

## Data Warehouse (DW)

Inteligência de Negócios  
Fatec Indaiatuba  
Prof. Piva

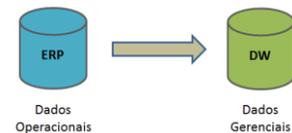
### Objetivos desse Encontro...

- ▶ Compreender as definições e conceitos básicos do Data Warehouse (DW)
- ▶ Entender as arquiteturas do DW
- ▶ Descrever os processos utilizados no desenvolvimento e gerenciamento do DW
- ▶ Explicar o papel do DW no apoio às decisões.
- ▶ Explicar a integração de dados e o processo de extração, transformação e carga (ETL)
- ▶ Descrever a necessidade de um DW em tempo real
- ▶ Compreender a administração do DW e as questões de segurança.

## INTRODUÇÃO

### Conceito

- ▶ Um Data Warehouse (armazém de dados) é um repositório físico onde os dados transacionais são organizados, consolidados e agrupados para apoiar o processo de tomada de decisão a partir de uma visão única dos fatos no tempo.



### Conceito

- ▶ Questões que devem ser respondidas pelo DW:
  - Nós possuímos montanhas de dados, no entanto, não conseguimos acessá-los
  - Nós precisamos visualizar os dados de todas as maneiras possíveis
  - É necessário facilitar o acesso às informações para os usuários de negócio.
  - Apenas mostre-me o que é importante.
  - Usar a informação para tomada da decisão
  - Duas pessoas apresentam a mesma métrica em uma reunião com diferentes números.

### Características do DW

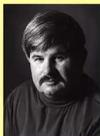
- ▶ Orientado por assunto
- ▶ Integrado aos dados transacionais
- ▶ Baseado em séries temporais
- ▶ Não volátil
- ▶ Escalável
- ▶ Multidimensional
- ▶ Padrão cliente/servidor
- ▶ Foco gerencial.

### Requisitos do DW

- ▶ Tornar a informação facilmente acessível
- ▶ Apresentar informações consistentes
  - Alta qualidade / confiáveis / íntegros
- ▶ Adaptável e flexível à mudanças
  - Necessidades dos negócios
- ▶ Proteger e tornar a informação segura
- ▶ Auxiliar no processo de tomada de decisão
- ▶ Ser aceito pela comunidade de negócios
  - Trazer os indicadores necessários
  - Desde o início, envolver todos os envolvidos.

### Características do DW

**Bill Inmon**



- Born 1945
- BSc from Yale
- MSc from New Mexico Uni.

**Ralph Kimball**



- Born 1944
- PhD from Stanford
- Commercialised DWs.

### Características do DW

Características	Inmon	Kimball
Arquitetura	De cima para baixo	De baixo para cima
Estrutura	DW para toda a empresa (atômico) alimenta os BD departamentais	Os Data Marts modelam um único processo de negócio Barramento de dados e ajuste dimensional
Complexidade	Complexo	Simple
Metodologia de desenvolvimento	Espiral	SGBDR
Projeto Físico	Detalhado	Simple /leve
Modelagem de Dados	Orientado por assunto ou dados	Orientado por processos
Acessibilidade do Usuário final	Baixa	Alta
Publico principal	Profissionais de TI	Usuários finais

### Características do DW

- ▶ Geralmente os Data Warehouse começam “departamentais” ou em “menor escala”, também denominados Data Marts
  - Data Marts Independentes
    - Um subconjunto de dados criados de um DW
  - Data Marts Dependentes
    - Um DW menor projetado por uma unidade estratégica de negócios ou departamento.

