

Aula 5: Dimensões e Hierarquias, Tabelas Fato e Agregações

Apresentação

Nesta aula serão apresentadas as hierarquias que podem ser construídas no modelo de dados dimensional, tanto no esquema Floco de Neve quanto no esquema Estrela. Serão apresentadas as técnicas para agregação e consolidação de dados para otimizar as consultas analíticas realizadas no Data Warehouse (DW).

Objetivos

- Examinar as possibilidades de navegação de dados com a construção de hierarquias, com as dimensões organizadas no esquema Floco de Neve ou em uma mesma dimensão no esquema Estrela;
- Descrever o processo de agregação e consolidação dos dados em tabelas Fatos com o objetivo de otimizar as consultas que serão submetidas.

Dimensões

Como já estudamos, as tabelas Dimensões contêm as descrições dos fatos ocorridos. Geralmente possuem muitas colunas de texto de baixa cardinalidade e seus atributos têm importante papel no filtro das consultas ou nas agregações dos dados contidos na tabela Fato.

Aprendemos que as dimensões podem ser normalizadas ou desnormalizadas quando elas estão normalizadas no esquema Floco de Neve, que especializa as dimensões em decomposição hierárquica. Assim, as dimensões são organizadas em níveis, em que uma se relaciona com a outra por meio da chave primária. Essa relação apresenta uma hierarquia que estabelece os níveis dentro de um contexto. A seguir, vamos entender como as hierarquias se comportam, quais são os benefícios e como elas são representadas no esquema Estrela.

