

INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO
DE SÃO PAULO

EDSON YASSUYUKI FUKUMA

Um Estudo sobre a Adequação do Modelo Estrela para
Datawarehouses em Evolução

São Paulo

2004

EDSON YASSUYUKI FUKUMA

Um Estudo sobre a Adequação do Modelo Estrela para
Datawarehouses em Evolução

Dissertação apresentada ao Instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT, para obtenção do título de Mestre em Engenharia da Computação.

Área de Concentração: Engenharia de Software

Orientadora: Dra. Edit Grassiani de Lino Campos

São Paulo

2004

Fukuma, Edson Yassuyuki

Um estudo sobre a adequação do modelo estrela para datawarehouses em evolução / Edson Yassuyuki Fukuma. São Paulo, 2004.

92p.

Dissertação (Mestrado em Engenharia de Computação) - Instituto de Pesquisas Tecnológicas do Estado de São Paulo. Área de concentração: Engenharia de Software.

Orientador: Dra. Edit Grassiani de Lino Campos

1. Datawarehouse 2. Modelo estrela 3. Tese I. Instituto de Pesquisas Tecnológicas do Estado de São Paulo. Centro de Aperfeiçoamento Tecnológico II. Título

CDU 004.652(043)
F964e

A minha esposa Akiko
pelo seu incentivo e apoio.

Agradeço à professora Edit
pela dedicação dispensada ao me orientar
e pelos seus valiosos conselhos.

RESUMO

A maior parte da bibliografia sobre datawarehouse (DW) recomenda sua construção segundo o modelo estrela. Entretanto, publicações a respeito do que ocorre quando este DW é inserido em um ambiente em evolução ainda são escassas.

O objetivo deste estudo é mostrar a evolução de um DW desenvolvido segundo o modelo estrela e suas implicações. Essa evolução acontece a partir de alterações nas regras de negócio que podem implicar em mudanças estruturais no esquema de dados. As mudanças frequentemente afetam os aspectos de adaptabilidade (em maior grau) e legibilidade do esquema inicial do DW. As mudanças estruturais são baseadas na análise de soluções propostas por diversos autores assim como na experiência do próprio autor.

Um estudo de caso ilustra a evolução de um DW criado no esquema estrela que, durante sua evolução, sofre diversas alterações estruturais. A cada incorporação de uma alteração o esquema distancia-se cada vez mais da configuração em estrela inicial. No final do estudo de caso, o esquema resultante transforma-se em uma constelação, indicando que o modelo estrela pode não ser o mais adequado para a modelagem de um DW em evolução.

Palavras-Chave: *Datawarehouse*; Evolução; Modelo Estrela; Adaptabilidade; Legibilidade.

ABSTRACT

The most part of the bibliography about DW suggests its construction according to the Star Model. On the other hand, issues describing the consequences about the use of a DW in an evolutionary context is very rare.

The purpose of this work is to demonstrate the consequences of using the star model in a DW in evolution. The evolution occurs through the business rules modification which can imply structural changes on the data scheme. The structural changes frequently affect the adaptability and the legibility of the initial scheme. The structural changes are based on the analysis of solutions proposals for several authors as well as in the experience of the author of this work.

A case study shows a DW modeled as a star schema. As changes progressively occur, the initial star schema degenerates. In the end, the final schema is transformed into a constellation, which shows that the star model may not be the most suitable for the development of a DW.

Keywords: Datawarehouse; Evolution; Star Model; Adaptability; Legibility

LISTA DE FIGURAS

Figura 2.1 – Exemplo de esquema Estrela	8
Figura 2.2 – Exemplo de Esquema Floco de Neve	11
Figura 2.3 – Exemplo de Esquema Constelação	12
Figura 3.1 – Exemplo do problema de relacionamento M:N	20
Figura 3.2 – Criação de tabelas pontes	22
Figura 3.3 – Criação de dimensão com mapa binário	25
Figura 3.4 – Indicação de grupos dentro da tabela fato	27
Figura 3.5 – Indicação de grupos em uma nova tabela fato	29
Figura 3.6 – Inclusão das mini-dimensões na tabela fato	30
Figura 3.7 – Combinações das instâncias da dimensão original em listas	32
Figura 3.8 – Exemplo do detalhamento dos fatos e dimensões	34
Figura 3.9 – Detalhamento de uma dimensão	35
Figura 3.10 – Exemplo para armazenamento de dados históricos na dimensão	37
Figura 3.11 – Armazenamento dos dados históricos na própria dimensão	38
Figura 3.12 – Armazenamento dos dados históricos em outra tabela fato	39
Figura 3.13 – Exemplo do problema de inclusão de uma nova dimensão	41
Figura 3.14 – Relacionamento da tabela fato com uma nova dimensão	42
Figura 3.15 – Relacionamento da tabela fato VENDA com a dimensão OPERADORA_CARTÃO	43
Figura 3.16 – Inclusão de uma nova tabela fato com a nova dimensão	44
Figura 3.17 – Inclusão de uma nova tabela fato relacionada com a tab. fato original	46
Figura 4.1 – DW de uma empresa de assistência médica.	50
Figura 4.2 – Inclusão dos atributos de internação diretamente na tabela fato	51
Figura 4.3 – Inclusão das informações de internação	52
Figura 4.4 – Inclusão das informações sobre pagamento	53
Figura 4.5 – Inclusão das informações do plano do associado	54
Figura 4.6 – Inclusão do histórico do plano do associado	56

Figura 4.7 – Detalhamento do pagamento com a inclusão da dimensão serviço.....	57
Figura 4.8 – Relacionamento da dimensão SERVIÇO com a tabela fato ASSOCIADO_PLANO	58
Figura 4.9 – Detalhamento do atendimento com a inclusão da dimensão serviço	59
Figura 4.10 – Inclusão das informações sobre o procedimento	60
Figura 4.11 – Tabela fato ATENDIMENTO e relacionadas	62
Figura 4.12 – Solução do relacionamento M:N entre a tabela fato e dimensão.....	63
Figura 4.13 – Esquema resultante do estudo de caso.....	64
Figura 5.1 – Esquema resultante do estudo de caso sem os atributos.....	68
Figura 5.2 – Esquema ER.....	69

LISTA DE TABELAS

Tabela 3.1 – Conteúdo da tabela fato ATENDIMENTO.....	22
Tabela 3.2 – Conteúdo da tabela ponte	
ELEMENTOS_DO_GRUPO_DIAGNÓSTICO	23
Tabela 3.3 – Valores por diagnóstico sem utilizar o fator de peso	23
Tabela 3.4 – Valores por diagnóstico utilizando o fator de peso.....	24
Tabela 3.5 – Conteúdo da tabela GRUPO_DIAGNÓSTICO_BINÁRIO da fig. 3.3	32
Tabela 3.6 – Conteúdo da tabela GRUPO_DIAGNÓSTICO_LISTA da fig. 3.7	32
Tabela 3.7 – Conteúdo da tabela fato VENDA da fig. 3.16.....	43
Tabela 4.1 – Relatório do número de associados emitido em abril	55
Tabela 4.2 – Relatório do número de associados emitido em maio.....	55

LISTA DE ABREVIATURAS

1:1	Relacionamento de 1 para 1
1:N	Relacionamento de 1 para N
DM	<i>Datamart</i>
DW	<i>Datawarehouse</i>
ER	Entidade-Relacionamento
FK	<i>Foreign Key</i>
M:N	Relacionamento de Muitos para Muitos
OLAP	<i>Online Analytical Processing</i>
OLTP	<i>Online Transaction Processing</i>
PK	<i>Primary Key</i>
SQL	<i>Structured Query Language</i>

SUMÁRIO

RESUMO

ABSTRACT

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

CAPITULO 1 – INTRODUÇÃO	1
1.1 – Motivação.....	1
1.2 – Objetivos	2
1.3 – Contribuições Esperadas	2
1.4 – Metodologia	3
1.5 – Organização do Trabalho	3
CAPITULO 2 – CONCEITOS PRINCIPAIS E ESTADO DA ARTE	5
2.1 – Introdução	5
2.2 – <i>Datawarehouse</i>	5
2.2.1 – <i>Datamart</i>	7
2.3 – OLTP e OLAP	7
2.4 – Modelo Estrela	8
2.4.1 – Tabela Fato.....	9
2.4.2 – Tabela Dimensão.....	10
2.4.3 – Métrica	10
2.4.4 – Modelo Floco de Neve.....	11
2.4.5 – Modelo Constelação.....	11
2.5 – Aspectos de Qualidade em um DW	12
2.5.1 – Adaptabilidade	13
2.5.2 – Legibilidade.....	13

2.5.3 – Eficiência, Conformidade e Custo	15
2.6 – Evolução de Sistemas de Informação	15
2.6.1 – Motivos que causam mudanças nas regras de negócio	16
2.6.2 – Novas necessidades geradas pela alteração nas regras de negócio	17
2.7 – Resumo e Considerações Gerais	18
CAPITULO 3 – ANÁLISE DA ADAPTABILIDADE E LEGIBILIDADE DO MODELO ESTRELA SUBMETIDO A ALTERAÇÕES.....	19
3.1 - Introdução	19
3.2 – Relacionamento M:N entre a tabela fato e dimensão.....	19
3.2.1 – Apresentação do problema	20
3.2.2 – Soluções para o relacionamento M:N entre a tabela fato e a dimensão.....	21
3.2.2.1 – Criação de tabelas ponte (Kimball, 1997).....	21
3.2.2.2 – Criação de uma nova tabela dimensão contendo atributos binários (Song <i>et al</i> , 2001)	25
3.2.2.3 – Indicação de grupos dentro da tabela fato (Song <i>et al</i> , 2001).....	26
3.2.2.4 – Indicação de grupos em uma nova tabela fato (Song <i>et al</i> , 2001)	28
3.2.2.5 – Pulverização da dimensão em N ‘mini-dimensões’ (Pedersen, 1999a).....	30
3.2.2.6 – Combinação das instâncias da dimensão em uma lista (Song <i>et al</i> , 2001)..	31
3.2.3 – Resumo das soluções	33
3.3 – Detalhamento de dimensões.....	34
3.3.1 – Apresentação do problema	34
3.3.2 – Solução para o detalhamento de dimensões.....	35
3.3.2.1 – Detalhamento da dimensão	35
3.3.3 – Resumo das soluções	36
3.4 – Armazenamento de dados históricos na tabela dimensão	36
3.4.1 – Apresentação do problema	36
3.4.2 – Soluções para o armazenamento de dados históricos na tabela dimensão.....	37
3.4.2.1 – A cada atualização de um atributo da dimensão gera-se uma nova instância na mesma tabela dimensão (Kimball, 1996).....	37
3.4.2.2 – Criação de uma nova tabela fato para armazenar os dados históricos	39

3.4.3 – Resumo das soluções	40
3.5 – Inclusão de uma nova dimensão	40
3.5.1 – Apresentação do problema	41
3.5.2 – Soluções para a inclusão de uma nova dimensão	41
3.5.2.1 – Relacionamento da tabela fato com uma nova dimensão	41
3.5.2.2 – Criação de uma nova tabela fato relacionada com a nova dimensão	42
3.5.2.3 – Criação de uma nova tabela fato relacionada com a tabela fato original e a nova dimensão	45
3.5.3 – Resumo das soluções	46
3.6 – Resumo e Considerações Gerais	47
CAPITULO 4 – ESTUDO DE CASO	49
4.1 – Introdução	49
4.2 – Descrição do DW existente	49
4.3 – Necessidades de evolução	50
4.3.1 – Mudança na legislação	50
4.3.2 – Inserção de informações sobre pagamento	53
4.3.3 – Determinação do plano do associado	54
4.3.4 – Necessidade de manter o histórico dos planos do associado	55
4.3.5 – Inclusão de um novo serviço	57
4.3.6 – Detalhamento do atendimento	59
4.3.7 – Determinação das doenças mais diagnosticadas	61
4.4 – O DW após a evolução	63
4.5 – Resumo e Considerações Gerais	64
CAPITULO 5 – ANÁLISE DOS RESULTADOS	66
5.1 – Introdução	66
5.2 – Comentários sobre a adaptabilidade do modelo estrela	66
5.3 – Comentários sobre a legibilidade do esquema estrela	67
5.4 – Análise da constelação	68
5.5 – Resumo e Considerações Finais	70

CAPITULO 6 – CONCLUSÃO	71
6.1 – Resumo.....	71
6.2 – Contribuições	72
6.3 – Sugestões para Pesquisas Futuras	73
REFERÊNCIAS BIBLIOGRÁFICAS.....	74
APÊNDICE 1 – SIMBOLOGIA DOS ESQUEMAS DIMENSIONAIS	
APÊNDICE 2 – SIMBOLOGIA DOS ESQUEMAS ER	

CAPITULO 1 – INTRODUÇÃO

1.1 – Motivação

Atualmente difunde-se que um *datawarehouse* (DW) deve ser construído segundo o modelo estrela (Krippendorf e Song, 1997), sendo esta posição defendida por diversos autores, entre eles Kimball (1997), Corey *et al* (2001), Kelly (2002), Winsberg (1996), Moody e Kortink (2000), Han e Kamber (2001), Schouten (1999) e Humphries, Hawkins e Dy (1999), devido principalmente à simplicidade e facilidade de utilização do esquema. Esta simplicidade é decorrente de sua estrutura baseada em uma tabela fato e suas respectivas dimensões (Ballard *et al*, 1998; Sign, 2001; Pedersen e Jensen, 1998; Winsberg, 1996; Chaudhuri e Dayal, 1997).

Quando um DW construído segundo o modelo estrela é submetido a alterações decorrentes de mudanças nas regras de negócio, pode ocorrer a necessidade de uma transformação estrutural que nem sempre é suportada por este tipo de modelo. Esta transformação normalmente distorce o esquema e o faz perder, ou pelo menos degenerar, a sua simplicidade. Um agravante para este caso é que mudanças em um DW, assim como o seu comportamento em relação a estas mudanças, costumam ser abordadas de forma incipiente na maior parte da bibliografia existente na área.

Alguns autores, no entanto, sugerem que a melhor maneira de se projetar um DW é utilizando o modelo Entidade-Relacionamento (ER). Esta sugestão baseia-se no fato que o modelo ER é mais flexível que a estrela (Inmon, 2000; Song *et al*, 2001; Krippendorf e Song, 1997), uma vez que o ER não se limita a ser visualizado em uma configuração específica como no caso da estrela.

O problema de pesquisa aqui levantado é verificar se o modelo estrela é o mais indicado para a construção de um DW.

1.2 – Objetivos

Este trabalho tem como objetivo mostrar a transformação de um DW, construído inicialmente segundo o modelo estrela, em um ambiente em evolução, sabendo-se que a maioria dos DW reais sofrem adaptações com o passar do tempo.

Inicialmente são apresentados os motivos mais comuns que podem ocasionar alterações nas regras de negócio. Essas alterações, por sua vez, provocam diferentes necessidades que devem ser refletidas no DW.

Diversos autores, assim como o autor deste trabalho, apresentam soluções para estas necessidades. A cada solução proposta geram-se mudanças estruturais e análises são feitas para verificar o quanto essas mudanças interferem na configuração do esquema estrela, e se ela prejudica a compreensão do esquema pelo usuário.

Um estudo de caso é analisado para acompanhar o comportamento do DW construído com base no modelo estrela em um ambiente em evolução, onde as soluções apresentadas são utilizadas. Após todas as alterações serem implementadas, é analisado o esquema resultante deste estudo.

Uma análise verificando se o esquema estrela é o mais adequado para ser utilizado em um DW em evolução termina este trabalho.

1.3 – Contribuições Esperadas

Este trabalho apresenta e comenta soluções para algumas necessidades de mudanças estruturais que podem ocorrer durante a evolução de um DW. Os comentários destas soluções visam auxiliar o analista a escolher as que melhor se aplicam ao seu projeto e que menos interfiram na configuração em estrela do esquema.

O estudo de caso mostra como o esquema de um DW, construído segundo o modelo estrela, cria artificios para se adequar a cada mudança na regra de negócio. Neste estudo, também é apresentado e analisado o esquema final do DW após sofrer todas as alterações a que foi submetido.

A principal contribuição deste trabalho é mostrar se o modelo estrela é a melhor opção para se construir um DW que geralmente sofre transformações ao longo do tempo.

1.4 – Metodologia

A metodologia para o desenvolvimento desta dissertação baseia-se na literatura sobre DW e nas avaliações dos argumentos que defendem ou criticam o modelo estrela para se construir um DW. Também são analisados os trabalhos que propõem soluções à implementação de mudanças estruturais, principalmente sob o ponto de vista de quanto estas soluções interferem na configuração do esquema e na facilidade de compreensão pelo usuário.

A modelagem de um DW real no esquema estrela também serve de base para a elaboração deste trabalho. Neste DW, são observadas as alterações que ocorrem ao longo do tempo nas regras de negócio e analisadas as suas conseqüências no esquema do DW. Além do caso real, é desenvolvido um estudo de caso que mostra as alterações estruturais que ocorrem no esquema quando submetido a um ambiente em evolução.

1.5 – Organização do Trabalho

O trabalho está organizado da seguinte forma:

- Capítulo 2 – Conceitos Principais e Estado da Arte
 - Conceituação de *datawarehouse* (DW);
 - Modelagem Dimensional, principais características e principais modelos (estrela, floco de neve e constelação);
 - Aspectos de qualidade de um DW; e
 - Evolução de Sistemas de Informação.
-

-
- Capítulo 3 – Análise da Adaptabilidade e Legibilidade do modelo estrela submetido a alterações.
 - Apresentação das novas necessidades ocasionadas pela alteração nas regras de negócio;
 - Sugestões de soluções para as novas necessidades, apresentando as respectivas mudanças estruturais; e
 - Conseqüências das mudanças estruturais sobre a configuração em estrela do esquema e na facilidade de compreensão do esquema pelo usuário.
 - Capítulo 4 – Estudo de Caso.
 - Apresentação de um estudo de caso de uma empresa de assistência médica que possui um DW desenvolvido segundo o modelo estrela;
 - Alterações nas regras de negócio e a conseqüente adequação do esquema às novas regras; e
 - Análise do esquema resultante após a execução das alterações.
 - Capítulo 5 – Análise dos Resultados.
 - Comentários sobre a adaptabilidade e legibilidade de um esquema estrela em geral; e
 - Análise do esquema final do estudo de caso após ser submetido a alterações.
 - Capítulo 6 – Conclusão.
 - Apresentação de um breve resumo;
 - Descrição das contribuições do trabalho; e
 - Sugestões para futuras pesquisas.
-

CAPITULO 2 – CONCEITOS PRINCIPAIS E ESTADO DA ARTE

2.1 – Introdução

Neste capítulo são apresentados os conceitos de *Datawarehouse* (DW), *Datamart* (DM), modelos dimensionais (estrela, floco de neve e constelação) e os principais aspectos de qualidade referentes ao DW utilizados neste trabalho (adaptabilidade e legibilidade).

Ao final do capítulo apresenta-se a principal questão a ser investigada: as novas necessidades que o DW deve contemplar (ou problemas a resolver) devido à evolução do DW.

2.2 – *Datawarehouse*

O DW apareceu inicialmente na década de 1980 (Winsberg, 1996; Ballard *et al*, 1998; Maier, 1999; Gupta, 1997; Moody e Kortink, 2000) como uma ferramenta de auxílio à tomada de decisões nas empresas.

O DW é um “armazém de dados” que guarda informações que auxiliam na tomada de decisões. Normalmente é construído em um base de dados a parte do banco de dados operacional da corporação (Chaudhuri e Dayal, 1997; Ballard *et al*, 1998). Para que essa base seja considerada um DW (Inmon, 1996), o conjunto de dados deve ser:

- orientado por assunto;
 - integrado;
 - não volátil;
 - variante no tempo;
 - útil no apoio a decisões gerenciais.
-

A **orientação por assunto** significa que o DW aborda as principais necessidades da organização por temas (Inmon, 2000a, Pedersen e Jensen, 1999a; Corey *et al*, 2001; Sign, 2001; Chaudhuri e Dayal, 1997).

Dados integrados significa o agrupamento, no DW, de todos os dados que uma organização possui sobre um determinado assunto (Inmon, 2000a, Pedersen e Jensen, 1999a; Corey *et al*, 2001; Sign, 2001; Chaudhuri e Dayal, 1997). Por exemplo, uma empresa pode ter três sistemas de vendas independentes, em plataformas diferentes, com nomes e códigos diferentes entre si. Integrar estes dados é tratá-los, agrupá-los e carregá-los no DW, sendo que, quanto mais sistemas existirem, maior será a dificuldade em realizar esta integração.

Não volátil quer dizer que o dado, uma vez inserido no DW, não se altera mais (Inmon, 2000a, Pedersen e Jensen, 1999a; Corey *et al*, 2001; Sign, 2001; Chaudhuri e Dayal, 1997). Por exemplo, uma venda realizada no passado terá seu valor armazenado no DW e não pode ser modificado. Resumidamente, uma informação só pode ser inserida e consultada do DW, não podendo ser atualizada ou excluída.

Variante no tempo significa que ao guardar a informação no DW, é registrado a referência deste dado. Esta variante no tempo é fundamentalmente importante quando se deseja realizar análises de tendências ao longo do tempo, como histórico ou previsão de vendas (Inmon, 2000a, Pedersen e Jensen, 1999a; Corey *et al*, 2001; Sign, 2001; Chaudhuri e Dayal, 1997).

