

PÓS-GRADUAÇÃO

Integração e fluxo de dados



PÓS-GRADUAÇÃO

Entrega de Dados

Bloco 1

Thiago Salhab Alves





► **Conformação de Dados**

Objetivos

- Compreender o processo de entrega de dados.
- Aprender sobre dimensões planas e dimensões flocos de neve.
- Aprender sobre dimensões de data e hora e dimensões grandes.



► **Conformação de Dados**

Entrega de Dados

- De acordo com Kimball e Caserta (2009), a etapa final do processo de ETL é a entrega de dados, em que são preparadas as estruturas de tabelas dimensionais de forma mais restrita.
- Todas as dimensões devem ser fisicamente contruídas para ter o conjunto mínimo de componentes.

► Entrega de Dados

- Segundo Kimball e Caserta (2009), a chave primária fica armazenada em um único campo contendo um valor inteiro único chamado valor substituto.
- O processo de ETL do *data warehouse* deve sempre criar e inserir as chaves substitutas, isto é, o *data warehouse* possui essas chaves e nunca permite que outra entidade as atribua.
- A chave primária de uma dimensão é usada para unir-se a tabela de fatos que devem preservar a integridade referencial. A chave de dimensão principal é unidade a uma chave estrangeira, correspondente na tabela de fatos .



► Entrega de Dados

- O componente final de todas as dimensões, de acordo com Kimball e Caserta (2009), além da chave primária e da chave natural, é o conjunto de atributos descritivos.
- O arquiteto do *data warehouse*, provavelmente, especificará um grande número de atributos descritivos para dimensões, como empregado, cliente e produto.
- Dados dimensionais para dimensões grandes e complexas, como cliente, fornecedor ou produto, são frequentemente extraídos de várias fontes, em diferentes momentos.

► Dimensões Planas e Dimensões de Flocos de Neve

- Segundo Kimball e Caserta (2009), se uma dimensão é normalizada, as hierarquias criam uma estrutura característica conhecida como floco de neve, com os níveis das hierarquias com relacionamentos muitos-para-um perfeitos.
- A diferença com a qual nos importamos é o impacto negativo que o modelo normalizado e com flocos de neve tem no ambiente do usuário final.

► Dimensões Planas e Dimensões de Flocos de Neve

Existem dois problemas:

- Primeiro, se os muitos relacionamentos restritos a um modelo hierárquico mudarem, o esquema da tabela normalizada e as junções declaradas entre as tabelas devem ser alteradas e o ambiente do usuário final deve ser recodificado em algum nível para que os aplicativos continuem funcionando.
- Segundo, os esquemas complexos são notórios por confundir os usuários finais, e um esquema normalizado requer o mascaramento dessa complexidade na área de apresentação do data warehouse.

PÓS-GRADUAÇÃO

Entrega de Dados

Bloco 2

Thiago Salhab Alves





► Dimensões de Data e Hora

- A dimensão de tempo mais comum e útil é a dimensão de data do calendário com a granularidade de um único dia.
- Apenas alguns desses atributos (como nome do mês e ano) podem ser gerados diretamente de uma expressão de data e hora do SQL.
- Feriados, dias úteis, períodos fiscais, números de semanas, sinalizadores de último dia do mês e outros atributos de navegação devem ser incorporados na dimensão de data do calendário e toda navegação de data deve ser implementada em aplicativos, usando os atributos dimensionais.



► Dimensões Grandes

- De acordo com Kimball e Caserta (2009), as dimensões mais interessantes em um *data warehouse* são as grandes dimensões, como cliente, produto ou local. Uma grande dimensão de cliente comercial, geralmente, tem milhões de registros e cem ou mais campos em cada registro.
- As dimensões realmente grandes quase sempre são derivadas de múltiplas fontes. Os clientes podem ser criados por um dos vários sistemas de gerenciamento de contas em uma grande empresa.



► Dimensões Grandes

- Na etapa de remoção de duplicatas, que faz parte do módulo de limpeza de dados, cada cliente deve ser identificado corretamente em fontes de dados originais separadas para que a contagem total de clientes esteja correta. Uma chave mestra natural para o cliente pode ter que ser criada pelo *data warehouse* nesse ponto.

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. Outro problema são os produtos que possuem prazo de validade curto, que acabam por vencer e não podem ser trocados pelos fornecedores.



► Conformação de Dados

- 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 tendo dificuldades para a tomada de decisões relacionado ao que o seu público-alvo realmente consome e, assim, evitar gastos desnecessários. Após o processo de extração e limpeza e conformação dos dados do sistema de vendas e controle de dados e do marketing das redes social, os dados necessitam ser entregues. Como podemos auxiliar quanto a organização no processo de entrega dos dados?



► **Conformação de Dados**

Resposta: criando um processo de entrega de dados, preparando as estruturas de tabelas dimensionais de forma mais restrita, construindo com o mínimo de componentes.

Dica do Professor

Bloco 4

Thiago Salhab Alves





► Conformação de Dados

Indicação de Leitura de Capítulo de Livro da Biblioteca Virtual:

Leitura do capítulo 5 (parte 4) 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.