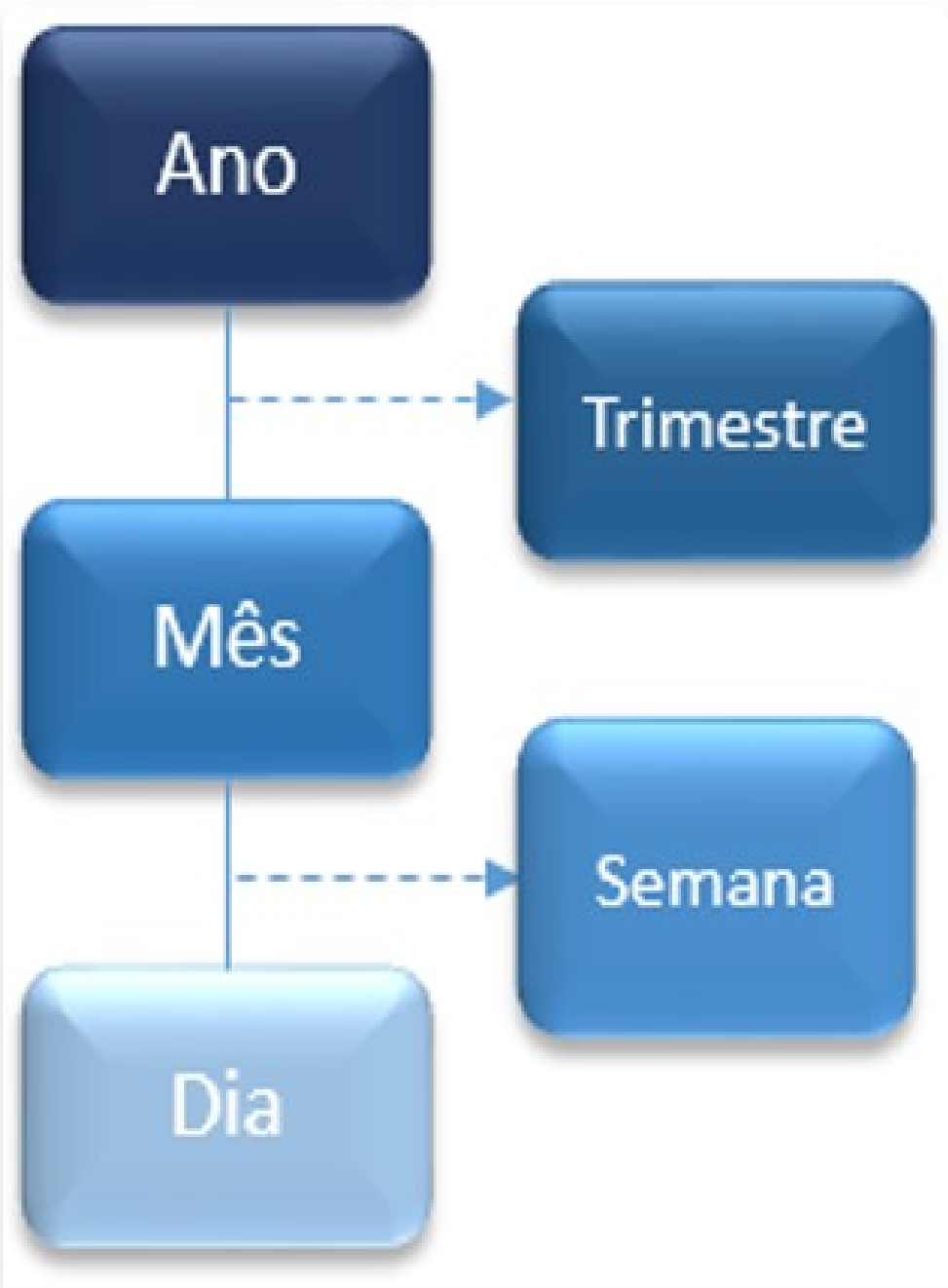
Hierarquias

Uma hierarquia estabelece a relação entre elementos de um conjunto de dados de um contexto. As hierarquias são organizadas em níveis, em que o nível mais alto é o elemento que agrupa os demais, que estão abaixo. Isso permite que os dados da tabela Fato sejam visualizados conforme a navegação na hierarquia, de modo agrupado (nível mais alto) ou detalhado (nível mais baixo).

Algumas dimensões possuem atributos que formam uma hierarquia. Esses atributos possuem um relacionamento que estabelece a ordem de prioridade dos elementos. Por exemplo, no modelo de dados dimensional do DW Supermercado, a dimensão Data possui uma hierarquia formada pelos elementos Dia, Mês e Ano.

A figura a seguir ilustra a hierarquia de Data, que é composta pelo Ano, no mais alto nível da hierarquia, o mês, no nível intermediário e o dia no nível mais detalhado. Com os dados básicos de data é possível criar elementos intermediários que agrupam o dado menos detalhado. Por exemplo, podemos agrupar os dias em semanas, os meses em bimestre, trimestre, semestre, entre outros.

Hierarquia de Data



Fonte: próprio autor

No esquema Estrela, os dados ficam desnormalizados em uma única tabela dimensão. Nesse caso, todos os atributos são colunas de uma mesma tabela. Veja o exemplo de outra hierarquia bastante utilizada nos projetos de Data Warehouse, a de Região geográfica, que relaciona os dados País, Região, Estado, Cidade e Bairro.