Servir de apoio a decisões gerenciais é uma das principais características do DW, pois as informações armazenadas devem auxiliar a corporação nas suas decisões, sendo os dados utilizados de maneira estratégica por qualquer nível gerencial da empresa (Inmon, 2000a, Pedersen e Jensen, 1999a; Corey *et al*, 2001; Sign, 2001; Chaudhuri e Dayal, 1997).

2.2.1 – Datamart

Um conceito muito próximo do DW é o de *Datamart* (DM), que também é um conjunto de dados integrado, não volátil, variante no tempo e que serve de apoio a decisões gerenciais. A diferença é que o DM é departamental e os seus dados são provenientes do DW (Corey *et al*, 2001; Han e Kamber, 2001; Winsberg, 1996, Moody e Kortink, 2000).

Resumidamente, o DM é um depósito de dados especializado construído através de tabelas ou de visões.

2.3 – OLTP e OLAP

A principal diferença entre sistemas OLTP (*Online Transaction Processing*) e sistemas OLAP (*Online Analytical Processing*) é o fato dos OLTP serem utilizados para processar transações operacionais do dia a dia das empresas, enquanto que sistemas OLAP servem de suporte à decisões gerenciais (Chaudhuri e Dayal, 1997; Schouten, 1999; Han e Kamber, 2001; DuMoulin, 2001; Tierstein, 2001).

Em sistemas OLTP processos e/ou ferramentas manipulam dados operacionais através de transações isoladas, utilizando dados simples (não agrupados) em operações de tempo real (Chaudhuri e Dayal, 1997). São sistemas utilizados de forma previsível e repetidamente, normalmente para a entrada e atualização de dados (DuMoulin, 2001). As transações são geralmente realizadas por consultas simples e pré definidas, testadas e otimizadas pela equipe técnica (Kelly, 1998; Corey *et al*, 2001). Exemplos de sistemas OLTP podem ser os que controlam transações bancárias de crédito e débito ou o agendamento de consultas médicas. Normalmente são dos sistemas OLTP que se originam os dados que alimentarão o DW.

Em sistemas OLAP processos e/ou ferramentas extraem e analisam dados gerenciais. Diferentemente do OLTP, o OLAP é utilizado para analisar dados sumarizados, em atividades de pesquisa (DuMoulin, 2001). Sistemas OLAP são

projetados para serem eficientes em pesquisas e flexíveis na geração de relatórios (DuMoulin, 2001).

Também, diferentemente dos sistemas OLTP, as análises dos sistemas OLAP são normalmente realizadas por consultas que o usuário define e constrói a medida de sua necessidade de informação e de maneira aleatória, o que dificulta sua otimização (Kelly, 1998; SyBase, 1998; Chaudhuri e Dayal, 1997).

2.4 – Modelo Estrela

O modelo estrela é o mais popular entre os modelos dimensionais e sua principal característica é a de ser facilmente compreendida pelo usuário (Dowling, Schuff e Louis, 2001). Sua aparência lembra uma estrela, com apenas uma tabela fato no centro e as dimensões em volta (Ballard *et al*, 1998; Sign, 2001; Pedersen e Jensen, 1998; Winsberg, 1996; Chaudhuri e Dayal, 1997). Um exemplo deste modelo é o mostrado na figura 2.1, onde a tabela COMPRA é a tabela fato e as demais são as dimensões (No apêndice 1 encontra-se uma explicação sobre a simbologia utilizada nos esquemas dimensionais).

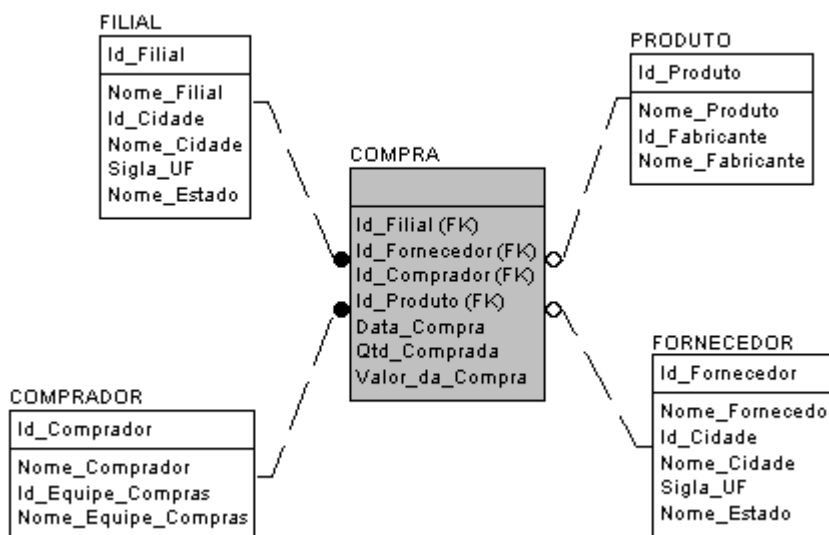


Figura 2.1 – Exemplo de esquema Estrela

Uma característica importante do modelo estrela são as dimensões desnormalizadas. Uma tabela desnormalizada é a que apresenta duas ou mais tabelas em apenas uma, redundando, assim, algumas das suas informações (Gupta, 1997; Winsberg, 1996; Sybase, 1998; Kelly, 1998). Na figura 2.1 pode-se observar que todas as dimensões estão desnormalizadas. Como exemplo, na dimensão FORNECEDOR, caso existam mais de um fornecedor na mesma cidade os dados dos atributos Id_Cidade e Nome_Cidade serão redundados. Apesar da redundância parecer uma desvantagem (ao aumentar o espaço utilizado para armazenar a mesma quantidade de informação), ela melhora a compreensão pelo usuário e a performance (velocidade) na extração de dados (Chaudhuri e Dayal, 1997; Winsberg, 1996; Gupta, 1997; Dowling, Schuff e Louis, 2001).

O modelo estrela apresenta ainda uma outra característica, que é normalmente ser focado em apenas um assunto.

Como visto, o modelo estrela é definido através de três conceitos básicos:

- a tabela fato;
- a dimensão, e
- a métrica.

Estes conceitos são definidos a seguir.

2.4.1 – Tabela Fato

A tabela fato é uma coleção de itens, fatos ou métricas que caracterizam dados de interesse do negócio (Husemann e Lechtenborger, Vossen, 2000; Ballard *et al*, 1998; Krippendorf e Song, 1997; Humphries, Hawkins e Dy, 1999; Kelly, 1998; Sybase, 1998). Nesta tabela se encontram as informações a serem analisadas. Exemplos de tabela fato podem ser: pedido, atendimento médico, etc.

Um aspecto importante na tabela fato é o nível de detalhamento dos seus dados. Essa característica, conhecida como granularidade (Humphries, Hawkins e

Dy, 1999; Ballard *et al*, 1998), pode ser grossa (menos detalhada) ou fina (mais detalhada). Ao definir esta característica, por consequência, determina-se também o fator de crescimento da tabela. Quanto mais detalhada for a informação, maior será o crescimento da tabela, uma vez que mais dados serão armazenados. Por exemplo, no caso de uma compra no supermercado, pode-se tanto armazenar o total da nota fiscal (menos detalhada), como armazenar todos os preços dos itens que o cliente comprou (mais detalhada). A boa prática não recomenda a existência de instâncias com granularidades diferentes na mesma tabela fato (Ballard *et al*, 1998).

2.4.2 – Tabela Dimensão

Nas tabelas dimensão (ou simplesmente dimensões) são estabelecidos os contextos em que um fato é analisado (Humphries, Hawkins e Dy, 1999; Ballard *et al*, 1998; Krippendorf e Song, 1997; Kelly, 1998; Sybase, 1998). Exemplos de dimensões podem ser: produto, filial, cliente, etc.

Um aspecto da modelagem dimensional é o relacionamento entre uma tabela fato e a dimensão que é de 1:N (um para muitos).

Em particular, a dimensão temporal é importante em um DW quando se deseja armazenar os acontecimentos no tempo, permitindo uma análise histórica. Esta dimensão pode ser representada em dias, meses, segundos ou em qualquer outra escala de tempo (Kelly, 1998; Humphries, Hawkins e Dy, 1999).

2.4.3 – Métrica

Métricas são as representações numéricas do fato que se deseja armazenar. A métrica pode ser o total de vendas, o preço da consulta médica, etc. (Ballard *et al*, 1998; Krippendorf e Song, 1997; Kelly, 1998; Sybase, 1998).

2.4.4 – Modelo Floco de Neve

O modelo Floco de Neve é uma variante do modelo estrela, no qual uma ou mais de suas dimensões são normalizadas, como mostrado na figura 2.2. Nesta figura, a tabela fato é a VENDA e o exemplo de dimensão normalizada é VENDEDOR e EQUIPE_VENDAS (Han e Kamber, 2001; Pedersen e Jensen, 1998; Ballard *et al*, 1998; Chaudhuri e Dayal, 1997; Kelly, 1998; Husemann e Lechtenborger, Vossen, 2000; Krippendorf e Song, 1997).

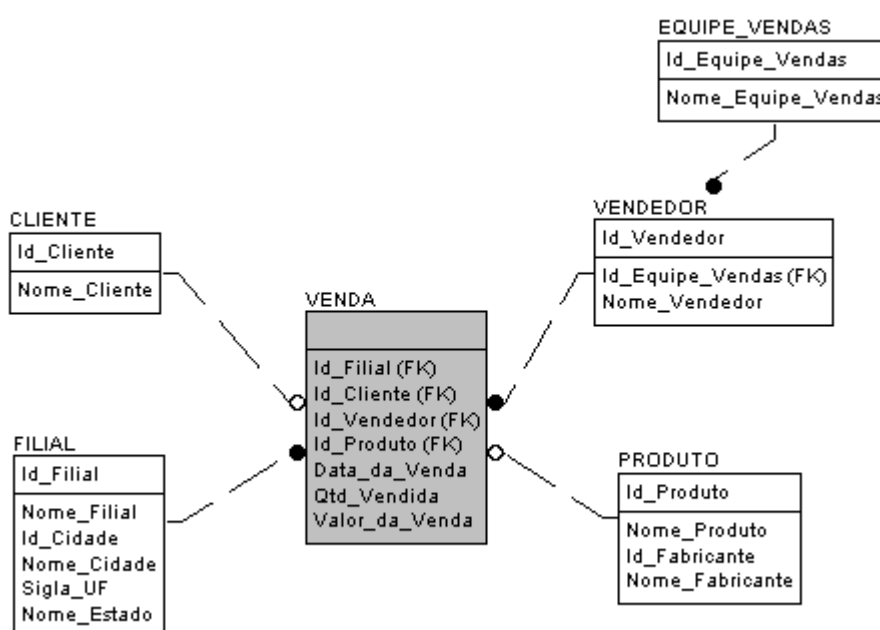


Figura 2.2 – Exemplo de Esquema Floco de Neve

O floco de neve, assim como a estrela, possui a característica de normalmente ser focado em apenas um assunto.

2.4.5 – Modelo Constelação

O último modelo apresentado é a constelação, que permite a representação de uma combinação dos esquemas estrela e/ou flocos de neve. (Han e Kamber, 2001; Chaudhuri e Dayal, 1997; Humphries, Hawkins e Dy, 1999). Na figura 2.3 observa-

se que existem duas tabelas fato, a VENDA e a COMPRA, sendo que algumas dimensões podem ser compartilhadas, como no caso FILIAL e PRODUTO.

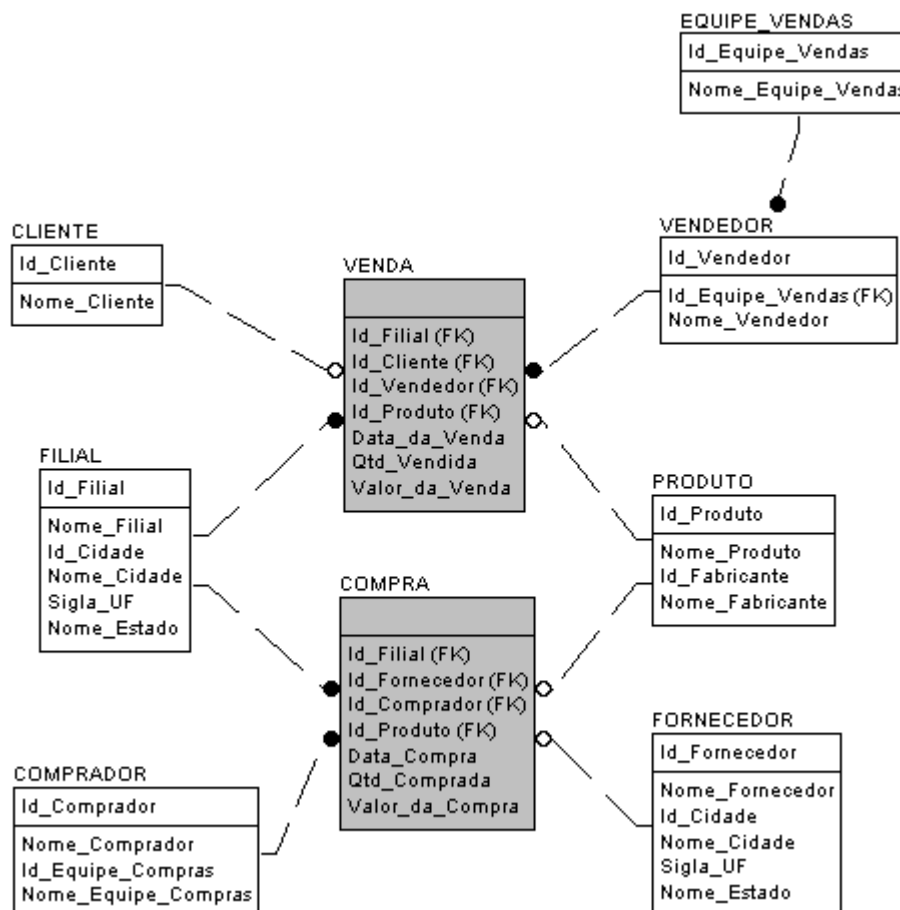


Figura 2.3 – Exemplo de Esquema Constelação

O modelo constelação, ao contrário da estrela e do floco de neve, possui a característica de focar em mais de um assunto.

2.5 – Aspectos de Qualidade em um DW

No desenvolvimento de um DW os aspectos de qualidade desempenham um papel importante para o seu sucesso. Dentre os diversos aspectos de qualidade

existentes, os que são considerados muito importantes para um DW são (Vassiliadis, 2000; Jarke *et al*, 1999; Hyatt e Rosenberg, 1996):

- a adaptabilidade,
- a legibilidade,
- a eficiência,
- a conformidade, e
- o custo.

O escopo deste trabalho abrange o estudo da adaptabilidade e da legibilidade.

2.5.1 – Adaptabilidade

A adaptabilidade é a qualidade que o modelo tem de manter as suas características, independentemente das mudanças a que é submetido.

No caso do modelo estrela, esta qualidade corresponde à inalterabilidade das seguintes características:

- ter apenas uma tabela fato;
- ter relacionamentos 1:N entre a tabela fato e dimensão; e
- ter dimensões desnormalizadas.

2.5.2 – Legibilidade

Legibilidade é o aspecto que se refere ao grau de facilidade com que o usuário consegue entender o esquema de dados.

Sobre este aspecto existe um artigo (Dowling, Schuff e Louis, 2001) bastante interessante onde é relatado um experimento comparando a legibilidade entre um esquema estrela e um esquema ER (Entidade-Relacionamento).

O estudo foi baseado em duas teses:

Tese 1: “Os usuários lembrarão melhor dos relacionamentos em um esquema estrela do que em um esquema ER”; e

Tese 2: “Os usuários lembrarão melhor das tabelas em um esquema estrela do que em um esquema ER”.

Para este experimento foram criados quatro esquemas: dois simples e dois complexos (os conceitos ‘simples’ e ‘complexos’ foram definidos a partir da quantidade de tabelas e relacionamentos, ou seja, quanto mais tabelas e relacionamentos mais complexos são os esquemas). Os esquemas eram:

- Uma estrela simples com 1 tabela fato e 9 dimensões (com um total de 9 relacionamentos);
- Um ER simples com 10 tabelas e 11 relacionamentos;
- Uma estrela complexa com 1 tabela fato e 18 dimensões (com um total de 18 relacionamentos); e
- Um ER complexo com 16 tabelas e 19 relacionamentos.

Ao criar estes esquemas foram tomados os devidos cuidados para que os dois esquemas simples representassem a mesma informação. Estes mesmos cuidados foram tomados para os complexos.

Para realizar o experimento, foram selecionados estudantes como usuários, divididos em 4 equipes. Cada equipe observou apenas um esquema por um determinado tempo e depois tentou reproduzi-los.

Como resultado, este estudo mostra que, em relação à segunda tese, os usuários tiveram praticamente a mesma facilidade em lembrar das tabelas nos esquemas estrela e ER. Em relação à primeira tese, contudo, os usuários tiveram uma dificuldade maior em lembrar dos relacionamentos do esquema ER. Dessa maneira, concluiu-se que o esquema estrela é mais legível para o usuário.

2.5.3 – Eficiência, Conformidade e Custo

Eficiência é o aspecto que determina o quão veloz é o DW na extração as informações que o usuário necessita (Inmon, 2002). Essa velocidade é dependente de fatores tão variados como o equipamento onde está localizado o DW, o software utilizado, a criação de tabelas agregadas (Kelly, 1998; Gupta, 1997; Sybase, 1998; Shouten, 1999), a granularidade (Ballard *et al*, 1998; Song *et al*, 2001; Song, 1999; Humphries, Hawkins e Dy, 1999), etc.

Conformidade é o aspecto que determina o quanto os dados do DW estão representados corretamente em relação aos dados provenientes dos sistemas de origem.

Custo, por sua vez, representa o investimento na construção e manutenção do DW. Este é outro aspecto muito amplo, ele também é dependente de inúmeros fatores tais como o volume e da taxa de crescimento dos dados armazenados, o equipamento utilizado, as manutenções necessárias nos dados e no esquema, etc. Existem, ainda, os custos indiretos, como o sucesso ou não da extração da informação necessária no prazo desejado.

Os aspectos de qualidade Eficiência, Conformidade e Custo foram apenas citados por serem considerados importantes pelo autor em um projeto de DW. Contudo, para este trabalho, os únicos aspectos relevantes são os de Adaptabilidade e Legibilidade.

2.6 – Evolução de Sistemas de Informação

A maioria dos sistemas de informações sofre alterações ao longo do tempo e, conseqüentemente, a maioria dos DW também sofre mudanças para se adequarem à nova realidade (Roddick *et al*, 2000; Eder e Koncilia, 2002).

A seguir, são apresentados os principais motivos que causam mudanças nas regras de negócio e as novas necessidades que o DW deve contemplar advindas destas mudanças.

2.6.1 – Motivos que causam mudanças nas regras de negócio

Dentre os possíveis motivos de alteração nas regras de negócio, pode-se citar (Roddick *et al*, 2000):

- Mudanças no ambiente interno da corporação:
 - Mudança de objetivos: Pode-se exemplificar esta mudança com uma corporação que focava os seus produtos no mercado interno e passa a ter o mercado externo como seu principal cliente.
 - Mudança de estrutura organizacional: Um exemplo deste caso é quando uma corporação que estava trabalhando hierarquicamente passa a trabalhar por projetos.
 - Otimização dos processos do negócio: Este caso pode ocorrer em uma corporação que possui uma linha de montagem e decide agregar algumas das tarefas.
 - Novas informações: Um exemplo desta mudança é quando a empresa lança um novo produto que precisa ser analisado sob uma nova visão ou perspectiva.
 - Mudança no ambiente externo da corporação:
 - Alteração na legislação: Como, por exemplo, a mudança na lei das S/A, a reforma tributária, etc.
 - Fusões: Quando empresas se fundem, geralmente elas possuem diferentes sistemas e métodos de trabalho que normalmente precisam ser equalizados.
 - Formação de grupos internacionais: Quando, por exemplo, se formar a ALCA, novas maneiras de se realizarem negócios poderão ser criadas.
-

2.6.2 – Novas necessidades geradas pela alteração nas regras de negócio

Com a alteração nas regras de negócio, apresentam-se como consequência novas necessidades que o DW deve contemplar (ou problemas a resolver).

A seguir são apresentadas algumas dessas novas necessidades identificadas pelo autor:

- O relacionamento M:N entre a tabela fato e dimensão:
 - Esse caso aparece quando o relacionamento 1:N entre as tabelas fato e dimensão não permitir ao DW refletir a informação correta. Um exemplo ocorre em um DW de vendas de uma empresa que decide comercializar imóveis. Como um imóvel pode ser adquirido por mais de um cliente, o relacionamento entre a tabela fato (venda) e a dimensão (cliente) passa a necessitar um relacionamento M:N;
 - O detalhamento das dimensões:
 - Essa necessidade é a mais comum e ocorre quando é preciso acrescentar novas informações em uma dimensão existente no DW. Como exemplo pode-se citar o caso em que um DW de compras necessitar armazenar o endereço de um fornecedor;
 - O armazenamento de dados históricos na tabela dimensão:
 - Esta situação acontece quando se torna necessário guardar, na tabela dimensão, uma informação passível de mudança ao longo do tempo. Como exemplo pode-se citar o mesmo caso anterior em que um DW de compras que agora necessita armazenar o endereço atual e passados de um fornecedor;
 - A inclusão de uma nova dimensão:
 - Esta necessidade ocorre quando se precisar analisar o fato sob um novo ponto de vista ou perspectiva. Como exemplo pode-se citar
-

um DW de vendas que necessita analisar um fato (uma venda) sob a ótica da forma de pagamento utilizada.

