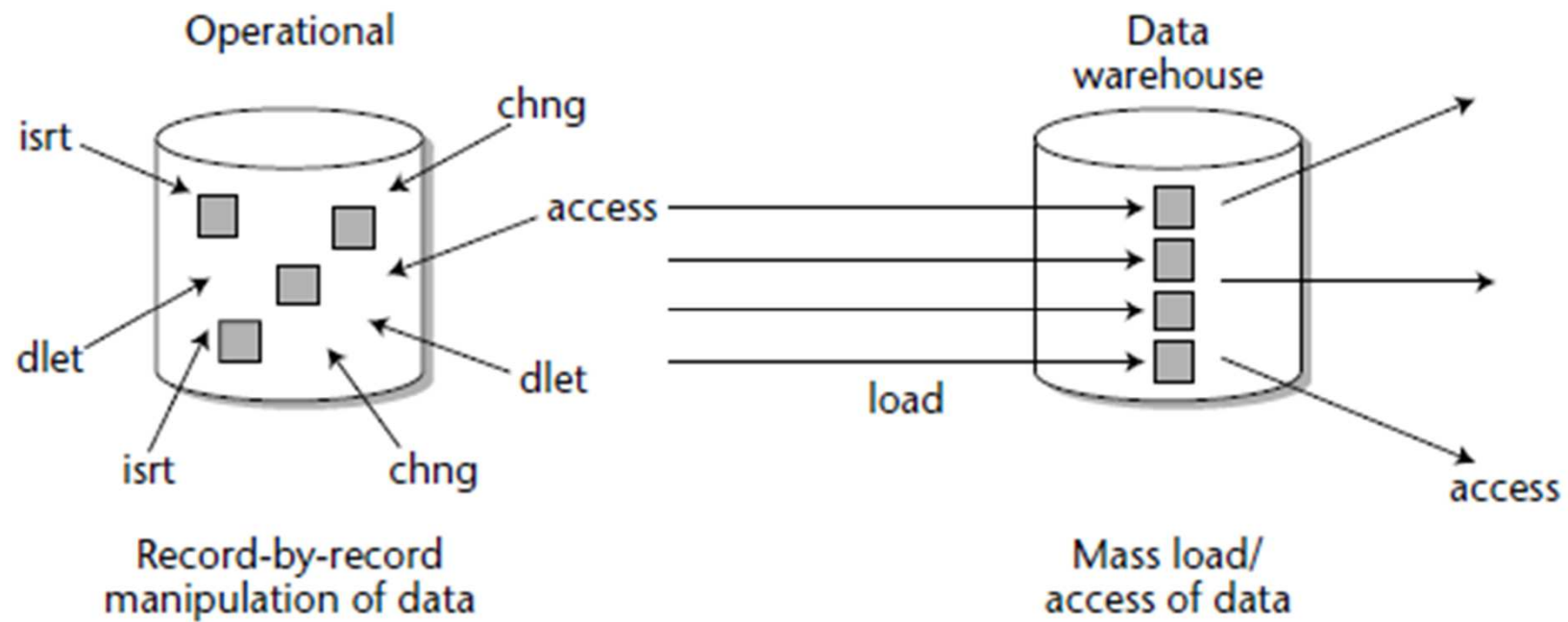
## Não volátil

---

- Significa que o Data Warehouse **permite apenas a carga inicial dos dados e consultas** a estes dados
- Após serem **integrados e transformados**, os dados são **carregados em bloco** para o Data Warehouse, para que estejam disponíveis aos usuários **para acesso**
- **No ambiente operacional**, ao contrário, os dados são, em geral, **atualizados registro a registro**, em múltiplas transações
  - Esta **volatilidade** requer um trabalho considerável para assegurar integridade e consistência através de atividades de rollback, recuperação de falhas, commits e bloqueios

# Não volátil

---



# Variante no tempo

---

- Refere-se ao fato **do dado** em um Data Warehouse referir-se a algum momento específico, significando que ele não é atualizável
- Enquanto que **o dado de produção é atualizado** de acordo com mudanças de estado do objeto em questão, refletindo, em geral, o estado do objeto no momento do acesso
- Em um Data Warehouse, a cada ocorrência de uma mudança, uma nova entrada é criada, para marcar esta mudança
  - O tratamento de **séries temporais** apresenta características específicas, que adicionam complexidade ao ambiente do Data Warehouse
  - Deve-se considerar que não apenas os dados têm uma característica temporal, mas também os metadados, que incluem definições dos itens de dados, rotinas de validação, algoritmos de derivação, etc.
  - Sem a manutenção do histórico dos metadados, as mudanças das regras de negócio que afetam os dados no Data Warehouse são perdidas, invalidando dados históricos

# Variante no tempo

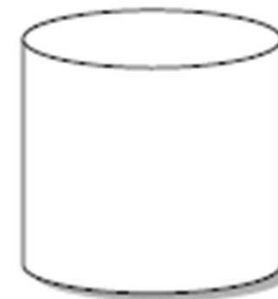
---

Operational



- Time horizon – current to 60–90 days
- Update of records
- Key structure may or may not contain an element of time

Data warehouse



- Time horizon – 5–10 years
- Sophisticated snapshots of data
- Key structure contains an element of time



# Granularidade

---

- “A granularidade de dados refere-se ao **nível de sumarização** dos elementos e de detalhe disponíveis nos dados, considerado **o mais importante aspecto do projeto** de um Data Warehouse.”
- Um nível de granularidade muito alto:
  - O **espaço em disco** e o **número de índices** necessários se tornam **bem menores**,
  - Há, porém, uma **diminuição** da **possibilidade de utilização dos dados** para atender a consultas detalhadas



# Requisitos básicos para um Data Warehousing

---

- DW: **Organizados em assuntos**
- DW: capacidade de **integração**
- DW: deve ser **flexível** o suficiente para **atender às exigências de mudança** rapidamente.
- Dados: considerados **não volátil** e deve ser **carregado em massa**.
- Dados: existem em **vários níveis de granularidade**.

## Questão 04 - ESAF - 2010 - MPOG - Analista de PO

### Tecnologia da Informação - Prova 2

---

Um ambiente *Datawarehouse*:

(a) é uma coleção de dados voláteis, invariantes em termos locais, integrados e orientados a um assunto, direcionados a decisões operacionais.

(b) ao ser projetado segundo a abordagem *bottom-up*, fundamenta-se nas necessidades de informação de todos os níveis gerenciais da empresa.

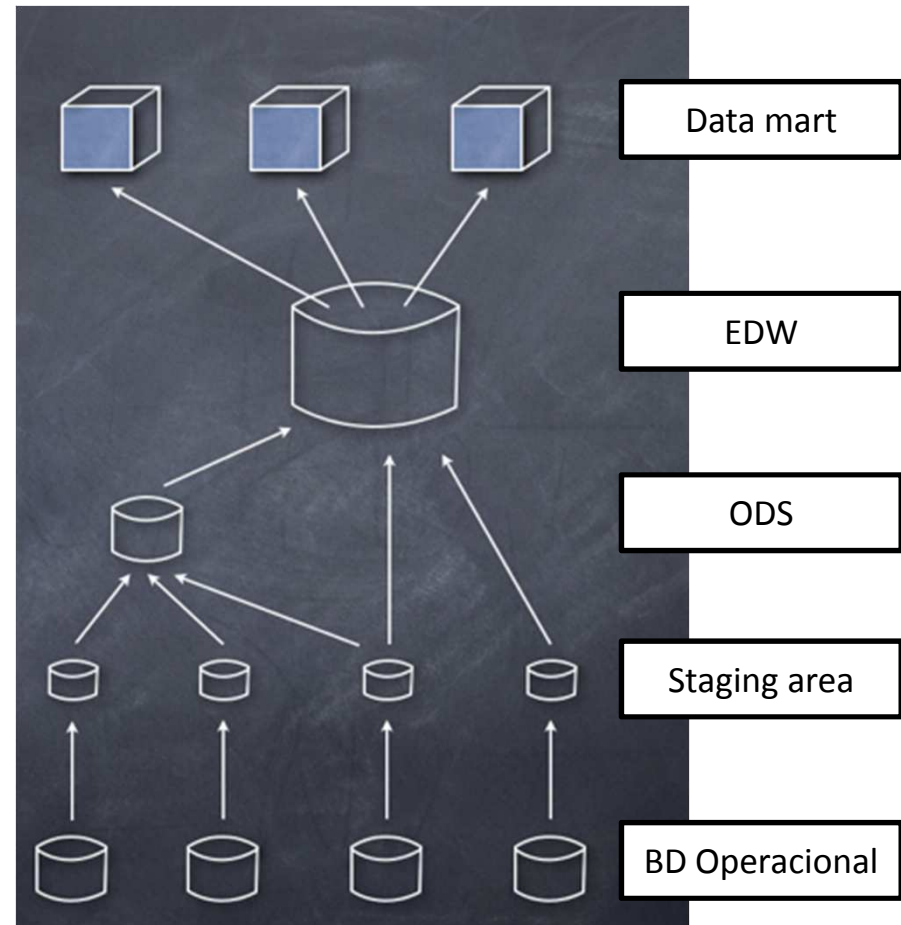
(c) para ser implementado, requer a remoção de metadados e a integração de dados existentes em um banco de dados.

(d) ao ser projetado, segundo a abordagem *top-down*, não comporta necessidades de informações gerenciais.

(e) é uma coleção de dados não-voláteis, invariantes em termos temporais, integrados e orientados a um assunto, utilizados no suporte a decisões gerenciais.

## Tipos de DW

- A indústria atualmente reconhece pelo menos três tipos diferentes de Data Warehouses:
  - Data Mart
  - Data Warehouse Empresarial (EDW)
  - Armazenamento de Dados Operacionais (ODS)



## Data Mart

---

- Tem um público mais focado e normalmente consiste em um subconjunto do EDW
  - O subconjunto pode ser definida pela **área geográfica** (por exemplo, apenas os dados da Alemanha), a **linha de produtos** (por exemplo, apenas produtos para o cabelo), ou **área funcional** (por exemplo, de produção)
- Dependente ou independente

Escopo reduzido

## Data Mart (Segundo date)

---

- “Um Data Mart é especializado e volátil”
  - **Especializado:** possui uma estrutura baseada em um ambiente, tema, situação, área, setor ou aplicação específica
    - Enquanto o EDW se baseia em várias fontes de diversas aplicações, fontes e situações para facilitar um suporte a decisão gerencial.
  - **Volátil:** dados são alterados frequentemente
    - Enquanto os do DW, por guardarem histórico, só são alterados quando uma carga foi feita de forma errada, mas não frequentemente como em um data mart (que é baseado em aplicações)

## Armazenamento de dados operacionais (ODS)

---

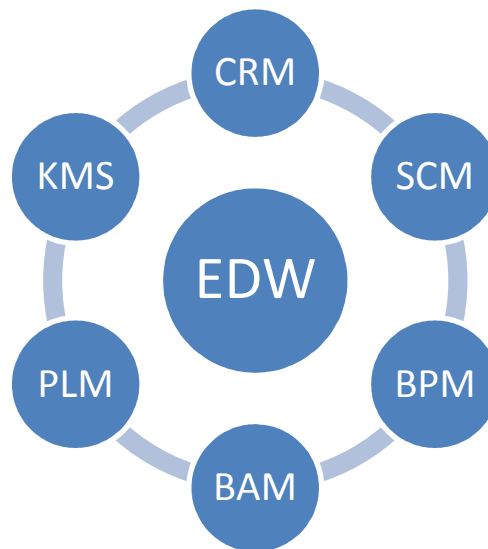
- O armazenamento de dados operacionais se concentra em um **prazo mais curto** para a análise e para o subconjunto é definido por tempo
  - Por exemplo, apenas os dados desta semana
- Um ODS consolida dados de vários sistemas de origem e fornece uma visão quase em **tempo real e volátil** dos dados atuais.

Prazo é reduzido e dados voláteis

# EDW – Enterprise Data Warehouse

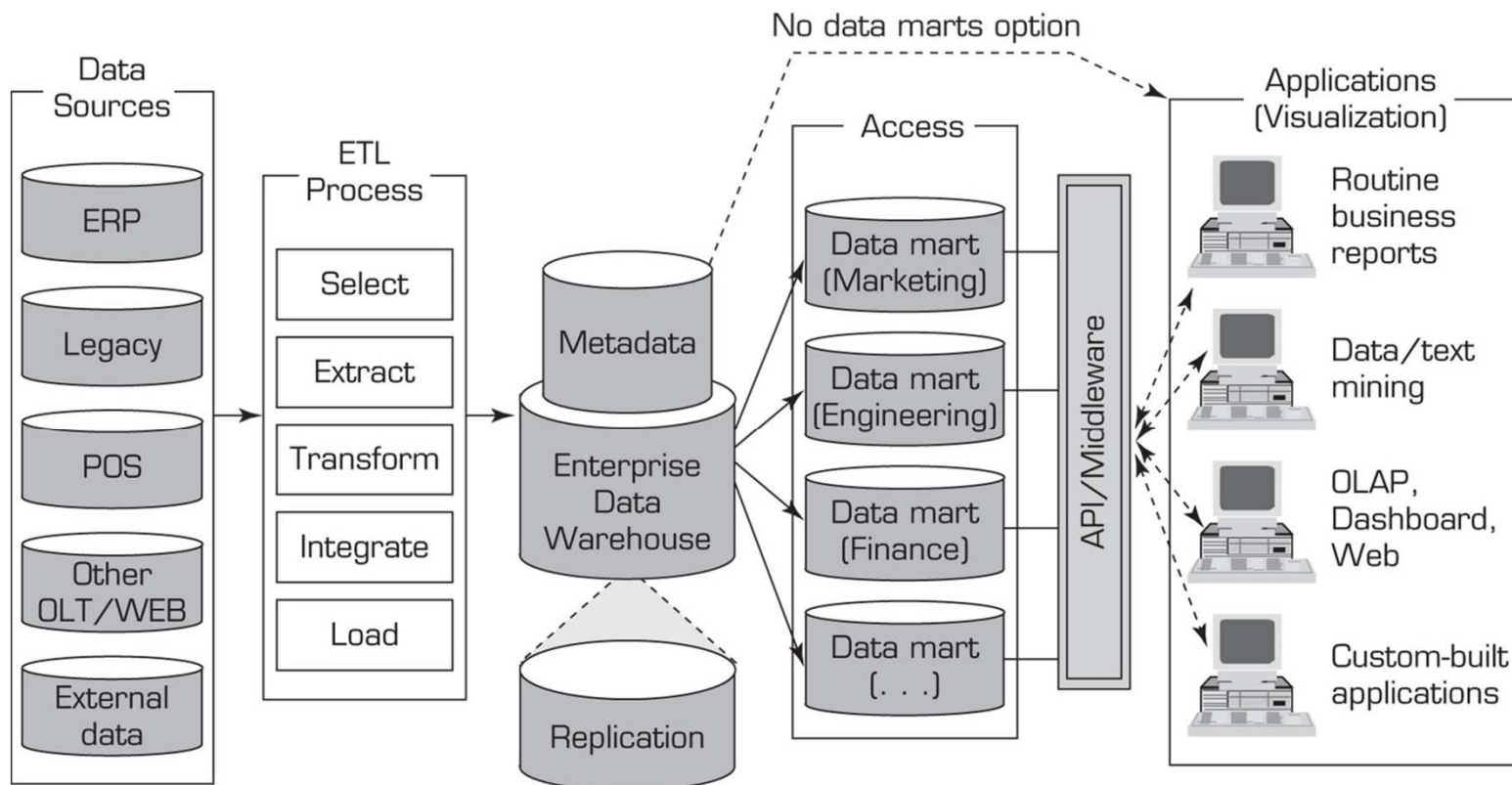
---

- É um DW de larga escala usado pela organização como um todo
- Congrega informações de diversas fontes de dados





# Processo de Data Warehousing



Copyright © 2011 Pearson Education, Inc. publishing as Prentice Hall

# Principais componentes

---

- As fontes de dados (Data sources)
- Extração e transformação de dados (Data extraction and transformation)
- Carga de dados (Data loading)
- Banco de dados abrangente (Comprehensive database)
- Metadados (Metadata)
- Ferramentas de middleware (Middleware tools)

## Questão 05 - ESAF - 2012 - Receita Federal - Analista Tributário da RF - Prova 2 - Área Informática

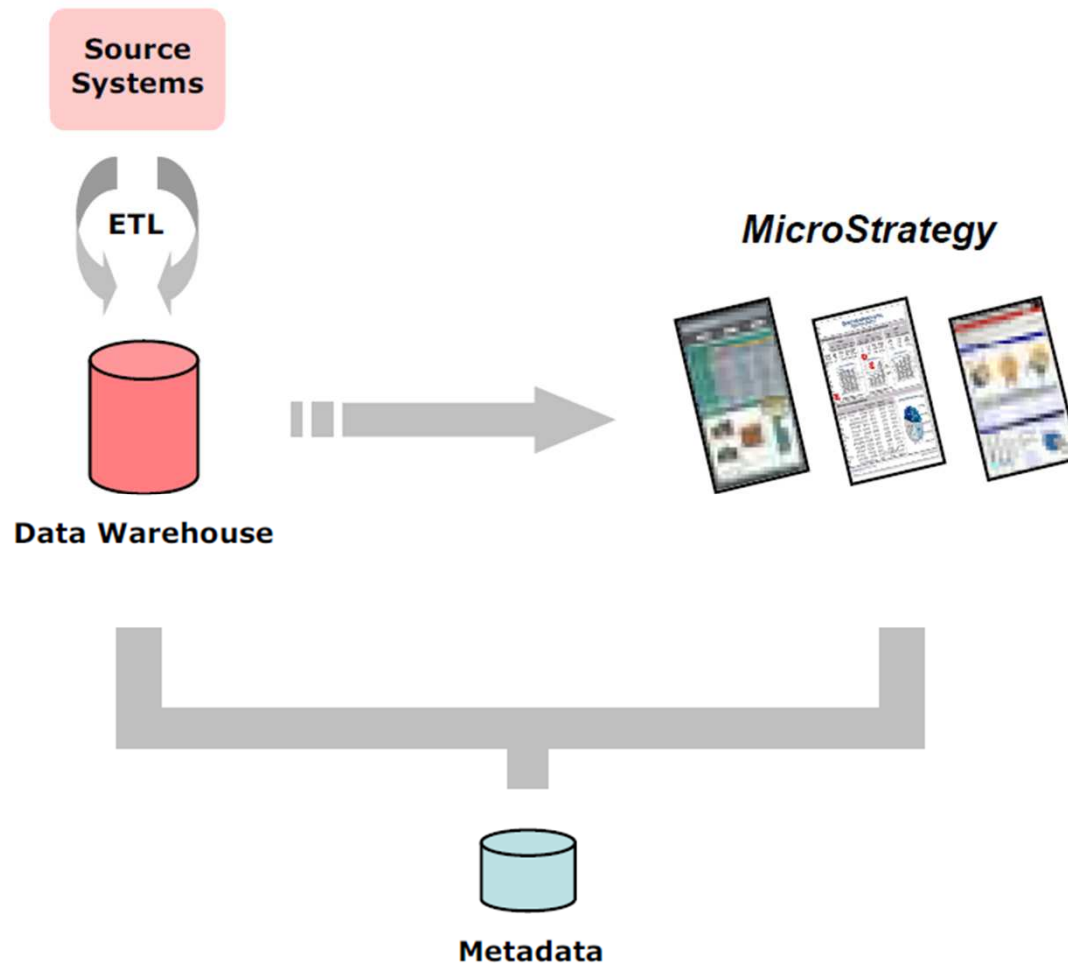
---

São componentes principais de um processo de data *warehousing*:

- (a) Fontes de dados. Extração de dados. Carregamento de dados. Banco de dados abrangente. Metadados. Ferramentas de *middleware*.
- (b) Fontes de usuários. Extração de relações. Carregamento de dados. Banco de dados focalizado. Metadados. Ferramentas de *upperware*.
- (c) Fontes de dados. Extração de dados. Monitoramento de dados. Banco de dados abrangente. Métodos de dados. Usuários *demiddleware*.
- (d) Fontes de dados homogêneos. Manipulação de dados. Carregamento de programas. Banco de dados equivalente. Polidados. Ferramentas de *middleware*.
- (e) Fontes de dados. Transformação de dados. Carregamento de sites de dados. Banco de dados abrangente. Meta-relações. Ferramentas de *middlehousing*.

# Componentes da arquitetura da MicroStrategy

---



# Arquitetura de DW - Turban

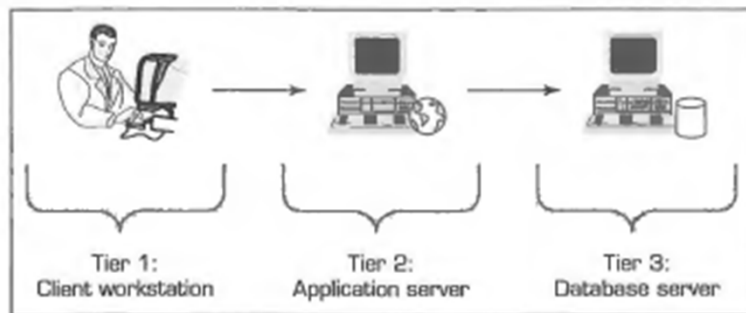


FIGURE 2.2 Architecture of a Three-Tier Data Warehouse.