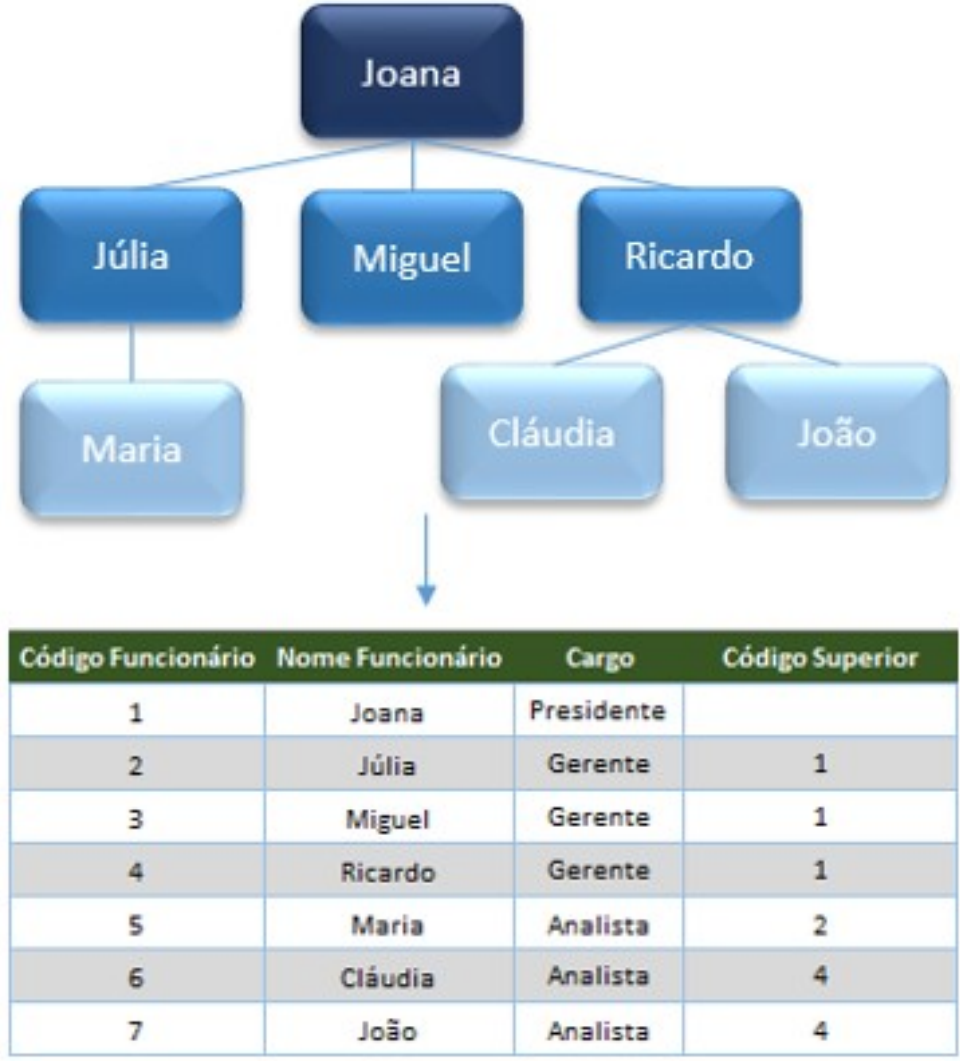
A figura mostra a Dimensão Região Geográfica e suas colunas disponibilizam toda a hierarquia de região contendo o País, que é o nível mais alto da hierarquia, a região, o estado, a cidade e o bairro, que é o nível mais granular dessa visão.

	Código Região	País	Região	Estado	Cidade	Bairro
Região Geográfica	1	Brasil	Suldeste	São Paulo	São Paulo	Morumbi
	2	Brasil	Suldeste	Rio de Janeiro	Rio de Janeiro	Centro
	3	Brasil	Suldeste	Paraná	Curitiba	Alto da Glória

Fonte: próprio autor

Hierarquia Região Geográfica

Algumas hierarquias podem ser representadas sem as repetições dos elementos nas linhas. Esse tipo de solução utiliza o autorrelacionamento na tabela, onde uma chave estrangeira aponta para a chave primária da tabela. Veja um exemplo da hierarquia de um quadro funcional. Observe que Joana é a presidente e por isso ela é o primeiro nível da hierarquia; abaixo dela estão os gerentes Júlia, que possui um subordinado, Miguel que não possui subordinado e Ricardo, que possui dois subordinados.



Fonte: próprio autor

A dimensão acomoda a relação de hierarquia por meio da coluna Código Superior, que é usada para retornar uma consulta sobre os funcionários e seus superiores. Para retornar os valores, a consulta deverá montar a hierarquia aplicando busca com recursividade, em que a coluna Código Superior será relacionada com a coluna Código Funcionário da mesma tabela.

Dica

As hierarquias são muito úteis para a navegação dos dados e possibilitam realizar as operações de Drill-Up e Drill Down, detalhando mais ou menos os dados contidos na tabela Fato, conforme vimos em aula anterior.

Carga das Dimensões

As dimensões são carregadas com os elementos contidos em tabela do banco de dados transacional. A primeira carga dos dados no Data Warehouse deve inserir todos os elementos existentes na tabela Origem, considerando algumas regras que possam existir, como carregar no DW somente os registros ativos. Um exemplo disso são clientes antigos, que, no sistema de origem, receberam status inativo por não realizarem compras nos últimos cinco anos.