### Características do DW

## Qual a diferença entre os dados em Sistemas Transacionais ou ERPs e os Data Warehouse?

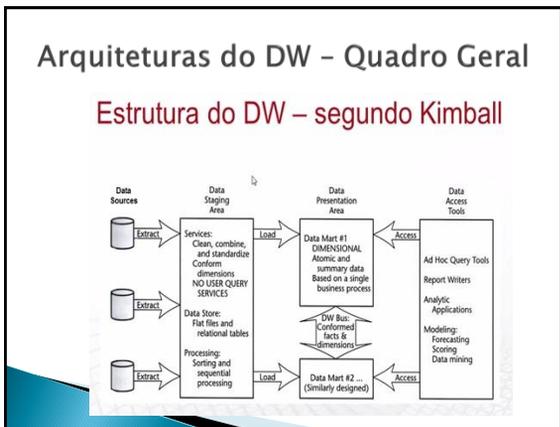
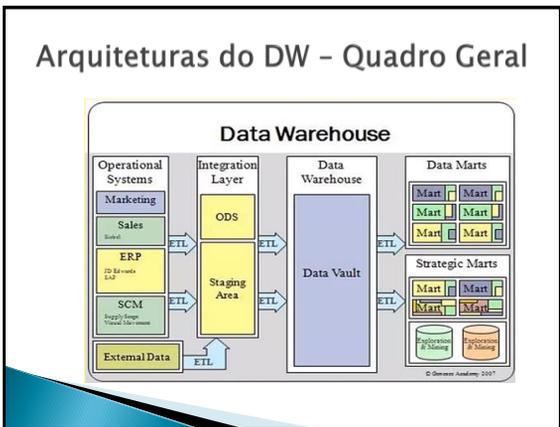
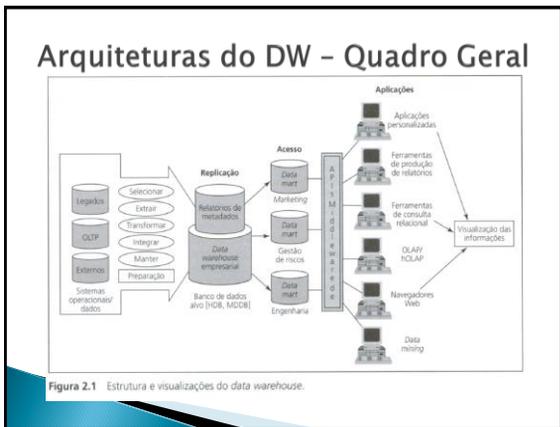
### Características do DW

**DW não estão em 3FN (Big Data)**

- ▶ Os dados devem ser “des”-normalizados para 2FN
- ▶ Significa que você obterá redundância de dados
- ▶ Significa que você precisará de mais local de armazenamento...
- ▶ ... Mas você poderá recuperar os dados mais rapidamente.
- ▶ O propósito do DW é prover dados agregados que com um formato adequado, facilitem o processo de decisão

# ARQUITETURAS

- ## Arquiteturas do DW
- ▶ Organizações coletam continuamente dados, informações e conhecimentos num ritmo cada vez mais acelerado e precisam armazená-los em sistemas informatizados
  - ▶ O número de usuários que precisam acessar estas informações continua crescendo e necessitam de maior confiabilidade e disponibilidade para a tomada de decisão.



- ## Arquiteturas do DW
- ▶ Por que trabalhar com DATA MARTS?

## Arquiteturas do DW

- ▶ **Por que trabalhar com DATA MARTS?**
- ▶ Não bagunce meus dados
- ▶ Torna-os simples para o usuário
- ▶ Problemas menores são mais fáceis de resolver!

## Arquiteturas do DW – Visão Geral

- ▶ Os principais componentes de um processo de armazenamento de dados são:
  - Fonte de dados
  - Extração de dados
  - Tratamento dos dados
  - Metadados
  - Carga dos dados
  - Ferramentas de middleware (acesso ao DW / consultas)

## Arquiteturas do DW

- ▶ Um DW é composto por três partes:
  - Servidor de dados (software + dados) ou o próprio DW
  - Software de aquisição de dados (servidor de aplicação)
  - Sistema gerenciador de dados (interface usuário) *Manutenção, ETL e análise.*

## Arquiteturas do DW

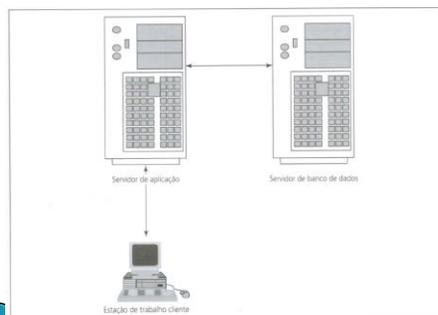


Figura 2.2 Arquitetura de um data warehouse de três camadas.

## Arquiteturas do DW

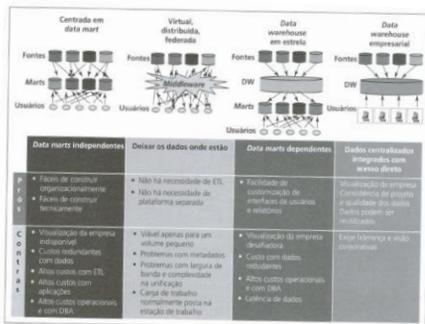


Figura 2.6 Arquiteturas alternativas para empreendimentos de data warehousing.

