
Newton
Quem se prepara, não para.

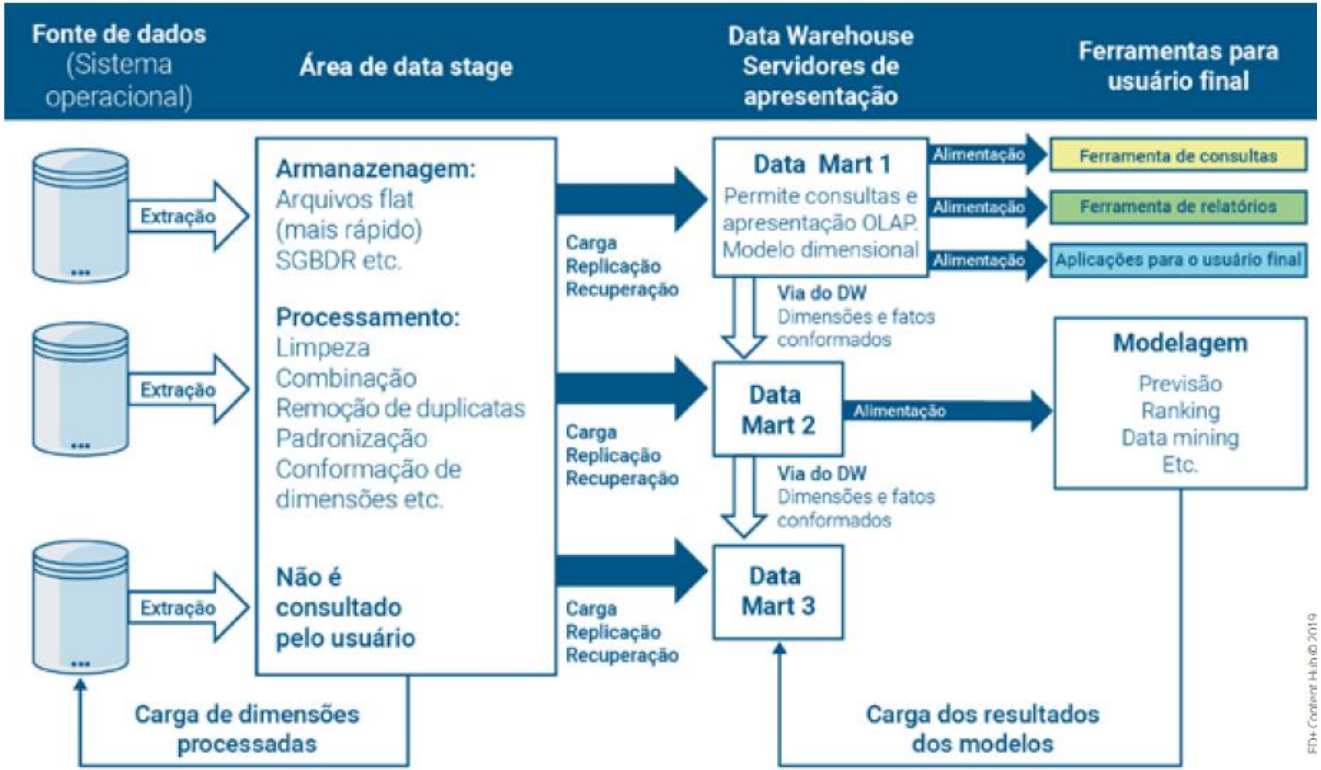
Business Intelligence

4º período

Professora: Michelle Hanne

Componentes de um Datawarehouse

Newton
Quem se prepara, não para.

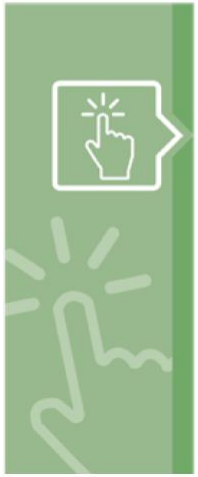


Processo de Projeto

Newton
Quem se prepara, não para.

