

PÓS-GRADUAÇÃO

Integração e fluxo de dados



PÓS-GRADUAÇÃO

Introdução à Extração, Transformação e Carga (ETL)

Bloco 1

Thiago Salhab Alves





► Introdução à Extração, Transformação e Carga (ETL)

Objetivos

- Compreender as definições e conceitos básicos dos *data warehouses*.
- Introduzir os processos de extração, transformação e carga (ETL – *Extract Transform Load*).
- Aprender sobre arquiteturas de *data warehousing*.



► *Data Warehouse*

Data Warehouse

- Por mais de trinta anos, são utilizadas aplicações de bancos de dados, com alto investimento em aplicativos, mas que são de difícil uso para fins analíticos.
- *Data Warehouse* é um conjunto de dados utilizado no suporte de tomada de decisões, sendo um repositório de dados atuais e históricos, orientado por assunto, integrado, variável no tempo e não-volátil.



► *Data Warehouse*

- *Data warehouse* tem por objetivo publicar dados da organização com suporte mais efetivo para a tomada de decisão.
- Os custos do *data warehouse* são táticos, mas seus benefícios para o suporte de tomada de decisão são estratégicos.
- Os dados estruturados são utilizados em atividades de processamento analítico, por exemplo, processamento analítico on-line (OLAP), *data mining*, sistemas de BI, consultas e geração de relatórios.




► *Data Warehouse*

- Um *data warehouse* é um sistema que extrai, limpa, conforma e entrega fonte de dados em armazenamento de dados dimensional e implementa consultas e análises para fins de tomada de decisão. Tipos de *data warehouse*:
- *Data mart*: é um subconjunto de um *data warehouse* voltado para uma determinada área, tais como, marketing ou operações.

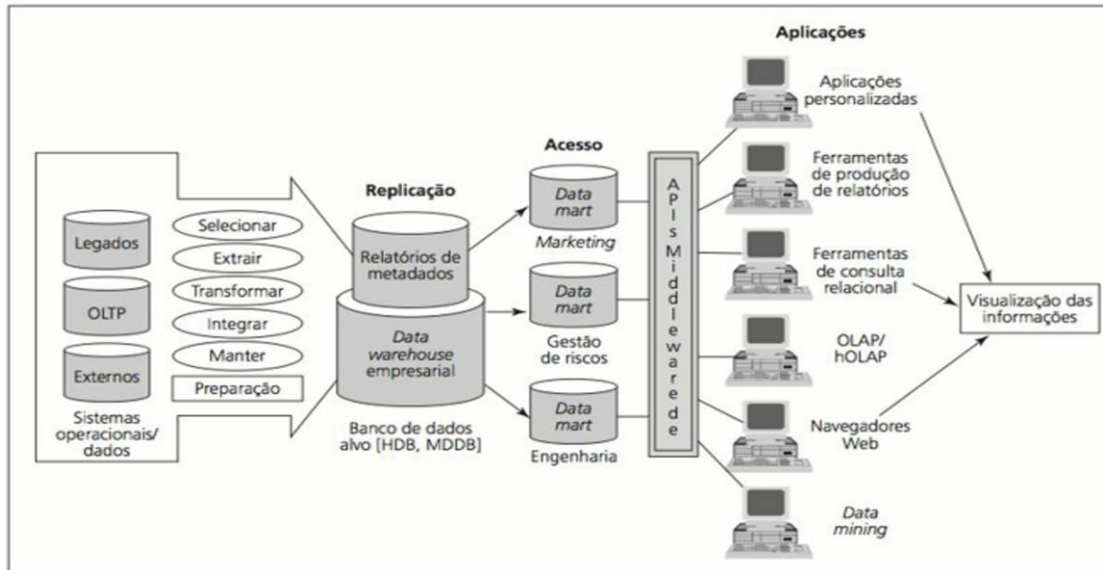


► *Data Warehouse*

- *Data store* operacional: arquiva informações recentes para uso, podendo ser atualizado durante as operações.
 - *Data warehouse* empresarial: *data warehouse* de grande escala e usado por toda a empresa no suporte à decisão.
- 

► Data Warehouse

Figura 1 – Conceito de *Data Warehouse*



Fonte: Turban et al. (2009, p. 61).



► Extração, transformação e carga (ETL)

Extração, transformação e carga (ETL)

- De acordo com Turban et al. (2009), a parte fundamental do processo de *data warehouse* é a extração, transformação e carga (ETL).
- Extração: leitura dos dados de um ou mais bancos de dados.
- Transformação: conversão dos dados extraídos para formato a ser usado no *data warehouse*.
- Carga: inserção dos dados no *data warehouse*.