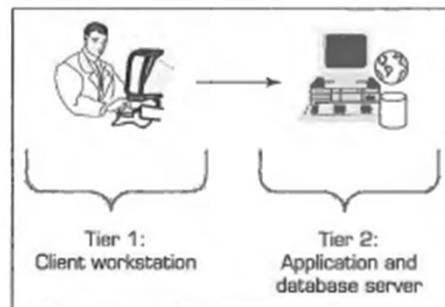
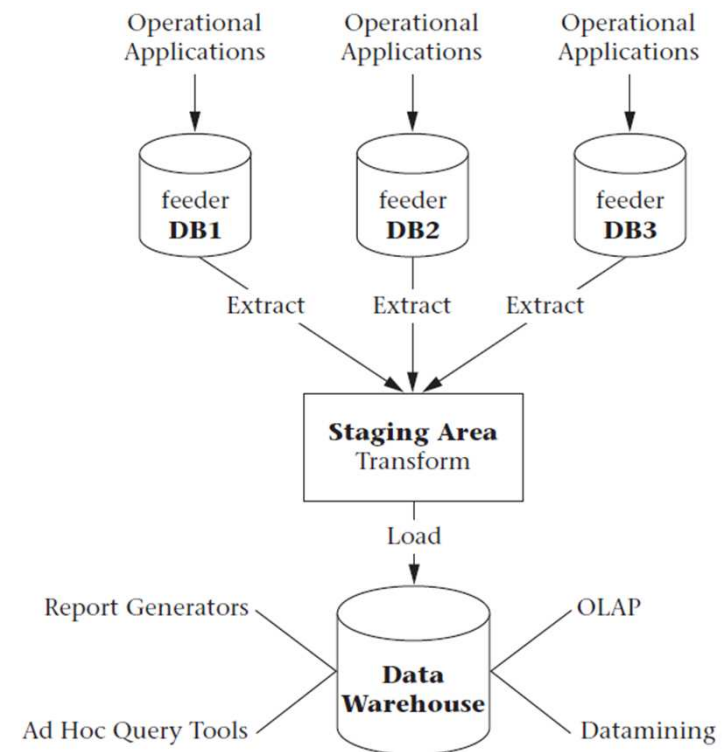


FIGURE 2.3 Architecture of a Two-Tier Data Warehouse.



Arquitetura básica de DW

## Questão 06 - ESAF - 2013 - MF - Analista de Finanças e Controle - Gestão em Infraestrutura de TI

---

A arquitetura de **data warehouse de três camadas** contém as seguintes camadas:

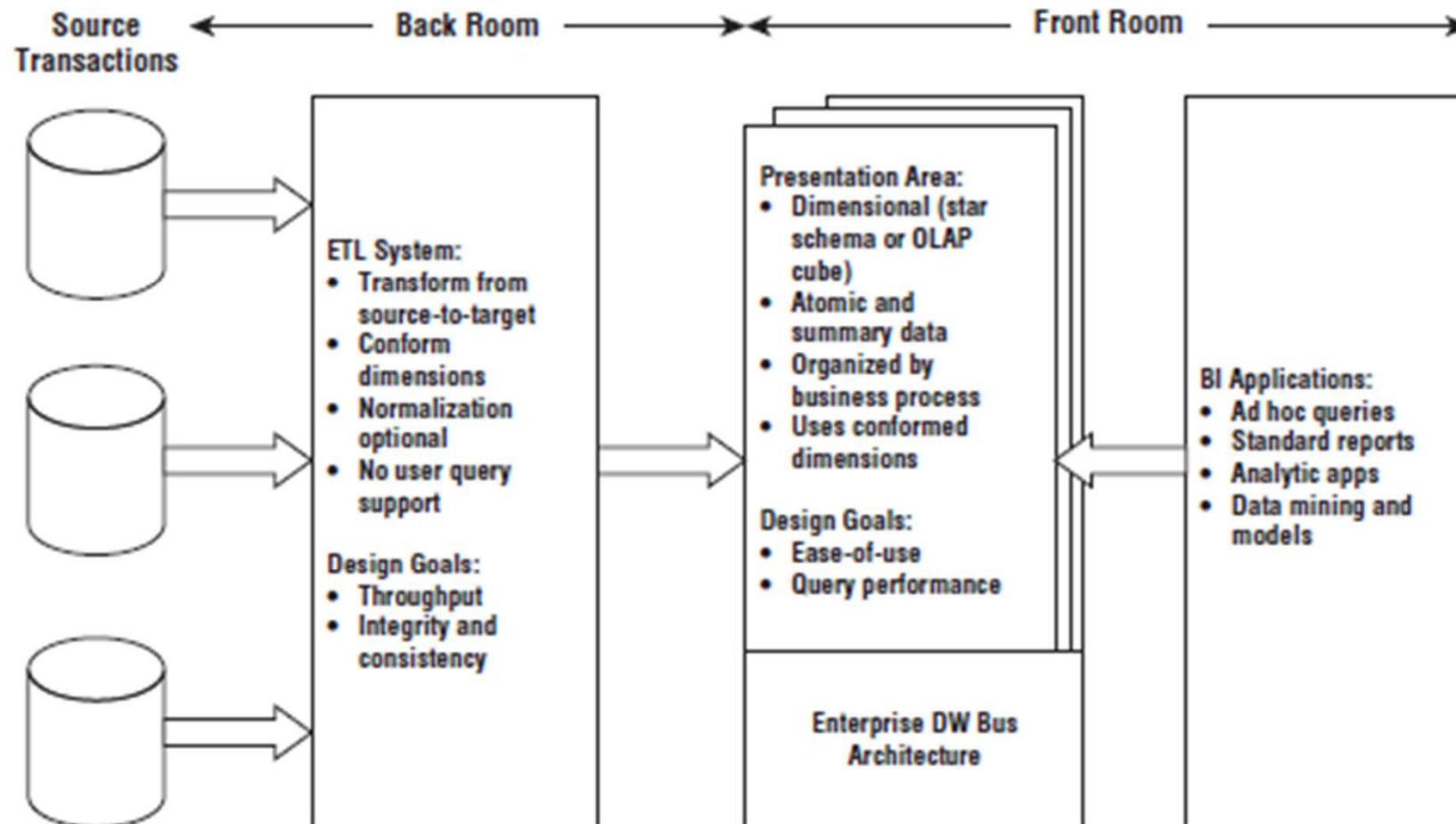
- (a) Primeira camada: sistema operacional com os dados e o software para aquisição.  
Segunda camada: especificação de dados.  
Terceira camada: servidor de monitoramento e cliente.
- (b) Primeira camada: sistema aplicativo.  
Segunda camada: data warehouse.  
Terceira camada: clientes preferenciais.
- (c) Primeira camada: sistema operacional com arquiteturas diferenciadas.  
Segunda camada: data center.  
Terceira camada: servidor de aplicação e cliente.
- (d) Primeira camada: sistema operacional com os dados e o software para aquisição.  
Segunda camada: data warehouse.  
Terceira camada: servidor de aplicação e cliente.
- (e) Primeira camada: sistema aplicativo com os dados e o software para aquisição.  
Segunda camada: data warehouse.  
Terceira camada: relatórios e telas de operação.

## Arquitetura de DW/BI do Kimball

---

- Há quatro **componentes** separados e distintos a serem considerados no ambiente de DW/BI:
  - **Sistemas operacionais de origem**
  - Sistema de ETL
  - Área de apresentação de dados
  - Aplicações de business intelligence

# Arquitetura de DW/BI do Kimball





# Fontes de dados operacionais

---

- Capturam as transações do negócio
- Estão do lado de fora do armazém de dados (DW)
  - Pouco ou nenhum controle sobre o conteúdo e o formato dos dados
- Prioridades: o desempenho e a disponibilidade
  - Consultas pequenas, registro a registro
  - Parte do fluxo de transação normal e severamente restringidos em suas demandas sobre o sistema operacional
  - Não são consultados nas formas amplas e inesperadas que os sistemas de DW/BI
- Mantém **poucos dados históricos**
  - Um bom DW pode aliviar os sistemas de origem de grande parte da responsabilidade de representar o passado.
- São aplicações de **propósito específico**, sem qualquer compromisso com o compartilhamento de dados comuns
  - Ex: produto, cliente, geografia ou calendário com outros sistemas operacionais da organização
  - Um sistema amplamente adotado entre aplicativos **Enterprise Resource Planning (ERP)** ou **sistema de gerenciamento de dados mestre** operacional poderiam ajudar a resolver essas deficiências

# Sistemas de ETL

---

- Sistema ETL do ambiente de DW / BI consiste
  - De uma área de trabalho
  - Da instanciação das estruturas de dados
  - De um conjunto de processos
- O sistema de ETL é **tudo que existe** entre os sistemas operacionais de origem e área de apresentação do DW/BI
- Extração
  - É o **primeiro passo** no processo de obtenção de dados para o ambiente de DW
  - Significa **ler e entender os dados de origem e copiar os dados necessários** para o sistema de ETL para posterior manipulação
  - Neste ponto , os dados pertence ao DW

# Sistemas de ETL

---

- Transformação
  - Limpeza dos dados
    - Corrige erros de ortografia
    - Resolve de conflitos de domínio
    - Lida com elementos faltante ou análise em formatos padrão
    - Combina dados de várias fontes
    - Trabalha a de-duplicação de dados
  - O sistema de ETL **agrega valor aos dados** com esta limpeza e conformidade as tarefas, **alterando os dados** e melhorando a qualidade
  - Além disso, essas atividades podem ser desenvolvidas para criar **metadados de diagnóstico** , acabam levando a **reengenharia de processos** visando melhorar a **qualidade dos dados** nos sistemas de origem ao longo do tempo

# Sistemas de ETL

---

- Carga
  - A **etapa final** do processo de ETL: a estruturação física e carregamento de dados em modelos multidimensionais
  - A principal missão do sistema de ETL é entregar as tabelas de dimensão e de fatos na etapa de entrega, esses subsistemas são críticos
  - Quando as tabelas dimensão e o fato em um modelo dimensional foram **atualizadas, indexadas**, fornecidas com agregados adequados, e ainda **qualidade assegurada**, a comunidade empresarial é notificada de que os novos dados foram publicados

## Área de apresentação de dados

---

- Dados do sistema DW/BI deve ser
  - Dimensionais
  - Atômicos (complementado por melhoram o desempenho de agregados)
  - Centrados no processos de negócios
  - Aderentes à arquitetura de barramento Enterprise Data Warehouse (EDW)

## Aplicações de business intelligence

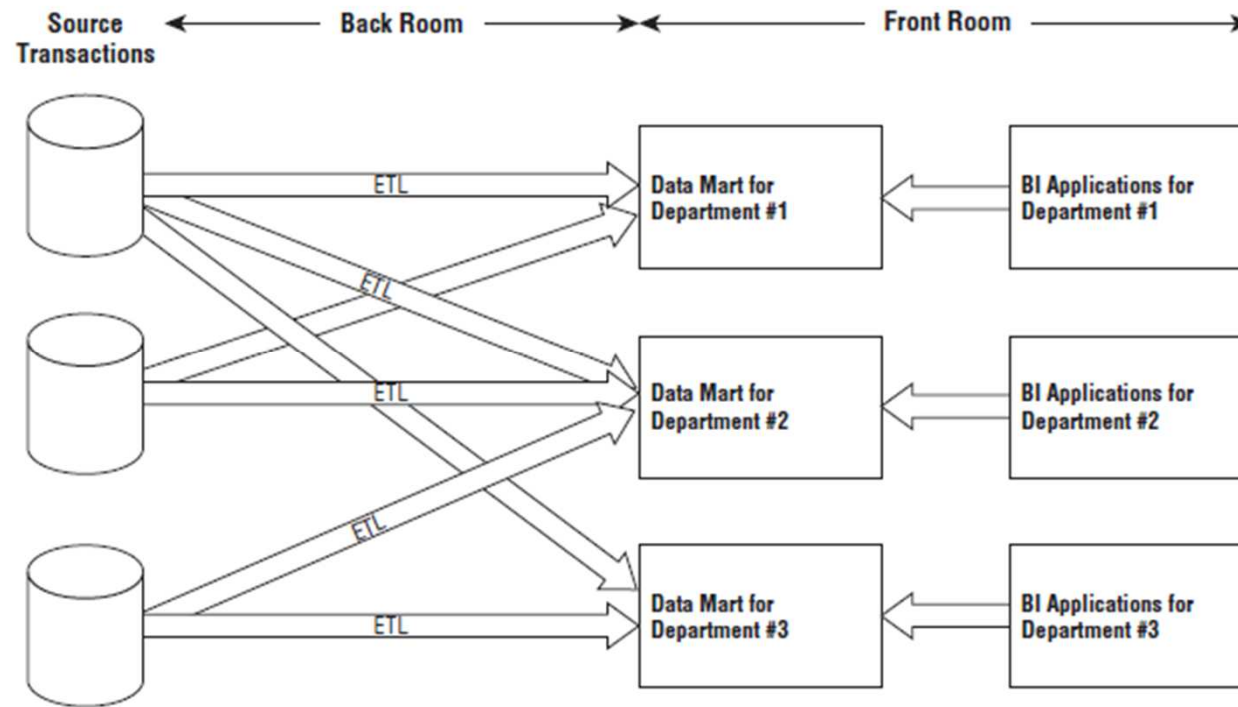
---

- O **último componente** principal da arquitetura de DW/BI do Kimball
- Refere-se à **gama de recursos** fornecidos para **usuários de negócios** para alavancar a área de apresentação para a tomada de decisão analítica
- Por definição, todas as aplicações de BI que consultam os dados estão na área de apresentação do DW/BI

# Arquiteturas alternativas de DW/BI

---

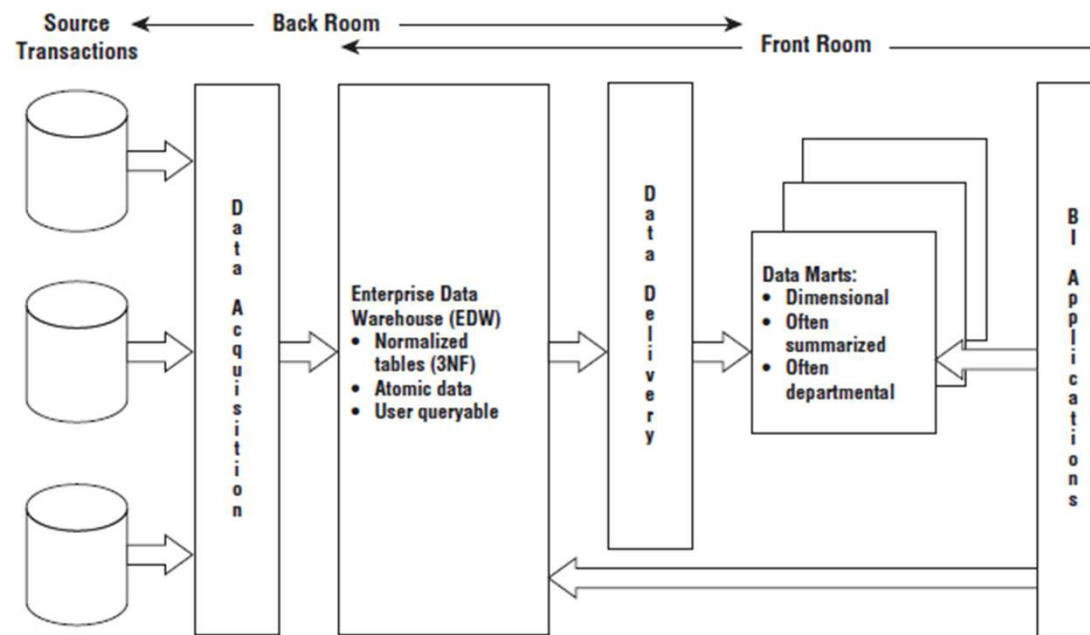
- Independent Data Mart Architecture



# Arquiteturas alternativas de DW/BI

---

- Hub-and-Spoke Corporate Information Factory  
Inmon Architecture

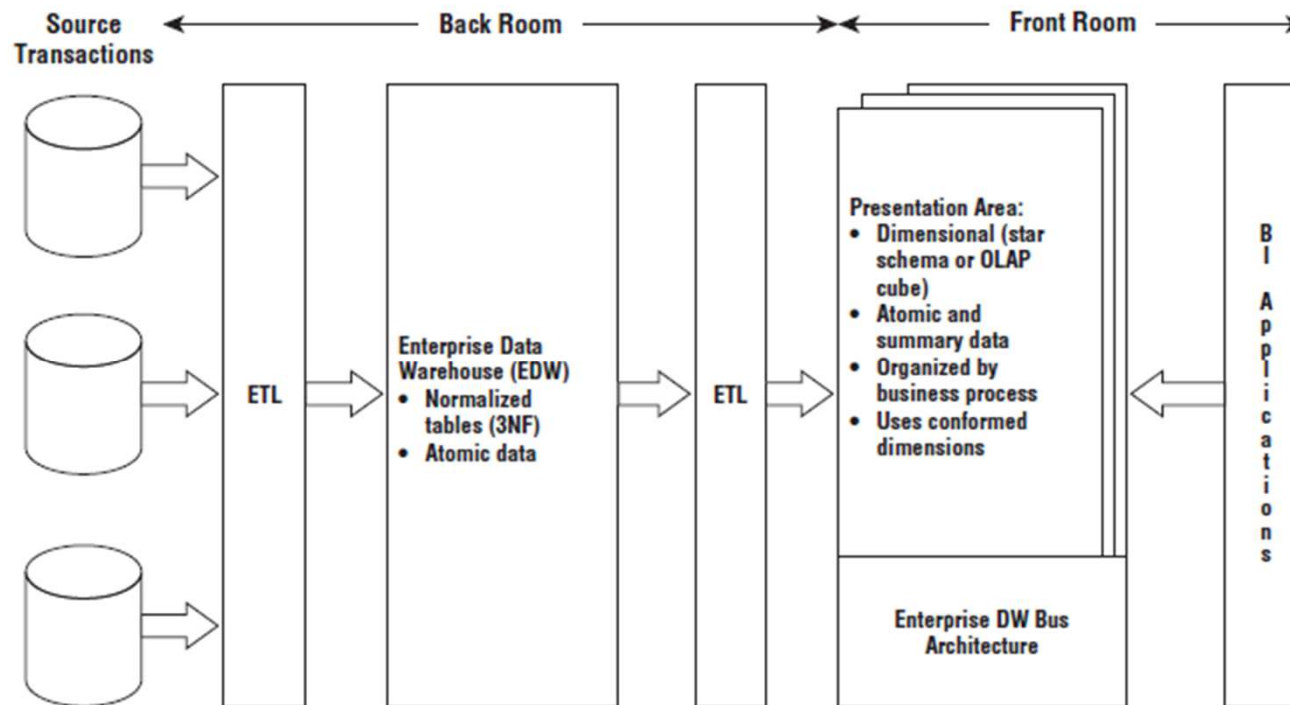




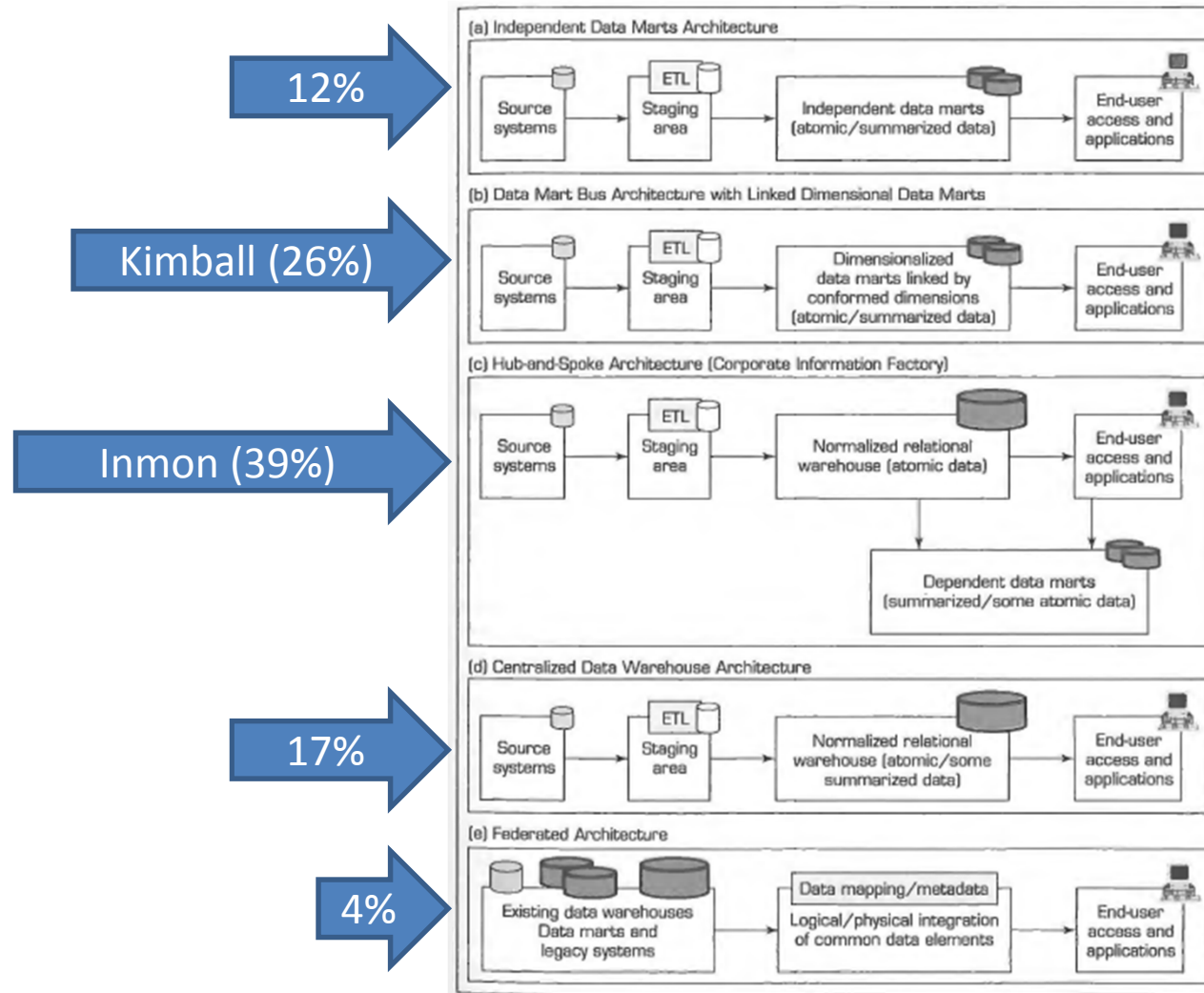
# Arquiteturas alternativas de DW/BI

---

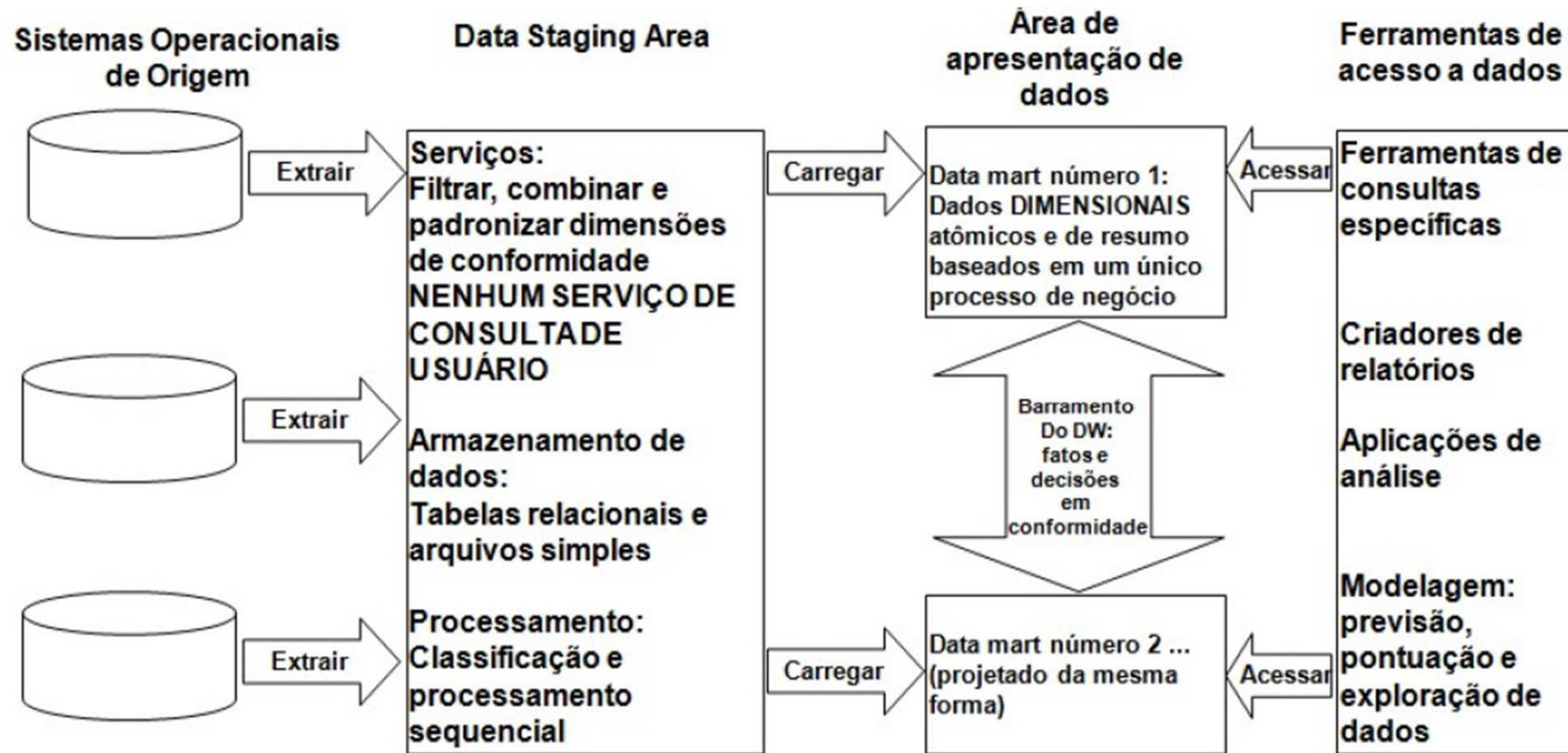
- Hybrid Hub-and-Spoke and Kimball Architecture