Processo de projeto dimensional em 4 etapas

As técnicas de modelagem dimensional de um data warehouse, se aplicadas corretamente, garantem que o desenho do **data warehouse** reflita a forma de pensar dos analistas de negócio e gerentes da empresa e possa ser usado eficazmente para atender os seus requisitos de negócio.



Você deve selecionar o processo de negócios a ser modelado. Um processo é uma atividade comercial natural realizada em sua organização que, normalmente, é suportada por um sistema de coleta de dados de origem. Ouvir seus usuários é o meio mais eficiente para selecionar o processo de negócios. As medições de desempenho que eles desejam analisar no data warehouse resultam de processos de medição de negócios. Exemplos de processos de negócios incluem compra de matérias-primas, pedidos, remessas, faturamento, estoque e contabilidade. Ao focar nos processos de negócios e não nos departamentos de negócios, podemos fornecer informações consistentes de maneira mais econômica em toda a organização. Se estabelecermos modelos dimensionais vinculados por departamentos, inevitavelmente duplicaremos os dados com diferentes rótulos e terminologia. Múltiplos fluxos de dados em modelos dimensionais separados nos tornarão vulneráveis a incompatibilidades de dados. A melhor maneira de garantir a consistência é publicar os dados uma vez. Uma única execução de publicação também reduz o esforço de desenvolvimento de Extração-Transformação-Carga (ETL).

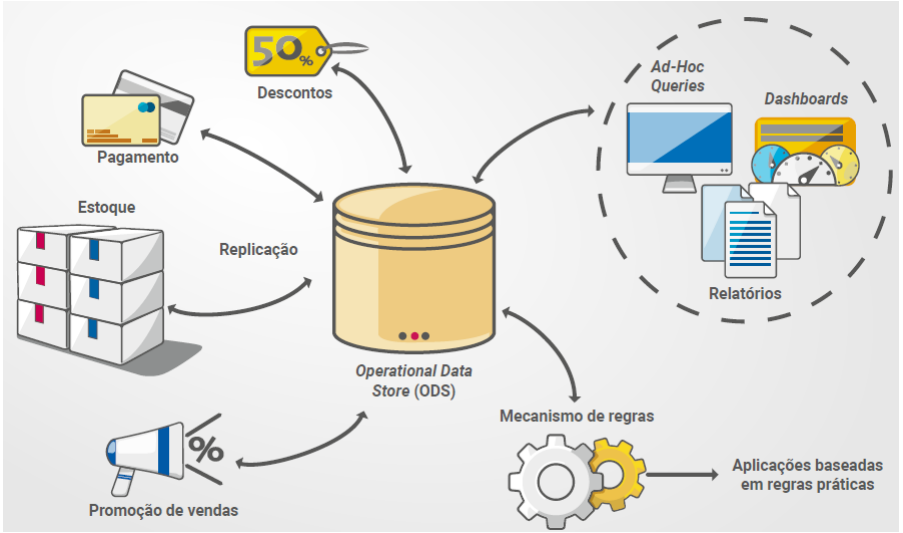
Encontre o grão do processo de negócios. Especifique exatamente o que ele representa na linha de tabela de fatos individuais. O grão transmite o nível de detalhe associado às medições da tabela de fatos. Ele fornece a resposta para a pergunta: **como você descreve uma única linha na tabela de fatos?** Exemplos na definição do grão: item de linha em um cupom fiscal, cartão de embarque individual para embarcar e extrato de conta bancária.

Você deve escolher as dimensões que se aplicam a cada linha da tabela de fatos. Dimensões resultam da pergunta: **como os empresários descrevem os dados resultantes do processo de negócios?** A intenção é decorar as tabelas de fatos com um conjunto de dimensões bastante representativo de todas as descrições possíveis que assumem valores únicos no contexto de cada medição. Se estiver claro sobre a granulação, as dimensões poderão ser identificadas com bastante facilidade.

Você deve identificar as transações numéricas que preencherão cada linha da tabela de fatos. Os fatos são determinados pela resposta à pergunta: **o que está sendo medido?** Os usuários de negócios estão profundamente interessados em analisar essas medidas de desempenho dos processos de negócios. Os fatos que claramente pertencem a um grão diferente devem estar em uma tabela de fatos separada. Fatos típicos são números aditivos numéricos, como quantidade solicitada pelo valor do custo ordinário.

