

PEDRO HENRIQUE TOLENTINO RODRIGUES

**MODELO DE IMPLEMENTAÇÃO DE
BUSINESS INTELLIGENCE NA ÁREA DE SAÚDE**

Monografia apresentada à
Escola Politécnica da
Universidade de São Paulo
para obtenção da certificação
no curso de MBA em
Tecnologia da Informação

Orientador:

Prof. Alexandre Ricardo Nardi

São Paulo

2007

MBA ITI

2007

R618m

DEDALUS - Acervo - EPEL



31500015854

FICHA CATALOGRÁFICA

M 2007 B

Rodrigues, Pedro Henrique Tolentino

**Modelo de implementação de Business Intelligence na área de saúde / P.H. T. Rodrigues. São Paulo, 2007.
57 p.**

Monografia (MBA em Tecnologia e Gestão na Produção de Edifícios) – Escola Politécnica da Universidade de São Paulo. Programa de Educação Continuada em Engenharia.

1.Sistemas de informação 2.Banco de dados I.Universidade de São Paulo. Escola Politécnica. Programa de Educação Continuada em Engenharia II.t.

DEDICATÓRIA

À minha esposa **Tatielly** pelo companheirismo e amor incondicional e aos meus pais **Pedro** e **Sandra** pela compreensão, paciência, incentivo e patrocínio durante toda esta jornada.

AGRADECIMENTOS

Inicialmente agradeço aos meus pais **Sandra e Pedro** pela educação, pelos ensinamentos e pela oportunidade que recebi de poder cursar o MBA e o incentivo de todos os familiares, em especial minha esposa **Tatielly**.

Agradeço ao Prof. **Alexandre Ricardo Nardi**, orientador deste trabalho, que esteve presente durante toda esta jornada final de Monografia. Sua amizade, suas orientações objetivas e o empenho de todo o seu conhecimento e experiência foram fundamentais para o desenvolvimento do trabalho.

Agradeço também a todos que participaram indiretamente deste processo:

- **Tadashi Kadomoto** pelos seus ensinamentos;
- **Yoshio Kadomoto** pelos seus brilhantes ensinamentos e aconselhamentos;
- **Líria Ohara** pelo acompanhamento, ensinamentos, paciência e dedicação nestes meses que passamos juntos e que valeram por uma vida inteira.
- Ao escritor **Deepak Chopra** que com sua obra contribuiu de forma muito positiva nos momentos difíceis;
- Aos amigos e familiares pela paciência e palavras de incentivo;

Agradeço a **Deus** por todas as dificuldades que nos trazem aprendizado e principalmente as oportunidades que tenho recebido.

Muito obrigado a todos.

RESUMO

Este estudo propõe a implementação de um Modelo de Business Intelligence que a partir de transações eletrônicas forneça informações relevantes aos participantes da área de saúde.

Nele serão analisadas opções de modelagem e processos, a partir de um ambiente transacional (OLTP) chegando até a construção de informações analíticas (OLAP). Nesta análise serão incluídas as vantagens e desvantagens dos modelos passíveis de serem utilizados, seguidos do detalhamento do Modelo escolhido para a aplicação deste trabalho.

A implementação deste Modelo de Business Intelligence é capaz de potencializar os benefícios do uso dos dados provenientes do ambiente transacional já existente, ampliando consideravelmente os novos modelos de consultas e conseqüentemente as informações geradas, ajudando as empresas da área de saúde tornarem-se mais competitivas.

Palavras-chave: Business Intelligence; Data Warehouse; Modelagem Multidimensional.

ABSTRACT

The purpose of this study considers the implementation of a Business Intelligence Model that supplies relevant information to the participants of the health area through electronic transactions.

Some options of modeling and processes will be analyzed from a transactional environment (OLTP) up to the construction of some analytical information (OLAP). This analysis includes the advantages and disadvantages of the models that can be used, followed by the detailing of the Model chosen for the application of this work.

The implementation of this Business Intelligence Model is able to increase the benefits of using the data proceeding from a transactional environment that already exists, extending considerably the new models of queries and consequently the generated information, helping the companies of the health area to become more competitive.