# Alternativas de Arquiteturas



# Elementos básicos do DW



## Questão 07 - CESGRANRIO - 2010 - ELETROBRÁS - Analista de Sistemas - FUNCIONAL SAP-ERP

---

Numa grande multinacional, os dados relativos aos sistemas transacionais são lidos, transformados e carregados para posterior apresentação sob a forma dimensional. O departamento de TI reservou uma área de armazenamento específica às informações lidas e transformadas, denominada *staging area*. A respeito dessa área de dados, afirma-se que

- (a) a *staging area* armazena os dados exatamente como estão representados no sistema transacional.
- (b) as tabelas da *staging area* não devem sofrer modificações estruturais ao serem carregadas para a presentation area.
- (c) os dados aí contidos estão prontos para serem inquiridos para geração de consultas e relatórios gerenciais.
- (d) os dados, na *staging area*, podem ser armazenados tanto de forma normalizada como na forma de arquivos flat.
- (e) cada tabela existente na *staging area* equivale a uma tabela normalizada na presentation area.

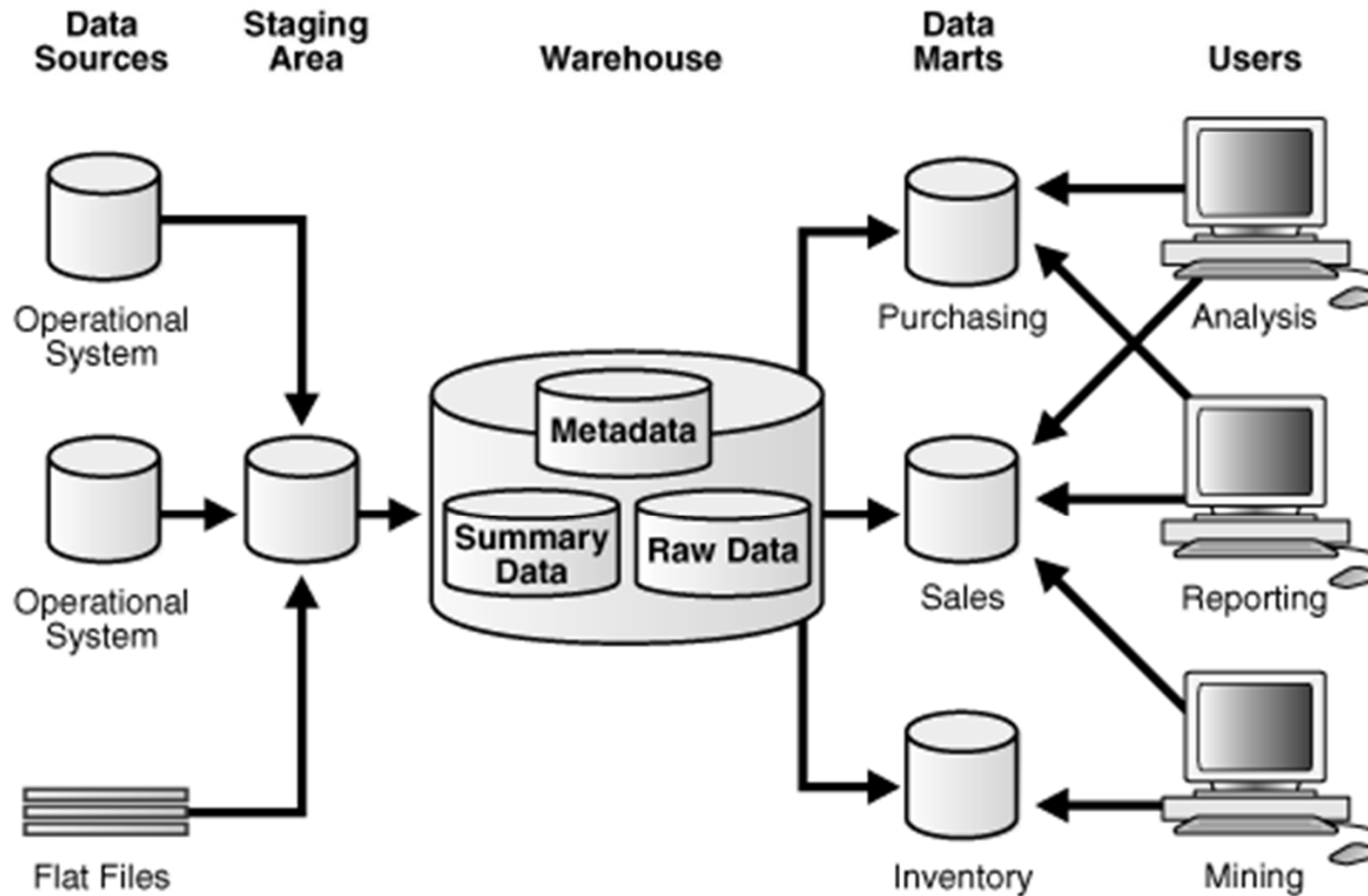
## Operational Data Storage (ODS)

---

- Ou **Staging Area (SA)** representa um armazenamento intermediário dos dados, promovendo a integração dos dados do ambiente operativo antes de sua atualização no DW
- Inicialmente, um ODS era considerado um repositório temporário que armazenava apenas informações correntes antes de serem carregadas para o DW, similar a uma cópia dos ambientes de sistemas transacionais em uma empresa
- Atualmente, *alguns autores passaram a denominá-lo* **Dynamic Data Storage (DDS)**

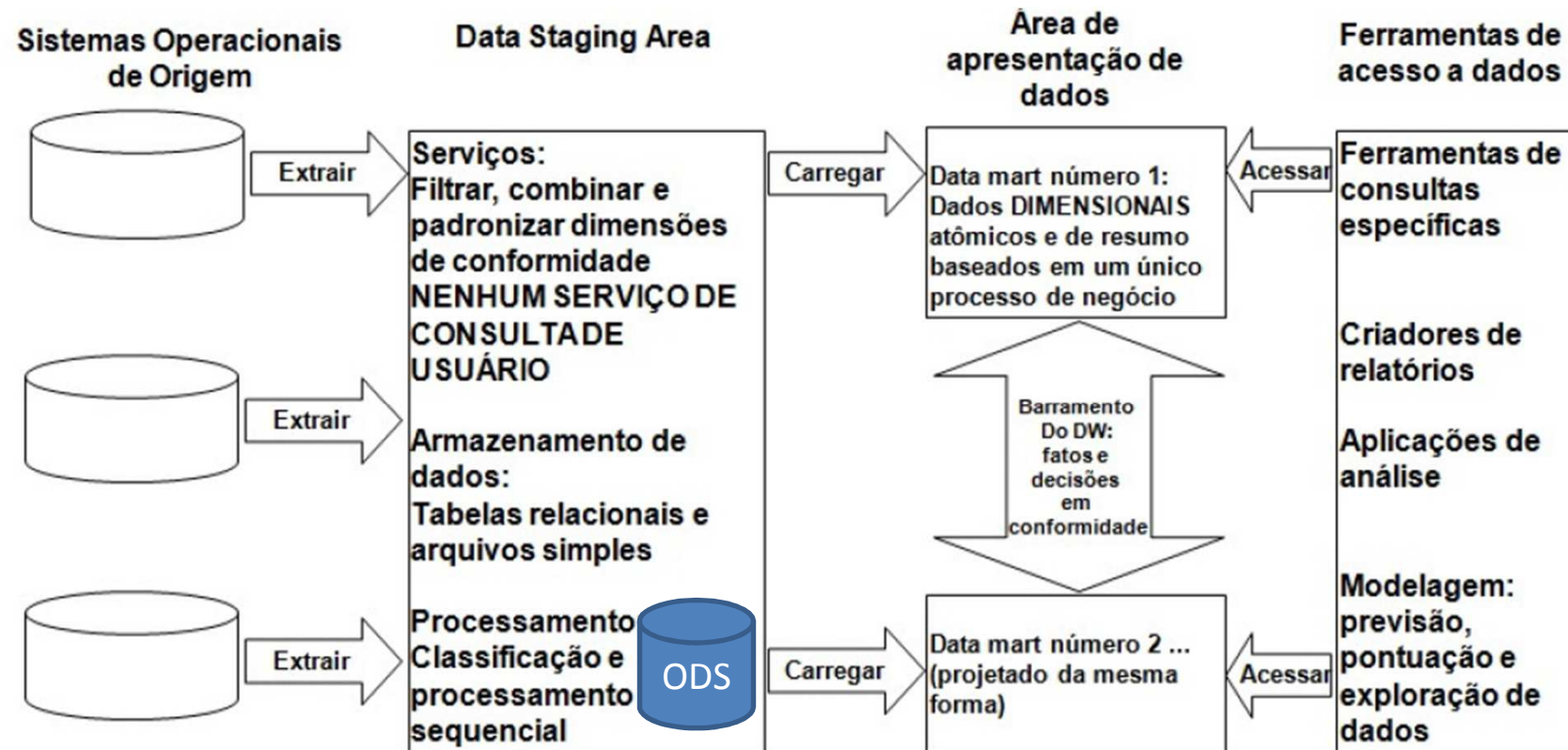
# Staging Area

Pos. 1



# Operational Data Storage (ODS)

Pos. 2





## Data Staging Area (Kimball)

---

- É tanto uma área de armazenamento como um conjunto de processos, e normalmente denomina-se ETL (*Extract – Transformation - Load*)
- O principal requisito de arquitetura é que ela não esteja acessível aos usuários e que não forneça serviços de consulta nem de apresentação



## Questão 08 - TRT - 11ª Região (AM) - Técnico Judiciário Tecnologia da Informação - 2012

---

No âmbito dos DWs, representa um armazenamento intermediário que facilita a integração dos dados de ambiente operativo antes da sua atualização no DW. Trata-se de

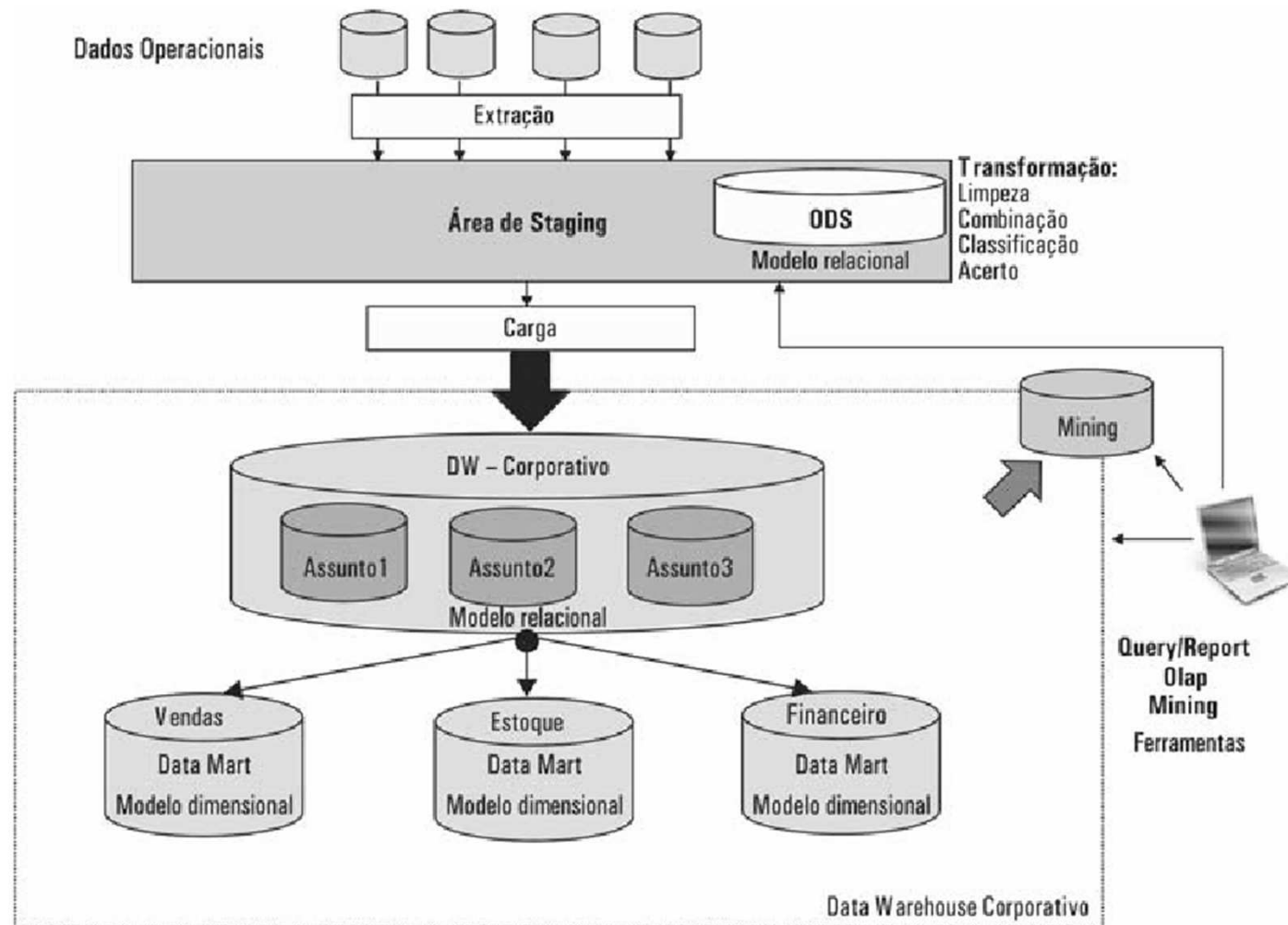
- (a) ODS.
- (b) ETL.
- (c) *Data Mart*.
- (d) *Star Schema*.
- (e) *Fact Table*.

# Polêmica entre Kimball e Inmon

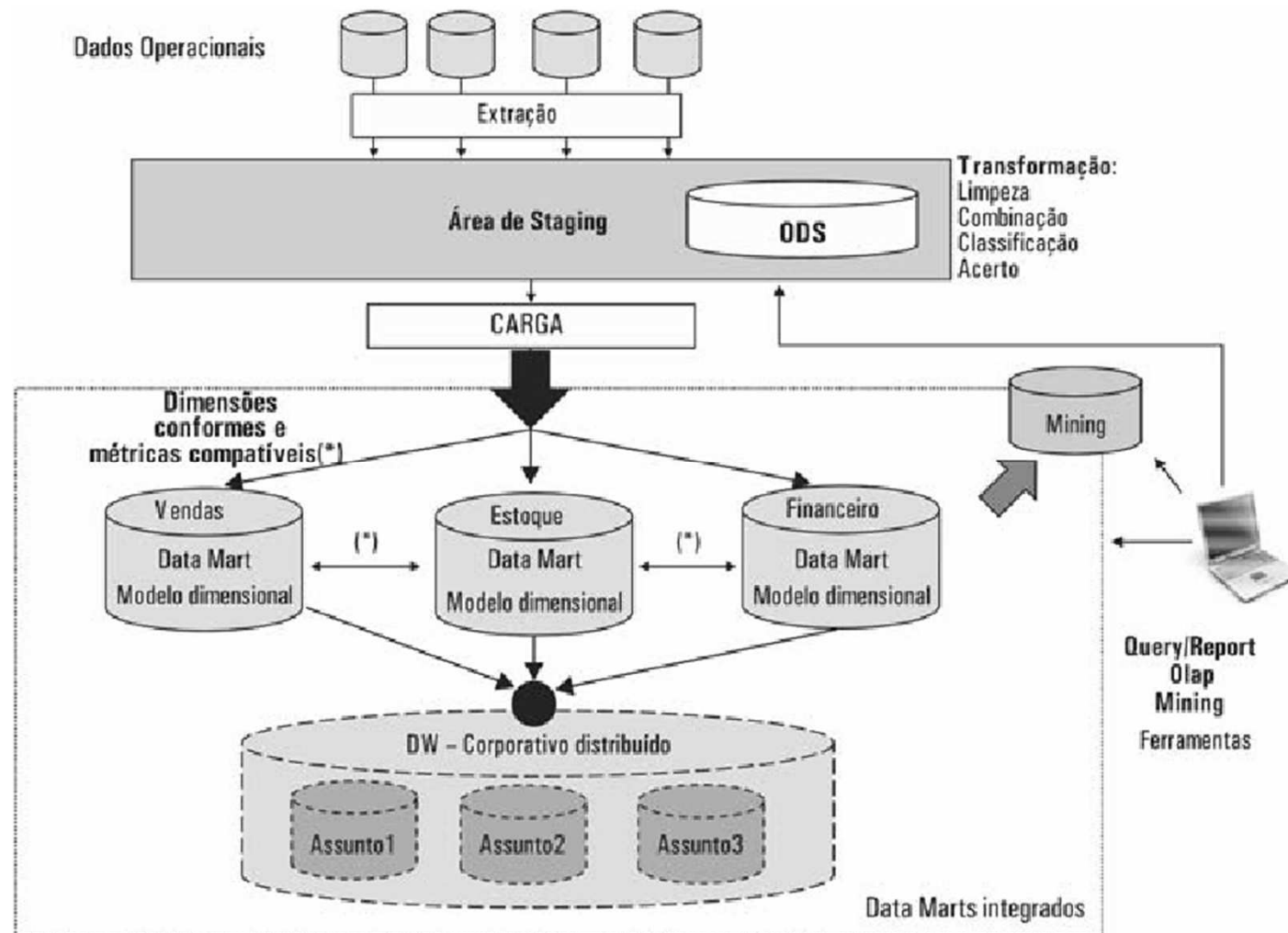
---

- Ralph Kimball
  - Pai do conceito de star schema
  - **BOTTOM-UP**
  - **Estilo mais simples e incremental**
- Bill Inmon
  - Pai do conceito de DW
  - **TOP-DOWN**
  - **A ênfase sempre foi um grande depósito central de informações**

# Inmon



# Kimball



# Diferenças essenciais entre as abordagens

Characteristic	Inmon	Kimball
<i>Methodology and Architecture</i>		
Overall approach	Top-down	Bottom-up
Architecture structure	Enterprise-wide (atomic) data warehouse "feeds" departmental databases	Data marts model a single business process, and enterprise consistency is achieved through a data bus and conformed dimensions
Complexity of the method	Quite complex	Fairly simple
Comparison with established development methodologies	Derived from the spiral methodology	Four-step process; a departure from RDBMS methods
Discussion of physical design	Fairly thorough	Fairly light
<i>Data Modeling</i>		
Data orientation	Subject or data driven	Process oriented
Tools	Traditional (ERD, data flow diagrams)	Dimensional modeling; a departure from relational modeling
End-user accessibility	Low	High
<i>Philosophy</i>		
Primary audience	IT professionals	End users
Place in the organization	Integral part of the corporate information factory	Transformer and retainer of operational data
Objective	Deliver a sound technical solution based on proven database methods and technologies	Deliver a solution that makes it easy for end users to directly query the data and still get reasonable response times

Sources: Based on M. Breslin, "Data Warehousing Battle of the Giants: Comparing the Basics of Kimball and Inmon Models," *Business Intelligence Journal*, Vol. 9, No. 1, Winter 2004, pp. 6-20; and T. Ariyachandra and H. Watson, "Which Data Warehouse Architecture Is Most Successful?" *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter 2006.

