

**PÓS-GRADUAÇÃO**

**Integração e fluxo  
de dados**



PÓS-GRADUAÇÃO

# Sistemas ETL de Tempo Real

**Bloco 1**

Thiago Salhab Alves





## ► Sistemas ETL de Tempo Real

### Objetivos

- Compreender o que são sistemas ETL de Tempo Real.
- Conhecer as aplicações de ETL de Tempo Real.
- Conhecer as abordagens ETL de Tempo Real.



## ► Sistemas ETL de Tempo Real

- Segundo Santos e Bernardino (2009), para atender à característica de tempo real, são necessárias estratégias diferentes das tradicionais para atualização do *data warehouse*.
- O ETL tradicional não está preparado para atualizações de baixa latência, sendo necessárias algumas adaptações para atingir este objetivo.



## ► Sistemas ETL de Tempo Real

- Sistemas ETL de Tempo Real referem-se ao software que move os dados de maneira assíncrona para um *data warehouse* com certa urgência, minutos após a execução da transação comercial.
- A abordagem em tempo real para armazenamento de dados pode traçar uma linhagem clara para o que foi, originalmente chamado de Armazenamento de Dados Operacional (ADO).
- As motivações dos ADOs originais eram semelhantes aos *data warehouses* modernos em tempo real, mas a implementação de *data warehouses* em tempo real reflete uma nova geração de hardware, software e técnicas.



## ► Sistemas ETL de Tempo Real

- O uso da partição lógica e física em tempo real, conforme descrito originalmente por Kimball (2009), é uma solução pragmática disponível para fornecer análises em tempo real a partir de um *data warehouse*.
- O ADO é um sistema semi-operacional e de apoio à decisão, tentando encontrar um equilíbrio entre a necessidade de suportar simultaneamente atualizações frequentes e consultas frequentes.



## ► Sistemas ETL de Tempo Real

- Usando essa abordagem, uma tabela de fatos em tempo real separada é criada, cuja granularidade e dimensionalidade corresponde à tabela de fatos correspondente no *data warehouse* estático (carregado por noite).



PÓS-GRADUAÇÃO

# Sistemas ETL de Tempo Real

Bloco 2

Thiago Salhab Alves







## ► Abordagens ETL de Tempo Real

- Segundo Kimball e Caserta (2009), o ETL convencional é extremamente eficaz para atender aos requisitos diários, semanais e mensais, de relatórios em lote.
- Transações novas ou alteradas (registros de fatos) são movidas em massa e as dimensões são capturadas como instantâneos pontuais para cada carga.



## ► Abordagens ETL de Tempo Real

- O ETL, portanto, não é uma técnica adequada para integração de dados ou aplicativos para organizações que precisam de relatórios de baixa latência ou para organizações que precisam de uma captura de alterações dimensionais mais detalhada.

## ► Abordagens ETL de Tempo Real

- ***Microbatch ETL***
- *Microbatch ETL* é muito semelhante ao ETL convencional, exceto pelo fato de que a frequência de lotes é aumentada, talvez com a mesma frequência que de hora em hora.
- Esses microbatchs frequentes são executados por meio de um processo ETL convencional e alimentam diretamente as partições em tempo real dos *data marts*.
- Uma vez por dia, as partições em tempo real são copiadas para os *data marts* estáticos e são esvaziadas.

## ► Abordagens ETL de Tempo Real

- **Integração de Aplicativos Corporativos**
- A Integração de Aplicativos Corporativos descreve o conjunto de tecnologias que suportam a verdadeira integração de aplicativos, permitindo que sistemas operacionais individuais interajam de maneiras novas e potencialmente diferentes do que foram originalmente projetados.
- Habilitam ferramentas para o *data warehouse* em tempo real, pois suportam a capacidade de sincronizar dados importantes, como informações do cliente em aplicativos, e fornecem meios eficazes para distribuir ativos de informações derivadas de *data warehouse*, como novos valores de segmentação de clientes em toda a empresa.

## ► Abordagens ETL de Tempo Real

- **Capturar, Transformar e Fluxo**
- É uma categoria relativamente nova de ferramentas de integração de dados, projetada para simplificar o movimento de dados em tempo real por meio de tecnologias de bancos de dados heterogêneos.
- A camada de aplicativo dos aplicativos transacionais é ignorada. Em vez disso, as trocas diretas do banco de dados com o banco de dados são executadas.
- Transações, tanto novos fatos quanto mudanças de dimensão, podem ser movidos diretamente dos sistemas operacionais para as tabelas intermediárias de *data warehouse* com baixa latência, normalmente, alguns segundos.

PÓS-GRADUAÇÃO

## Teoria em Prática

Bloco 3

Thiago Salhab Alves





## ► Conformação de Dados

- 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 lançamento, o que, muitas vezes, acaba por comprometer o resultado financeiro, por investir em produtos com baixa procura. 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, entretanto, está tendo dificuldades para a tomada de decisões relacionadas ao que o seu público-alvo realmente consome e, assim, evitar gastos desnecessários.





## ► Conformação de Dados

- Após o processo de extração e limpeza, conformação e entrega dos dados do sistema de vendas e controle de dados e do marketing das redes social, constatou-se que os dados deveriam ser armazenados em tempo real. Como podemos auxiliar quanto a organização no processo armazenamento em tempo real?
- Resposta: pode-se utilizar um *microbatch ETL*, que é muito semelhante ao ETL convencional, exceto pelo fato de que a frequência de lotes é aumentada, talvez com a mesma frequência que de hora em hora.

## Dica do Professor

### Bloco 4

Thiago Salhab Alves





## ► Conformação de Dados

Indicação de Leitura de Capítulo de Livro da Biblioteca Virtual:

Leitura do capítulo 11 do livro:

- KIMBALL, L., R.; CASERTA, J. **The Data Warehouse ETL Toolkit**: Practical Techniques for Extracting, Cleaning, Conforming, and Data Delivering Data. Indianapolis: Wiley Publishing, 2009.



## ► Referências Bibliográficas

KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL Toolkit**: Practical Techniques for Extracting, Cleaning, Conforming, and Data Delivering Data. Indianapolis: Wiley Publishing, 2009.

SANTOS R. J., BERNARDINO J., 2009. Optimizing data warehouse loading procedures for enabling useful-time data warehousing. *In: Proceedings of the 2009 International Database Engineering & Applications Symposium (IDEAS '09)*. ACM, New York, NY, USA, 292-299. DOI=<http://dx.doi.org/10.1145/1620432.1620464>.