2.7 – Resumo e Considerações Gerais

Neste capítulo são apresentadas as definições dos principais conceitos envolvidos com DW. São mostrados os principais modelos dimensionais (estrela, floco de neve e constelação) e suas características básicas.

Dentre aos aspectos de qualidade do DW são apresentados a adaptabilidade e a legibilidade. No caso do modelo estrela mostra-se que a sua adaptabilidade corresponde a inalterabilidade das seguintes características: ter apenas uma tabela fato; ter relacionamentos 1:N entre a tabela fato e dimensão; e ter dimensões desnormalizadas. Quanto a legibilidade, constata-se que o modelo estrela é melhor compreendido pelo usuário que o ER (Dowling, Schuff e Louis, 2001).

Em seguida são apresentados alguns motivos que alteram as regras de negócio. Como consequência destas alterações, são mostradas as novas necessidades que o DW deve contemplar para se adequar à nova realidade.

No capítulo 3 é feita uma análise da adaptabilidade e legibilidade do esquema estrela submetido a alterações decorrentes das mudanças das regras de negócio.

CAPITULO 3 – ANÁLISE DA ADAPTABILIDADE E LEGIBILIDADE DO MODELO ESTRELA SUBMETIDO A ALTERAÇÕES

3.1 - Introdução

A maioria dos DW evolui e, conseqüentemente, refletem estas evoluções em seus esquemas. Como parte deste processo de evolução, estes esquemas podem sofrer mudanças estruturais que podem acarretar a criação, exclusão ou modificação das tabelas e seus respectivos elementos.

Estas mudanças estruturais, contudo, podem comprometer a adaptabilidade do modelo, haja vista a criação de artifícios nem sempre permitidos nesta modelagem. Outros aspectos, como a legibilidade do modelo, também podem ser comprometidos, pois se altera a forma com que o usuário está acostumado a analisar os dados.

Como visto no capítulo anterior, algumas novas necessidades que o DW deve contemplar (ou problemas a serem resolvidos) identificadas neste trabalho são:

- O relacionamento M:N entre a tabela fato e dimensão;
- O detalhamento de dimensões;
- O armazenamento de dados históricos na tabela dimensão; e
- A inclusão de uma nova dimensão.

A seguir são estudadas estas novas necessidades e as soluções propostas para que o DW possa contemplá-las.

3.2 – Relacionamento M:N entre a tabela fato e dimensão

A necessidade do relacionamento M:N entre tabelas fato e dimensão é um evento que comumente ocorre em projetos reais e o esquema estrela nem sempre consegue absorver de forma satisfatória esta mudança (Song *et al*, 2001). Para

exemplificar a necessidade do relacionamento M:N entre a tabela fato e dimensão é apresentado a seguir um caso em que esta mudança é necessária.

3.2.1 – Apresentação do problema

Uma empresa de saúde possui um DW que contém as informações dos atendimentos médicos efetuados pelos seus associados. Este DW armazena as informações sobre o paciente atendido, o local, o tipo de atendimento recebido, o diagnóstico, além da data/hora e do custo do atendimento. De acordo com estas regras de negócio o esquema estrela do DW é apresentado na figura 3.1.

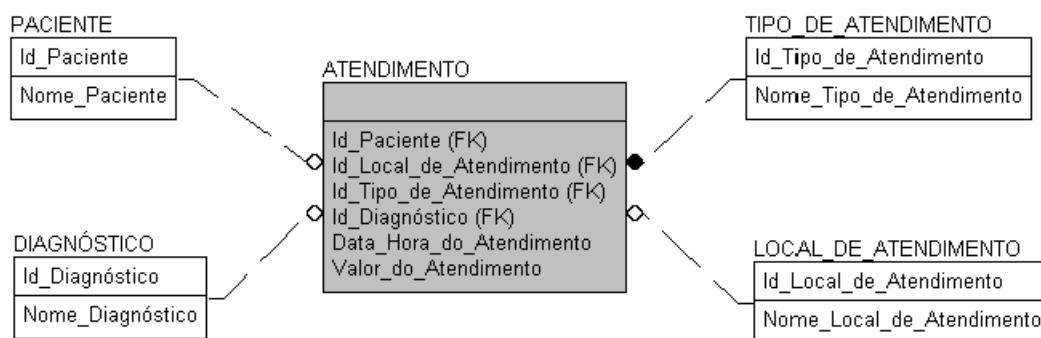


Figura 3.1 – Exemplo do problema de relacionamento M:N

No esquema da figura 3.1, a dimensão PACIENTE identifica quem é atendido; a dimensão LOCAL_DE_ATENDIMENTO identifica onde ocorre o atendimento e contém o nome de um hospital, clínica ou laboratório; a dimensão TIPO_DE_ATENDIMENTO mostra qual o atendimento realizado que pode ser uma consulta, internação ou exame; e a dimensão DIAGNÓSTICO mostra a doença que é diagnosticada no atendimento (câncer, pneumonia, meningite, etc). Na tabela fato ATENDIMENTO ainda existe a data/hora em que ocorre o atendimento e o valor do atendimento.

Com o passar do tempo, ocorre uma mudança na regra de negócio: um atendimento realizado por um paciente que originalmente poderia resultar em apenas um diagnóstico passou a poder resultar em um ou mais diagnósticos. Isto implica em mudar o relacionamento da tabela fato ATENDIMENTO com a dimensão DIAGNÓSTICO de 1:N para M:N.

3.2.2 – Soluções para o relacionamento M:N entre a tabela fato e a dimensão

Existem diversas soluções (Song *et al*, 2001; Pedersen, 1999a; Kimball, 1997) para resolver o relacionamento M:N entre a tabela fato e a dimensão. A seguir são apresentadas algumas das soluções possíveis.

3.2.2.1 – Criação de tabelas ponte (Kimball, 1997)

Criar tabelas ponte consiste em inserir duas novas tabelas entre a tabela fato e a dimensão, realizando estas novas tabelas a função de um relacionamento M:N entre a tabela fato e dimensão (Kimball, 1997).

Na figura 3.2 é apresentado um esquema onde entre a dimensão DIAGNÓSTICO e a tabela fato ATENDIMENTO são inseridas as tabelas pontes GRUPO_DIAGNÓSTICO_PONTE e ELEMENTOS_DO_GRUPO_DIAGNÓSTICO. Desta forma, um atendimento possui um grupo de diagnóstico que é composto por um ou mais diagnósticos.

As dimensões que não são relevantes para a solução do problema não são mostradas na figura 3.2 e os nomes das tabelas e atributos novos estão em negrito.

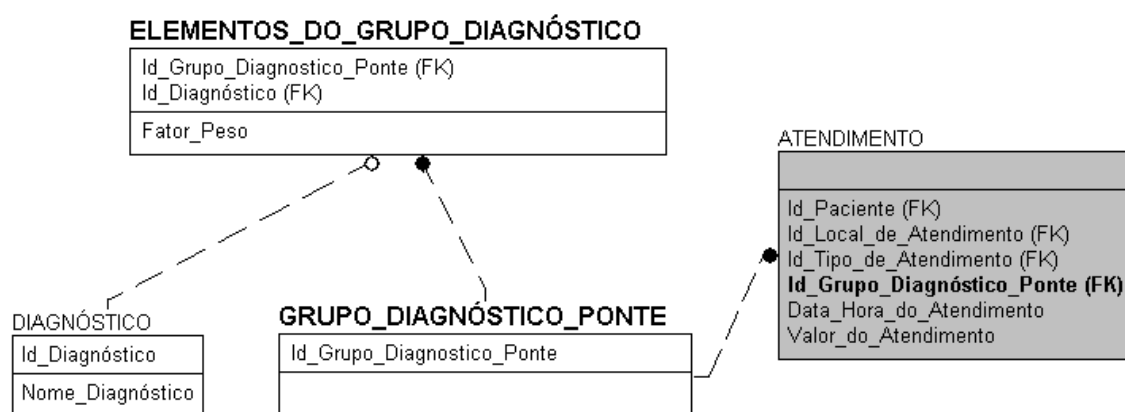


Figura 3.2 – Criação de tabelas pontes

Comparando a tabela fato ATENDIMENTO da figura 3.1 com a mesma tabela da figura 3.2 nota-se que na tabela fato da figura 3.2 a coluna Id_Diagnóstico é substituída pela coluna Id_Grupo_Diagnóstico_Ponte. Esta nova coluna passa a representar um agrupamento de diagnósticos possíveis para uma instância da tabela fato, e, sendo assim, para cada grupo de diagnóstico pode haver um ou mais diagnósticos associados. Na tabela ponte ELEMENTOS_DO_GRUPO_DIAGNÓSTICO também é observada a existência da coluna Fator_Peso. O fator de peso representa a fração de participação de uma instância no grupo.

Para exemplificar melhor a importância do fator de peso para esta solução são adotados, como conteúdo da tabela fato ATENDIMENTO, os dados da tabela 3.1 e como conteúdo da tabela ponte ELEMENTOS_DO_GRUPO_DIAGNÓSTICO, os dados da tabela 3.2.

Id_Paciente	Id_Grupo_Diagnóstico_Ponte	Valor_do_Atendimento	Demais Atributos
Paciente 1	GD 1	500,00	...
Paciente 2	GD 2	200,00	...

Tabela 3.1 – Conteúdo da tabela fato ATENDIMENTO

Id_Grupo_Diagnóstico	Id_Diagnóstico	Fator_Peso
GD1	D1	0,3
GD1	D2	0,3
GD1	D3	0,4
GD2	D3	0,8
GD2	D4	0,2

Tabela 3.2 – Conteúdo da tabela ponte ELEMENTOS_DO_GRUPO_DIAGNÓSTICO

Suponha que se deseja saber qual o total de gasto por diagnóstico. Inicialmente é executada uma consulta no esquema da figura 3.2 em que não é utilizado o fator de peso e que tem como resultado os valores da tabela 3.3. A consulta utilizada é apresentada a seguir:

```
select DIAGNOSTICO.Nome_Diagnostico,
       sum(ATENDIMENTO.Valor_do_Atendimento) Valor_Total_Diagnóstico
from ATENDIMENTO, DIAGNÓSTICO, ELEMENTOS_DO_GRUPO_DIAGNOSTICO
where DIAGNÓSTICO.Id_Diagnostico = ELEMENTOS_DO_GRUPO_DIAGNOSTICO.Id_Diagnostico and
      ELEMENTOS_DO_GRUPO_DIAGNOSTICO.Id_Grupo_Diagnostico =
      ATENDIMENTO.Id_Grupo_Diagnostico;
```

Nome_Diagnóstico	Valor_Total_Diagnóstico
D1	500,00
D2	500,00
D3	700,00
D4	200,00

Tabela 3.3 – Valores por diagnóstico sem utilizar o fator de peso

Analisando-se os resultados da consulta na tabela 3.3 observa-se que os resultados estão incorretos, pois somando-se o Valor_Atendimento da tabela 3.1 ter-se-á como resultado o valor de 700,00 enquanto que somando-se o Valor_Total_Diagnóstico da tabela 3.3 o resultado é o valor 1.900,00.

Em contrapartida utilizando-se o fator de peso conforme a consulta a seguir, tem-se como resultado a tabela 3.4 onde é mostrado o valor correto associado a cada diagnóstico.

```
select DIAGNOSTICO.Nome_Diagnostico,
       sum(ATENDIMENTO.Valor_do_Atendimento * ELEMENTOS_DO_GRUPO_DIAGNOSTICO.Fator_Peso)
       Valor_Total_Diagnóstico
from ATENDIMENTO, DIAGNÓSTICO, ELEMENTOS_DO_GRUPO_DIAGNOSTICO
where DIAGNÓSTICO.Id_Diagnostico = ELEMENTOS_DO_GRUPO_DIAGNOSTICO.Id_Diagnostico and
      ELEMENTOS_DO_GRUPO_DIAGNOSTICO.Id_Grupo_Diagnostico =
      ATENDIMENTO.Id_Grupo_Diagnostico;
```

Diagnóstico	Valor_Total_Diagnóstico
D1	150,00
D2	150,00
D3	360,00
D4	40,00

Tabela 3.4 – Valores por diagnóstico utilizando o fator de peso

A solução proposta [Kimball, 1997] possui as seguintes características:

Adaptabilidade: Nesta solução são criadas estruturas adicionais que não estão previstas no modelo estrela.

Legibilidade: A criação das tabelas ponte dificulta a legibilidade do esquema para as pesquisas onde o usuário necessita utilizar os novos relacionamentos criados pelas novas tabelas ponte. A forma com que é utilizado o fator de peso também prejudica a legibilidade do esquema, uma vez que é mais uma informação a ser registrada e gerenciada no DW.

3.2.2.2 – Criação de uma nova tabela dimensão contendo atributos binários (Song *et al*, 2001)

Esta solução consiste em criar uma nova dimensão representando um mapa binário que indica a existência ou não das instâncias da dimensão original. Dimensão original significa a tabela dimensão antes de submetê-la a qualquer alteração para adequar o esquema ao relacionamento M:N. Nesta nova tabela, cada elemento da dimensão original se transforma em um atributo (Song *et al*, 2001).

O exemplo apresentado na figura 3.3 resulta da eliminação da dimensão (original) DIAGNÓSTICO da figura 3.1 criando-se a nova dimensão GRUPO_DIAGNÓSTICO_BINÁRIO, onde a cada diagnóstico corresponde um atributo binário identificado como Possui_Diagnóstico_1 a Possui_Diagnóstico_12, que pode ser verdadeiro ou falso. A combinação dos atributos mostra quais diagnósticos o paciente possui.

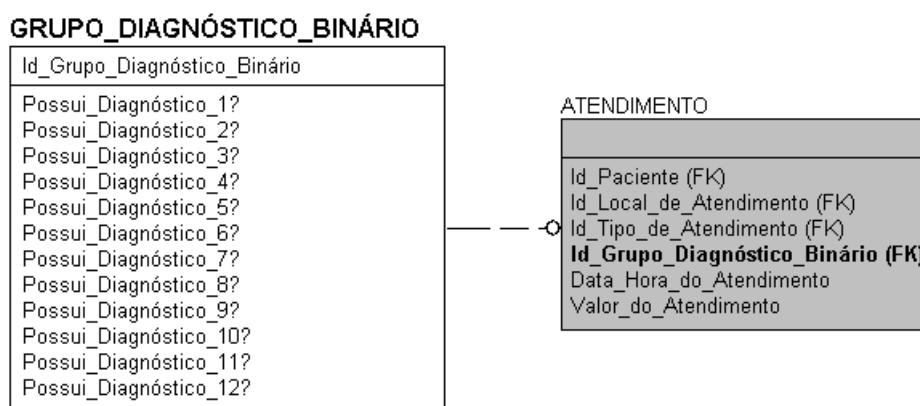


Figura 3.3 – Criação de dimensão com mapa binário

Comparando a tabela fato ATENDIMENTO da figura 3.1 com a da figura 3.3 nota-se que é excluída a coluna Id_Diagnóstico e incluída a coluna Id_Grupo_Diagnóstico_Binário que representa o grupo de diagnósticos do paciente. Para identificar o que é diagnosticado em um paciente basta verificar os atributos

Possui_Diagnóstico_1 a Possui_Diagnóstico_12, da dimensão GRUPO_DIAGNÓSTICO_BINÁRIO e verificar quais atributos estão com valor verdadeiro.

Nesta solução a tabela GRUPO_DIAGNÓSTICO_BINÁRIO pode ser carregada com todas as combinações necessárias previamente, ou gradativamente, conforme a necessidade. A título de exemplo, como são 12 diagnósticos diferentes na tabela GRUPO_DIAGNÓSTICO_BINÁRIO podem existir 4096 instâncias.

Adaptabilidade: Nesta solução, mesmo alterando a tabela fato e a tabela dimensão, a configuração em estrela do esquema é preservada, pois não é alterado o relacionamento 1:N entre a tabela fato e a dimensão.

Legibilidade: A legibilidade pode ser prejudicada no caso da dimensão possuir muitos atributos, uma vez que a combinação destes pode levar o total de instâncias da tabela a 2^N , onde N é o total de atributos da tabela. Como exemplo, caso existissem 40 atributos na dimensão, seriam necessários um trilhão de instâncias para se fazer todas as combinações.

3.2.2.3 – Indicação de grupos dentro da tabela fato (Song *et al*, 2001)

As soluções anteriores consistiam em manter a semântica da tabela fato inalterada, isto é, continuar armazenando o evento original, alterando-se a semântica da dimensão da seguinte maneira: uma instância da nova dimensão (com a semântica alterada) armazenava um grupo de instâncias da dimensão original. Alterar a semântica de uma tabela significa mudar o que representa uma instância da tabela.

A solução aqui apresentada consiste em manter inalterada a semântica da dimensão original e mudar a semântica da tabela fato. Com isso, a tabela fato passa a ter uma granularidade mais fina ao armazenar não mais o evento original, mas um novo evento que é relacionado a apenas uma instância da dimensão original. Desta maneira, um evento originalmente armazenado na tabela fato original pode ser

desmembrado em diversos eventos na nova tabela fato com a semântica alterada (Song *et al*, 2001).

A figura 3.4 apresenta esta solução, onde se pode observar que a tabela fato original ATENDIMENTO da figura 3.1 é substituída pela nova tabela fato ATENDIMENTO_DIAGNÓSTICO, que tem uma semântica diferente, pois armazena as informações a respeito do diagnóstico e não mais as do atendimento. Neste exemplo, cada instância da dimensão GRUPO_ATENDIMENTO, representa um atendimento.

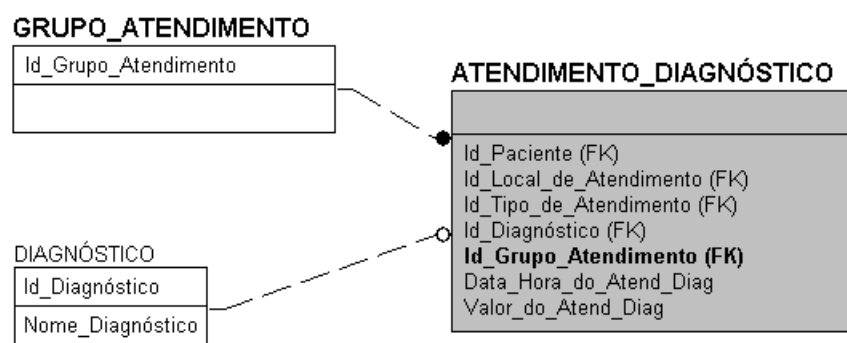


Figura 3.4 – Indicação de grupos dentro da tabela fato

A maior diferença entre a tabela fato ATENDIMENTO apresentada na figura 3.1 e a tabela fato ATENDIMENTO_DIAGNÓSTICO da figura 3.4 é que na tabela fato ATENDIMENTO cada instância representa um atendimento do paciente, enquanto que na tabela ATENDIMENTO_DIAGNÓSTICO cada instância representa as informações relativas a um diagnóstico de um atendimento. Com isso, o Valor_do_Atendimento da figura 3.1 deixa de existir e é criado o Valor_do_Atend_Diag que é o valor do atendimento por diagnóstico.

Para saber o valor de um atendimento, é necessário somar o Valor_do_Atendimento_Diag dos registros da tabela fato ATENDIMENTO_DIAGNÓSTICO com o mesmo Id_Grupo_Atendimento.

Adaptabilidade: Nesta solução é alterada a tabela fato e criada uma nova estrutura, mas não se altera a configuração em estrela do esquema, mantendo o relacionamento 1:N entre a tabela fato e a dimensão.

Legibilidade: Nesta solução, ao alterar-se a semântica da tabela fato afinando a sua granularidade, a legibilidade é prejudicada porque pode haver confusão com esta diferença de granularidade entre a tabela fato original e a nova tabela fato.

3.2.2.4 – Indicação de grupos em uma nova tabela fato (Song *et al*, 2001)

Esta solução é semelhante à solução anterior 3.2.2.3 no que diz respeito a alterar a semântica da tabela fato e não mais da tabela dimensão. A diferença entre as duas é que enquanto a solução anterior altera a semântica da tabela fato original, esta preserva a semântica desta tabela e cria uma nova tabela fato com a nova semântica de granularidade mais fina (Song *et al*, 2001).

A figura 3.5 apresenta um exemplo desta solução, onde para armazenar as informações referentes ao diagnóstico, é criada a nova tabela fato ATENDIMENTO_DIAGNÓSTICO_NOVO que é idêntica à tabela fato ATENDIMENTO_DIAGNÓSTICO da solução 3.2.2.3 representado na figura 3.4. Neste exemplo, para relacionar a tabela fato ATENDIMENTO_DIAGNÓSTICO_NOVO com a tabela fato original ATENDIMENTO é utilizada a dimensão GRUPO_ATENDIMENTO_NOVO, onde cada instância desta dimensão representa um atendimento.

Comparando a tabela fato ATENDIMENTO da figura 3.1 com a tabela fato ATENDIMENTO da figura 3.5, esta última não possui mais a referência para a dimensão DIAGNÓSTICO, pois os diagnósticos dos pacientes agora são armazenados na nova tabela fato ATENDIMENTO_DIAGNÓSTICO_NOVO. Esta nova tabela fato também possui o atributo Valor_do_Atendimento_Diag_Novo, que é o valor do diagnóstico em um atendimento. A dimensão GRUPO_

ATENDIMENTO_NOVO serve para criar um relacionamento 1:N entre a tabela fato ATENDIMENTO e ATENDIMENTO_DIAGNÓSTICO_NOVO.