## Questão 09 - CESPE - 2009 - TCU - Analista de Controle Externo - Tecnologia da Informação

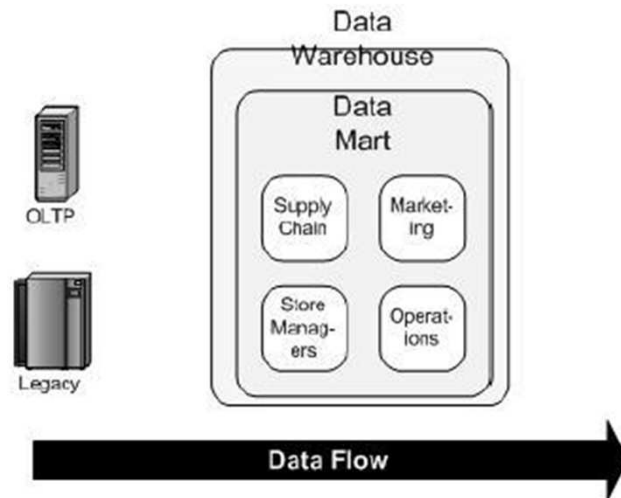


Figura I

Com base nas figuras I e II ao lado, e nos conceitos de datawarehousing, julgue os itens subsequentes.

[135] Na figura II, que corresponde à visão de Kimball, o datawarehousing é visto como uma constituição de data marts, cujo foco é entregar objetivos do negócio para departamentos na organização.

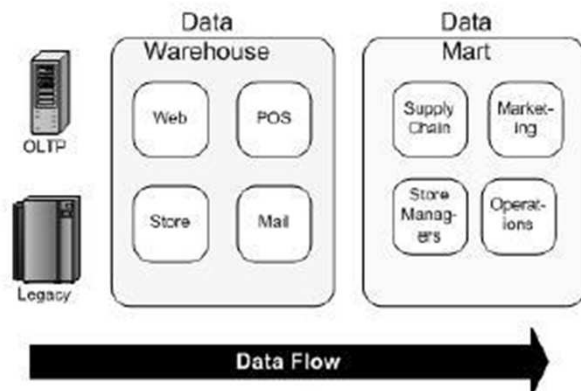
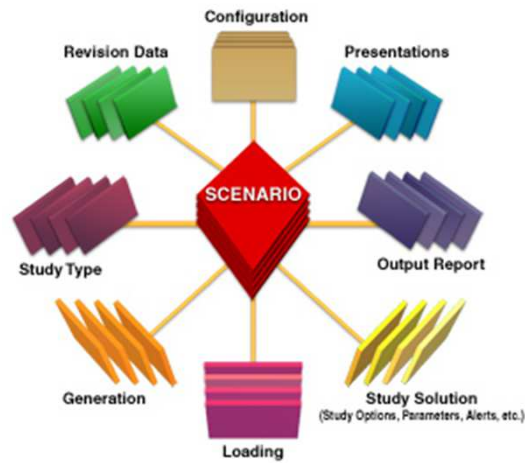


Figura II

[136] Na figura I, o datawarehouse é criado com base em um enfoque subject-by-subject. Dessa forma, o desenvolvimento de um datawarehouse poderia começar com dados de uma loja online, por exemplo, e outros subjects seriam adicionados ao datawarehouse conforme as necessidades. Nesse enfoque, o data mart é a criação de uma subject area do datawarehouse.



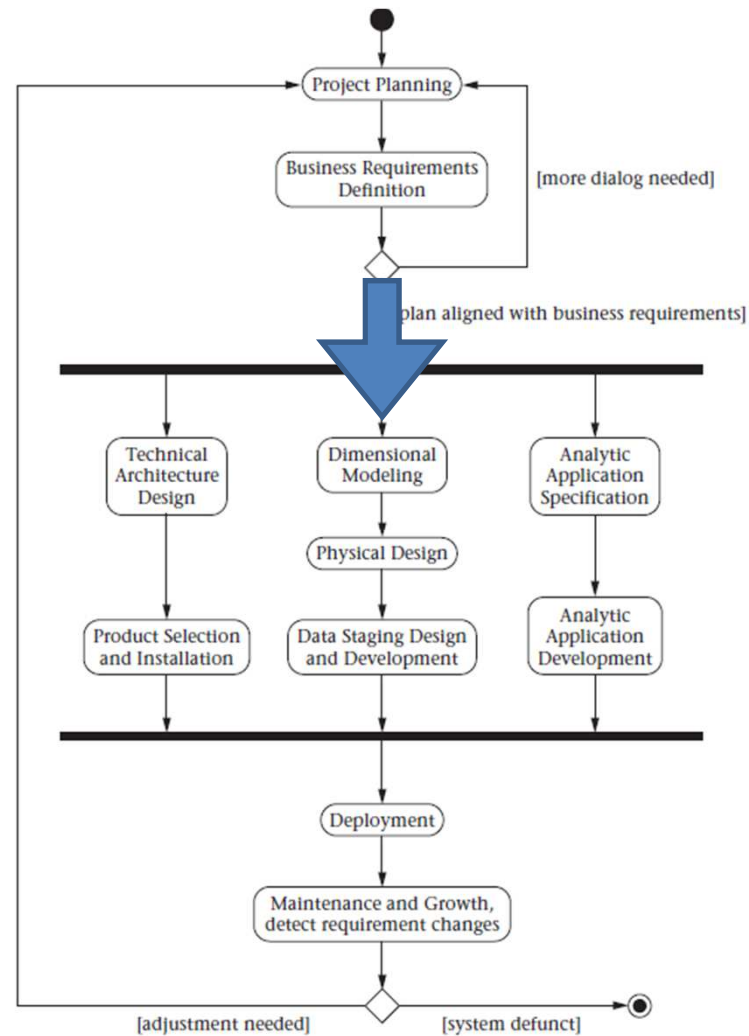
# Modelagem multidimensional

---

Ambos os modelos 3NF e multidimensionais podem ser representados em diagramas ER pois são constituídos por tabelas relacionais

A principal diferença entre **os modelos** na 3NF e o multidimensional é o **grau de normalização**

# O ciclo de vida de um DW





# Motivação

---

- Um modelo dimensional contém as mesmas informações que um modelo normalizado
- Os pacotes de dados oferecem um formato com as seguintes preocupações:
  - Facilidade de compreensão ao usuário
  - Desempenho da consulta
  - Resiliência à mudança.



## Requisitos básicos

---

Imagine um executivo que descreve o seu negócio como, "Nós vendemos *produtos* em vários *mercados* e medimos o nosso desempenho ao longo do *tempo*."



## Orientado por assunto

Projetistas multidimensionais devem ouvir atentamente a ênfase no **produto, mercado e tempo.**



# Modelo conceitual dimensional

---

- Tabelas Fato
  - Servem para **o armazenamento, medidas** (quase sempre) **numéricas** associadas a **eventos** de negócio
- Tabelas Dimensão
  - Representa **entidades de negócios** e constituem as estruturas de entrada que servem para armazenar informações como tempo, geografia, produto, cliente
  - As tabelas dimensão têm uma relação 1:N com a tabela fato

## Tabela Fato

---

- Armazena as **medições de desempenho** decorrentes de eventos dos **processos de negócios** de uma organização
  - Representa **uma medida** de negócios
- Uma tabela fato contém vários **fatos**, correspondentes a **cada uma das suas linhas**
  - Cada linha corresponde a um evento de medição
- Os dados em cada linha estão a um nível específico de detalhe, referido como **o grão**
  - Ex: uma linha por produto vendido numa operação de venda

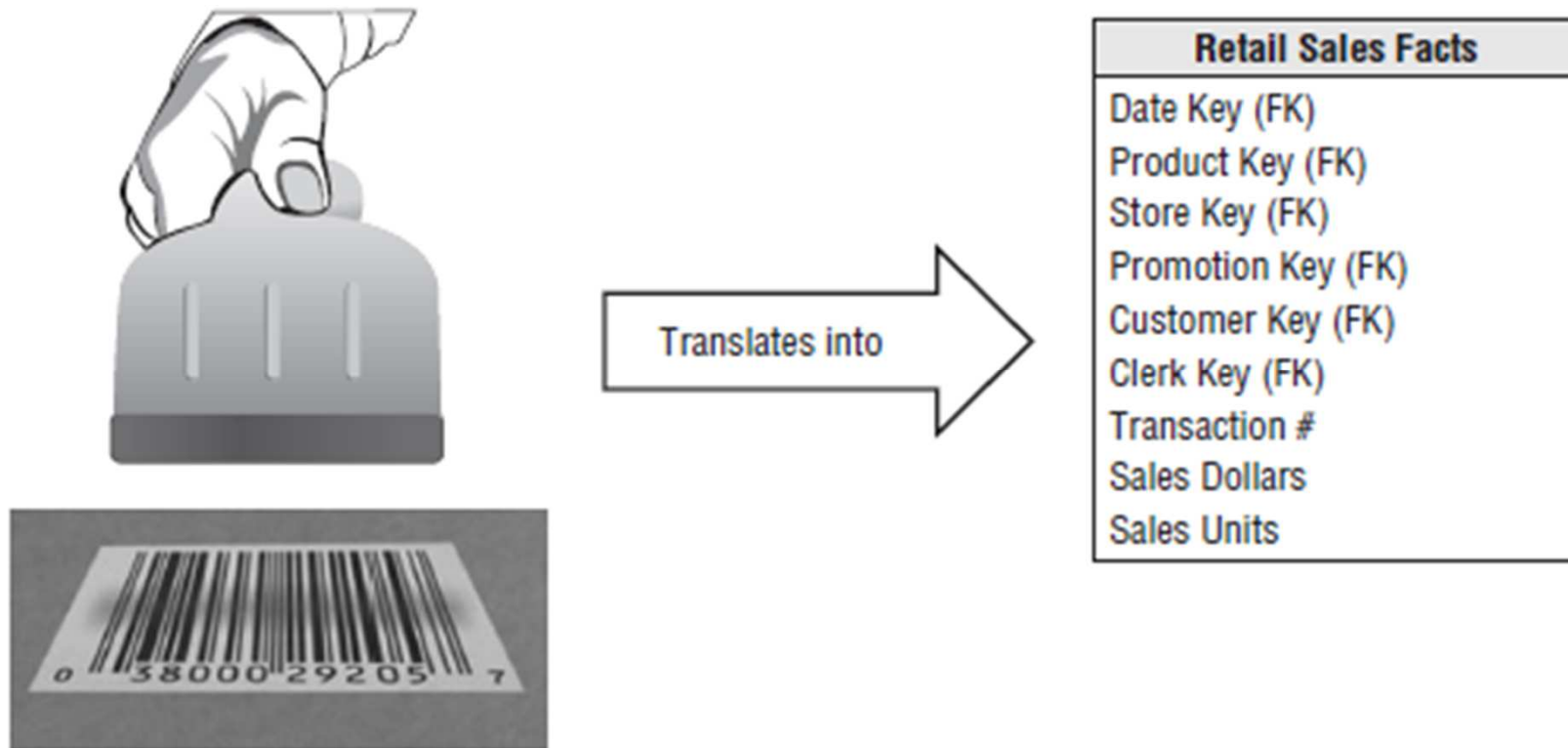
## Fatos para medições

---

- Uma única linha da tabela fato tem uma relação **um-para-um com o evento de medição**, como descrito pela **granularidade** da tabela fato
- Uma tabela fato corresponde a **um evento físico observável**, e não às exigências de um relatório específico
- Dentro de uma tabela fatos, apenas fatos consistentes com a granularidade definida são permitidos
  - Por exemplo, em uma transação de vendas no varejo, a quantidade de um produto vendido e seu preço são bons fatos

# Tabela Fato

---



## Tabelas dimensão (Contexto descritivo)

---

- São companheiros integrais para uma tabela de fatos
- Contém **o contexto textual** associado a um evento de medição dos processos de negócios
- Descrevem o "**quem, o que, onde, quando, como e porquê**" associado ao evento

## Dimensões para o contexto descritivo

---

- Contém **os atributos descritivos** usados pelas aplicações de BI para **filtrar e agrupar** os fatos
- Com a **granularidade** de uma tabela de fato definida, todas as possíveis dimensões podem ser identificados.
- Sempre que possível, a **dimensão** deve ter **um valor único** associado a uma **determinada linha da tabela fato**
- Tabelas dimensão são chamados a "**alma**" do DW, pois eles contêm **os pontos de entrada e rótulos descritivos** que permitem ao sistema de DW/BI para ser aproveitado para a análise de negócios
- Um esforço é necessário para o desenvolvimento e o gerenciamento da tabelas de dimensão, pois elas **são os condutores da experiência de BI do usuário**



## Um exemplo de tabela dimensão

---

- Dimensões fornecem os pontos de entrada para os dados, e as **etiquetas finais** e os **agrupamentos** em todas as análises de DW/BI

Product Dimension
Product Key (PK)
SKU Number (Natural Key)
Product Description
Brand Name
Category Name
Department Name
Package Type
Package Size
Abrasive Indicator
Weight
Weight Unit of Measure
Storage Type
Shelf Life Type
Shelf Width
Shelf Height
Shelf Depth
...

## Sobre as tabelas

---

- O dilema de designer
  - Uma quantidade numérica é um fato ou um atributo de dimensão?
    - É uma decisão difícil!
  - Observações numéricas **continuamente valorizadas** são quase sempre os medidas da tabela **fato**
  - Observações numéricas **discretas e extraídas de uma lista pequena**, quase sempre são **atributos de dimensão**

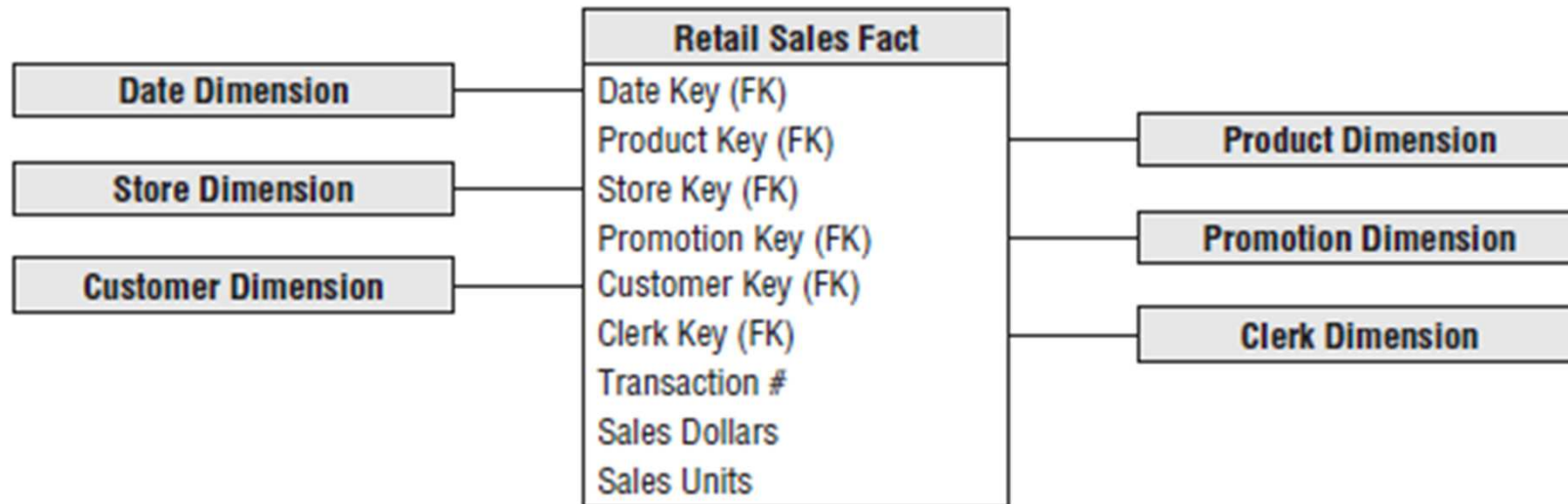
## Fatos e dimensões: Juntos no Star Schema

---

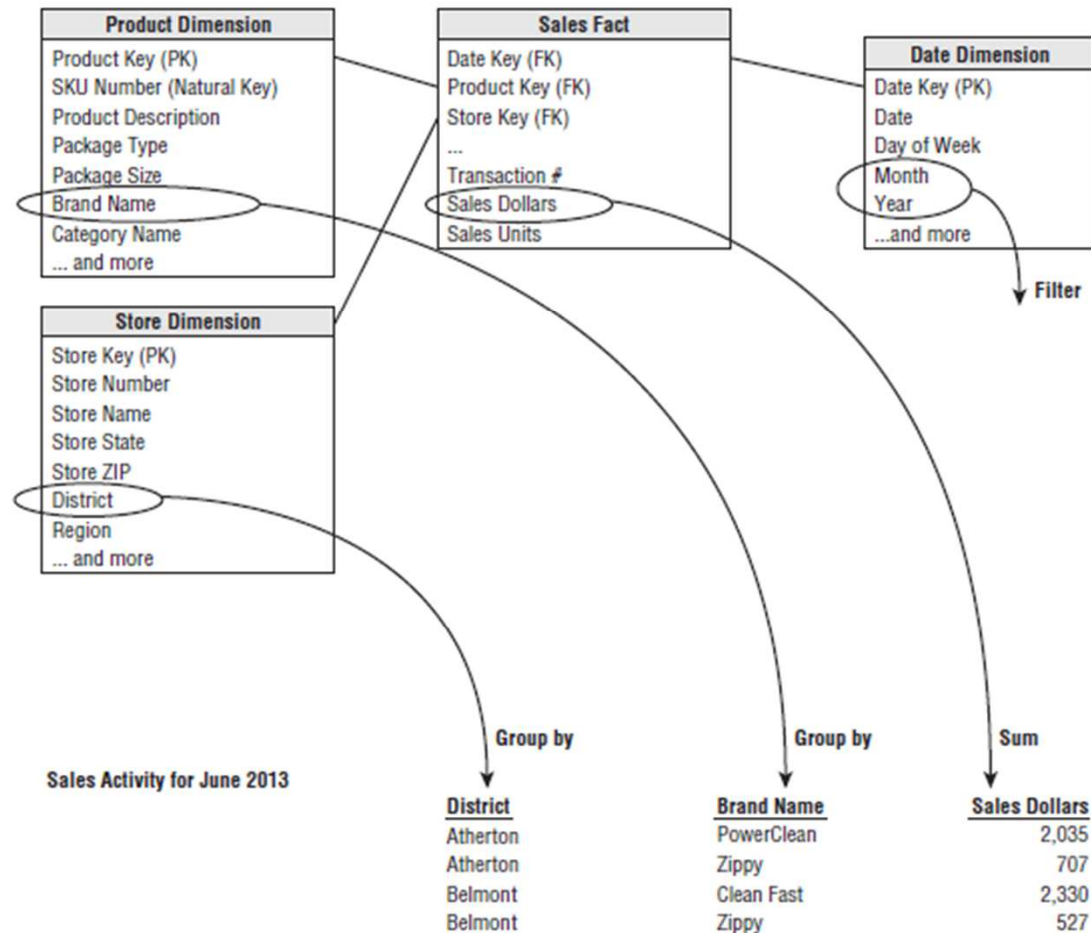
- Cada processo de negócio é representado por um modelo dimensional que consiste em:
  - **Uma tabela fato** contendo medições numéricas do evento
  - Cercada por um **conjunto de tabelas dimensão** que contêm o contexto no momento em que ocorreu o evento.
- Esta característica de **estrutura estrela (star schema)** é muitas vezes chamado de junção estrela (star-join), um termo que remonta aos primórdios de bancos de dados relacionais

# Tabela fato e dimensões juntas

---



# Um simples relatório



```
SELECT
    store.district_name,
    product.brand,
    sum(sales_facts.sales_dollars) AS "Sales Dollars"
FROM
    store,
    product,
    date,
    sales_facts
WHERE
    date.month_name="January" AND
    date.year=2013 AND
    store.store_key = sales_facts.store_key AND
    product.product_key = sales_facts.product_key AND
    date.date_key = sales_facts.date_key
GROUP BY
    store.district_name,
    product.brand
```

## Questão 10 - CESGRANRIO - 2011 - Petrobrás - Analista de Sistemas Júnior - Engenharia de Software

---

O modelo dimensional, implementado em sistemas de Data Warehouse, tornou-se uma importante ferramenta de análise de negócios. Nesse modelo, (a) apenas dados sumarizados são usados.

(b) as linhas das tabelas dimensão representam as medidas de negócio.

(c) as tabelas fato expressam relacionamentos de muitos para muitos entre as tabelas dimensão.

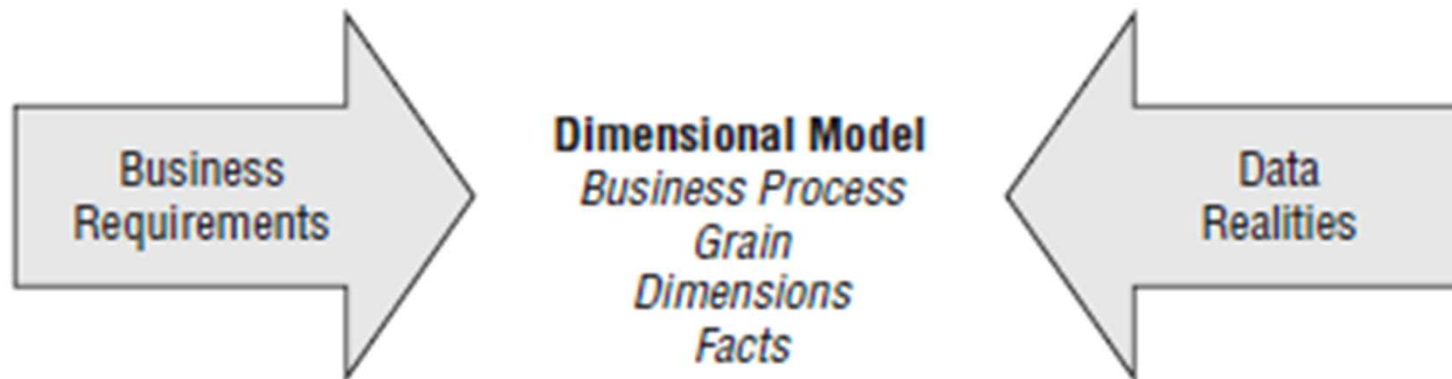
(d) as tabelas fato apresentam dados em diferentes granularidades.