ODS

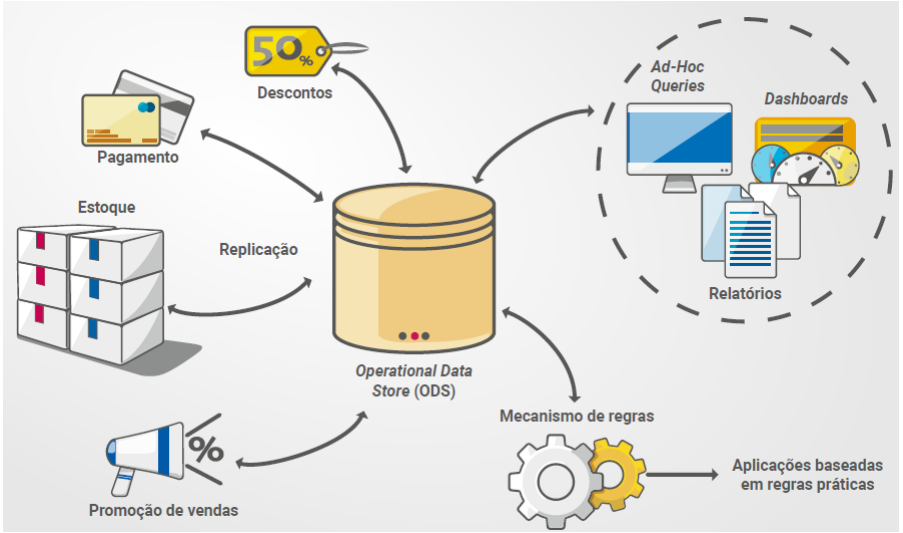
Os dados operacionais (ODS – ***Operational Data Store***) se encaixam no diagrama de componentes do data warehouse. Os *ODSs* são cópias atualizadas e um tanto integradas dos dados operacionais (ERP). A frequência de atualização e o grau de integração de um *ODS* varia de acordo com os requisitos específicos. Os ODSs são criados para oferecer suporte a interações em tempo real, especialmente em aplicativos de gerenciamento de relacionamento com clientes (CRM), como acessar o itinerário da sua viagem, em um site ou histórico de serviços ao ligar para o suporte ao cliente



ODS

Características do Operational Data Store:

1. Possibilitar a integração de dados de várias aplicações;
2. Ter desempenho na hora de armazenar seus dados e, principalmente, na hora de consultas sobre esses dados;
3. Ter dados de negócio atualizados e ao mesmo tempo servir para processos decisórios.



Normalização de Dados

Newton
Quem se prepara, não para.

Primeira forma normal – 1FN: uma relação está na 1FN quando os atributos são atômicos, o ue significa que as tabelas não podem ter valores repetidos, nem os atributos podem possuir mais de um valor.

Exemplo: Fornecedor = {ID + ENDEREÇO + TELEFONES}

Porém, um fornecedor poderá ter mais de um número de telefone, sendo assim o atributo TELEFONES é multivalorado. Para normalizar, é necessário:

1. Identificar a chave primária e a coluna que possui dados repetidos e removê-los;
2. Construir uma outra tabela com o atributo em questão, no caso TELEFONES. Mas não se esquecendo de fazer uma relação entre as duas tabelas: Fornecedor = {ID + ENDEREÇO} e TELEFONE (nova tabela) = {Fornecedor_ID (chave estrangeira) + TELEFONE}.

Normalização de Dados

Newton
Quem se prepara, não para.

Segunda forma normal – 2FN: primeiramente, para estar na 2FN é preciso estar também na 1FN. 2FN define que os atributos normais, ou seja, os não chave, devem depender unicamente da chave primária da tabela. Assim, como as colunas da tabela que não são dependentes dessa chave devem ser removidas da tabela principal, cria-se uma nova tabela utilizando esses dados.