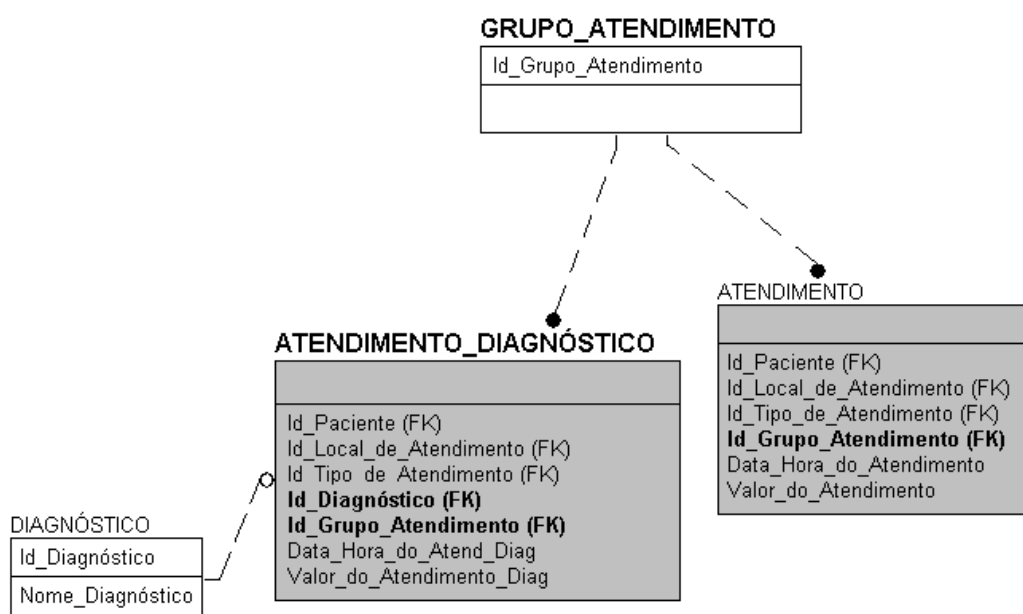


Figura 3.5 – Indicação de grupos em uma nova tabela fato

Neste novo esquema observam-se dois esquemas estrelas, sendo um com a tabela fato ATENDIMENTO_DIAGNÓSTICO_NOVO e outro com a tabela fato ATENDIMENTO, separando-se assim, os assuntos de atendimento e de diagnósticos por atendimento em dois esquemas diferentes.

Uma solução alternativa, neste caso, é transformar a tabela fato ATENDIMENTO em uma dimensão da tabela fato ATENDIMENTO_DIAGNÓSTICO_NOVO. Como executar esta transformação pode ser verificado adiante no item 3.5.2.3.

Adaptabilidade: Esta solução prejudica a adaptabilidade do esquema estrela, pois, ao criar uma nova tabela fato, o transforma em um esquema constelação. Também é criada uma estrutura adicional para relacionar as tabelas fato.

Legibilidade: A transformação para um esquema constelação pode prejudicar a legibilidade, pois o usuário pode necessitar consultar os dois esquemas para obter a informação desejada.

3.2.2.5 – Pulverização da dimensão em N ‘mini-dimensões’ (Pedersen, 1999a)

Esta solução consiste em pulverizar a dimensão original em ‘mini-dimensões’ para cada instância da dimensão original. Em seguida relacionam-se estas ‘mini-dimensões’ com a tabela fato. Cada ‘mini-dimensão’ é uma tabela contendo uma instância da dimensão original, possuindo apenas um atributo binário de existência.

Na figura 3.6 é apresentado um exemplo da pulverização da dimensão DIAGNÓSTICO da figura 3.1, criando as ‘mini-dimensões’ SITUAÇÃO_DIAGNÓSTICO_1 a SITUAÇÃO_DIAGNÓSTICO_4, representando as instâncias da dimensão DIAGNÓSTICO (para este exemplo, a dimensão DIAGNÓSTICO possui apenas quatro instâncias).

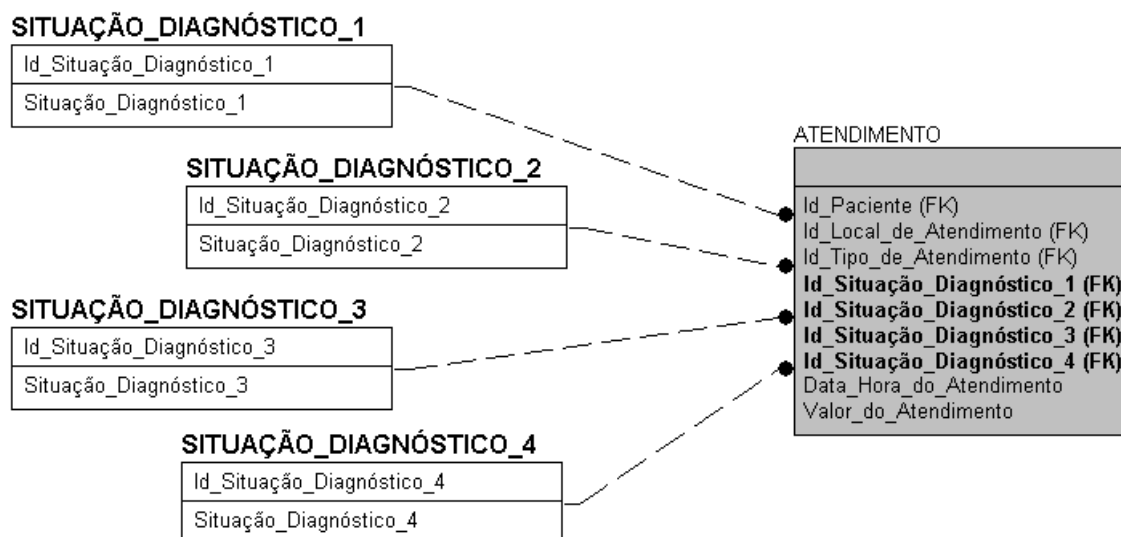


Figura 3.6 – Inclusão das mini-dimensões na tabela fato

Cada ‘mini-dimensão’ SITUACÃO_DIAGNÓSTICO_X contém um atributo binário Situação_Diagnóstico_x, que indica a presença ou não deste diagnóstico.

Para saber quais diagnósticos foram feitos no atendimento, é preciso verificar quais Id_Situação_Diagnósticos_x da tabela fato ATENDIMENTO representam a Situação_Diagnóstico_x igual a verdadeiro da sua respectiva ‘mini-dimensão’.

Adaptabilidade: Mesmo criando diversas estruturas, esta solução mantém a configuração em estrela do esquema ao conservar os relacionamentos 1:N entre a nova tabela fato e as ‘mini-dimensões’.

Legibilidade: A legibilidade pode ser prejudicada se existirem muitas ‘mini-dimensões’, pois dependendo da necessidade do usuário, este precisará verificar todas as ‘mini-dimensões’ para realizar as suas pesquisas.

3.2.2.6 – Combinação das instâncias da dimensão em uma lista (Song *et al*, 2001)

Esta solução consiste em criar uma nova tabela dimensão que armazene em listas, as combinações das instâncias da dimensão original, sendo estas listas uma instância de um atributo tipo texto (Song *et al*, 2001). Esta solução, ao armazenar as combinações das instâncias da dimensão original, é semelhante à do item 3.2.2.2.

Como exemplo, a figura 3.7 apresenta um esquema em que é eliminada a dimensão original DIAGNÓSTICO da figura 3.1 e criada a nova dimensão GRUPO_DIAGNÓSTICO_LISTA. Nesta tabela, o atributo Lista_de_Diagnósticos contém a combinação das instâncias da dimensão original DIAGNÓSTICO e, para identificar o que é diagnosticado no paciente, basta verificar se o diagnóstico está inserido dentro desta lista.

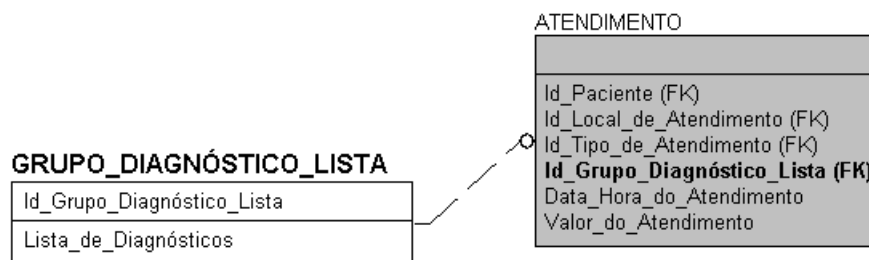


Figura 3.7 – Combinações das instâncias da dimensão original em listas

Para uma melhor comparação entre esta solução e a do item 3.2.2.2, a tabela 3.5 apresenta algumas instâncias da tabela GRUPO_DIAGNÓSTICO_BINÁRIO da figura 3.3 e a tabela 3.6 apresenta as instâncias da tabela GRUPO_DIAGNÓSTICO_LISTA.

Id_Grupo_Diagnóstico_Binário	Possui_Diagnóstico_1 (Câncer)	Possui_Diagnóstico_2 (Enxaqueca)	Possui_Diagnóstico_3 (Úlcera)	Possui_Diagnóstico_4 (AIDS)
ID 1	POSSUI	NÃO POSSUI	NÃO POSSUI	POSSUI
ID 2	NÃO POSSUI	NÃO POSSUI	POSSUI	POSSUI
ID 3	POSSUI	POSSUI	POSSUI	NÃO POSSUI

Tabela 3.5 – Conteúdo da tabela GRUPO_DIAGNÓSTICO_BINÁRIO da fig. 3.3

Id_Grupo_Diagnóstico_Lista	Lista_de_Diagnósticos
ID 1	Câncer; AIDS
ID 2	Úlcera; AIDS
ID 3	Câncer; Enxaqueca; Úlcera

Tabela 3.6 – Conteúdo da tabela GRUPO_DIAGNÓSTICO_LISTA da fig. 3.7

Comparando-se as tabelas, nota-se que o Id_Grupo_Diagnóstico_Binário igual a 1 da tabela 3.5 representa o mesmo conjunto de diagnósticos que o Id_Grupo_Diagnóstico_Lista igual a 1 da tabela 3.6.

Nesta solução, assim como na solução do item 3.2.2.2, a tabela GRUPO_DIAGNÓSTICO_LISTA pode ser carregada previamente com todas as combinações necessárias ou gradativamente, conforme a necessidade.

Adaptabilidade: Nesta solução a tabela fato é alterada, mas se conserva o relacionamento 1:N entre a tabela fato e a dimensão, preservando-se a configuração em estrela esquema.

Legibilidade: A legibilidade pode ser prejudicada, uma vez que o usuário está consultando uma dimensão com a semântica alterada, pois as instâncias da nova dimensão passam a representar os grupos de dimensão original, ao invés da dimensão original.

3.2.3 – Resumo das soluções

Analisando as soluções sob a ótica da adaptabilidade, verificou-se que a maioria delas prejudicam a configuração estrela do esquema inicial. Esta configuração normalmente é prejudicada ao transformar o esquema em uma constelação ou criando-se estruturas adicionais.

Sob a ótica da legibilidade, todas as soluções propostas a prejudicam em maior ou menor grau, pois altera-se a semântica da tabela fato (diminuindo a sua granularidade) ou da dimensão (passando a representar os grupos de dimensão original) ou ainda porque é alterada a configuração estrela do esquema. Verificou-se também que as soluções que mantêm a configuração estrela do esquema apresentam o inconveniente de poderem ser utilizadas apenas em dimensões com pequeno número de instâncias.

Com isso, para resolver o problema levantado, não se encontrou nenhuma solução satisfatória tanto para a adaptabilidade quanto para a legibilidade, indicando que o esquema estrela não resolve satisfatoriamente o problema de relacionamento M:N entre as tabelas fato e dimensão.

3.3 – Detalhamento de dimensões

Uma alteração que normalmente ocorre é o detalhamento de uma dimensão existente no DW, sem interferir na informação armazenada na tabela fato.

Para ilustrar esta necessidade, apresenta-se a seguir uma situação onde é necessário realizar esta mudança estrutural.

3.3.1 – Apresentação do problema

Uma rede de lojas de materiais de construção possui um DW que controla a venda dos seus produtos. No DW são armazenadas as informações sobre as vendas, as filiais onde estas são realizadas, os vendedores, os produtos vendidos e os clientes, além das datas/horas e os valores da vendas. O esquema estrela do DW desta rede é o apresentado na figura 3.8.

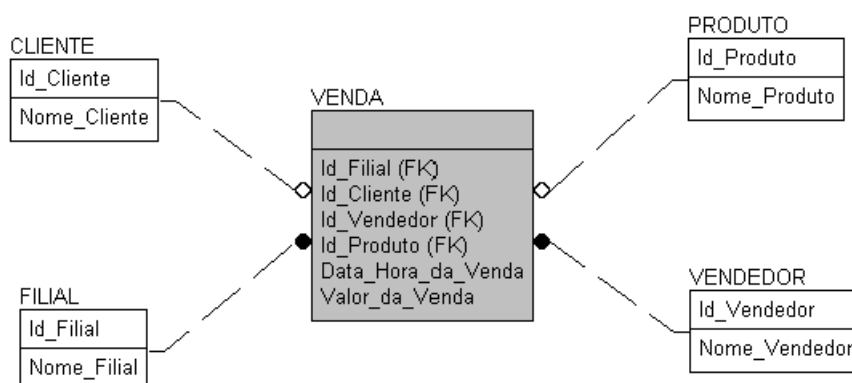


Figura 3.8 – Exemplo do detalhamento dos fatos e dimensões

No esquema apresentado na figura 3.8 a tabela fato é representada pela tabela VENDA e as dimensões são CLIENTE, FILIAL, VENDEDOR E PRODUTO. Uma necessidade que esta empresa possui é acrescentar informação sobre o fabricante dos produtos, sendo que cada produto é relacionado a apenas um fabricante.

3.3.2 – Solução para o detalhamento de dimensões

A solução para este caso é apresentada a seguir.

3.3.2.1 – Detalhamento da dimensão

O detalhamento de uma dimensão já existente é uma mudança simples de ser executada. Esta solução consiste em apenas incluir os novos atributos na dimensão.

No exemplo da rede de lojas de materiais de construção, esta solução é válida para a inclusão da informação do fabricante dos produtos, pois esta informação está diretamente relacionada com a dimensão PRODUTO. Com isso, nem a semântica da tabela fato VENDA, nem a da dimensão PRODUTO são alteradas. A figura 3.9 apresenta a informação do fabricante, identificado como o atributo Nome_Fabricante inserido na tabela PRODUTO.

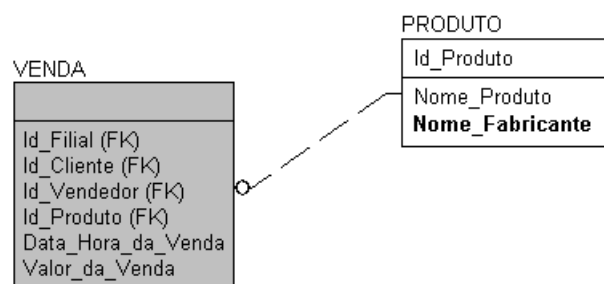


Figura 3.9 – Detalhamento de uma dimensão

Adaptabilidade: Nesta solução, apesar de ser alterada a dimensão, não é alterada a configuração em estrela do esquema.

Legibilidade: Nesta solução a legibilidade não é prejudicada.

3.3.3 – Resumo das soluções

Esta solução não prejudica a adaptabilidade do esquema, uma vez que a configuração em estrela é conservada.

A legibilidade também não é prejudicada, pois a semântica da tabela fato e a da dimensão permanecem inalteradas.

Assim, pode-se verificar que o esquema estrela resolve satisfatoriamente o detalhamento de dimensões.

3.4 – Armazenamento de dados históricos na tabela dimensão

Costuma-se assumir que os valores dos atributos de dimensões numa estrela não se alteram com o tempo, mas isto nem sempre é verdade em DW reais (Ballard *et al*, 1998; Kimball, 1996).

Para ilustrar este caso, é apresentada a seguir uma situação onde este caso ocorre.

3.4.1 – Apresentação do problema

O exemplo a seguir mostra um caso em que é necessário o armazenamento de dados históricos na tabela dimensão.

Uma empresa que comercializa materiais para escritório possui o esquema estrela mostrado na figura 3.10 onde é apresentada uma tabela fato de pedidos.

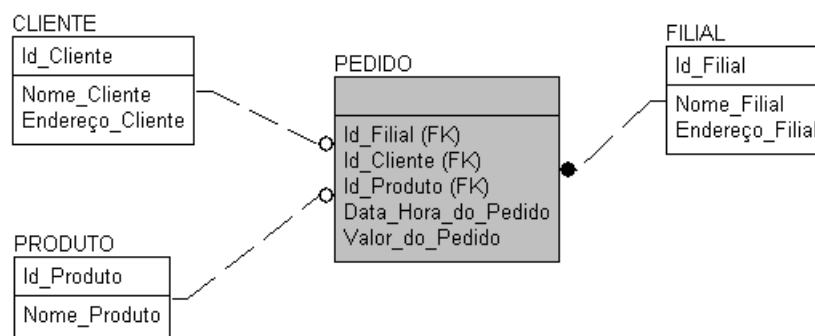


Figura 3.10 – Exemplo para armazenamento de dados históricos na dimensão

Por este esquema a empresa determina qual filial possui mais pedidos, em quais filiais um cliente costuma realizar os seus pedidos, etc. Mas, ao analisar os resultados, a empresa deparou-se com o seguinte problema: o cliente A, que até certa data sempre fazia seus pedidos na filial 1, começou a fazer seus pedidos na filial 2. A questão é: será que o cliente A começou a realizar os seus pedidos na filial 2 porque mudou-se para um endereço próximo da filial 2?

Este esquema não permite saber se o cliente mudou ou não de endereço, pois na dimensão CLIENTE é armazenado apenas o endereço atual do cliente.

3.4.2 – Soluções para o armazenamento de dados históricos na tabela dimensão

A seguir são mostradas duas soluções, sendo que uma baseia-se no artigo (Kimball, 1996) e a outra na experiência do próprio autor.

3.4.2.1 – A cada atualização de um atributo da dimensão gera-se uma nova instância na mesma tabela dimensão (Kimball, 1996)

A primeira solução é gerar uma nova instância na própria tabela dimensão a cada mudança do valor de um atributo (Kimball, 1996).

Adotando esta solução no exemplo para armazenar os dados históricos de endereço do cliente, são incluídos na dimensão CLIENTE os atributos

Código_Cliente, para identificar um cliente, a Data_Início e a Data_Fim que informam a data de início e fim de uma instância da tabela fato. O esquema resultante é apresentado na figura 3.11.

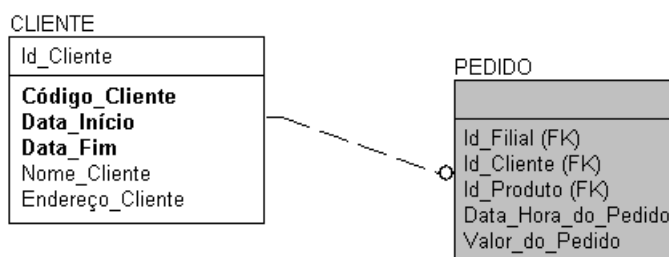


Figura 3.11 – Armazenamento dos dados históricos na própria dimensão

Nesta solução, cada vez que o cliente mudar de endereço, é gerada uma nova instância na tabela CLIENTE com os novos dados de endereço e é preenchida a Data_Início desta instância. A instância que se torna antiga, tem a sua Data_Fim preenchida também. Com isso, esta dimensão pode possuir mais de uma instância com os dados do mesmo cliente (uma para cada endereço diferente), e para determinar qual a instância atual, verifica-se qual das instâncias possui a data de início mais recente.

Como foi visto que pode existir mais de uma instância para um mesmo cliente, a maneira utilizada para determinar o número de clientes existentes na dimensão é através do atributo Código_Cliente, que é único por cliente.

Adaptabilidade: Esta solução, mesmo alterando a dimensão, não altera a configuração estrela do esquema.

Legibilidade: A legibilidade do esquema resultante fica prejudicada ao alterar a semântica da dimensão, que passa a armazenar várias instâncias para a mesma dimensão original. Além disso, caso os valores da dimensão sejam

alterados freqüentemente, ocorre uma alta redundância de dados, o que também prejudica a legibilidade.

3.4.2.2 – Criação de uma nova tabela fato para armazenar os dados históricos

A segunda solução para este problema é transferir para uma nova tabela fato todos os atributos da dimensão que podem se tornar históricos.

Utilizando esta solução no exemplo do endereço do cliente, é criada a nova tabela fato **CLIENTE_ENDEREÇO** como mostra a figura 3.12.

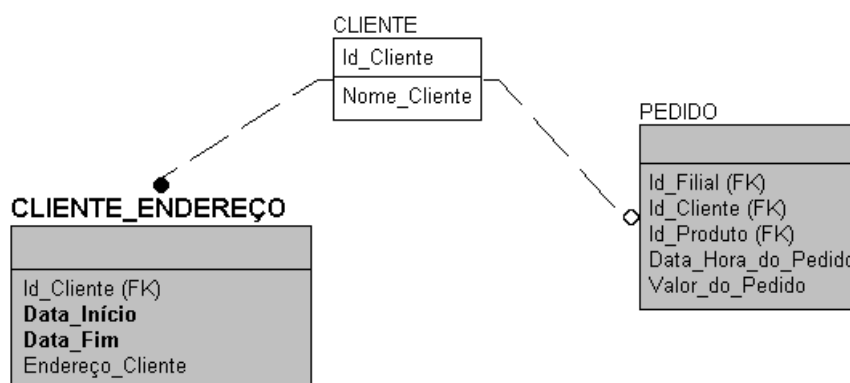


Figura 3.12 – Armazenamento dos dados históricos em outra tabela fato

Comparando a dimensão CLIENTE da figura 3.10 com a da figura 3.12, nota-se que esta última possui somente os atributos que não se alteram com o tempo e que o atributo **Endereço_Cliente** é transferido para a nova tabela fato **CLIENTE_ENDEREÇO**. Esta nova tabela fato também possui os atributos **Data_Início** e **Data_Fim** para indicar a partir de qual data este endereço é válido.