Keywords: Business Intelligence; Data Warehouse; Multidimensional Data Modeling.

SUMÁRIO

1. INTRODUÇÃO.....	1
1.1. OBJETIVO.....	1
1.2. MOTIVAÇÃO.....	1
2. CONCEITOS	3
2.1. AMBIENTE TRANSACIONAL.....	3
2.2. BUSINESS INTELLIGENCE E AMBIENTE ANALÍTICO.....	4
2.2.1. <i>Data Warehouse x Data Mart x Staging Area</i>	5
2.2.2. <i>Modelagem Multidimensional</i>	9
2.2.3. <i>Processo de Construção</i>	20
3. A ÁREA DE SAÚDE	22
4. SISTEMAS TRANSACIONAIS NA ÁREA DE SAÚDE	25
5. MODELO DE BUSINESS INTELLIGENCE NA ÁREA DE SAÚDE.....	27
5.1. AMBIENTE ANALÍTICO.....	28
5.2. STAGING AREA - ODS	35
5.3. ETL – EXTRAÇÃO TRANSFORMAÇÃO E CARGA	37
5.4. DATA MART / DATA WAREHOUSE	44
5.5. MODELOS DE REPRESENTAÇÃO FÍSICA.....	47
6. CONCLUSÃO.....	54
7. TRABALHOS FUTUROS	55
REFERÊNCIAS.....	56

LISTA DE FIGURAS

FIGURA 1 - REPRESENTAÇÃO DO AMBIENTE TRANSACIONAL	4
FIGURA 2 - ELEMENTOS BÁSICOS DE UM DATA WAREHOUSE.....	5
FIGURA 3 - EXEMPLOS DE NÍVEIS DE GRANULARIDADE.....	13
FIGURA 4 - CUBO DE DADOS COM 3 DIMENSÕES	15
FIGURA 5 - REPRESENTAÇÃO ESQUEMA ESTRELA	16
FIGURA 6 - REPRESENTAÇÃO ESQUEMA FLOCO DE NEVE	17
FIGURA 7 - ARQUITETURA DO MODELO PROPOSTO	20
FIGURA 8 - COMPOSIÇÃO DO MODELO INICIAL	26
FIGURA 9 - IMPLEMENTAÇÃO TOP_DOWN	30
FIGURA 10 - IMPLEMENTAÇÃO BOTTOM-UP	31
FIGURA 11 - COMPOSIÇÃO DO MODELO ANALÍTICO PROPOSTO	34
FIGURA 12 - MODELO DE STAGING TRANSAÇÕES	36
FIGURA 13 - MODELO DE STAGING FINANCEIRO	37
FIGURA 14- PROCESSO DE ETL DE UM DATA WAREHOUSE	39
FIGURA 15 - MODELO DO DATA MART TRANSAÇÕES	45
FIGURA 16 - MODELO DO DATA MART FINANCEIRO	46
FIGURA 17 - MODELO DO DATA WAREHOUSE PARCIAL PARA ÁREA DE SAÚDE	47

LISTA DE TABELAS

TABELA 1 - DIFERENÇAS ENTRE MODELO MULTIDIMENSIONAL E ENTIDADE-RELACIONAMENTO	7
TABELA 2 - COMPARATIVO ENTRE OLTP E OLAP	28
TABELA 3 - COMPARATIVO DA IMPLEMENTAÇÃO TOP-DOWN	30
TABELA 4 - COMPARATIVO DA IMPLEMENTAÇÃO BOTTOM-UP.....	31
TABELA 5 - MAPA DE COMPARAÇÃO DAS CARACTERÍSTICAS DA ARQUITETURA OLAP	51

LISTA DE ABREVIATURAS

ETL	<i>Extraction Transform Load</i>
ODS	<i>Operational Data Store</i>
OLAP	<i>On-line Analytical Process</i>
OLTP	<i>On-line Transaction Processing</i>
POS	<i>Point of Sales</i>
URA	Unidade de Resposta Audível
SGBD	Sistema Gerenciador de Banco de Dados
DW	<i>Data Warehouse</i>
DM	<i>Data Mart</i>

1. INTRODUÇÃO

A elaboração deste trabalho vem contribuir no atendimento as demandas de se produzir informações analíticas a partir de um ambiente transacional (OLTP) na área de saúde, com a criação de um Modelo de Implementação de Business Intelligence para prover apoio às tomadas de decisões, auxiliar na definição das estratégias futuras e promover avaliação de resultados históricos baseados em transações eletrônicas.