(e) os fatos mais úteis são textuais, provendo informação ao usuário final.

# Processo de design multidimensional

---

- Selecionar o processo de negócio
- Definir a granularidade
- Identificar as dimensões
- Identificar os fatos



## Processo de negócio

---

- Eventos de processos de negócios podem **gerar ou capturar as métricas de desempenho** que se **traduzem em fatos** em uma tabela de fatos
- A maioria das tabelas de fatos concentram-se nos resultados de **um único processo de negócio**
- A escolha do processo é importante!
  - **Define um objetivo** de concepção específica e **permite que a granularidade, as dimensões e os fatos** sejam declarados



# Passo 1: Selecionando o processo de negócio

---

- O primeiro passo é **decidir o processo de negócio** para o modelo, combinando a compreensão dos **requisitos de negócio** com a análise dos **dados de origem disponíveis**
  - O **primeiro projeto de BI/DW** deve se concentrar no processo de negócio que é **o mais crítico** para **os usuários de negócios**
  - **Viabilidade** abrange uma série de considerações, incluindo a **disponibilidade e qualidade dos dados**, e ainda a **capacidade organizacional**
- Nosso exemplo: Suponha que os gestores querem entender melhor as compras dos clientes
  - O processo de negócio que você está modelando é baseado nas **transações de vendas de varejo**
  - Esta informação permite que **os usuários de negócios** analisar, por exemplo, quais produtos estão vendendo em que as lojas em que dias em que condições promocionais

# Granularidade

---

- Declarar a granularidade é um passo fundamental em um projeto dimensional
- Estabelece exatamente o que uma única linha da tabela fato representa
  - Torna-se um contrato vinculativo sobre o design
  - Deve ser declarada **antes de escolher dimensões ou fatos**, porque cada **dimensão candidata ou fato relevante** deve ser **consistente** com a **granularidade**
- A atomicidade do grão refere-se ao nível mais baixo no qual os dados são capturados por um determinado processo de negócio
  - Kimball encoraja a começar pelos dados de grãos atômicos, porque **resiste ao ataque de consultas de usuários imprevisíveis**
  - Sumarização *roll-up* são importantes para o ajuste de desempenho, mas eles pressupõem perguntas comuns do negócio
  - Cada **proposta de granularidade** da tabela fato resultada em **uma tabela física separada**, diferentes granularidades não devem ser misturados na mesma tabela fato

## Passo 2: Declare a granularidade

---

- Qual o **nível de detalhe dos dados** devem ser disponibilizados no modelo dimensional?
- Você deve desenvolver modelos dimensionais que representem as informações **mais detalhadas, atômicas capturadas** por um processo de negócio
- Um sistema de DW/BI exige quase sempre os dados expressos no mais baixo possível de granularidade
  - Não porque as consultas querem ver linhas individuais, mas **porque as consultas precisam cortar os detalhes de formas muito precisas**

## Passo 3: Identificando as dimensões

---

- A declaração cuidadosa da **granularidade determina a dimensão** primária da tabela fato
- Em seguida, adiciona-se mais dimensões para a tabela fato, **se essas dimensões adicionais tiverem apenas um valor** para cada combinação das dimensões principais
  - Se a dimensão adicional viola a granularidade, fazendo com que linhas adicionais da tabela fato sejam geradas, a dimensão precisa ser desclassificado ou a granularidade precisa ser revista
- As dimensões descritivas aplicam ao nosso caso:
  - Data, produto, loja, promoção, caixa, e forma de pagamento
  - Além disso, o número do bilhete de transação é incluída como uma dimensão especial, dita **dimensão degenerada** para números de transação

## Passo 4: Identificando os fatos

---

- A quarta e última etapa do projeto é a determinação cuidadosa dos fatos que aparecerão na tabela fato
  - Mais uma vez, a declaração de granularidade ajuda a ancorar o raciocínio
- Ao considerar **os fatos em potencial**, você pode descobrir novamente ajustes precisam serem feitos tanto para **granularidade** quanto para a escolha das dimensões
- Os dados coletados pelo sistema incluem a quantidade de vendas por unidade regular, o desconto, preços líquidos pagos, valores da vendas em dólares
- O valor da vendas é igual à quantidade de vendas multiplicado pelo preço unitário. Da mesma forma, o valor do desconto é a quantidade de vendas multiplicado pelo valor do desconto unitário
  - Alguns sistemas sofisticados também fornecem um custo dólar padrão para o produto entregue para a loja pelo vendedor.
  - Presumindo que este fato custo é prontamente disponível e não requer uma iniciativa baseado em atividade heroica de custo adicional, você pode incluir o valor de custo estendido na tabela de fatos

## Revisitando a tabela fato

---

- Uma tabela fato contém as **medidas numéricas** produzidos por um evento de medição operacional no mundo real
- No nível mais baixo de granularidade, uma linha da tabela fato **corresponde a um evento** de medição (fato) e vice-versa
- Além das medidas numéricas, uma tabela fato sempre **contém chaves estrangeiras para cada uma das suas dimensões associadas**, bem como as chaves de **dimensão degenerados** opcionais
- As tabelas fatos são o principal alvo de **computações e agregações dinâmicas** decorrentes das consultas

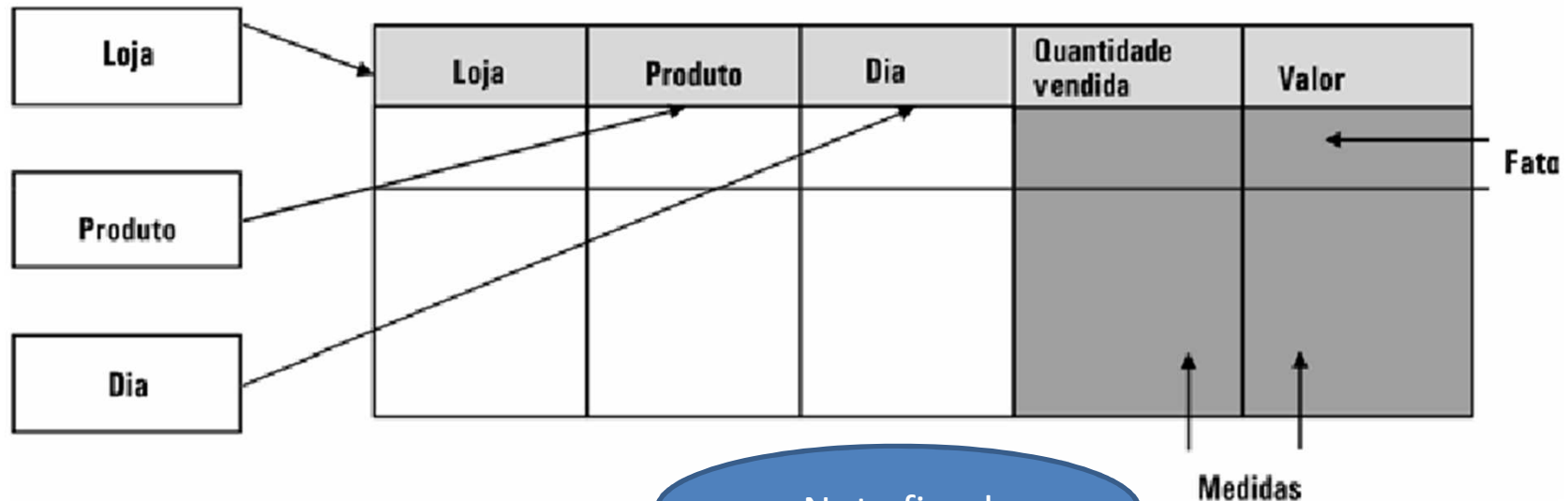
# Técnicas para a tabela fato

---

- Atributos
  - Medidas: Aditivas, semi-aditivas e não-aditivas
  - Nulos na tabela fato
  - Fatos conformes (Conformed Facts)
- Tipos de tabela fato
  - Tabela fato de transações (Transaction Fact Tables)
  - Tabela de **Snapshot Periódicos**
  - Tabela de **Snapshot Acumulativos**
- Tabela fato sem fato (Factless Fact Tables)
- Tabela fato consolidada (Consolidated Fact Tables)

# Composição básica de uma tabela fato

Tabelas  
Dimensão



Taxonomia da tabela fato:

- Fatos transações
- Fatos periódicos
- Fatos acumulados

Nota fiscal

Vendas no Dia

Valor-Gasto-Realizado



## Questão 11 - CESGRANRIO - 2008 - Petrobrás - Analista de Sistemas Júnior - Processos de Negócios

---

A empresa passou a sortear cupons de desconto para alguns clientes, os quais dão direito a um desconto nas compras em uma determinada data. A informação sobre que clientes possuem cupons para que datas é mantida de forma independente e consolidada no processo de extração, transformação e carga, resultando em um campo, na tabela fato, indicando se a venda foi realizada com o desconto ou não. A solução parecia atender bem às demandas dos usuários do data warehouse, até que um deles tentou realizar uma consulta para saber quais clientes não haviam realizado compras, mesmo tendo um cupom de desconto para a data. Este tipo de demanda tipicamente será resolvido introduzindo, no data warehouse, uma

- (a) tabela de fatos complementares (complimentary fact table).
- (b) tabela de fatos sem dimensão (dimensionless fact table).
- (c) tabela de fatos sem fatos (factless fact table).
- (d) dimensão multivalorada (multivalued dimension).
- (e) dimensão degenerada (degenerated dimension).

# Técnicas básicas para tabelas dimensão

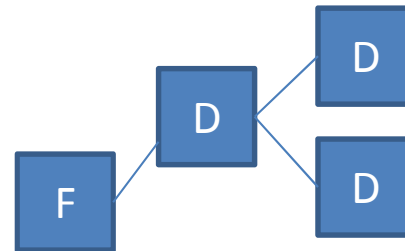
---

- Estrutura da tabela dimensão
  - Chaves:
    - Chave artificial (substituta) da dimensão (Surrogate Keys)
    - Chave natural, durável e supernatural
  - Dimensão degenerada
  - (Des)Normalização
    - Dimensões achatadas (Denormalized Flattened Dimensions)
  - Múltiplas hierarquias nas dimensões (Multiple Hierarchies in Dimensions)

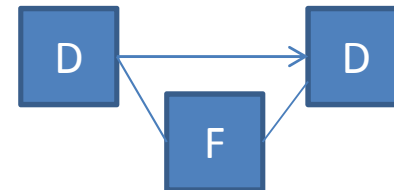
# Técnicas básicas para tabelas dimensão

---

- Flags e indicadores como atributo textuais
  - Dimensão Junk
- Atributos nulos nas dimensões
- Dimensões data (PK:AAAAMMDD) e calendários
- Papéis de dimensões
- Dimensão Snowflake



- Dimensão Outrigger



## Questão 12 - CESPE - 2012 - TCE-ES - Auditor de Controle Externo - Tecnologia da Informação

---

Acerca de data warehousing e OLAP, julgue os itens seguintes.

[134] Na modelagem multidimensional da tabela fato, a chave da dimensão tempo deve ser sempre representada como parte da chave primária

[As 10 Regras Essenciais para a Modelagem de Dados Dimensional](#)

---

## Questão 13 - CESPE - 2010 - TRT - 21ª Região (RN) - Analista Judiciário - Tecnologia da Informação

---

Acerca de sistemas de suporte a decisão e *data warehousing*, julgue os itens a seguir.

[1] Em um *data warehouse*, cada linha em uma tabela fato corresponde a uma medida, representada por um valor aditivo, em que necessariamente essas medidas não compartilham a mesma granularidade.

# Exemplo do modelo dimensional clássico

**Cod-Loja (PK) + Atributos**



**Cod-Dia (PK) + Atributos**



**Cod-Produto (PK) + Atributos**

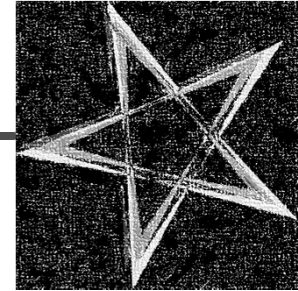
Esquema Dimensional:

 Tabela Fatos-Chaves mais  
Fatos Tratados Numericamente

 Tabelas Dimensão-Chaves mais  
Atributos de Filtros

# Revisando: Star Schema x Snow Flake

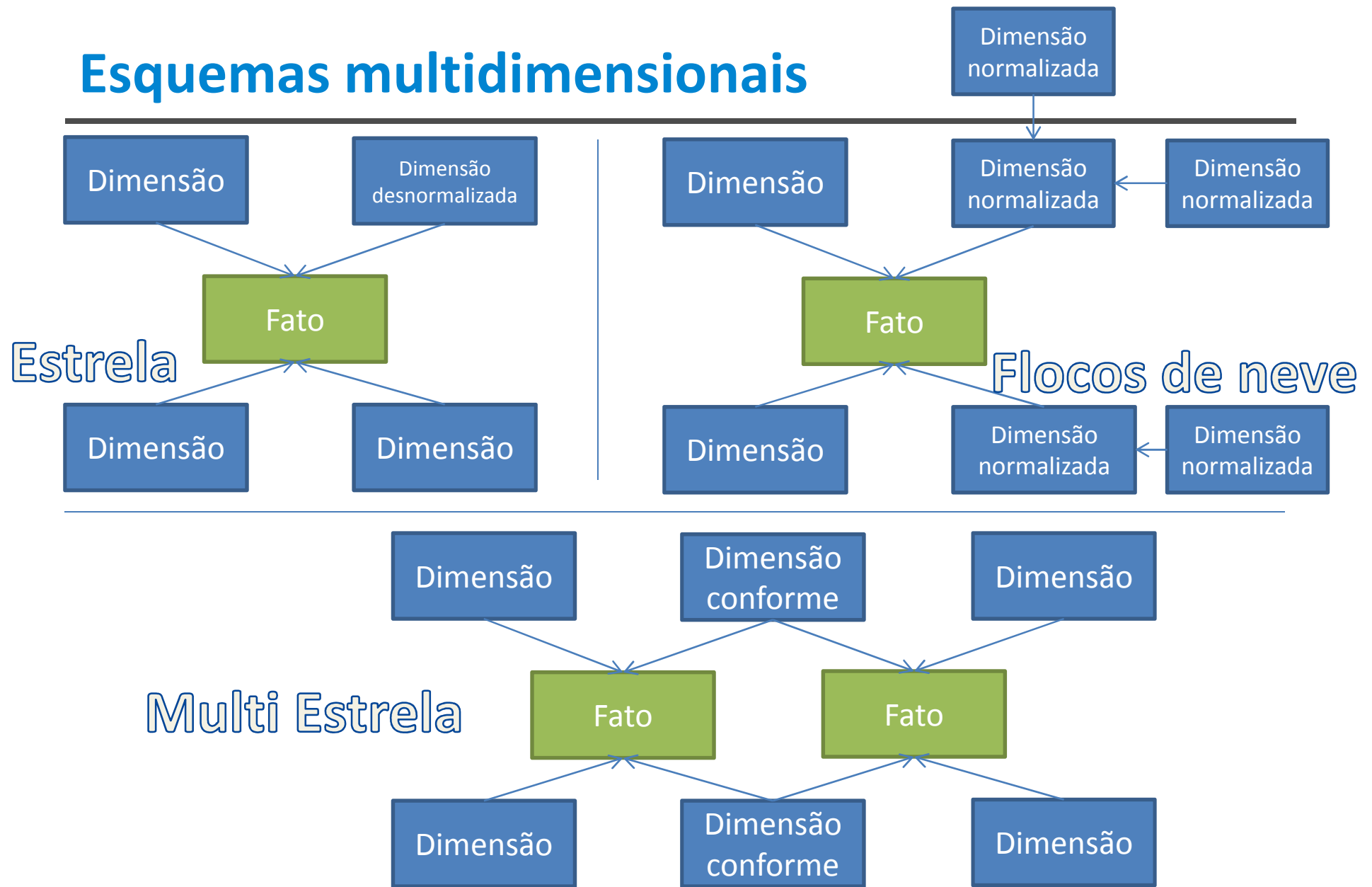
---



- Star Schema
  - Esquema estrela: abordagem que recomenda a não normalização das tabelas dimensão
- Snow Flake
  - Esquema de flocos de neve: abordagem que recomenda a normalização das tabelas dimensão.



# Esquemas multidimensionais





## Questão 14 - CESGRANRIO - 2006 - Petrobrás - Analista de Sistemas Pleno - Engenharia de Software

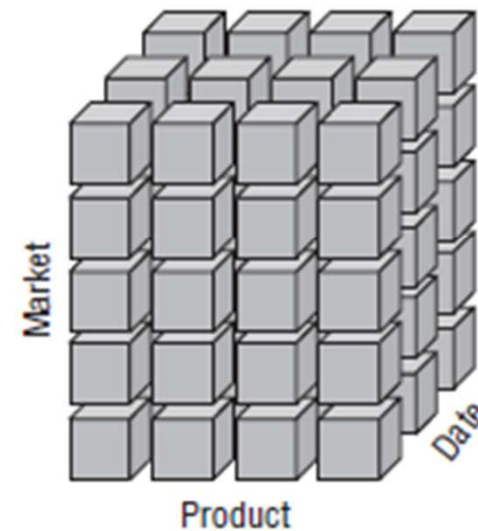
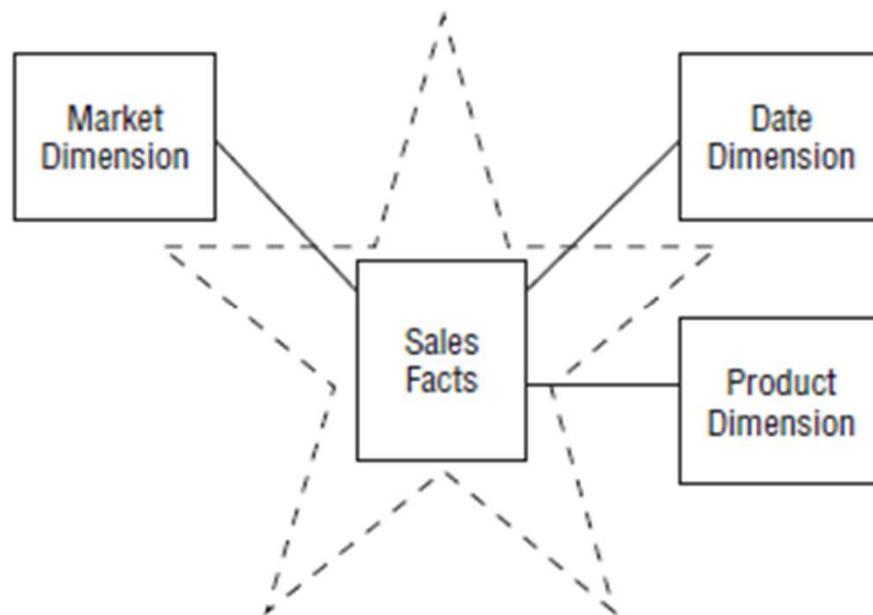
---

Bases de dados de sistemas de informações gerenciais são projetadas, muitas vezes, em um esquema denominado *estrela* que:

- (a) é uma arquitetura física que permite definir somente estruturas unidimensionais de dados.
- (b) é composta por uma tabela central, denominada de *dimensão*, e várias tabelas periféricas a ela relacionadas, denominadas  *fatos*.
- (c) apresenta tabelas periféricas que se encontram desnormalizadas.
- (d) faz com que a expansão e a evolução da base de dados necessite de grande atividade de manutenção.
- (e) é um esquema onde o número de junções realizadas é relativamente maior que o realizado em bases de dados relacionais convencionais.

# Star Schemas e OLAP Cubes

---



# Star Schemas e OLAP Cubes

---

- Esquemas estrela são estruturas multidimensionais implantados em um SGBD relacional (RDBMS)
  - Consistem em **uma tabela fato** ligados um conjunto de **tabelas de dimensões** associadas através de **relações de chave primária/estrangeira**
- O cubo de processamento analítico on-line (OLAP) é uma estrutura multidimensional implementada em um banco de dados multidimensionais, que pode ser equivalente em conteúdo, ou mais frequentemente derivado de um esquema em estrela relacional
  - Um cubo OLAP contém atributos dimensionais e fatos, mas é acessado através de **linguagens com mais capacidade analítica** do que SQL, como **XMLA e MDX**
  - Cubos OLAP são incluídos nesta lista de **técnicas básicas** porque um cubo OLAP é muitas vezes o último passo na implantação de um sistema de DW/BI dimensional, ou pode existir como uma estrutura agregada com base em um esquema estrela relacional mais atômico

# Considerações sobre OLAP

---

- Aqui estão algumas coisas para manter em mente se você implantar dados em cubos OLAP:
  - Um **esquema em estrela** hospedado em um BD relacional é uma **boa base física** para a construção de um cubo OLAP, e é geralmente considerado como uma base mais estável para **backup e recuperação**
  - Cubos OLAP eram tradicionalmente conhecidos por vantagens extremas de desempenho comparados com RDBMSs, mas essa distinção tornou-se menos importante com avanços tecnológicos
  - **Estruturas de dados de cubo OLAP** são as **mais variadas** entre os **diferentes fornecedores** de SGBDs relacionais, assim, os detalhes da implantação final muitas vezes dependem do fornecedor foi escolhido
  - Normalmente, **é mais difícil de portar aplicações de BI** entre as diferentes ferramentas OLAP do que portar aplicações de BI entre diferentes bancos de dados relacionais.