Exemplo: FORNECEDOR_PRODUTO = {ID_FORNECEDOR + ID_PRODUTO + FRETE + DESCRICAO_PRODUTO}
Como se pode observar, o atributo **DESCRICAO_PRODUTO** não depende unicamente da chave primária ID_FORNECEDOR, mas somente da chave ID_PRODUTO. Para normalizar, é necessário:
1. Identificar os dados não dependentes da chave primária (nesse exemplo DESCRICAO_ PRODUTO) e removê-los;
2. Construir uma nova tabela com os dados em questão: FORNECEDOR_PRODUTO = {ID_ FORNECEDOR + ID_PRODUTO + FRETE} e PRODUTO (nova tabela) = {ID_PRODUTO + DESCRICAO_PRODUTO}.

Normalização de Dados

Newton
Quem se prepara, não para.

Terceira forma normal – 3FN: para estar na 3FN é preciso estar também na 2FN. 3FN define que todos os atributos dessa tabela devem ser funcionalmente independentes uns dos outros, ao mesmo tempo que devem ser dependentes exclusivamente da chave primária da tabela. 3NF foi projetada para melhorar o desempenho de processamento dos bancos de dados e minimizar os custos de armazenamento.

Exemplo: VENDEDOR = {ID + NOME + VALOR_SALARIO_FIXO + VALOR_PERCENTUAL_COMISSAO}

Como saber o valor da COMISSAO, vai depender da venda representada pelo valor salário fixo, logo o atributo normal VALOR_PERCENTUAL_COMISSAO é dependente do também atributo normal VALOR_SALARIO_FIXO. Para normalizar, é necessário:

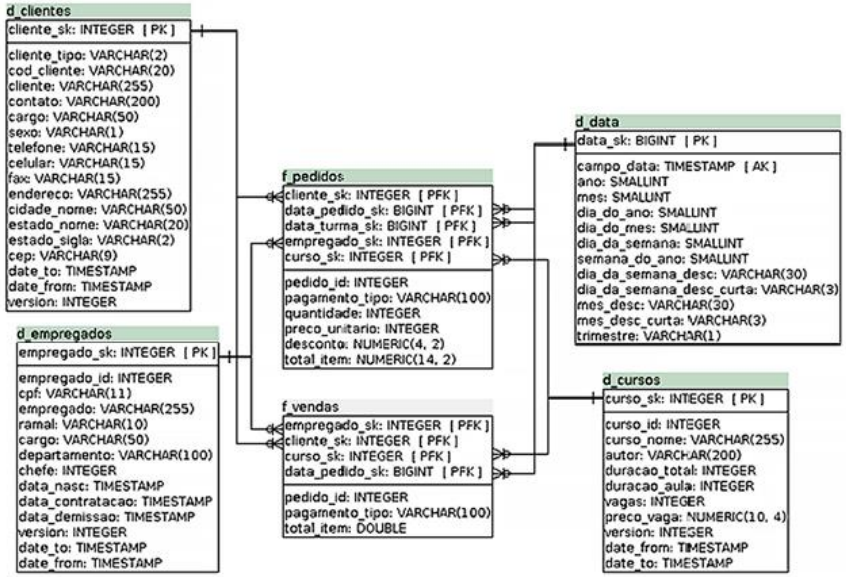
1. Identificar os dados dependentes de outros (nesse exemplo VALOR_SALARIO_FIXO);
2. Removê-los da tabela. Esses atributos poderiam ser definitivamente excluídos – deixando para a camada de negócio a responsabilidade pelo seu cálculo – ou até ser movidos para uma nova tabela e referenciar a principal (VENDEDOR).

Desnormalização

Newton
Quem se prepara, não para.

Um modelo dimensional é formado por uma tabela com uma chave composta, **denominada tabela de fatos**, e um conjunto **de tabelas menores conhecidas como tabelas de dimensão**, que **possuem chaves simples (formadas por uma única coluna).** A chave da tabela de fatos é uma combinação das chaves das tabelas de dimensão, isso faz com que a representação gráfica do modelo dimensional se assemelhe a uma estrela. Modelos dimensionais reais do mundo dos negócios geralmente possuem entre 4 e 15 dimensões.