Nesta solução, quando é inserido um novo cliente na dimensão CLIENTE o seu endereço também é inserido na tabela fato CLIENTE_ENDEREÇO, bem como a **Data_Início**. A cada vez que o cliente alterar o endereço, somente uma nova instância na tabela fato CLIENTE_ENDEREÇO será inserida, preenchendo-se a **Data_Início**

desta instância. A instância que se tornou antiga na tabela deverá ter a Data_Fim preenchida também.

Observando novamente a figura 3.12 pode-se notar que o esquema é transformado em um esquema constelação.

Adaptabilidade: Esta solução transforma o esquema em constelação ao criar uma nova tabela fato.

Legibilidade: A transformação para o esquema constelação pode prejudicar a legibilidade, uma vez que o usuário pode necessitar consultar as duas tabelas fato dependendo da pesquisa realizada.

3.4.3 – Resumo das soluções

Em relação à adaptabilidade, somente a primeira solução mantém a configuração em estrela do esquema. A outra solução prejudica a adaptabilidade pois transforma o esquema em constelação.

Com relação a legibilidade, a primeira solução é prejudicada pois a semântica da dimensão é alterada, enquanto que a outra, ao transformar-se em constelação, pode também ter a legibilidade prejudicada dependendo da consulta realizada.

Observando-se estas características das soluções, pode-se notar que nenhuma delas é satisfatória, o que mostra que o esquema estrela não suporta bem este tipo de necessidade.

3.5 – Inclusão de uma nova dimensão

Aumentar o escopo do DW com o intuito de poder analisar os dados sob outra perspectiva também é uma alteração que normalmente ocorre.

A seguir é apresentado um exemplo deste problema.

3.5.1 – Apresentação do problema

O exemplo a seguir utilizará a mesma rede de lojas de materiais de construção do item 3.3.1, depois de resolvido o detalhamento da dimensão. O esquema estrela deste exemplo é apresentado na figura 3.13.

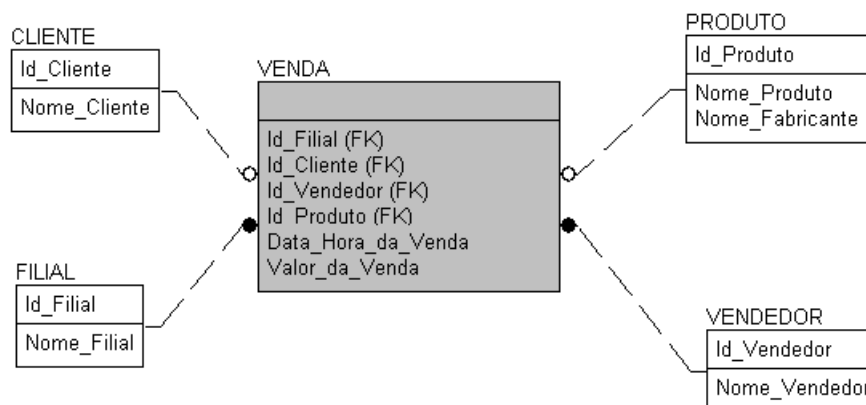


Figura 3.13 – Exemplo do problema de inclusão de uma nova dimensão

Com a evolução dos negócios a empresa necessita detalhar mais a informação sobre as vendas, incluindo as formas de pagamentos (dinheiro, cheque ou cartão de crédito) para poder melhor controlar o seu fluxo de caixa. Quando o pagamento for em cartão, a empresa também precisa saber qual é a operadora do cartão (Visa, Credicard, Diners, etc).

3.5.2 – Soluções para a inclusão de uma nova dimensão

A seguir são apresentadas as soluções.

3.5.2.1 – Relacionamento da tabela fato com uma nova dimensão

Este caso ocorre quando há necessidade de detalhar todas as instâncias da tabela fato.

No exemplo da rede de lojas de materiais de construção uma nova dimensão FORMA DE PAGAMENTO é inserida como mostra a figura 3.14.

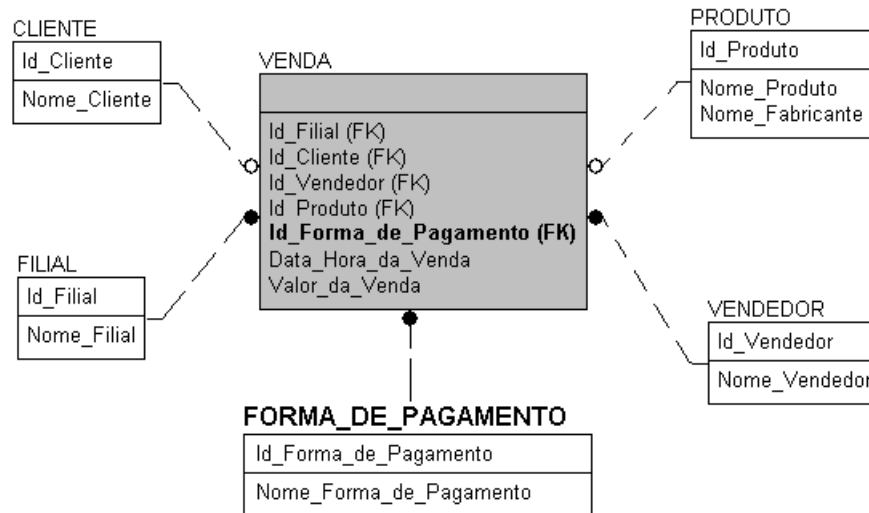


Figura 3.14 – Relacionamento da tabela fato com uma nova dimensão

Comparando a tabela fato VENDA da figura 3.13 com a tabela fato da figura 3.14 pode-se notar que é incluído o atributo Id_Forma_Pagamento que referencia a nova dimensão FORMA_DE_PAGAMENTO na tabela da figura 3.14, detalhando assim todas as instâncias da tabela fato.

Adaptabilidade: Nesta solução, a alteração da tabela fato não prejudica a configuração em estrela do esquema.

Legibilidade: Esta solução não prejudica a legibilidade do esquema.

3.5.2.2 – Criação de uma nova tabela fato relacionada com a nova dimensão

No caso anterior, o problema é facilmente resolvido, pois todas as instâncias da tabela fato necessitam ser detalhados pela nova dimensão, ou seja todas as instâncias possuem a mesma granularidade.

Mas existem casos onde o simples relacionamento da tabela fato com a nova dimensão pode gerar instâncias com granularidades diferentes dentro da tabela fato. Se, no exemplo, uma dimensão de operadora de cartão fosse relacionada à tabela fato VENDA, ocorreria este problema. Para ilustrar esta situação, a figura 3.15 apresenta um esquema em que a dimensão OPERADORA_CARTÃO está relacionada com a tabela fato VENDA e a tabela 3.7 apresenta um exemplo das instâncias desta tabela.

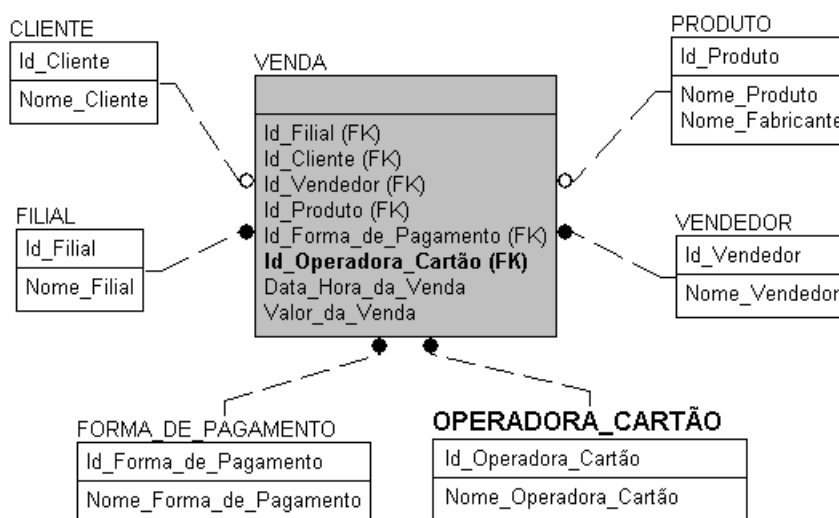


Figura 3.15 – Relacionamento da tabela fato VENDA com a dimensão OPERADORA_CARTÃO

Instância	Cliente	Produto	Forma de Pagamento	Operadora do Cartão	Demais Atributos
1	Cliente 1	Produto 1	Dinheiro		...
2	Cliente 2	Produto 2	Cheque		...
3	Cliente 3	Produto 3	Cartão	VISA	...

Tabela 3.7 – Conteúdo da tabela fato VENDA da fig. 3.16

Ao analisar a tabela 3.7 nota-se que as instâncias 1 e 2 possuem granularidades diferentes da instância 3, pois nesta última a forma de pagamento utilizada (cartão) tornou necessária a informação da operadora.

Por causa desta diferença de granularidades, não é correto relacionar a dimensão OPERADORA_CARTÃO com a tabela fato VENDA.

Para resolver este problema é apresentada uma nova solução, consistindo na criação de uma nova tabela fato que se relaciona com todas as dimensões da tabela fato original, mais a nova dimensão, e inserir nela somente as instâncias da tabela fato original que necessitarem ter a granularidade mais fina.

Adotando esta solução ao exemplo da figura 3.15, são incluídas a tabela fato VENDA_CARTÃO e as dimensões OPERADORA_CARTÃO e IDENTIFICADOR_VENDA, sendo esta última utilizada para criar um relacionamento 1:1 (um por um) entre as tabelas fato VENDA e VENDA_CARTÃO, obtêm-se o esquema apresentado na figura 3.16.

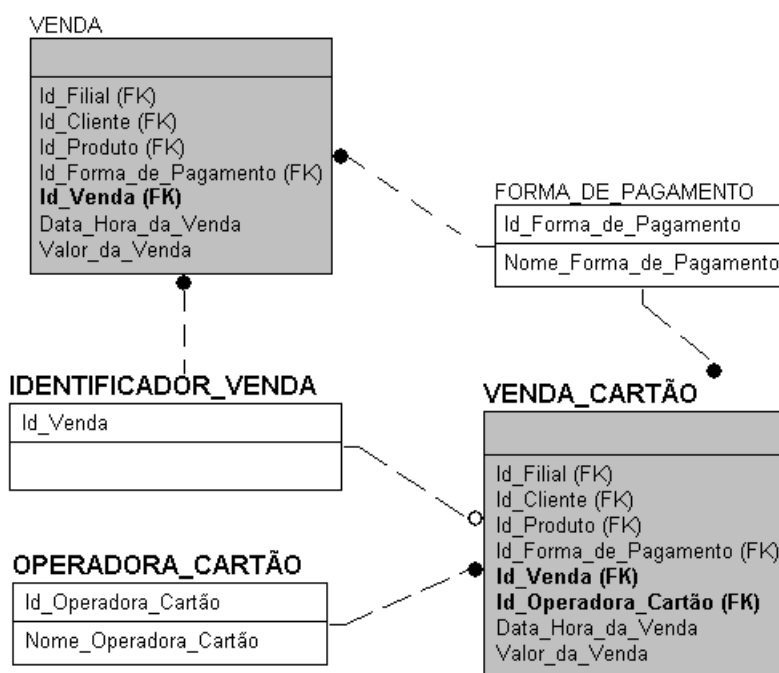


Figura 3.16 – Inclusão de uma nova tabela fato com a nova dimensão

Observando a figura 3.16 pode-se notar que existem dois esquemas estrelas, um com a tabela fato VENDA e outra com a tabela fato VENDA_CARTÃO.

Adotando esta solução, a tabela fato VENDA não tem sua granularidade alterada e, na tabela fato VENDA_CARTÃO estão inseridas as vendas que foram realizadas com cartão.

Adaptabilidade: Nesta solução é prejudicada a adaptabilidade do esquema estrela, pois, ao criar uma nova tabela fato, este esquema é transformado em uma constelação.

Legibilidade: A transformação para um esquema constelação pode prejudicar a legibilidade quando o usuário necessitar consultar os dois esquemas para obter a informação desejada.

3.5.2.3 – Criação de uma nova tabela fato relacionada com a tabela fato original e a nova dimensão

Na solução anterior, observa-se que existe uma redundância de dados na nova tabela fato, pois ela se relaciona com todas as dimensões da tabela fato original.

Ao invés da nova tabela fato se relacionar com todas as dimensões da tabela fato original, a nova pode relacionar-se diretamente com a tabela fato original. Para que isto possa ocorrer, a tabela fato original é transformada numa dimensão da nova tabela fato. Para a transformação da tabela fato original numa dimensão, cria-se uma PK (*Primary Key*) na tabela fato original, da mesma maneira como se cria em uma dimensão.

Utilizando esta solução no exemplo da figura 3.14, inclui-se a dimensão FORMA_DE_PAGAMENTO e a tabela fato VENDA_CARTÃO_NOVA, bem como uma PK na tabela fato VENDA. Em seguida, relaciona-se a tabela VENDA com a VENDA_CARTÃO_NOVA, como apresentado na figura 3.17.

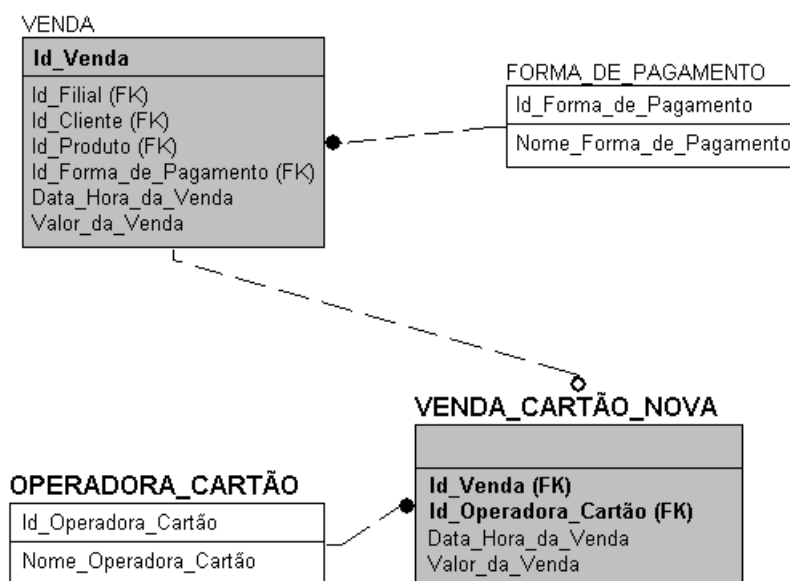


Figura 3.17 – Inclusão de uma nova tabela fato relacionada com a tab. fato original

Comparando esta solução com a da figura 3.17, observa-se, neste caso, que não foi necessário incluir a dimensão IDENTIFICADOR_VENDA, pois a PK da tabela fato VENDA executa a sua função de relacionar as duas tabelas fato.

Adaptabilidade: Nesta solução, a adaptabilidade do esquema estrela é prejudicada, pois, ao criar uma nova tabela fato este esquema é transformado em uma constelação.

Legibilidade: Nesta solução, tanto a transformação para um esquema constelação quanto a utilização da tabela fato original como dimensão da nova tabela fato podem prejudicar a legibilidade, dependendo da pesquisa realizada pelo usuário.

3.5.3 – Resumo das soluções

Somente a solução relacionamento da tabela fato com a nova dimensão do item 3.5.2.1 não prejudica a adaptabilidade do esquema estrela, uma vez que as

outras soluções apresentadas não conseguem resolver o problema sem transformar o esquema em uma constelação.

No aspecto de legibilidade, somente a mesma solução não a prejudica, pois as demais, ao transformarem o esquema em constelação, podem dificultar a sua compreensão pelo usuário.

A partir destas observações, nota-se que o esquema estrela suporta adaptações necessárias somente no caso onde seja necessário incluir uma nova dimensão que não cria instâncias com granularidades diferentes na tabela fato. Para os demais casos, este esquema falha ao tentar.

3.6 – Resumo e Considerações Gerais

Neste capítulo foram apresentadas e exemplificadas as mudanças estruturais identificadas pelo autor que podem ocorrer durante a evolução de um DW. Estas mudanças estruturais são: o relacionamento M:N entre a tabela fato e dimensão; o detalhamento de dimensões; o armazenamento de dados históricos na tabela dimensão; e a inclusão de uma nova dimensão.

Foram apresentadas soluções para cada mudança, sendo que para cada solução é mostrada a sua influência na legibilidade e adaptabilidade do esquema estrela.

Observa-se que dentre os problemas apresentados, apenas no item que se refere ao detalhamento de dimensões é possível apresentar soluções que não causam prejuízo tanto na adaptabilidade quanto na legibilidade do esquema estrela. Para os demais problemas, a quase totalidade das sugestões não conserva a configuração em estrela do esquema e praticamente todas prejudicam a legibilidade.

No capítulo 4 é apresentado um estudo de caso de uma empresa de atendimento médico detentora de um DW em esquema estrela, inserido em um ambiente em evolução. Para mostrar a evolução do DW são criadas situações em que as regras de negócio são alteradas.

A partir destas alterações nas regras de negócio, são determinadas quais soluções dentre as apresentadas neste capítulo podem ser utilizadas para resolver os novos problemas gerados.

CAPITULO 4 – ESTUDO DE CASO

4.1 – Introdução

Neste capítulo é apresentado um estudo de caso de uma empresa de assistência médica que possui um DW construído segundo o modelo estrela e submetido a um ambiente em evolução. O estudo de caso é utilizado para testar a adaptabilidade e legibilidade do DW frente a situações em que as regras de negócio definidas na sua construção sofrem alterações. A partir destas alterações, utilizam-se as soluções propostas no capítulo 3 para adequar o esquema à nova realidade.

4.2 – Descrição do DW existente

Uma empresa de assistência médica tem, como principal objetivo, oferecer a seus associados (clientes) todo o atendimento médico de que ele precisar. Para tanto, a empresa possui uma rede de prestação de serviços formada por diversos hospitais, clínicas e laboratórios.

Para analisar a utilização da rede de atendimento pelos seus associados, a empresa possui o DW da figura 4.1. Neste esquema são armazenadas as informações sobre os clientes atendidos, o local, o tipo, o diagnóstico, a data e o custo do atendimento.

Na figura 4.1 a tabela fato ATENDIMENTO mostra os atendimentos que os associados recebem. Neste esquema, as dimensões são: ASSOCIADO, LOCAL_DE_ATENDIMENTO, TIPO_DE_ATENDIMENTO e DIAGNÓSTICO. Na dimensão ASSOCIADO tem-se o Nome_Associado, a Data_Início (data em que o associado se vinculou à empresa), a Situação_Associado (identificação se ele está ativo ou cancelado sob o ponto de vista da empresa) e outros dados pessoais; a dimensão LOCAL_DE_ATENDIMENTO identifica o hospital, clínica ou laboratório em que o associado é atendido; o TIPO_DE_ATENDIMENTO (internação, consulta ou exame); e a dimensão DIAGNÓSTICO identifica a doença

diagnosticada durante um atendimento. Na tabela fato ATENDIMENTO ainda existem a Data_Hora_do_Atendimento e o Valor_do_Atendimento.

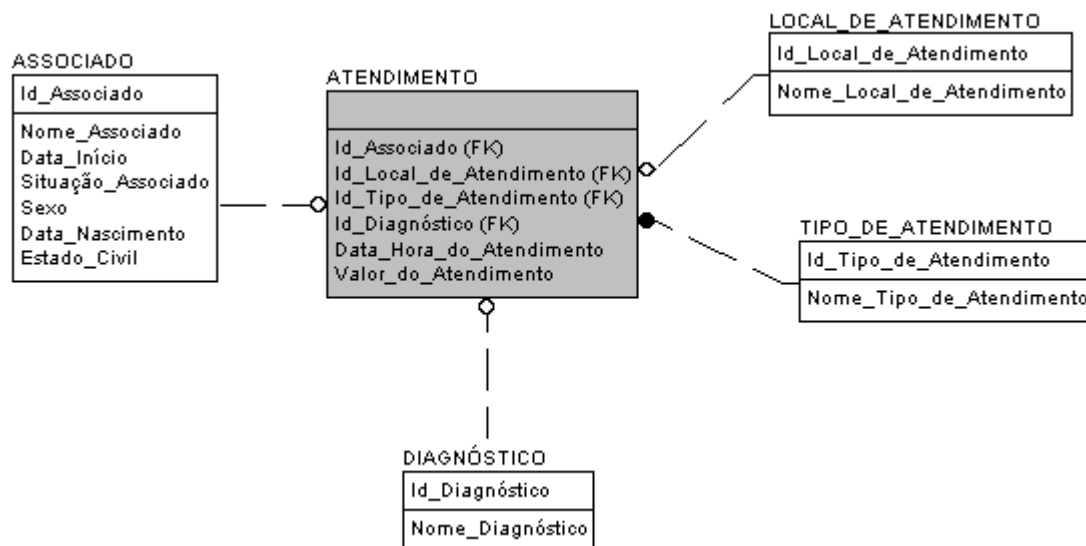


Figura 4.1 – DW de uma empresa de assistência médica.