# Tabela fato agregada ou cubos OLAP?

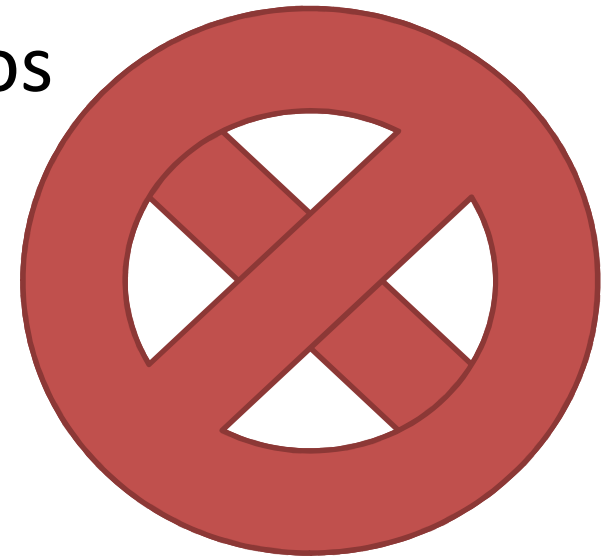
---

- Tabelas fato agregadas são formadas pela agregação de valores numéricos de dados atômicos da tabela de fatos construídas apenas para acelerar o desempenho da consulta
  - Estas tabelas devem estar **disponíveis para a camada de BI**, ao mesmo tempo em que as tabelas fatos atômica de modo que as ferramentas de BI possam escolher qual o nível agregação apropriada no momento da consulta
  - Este processo, conhecido como **navegação agregada**, deve ser aberto para que **todas as ferramentas** de relatório, ferramenta de consulta, e aplicações de BI
  - Um projeto de **tabelas agregadas** deve **comportar-se** como **índices** de banco de dados, que **aceleram o desempenho** da consulta, mas não são encontrados diretamente pelas aplicações de BI ou usuários empresariais
  - Tabelas de fatos agregados contém **chaves estrangeiras** para dimensões encolhidas, bem como **os fatos agregados** criados a partir da soma das medidas de tabelas de fatos mais atômicas
- **Cubos OLAP** agregados com medidas resumidas são frequentemente construídos da mesma forma que os agregados relacionais, mas os cubos OLAP são feitos para serem **acessados diretamente pelos usuários corporativos**

## Mitos da modelagem multidimensional

---

- Mito 1: são apenas para dados de resumo
- Mito 2: são departamentais e não da empresa
- Mito 3: não são escaláveis
- Mito 4: são apenas para uso previsível
- Mito 5: não podem ser integrados





# Extração, transformação e carga (ETL)

---

Um ambiente de DW **devidamente projetado** possui um *trade off* entre o trabalho na sala de aplicações de BI e o trabalho do sistema de ETL.

O **primeiro** deve ser feito **repetidas vezes** pelos **usuários de negócios**, enquanto o **segundo** é feito **uma vez** pela **equipe de ETL**

## Motivação

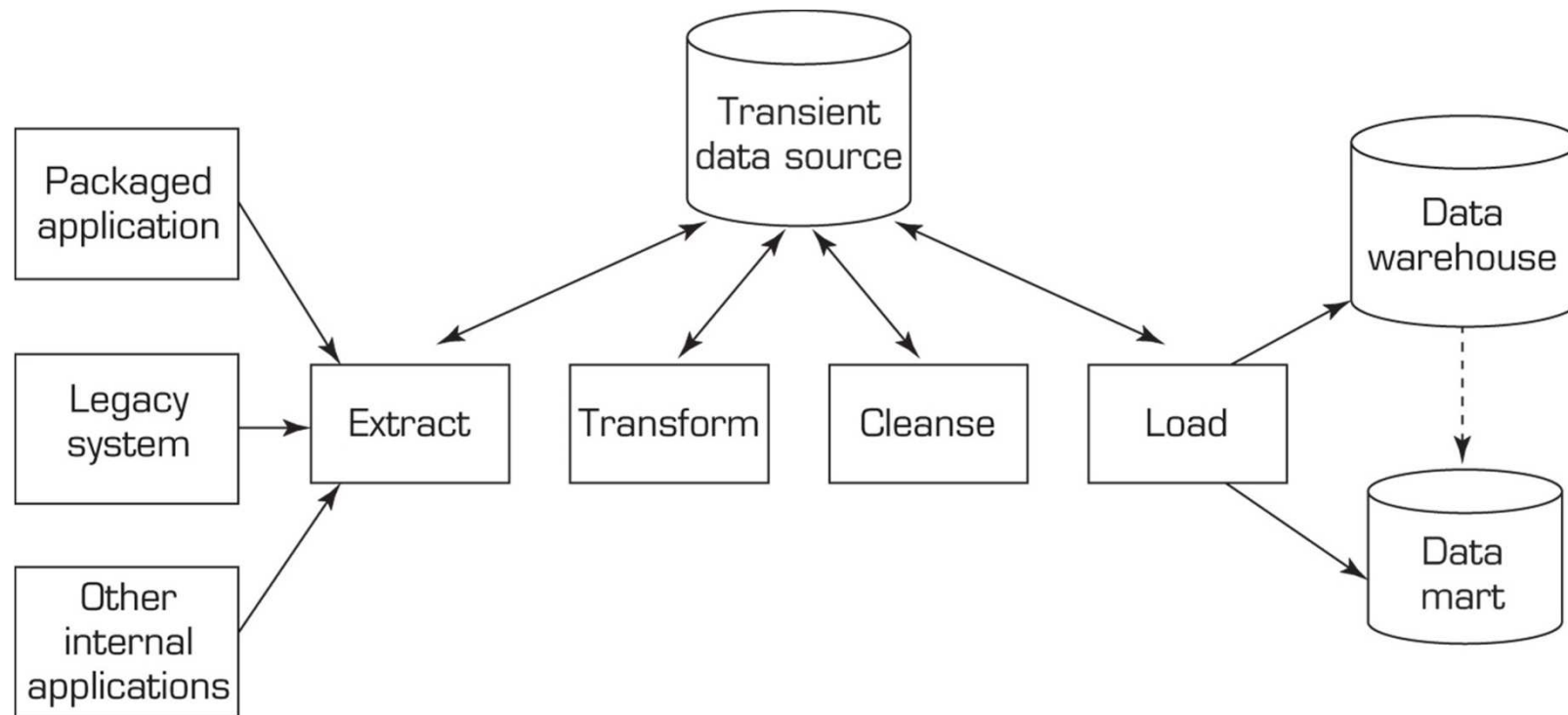
---

- O processo de ETL é o processo **mais crítico e demorado** na construção de um DW
  - ETL e as ferramentas de limpeza de dados consomem **um terço do orçamento** num projeto de DW
  - **80% do tempo de desenvolvimento** de um DW consiste no processo de ETL



# O processo de ETL

---



## Requisitos de ETL (10)

---

1. Necessidades de negócio
2. Conformidade (Contexto legal)
3. Qualidade dos dados
4. Segurança
5. Integração dos dados (MDM)
6. Latência do dados
7. Arquivamento e *Lineage*
8. Interfaces de entrega de BI
9. Habilidades disponíveis
10. ETL Legado

## Questão 15 - FCC - 2011 - TRE-PE - Analista Judiciário - Análise de Sistemas

---

Um processo importante que ocorre em relação à formação de um data *warehouse* é a obtenção dos dados de uma ou mais bases de dados da origem. Deve ser rigoroso para evitar a deformação e/ou a perda dos dados quando passados da fonte original para o destino. Trata-se de

- (a) MINING.
- (b) DATA MART.
- (c) MOLAP.
- (d) STAR.
- (e) ETL.

## Questão 16 - CESGRANRIO - 2010 - Petrobrás - Analista de Sistemas Júnior - Processos de Negócios

---

No contexto de Data Warehouses, o processo de Extração, Transformação e Carga (ETC)

- (a) revela-se como uma das etapas importantes do processo de criação do data warehouse, já que sua função é obter automaticamente os conhecimentos necessários para a padronização dos dados em modelos multidimensionais.
- (b) produz, ao seu término, uma série de tabelas (chamadas fatos) que se caracterizam por possuírem dados normalizados até a 3ª forma normal.
- (c) apresenta, como algumas de suas tarefas, filtragem, integração, conversão, condensação e derivação dos dados de entrada, que podem ser originários de diversas fontes, inclusive externas aos sistemas OLTP da organização.
- (d) leva em consideração o modelo conceitual de dados das fontes de dados, que é geralmente expresso como modelo entidade-relacionamento.
- (e) considera somente os dados provenientes de sistemas OLTP como válidos para o processo e, caso exista a necessidade de consideração de dados externos, estes devem ser importados para os sistemas legados

## Componentes do ETL

---

- Descrevem as etapas do processo tem quatro componentes principais:
  - Extração
  - Limpeza e conformidade
  - Entrega
  - Gestão

# Subsistemas de ETL (34 subsistemas)

---

- **Extração**
  - **Coleta de dados brutos** a partir dos sistemas de origem e, geralmente, gravação em disco no ambiente de ETL antes de qualquer reestruturação dos dados. Subsistemas de 1 a 3 devem apoiar o processo de extração (3)
- **Limpeza e conformidade**
  - O envio de dados de origem através de uma série de **etapas de processamento** no sistema de ETL para **melhorar a qualidade dos dados** recebidos a partir da fonte , e a **fusão de dados** a partir de duas ou mais fontes de criar e aplicar **dimensões conformes e métricas conformadas** . Subsistemas de 4 a 8 descrevem a arquitetura necessária para suportar os processos de limpeza e conformidade (5)
- **Entrega**
  - A **estruturação física e carregamento dos dados** em modelos dimensionais de destino no servidor de apresentação. Subsistemas de 9 a 21, fornecem a capacidade de entrega dos dados (13)
- **Gestão**
  - **Gestão dos sistemas e processos** relacionados do ambiente de ETL de uma forma coerente. Subsistemas de 22 a 34 descrevem os componentes necessários para apoiar o gerenciamento contínuo do sistema de ETL (13)

# Extração: Trazendo os dados para o DW

---

## 1. Sistema de extração

- Obviamente, a extração de dados dos sistemas de origem é um componente fundamental da arquitetura de ETL.
- O sistema de ETL deve esperar dados a partir de uma ampla variedade de sistemas que envolve diferentes tipos de dados e outros desafios inerentes.

## Extração: Trazendo os dados para o DW

---

### 2. *Data profiling* (perfil de dados)

- Técnica de análise dos dados para descrever o seu **conteúdo, consistência e estrutura**
- Fornece a equipe de ETL uma **orientação** de quanto a **limpeza de dados** é importante para o projeto
- Toda vez que você executar uma consulta SELECT DISTINCT investigativo em um campo do banco de dados, você está traçando o “perfil de dados”



## Extração: Trazendo os dados para o DW

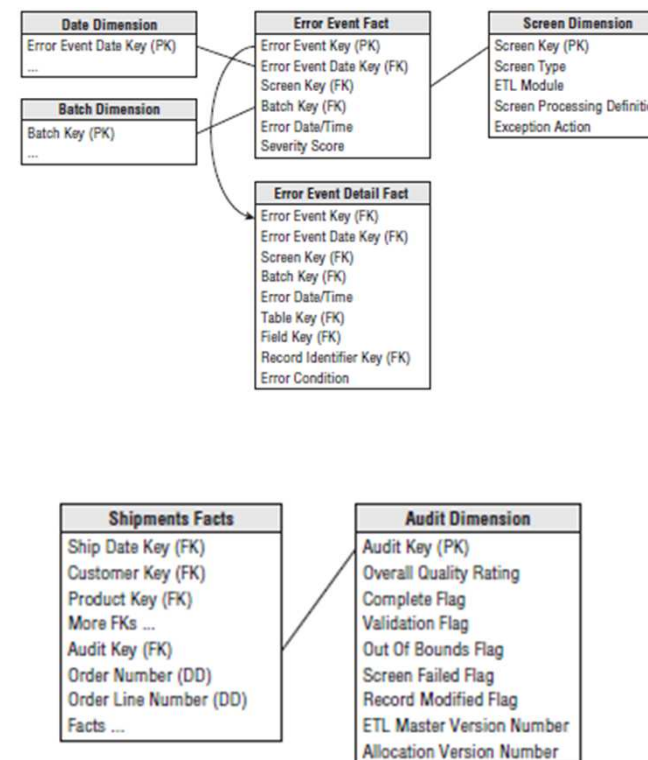
---

### 3. Sistema de captura das alterações nos dados

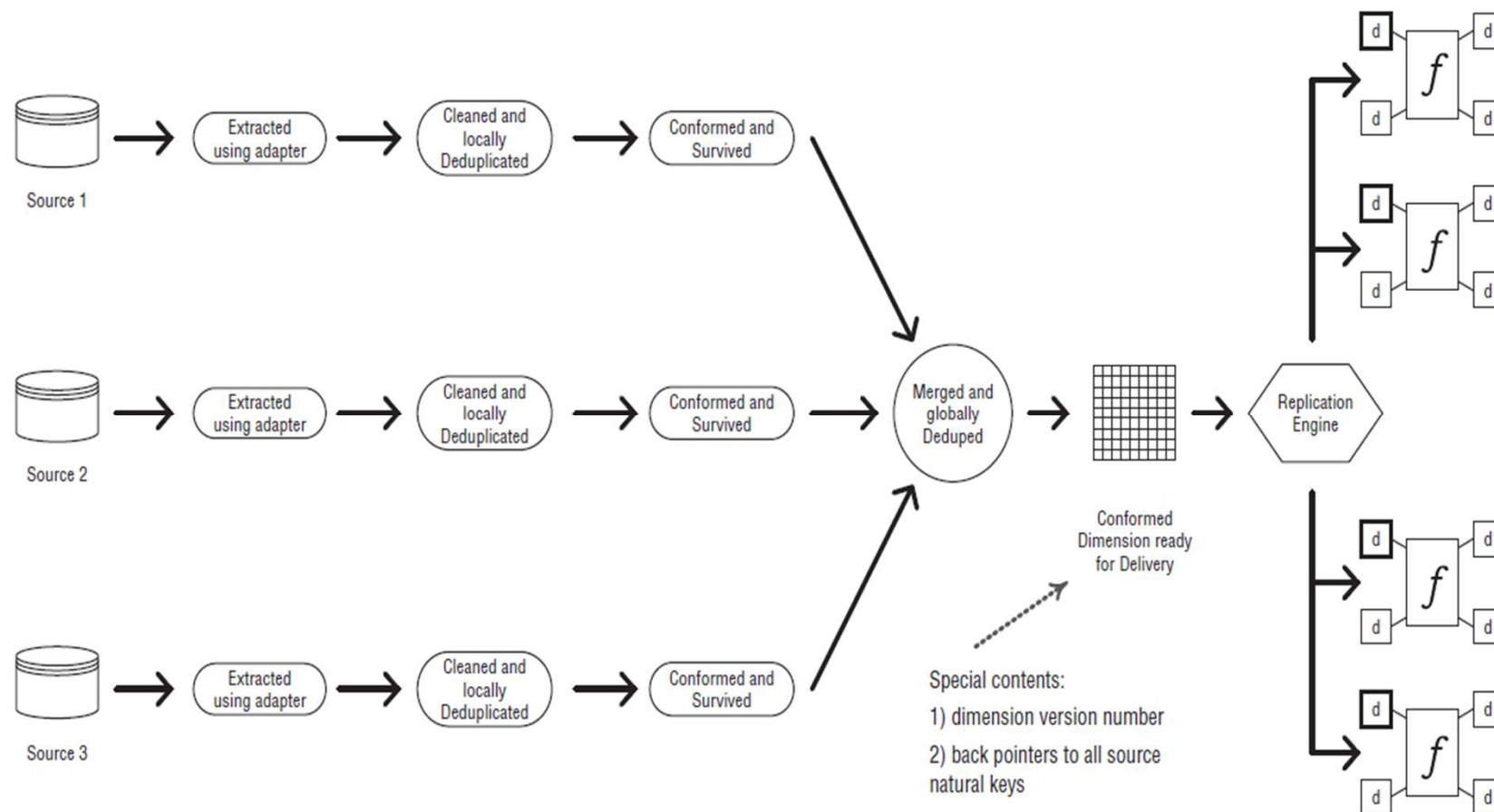
- Capacidade de transferir apenas **as alterações relevantes** da fonte de dados desde a última atualização
- Isolar os dados de origem mais recente é chamado de captura de dados de alteração (CDC)
- A ideia por trás do CDC é bastante simples: basta **transferir os dados que foram alterados** desde a última carga.
- Existem várias maneiras de capturar as alterações de dados de origem, cada uma eficaz na situação adequada, incluindo:
  - Colunas de auditoria (Audit Columns)
  - Extração de tempo (Timed Extracts)
  - Comparação diferencial total (Full Diff Compare)
  - Log do banco de dados (Database Log Scraping)
  - Monitorar a fila de mensagens (Message Queue Monitoring)

# Limpeza e Conformidade de dados

1. Sistema de limpeza de dados
2. Esquema de eventos de erro
3. Dimensão de auditoria
4. Sistema de deduplicação
5. Sistema de conformidade



# Fluxo do processo de conformidade



## Entrega: Prepare-se para Apresentação

---

1. Gestão da alteração lenta da dimensão
2. Gerador de chave substituta
3. Gerencia de hierarquia
4. Gerencia de dimensões especiais
  - *Date/Time Dimensions, Junk(lixo) Dimensions, Mini-Dimensions, Shrunk Subset Dimensions, Small Static Dimensions*
5. Construtores da tabela fato
6. Pipeline da chave substituta
7. Construção da tabela ponte para dimensão multivaloradas

## Entrega: Prepare-se para Apresentação

---

- 8. Gerenciador de dados atrasados
- 9. Sistema de gerenciamento das dimensões
- 10. Sistemas provedor dos fatos
- 11. Construção dos agregados
- 12. Construção do cubo OLAP
- 13. Gerencia da propagação de dados

## Gerenciando o ambiente de ETL

---

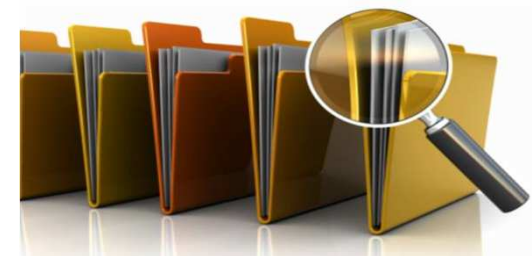
- Para atingir este objetivo, o sistema de ETL deve trabalhar constantemente para cumprir três critérios:
  - **Confiabilidade** - Os processos de ETL deve sempre ser executado consistentemente. Eles devem fornecer dados em **tempo hábil, confiáveis e em qualquer nível de detalhe**
  - **Disponibilidade** - O DW deve cumprir seus **acordos de nível de serviço (SLAs)**. O armazém deve estar pronto e disponível, como prometido
  - **Gerenciamento** - Um DW de sucesso nunca está pronto. Ele **cresce constantemente e muda** juntamente com o negócio. Os processos de ETL também precisam **evoluir**

# Gerenciando o ambiente ETL

---



1. Agendador de *Jobs*
  - *Job definition, Job scheduling, Metadata capture, Logging, Notification*
2. Sistema de backup
3. Sistema de recuperação e *restart*
4. Sistema de controle de versão
5. Sistema de migração de versão
6. *Workflow Monitor*



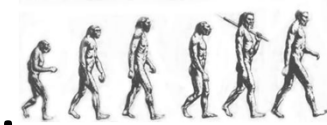
# Gerenciando o ambiente ETL

---

7. Sistema de ordenação



8. Analisador de ancestrais e dependência



9. Sistemas para problemas de escala



10. Sistema de paralelismo/pipeline



11. Sistema de segurança



12. Gerenciamento de conformidade



13. Gerenciar repositório de Metadados



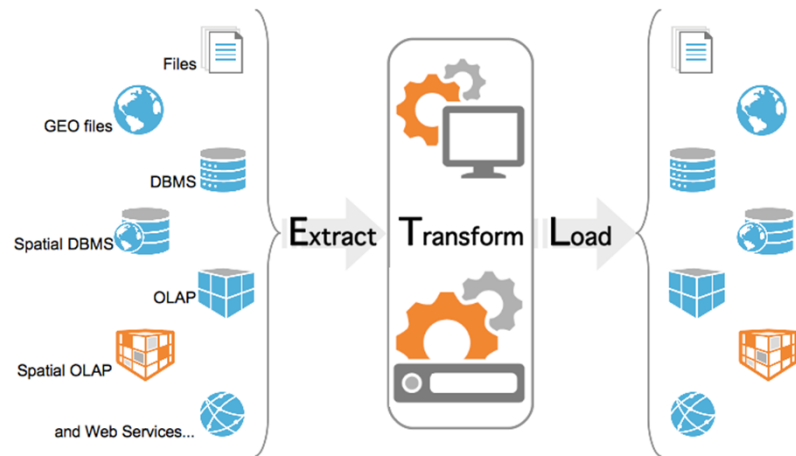


# Processo de ETL

---

- Desenvolva um plano
  1. Desenvolva um plano de alto nível
  2. Escolha a ferramenta de ETL
  3. Desenvolva um conjunto de estratégias padrão
  4. *Drill Down by Target Table*
- Desenvolver o documento de especificação do ETL
- A primeira carga dos dados
  7. Popular as tabelas dimensão com os dados históricos
  8. Executar a carga histórica da tabela fato
- O processo incremental do ELT
  7. Processamento incremental da tabela dimensão
  8. Processamento incremental da tabela fato
  9. Carga das tabelas agregadas e dos cubos OLAP
  10. Operação e automação do sistema de ETL





**ITnerante** 

**TIMASTERS** 

# ETL na prática

---



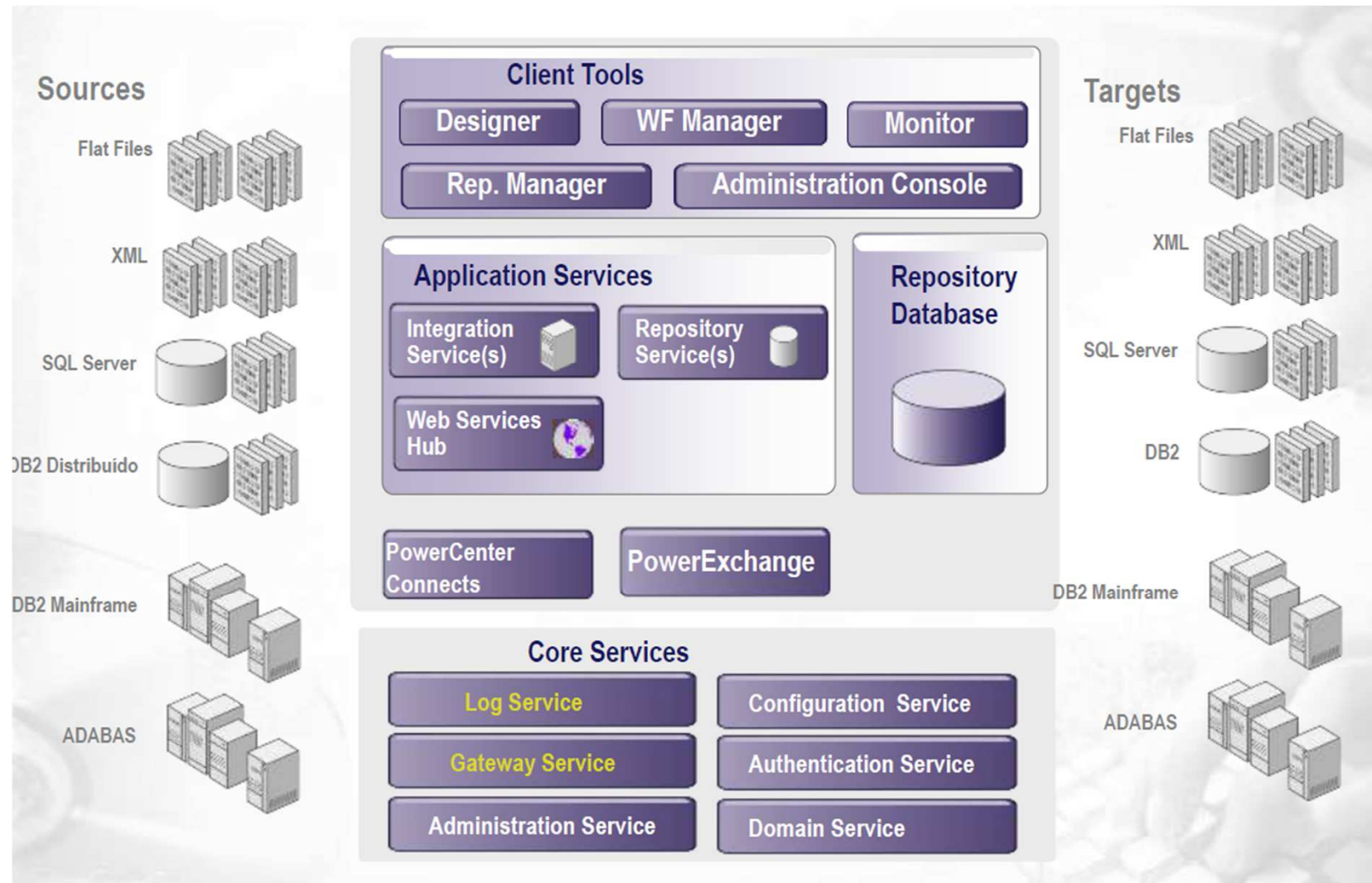
Prof. Thiago Cavalcanti

# Utilização do ETL

---

- No BCB é utilizado em
  - Migrações de dados
  - Organização de dados
  - Sincronização de dados
  - Criação de Datamarts e Warehouses
  - Conversão de tecnologia de banco de dados

# Informatica Power Center



# Processo de construção

---

## 1. Importação de definições

Sources

Targets

## 2. Construção de mapas

## 3. Configuração dos workflows

## 4. Execução dos workflows

## 5. Acompanhamento das execuções

# Power Center - Ferramentas

---



**Designer** – Usado para criar mapeamentos que, logicamente, definem o que deve ser feito e como. Mapeamentos de definem as fontes, as destinos e as transformações que se pretende executar nos dados por meio de um ambiente gráfico de *drag and drop*.