1.1. Objetivo

Este estudo irá propor a implementação de um Modelo que forneça informações relevantes aos participantes da área de saúde, que utilizam de autorização eletrônica das transações provenientes de execução de procedimentos médicos.

Serão analisadas opções de modelagem e processos, a partir de um ambiente transacional (OLTP) chegando até a construção de informações analíticas (OLAP). Nesta análise serão incluídas as vantagens e desvantagens dos modelos passíveis de serem utilizados, seguidos do detalhamento do Modelo escolhido para a aplicação deste trabalho.

Este trabalho está delimitado em apresentar parte do modelo final de Data Warehouse para área de saúde, de forma que será aprofundado somente no que tange o tratamento dos dados pertinentes a transações.

1.2. Motivação

Em um ambiente globalizado, cada vez mais as empresas necessitam de informações para reagir aos problemas e oportunidades de negócio que surgem no horizonte, independentemente da área de negócio ou do porte das mesmas.

Ainda hoje a integração entre operadoras de saúde e prestadores ainda é muito incipiente e trata-se de uma área com grande potencial. De

acordo com estimativas, são transacionadas mensalmente 30 milhões de contas médicas, sendo que somente 13% delas trafegam eletronicamente[1].

A perspectiva das grandes Empresas de Medicina Suplementar é realizar 80% das transações eletronicamente nos próximos cinco anos. “As empresas de planos de saúde sentem cada vez mais a pressão por redução de custos, o que as leva a buscar recursos para aumentar o controle”, destaca Leite[1].

Assim para ajudar empresas participantes da área de saúde na transformação dos seus dados de transações eletrônicas em informações, este trabalho irá propor um Modelo de Implementação de Business Intelligence que possa gerar diferencial de negócio.

A implementação deste Modelo de Business Intelligence é capaz de potencializar os benefícios do uso das informações provenientes do ambiente transacional já existente.

2. CONCEITOS

Nos próximos tópicos deste capítulo serão abordados alguns conceitos básicos acerca de Ambiente Transacional e Business Intelligence como parte do contexto de um Ambiente Analítico que irão auxiliar na compreensão do desenvolvimento do trabalho.

2.1. Ambiente Transacional

Geralmente, os sistemas do ambiente operacional que apóiam os usuários em suas funções do dia-a-dia são chamados OLTP (On-Line Transaction Processing), e seu principal objetivo é executar o maior número de transações possíveis no menor tempo de processamento. Em geral, os sistemas OLTP são pouco flexíveis em relação à quantidade de relatórios e consultas, devido às limitações impostas por seu modelo de dados e à linguagem SQL.

Na figura abaixo podemos observar um exemplo da composição de um Ambiente OLTP, o qual tomaremos como referência e onde ocorrem a captura e o tratamento de transações eletrônicas na área de saúde.

Como podemos observar, a figura é composta por três modalidades de meio de captura onde as transações são geradas. Estas transações são concentradas e roteadas até o servidor principal, onde neste ocorre o controle da fila de transação e em seguida após o tratamento e resposta os dados são armazenados em banco de dados OLTP.

Existe ainda um processo de replicação que copia as transações no segundo servidor para garantir assim a integridade e disponibilidade de acesso aos dados.