4.3 – Necessidades de evolução

A seguir são apresentadas as mudanças nas regras de negócio decorrentes de novas necessidades, e para adequar o DW a estas alterações, são utilizadas as adaptações sugeridas no capítulo anterior.

4.3.1 – Mudança na legislação

Segundo a legislação vigente, toda empresa de assistência médica é obrigada a enviar regularmente informações sobre seus atendimentos para a Agencia Nacional de Saúde (ANS). Com o passar do tempo, a ANS passou a exigir as seguintes informações sobre os atendimentos de internação: quantidade de diárias de internação e tipo de acomodação. O tipo de acomodação pode ser quarto, enfermaria ou unidade de terapia intensiva.

Para contemplar esta situação é preciso decidir em qual estrutura do esquema deve-se inserir estas informações. Num primeiro momento pode-se imaginar que criar novos atributos na tabela fato ATENDIMENTO e relacionando-a com a nova dimensão TIPO_DE_ACOMODAÇÃO, como apresentado na figura 4.2, resolveria o problema.

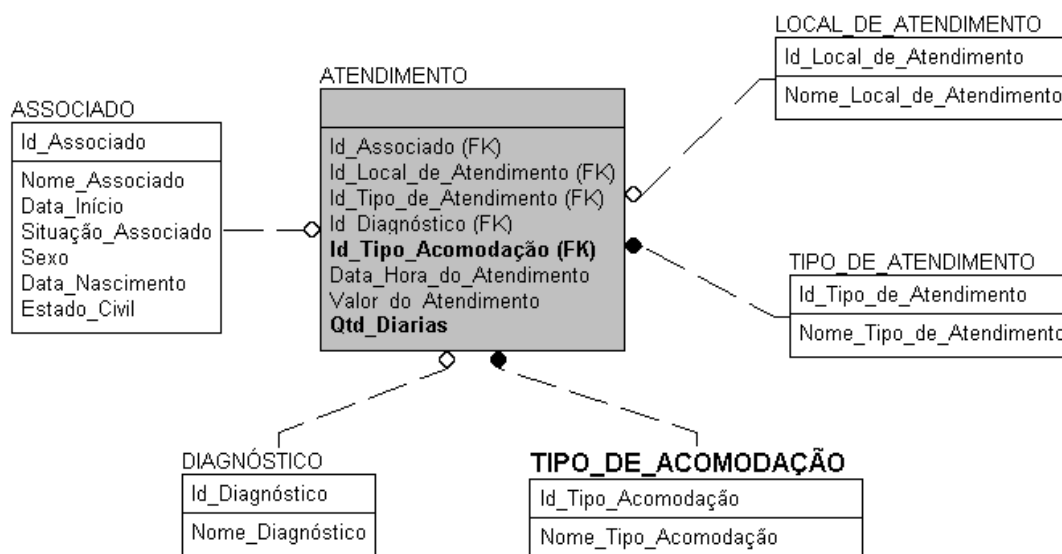


Figura 4.2 – Inclusão dos atributos de internação diretamente na tabela fato

A solução apresentada na figura 4.2, no entanto, teria como consequência armazenar granularidades diferentes dentro da tabela fato ATENDIMENTO para as internações e os demais atendimentos, uma vez que os atributos **Qtd_Diárias** e **Id_Tipo_Acomodação** não seriam utilizados para todos os atendimentos.

Como a solução apresentada na figura 4.2 não é satisfatória, uma outra sugestão é armazenar as informações de internações em uma nova tabela fato, não causando assim nenhum problema de granularidade para a tabela fato ATENDIMENTO.

Adotada esta solução, é criada a nova tabela fato ATENDIMENTO_INTERNAÇÃO com os atributos **Qtd_Diárias** e **Id_Tipo_Acomodação**. Porém torna-se necessário identificar a qual atendimento estas novas informações estão

relacionadas e para isto, é criada uma PK (*Primary Key*) *Id_Atendimento* na tabela fato *ATENDIMENTO* transformando-a numa dimensão da nova tabela fato. Também é inserida a dimensão *TIPO_DE_ACOMODAÇÃO* e o esquema resultante é representado na figura 4.3.

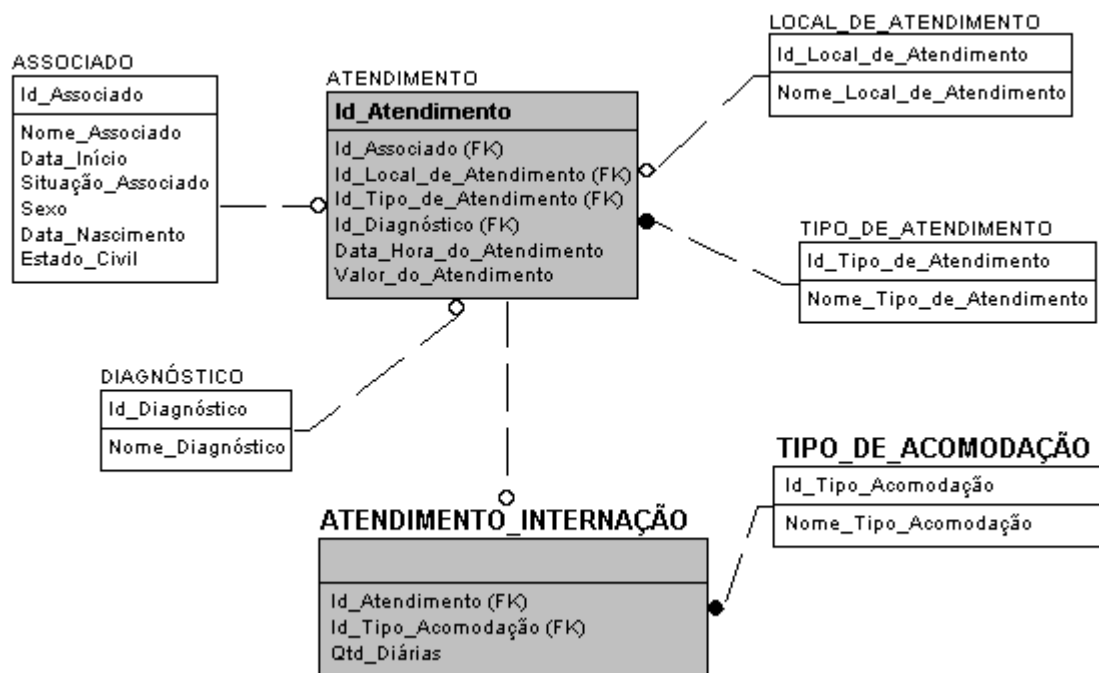


Figura 4.3 – Inclusão das informações de internação

Com esta solução, quando ocorrer um atendimento que não seja internação será incluída uma instância apenas na tabela fato *ATENDIMENTO*, enquanto que se ocorrer uma internação, serão incluídas instâncias nas duas tabelas fato.

Este problema ocasionado pela mudança na legislação é similar ao problema ‘Inclusão de uma Nova Dimensão’ (item 3.5) apresentado no capítulo anterior, e a solução aqui adotada é semelhante a que cria uma nova tabela fato relacionada com a tabela fato original e a nova dimensão (item 3.5.2.3).

4.3.2 – Inserção de informações sobre pagamento

O associado, para ter direito ao atendimento, deve se vincular à empresa e realizar o pagamento das mensalidades. Com o intuito de analisar a receita, a empresa decide inserir os dados sobre os pagamentos realizados pelos seus associados no DW.

Contudo, o DW existente armazena apenas informações relativas ao atendimento dos associados e, como os dados de pagamentos de mensalidades não são relacionados com os de atendimento, é necessário construir um novo esquema estrela.

Desta forma, é inserido um novo esquema estrela com a nova tabela fato PAGAMENTO, com os atributos Data_Pagamento e Valor_Pagamento. Esta nova tabela fato também relaciona-se com uma única dimensão, a ASSOCIADO. O novo esquema é apresentado na figura 4.4.

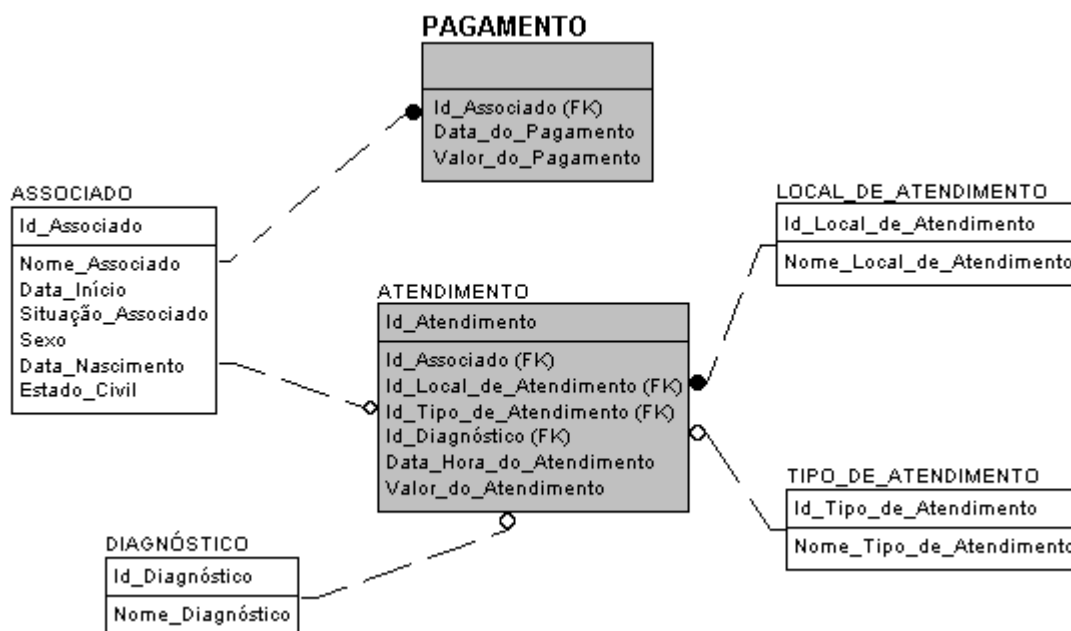


Figura 4.4 – Inclusão das informações sobre pagamento

4.3.3 – Determinação do plano do associado

A empresa possui diversos planos que determinam as redes de atendimento e benefícios a que os associados tem direito. Cada associado está vinculado a um plano, que por sua vez determina o valor da mensalidade.

Este DW, contudo, não possui a informação sobre o plano do associado e esta informação se torna necessária para realizar algumas análises.

Como cada associado pode possuir apenas um plano, uma solução para este problema é detalhar a dimensão ASSOCIADO inserindo o atributo Nome_Plano, conforme apresentado na figura 4.5.

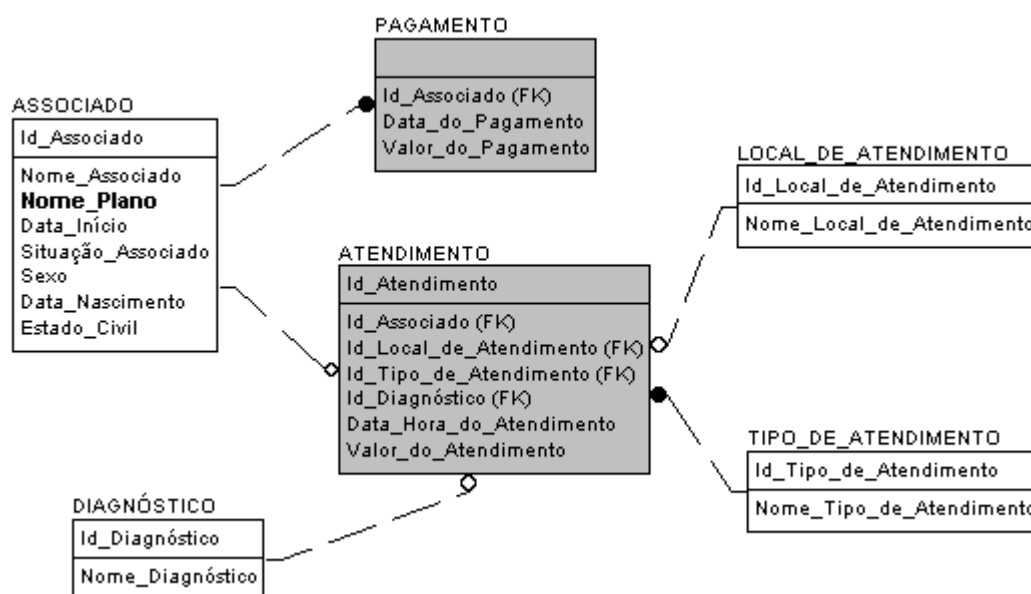


Figura 4.5 – Inclusão das informações do plano do associado

A necessidade de incluir a informação a respeito do plano torna este problema parecido com o problema ‘Detalhamento de dimensões’ (item 3.3) e a solução adotada é a única que o capítulo anterior sugere no item 3.3.2.1.

4.3.4 – Necessidade de manter o histórico dos planos do associado

A partir de um determinado momento, a empresa passa a analisar a evolução do número de associados por plano. Para fazer esta análise é criado um relatório que mostra o número de associados divididos por mês e plano, para um período de três meses. A tabela 4.1 apresenta um exemplo deste relatório emitido no mês de abril, mostrando o período de janeiro a março.

Nome do Plano	Janeiro	Fevereiro	Março
Plano 1	500	600	600
Plano 2	200	200	250
Plano 3	300	300	350
Total de Associados	1000	1100	1200

Tabela 4.1 – Relatório do número de associados emitido em abril

Entretanto, se o mesmo relatório retratando o mesmo período (de janeiro a março) for emitido no mês de maio, o número de associados por plano por mês será diferente (mas não se altera o número total de associados por mês). A tabela 4.2 apresenta outro exemplo deste relatório agora emitido no mês de maio mostrando o período de janeiro a março.

Nome do Plano	Janeiro	Fevereiro	Março
Plano 1	550	650	650
Plano 2	200	250	300
Plano 3	250	200	250
Total de Associados	1000	1100	1200

Tabela 4.2 – Relatório do número de associados emitido em maio

Comparando a tabela 4.2 com a tabela 4.1, observa-se que na tabela 4.1 o número de associados do plano 1 para o mês de janeiro é 500 enquanto que na tabela 4.2 neste mesmo plano neste mesmo mês o número é de 550.

Isto acontece porque os associados migraram de um plano para outro entre os meses de abril e maio e a tabela dimensão ASSOCIADO armazena apenas a informação atual do associado, ocasionando assim esta diferença nos resultados. Neste exemplo pode-se observar que no intervalo entre as impressões (entre o meses de abril e maio) 50 pessoas migraram do plano 3 para o plano 1.

Para poder analisar a informação ao longo do tempo, é necessário armazenar o histórico de migração de planos pelo associado.

Para armazenar este histórico é criada uma nova tabela fato ASSOCIADO_PLANO que se relaciona com a dimensão ASSOCIADO e é transferido para esta tabela fato o atributo Nome_Plano. Para controlar quando se inicia e se cancela um plano, também são inseridos os atributos Data_Inicio_Plano e Data_Fim_Plano. O esquema resultante é apresentado na figura 4.6

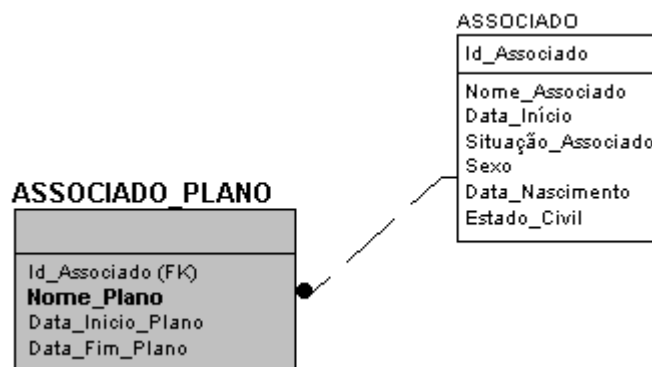


Figura 4.6 – Inclusão do histórico do plano do associado

Este armazenamento de históricos dos planos é um problema semelhante ao ‘Armazenamento de dados históricos na tabela dimensão’ do item 3.4 e a solução

adotada aqui é similar à que cria uma nova tabela fato para armazenar dados históricos (item 3.4.2.2).

4.3.5 – Inclusão de um novo serviço

Visando conquistar novos mercados, a empresa decide comercializar outros serviços, em particular a assistência odontológica.

A incorporação da assistência odontológica incorre na necessidade de vincular os associados aos seus serviços e planos, sendo que cada serviço possui o seu próprio plano.

Com esta nova realidade, a primeira necessidade é controlar os pagamentos, identificando para qual serviço a mensalidade é paga.

Para resolver este problema é detalhada a tabela fato PAGAMENTO, inserindo um atributo que identifica para qual serviço a mensalidade está sendo paga. No esquema da figura 4.7 é mostrada a tabela fato PAGAMENTO relacionando-se com a nova dimensão SERVIÇO.

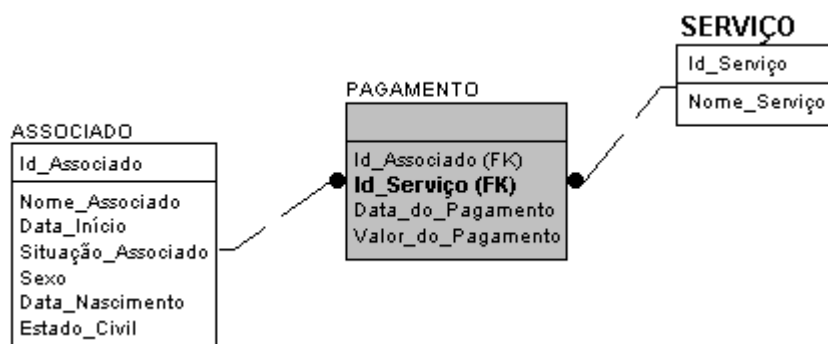


Figura 4.7 – Detalhamento do pagamento com a inclusão da dimensão serviço

Esta necessidade é semelhante ao problema ‘Inclusão de uma nova dimensão’ (item 3.5) e a solução utilizada aqui é equivalente ao relacionamento da tabela fato com uma nova dimensão do item 3.5.2.1.

A segunda necessidade é identificar, para cada serviço, qual plano o associado possui. Como neste esquema existe uma tabela fato (ASSOCIADO_PLANO) que já identifica este plano, é preciso apenas detalhá-la, inserindo um atributo que identifique o serviço. Realizada esta alteração, o esquema resultante é apresentado na figura 4.8.

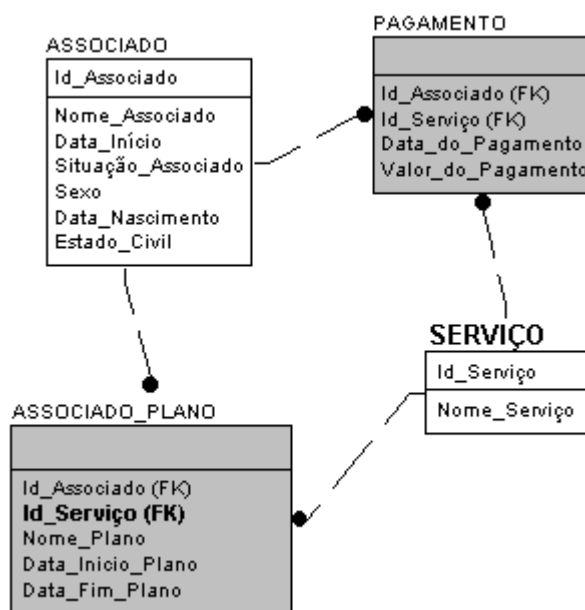


Figura 4.8 – Relacionamento da dimensão SERVIÇO com a tabela fato ASSOCIADO_PLANO

A necessidade de identificar para cada serviço qual plano o associado possui também é semelhante ao problema ‘Inclusão de uma nova dimensão’ (item 3.5) e a solução utilizada aqui é equivalente ao relacionamento da tabela fato com uma nova dimensão do item 3.5.2.1 proposta no capítulo 3.

A última necessidade devida à incorporação da assistência odontológica ainda não atendida é classificar os atendimentos por serviço. Neste caso, assim como nos dois últimos, é necessário incluir uma dimensão na tabela fato. Para isso, inclui-se o atributo que identifica o serviço na tabela fato ATENDIMENTO e o esquema resultante é apresentado na figura 4.9.