Após a primeira carga dos dados, os elementos são atualizados conforme forem atualizados no sistema de origem. Normalmente, as alterações nas dimensões acontecem esporadicamente. No entanto, se pensarmos em uma grande empresa e em sua dimensão de clientes, com 30 mil clientes ativos, ela pode ser atualizada diariamente, já que o volume de dados é muito grande.

Além disso, uma dimensão desse porte pode receber novos cliente diariamente e ainda cancelar os que estão ativos.

Exemplo

Por exemplo, uma operadora de telefonia que diariamente vende pacotes de telefone para novos clientes ou vários outros migram para outra operadora e precisam ser cancelados. Os registros desses clientes precisam ser atualizados com a situação atual, com a data de cancelamento, entre outros.

Dimensões históricas

Alguns negócios precisam, além de acompanhar os fatos ao longo do tempo, manter o histórico de atributos para analisar os dados conforme a situação em um momento do tempo, mas não em relação ao fato e sim ao contexto que o descreve. Para uma organização, pode ser importante acompanhar a situação de estado civil dos clientes, não somente se estão casados, mas se quando tornaram-se clientes eram solteiros.

Para resolver essas questões há algumas soluções que podem ser aplicadas, como criar uma tabela de suporte que armazene as informações históricas dos atributos, relacionando a dimensão e quais são as alterações realizadas; para cada dimensão criar uma tabela de suporte que guarde apenas as alterações do atributo desejado; ou, ainda, na própria dimensão criar colunas para informar a data de início de vigência e data fim de vigência do registro, em que o registro com a data de fim de vigência fechada mantém os dados sem alteração e o novo registro com nova data de início de vigência recebe as alterações realizadas pelo sistema origem.

Veja um exemplo de alteração direta no registro existente (Dimensão Cliente 1) e um exemplo de alteração com datas de início e fim de vigência do registro.

Dimensão Cliente 1

Identificador Cliente	Código Cliente	Nome Cliente	Estado Civil	Data Nascimento	Data Cadastro	Situação Cliente
1	110506	João	Solteiro	10/12/1977	01/05/2000	Ativo
2	150063	Maria	Solteiro	15/04/1970	10/03/2004	Ativo
3	108024	José	Divorciado	15/06/1965	04/07/1998	Ativo

Dimensão Cliente 2

Identificador Cliente	Código Cliente	Nome Cliente	Estado Civil	Data Nascimento	Data Cadastro	Situação Cliente	Data Inicio Vigência	Data Fim Vigência
1	1	João	Solteiro	10/12/1977	01/05/2000	Ativo	01/05/2000	
2	2	Maria	Solteiro	15/04/1970	10/03/2004	Ativo	10/03/2004	
3	3	José	Solteiro	15/06/1965	04/07/1998	Inativo	04/07/1998	10/03/2001
4	3	José	Casado	15/06/1965	04/07/1998	Inativo	11/03/2001	26/06/2010
5	3	José	Divorciado	15/06/1965	04/07/1998	Ativo	27/06/2010	

Fonte: próprio autor

Dimensão Cliente

Com as colunas de data de início e fim de vigência é possível saber que do dia 11/03/2001 ao dia 26/06/2010 José constava na base de clientes com o status Casado. Esse tipo de informação, pode, por exemplo, ser útil para a oferta de produtos destinados à situação em que o cliente se encontra.

As tabelas Fatos armazenam as métricas que serão descritas pelas Dimensões. Quanto maior a granularidade da tabela Fato, maior a quantidade de registros nela contidos. O grão atômico armazenado na tabela fato garante que qualquer consulta submetida pelo usuário poderá ser respondida.

Para agregar os dados contidos na tabela fato, algumas funções são aplicadas às métricas para que os dados sejam apresentados conforme a necessidade do usuário. Essas funções realizam somas nas métricas (SUM), apresentam o menor valor ou o maior valor de uma métrica (MIN e MAX), realizam contagem (COUNT) e aplicam médias (AVG).

Saiba mais

Realizar esses cálculos em uma tabela fato com milhões de registros, no grão mais baixo, é muito oneroso para as ferramentas de visualização. O relatório ou dashboard demora um certo tempo para a exibir o resultado da consulta e isso pode causar transtornos pela demora da informação. Para revolver esse problema, podemos trabalhar com a agregação dos dados para melhorar a performance das consultas.