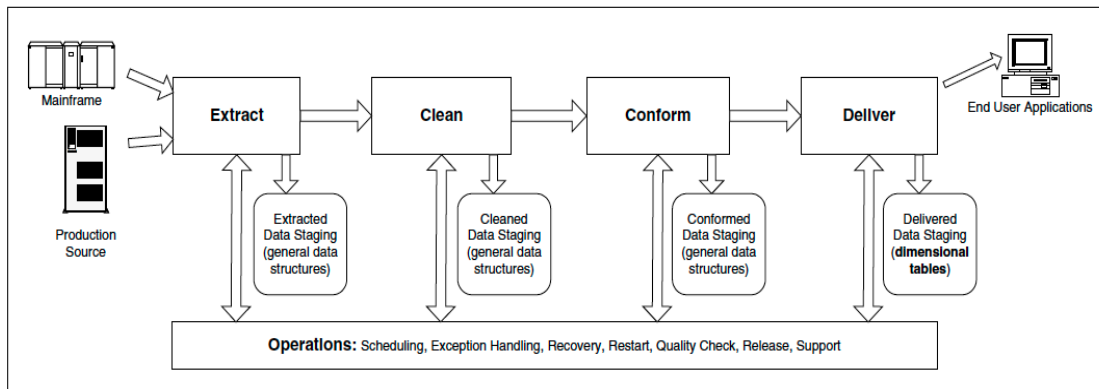


► Extração, transformação e carga (ETL)

- ETL é extremamente importante na integração de dados, e tem por objetivo carregar dados integrados e limpos no *data warehouse* – dados que podem vir de qualquer fonte.

► Extração, transformação e carga (ETL)

Figura 2 – Processos de ETL



Fonte: Kimball e Caserta (2009, p. 18).

PÓS-GRADUAÇÃO

Introdução à Extração, Transformação e Carga (ETL)

Bloco 2

Thiago Salhab Alves





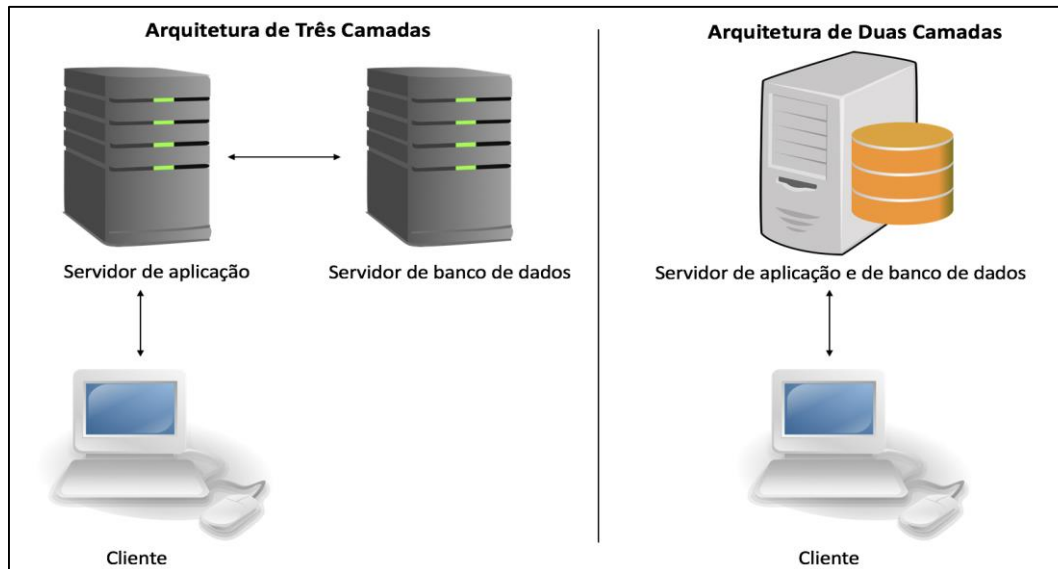
► Extração, transformação e carga (ETL)

- Segundo Turban et al. (2009), o processo de carregar dados para um *data warehouse* pode ser realizado por meio de ferramentas de transformação de dados. As ferramentas devem permitir:
 - Ler e gravar um número ilimitado de arquiteturas de fontes de dados.
 - Captura e entrega automática de metadados.
 - Histórico de conformidade com padrões abertos.
 - Interface fácil de usar para o desenvolvedor e usuário final.

