

Data Warehouse

Compras

Caroline B. Perlin

Agenda

- O processo de compra
- Requisitos de compras
- Transações de compra
- Tabela de fatos
- Slowly Changing Dimensions (SCD)
- Técnicas para lidar com SCD
 - Abordagens puras
 - Abordagens híbridas
- Dimensões que mudam mais rapidamente (*overview*)

O processo de compra

- Oportunidade de economia de custos, com o planejamento da demanda
- Envolve os seguintes itens:
 - Contratos de negociação
 - Emissão de requisição de compra e ordem de compra
 - Controle de recebimento e pagamentos

Requisitos de compras

- Que materiais são comprados com maior frequência?
 - Existem quantos fornecedores destes produtos?
 - A qual preço cada fornecedor oferece o produto?
 - Em qual unidade de medida?
- Podemos reduzir o preço fazendo uma compra maior?
- Os funcionários compram dos fornecedores preferenciais ou burlam contratos firmados?
- O preço do produto é o mesmo que foi negociado anteriormente com o fornecedor?
- As entregas são feitas no prazo?
- Qual a porcentagem de desistência de pedidos?

Transações de compra

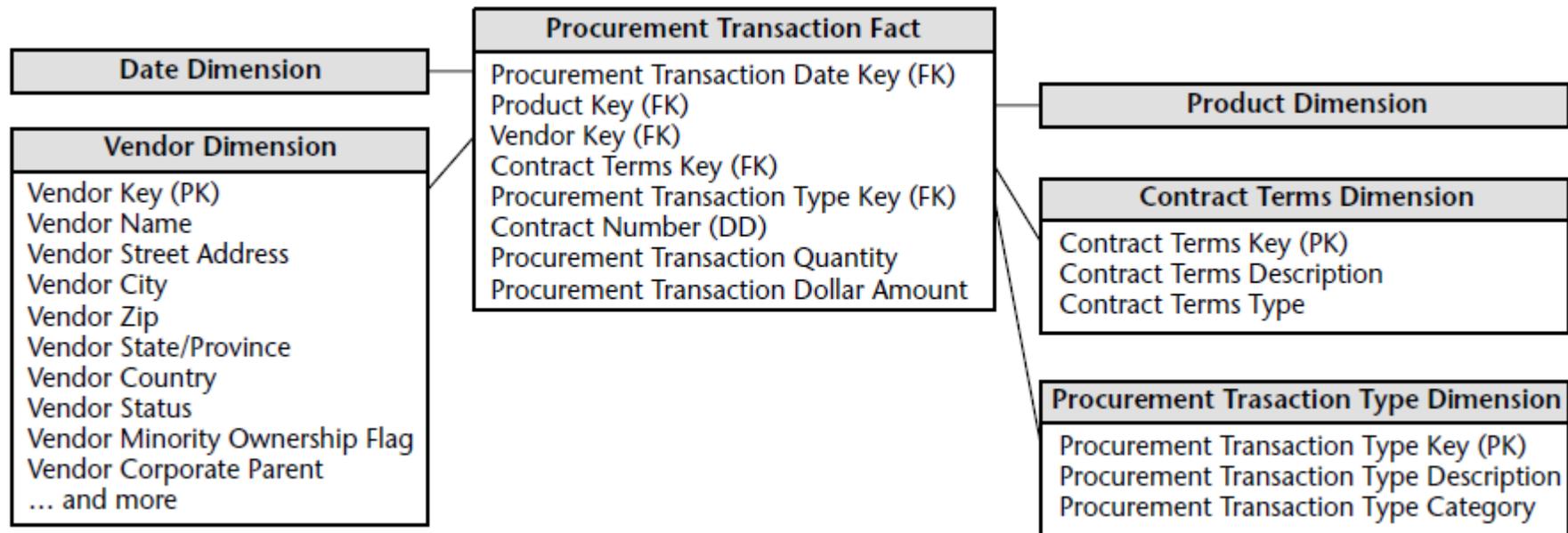


Tabela de fatos

- **Opção 1:** tabela de fatos de várias transações.
- **Opção 2:** várias tabelas de fatos de 1 transação.
- Mais tempo de gerenciamento e administração pois há mais tabelas para carregar, indexar e agregar.