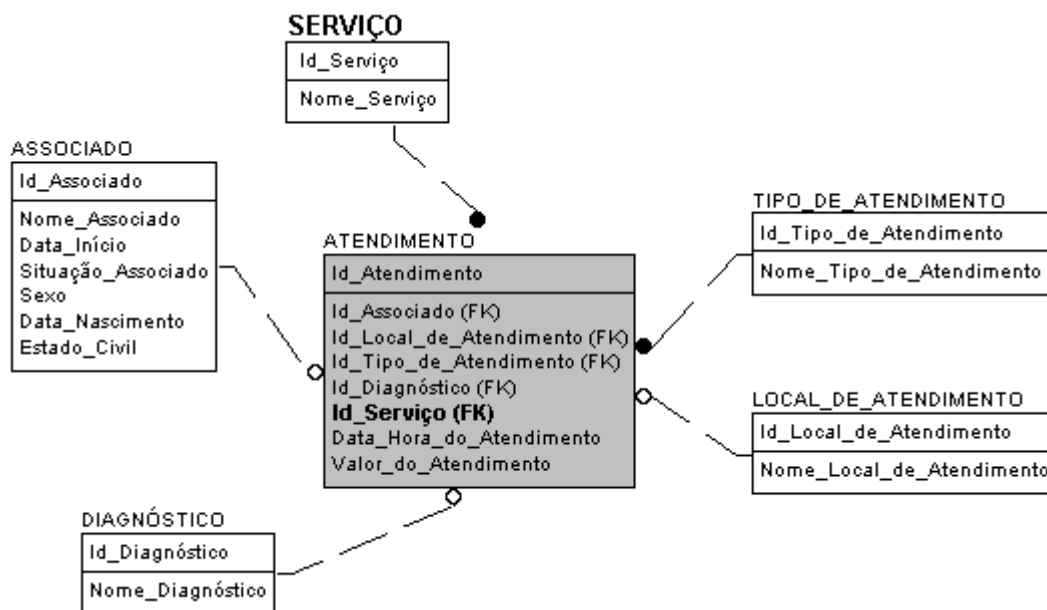


Figura 4.9 – Detalhamento do atendimento com a inclusão da dimensão serviço

4.3.6 – Detalhamento do atendimento

Feitas algumas análises, a empresa verifica que os atendimentos realizados em certos locais são mais caros que em outros. Para poder pesquisar esta diferença de valores, a empresa precisa detalhar o atendimento. O detalhamento de um atendimento consiste em incluir os procedimentos executados e seus custos. Procedimento é tudo aquilo que é executado em um atendimento; por exemplo, em uma internação, pode ser uma cirurgia, uma visita hospitalar, um aplicação de injeção, etc.; cada atendimento envolve um ou mais procedimentos. Para esta empresa, somente os dados do serviço de assistência médica são detalhados por procedimentos.

Para solucionar este caso é necessário detalhar a tabela fato ATENDIMENTO; contudo a informação sobre o procedimento não pode ser inserida diretamente nesta tabela, pois isso iria acarretar em um armazenamento de diferentes granularidades entre os serviços de assistência médica e odontológica.

Então, uma solução viável é criar uma nova tabela fato ATENDIMENTO_ PROCEDIMENTO para armazenar estas informações. O esquema da figura 4.10

apresenta esta nova tabela fato e a nova dimensão PROCEDIMENTO. Assim como na solução da inclusão dos dados de internação (item 4.3.1), nesta solução a tabela fato ATENDIMENTO também se torna uma dimensão desta nova tabela fato.

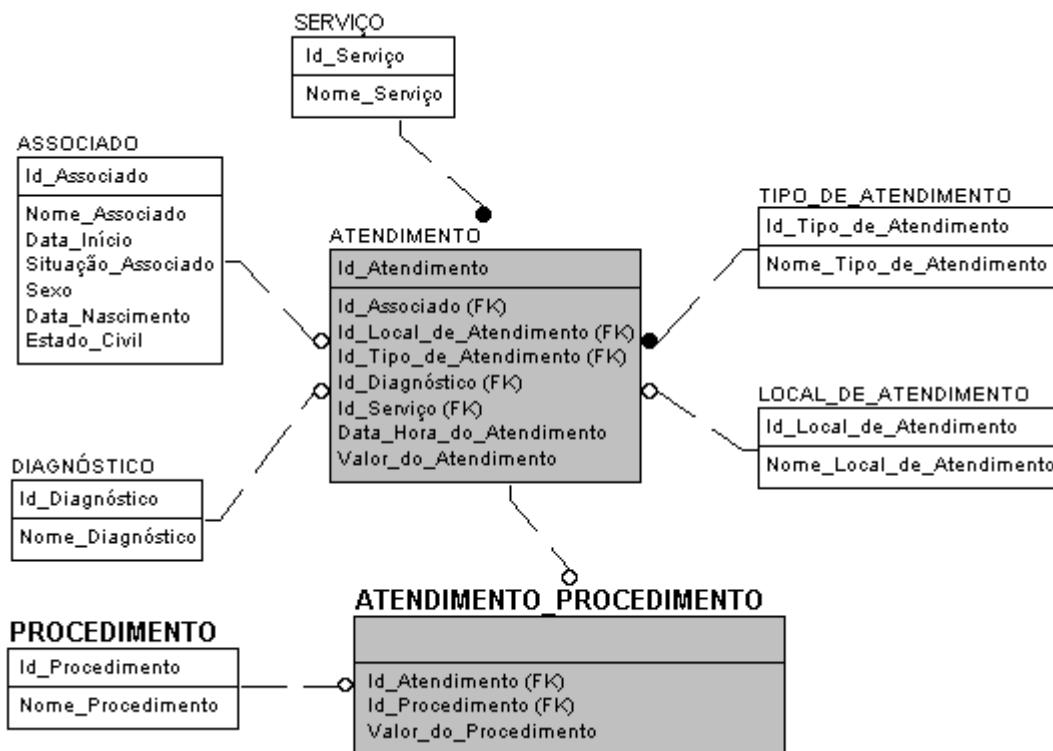


Figura 4.10 – Inclusão das informações sobre o procedimento

Observando a tabela fato **ATENDIMENTO_PROCEDIMENTO** da figura 4.10 nota-se que foi incluído o atributo **Valor_do_Procedimento** que armazena o valor do seu custo.

Este problema é semelhante ao item 3.5 (Inclusão de uma nova Dimensão) e a solução utilizada identifica-se com a solução ‘Criação de uma nova tabela fato relacionada com a tabela fato original e a nova dimensão’ do item 3.5.2.3.

4.3.7 – Determinação das doenças mais diagnosticadas

Até este ponto a cada atendimento diagnosticava-se apenas uma doença. Com o tempo, entretanto, a empresa passou a admitir mais de uma doença por atendimento. O DW, entretanto, limitava o armazenamento a apenas um por atendimento, prejudicando desta maneira as pesquisas dos usuários.

A necessidade de armazenar todos os diagnósticos tem como solução a alteração do relacionamento entre a tabela fato ATENDIMENTO e a dimensão DIAGNÓSTICO de 1:N para M:N.

Entretanto, como visto no capítulo 3, existem diversas soluções para o relacionamento M:N entre as tabelas fato e dimensão. Porém, muitas das soluções propostas são aplicáveis apenas quando a dimensão possui poucas instâncias, o que não ocorre com a dimensão DIAGNÓSTICO desta empresa.

Outra solução seria criar tabelas pontes (item 3.2.2.1), gerando o inconveniente de prejudicar significativamente a adaptabilidade e legibilidade.

Poderia-se ainda utilizar a solução que indica grupos dentro da tabela fato (item 3.2.2.3), mas como esta altera a semântica da tabela, pois armazenando os dados de diagnósticos por atendimento, altera significativamente do esquema até aqui desenvolvido. A figura 4.11 apresenta o esquema com tabela fato ATENDIMENTO e as tabelas associadas onde verifica-se este problema.

Observando a figura 4.11, nota-se que se a semântica da tabela fato ATENDIMENTO for alterada para armazenar os diagnósticos por atendimento, as demais tabelas fatos devem ser adaptadas a esta nova semântica. Devido a este problema, esta solução também não é a mais indicada.

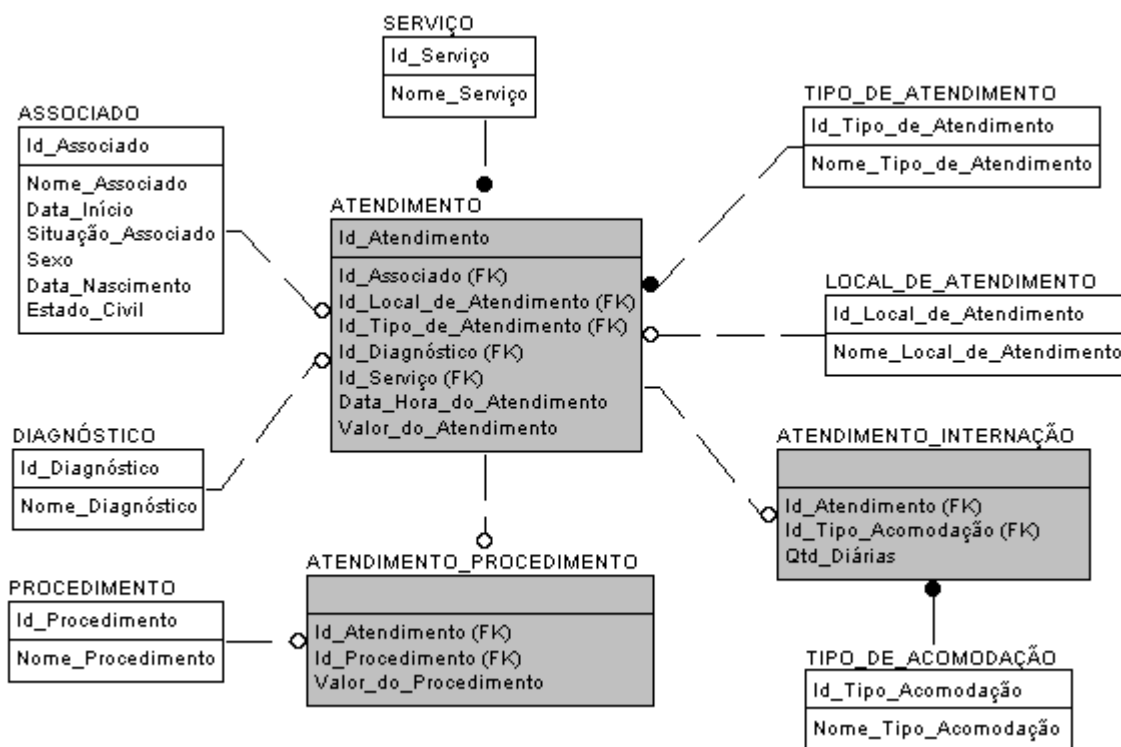


Figura 4.11 – Tabela fato ATENDIMENTO e relacionadas

A solução que indica grupos em uma nova tabela fato (item 3.2.2.4) é a que menos interfere no esquema, tendo uma melhor adaptabilidade e boa legibilidade.

Adotada esta solução, é criada a nova tabela fato **ATENDIMENTO_DIAGNÓSTICO** e relacionada com as tabelas fato **ATENDIMENTO** e dimensão **DIAGNÓSTICO** conforme apresentado na figura 4.12.

Comparando as figuras 4.12 e 3.5 (figura do item 3.2.2.4), observa-se que na figura 4.12, as tabelas fato estão relacionadas diretamente entre si, enquanto que na figura 3.5 elas possuem uma tabela auxiliar para desempenhar esta função. Esta diferença existe pois a tabela fato **ATENDIMENTO** da figura 4.12 já possuía uma PK, o que auxiliou na simplificação deste relacionamento.

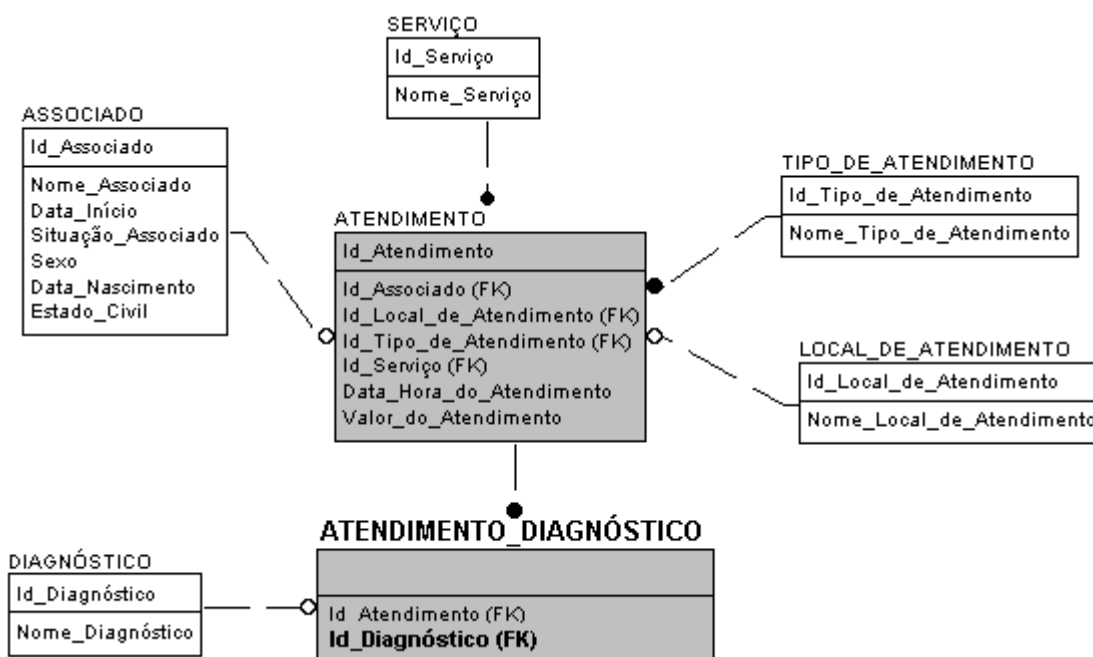


Figura 4.12 – Solução do relacionamento M:N entre a tabela fato e dimensão

4.4 – O DW após a evolução

Após serem realizadas todas as mudanças necessárias no esquema para contemplar a evolução do DW da empresa de assistência médica, o esquema resultante é o apresentado na figura 4.13.

Analisando o esquema resultante, observa-se que ele é transformado em uma constelação, isto devido ao fato de que a maioria das soluções cria novas tabelas fato, mostrando que a adaptabilidade do esquema inicial falha ao não conseguir preservar a sua configuração em estrela.

Em relação à legibilidade, esta solução, ao possuir diversas tabelas fatos relacionadas entre si, cria um esquema que dificulta a sua compreensão pelo usuário.

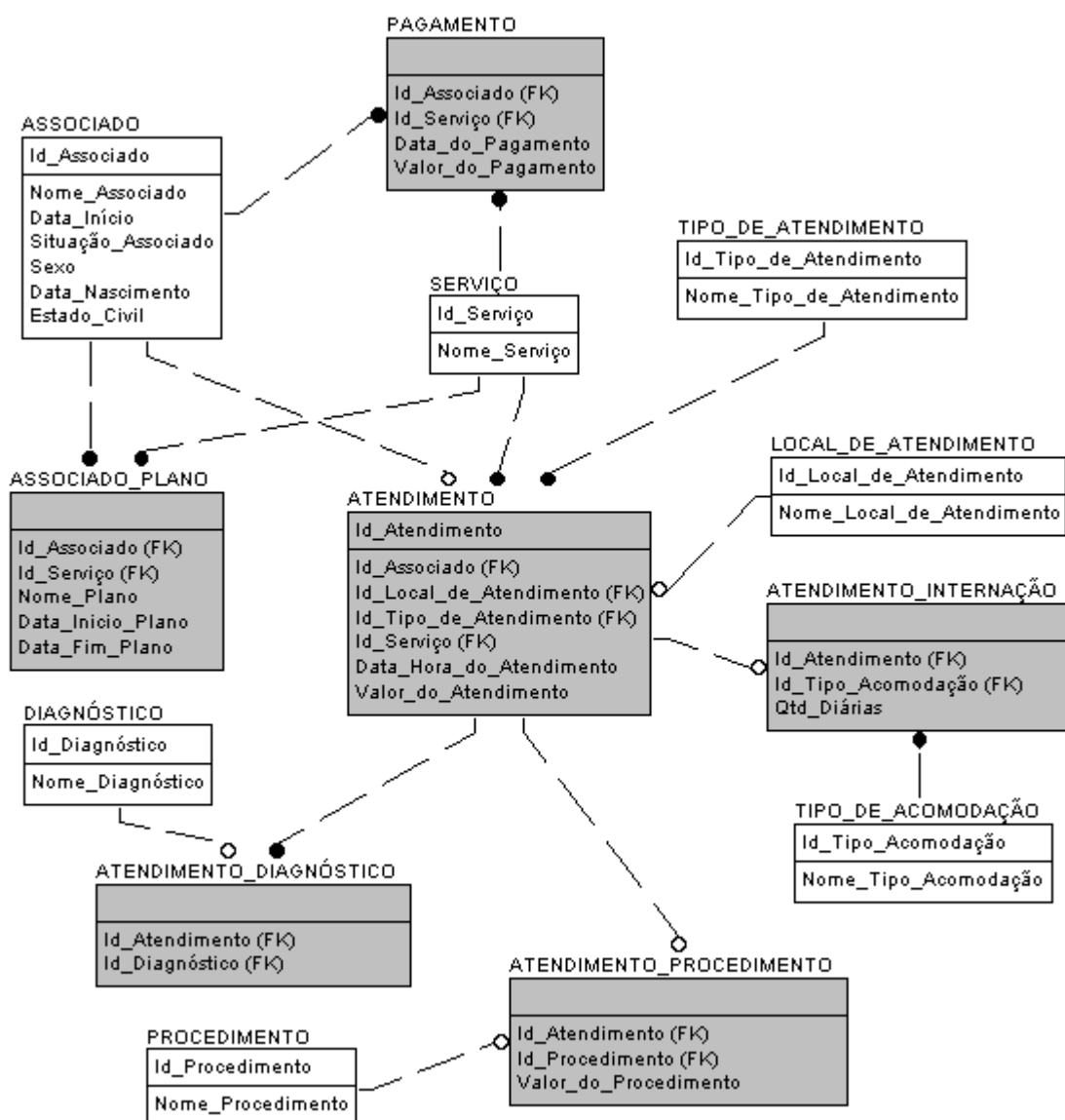


Figura 4.13 – Esquema resultante do estudo de caso

4.5 – Resumo e Considerações Gerais

Neste capítulo foi apresentado um estudo de caso em que o DW de uma empresa de assistência médica foi submetido a uma série de mudanças provocadas por alterações nas regras de negócio mostrando assim o seu comportamento em um ambiente em evolução.

A cada mudança nas regras de negócio, soluções são propostas e, depois de implementadas, são comparadas com as sugeridas no capítulo anterior.

Observando a figura 4.13 nota-se que em termos de adaptabilidade o esquema final deixa de ser uma estrela e transforma-se em uma constelação sendo que a legibilidade do esquema é prejudicada pelo grande número de relacionamentos existentes entre as tabelas.

No próximo capítulo é feita uma análise dos resultados obtidos e também algumas considerações sobre a conveniência da adoção da estrela como modelo para a representação de esquemas corporativos.

CAPITULO 5 – ANÁLISE DOS RESULTADOS

5.1 – Introdução

Observando o que ocorre com o esquema estrela após as mudanças executadas tanto no capítulo 3 quanto no 4, nota-se que este esquema é degenerado na maioria das alterações a que é submetido.

A seguir, o modelo estrela é analisado sob a ótica da adaptabilidade e legibilidade.

5.2 – Comentários sobre a adaptabilidade do modelo estrela

A adaptabilidade do modelo estrela, como visto no capítulo 2, é a qualidade que o modelo tem de manter as suas características, independentemente das mudanças a que é submetido, lembrando que algumas destas principais características são a de possuir apenas uma tabela fato, e a de manter o relacionamento 1:N entre a tabela fato e as dimensões.

Nas modificações realizadas nos capítulos 3 e 4, o DW teve, na maioria das vezes, o seu esquema estrela original degenerado pela criação de mais de uma tabela fato. Como consequência destas alterações o esquema inicial é transformado em uma constelação, de forma a evitar granularidades diferentes nas métricas de uma mesma tabela fato.

Assim sendo, existem diversas situações em que torna-se necessário alterar o esquema, de forma que este perde as características do modelo estrela. Essa degeneração evidencia que o modelo estrela é pouco flexível a mudanças.

5.3 – Comentários sobre a legibilidade do esquema estrela

Em um DW em evolução, como visto desde o capítulo 3, a maioria das soluções propostas acaba tendo a sua legibilidade prejudicada pela criação de novas tabelas fato, muitas vezes relacionadas entre si. Estes relacionamentos adicionais tornam o esquema mais complexo do que a estrela original.

Analisando o esquema da figura 5.1 (que é o mesmo esquema apresentado na figura 4.13, mas sem os atributos) sob o aspecto da legibilidade, pode-se observar que a existência de relacionamentos entre tabelas fato pode dificultar a compreensão do usuário. Essa dificuldade advém da existência de uma tabela fato entre a dimensão e a tabela fato que se deseja pesquisar, pois, com isso, é necessário realizar mais *joins* entre as tabelas. Como exemplo, se o usuário necessitar saber quais procedimentos foram realizados em um local de atendimento, é necessário iniciar a pesquisa pela tabela fato ATENDIMENTO_PROCEDIMENTO depois relacioná-la com a tabela fato ATENDIMENTO e finalmente relacionar esta última com as dimensões PROCEDIMENTO e LOCAL_DE_ATENDIMENTO.

Mesmo optando por soluções que não prejudiquem muito a legibilidade, nem sempre é possível ter um esquema final em que a estrela não fique comprometida, pois naturalmente qualquer aumento de escopo do DW torna-o mais complexo e menos legível.

Como visto no capítulo 2, segundo Dowling (Dowling, Schuff e Louis, 2001) um esquema estrela é mais legível que um esquema ER por possuir relacionamentos mais fáceis de serem lembrados pelos usuários finais.

Um ambiente em evolução, no entanto, acaba invalidando esta afirmação impondo a necessidade de modelos mais flexíveis, que permitam a representação de esquemas mais gerais que de fato possam representar as regras de negócio em toda a sua complexidade, e que permitam incorporar as mudanças decorrentes da evolução sem lançar-se mão de artifícios para tal, como no caso estrela, na tentativa, muitas vezes frustrada, de conservar sua configuração original.