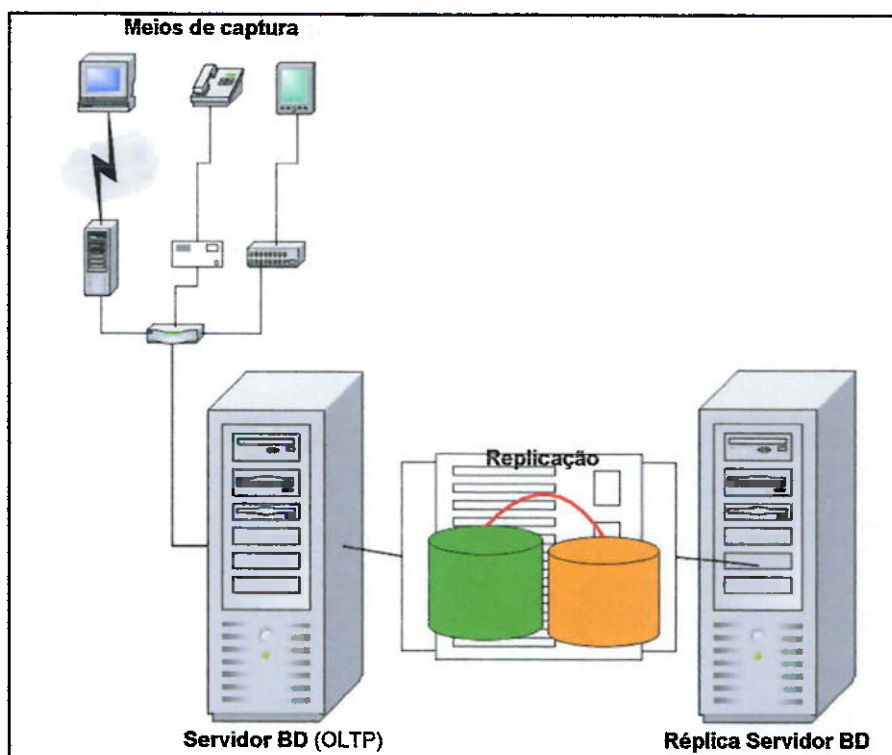


Figura 1 - Representação do Ambiente Transacional
Fonte: Elaborado pelo Autor

Conforme definimos acima, este ambiente OLTP representado na *Figura 1* foi projetado para atender e armazenar o maior número de transações possíveis.

Ampliaremos os detalhes sobre o ambiente proposto como modelo no *item 4*.

2.2. Business Intelligence e Ambiente Analítico

Segundo Machado[2], Business Intelligence é um conjunto de tecnologias orientadas a disponibilizar informação e conhecimento em uma empresa.

Essas tecnologias fazem parte do contexto de um Ambiente Analítico, também definido como OLAP (*On-line Analytical Process*).

Inmon[3] nos dá uma definição bastante completa sobre o OLAP:

(...) é uma tecnologia de software que permite a analistas, gerentes e executivos obterem os dados de uma forma rápida, consistente e com

acesso interativo para uma grande variedade de possíveis visões da informação na empresa. Mais sucintamente, OLAP é um conjunto de funcionalidades que tem, como principal objetivo, facilitar a análise multidimensional. Análise Multidimensional por sua vez significa a análise da combinação de dados em suas várias dimensões segundo Machado[2].

Sistemas OLAP fornecem uma visão multidimensional dos dados não importando como estes dados estão fisicamente armazenados. Os dados são percebidos pelo usuário como um cubo multidimensional, que será definido no item 2.2.2, onde cada célula contém um valor ou medida. A seguir temos uma representação apresentada por Kimball[4] para dar visão dos elementos básicos que compõem um Data Warehouse.