## Arquiteturas Web do DW

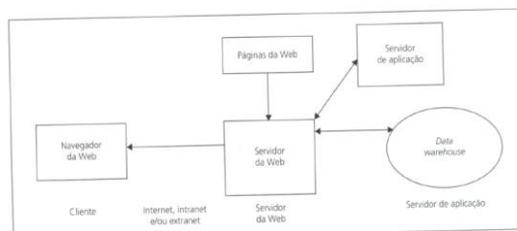
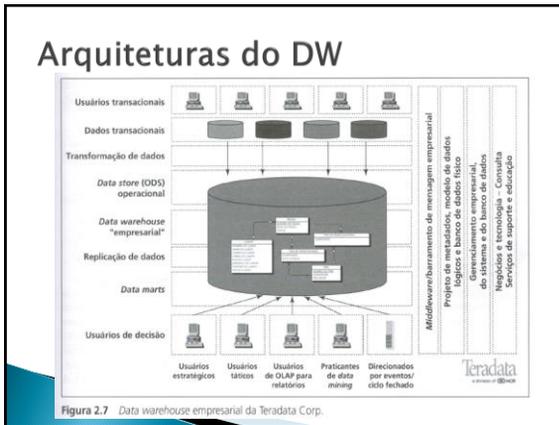


Figura 2.4 Arquitetura de um data warehouse baseado na Web.



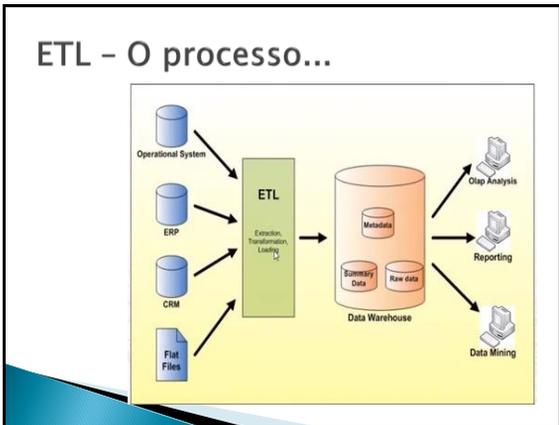
- ### Arquiteturas do DW
- ▶ Questões a considerar na escolha da arquitetura técnica:
    - Qual SGBD usar?
    - Processamento local, paralelo, particionado ou nuvem
    - Ferramentas de manipulação (ETL)
    - Ferramentas analíticas (BA)
    - Infraestrutura de rede

- ### Arquiteturas do DW
- ▶ Questões a considerar na escolha da gestão de processos:
    - Interdependências dos dados entre unidades organizacionais
    - Necessidade de informação para gestão estratégica
    - Natureza das tarefas dos usuários finais
    - Restrição de recursos financeiros
    - Questões técnicas (Compatibilidade dos sistemas existentes e capacidade da equipe técnica)
    - Fatores sociais e políticos.

### EXTRAÇÃO, TRANSFORMAÇÃO E CARGA

ETL – Extration, Transformation and Loading

- ### ETL – Conceito...
- ▶ É um processo de armazenamento de dados no DW que consiste da:
    - **Extração**  
leitura dos dados a partir de um banco de dados
    - **Transformação**  
conversão dos dados extraídos a partir de sua forma anterior para um formato novo para ser armazenado no DW
    - **Carga**  
colocar os dados no DW.



## ETL - o processo...

- ▶ O processo de integração do DW com os sistemas transacionais compreende três etapas:
  - Acesso aos dados para extração
  - Estruturação dos dados e transformação
  - Automação de carga no DW.

## ETL - o processo...

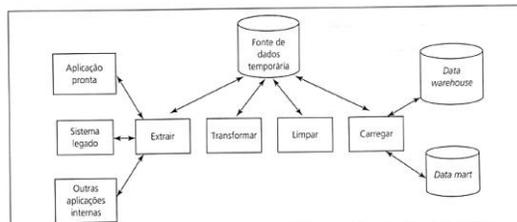
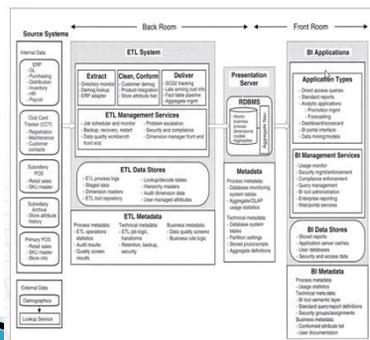
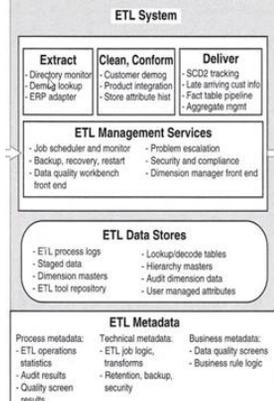


Figura 2.8 O processo de ETL.

## ETL - o processo...



## ETL o processo...



## ETL - escolha da ferramenta

- ▶ Critérios importantes na escolha de uma ferramenta de ETL:
  - Capacidade de ler e escrever para um número ilimitado de arquiteturas e fonte de dados
  - Geração de scripts e automação
  - Conformidade com os padrões abertos
  - Interface simples e fácil para o potencializar a curva de aprendizado

## BENEFÍCIOS DO DW

### Benefícios Diretos do DW

- ▶ Permite que os usuários finais executam várias análises
- ▶ Permite uma visão consolidada dos dados corporativos
- ▶ Simplificação de acesso aos dados
- ▶ Necessidade de integração das unidades informação e sistemas legados.