Como decidir entre estas opções?

- Requisitos de usuário
 - Usuários analisam várias transações de uma vez ou cada uma em separado?
- Processos de negócio
 - São exclusivos? Ex.: tratar ordens de compra diferente de receber produto indica que podemos usar tabelas de fatos separadas.
- Sistemas de origem
 - Vários sistemas de origem sugere tabelas de fatos separadas.
- Dimensão dos dados
 - Se as dimensões se aplicam a somente alguns tipos de transação, leva a optar por tabelas de fatos separadas.

Exemplo de tabela de fatos separada

Date Dimension	Purchase Requisition Fact	Product Dimension
Vendor Dimension	Requisition Date Key (FK) Requested Date Key (FK) Product Key (FK) Vendor Key (FK) Contract Terms Key (FK) Requested By Key (FK) Contract Number (DD) Purchase Requisition Number (DD) Purchase Requisition Quantity Purchase Requisition Dollar Amount	Contract Terms Dimension
Employee Dimension		Received Condition Dimension
		Discount Taken Dimension
	Purchase Order Fact Requisition Date Key (FK) Requested Date Key (FK) Purchase Order Date Key (FK) Product Key (FK) Vendor Key (FK) Contract Terms Key (FK) Requested By Key (FK) Purchase Agent Key (FK) Contract Number (DD) Purchase Requisition Number (DD) Purchase Order Number (DD) Purchase Order Quantity Purchase Order Dollar Amount	
	Shipping Notices Fact Shipping Notification Date Key (FK) Ship Date Key (FK) Requested Date Key (FK) Product Key (FK) Vendor Key (FK) Contract Terms Key (FK) Requested By Key (FK) Purchase Agent Key (FK) Contract Number (DD) Purchase Requisition Number (DD) Purchase Order Number (DD) Shipping Notification Number (DD) Shipped Quantity	
	Warehouse Receipts Fact Warehouse Receipt Date Key (FK) Ship Date Key (FK) Requested Date Key (FK) Product Key (FK) Vendor Key (FK) Received Condition Key (FK) Warehouse Clerk (FK) Purchase Requisition Number (DD) Purchase Order Number (DD) Shipping Notification Number (DD) Received Quantity	
	Vendor Payment Fact Payment Date Key (FK) Ship Date Key (FK) Warehouse Receipt Date Key (FK) Product Key (FK) Vendor Key (FK) Contract Terms Key (FK) Discount Taken Key (FK) Contract Number (DD) Purchase Requisition Number (DD) Purchase Order Number (DD) Shipping Notification Number (DD) Accounts Payable Check Number (DD) Vendor Payment Quantity Vendor Gross Payment Dollar Amount Vendor Payment Discount Dollar Amount Vendor Net Payment Dollar Amount	

- Pontos importantes para esta decisão:
 - Atividades são processos de negócio distintos
 - Dados vêm de diferentes sistemas de origem
 - Dimensionalidade única para os vários tipos de transação.

Motivação

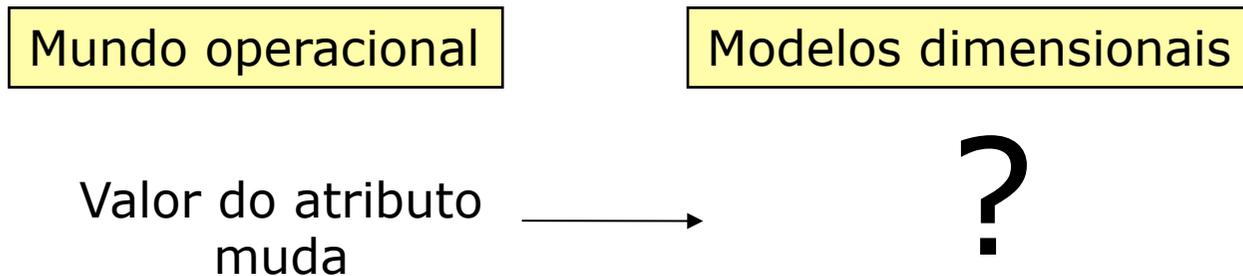
- Até agora consideramos:
 - Cada dimensão independente das demais no sentido lógico
 - Dimensões independentes do tempo

MAS

- Os atributos de dimensão mudam lentamente com o passar do tempo

Dimensões que mudam lentamente – Slowly Changing Dimensions (SCD)

- Solução: uma estratégia para cada atributo da tabela de fatos lidar com a mudança.



