



Sistemas de informação

enterprise analytics and data warehousing

Prof^o Fabiano J. Cury Marques



enterprise analytics and data warehousing

Modelagem Dimensional Parte IV



- X Normalização de dimensões Snowflaking
- **X** Erro de Projeto: Muitas dimensões
- X Referências





Normalização de dimensões - Snowflaking





- X As tabelas dimensão desnormalizadas com repetições de valores textuais podem fazer com que os projetistas de base de dados sintam-se desconfortáveis
- Ex.: 150.000 produtos pertencem a apenas 50 departamentos distintos. Em vez de armazenar redundantemente a descrição do departamento com 20 bytes na dimensão Produto, projetistas com bagagem de normalização querem armazenar um código de departamento com 2 bytes apenas e então criar uma nova tabela dimensão Departamento que irá decodificar os departamentos
- > De fato, para estes projetistas, isso serviria para todos os descritores de nossas tabelas
- Além disso tem o fator da manutenção, mas lembre-se, no DW, isso ocorre lá no repositório Stage e não no DDS





✗ A normalização das tabelas dimensão tipicamente são chamadas de snowflaking

✗ Atributos redundantes são removidos da tabela dimensão desnormalizada e colocados em tabelas dimensão secundárias

POS Retail Sales Transaction Fact

Date Key (FK)
Product Key (FK)
Store Key (FK)
Promotion Key (FK)
POS Transaction Number (DD)
Sales Quantity
Sales Dollar Amount
Cost Dollar Amount
Gross Profit Dollar Amount

Product Dimension

Product Key (PK)
Product Description
SKU Number (Natural Key)
Brand Key (FK)
Package Type Key (FK)
Fat Content
Weight
Weight Units of Measure
Storage Type Key (FK)
Shelf Width
Shelf Height
... and more

Brand Dimension

Brand Key (PK) Brand Description Category Key (FK)

Package Type Dimension

Package Type Key (PK) Package Type Description

Storage Type Dimension

Storage Type Key (PK) Storage Type Description Shelf Life Type Key (FK)

Category Dimension

Category Key (PK) Category Description Department Key (FK)

Department Dimension

Department Key (PK)
Department Description

Shelf Life Type Dimension

Shelf Life Type Key (PK) Shelf Life Type Description





- Apesar do snowflaking ser uma extensão legal do modelo dimensional, em geral, encoraja-se a resistir a tentação de normalizar seu modelo dimensional dado nossas duas principais premissas de projeto:
 - Facilidade de Uso
 - Desempenho
- As tabelas snowflaking tornam a representação muito mais complexa. Além disso, os otimizadores das bases de dados terão muito mais trabalho com inúmeros joins, gerando maior lentidão no desempenho das consultas
- Se o problema for espaço observe que a economia de uma dimensão snowflaked é mínimo: no nosso exemplo anterior, 150.000 * 20 bytes trocados por 150.000 * 2 bytes economizam APENAS 2,7 MB. Enquanto isso temos fatos ocupando facilmente 10 GB



X Dilema de modelagem ...

VOLUME DE DADOS ARMAZENADO!

VS

DESEMPENHO NA CONSULTA DOS DADOS

FACILIDADE DE USO

Se o problema for espaço observe que a economia de uma dimensão snowflaked é mínimo: no nosso exemplo anterior, 150.000 * 20 bytes trocados por 150.000 * 2 bytes economizam APENAS 2,7 MB. Enquanto isso temos fatos ocupando facilmente 10 GB





Muitas Dimensões





- ✗ A tabela fato no esquema dimensional é naturalmente altamente normalizada e compacta
- X Não há como normalizar mais ainda relacionamentos muitos-para-muitos complexos pois as dimensões não são correlacionadas
- **X** É interessante notar que, apesar de desconfortáveis com tabelas dimensão desnormalizadas, alguns projetistas são tentados a desnormalizar a tabela fato
- **X** Em vez de ter a chave estrangeira de um produto específico, inclui-se chaves estrangeiras para elementos analisados com frequência na hierarquia de produtos. Ex.: marca, categoria e departamento





X Antes que perceba, sua tabela fato que deveria ser compacta já está relacionando dezenas de tabela dimensão



- X Normalmente as tabelas com estas características são chamadas de centopéia
- Lembre-se que apesar de não estarmos preocupados com o espaço consumido pelas tabelas dimensão, as tabelas fato sim devem ter esse fator levado em consideração
- A maior parte dos processos de negócio podem ser representados com menos de 15 dimensões na tabela fato. Se passar muito disso, deve-se verificar se não há como combinar dimensões correlacionadas em uma só (ex.: níveis hierárquicos)



Dimensão Data Dimensão Promoção

Dimensão Produto

Combinação máxima possível:

Fato Venda

Dimensão Local

Dimensão Vendedor Dimensão Cliente

= 111

Date Dimension

Week Dimension

Month Dimension

Quarter Dimension

Year Dimension

Fiscal Year Dimension

Fiscal Month Dimension

Store Dimension

Store County Dimension

Store State Dimension

Store District Dimension

Store Region Dimension

Store Floor Plan Dimension

POS Retail Sales Transaction Fact

Date Key (FK)

Week Key (FK)

Month Key (FK)

Quarter Key (FK)

Year Key (FK)