Tabela Fato Agregada

A agregação de dados é um resumo dos dados da tabela fato que visa um bom desempenho das consultas do Data Warehouse. Esse recurso não traz custos ao projeto e pode evitar o investimento em hardware. Assim, o custo benefício torna-se muito atraente para a melhoria de desempenho.

Atenção

A tabela Fato Agregada armazena informações pré-calculadas de acordo com nível de granularidade, que é mais alto do que a tabela Fato Transacional, que apresenta os dados no nível mais detalhado. Por esse motivo, esse tipo de tabela apresenta o volume de dados menor do que a tabela Fato Transacional.

A carga e a atualização dos dados podem ser feitas com os dados lidos diretamente do sistema origem ou das tabelas fatos transacionais. Se a origem da agregação for as tabelas Fatos Transacionais, as tabelas agregadas não precisam de tabelas temporárias para dar suporte ao processo ETL, pois os dados armazenados já estão validados. No entanto, se os dados são lidos diretamente no sistema transacional, existe a possibilidade de ocorrer problema. Então, é recomendado que a leitura seja feita com base na tabela temporária da tabela Fato.

Para criar uma tabela Fato Agregada você precisa analisar quais são os fatos submetidos ou que serão constantemente submetidos às funções de agregação e por quais atributos eles são ou serão agregados. Esse conjunto de dados é um forte candidato a virar uma tabela Fato Agregada. Caso a consulta submetida apresente os dados de forma resumida, mas contenha um atributo que não está presente na tabela Fato Agregada, a consulta é direcionada para a tabela Fato Transacional que contenha a combinação solicitada.

O exemplo a seguir apresenta os dados da tabela Fato Vendas no nível de número do pedido, produto, cliente e data da venda.

Tabela fato Vendas a Varejo

	num_pedido	sk_produto_vendas	sk_categoria_vendas	sk_cliente_vendas	sk_data_vendas	qt_produto_vendido	vl_produto_vendido
1	10003	1	6	200	20200917	5	R\$ 10,52
2	10004	2	8	100	20200917	1	R\$ 23,15
3	10008	1	6	300	20200917	3	R\$ 10,52
4	10015	2	8	200	20200917	2	R\$ 23,15
5	10110	1	6	100	20200917	10	R\$ 10,52
6	10160	1	6	300	20200917	1	R\$ 10,52

Fonte: próprio autor

Vamos construir a tabela Fato Agregada baseada em uma das análises realizadas pelos gerentes do Supermercado: quais são os produtos mais vendidos no verão?

Observe que a necessidade é avaliar a venda de produtos para construir a tabela agregada para análises referentes aos produtos. Podemos agregar os dados preservando apenas as dimensões que fazem parte do contexto da análise. Nesse caso, vamos manter o produto, a categoria do produto, a data da venda e as métricas. Veja o resultado.

	sk_produto_vendas	sk_categoria_vendas	sk_data_vendas	qt_produto_vendido	vl_produto_vendido
1	1	6	20200917	19	R\$ 10,52
2	2	8	20200917	3	R\$ 23,15

Fonte: próprio autor

Tabela fato agregada
Vendas Produto

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

Tabelas Fatos

Esse comportamento é refletido no modelo de dados dimensional que receberá o desenho da tabela agregada mantendo apenas o relacionamento com as dimensões que atenderão às análises focadas no produto. A tabela agregada agr_vendas_produto mantém apenas as chaves da dimensão Produto e da dimensão Data.