► Extração, transformação e carga (ETL)

- Arquiteturas de *Data Warehouse*

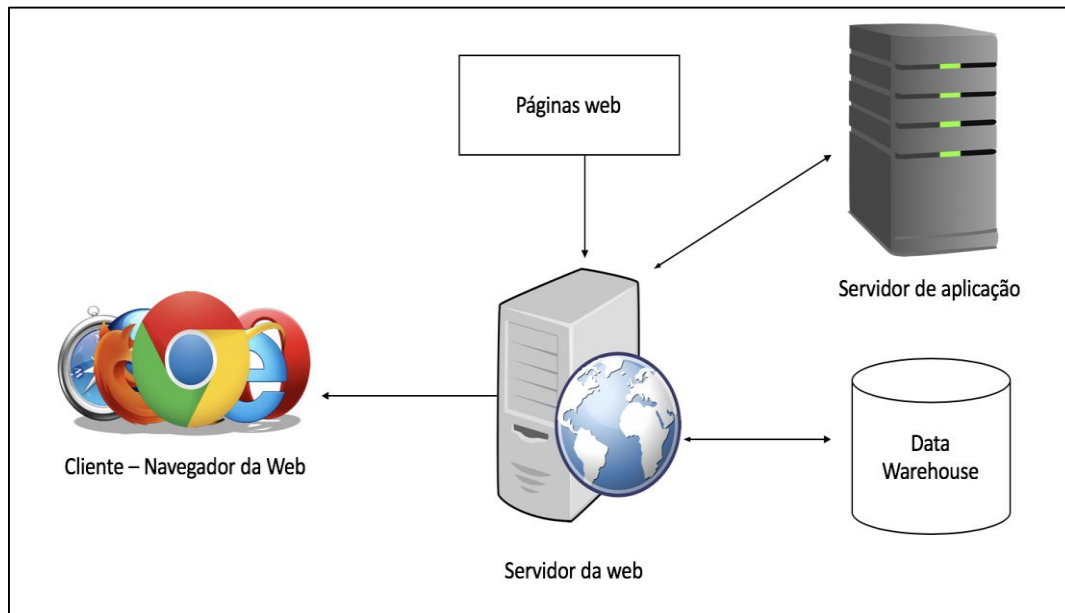
Figura 3 – arquitetura *Data Warehouse* com duas e três camadas



Fonte: elaborada pelo autor.

► Extração, transformação e carga (ETL)

Figura 4 – Arquitetura *Data Warehouse* baseado na web



Fonte: elaborada pelo autor.



► Extração, transformação e carga (ETL)

- Qual sistema de gerenciamento de banco de dados (SGBD) usar?
- A maioria dos *data warehouses* usam SGBD relacional, tais como Oracle, SQL Server e DB2, que são produtos que suportam arquiteturas cliente/servidor e baseado na web.
- O processamento será paralelo e/ou os dados serão particionados?
- Processamento paralelo permite que múltiplas CPUs processem solicitações de consultas ao *data warehouse*, simultaneamente, e oferece escalabilidade.

PÓS-GRADUAÇÃO

Teoria em Prática

Bloco 3


Thiago Salhab Alves





► Teoria em Prática

Uma empresa nacional de revenda de cosméticos está enfrentando alguns problemas financeiros. Dado o grande volume de produtos lançados pelo setor de cosméticos, a empresa está tendo dificuldades em acompanhar a demanda por produtos de lançamento, o que muitas vezes acaba por comprometer o resultado financeiro por investir em produtos com baixa procura. Outro problema são os produtos que possuem prazo de validade curto, que acabam por vencer e não podem ser trocados pelos fornecedores.





► Teoria em Prática

- Hoje, a empresa conta com um sistema de vendas e controle de estoque, com banco de dados relacional e um processo de marketing pelas redes sociais. Porém, está com dificuldades para a tomada de decisões relacionada ao que o seu público-alvo realmente consome, a fim de evitar gastos desnecessários.
- Como podemos auxiliar a organização a reverter essa situação?



► Teoria em Prática

- R: Uma das possibilidades de solução é criar um *data warehouse*, utilizando o processo de ETL, que extraia, transforme e carregue os dados vindos da base de dados do sistema já existente, utilizando uma arquitetura de duas camadas. Existem outras possibilidades inclusive sem o desenvolvimento de software.

Dica do Professor

Bloco 4

Thiago Salhab Alves





► Dica do Professor

Indicação de leitura: capítulo 2, *Data Warehousing*, do livro de Turban et al. (2009), disponível na Biblioteca Virtual:

- TURBAN, Efrain et al. **Business Intelligence: Um Enfoque Gerencial para a Inteligência do Negócio**. Porto Alegre: Bookman, 2009.



► Referências

KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL Toolkit**: Practical Techniques for Extracting, Cleaning, Conforming, and Data Delivering Data. Indianapolis: Wiley Publishing, 2004.

TURBAN, Efrain et al. **Business Intelligence**: Um Enfoque Gerencial para a Inteligência do Negócio. Porto Alegre: Bookman, 2009.