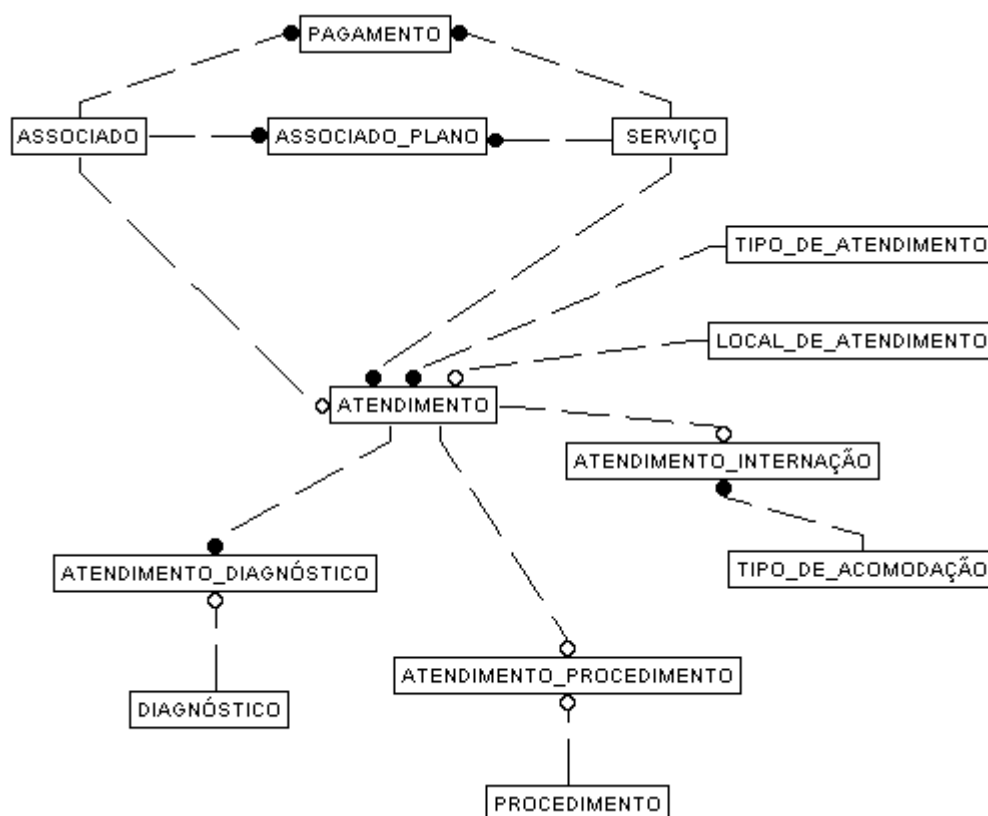


Figura 5.1 – Esquema resultante do estudo de caso sem os atributos

5.4 – Análise da constelação

Observando o esquema constelação da figura 5.1, nota-se que este esquema é bem mais flexível que a estrela, uma vez que existem diversas tabelas fatos, cada qual representando um assunto diferente e diversas dimensões que são compartilhadas pelas tabelas fato. Além disso, pode-se observar que existem relacionamentos entre tabelas fato e a tabela ATENDIMENTO tornou-se uma dimensão destas tabelas fato.

Atentando-se às figuras 4.13 e 5.1, estes esquemas podem ser reinterpretados sob a ótica do modelo ER. Assim, as tabelas fato, com exceção da ATENDIMENTO, são associações entre duas entidades, que no caso estrela corresponde a duas dimensões. A tabela fato ATENDIMENTO, por possuir uma PK, pode ser considerada uma entidade. O esquema ER resultante é apresentado na figura 5.2 (No

apêndice 2 encontra-se uma explicação sobre a simbologia utilizada no esquema ER).

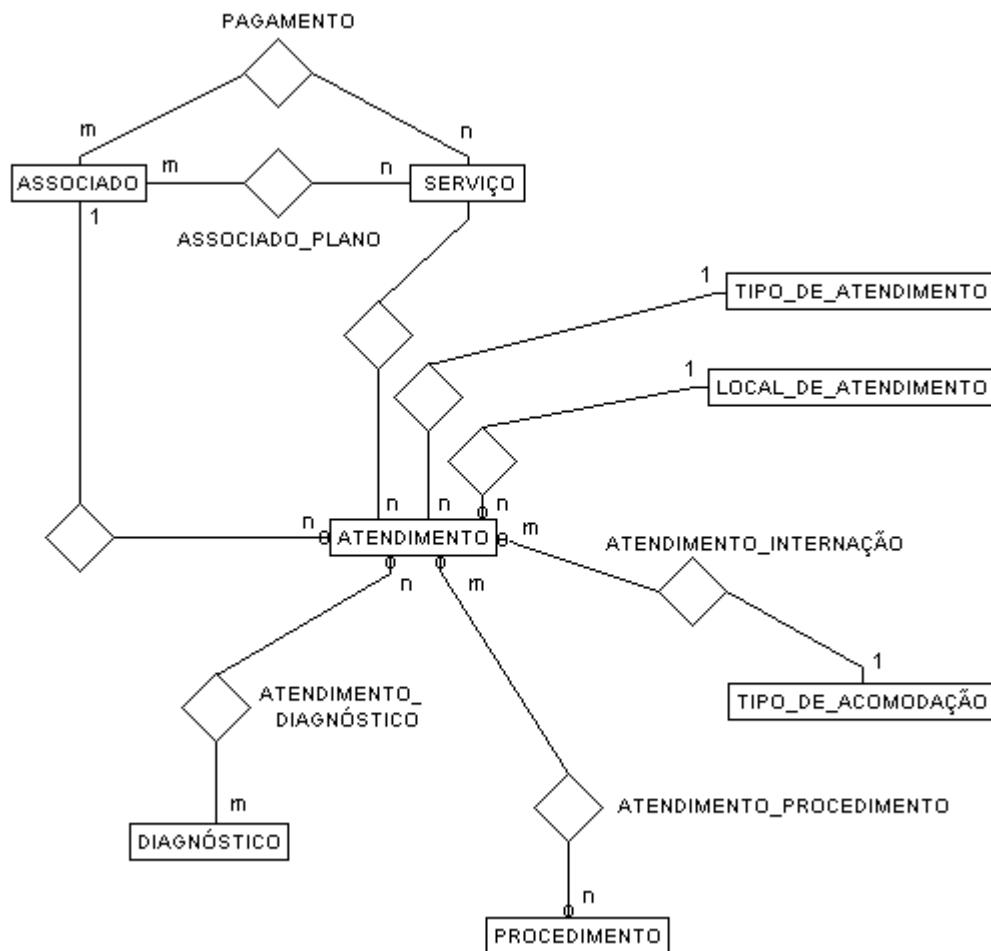


Figura 5.2 – Esquema ER

Comparando-se o esquema ER representado na figura 5.2 e a constelação da figura 5.1, percebe-se grande similaridade entre ambos e verifica-se que eles são mais adequados para a modelagem de um DW, uma vez que permitem fazer representações mais simples e são mais flexíveis, pois ambos permitem a incorporação de mudanças com o acréscimo (ou eliminação) de tabelas sem ter que se restringir rigidamente a uma configuração em particular, como no caso estrela. Entretanto, nota-se que no ER percebe-se a diferença entre entidades e

relacionamentos, o que não pode ser visto na constelação, tornando assim o ER semanticamente mais rico.

A partir do esquema ER da figura 5.2, verifica-se que é possível derivar visões em esquemas estrela. Essas visões, por serem mais simples de serem utilizadas e compreendidas, podem ser aplicadas em datamarts (DM) que foquem apenas um assunto.

5.5 – Resumo e Considerações Finais

Neste capítulo, a discussão sobre adaptabilidade e legibilidade do esquema estrela mostra que este é pouco flexível a mudanças. Essa falta de flexibilidade permite constatar que o esquema estrela se degenera muito facilmente quando submetido a um ambiente em evolução.

Comparando o esquema constelação resultante do estudo de caso com um esquema ER correspondente, nota-se que existe uma grande similaridade entre os dois, sendo que o esquema ER é semanticamente mais rico.

Das análises realizadas em um DW em evolução, foi verificado que o esquema estrela inicial não conseguiu conservar a sua configuração original. Também é possível concluir que o modelo ER é o mais adequado como modelo de um DW corporativo. Entretanto, apesar da inadequação do modelo estrela quanto à adaptabilidade, este tem a grande vantagem de possuir uma boa legibilidade o que o confirma como apropriado para a construção de visões, permitindo o seu uso em datamarts derivados do DW.

Este resultado vem a contrariar diretamente os autores Kimball (1997), Corey *et al* (2001), Kelly (2002), Winsberg (1996), Moody e Kortink (2000), Han e Kamber (2001), Schouten (1999) e Humphries, Hawkins e Dy (1999), que sugerem a utilização de um esquema estrela em DW corporativo.

CAPITULO 6 – CONCLUSÃO

6.1 – Resumo

A maior parte da bibliografia que trata sobre DW sugere apenas o uso do modelo estrela como modelo principal de um DW.

Para analisar este fato foram definidos os conceitos adotados pelo autor neste trabalho: DW, modelos estrela, floco de neve e constelação e os atributos adaptabilidade e legibilidade.

Em seguida foram apresentadas as condições que podem ocasionar mudanças nas regras de negócio que podem gerar novas necessidades que devem ser contempladas pelo DW.

Como resultado deste estudo, foram apresentadas soluções para estas novas necessidades, sendo algumas baseadas em artigos (Song et al, 2001; Pedersen, 1999a; Kimball, 1997; Kimball, 1996) e outras sugeridas pelo autor deste trabalho. Essas soluções acarretam mudanças estruturais no esquema estrela e foram comentadas quanto aos aspectos de adaptabilidade e legibilidade. Analisando os comentários verificou-se que a grande maioria degenerava o esquema estrela inicial e prejudicava a legibilidade. Assim foi possível mostrar que somente poucas soluções eram satisfatórias tanto para a adaptabilidade e legibilidade, concluindo que o esquema estrela não consegue contemplar adequadamente a maioria das necessidades decorrentes do processo de evolução.

Para ilustrar o comportamento de um DW em evolução construído inicialmente segundo o modelo estrela, foi realizado um estudo de caso de uma empresa de assistência médica onde seu DW foi submetido a uma série de mudanças. Representadas todas as alterações o esquema resultante mostrou o modelo estrela transformou-se em uma constelação.

A partir desta constatação foi possível compará-lo a um esquema ER (representando as mesmas informações) e observar a sua similaridade. Aproveitando-se essa semelhança e a característica de flexibilidade do ER, conclui-se que este esquema é adequado para ser utilizado em DW corporativos sujeitos a evolução. Dessa forma, o resultado deste trabalho é diferente da recomendação usual dos autores Kimball (1997), Corey *et al* (2001), Kelly (2002), Winsberg (1996), Moody e Kortink (2000), Han e Kamber (2001), Schouten (1999) e Humphries, Hawkins e Dy (1999), que sugerem o uso do esquema estrela em DW corporativos.

Finalmente, foi visto que o modelo estrela, por ser mais simples de ser utilizado e compreendido, pode ser aplicado em DM que foquem apenas em um assunto.

6.2 – Contribuições

Este trabalho apresentou soluções para algumas necessidades de mudanças estruturais que podem ocorrer durante a evolução de um DW comentando-as sob os aspectos de adaptabilidade e legibilidade visando auxiliar o analista a escolher as soluções que melhor se aplicam ao seu projeto.

O estudo de caso mostrou que foi necessário criar artificios para adequar o esquema do DW (construído segundo o modelo estrela) ao processo de evolução que no final degenerou o esquema transformando-o numa constelação. Com isso mostrou-se que o esquema estrela não suporta as adaptações necessárias durante a evolução.

A principal contribuição deste trabalho foi mostrar que o modelo ER é a melhor opção para se construir DW em evolução, uma vez que este modelo é flexível o bastante para incorporar as mudanças decorrentes da evolução. Além disso, comparando o modelo ER com uma constelação, observou-se que o ER se apresenta como modelo semanticamente mais rico.

6.3 – Sugestões para Pesquisas Futuras

Algumas sugestões para futuras pesquisas que podem ser feitas a partir deste trabalho são:

- Adoção do modelo constelação como opção, analisando todos os aspectos de qualidade;
 - Derivação de esquemas estrelas a partir de um DW, construído segundo o modelo ER, com ou sem materialização; e
 - Construção de estrelas com visão materializada – e a manutenção delas com mudanças no DW corporativo;
-

REFERÊNCIAS BIBLIOGRÁFICAS

BALLARD, C. *et al.* **Data Modeling Techniques for Data Warehousing.** International Technical Support Organization – IBM; 1998.
Disponível em:
<http://publib-b.boulder.ibm.com/Redbooks.nsf/RedbookAbstracts/sg242238.html?Open>
Acesso em 21.11.2002

CHAUDHURI, S.; DAYAL, U. **An Overview of Data Warehousing and OLAP Technology.** 1997.
Disponível em: http://www.acm.org/sigmod/record/issues/9703/Chaudhuri_e_Dayal.ps
Acesso em 21.11.2002

COREY, M. *et al.* **Oracle 8i Data Warehouse.** Oracle Press; 2001

DOWLING, K.; SCHUFF, D.; LOUIS, R. **Star Join Schemas versus Normalized Relational Schemas: Does it Really Make a Difference to End-Users?** Arizona State University, Temple University; 2001.
Disponível em :
http://mis.temple.edu/schuff/research/documents/StarJoinER_AMCIS2001.pdf
Acesso em 21.11.2002

DuMOULIN, R. **Architecting Data Warehouse for Flexibility, Maintainability, and Performance.** Encore Development; 2001.
Disponível em: http://www.dbatoolbox.com/wp2001/dssdw/arch_dw.pdf
Acesso em 23.04.2003

EDER J.; KONCILIA, C. **Representing Knowledge about Changes in Data Warehouse Structure.** University of Klagenfurt – Dep. Of Informatics-Systems; 2002.

GUPTA, V. R. **An Introduction to Data Warehousing.** 1997.
Disponível em: <http://www.bac-atl.com/library/bi/dwintro.pdf>
Acesso em 21.11.2002

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques.** Simon Fraser University - Morgan Kaufmann Publishers; 2001.

HUMPHRIES, M.; HAWKINS, M. W.; DY, M. C. **Datawarehousing Architecture and Implementation.** Prentice Hall; 1999.

HUSEMANN, B.; LECHTENBORGER, J.; VOSSEN, G. **Conceptual Data Warehouse Design.** 2000.

Disponível em:

<http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-28/paper6.pdf>

Acesso em 27.11.2002

HYATT, L. E.; ROSENBERG, L. H. **A Software Quality Model and Metrics for Identifying Project Risks and Assessing Software Quality**. 1996.

Disponível em: http://satc.gsfc.nasa.gov/support/STC_APR96/qualitiy/stc_qual.PDF

Acesso em 21.11.2002

INMON, W. H. **Building the Data Warehouse**. Wiley Computer Publishing. 2nd Edition, 1996.

_____. **The problem with Dimensional Modeling**. DM Review; Maio 2000.

Disponível em: <http://www.billinmon.com/library/articles/artdimmd.asp>

Acesso em 21.11.2002

_____. **What is Data Warehouse**. 2000a.

Disponível em: <http://www.inmoncif.com/library/whiteprs/earlywp/ttdw.pdf>

Acesso em 22.04.2003

_____. **The Issues of the eBusiness Infrastructure**. 2002.

Disponível em: <http://www.billinmon.com/library/articles/artissue.asp>

Acesso em 07.11.2002

JARKE, M. *et al.* **Architecture and Quality in Data Warehouses: an Extended Repository Approach**. 1999.

Disponível em:

<http://citeseer.nj.nec.com/cache/papers/cs/11574/http:zSzzSzwww-i5.informatik.rwth-aachen.dezSz~quixzSzpaperszSzcaise98-infsys.pdf/jarke99architecture.pdf>

Acesso em 21.11.2002

KELLY, T. J. **Dimensional Data Modeling**. Sybase Professional Services; 1998.

Disponível em:

http://www.gca.net/solutions/whitepapers/sybase/syb_dim_data_mod.html

Acesso em 21.11.2002

KIMBALL, R. **Slowly Changing Dimensions**. DBMS; april 1996.

Disponível em: <http://www.dbmsmag.com/9604d05.html>

Acesso em 15.04.2002

_____. **A Dimensional Modeling Manifesto**. DBMS; august 1997.

Disponível em: <http://www.dbmsmag.com/9708d15.html>

Acesso em 01.05.2002

_____. **Help for Dimensional Modeling**. DBMS; august 1998.

Disponível em: <http://www.dbmsmag.com/9808d05.html>

Acesso em 31.05.2002

KRIPPENDORF, M.; SONG, II. **The Translation of Star Schema into Entity-Relationship Diagrams**. 1997.

Disponível em: <http://citeseer.nj.nec.com/jarke99architecture.html>

Acesso em 21.11.2002

MAIER, R. **Evaluation of Data Modeling**. 1999.

Disponível em:

[http://www-wi.uni-regensburg.de/\(de\)/forschung/engl_pub/adbis99_maier.pdf](http://www-wi.uni-regensburg.de/(de)/forschung/engl_pub/adbis99_maier.pdf)

Acesso em 27.11.2002

MOODY, D. L.; KORTINK, M. A. R. **From enterprise models to dimensional models: a methodology for data warehouse and data mart design**. Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000); 2000.

Disponível em:

<http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-28/paper5.pdf>

Acesso em 27.02.2002

PEDERSEN, T. B.; JENSEN, C. S. **Multidimensional Data Modeling for Complex Data**. 1998.

Disponível em: http://citeseer.nj.nec.com/bachPedersen_e

[Jensen98multidimensional.html](http://citeseer.nj.nec.com/bachPedersen_eJensen98multidimensional.html)

Acesso em 26.11.2002

_____. **Research Issues in Clinical Data Warehousing**. 1999a.

Disponível em: http://citeseer.nj.nec.com/Pedersen_eJensen98research.html

Acesso em 26.11.2002

RODDICK, J. *et al.* **Evolution and Change in Data Management – Issues and Directions**. School of Informatics and Engineering, Flinders University of South Australia; 2000.

SCHOUTEN, H. **Analysis and design of data warehouses**. 1999.

Disponível em:

<http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-19/>

Acesso em 26.11.2002

SIGN, H. S. **Data Warehouse – Conceitos Tecnologias, Implementação e Gerenciamento**. Makron Books; 2001.

SONG, Y. II.; SHULTZ, K. L. **Data Warehouse Design for E-commerce Environment**. 1999.

Disponível em: <http://citeseer.nj.nec.com/song99data.html>

Acesso em 26.11.2002

SONG, Y. II. *et al.* **Analysis of Many to Many Relationships Between Fact and Dimension Tables in Dimensional Modeling**. Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2001); 2001.

Disponível em:

http://www.cis.drexel.edu/faculty/song/Papers/Song_M_N_DMDW_final.pdf

Acesso em 27.11.2002

SYBASE PowerDesigner Warehouse Architect - The Model for DataWarehousing Solutions. A Technical Whitepaper from Sybase, Inc; 1998.

Disponível em:

<http://www.sybase.com/content/1000844/PowerDesignerWarehouseArchitect.pdf>

Acesso em 27.11.2002

TIERSTEIN, L. Data Mart Database Design for OLTP Data Modelers. Constellar Corp; 2001.

Disponível em: http://www.dbatoolbox.com/wp2001/dssdw/datamart_design.pdf

Acesso em 23.04.2003

VASSILIADIS, P Data Warehouse Modeling and Quality Issues. National Technical University of Athens; 2000.

Disponível em: <http://citeseer.nj.nec.com/vassiliadis00data.html>

Acesso em 27.11.2002

WINSBERG, P. Modeling the Datawarehouse and Data Mart. InfoDB Volume 10 No. 3; 1996

Disponível em: <http://www.dbaint.com/pdf/v10n31.pdf>

Acesso em 27.11.2002

APÊNDICE 1 – SIMBOLOGIA DOS ESQUEMAS DIMENSIONAIS

A simbologia utilizada nos esquemas dimensionais neste trabalho é a seguinte:

Tabela

Chave 1
Chave 2
Atributo 1
Atributo 2

As tabelas são divididas em duas partes. Na divisão superior da tabela é informada a PK (*Primary Key*), que no exemplo são os atributos Chave 1 e Chave 2 e, na inferior os demais atributos.

Tabela Dimensão

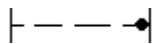
Chave 1

Tabela Dimensão (cor branca)

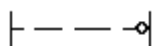
Tabela Fato

Chave 1

Tabela Fato (cor cinza)



Relacionamento 1:N partindo da esquerda para a direita em que todos os elementos da tabela origem se relacionam a um ou mais elementos da tabela destino.



Relacionamento 1:N partindo da esquerda para a direita em que nem todos os elementos da tabela origem se relacionam a um ou mais elementos da tabela destino.

(FK)

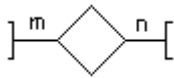
Foreign Key

APÊNDICE 2 – SIMBOLOGIA DOS ESQUEMAS ER

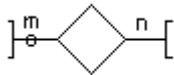
A simbologia utilizada nos esquemas ER neste trabalho é a seguinte:

ENTIDADE

Entidade



Relacionamentos entre duas entidades. As letras m e n indicam a cardinalidade. Existem também relacionamentos 1:N; M:1 e 1:1.



O círculo no relacionamento indica a opcionalidade do relacionamento