CHAVES SUBSTITUTAS – SURROGATE KEYS (SK)



As **chaves substitutas** (surrogate keys) devem ser aplicadas em modelos dimensionais. Por exemplo, ao primeiro registro de produção é atribuída uma chave substituta do produto com o valor 1, o próximo registro do produto recebe a chave 2 do produto e assim por diante. As chaves substitutas meramente servem para unir as tabelas de dimensões à tabela de fatos.

Inicialmente, pode ser mais rápido implementar um modelo dimensional usando códigos operacionais, mas as chaves substitutas definitivamente recompensam a longo prazo.

CHAVES SUBSTITUTAS – SURROGATE KEYS (SK)

Chaves substitutas protegem o ambiente do data warehouse de mudanças operacionais e permitem que a equipe do BI mantenha o controle do ambiente, em vez de ser prejudicada pelas regras operacionais para gerar, atualizar, excluir, reciclar e reutilizar códigos de produção.

A chave substituta é o menor número inteiro possível, garantindo ao mesmo tempo que acomodará a cardinalidade futura ou o número máximo de linhas na dimensão confortavelmente. **Geralmente, o código operacional é uma sequência de caracteres alfanuméricos volumosos.**

DIMENSÕES DE MODIFICAÇÃO LENTA - SLOWLY CHANGING DIMENSION (SCD)

Newton
Quem se prepara, não para.

Embora os atributos da tabela de dimensão sejam relativamente estáticos, eles não são corrigidos para sempre. Os atributos de dimensão mudam, embora de maneira bastante lenta, ao longo do tempo. Você precisa rastrear as alterações.

É inaceitável colocar tudo na tabela de fatos ou tornar cada dimensão dependente do tempo para lidar com essas alterações. **Existem técnicas de SCDs para lidar com essas mudanças dos atributos.**

DIMENSÕES DE MODIFICAÇÃO LENTA - SLOWLY CHANGING DIMENSION (SCD)

Newton
Quem se prepara, não para.

SCD tipo 1 sem histórico: para esse tipo de dimensão que muda lentamente, você simplesmente substitui os valores de dados existentes por novos. Isso facilita a atualização da dimensão e limita o crescimento da tabela de dimensões a apenas novos registros. A desvantagem disso é que você perde o valor histórico dos dados, porque a dimensão sempre conterá os valores atuais para cada atributo. Por exemplo, você tem uma dimensão de loja que possui um atributo para uma região geográfica. Se houver uma reformulação nos limites regionais, algumas lojas poderão passar de uma região para outra.

REGISTRO ORIGINAL			
Surrogate key	**ID**	**Nome**	**Região geográfica**
123	VA-13	ACME Products	**Nordeste**
234	PA-07	Ace Products & Services	Nordeste

ALTERAÇÃO DE REGISTRO			
Surrogate key	**ID**	**Nome**	**Região geográfica**
123	VA-13	ACME Products	**Centro-Oeste**
234	PA-07	Ace Products & Services	Nordeste

DIMENSÕES DE MODIFICAÇÃO LENTA - SLOWLY CHANGING DIMENSION (SCD)

Newton
Quem se prepara, não para.

SCD tipo 2 com histórico: este é o tipo mais comum de dimensão de alteração lenta. Para esse tipo de dimensão que muda lentamente, adicione um novo registro que inclua a alteração e marque o registro antigo como inativo. Isso permite que a tabela de fatos continue a usar a versão antiga dos dados para fins de relatórios históricos, deixando os dados alterados no novo registro para impactar apenas os dados de fatos daquele ponto em diante.

REGISTRO ORIGINAL

Surrogate key	ID	Nome	Região geográfica	*Flag* registro ativo	Data início	Data fim
123	VA-13	ACME Products	**Nordeste**	SIM	**28/03/2018**	**31/12/9999**
234	PA-07	Ace Products & Services	Nordeste	SIM	08/05/2018	31/12/9999

INCLUSÃO/ALTERAÇÃO DE REGISTRO

Surrogate key	ID	Nome	Região geográfica	*Flag* registro ativo	Data início	Data fim
123	VA-13	ACME Products	Nordeste	**NÃO**	28/03/2018	**28/07/2019**
234	PA-07	Ace Products & Services	Nordeste	SIM	08/05/2018	31/12/9999
784	**VA-13**	**ACME Products**	**Centro-Oeste**	**SIM**	**29/07/2019**	**31/12/9999**

Modelagem Multidimensional

Newton
Quem se prepara, não para.