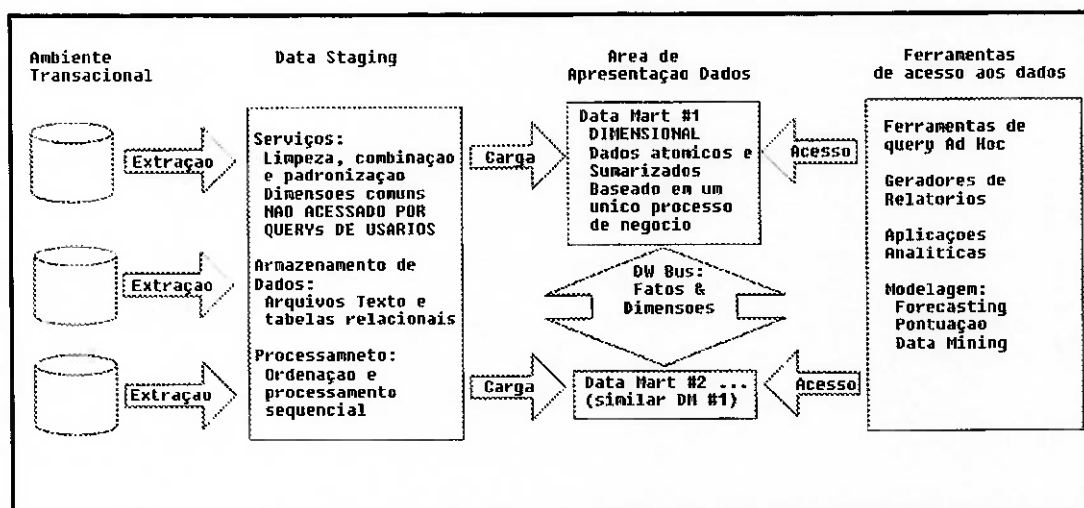


Figura 2 - Elementos básicos de um Data Warehouse
Fonte: Ralph Kimball[4]

A seguir abordaremos os principais componentes do ambiente OLAP.

2.2.1. Data Warehouse x Data Mart x Staging Area

O **Data Warehouse** proporciona a integração dos dados na Corporação. Dessa forma não é preciso confrontar várias aplicações em busca de uma inteligência nos negócios, basta usar ferramentas analíticas apropriadas.

A definição clássica criada por Inmon[3] é a seguinte:

Data Warehouse é uma coleção de dados orientada por assuntos, integrada, variante no tempo e não volátil. Que tem por objetivo dar suporte aos processos de tomada de decisão.

Data Warehouse é um processo em andamento que aglutina dados de fontes heterogêneas, incluindo dados históricos e dados externos para atender à necessidade de consultas estruturadas e ad-hoc, relatórios analíticos e de suporte de decisão, conforme Harjinder[5];

Kimball[4] define assim: *Data Warehouse um conjunto de ferramentas e técnicas de projeto, que quando aplicadas às necessidades específicas dos usuários e aos bancos de dados específicos permitirá que planejem e o construam.*

Desta forma o repositório das informações constitui o núcleo do ambiente do DW, onde estarão contidos todos os dados extraídos dos sistemas transacionais, formando assim a base necessária para o processo de tomada de decisão. A primeira fase no projeto do repositório é a definição do modelo de dados. A modelagem de dados do DW consiste da forma pela qual as informações do DW são representadas e armazenadas[6].

A tradicional abordagem entidade-relacionamento (E/R), utilizada com sucesso no projeto de sistemas para o ambiente operacional, não é adequada para a construção de um DW. No modelo E/R, a estrutura é otimizada para recuperar, criar e atualizar registros individuais em tempo real e preservar a integridade dos dados. No DW, a utilização dos dados é diferente. As consultas realizadas recuperam um grande número de registros e os resumem durante o processamento. Neste caso, a estrutura normalmente não é normalizada para evitar junções de muitas tabelas, obtendo-se assim uma performance superior[4].

Na Tabela 1 podemos observar as diferenças entre as técnicas de modelagem Entidade/Relacionamento e modelagem Multidimensional, que será conceituada no *item 2.2.2*.

Tabela 1 - Diferenças entre Modelo Multidimensional e Entidade-Relacionamento

Fonte: Adaptado de[7].

Uso	Modelo Multidimensional	Modelo Entidade/Relacionamento
Processamento de Transações	Não	Sim
Alvo de limpeza de dados	Não	Sim
Pré-consolidação de dados finais	Sim	Não
Consulta <i>Ad Hoc</i>	Sim	Não
Relatórios	Sim	Não
Exploração de dados	Sim	Não
Previsão	Sim	Não
Investigar através de dados distribuídos	Sim	Não
Armazenamento ciente agregado de dados	Sim	Não
Estrutura previsível, padrão.	Sim	Não
Arquitetura e administração distribuída	Sim	Não
Ferramenta de suporte explícita para usuário final	Sim	Não
Compatibilidade com OLAP	Sim	Não
Dimensões lentamente em alteração	Sim	Não

Como um DW é um banco de dados orientado somente para consulta de seus dados, a orientação da técnica criou os denominados modelos estrela ou esquema estrela. Esta formação é de fundamental importância para que um projeto de DW resulte em um armazém de dados organizado e acessível, com as informações necessárias à gestão de negócios, e não ao controle do negócio segundo Machado[2].