Técnicas básicas de lidar com dimensões que mudam lentamente

- Abordagens puras
 1. Substituir o valor
 2. Adicionar uma linha de dimensão
 3. Adicionar uma coluna de dimensão
- Abordagens híbridas
 - a. Alterações previsíveis com sobreposição de várias versões
 - b. Alterações imprevisíveis com sobreposição de versão única

Técnicas básicas de lidar com dimensões que mudam lentamente

- **Abordagens puras**
 1. Substituir o valor
 2. Adicionar uma linha de dimensão
 3. Adicionar uma coluna de dimensão
- **Abordagens híbridas**
 - a. Alterações previsíveis com sobreposição de várias versões
 - b. Alterações imprevisíveis com sobreposição de versão única

Técnica 1: Substituir o valor

Substituir o valor antigo pelo correto na linha de dimensão.

Chave	Descrição	Depto	SKU*
12345	IntelliKidz1.0	Educação	ABC922-Z

Modificado para:

Chave	Descrição	Depto	SKU*
12345	IntelliKidz1.0	Estratégia	ABC922-Z

*SKU- Stock Keeping Unit (chave natural)

Técnica 1: Substituir o valor

- Vantagem
 - Rápido e fácil
- Desvantagem
 - Perdemos o histórico de mudanças de um atributo.
 - Se as vendas aumentarem após a mudança do departamento, não teremos informações para explicar o aumento do desempenho.
- As agregações preexistentes baseadas no valor do departamento precisarão ser recriadas.

Técnica 2: Adicionar uma linha de dimensão

- Técnica predominante.
- Criar uma nova linha de dimensão para o produto e refletir o novo valor do depto.

Chave	Descrição	Depto	SKU
12345	IntelliKidz1.0	Educação	ABC922-Z
25984	IntelliKidz1.0	Estratégia	ABC922-Z

- Observe que cada uma das chaves substitutas identifica um perfil de atributo exclusivo que era verdadeiro por um período de tempo.

Técnica 2: Adicionar uma linha de dimensão

- Na tabela de fatos:
 - Linha do IntelliKidz anteriores a 19/10/2009 → chave do produto 12345, quando o produto era agrupado no departamento Educação.
 - Linha do IntelliKidz posteriores a 19/10/2009 → chave do produto 25984, o produto é agrupado no departamento Estratégia.

Técnica 2: Adicionar uma linha de dimensão

- A técnica 2 segmenta perfeitamente o histórico para justificar a alteração.
 - Para diferenciar os 2 perfis de produto → atributo departamento
 - Se quisermos o histórico completo do produto → atributo descrição
 - Contagem do número de produtos → atributo SKU.

Chave	Descrição	Depto	SKU
12345	IntelliKidz1.0	Educação	ABC922-Z
25984	IntelliKidz1.0	Estratégia	ABC922-Z

Técnica 2: Adicionar uma linha de dimensão

- Vantagens
 - Segmenta perfeitamente o histórico para justificar a alteração
 - Podemos controlar normalmente quantas mudanças de dimensão forem necessárias.
 - Não há necessidade de criar uma junção secundária confusa baseada em datas.
- Desvantagens
 - Crescimento acelerado da tabela de dimensão. Talvez esta técnica não seja apropriada para tabelas de dimensão que excederam 1 milhão de linhas.

Técnica 3: Adicionar uma coluna de dimensão

- Criar uma nova coluna para capturar a alteração do atributo.

Chave	Descrição	Depto atual	Depto ant	SKU
12345	IntelliKidz1.0	Estratégia	Educação	ABC922-Z

- Utilizada com pouca frequência.
- Distingue da técnica 2 porque tanto a descrição atual quanto a anterior podem ser consideradas verdadeiras ao mesmo tempo.

Técnica 3: Adicionar uma coluna de dimensão

- Vantagens
 - Necessidade de usar 2 modos de visão do mundo ao mesmo tempo.
 - Podemos controlar normalmente quantas mudanças de dimensão forem necessárias.
- Desvantagens
 - Quando há necessidade de controlar várias alterações imprevisíveis.