Essa técnica é utilizada em projetos de BI, aplicados nos dados relacionais. A modelagem multidimensional (**também conhecida como modelagem dimensional**) é baseada nas estruturas de **Fatos e Dimensões**. Esse modelo de trabalho é ideal para estruturação de dados em um ***Data Warehouse*** **(DW)**.

A dimensão é o que dá personalidade e qualidade aos "Fatos" ocorridos, é a dimensão que nos permite visualizar as informações por diversos aspectos.

Assim é possível estruturar os dados em cubos (junção entre dimensões e fatos) por assuntos, de forma a entregar mais resultados para o usuário que vai consumir essas informações.

Fonte: https://dadosaocubo.com/modelagem-de-dados-para-business-intelligence/

Modelagem Multidimensional

Newton
Quem se prepara, não para.

Tabela Fato: São aquelas que nos trazem fatos ou eventos que ocorreram**. Exemplos desse tipo de tabela são vendas, saldos, ordens e outros valores quantitativos**. Como característica, nossa **fTabela** normalmente **contém milhões de linhas**, e está continuamente incluindo registros. **Também é muito comum ter uma coluna com informações de data**. Outra informação muito importante é que nossa **Tabela Fato nos trará a chave conhecida como *Primary Key*, que será a chave de relação com o nosso outro tipo de tabela**. Esta tabela é o coração do relatório.

Tabela Dimensão: São tabelas que vão nos trazer pontos nos quais os fatos serão analisados. Na nossa **dTabela, teremos descrições dos eventos que foram trazidos em nossa Tabela Fato**, e não seria errado pensar em Tabela Dimensão como cadastros. Também temos nossa coluna ***Primary Key*, chave para fazer o relacionamento entre as tabelas Dimensão e Fato**, contendo informações com registros únicos como o ID. Como exemplos de dimensão temos: datas, produtos, países, clientes etc.

Fonte: https://www.dataside.com.br/dataside-community/power-bi/modelagem-basica-de-dados-1