A modelagem Multidimensional busca modelar dados para aprimorar o entendimento e o desempenho. Tanto os modelos multidimensionais como os E/R têm seu lugar adequado no DW segundo Kimball e Merz[7]. A *Tabela 1* resume as diferenças entre as duas técnicas.

O **Data Mart** por definição de Machado[2], representa um subconjunto de dados do Data Warehouse. Estes subconjuntos são direcionados a um departamento ou uma área específica de processos do negócio. Uma das principais vantagens da sua utilização é a possibilidade de retorno rápido, garantindo um maior envolvimento do usuário final, capaz de avaliar os benefícios extraídos de seu investimento.

Segundo Inmon[3], um Data Mart pode ser definido como um SGBD multidimensional que fornece uma estrutura bastante flexível de acesso a dados. Enquanto o Data Warehouse extrai, transforma e limpa os dados dos sistemas transacionais, mantendo-os integrados em quantidades massivas e em seu nível mais baixo, o Data Mart se serve destes dados, extraindo informações para um departamento ou uma área de negócio, oferecendo flexibilidade e controle ao usuário final, pois com o Data Mart é possível segmentar e agrupar dados de diversas maneiras.

Já Kimball[8], define que Data Mart é um subconjunto do Data Warehouse, direcionado a um departamento ou área de negócio. Para o autor, o conjunto de todos os Data Marts da organização, construídos de forma incremental, compartilhando dimensões e fatos comuns, segundo um planejamento prévio, formam o Data Warehouse lógico da organização. A definição de Data Mart apresentada por Kimball[8], será utilizada no desenvolvimento deste trabalho.

Temos também o componente **Staging Area**, também conhecido como ODS (Operational Data Store).

O Staging é uma base de dados de construção de DW onde os dados dos sistemas fontes são preparados e integrados, e onde para alguns autores se aplica também a consultas cotidianas no nível operacional voltado para a comunidade de escritório que toma decisões do dia-a-dia em questões detalhadas e imediatas operando em um nível mais alto do que as aplicações operacionais[9].

Kimball[8] se opõe a esta abordagem e considera que o Staging é válido e necessário quando usado como ponto de integração dos sistemas transacionais. Esta definição de Staging Area apresentada por Kimball[8], será utilizada no desenvolvimento deste trabalho.

2.2.2. Modelagem Multidimensional

A modelagem Multidimensional é a técnica de projeto mais freqüentemente utilizada para a construção de um ambiente Data Warehouse. O objetivo é buscar um padrão de apresentação de dados que seja facilmente visualizado pelo usuário final e que possua um bom desempenho para consultas.

Segundo Machado[2] a Modelagem Multidimensional é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios. É utilizada especialmente para sumarizar e reestruturar dados e apresentá-los em visões que suportem a análise dos valores desses dados.

Kimball[4] diz que a modelagem Multidimensional é uma metodologia que permite modelar logicamente dados para melhorar o desempenho de consultas e prover facilidade de utilização a partir de um conjunto de eventos básicos de medição. Os modelos multidimensionais são compreensíveis, previsíveis, ampliáveis e resistentes por se manterem fiéis à simplicidade. Possuem ainda uma perspectiva voltada para as necessidades analíticas da empresa, especialmente seu formato simétrico, em que todas as dimensões normalmente são iguais pontos de entrada na tabela de fatos.

Os modelos Multidimensionais são a base de muitos aprimoramentos de desempenho de SGBDs, inclusive agregações e métodos de indexação avançados.

O modelo multidimensional é formado por uma tabela central (Tabela Fatos) e várias outras tabelas a ela interligadas (Tabela Dimensão), sempre por meio de chaves (Chaves Especiais), que associam o fato a uma dimensão do cubo.

A seguir veremos os principais conceitos relacionados aos Modelos Multidimensionais e sua utilização:

Dimensões e Tabela Dimensão

Uma tabela de dimensão contém as descrições textuais do negócio, e possui as informações necessárias para análises ao longo de dimensões. Seus atributos são fonte das restrições das consultas, agrupamento dos resultados, e cabeçalhos para relatórios. As dimensões são os aspectos pelos quais se pretende observar as métricas relativas ao processo que está sendo modelado.

