Slowly Changing Dimensions

Slowly Changing Dimensions (SCD) consiste em técnicas para gerenciamento da história de dados dimensionais em um DW. Deve-se ter em mente que as dimensões associadas às Tabelas Fato são afetadas com o passar do tempo. A mudança nos valores de dimensões pode ocorrer às vezes, em alguns casos, meramente para a correção de erros. Mas, às vezes, uma mudança em alguma descrição representa uma mudança verdadeira ocorrida em algum ponto do tempo, como a mudança do nome de um produto, ou de um cliente, por exemplo. Como tais mudanças ocorrem de forma inesperada, e esporadicamente, e com bem menos frequência que medidas em Tabelas Fato, o tópico é denominado *Slowly Changing Dimensions* (SCDs) (Kimball e Ross, 2013).

Os representantes de governança e administração de dados da empresa devem estar ativamente envolvidos nas decisões sobre o tratamento de mudanças em valores de dimensões; A TI não deve fazer determinações por conta própria (Kimball e Ross, 2013).

Kimball e Ross (2013) afirmam que são necessárias apenas três respostas básicas em caso de mudança de descrição em algum atributo de dimensão:

- Tipo 1: Substituição (*Overwrite*):
  - Destruição da história. No exemplo acima, relatórios que filtram ou agrupam pelo campo “cidade” serão afetados. Usuários de negócio precisam ser avisados de que isso pode ocorrer. O DW precisa de uma política explícita e visível de utilização do Tipo 1, como por exemplo: “Erros serão corrigidos”, e/ou, “Não mantemos a história deste campo, mesmo se o seu valor for alterado”.
  - Propagação da mudança. Todas as cópias distribuídas que dependem da dimensão deverão replicar as mudanças simultaneamente.

    - Suponha que em um dia qualquer, o atributo “cidade” de Ralph Kimball, na Tabela Dimensão “empregado”, foi alterado de “Santa Cruz” para “Boulder Creek”. Mais tarde, somos avisados que isso se tratava de uma correção de erro, e não de uma mudança de localização. Neste caso, decidimos por atualizar o valor de “cidade”. Este é um exemplo clássico de uma mudança de Tipo 1. Adequado para casos de correção de erros e para situações em que a mudança não afete a história.
    - Embora o Tipo 1 seja o tratamento mais simples e direto para lidar com alterações em valores de dimensões, os seguintes pontos devem ser observados:

- Tipo 2: Adição de um novo registro de dimensão:
  - O Tipo 2 requer que generalizemos a chave primária de “empregado”. Neste caso, é recomendado que a chave primária da relação “empregado” seja algo desconectado da chave primária da relação original, proveniente do sistema relacional. Indica-se, por exemplo, a criação de uma chave primária sequencial. Este tipo de chave primária é chamado *surrogate key* (chave substituta).

    - Vamos alterar o cenário anterior e supor que Ralph Kimball de fato se mudou para uma nova cidade em 18/07/2020. Digamos que a política da organização consiste em manter no DW o histórico exato de endereço residencial de cada funcionário. Esta então seria uma situação clássica para aplicação do Tipo 2.
    - O Tipo 2 exige que seja incluído um novo registro na relação “empregado”, referente ao novo valor para o campo “cidade”. Este tratamento resulta nos seguintes efeitos:

- Tipo 3: Adição de um novo atributo.
    - Embora os tipos 1 e 2 vistos anteriormente sejam as técnicas principais utilizadas para tratamento de mudanças em dimensões, precisaremos de uma terceira técnica para abarcar situações em que precisamos alternar entre realidades distintas. Diferentemente de atributos físicos que podem ter apenas um único valor em determinado momento, alguns atributos atribuídos pelo usuário podem legitimamente ter mais de um valor por vez, dependendo do ponto de vista do observador. Por exemplo, uma categoria de produto pode ter mais de uma interpretação. Em uma papelaria, uma caneta de marcação pode ser assimilada na categoria de bens domésticos, ou na categoria de suprimentos artísticos. Usuários e aplicações de BI precisam ter a possibilidade de escolha, em tempo de execução, de qual realidade será aplicada naquele momento.
    - O requisito por uma realidade alternativa de visualização de um atributo de dimensão é geralmente acompanhado do requisito de se permitir a separação de versões da realidade nos dados passados e futuros, mesmo que a requisição da inclusão dessa nova realidade tenha chegado ao DW no momento presente.
    - Tais requisitos podem ser alcançados por meio da adição de uma nova coluna para cada realidade alternativa do atributo de dimensão. Com esta abordagem, usuários e aplicações de BI podem, a qualquer momento, alterar a realidade por meio da seleção do campo criado para a realidade desejada.

É comum uma mesma dimensão possuir estratégias de tratamento de mudanças específicas para cada atributo (Kimball e Ross, 2013).

Revisão #2
Criado 2022-03-04 14:55:18 UTC por FLAVIO LOPES DE MORAIS
Atualizado: 2022-03-08 08:24:45 UTC por FLAVIO LOPES DE MORAIS