Técnicas básicas de lidar com dimensões que mudam lentamente

- **Abordagens puras**
 1. Substituir o valor
 2. Adicionar uma linha de dimensão
 3. Adicionar uma coluna de dimensão
- **Abordagens híbridas**
 - a. Alterações previsíveis com sobreposição de várias versões
 - b. Alterações imprevisíveis com sobreposição de versão única

a. Alterações previsíveis com sobreposição de várias versões

- Suponha uma empresa de vendas revisando o mapa de seus distritos anualmente.
 - Após um período de 5 anos, a empresa foi reorganizada 5 vezes.
 - Requisitos de relatórios de venda anual usando um mapa de distritos de cada ano.

Decisão de qual técnica usar

- Não podemos usar a técnica 2 porque ela particiona o histórico.
- Não podemos usar a técnica 3 porque queremos usar mais de dois mapas ao mesmo tempo.
- **Solução:** generalizar a técnica 3 para ter cinco versões do atributo de distrito para cada representante de vendas.

Exemplo

Dimensão Rep. De Vendas
Chave do rep. De vendas
Endereço do rep. De vendas
Distrito atual
Distrito 2001
Distrito 2000
Distrito 1999
Distrito 1998
...e muito mais

- Cada linha da dimensão Rep. De vendas incluiria todas as atribuições distritais anteriores.
- O usuário poderia escolher fazer o agrupamento dos fatos de venda com qualquer dos 5 mapas de distrito.
- Ex.: Se um rep. De vendas fosse do ano 2000, nos anos 1999 e 1998 ele teria o valor N/A.

b. Alterações imprevisíveis com sobreposição de versão única

- Preservar a precisão histórica em torno das alterações de atributos imprevisíveis, ao mesmo tempo mantendo a capacidade de relatar dados históricos de acordo com os valores atuais.
- Assim, no exemplo do software IntelliKidz, a configuração inicial seria:

Chave	Descrição	Depto atual	Depto ant	SKU
12345	IntelliKidz1.0	Educação	Educação	ABC922-Z

Alterações imprevisíveis com sobreposição de versão única

- Quando os departamentos são reestruturados e o IntelliKidz muda para o depto de estratégia, é usada a técnica 2 para inserir uma nova linha:

Chave	Descrição	Depto atual	Depto ant	SKU
12345	IntelliKidz1.0	Estratégia	Educação	ABC922-Z
25984	IntelliKidz1.0	Estratégia	Estratégia	ABC922-Z

Alterações imprevisíveis com sobreposição de versão única

- Se o IntelliKidz fosse movido para o departamento de software “Pensamento Crítico”, teríamos:

Chave	Descrição	Depto atual	Depto ant	SKU
12345	IntelliKidz1.0	Pensamento Crítico	Educação	ABC922-Z
25984	IntelliKidz1.0	Pensamento Crítico	Estratégia	ABC922-Z
31726	IntelliKidz1.0	Pensamento Crítico	Pensamento Crítico	ABC922-Z

Alterações imprevisíveis com sobreposição de versão única

- Com esta abordagem híbrida, criamos:
 - Uma nova linha para capturar a alteração (técnica 2)
 - Uma nova coluna para controlar a atribuição atual (técnica 3)
- As alterações subsequentes são tratadas como a técnica 1.
- Esta técnica permite controlar precisamente as alterações históricas e ao mesmo tempo dá suporte ao agrupamento do histórico com base nas atribuições atuais.

Dimensões que mudam mais rapidamente

- Se um atributo mudar mensalmente, não poderemos aplicar as técnicas de SCD.
- Uma técnica que pode ser utilizada é separar os atributos que mudam mais rapidamente em uma ou mais dimensões.
- Na tabela de fatos, haveria 2 chaves externas:
 - Uma para cada tabela de dimensão principal
 - Outra para os atributos que mudam mais rapidamente
 - As tabelas de dimensão seriam associadas uma com a outra toda vez que colocássemos uma linha na tabela de fatos.
- Mais detalhes sobre esta abordagem no Cap. 6.



P E R G U N T A S ?