Segundo Kimball[4] a qualidade do banco de dados é proporcional à qualidade dos atributos de dimensões, portanto devem ser dedicados tempo e atenção à sua descrição, ao seu preenchimento e à garantia da qualidade dos valores em uma coluna de atributos.

Cada dimensão é definida com uma chave primária. Essa chave é a base da integridade referencial no relacionamento com a tabela de fatos. Uma tabela de dimensão é composta também de atributos. Os melhores atributos são textuais e distintos, e devem consistir de palavras reais, evitando-se o uso de códigos e abreviações. Esses atributos descrevem as linhas na tabela de dimensão.

Kimball[4] ainda ressalta que é muito importante que os atributos das tabelas de dimensão sejam preenchidos com valores descritivos ao invés de códigos sem sentido, criptografados ou abreviados.

Cada dimensão possui sua própria granularidade, sendo que ela não pode ser menor que a granularidade da tabela de fatos a ela associada. As tabelas de dimensão podem ser classificadas de diversas maneiras, como as dimensões *Dirty*, *Slowly Changing*, *Tempo* e *Compartilhada*, entre outras menos importantes.

As *dimensões Dirty* são aquelas inconsistentes, ou seja, que representam uma falha de integridade na base de dados operacional. A consequência deste problema é a possibilidade de existirem registros duplicados, cabendo ao processo de limpeza do DW esta verificação.

As *dimensões Slowly Changing* (dimensões de modificações lentas) são aquelas que apresentam tendências a sofrer modificações graduais ao longo do tempo. Elas podem ser manipuladas por três técnicas distintas.

- A primeira técnica substitui valores antigos do registro da tabela dimensão por novos valores, o que a faz perder a capacidade de rastrear valores históricos.
- A segunda técnica consiste em adicionar um registro à dimensão. Este registro deve conter os novos valores do atributo, segmentando o histórico entre a descrição antiga e a nova descrição com precisão.
- Já a terceira técnica cria novos campos no registro original da dimensão para incluir os novos valores do atributo, mantendo seus valores originais, permitindo assim descrever o histórico anterior e posterior à mudança, tanto em relação aos valores originais do atributo quanto aos valores atuais.

As *dimensões Tempo* são as mais comuns em DW's, tendo em vista que uma de suas principais características é ser variante ao tempo. Ela associa-se a praticamente todas as tabelas de fatos[4].

As *dimensões Compartilhadas* são aquelas comuns a dois ou mais esquemas multidimensionais, cujos valores são consistentes entre estes esquemas, tendo o mesmo significado

para todas as possíveis tabelas de fatos às quais elas são conectadas. Um conjunto de tabelas de fatos que utilizam dimensões compartilhadas é conhecido como uma constelação de fatos.

— Fatos e Tabela Fatos

A tabela de fatos é a principal tabela de um modelo multidimensional, onde as medições numéricas de interesse da empresa estão armazenadas segundo Kimball[4]. A palavra "fato" representa uma unidade dos processos que estamos modelando, como quantidades, valores e indicadores num determinado momento.

A tabela de fatos registra os fatos que serão analisados. É composta por uma chave primária (formada por uma combinação única de valores de chaves de dimensão) e pelas métricas de interesse para o negócio.

As dimensões indicam a forma como as medidas serão vistas, ou seja, são os aspectos pelos quais se pretende observar as métricas. A intersecção das chaves de dimensão define a granularidade da tabela de fatos, e é importante que todas as medidas desta tabela tenham a mesma granularidade.