Tabela Fato

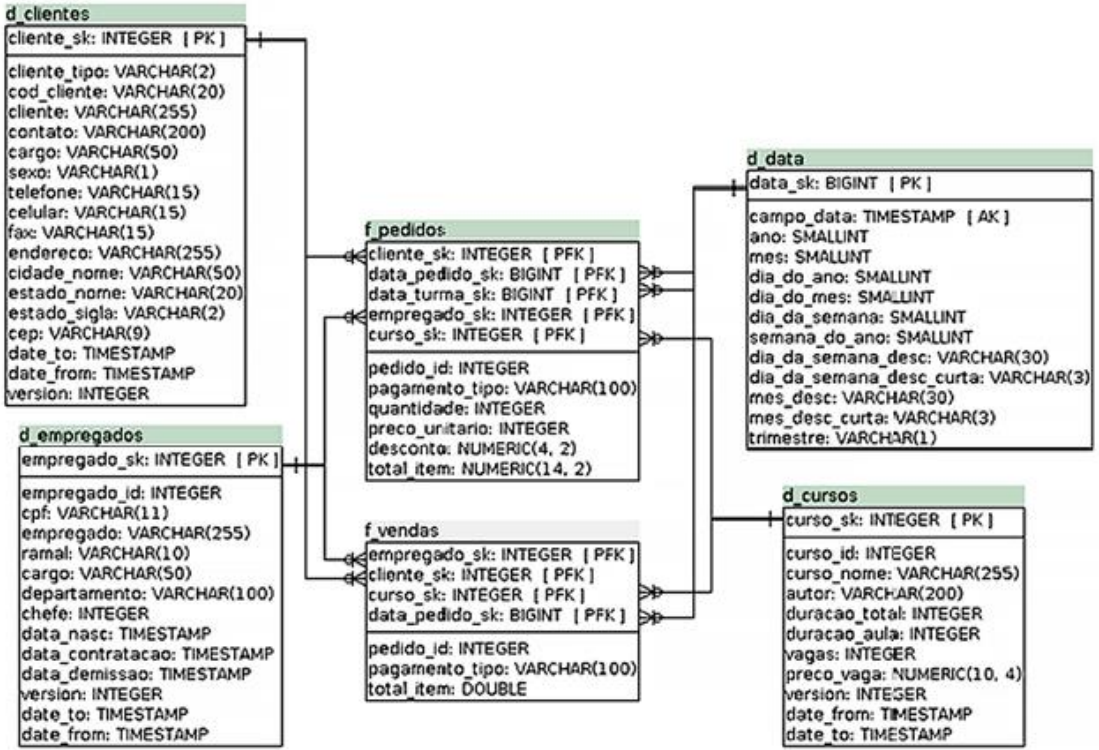


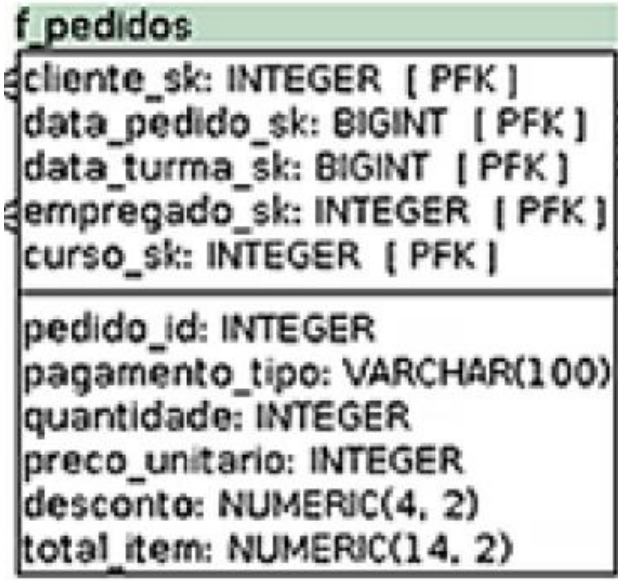
Tabela Fato: sintetizar o relacionamento existente entre as diversas dimensões. Isso ocorre porque a chave da tabela de fatos é a associação das chaves primárias das tabelas de dimensões.

Tabela Fato

Newton
Quem se prepara, não para.

Um modelo dimensional faz distinção entre fatos e atributos. Um atributo é usualmente alguma coisa que é conhecida com antecedência. Um fato é uma observação do mercado. Muitos fatos no mundo dos negócios são numéricos, embora alguns possam conter texto.

Algumas vezes, um valor numérico como “preço unitário” parece ser um atributo da dimensão curso, pois é uma constante conhecida antecipadamente. Porém, verifica-se que o atributo preço unitário sofre alteração durante o ano, o que leva a alterá-lo na fase de projeto para um fato. **Considerar quase todos os campos numéricos de pontos flutuantes como fatos.**



Atributo

Atributos são geralmente campos textos, os quais descrevem uma característica de algo tangível. Os atributos de dimensões oferecem o conteúdo da maioria das respostas solicitadas pelos usuários. Pode-se dizer que a qualidade do data warehouse é medida pela qualidade dos atributos das dimensões. Em uma tabela de dimensão, campos textos descrevem os membros de uma dimensão particular. A meta do data warehouse é criar atributos de tabelas de dimensão com as seguintes características: **eloquente, descritivo, completo, com qualidade garantida, indexado, disponível e documentado.**

Referências

- GONÇALVES, Glauber Rogério Barbieri. Sistemas de Informação. Porto Alegre: SAGAH, 2017. ISBN digital 9788595022270.
- INMON, W. H. Building the Data Warehouse. 4. ed. Indianápolis: Wiley, 2005. ISBN 0-7645-9944-5.
- MORAIS, I. S. et al. Introdução a Big Data e Internet das Coisas. Porto Alegre: SAGAH, 2018. ISBN digital 9788595027640.
- STAIR, Ralph M.; REYNOLDS, George W. Princípios de Sistemas de Informação. 11. ed. São Paulo: Cengage Learning, 2016. ISBN digital 9788522124107.
- TURBAN, Efraim et al. Business Intelligence: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009. ISBN digital 9788577804252.