### Benefícios Indiretos do DW

- ▶ Melhor conhecimento do negócio
- ▶ Criação de uma linguagem comum
- ▶ Agilidade no processo decisório
- ▶ Oportunidade de revisão dos processos de negócio
- ▶ Cultura de dados (qualidade)
- ▶ Redução dos conflitos analíticos
- ▶ Vantagem competitiva.

### Fatores Críticos de Sucesso...

- ▶ Fatores importantes para a seleção de fornecedores de DW:
  - Experiência na indústria
  - Consultores qualificados
  - Capacidade financeira
  - Participação no mercado
  - Experiências anteriores
  - Reputação.

### Fatores Críticos de Sucesso...

- ▶ Problemas na implantação de um DW:
  - Implementação de um DW é geralmente um enorme esforço que deve ser planejado e executado de acordo com os métodos estabelecidos pelo projeto
  - Há muitas facetas para o ciclo de vida do projeto e nenhuma pessoa será um especialista em todas as áreas.

### Fatores Críticos de Sucesso...



Figura 2.10 A evolução das decisões corporativas.

## MELHORES PRÁTICAS DO DW

## Desenvolvimento...

- ▶ Podemos considerar duas abordagens para o desenvolvimento e implantação de um DW:
  - Modelo de Inmon abordagem corporativa (top-down)
  - Modelo de Kimball abordagem Data Mart (bottom-up)
- ▶ Não existe a melhor estratégia...

## Desenvolvimento...

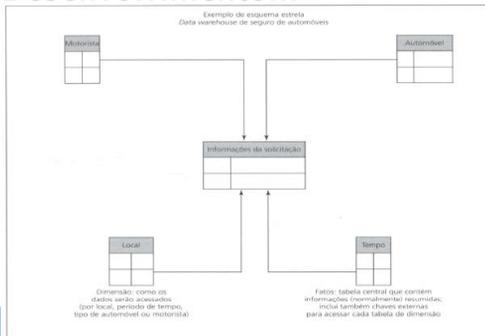


Figura 2.9 Esquema estrela.

## Desenvolvimento...

- ▶ Considerações sobre o desenho do DW:
  - **Granularidade**  
definição do mais alto nível de detalhe que será suportado pelo DW
  - **Navegação**  
filtros e dimensões para cada necessidade de análise do BI
  - **Segurança**  
nível de acesso aos dados e backup.

## Desenvolvimento...

- ▶ Considerações sobre o desempenho do DW:
  - Nível de serviço acordado (SLA)
  - Identificação das fontes de dados e suas políticas de governança
  - Qualidade dos dados
  - Desenho do modelo lógico
  - Ferramentas de ETL.

## Desenvolvimento...

- ▶ Considerações sobre a plataforma do DW:
  - Banco de dados relacional
  - Rede e servidores (infraestrutura)
  - Comunicação e reconciliação
  - Planejamento de backup
  - Suporte técnico e especificidades.

## Melhores Práticas...

- ▶ Sustentabilidade de um DW:
  - O projeto deve atender a estratégia corporativa e os objetivo de negócio
  - Os executivos, gestores e usuários devem estar envolvidos no projeto
  - Gerenciar as expectativas e resultados
  - Deve ser construído de forma incremental
  - Selecionar corretamente a equipe de TI
  - Cuidado com o pessoal de banco de dados.

### Melhores Práticas...

- ▶ Sustentabilidade de um DW, cont.:
  - Estabelecer bom relacionamento com os fornecedores e parceiros
  - Cuidado na manipulação e carga
  - Respeitar os requisitos
  - Identificar um patrocinador
  - Pensar no desempenho, capacidade e escalabilidade (Big Data)
  - Considerar a mineração de dados
  - Ficar politicamente alerta.

### Melhores Práticas...

- ▶ Os projetos de DW fracassam quando:
  - Questões culturais são ignoradas
  - Arquitetura imprópria
  - Objetivos poucos claros do negócio
  - Falta de informação
  - Expectativas irrealistas
  - Baixo nível de sumarização dos dados
  - Baixa qualidade dos dados
  - Estrutura relacional confusa.

### Melhores Práticas...

- ▶ Segurança de dados num DW deve se concentrar em quatro áreas principais:
  - Políticas corporativas eficazes e de procedimentos de segurança
  - Segurança lógica e física visando restringir o acesso aos dados
  - Limitar o acesso físico ao ambiente
  - Relatórios de acesso e controle interno de privacidade.