**Workflow Manager** – Usado para criar, agendar e executar fluxos de trabalho. Um fluxo de trabalho é um conjunto de instruções que descrevem como e quando executar as tarefas relacionadas com a extração, transformação e carregamento de dados.



**Workflow Monitor** – Usado para monitorar graficamente o status de fluxos de trabalho regulares e execução para cada servidor Power Center. Você pode ver quais as tarefas que tiveram êxito ou falha.



**Repository Manager** – Usado para criar e administrar o repositório de metadados. Você pode criar usuários e grupos, atribuir privilégios e permissões para eles e criar pastas que contem os metadados.

# Power Center - Ferramentas

---



**Designer** - Definição abstrata de Origens e Destinos. Só existe a definição das tabelas



**Workflow Manager** – Configuração de caminhos e Conexões

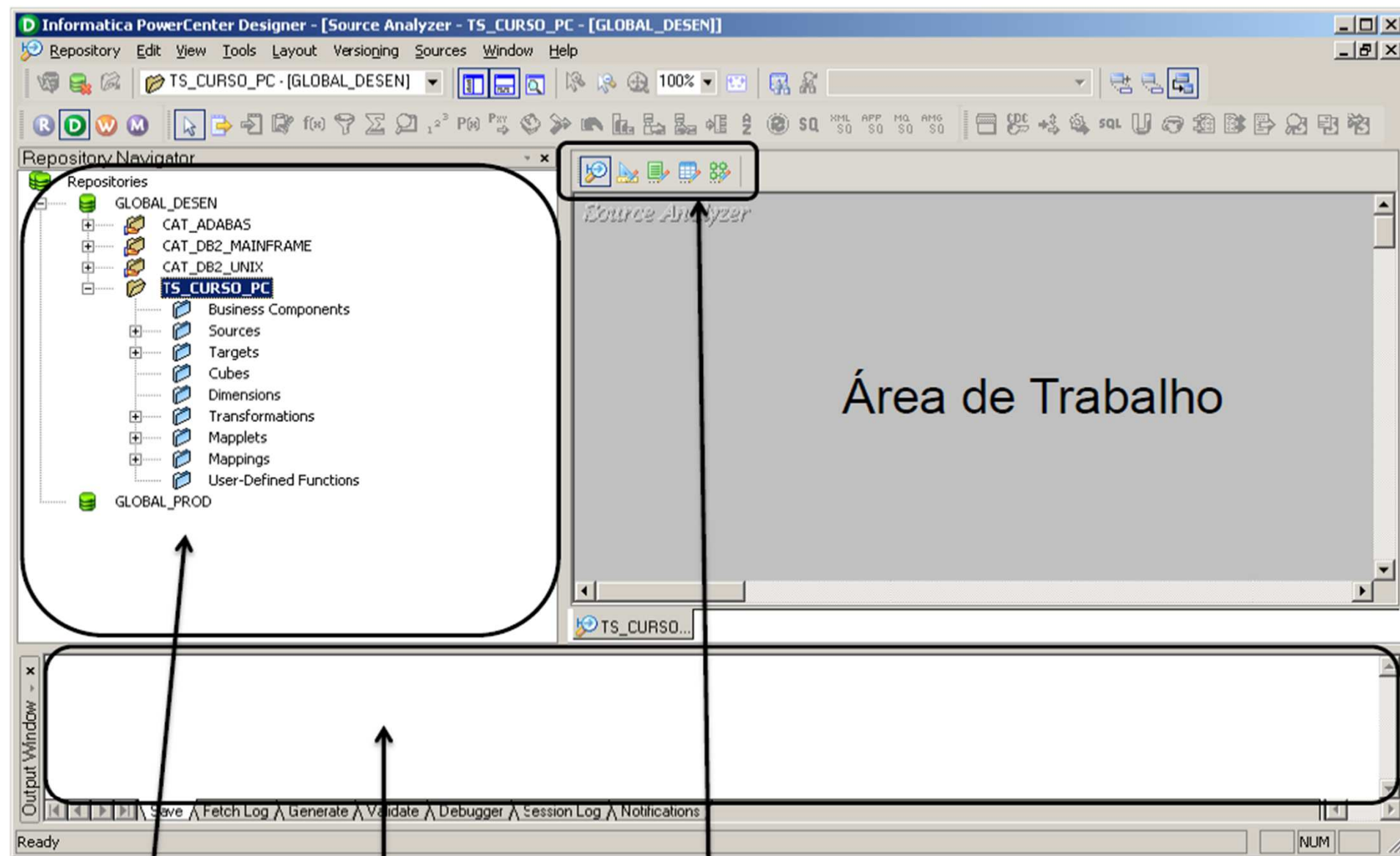


**Workflow Monitor** – Acompanhamento e verificação de Erros em Execuções



**Repository Manager** – Manipulação de objetos, versões, deploys, cópias





Navegação

Mensagens

Source Analyzer  
Target Designer  
Transformation Developer  
Maplet Designer  
Mapping Designer



Navegação

Execução

The screenshot displays the Informatica PowerCenter Workflow Monitor interface. The top menu bar includes Repository, Edit, View, Tools, Task, Filters, and Help. A toolbar with various icons is located below the menu. The main window is divided into several sections:

- Workflow Tree:** A tree view on the left showing the workflow structure. The task `s_MPP_CARREGA_INSTITUICAO_ARNALDO` is selected and highlighted in blue.
- Task Execution Table:** A table showing the execution status of tasks. The task `s_MPP_CARREGA_INSTITUICAO_ARNALDO` is listed with a duration of 00:00:08 and a status of Succeeded.
- Task Details:** A section below the table showing details for the selected task, including Source/Target Statistics, Partition Details, and Performance.
- Messages:** A section at the bottom showing a list of messages, including task updates and decisions.

Red boxes highlight the workflow tree, the task execution table, the task details section, and the messages section. Arrows point from the labels 'Navegação', 'Execução', 'Propriedades da Execução', and 'Mensagens' to these respective sections.

Task Name	Duration	Status
TR_G05_SAR		
TR_G05_SGS_SUPORTE		
TR_G05_JUD_BACENJUD		
TR_G05_SML		
TS_CURSO_PC		
TS_CURSO_PC_INSTRUTOR		
WKF_CARREGA_INSTITUICAO	00:00:08	Succeeded
s_MPP_CARREGA_INSTITUICAO_ARNALDO	00:00:06	Succeeded
TS_G00_TST		
TS_G00_TST_CAMBIO		
GLOBAL_PROD		
REPO_ZZZ		

Transformation Name	Node	Applied Rows	Affected Rows	Rejected Rows	Throughput (Rows/Sec)	Throughput (Bytes/Sec)	Bytes	Last Error Code	Last Error
INSTITUICAO_FI...	node01_sbcd...	223	223	0	223	27652	27652	0	No errors
SQL SQ_FF_INSTITU...	node01_sbcd...	223	223	0	223	19624	19624	0	No errors

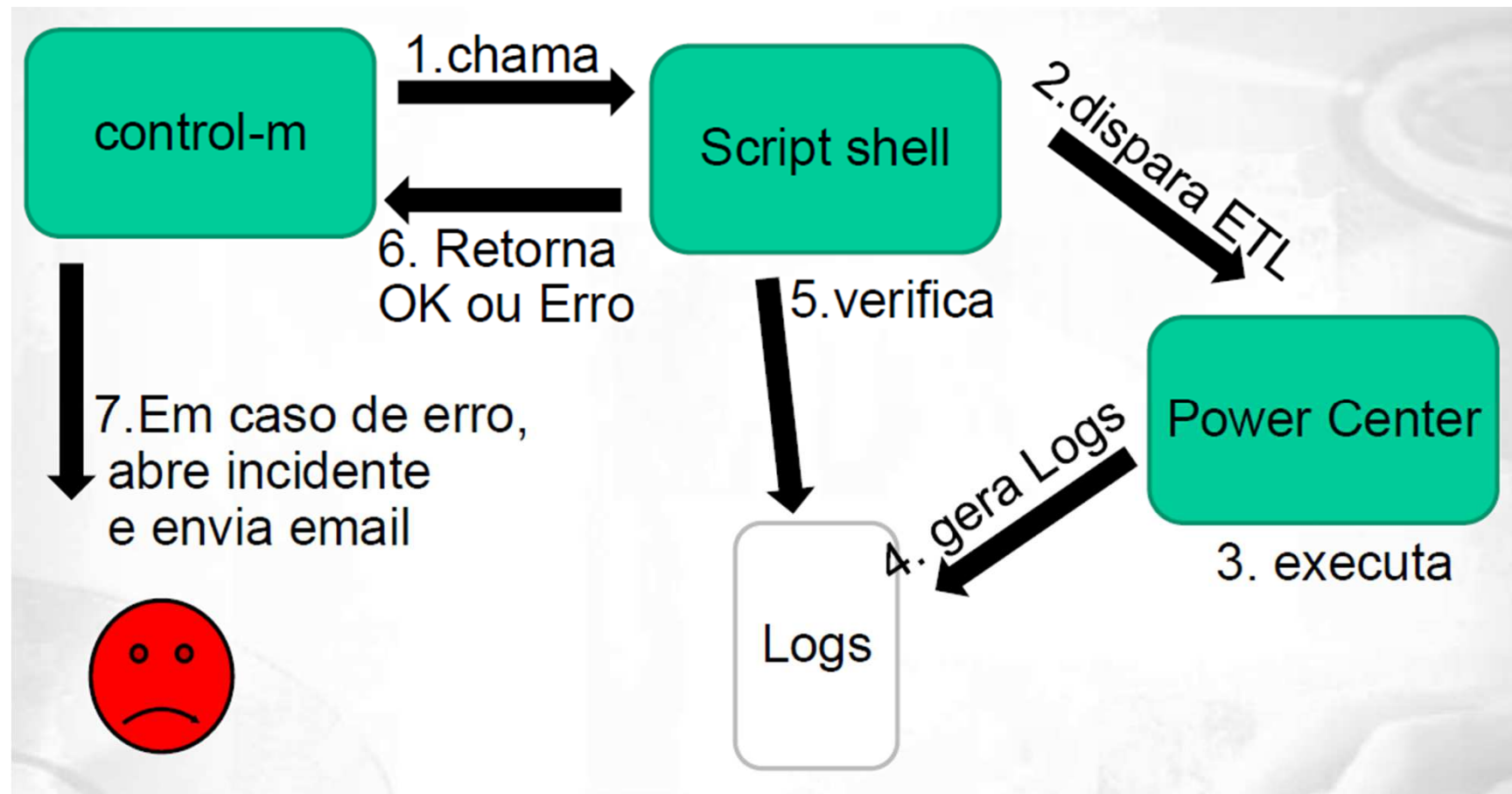
Messages:

- (Integration\_GLOBAL\_DESEN 24/03/2011 10:50:34) Task Update: Start (Succeeded) Start time: 24/03/2011 10:35:13 End time: 24/03/2011 10:35:13
- (Integration\_GLOBAL\_DESEN 24/03/2011 10:50:34) Task Update: s\_MPP\_SMM\_DRL\_2150\_GERA\_LISTA\_v02 (Succeeded) Start time: 24/03/2011 10:35:13 End time: 24/03/2011 10:35:19
- (Integration\_GLOBAL\_DESEN 24/03/2011 10:50:34) Task Update: Decision (Succeeded) Start time: 24/03/2011 10:35:22 End time: 24/03/2011 10:35:22
- (Integration\_GLOBAL\_DESEN 24/03/2011 10:50:34) Task Update: Start (Succeeded) Start time: 24/03/2011 10:35:13 End time: 24/03/2011 10:35:13

Propriedades da Execução

Mensagens

# Fluxo de produção





## Questões

---



# Ementa

---

- Módulo 01 – Soluções de suporte à decisão
  - Inteligência de negócios
    - Business intelligence (BI)
  - Data Warehouses e Data Marts
  - Processo de Data Warehousing
  - Modelagem multidimensional
  - Extração, transformação e carga (ETL)



***ITnerante*** 

**TIMASTERS** 

# Business Intelligence (BI)

---

## Questão 17 - CETRO - 2013 - ANVISA - Analista Administrativo - Área 5 - Prova

---

Quanto aos bancos de dados de suporte à decisão, assinale a alternativa correta.

- (a) São, principalmente (embora não totalmente), apenas de leitura (*read-only*).
- (b) São, principalmente, livres de qualquer tipo de redundância.
- (c) Sua principal preocupação é a integridade.
- (d) Por questões de eficiência no armazenamento, não são indexados.
- (e) Não devem ser muito grandes devido à constante manutenção exigida e às transações comerciais que não devem ser acumuladas com o tempo.

## Questão 18 - ESAF - 2013 - DNIT - Analista Administrativo - Tecnologia da Informação

---

O componente final do processo de *Business Intelligence* é

- (a) *Business balance management* (BBM).
- (b) *Executive office team* (EOT).
- (c) *Business performance management* (BPM).
- (d) *Priority statement board* (PSB).
- (e) *Business advisory management* (BAM).

## Questão 19 - ESAF - 2013 - MF - Analista de Finanças e Controle - Gestão em Infraestrutura de TI

---

Assinale a opção correta relativa a *Business Intelligence (BI)*

- (a) A cultura organizacional deve ser transformada pela alta administração antes das iniciativas de *BI*.
- (b) O *BI* transforma a tomada de decisões em processos mais orientados a processos.
- (c) A cultura organizacional é irrelevante para as iniciativas de *BI*.
- (d) O *BI* transforma a tomada de decisões em processos orientados à inteligência dos dirigentes.
- (e) O *BI* não pode ser simplesmente um exercício técnico da área de Tecnologia da Informação.



## Questão 20 - FCC - 2013 - TRT - 9ª REGIÃO (PR) - Técnico Judiciário - Tecnologia da Informação

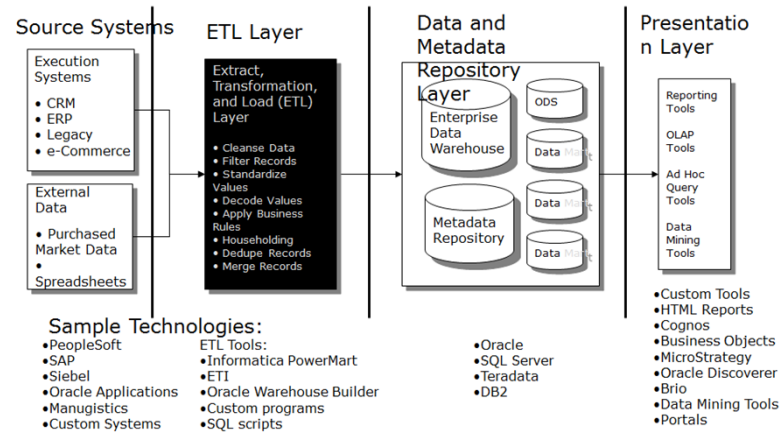
Com o crescente aumento do volume de dados, surge a necessidade de ferramentas e mecanismos que permitam que eles possam ser analisados de forma otimizada, uma vez que armazenam toda a trajetória da empresa. Uma solução é a utilização de **I**, que, em síntese, é utilizado para armazenar conjuntos de dados organizados por assuntos, mantendo todo um histórico corporativo. Outro recurso muito utilizado e dos mais importantes quando o objetivo é a busca de conhecimento, é o **II**, que é um processo que consiste na identificação de informações relevantes que estão presentes em grandes bancos de dados ou repositórios, geralmente realizado em três etapas: a exploração, a definição dos padrões e a validação dos dados. Estas ferramentas e técnicas fazem parte do **III**, definido como um conjunto de métodos e conceitos que podem ser implementados através de softwares com o intuito de utilizar os dados importantes da organização para auxiliar no processo de tomada de decisões, proporcionando melhorias para a alta administração. Outra tecnologia que pode prover uma melhor e mais flexível análise das informações, é o **IV**, que permite uma visão conceitual de forma multidimensional das informações da organização, de maneira que as informações possam ser visualizadas e analisadas de diferentes perspectivas pelo usuário.

(<http://www.devmedia.com.br/mineracao-de-dados-data-warehouse-data-mining-bi-e-olap-atraves-do-fastcube-revista-clubedelphi-146/26537>)

A revista ClubeDelphi publicou o seguinte artigo(acima): As lacunas que completam corretamente o texto estão expressas em

- a) I Data Warehouse      II Data Mart      III CPM      IV Dashboard
- b) I SGBD      II Data Mining      III Data Warehouse      IV OLAP
- c) I OLAP      II Business Intelligence      III Data Mining      IV Dashboard
- d) I Data Warehouse      II Data Mining      III Business Intelligence      IV OLAP
- e) I OLAP      II Data Mart      III Data Warehousing      IV Business Intelligence

# Data Warehousing - Architecture



**ITnerante**

**TIMASTERS**

## Data Warehouses e Data Marts

## Questão 21 - CESGRANRIO - 2011 - Transpetro - Analista de Sistemas Júnior

---

Os conceitos e aplicações de data warehouse e visões de bancos de dados relacionais de sistemas transacionais possuem pontos em comum, e, por conta deles, eventualmente são erroneamente considerados sinônimos. Uma das diferenças entre visões de bancos de dados relacionais e data warehouses reside no fato de que visões relacionais de bancos de dados transacionais

- (a) implementam como uma de suas componentes a dimensão tempo, que nem sempre está presente em data warehouses.
- (b) permitem a combinação dinâmica de seus dados, em diferentes perspectivas, pela aplicação de operações pivoting e roll up.
- (c) podem ser indexadas, ao contrário de data warehouses, que, por sua característica multidimensional, são armazenados em árvores AVL.
- (d) podem, diferentemente de data warehouses, ser implementadas com técnicas top down e bottom up.
- (e) possuem estados com conteúdo dinâmico, já que elas dependem dos estados das relações que referenciam, diferentemente de data warehouses, de natureza não volátil.

## Questão 22 - CESPE - 2012 - MPE-PI - Analista Ministerial - Informática - Cargo 6

---

Acerca de conceitos básicos, arquiteturas e aplicações de data warehouse e datamining e técnicas de modelagem e otimização de bases de dados multidimensionais, julgue os itens que se seguem.

[82] Um cubo de dados é a representação multidimensional dos dados não agregados na qual é necessário que as dimensões tenham o mesmo tamanho.