Fiscal Year (FK)

Fiscal Month (FK)

Product Key (FK) Brand Key (FK)

Subcategory Key (FK)

Category Key (FK)

Department Key (FK)

Package Type Key (FK)

Store Key (FK)

Store County (FK) Store State Key (FK)

Store District Key (FK)

Store Region Key (FK)

Store Floor Plan (FK)

Promotion Key (FK)
Promotion Reduction Type (FK)

Promotion Media Type (FK)

POS Transaction Number (DD)

Sales Quantity

Sales Dollar Amount Cost Dollar Amount

Gross Profit Dollar Amount

Product Dimension

Brand Dimension

Subcategory Dimension

Category Dimension

Department Dimension

Package Type Dimension

Promotion Dimension

Promotion Reduction Type

Promotion Media Type



QUANTIDADE DE DIMENSÕES



X Dilema de modelagem ...

EXPRESSIVA
QUANTIDADE DE
CHAVES
EXTRANGEIRAS NA
TABELA FATO

VS

DIVERSIDADE DE DIMENSÕES



scd

Slowly Changing Dimension



SLOWLY CHANGING DIMENSION (SCD)



- ✗ SCD é uma técnica utilizada para armazenar os valores históricos de atributos de uma dimensão
- SOS valores dos atributos dimensionais mudam com o tempo. Quando eles mudam, você pode sobreescrever os valores antigos com os novos, ou pode preservar o valor antigo
- **X** Existem dois métodos para se preservar o valor antigo:
- X Armazená-los como linhas
- X Armazená-los como colunas



SLOWLY CHANGING DIMENSION (SCD)



- Suponha que a loja 7 estava na região 1, mas agora está na região 2. Você quer armazenar a informação histórica, o fato de que ela foi da região 1
- ✗ Você pode fazer isso armazenando a informação como linhas ou coluna

X Como linha:

key	store	region	status	
1	7	1	expired	
2	7	2	active	

X Como coluna:

key	store	current_region	old_region	effective_date
1	7	2	1	11/18/2007

- Neste caso a questão é, o que fazer se a região mudar novamente?
- Você pode criar mais uma coluna old_region2
- Esta é a principal desvantagem em relação ao anterior



TIPOS DE SCD



- **X** SCD tipo 1
 - O Sobreescreve os valores antigos dos atributos, assim, os valores antigos não são mantidos
- X SCD tipo 2
 - Mantém os valores antigos criando uma nova linha para cada mudança
- X SCD tipo 3
 - Mantém os valores antigos os colocando em outra coluna



SLOWLY CHANGING DIMENSION (SCD)



X Dilema de modelagem ...

ARMAZENAMENTO
DE DADOS POR
LINHA OU
COLOUNA

VS

FLEXIBILIDADE DE CONSTRUÇÃO DE CONSULTAS PARA ACESSO AO DADO



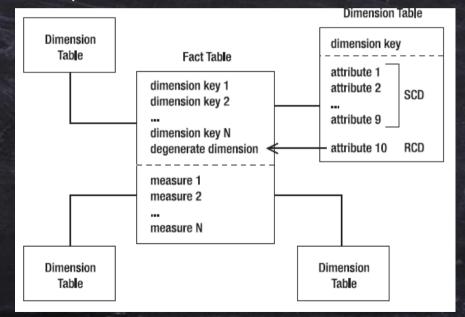
E A RAPIDLY CHANGING DIMENSIONS (RCD)?



Um critério para definir se um atributo muda lentamente é o tempo de sua atualização, algo como uma vez a cada trimestre, bimestre no mínimo

Mais do que isso, não é SCD, então trata-se diferentemente, tirando o atributo que muda com grande frequência da tabela dimensão e colocando diretamente na

fato





RAPIDLY CHANGING DIMENSIONS (RCD)



X Dilema de modelagem ...

ALTERAÇÃO RAPIDA DE DADOS NA TABELA DIMENSÃO

VS

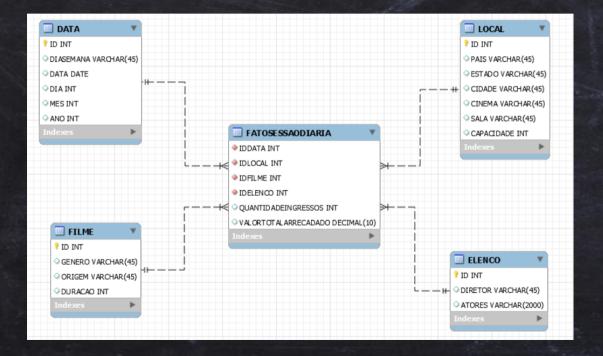
DEFINIÇÃO DE NOVA MEDIDA NA TABELA FATO



Exercício:



➤ De forma livre, sugira um redesenho do DataMart abaixo para ao menos uma situação hipotética que possa utilizar snowflaking:







Referências

- ✗ KIMBALL, R., ROSS, M. The Data Warehouse Toolkit. 2ª ed., John Wiley Professional, 2002.
- MACHADO, F. N. R. Tecnologia e Projeto de Data Warehouse. 1ª ed., São Paulo: Ed. Érica, 2004.





Copyright © 2019 Prof. MSc. Eng. Wakim B. Saba

https://br.linkedin.com/in/wakimsaba

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).