Modelo da tabela fato agregada agr_vendas_produto

agr_vendas_produto	
sk_produto_vendas:	INTEGER NOT NULL [PFK]
sk_data_vendas:	CHAR(8) NOT NULL [PFK]
qt_produto_venda:	INTEGER NOT NULL
vl_produto_venda:	DECIMAL(10, 2) NOT NULL

 Fonte: próprio autor

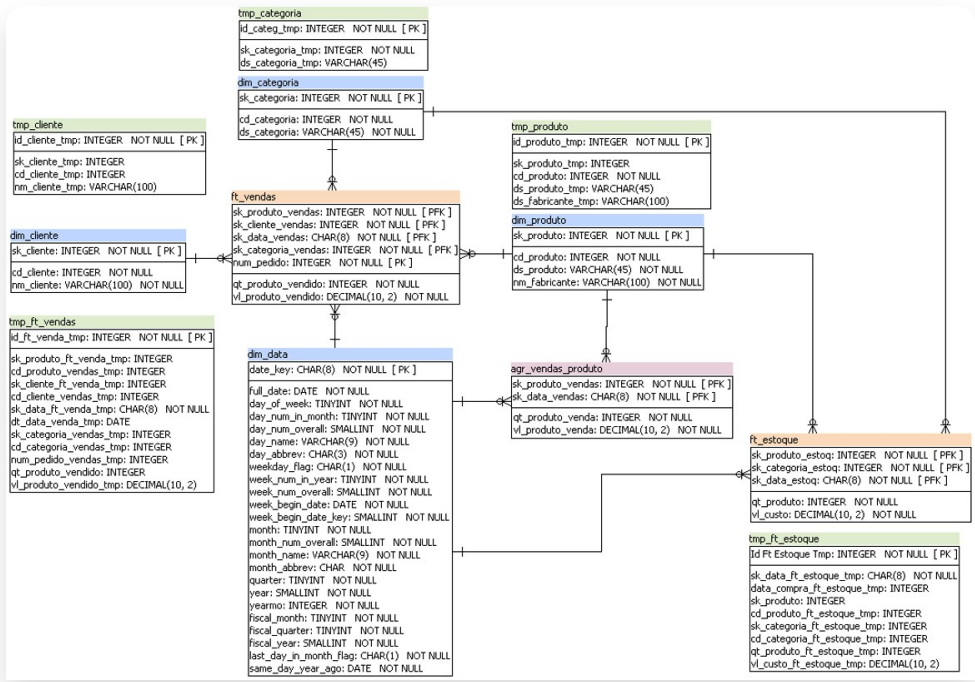
Na tabela Fato Agregada a métrica qt_produto_venda representa o somatório das quantidades de produtos que foram vendidos no grão dia e a métrica vl_produto_venda representa o somatório dos valores pagos pelos produtos no grão dia.

A tabela Fato Agregada é transparente para os usuários do DW, pois a ferramenta de visualização direciona as consultas para as agregadas sem que haja interferência e conhecimento de quem está submetendo a consulta.

Saiba mais

Ainda sobre as agregações, vimos em aula anterior que existem algumas métricas que não podem ser sumarizadas. São as métricas não aditivas ou ainda aquelas que só podem ser sumarizadas por determinadas visões, as semiaditivas. Fique sempre atento sobre quais métricas você irá aplicar à agregação para que não obtenha valores errados ou distorcidos.

Agora, adicione a tabela Fato Agregada ao modelo de dados dimensional e relacione com as dimensões Produto e Data. Para as tabelas Fato Agregadas utilize o prefixo agr_.



Fonte: próprio autor

Modelo de Dados Dimensional com
tabela Fato Agregada Vendas Produto

Tabela Fato Consolidada

Clique no botão acima.

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

Tabela Fato Consolidada

Algumas necessidades do negócio envolvem análises compostas por métricas armazenadas em diferentes tabelas Fatos. Muito semelhante à tabela Fato Agregada, a tabela Fato Consolidada agrega dados unindo dados contidos em mais de uma tabela Fato. Para que isso aconteça, os dados precisam estar no mesmo nível de granularidade de uma visão comum aos dois assuntos. Uma visão muito utilizada nesse tipo de análise é a visão data.

Para exemplificar esse conceito, vamos utilizar outra análise do cenário Supermercado: quais são os fabricantes dos produtos que oferecem maior lucro na comercialização de seus itens?

Para que seja possível avaliar o lucro referente aos produtos, é necessário consultar dados na tabela fato Vendas, onde está o valor do produto vendido, e na tabela fato Estoque, onde está o preço de custo do produto. Contudo, a tabela Fato Vendas está no grão Dia e a tabela Fato Estoque está no grão Mês. Isso significa que só será possível avaliar o lucro do produto por mês, devido à granularidade da segunda tabela Fato.