[84] Um data warehouse é um repositório de informações de uma fonte única, armazenadas sob diversos esquemas, em um só local, e que fornecem dados consolidados

## Questão 23 - CESGRANRIO - 2011 - Petrobrás - Analista de Sistemas Júnior - Engenharia de Software

---

Entre os requisitos que devem ser levantados na construção de um data warehouse se inclui a latência de dados, que descreve a(o)

- (a) rapidez de entrega dos dados ao usuário final
- (b) rapidez da carga inicial do data warehouse
- (c) rapidez da perda de validade do dado
- (d) tempo total de uma consulta
- (e) tempo entre consultas do usuário

## Questão 24 - CESGRANRIO - 2012 - Petrobrás - Analista de Sistemas Júnior - Engenharia de Software

---

Data warehouses são

- (a) tipicamente multidimensionais
- (b) materializados sob demanda
- (c) extratos dos bancos de dados subjacentes
- (d) visões parciais dos bancos de dados subjacentes
- (e) visões normalizadas dos bancos de dados subjacentes

## Questão 25 - CESGRANRIO - 2010 - EPE - Analista de Gestão Corporativa - Tecnologia da Informação

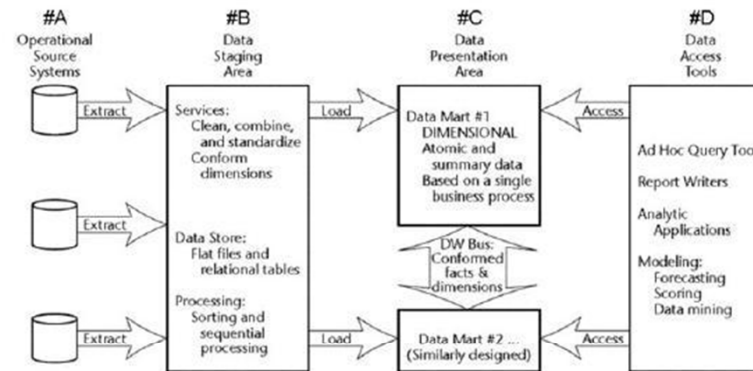
---

Em uma reunião técnica sobre a construção de um *Data Mart* de vendas de uma empresa, discute-se o nível de detalhamento dos dados dentro do banco de dados. Essa característica é conhecida como

- (a) composição.
- (b) integração.
- (c) volatilidade.
- (d) independência.
- (e) granularidade.



# Questão 26 - CESPE - 2008 - TCU - Analista de Controle Interno – Tecnologia da Informação



Ralph Kimball. The datawarehouse toolkit. Wiley, EUA. 2ª ed, 2002.

Ralph Kimball. The datawarehouse toolkit. Wiley, EUA. 2ª ed, 2002.

A figura acima apresenta os elementos básicos da arquitetura de um sistema de suporte a decisão que foi, de forma fictícia, implantado em uma organização de monitoramento, usando-se a tecnologia de SGBDs relacionais. Esse sistema atende ao objetivo de se aprimorarem a gestão e o uso de recursos de TI em mais de 3.000 órgãos públicos federais, não tendo sido consideradas as subunidades que compõem esses órgãos.

O principal negócio dessa organização é o monitoramento e controle da governança de TI desses órgãos, por meio de: aplicação e análise de questionários com questões fechadas (objetivas) e abertas (subjetivas) que, disponíveis via *web*, abordem aspectos relacionados à governança de TI conforme o modelo COBIT; análise de informações oriundas de coletores de dados *online* sobre os níveis de serviços e o emprego de recursos de TI nas organizações; prestação de serviço *online* visando à orientação e à normatização de compras de bens e serviços de TI; auditorias regulares.

Algumas das informações produzidas a cada dia pelos coletores de dados *online* de cada organização monitorada são: quantidade de computadores *desktop* em funcionamento dentro da rede; horas de disponibilidade de cada um dos serviços *online* prestados pela organização (sitio *web*, servidor de *webmail*, FTP, sistemas aplicativos, atendimento a usuários por meio de birô de serviços etc.); despesas e investimentos planejados (empenhados) e executados na aquisição de bens e serviços de TI registrados no sistema de administração financeira. Para bem planejar e desempenhar sua missão, a organização de monitoramento implementou um sistema de apoio à decisão, conforme a arquitetura esboçada na figura acima.



## Questão 26 - CESPE - 2008 - TCU - Analista de Controle Interno – Tecnologia da Informação

---

Julgue o item abaixo

[1] Os dados de controle de processamento da camada #B serão acessíveis a poucos funcionários servidores da organização de monitoramento e controle. A fim de agilizar o processo de extração, transformação e carga dos dados operacionais, estes devem ser armazenados de forma denormalizada na camada #B.

## Questão 27 - CESPE - 2010 - BASA- Técnico Científico - Tecnologia da Informação - Análise de Sistemas

---

Os data warehouses são sistemas computacionais que armazenam os dados corporativos de forma consolidada, disponibilizando informações com maior eficiência, consistência e segurança para a tomada de decisão dos gestores. Acerca dos data warehouses, julgue os itens que se seguem.

[1]Um Data Mart é uma reunião de vários data warehouses, a fim de fornecer visão mais ampla dos dados.

## Questão 28 - FCC - 2009 - TJ-PA - Analista Judiciário - Análise de Sistemas - Desenvolvimento

---

As principais distinções entre um data mart e um data warehouse são as de que um data mart é

- (a) orientado por evento e integrado.
- (b) orientado por assunto e não volátil.
- (c) especializado e orientado a evento.
- (d) integrado e não volátil.
- (e) especializado e volátil.

## Questão 29 - CESPE - Técnico – Programação e Controle de Tecnologia da Informação – Serpro 2013

---

Com relação a gerenciamento eletrônico de documentos (GED) e datawarehouse, julgue os itens que se seguem.

[91] A arquitetura de datawarehouse originou-se da arquitetura de software, que abrange desde a definição da estratégia de negócio até a sua implementação.

[92] Denomina-se de análise dimensional o processo de análise mediante o qual dados originários são selecionados das bases operacionais e modelados no ambiente de datamining, de acordo com as perspectivas ou visões elegidas para a realização.

## Questão 30 - INSTITUTO CIDADES - 2009 - UNIFESP - Analista de Tecnologia da Informação

---

Acerca de Data Warehouse, marque a alternativa correta:

(a) Data Warehouse (ou armazém de dados, ou depósito de dados no Brasil) é um sistema de computação utilizado para armazenar informações relativas às atividades de uma organização em sistemas operacionais, de forma consolidada.

(b) O Data Warehouse possibilita a análise de grandes volumes de dados, coletados dos sistemas transacionais (OLTP).

(c) Por definição, os dados em um Data Warehouse são voláteis, ou seja, mudam. Os dados estão disponíveis para leitura e escrita e podem ser alterados.

(d) As ferramentas OLAP (Online Analytical Processing) tem como função de desfragmentar os dados de um data warehouse, possibilitando uma estrutura adequada tanto para as pesquisas como para a apresentação das informações.

## Questão 31 - INSTITUTO CIDADES - 2009 - UNIFESP - Analista de Tecnologia da Informação

---

Ainda acerca de Data Warehouse, marque a alternativa **INCORRETA**:

- (a) Esse sistema surgiu como conceito acadêmico durante a década de 1950.
- (b) Atualmente, por sua capacidade de sumarizar e analisar grandes volumes de dados, o Data Warehouse é o núcleo dos sistemas de informações gerenciais, e de apoio à decisão das principais soluções de business intelligence do mercado.
- (c) Um repositório de metadados é uma ferramenta essencial para o gerenciamento de um Data Warehouse no momento de converter dados em informações para o negócio.
- (d) O histórico das atualizações normalmente é mantido pelo próprio banco de dados, mas define um elemento de metadado, indicando as datas de atualização dos dados, podendo ajudar o usuário a verificar a atualidade dos dados e a consistência da dimensão tempo do Data Warehouse.

## Questão 32 - CESPE - 2010 - MPU - Analista de Informática - Perito

---

Acerca de *data warehousing*, *data mining* e OLAP (*online analytical processing*), julgue os itens que se seguem.

[1] Dados para a formação de *data warehouses* podem ser provenientes de e-mails, conversas de telefone, documentos, planilhas. A abordagem linguística utiliza a análise de *strings* de caracteres para classificar o texto em determinadas categorias, que são definidas previamente de acordo com o objetivo da empresa.

## Questão 33 - CESPE - 2010 - MPU - Analista de Informática - Perito

---

Acerca de *data warehousing*, *data mining* e OLAP (*online analytical processing*), julgue os itens que se seguem.

[1] Na modelagem de *data warehouses*, o uso de taxonomia facetada como método de descrição multidimensional e agrupamento da informação por meio de seus assuntos ou atributos permite a análise de dados não estruturados em sua forma original.



## Questão 34 - CESGRANRIO - 2010 - ELETROBRÁS - Analista de Sistemas - FUNCIONAL SAP-ERP

---

Os dados armazenados em um data warehouse, que fornecem informações para um ambiente multidimensional para tomada de decisões de uma empresa, não estão normalizados. Essa redundância de dados, criada propositalmente, tem como objetivo

- (a) oferecer segurança na obtenção da informação para o ambiente multidimensional.
- (b) melhorar o desempenho de consultas complexas sobre o data warehouse.
- (c) permitir a distribuição das informações por data marts departamentais.
- (d) facilitar a agregação de dados atômicos.
- (e) relacionar, com maior acuidade, os dados atômicos com os transacionais.



# Modelagem multidimensional

---

## Questão 35 - CESPE - 2010 - TRE-MT - Técnico Judiciário - Programação de Sistemas

---

Com relação aos conceitos de modelagem dimensional, assinale a opção correta.

- (a) Na abordagem utilizada por Kimball, todos os dados de uma empresa são reunidos em um repositório central, caracterizado por um projeto dimensional, e que pode ser consultado diretamente.
- (b) A abordagem de Inmon consolida todas as informações de uma empresa em um repositório central chamado data warehouse corporativo. Essa abordagem é caracterizada por um projeto na terceira forma normal, no qual os dados são consultados diretamente por aplicativos de data warehouse.
- (c) Na abordagem de Kimball, os data marts são adaptados às necessidades e pontos de vista de um grupo de negócios específico, suportando a modelagem dimensional, e são consultados por aplicativos de data warehouse.
- (d) A abordagem de data marts autônomos tem foco empresarial e atende às necessidades dentro de uma área de assunto. Seu desenvolvimento necessita explorar o uso de elementos de dados comuns em toda a empresa.
- (e) Se, em um sistema transacional usando-se um esquema estrela, um cliente, identificado por `Codigo_do_cliente` 1010, modificar o seu endereço da empresa sede, então, para fins analíticos, não será possível acompanhar a história de endereços desse cliente, pois, nessa abordagem, depende-se de `Codigo_do_cliente` para identificar uma linha exclusiva na dimensão cliente, de tal forma que não é possível armazenar várias versões do cliente.

## Questão 36 - CESPE - 2010 - TRT - 21ª Região (RN) - Analista Judiciário - Tecnologia da Informação

---

Acerca de sistemas de suporte a decisão e data warehousing, julgue os itens a seguir.

[70] Em um data warehouse, cada linha em uma tabela fato corresponde a uma medida, representada por um valor aditivo, em que necessariamente essas medidas não compartilham a mesma granularidade.

[71] Em um modelo do tipo estrela (star schema), devido à ligação entre as tabelas dimensionais e suas respectivas fontes de dados, as dimensões são dependentes de códigos operacionais de produção. Desse modo, nessas tabelas, convencionou-se usar como chave primária as mesmas utilizadas no ambiente de produção — origem dos dados.

## Questão 37 - CESPE - 2013 - INPI - Analista de Planejamento – Desv. e Manutenção de Sistemas

---

Acerca de normalização de dados, julgue o item seguinte.

[1] Ao se realizar uma modelagem para Data Warehouse, a forma normalizada Snow Flake permite que as dimensões sejam modeladas na 3FN (terceira forma normal). Dessa forma, todo modelo Snow Flake na 3FN estará necessariamente também na forma normal BCNF (*boyce-codd*), visto que ambas são baseadas no conceito de dependência transitiva.

## Questão 38 - CESGRANRIO - 2010 - ELETROBRÁS - Analista de Sistemas - FUNCIONAL SAP-ERP

---

Um data mart do departamento de vendas de uma grande empresa possui uma tabela fato com cinco métricas associadas a oito diferentes dimensões. O analista de TI está em dúvida sobre que modelagem utilizar. Considerando-se as opções snowflake e star schema, qual a desvantagem de se utilizar, nessa situação, a modelagem snowflake?

- (a) A modelagem star schema não contempla a possibilidade de armazenamento de atributos para cada dimensão, minimizando a quantidade de memória utilizada.
- (b) Uma maior quantidade de joins entre as tabelas de dimensões e as tabelas de hierarquia associadas será realizada, ocasionando queda de desempenho no processo de agregação das métricas.
- (c) As chaves estrangeiras, relativas às dimensões armazenadas na tabela fato, devem corresponder às hierarquias existentes nessas dimensões, ocasionando maior quantidade de registros na tabela fato.
- (d) Na modelagem snowflake, a manutenção dos dados dimensionais se torna mais dispendiosa, tendo em vista que as tabelas dimensionais não estão normalizadas.
- (e) Diferentemente da modelagem star schema, na modelagem snowflake, as chaves estrangeiras na tabela fato não apontam para o nível mais baixo (atômico) da hierarquia das dimensões associadas.

## Questão 39 - FUNCAB - 2010 - PRODAM-AM - Analista de Tec. da Informação - Banco de Dados

---

Sejam as seguintes assertivas sobre esquemas multidimensionais para a modelagem de dados para *Data Warehouses*:

- I. O esquema **snowflake** é uma variação do esquema **estrela** no qual as tabelas dimensionais de um esquema **estrela** são organizadas em uma lista de tabelas não normalizadas.
- II. O esquema **estrela** consiste em uma tabela de fato com uma única tabela para cada dimensão.
- III. O esquema onde um conjunto de tabelas de fatos compartilha algumas tabelas de dimensão é conhecido por **hipercubo**.

Marque a alternativa correta em relação às assertivas acima.

- (a) Apenas a assertiva I é verdadeira.
- (b) Apenas a assertiva II é verdadeira.
- (c) Apenas a assertiva III é verdadeira.
- (d) Todas as assertivas são verdadeiras.
- (e) Todas as assertivas são falsas.

## Questão 40 - FEPESE - 2010 - SEFAZ-SC - Auditor Fiscal da Receita Estadual - Tecnologia da Informação

---

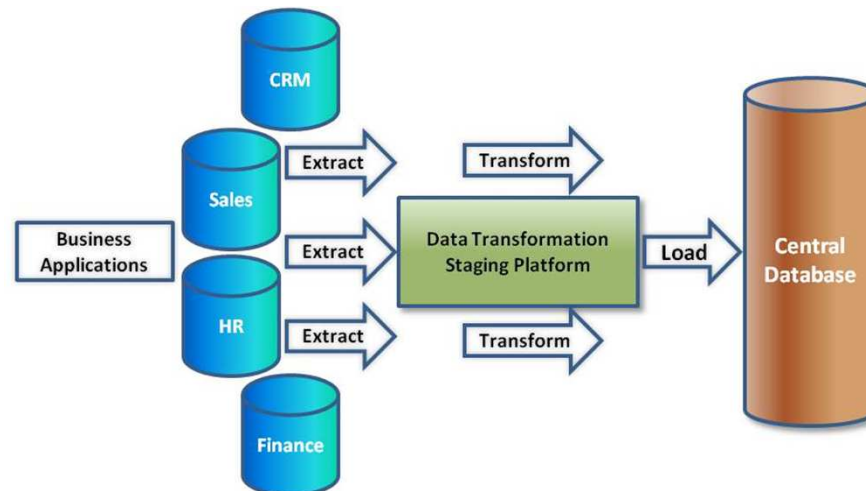
Assinale a alternativa **correta** a respeito de *Data Warehouse* (DW) e modelagem multidimensional.

- (a) Um esquema multidimensional, composto por fatos e dimensões, não pode ser modelado em um banco de dados relacional.
- (b) Uma dimensão definida em uma modelagem multidimensional deve estar associada a um único fato.
- (c) Operações OLAP podem ser executadas tanto para fins analíticos quanto para fins de atualização de dados.
- (d) O volume de dados de um DW geralmente é superior ao volume de dados de um BD transacional.
- (e) Um *Data Mart* é um componente de um *DW* que não possui dados históricos.



# Extração, transformação e carga (ETL)

---



## Questão 41 - CESPE - 2012 - MPE-PI - Analista Ministerial - Informática - Cargo 6

---

Acerca de fundamentos de administração de dados, fundamentos de banco de dados e conceitos envolvidos em data warehouse, julgue os itens seguintes.

[65] Na construção de um data warehouse, o processo extract, transform and load (ETL), normalmente, é o menos crítico e o mais rápido. Esse processo envolve a extração dos dados, na sua transformação, e, eventualmente, influencia na limpeza desses dados.

[66] Em modelo de banco de dados, a normalização estabelece um conjunto de regras que tem por objetivo maximizar as anomalias de modificação dos dados e aumentar a flexibilidade em sua utilização.

## Questão 42 - CESGRANRIO - 2010 - ELETROBRÁS - Analista de Sistemas Júnior - Engenharia de Software

---

Os dados transacionais de uma grande empresa estão dispostos em bases heterogêneas. Para que dados carregados no data warehouse sejam considerados adequados à formação de tabelas fato e dimensionais, a etapa de transformação pode realizar procedimentos de limpeza nesses dados. Que problema **NÃO** é resolvido por esse tipo de procedimento?

- (a) Existência de valores não previstos.
- (b) Filtragem de dados a serem carregados.
- (c) Ausência de dados.
- (d) Inconsistência de dados.
- (e) Duplicação de dados.

## Questão 43 - FCC - 2012 - TST - Analista Judiciário - Análise de Sistemas

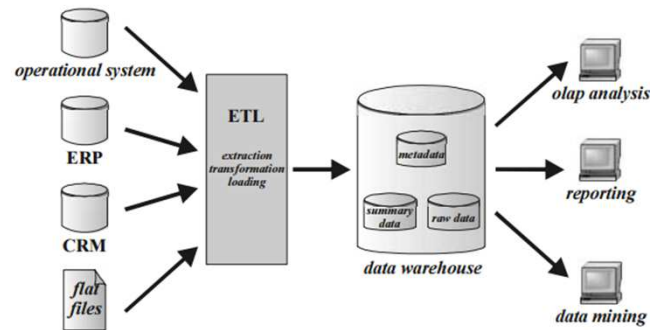
---

O processo de ETL em uma Data Warehouse possui várias fases. Em uma destas fases é efetuada a

- (a) extração dos dados dos sistemas de origem.
- (b) introdução de novos produtos no mercado.
- (c) validação das interfaces de usuário.
- (d) criação de diagramas estáticos e comportamentais das classes e atributos.
- (e) definição dos custos e prazos.

## Questão 44 - CESPE/UnB – Banco da Amazônia S.A. - 2012

---



Internet: <[www.infoescola.com/informatica/data-warehouse](http://www.infoescola.com/informatica/data-warehouse)>.

Considerando a figura acima, que representa um esquema básico do processo de desenvolvimento de um armazém de dados (data warehouse), julgue os itens que se seguem.

[110] O processo de extração, transformação e carga (ETL) ilustrado permite extrair dados de diversas fontes de dados e migrá-los para o data warehouse, mantendo sua estrutura e normalização originais.

[111] Os modelos de dados usados no data warehouse são frequentemente constituídos por modelos dimensionais, em que um mesmo dado pode ser fisicamente gravado mais de uma vez. Entre as razões de esses modelos serem utilizados, destaca-se o aumento do desempenho em consultas a grandes volumes de dados associados a realizações de uma dimensão do modelo.

## Questão 45 - CESPE - 2011 - Correios - Analista de Correios - Analista de Sistemas – Desen. de Sistemas

---

Acerca de soluções de suporte a decisão, julgue os itens seguintes.

[1] As ferramentas de software ETL (extract transform load) têm como função a extração de dados de diversos sistemas, a transformação desses dados de acordo com as regras de negócio e a carga dos dados em um data mart ou um DW.

[2] Em um ambiente data warehouse (DW), é possível a análise de grandes volumes de dados, os quais ficam disponíveis para serem alterados e manipulados pelo usuário.

## Questão 46 - CESGRANRIO - 2009 - Casa da Moeda - Desenvolvimento de Sistemas

---

Uma empresa deseja criar uma base de dados para apoio à tomada de decisão. As informações corporativas encontram-se em 4 bancos de dados distintos e possuem representações diferentes para atributos de uma pessoa física. Por exemplo, em determinado banco de dados, uma pessoa física do sexo masculino é representada por um atributo com valor M, enquanto que em outro banco de dados, com valor 1. Em que etapa de um procedimento de ETL para essa base de dados deve ocorrer uma padronização desse tipo de valor?

- (a) Extração.
- (b) Deleção.
- (c) Transformação.
- (d) Paralelismo.
- (e) Carga.

## Questão 47 - CESGRANRIO - 2009 - Casa da Moeda Banco de Dados

---

Observe as fontes de dados a seguir.

I - Tabela de um banco de dados relacional.

II - Arquivo XML.

III - Arquivo TXT flat.

Utilizando procedimentos de ETL (*Extract, Transform e Load*) para carga de um DW, é possível ler dados das fontes

(a) I, apenas.

(b) II, apenas.

(c) III, apenas.

(d) I e III, apenas.

(e) I, II e III.



# Perguntas

---



# Muito obrigado!

---

A close-up photograph of a hand holding a black pen, having just finished writing the words "THANK YOU!" in a bold, cursive, hand-drawn style on a white surface. The pen is positioned at the bottom right of the text.



Toda a educação, no momento, não parece motivo de alegria, mas de tristeza. Depois, no entanto, produz naqueles que assim foram exercitados um fruto de paz e de justiça.

Prof. Thiago Cavalcanti

**Hebreus (12:11)**

# Banco de dados

## Suporte a decisão – Business Intelligence

---

**Curso Preparatório - ITnerante**

*Prof. Thiago Cavalcanti*



# Gabarito

---

- |                    |                   |                     |
|--------------------|-------------------|---------------------|
| 1. E               | 17. A             | 33. [1] C           |
| 2. C               | 18. C             | 34. B               |
| 3. D               | 19. E             | 35. A               |
| 4. E               | 20. D             | 36. [70] F [71] F   |
| 5. A               | 21. E             | 37. [1] F           |
| 6. D               | 22. [82] C [84] C | 38. B               |
| 7. D               | 23. A             | 39. B               |
| 8. A               | 24. A             | 40. D               |
| 9. [135] F [136] F | 25. E             | 41. [65] F [66] F   |
| 10. C              | 26. [1] Certo     | 42. B               |
| 11. C              | 27. Falso         | 43. A               |
| 12. [134] C        | 28. E             | 44. [110] F [111] C |
| 13. [1] F          | 29. [91] E [92] C | 45. [1] C [2] F     |
| 14. C              | 30. B             | 46. C               |
| 15. E              | 31. A             | 47. E               |
| 16. C              | 32. [1] F         |                     |