A solução para esse problema é agregar o dado da tabela Fato Venda para o mês e então consolidar com a tabela Fato Estoque. A tabela Fato Estoque é relacionada à dimensão Data sempre pelo dia 1º de cada mês, representando o mês.

Tabelas Fato Vendas e Estoque

 Fonte: próprio autor

As métricas calculadas Valor Receita Total e Valor Custo Total são adicionadas à tabela Fato Consolidada, assim como a métrica calculada Lucro, em que o Lucro é obtido do cálculo Valor Receita Total menos o Valor Custo Total.

 Fonte: próprio autor

Logo, para o produto 1, o lucro obtido no mês 09/2020 é igual a 43%.

Agora adicione a tabela consolidada ao modelo de dados dimensional do DW Supermercado e relacione-a com as dimensões adequadas. Após a inclusão das tabelas no modelo, gere o script e execute apenas o DDL das novas tabelas e das constraints no SGBD.



Estudamos nesta aula alguns conceitos importantes na construção do DW, as estruturas das hierarquias de atributos nas dimensões e as tabelas Fato de agregação e consolidação de dados.

Agora, vamos fixar o entendimento!

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

Atividades

1 - Os principais elementos do modelo de dados dimensional são as tabelas Dimensões e as tabelas Fato. Sobre as dimensões:

- a) Podem ser organizadas no esquema Floco de Neve formando uma hierarquia que possibilita a navegação dos dados.
 - b) Não disponibilizam atributos em hierarquia.
 - c) Só podem formar hierarquias utilizando o esquema Floco de Neve.
 - d) No esquema Estrela a dimensão não pode ter hierarquias.
 - e) Apenas uma dimensão do esquema Floco de Neve armazena uma hierarquia.
-

2 - A tabela Fato Agregada armazena informações pré-calculadas em nível mais alto do que a tabela Fato Transacional. Sobre as tabelas agregadas é correto afirmar que:

- a) São mais volumosas do que as tabelas Fato Transacional.
 - b) Possuem um nível de detalhamento igual ao da Fato Transacional.
 - c) São menos volumosas do que as tabelas Fato Transacional e possuem alto nível de detalhamento.
 - d) Apresenta maior nível de granularidade e menor volume de dados do que as tabelas Fato Transacional.
 - e) Apresentam maior volume de dados do que as tabelas Fato Transacional e menor nível de detalhamento.
-

3 - As tabelas Fato Consolidadas agregam dados unindo o conteúdo contido em mais de uma tabela Fato. Para isso:

- a) As tabelas Fato podem ser consolidadas sem qualquer restrição de dados.
 - b) As tabelas Fato devem possuir uma visão comum aos dois assuntos para consolidar dados de níveis de granularidade diferentes.
 - c) As tabelas Fato devem possuir uma visão comum aos dois assuntos e os dados consolidados devem estar no mesmo nível de granularidade.
 - d) Não precisa de uma visão comum aos dois assuntos, bastando apenas estarem no mesmo DM.
 - e) Todas as visões contidas nas tabelas Fatos Transacionais devem ser levadas para a tabela Fato Consolidada.
-

Jogadores informais¹

Usuários que jogam usando o equipamento de amigos ou adquirem jogos piratas ou ainda só jogam jogos gratuitos na internet.

Referências

KIMBALL, M. R. R. **The Data Warehouse Toolkit - The Definitive Guide to Dimensional Modeling**. 3. ed. Indianapolis: John Wiley Sons, 2013.

PITON, R. **Data Warehouse Passo a Passo – O guia prático de como construir um Data Warehouse do zero**. Porto Alegre: Edição do Autor, 2018.

Próxima aula

- O Processo ETL;
- As técnicas de extração de dados;
- Fontes de dados e ferramentas de ETL.

Explore mais

- O Capítulo 2 do livro de Kimball (2013) traz um resumo sobre modelagem dimensional, com seções sobre hierarquias em Dimensões e Fatos Agregados.
- O livro de Piton (2018) também aborda esses temas nas seções Dimensão hierárquica: Pai-filho e Fato Agregada.