Segundo Inmon[3] *a granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no Data Warehouse. Quanto mais detalhe, mais baixo o nível de granularidade. Quanto menos detalhe, mais alto o nível de granularidade, conforme podemos observar na Figura 3.*

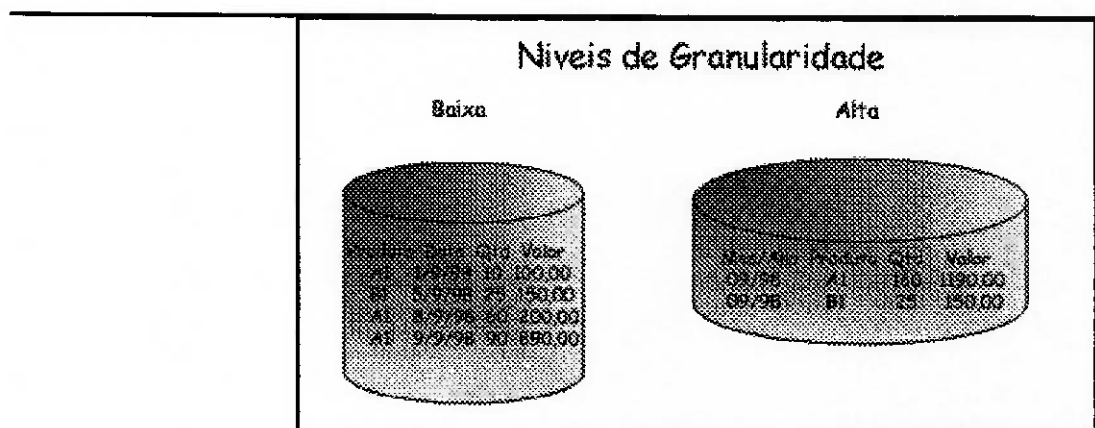


Figura 3 - Exemplos de níveis de granularidade
 Fonte: DWBrasil[10]

A granularidade da tabela de fatos é um fator que influencia na performance do DW. Ela diz respeito a um total calculado a partir de um determinado fator, seja ele temporal ou qualquer outra forma de dimensão mensurável, o que vai influenciar na quantidade de informações armazenadas na tabela de fatos. Assim sendo, quanto maior o nível de detalhes, mais baixo o nível de granularidade e maior o espaço em disco requerido para armazenar a tabela.

— Chaves Especiais

As chaves, que serão responsáveis pelas junções das tabelas lógicas, podem ser definidas como: composta, estrangeira, generalizada, primária e substituta (surrogate key).

Segundo Kimball[8], a definição para cada uma dessas chaves é a seguinte:

- ✓ Chave composta: Uma chave em uma tabela do banco de dados, composta de vários campos. Por exemplo, a chave geral em uma típica tabela de fatos, é um subconjunto de chaves estrangeiras na tabela de fatos.

-
- ✓ Chave estrangeira: Um campo em uma tabela de banco de dados relacional cujos valores são aproveitados a partir de valores de uma chave primária em outra tabela.
 - ✓ Chave generalizada: Uma chave primária da tabela de dimensões que foi criada ao generalizar uma chave original de produção como um número de produto ou um número de cliente, normalmente ao criar uma chave substituta.
 - ✓ Chave primária: Um campo, em uma tabela de banco de dados, que é único e diferente para cada registro na tabela.
 - ✓ Chave substituta: Uma chave artificial, normalmente um inteiro consecutivamente atribuído, que é utilizado em um modelo multidimensional para concatenar uma tabela de dimensões a uma tabela de fatos. Na tabela de dimensões, a chave substituta é a chave primária. Na tabela de fatos, a chave substituta é uma chave estrangeira e pode ser parte da chave primária da tabela de fatos.

O ponto a favor de uma chave substituta é que não há absolutamente nenhuma semântica no valor da chave. É somente um veículo para unir a tabela de dimensão a uma tabela de fatos[7]. Se um objeto sendo referido é desconhecido, não medido, inaplicável, corrompido ou ainda não acontecido, uma chave semântica significativa de produção não pode ser aplicada. Contudo, o DW tem que fornecer uma chave. As chaves substitutas isolam o DW de alterações na administração de chaves nos dados de produção, além do que podem tratar com facilidade dimensões que se alteram lentamente.

_ Cubos

A representação dos dados de um Data Warehouse é estruturada como em “*cubo de dados*”, ilustrado pela *Figura 4*.

Esta estrutura é chamada modelo multidimensional e implementa-se a partir de estruturas conhecidas tais como esquema estrela (Star Schema) e esquema floco de neve (Snowflake Schema), que veremos a seguir.

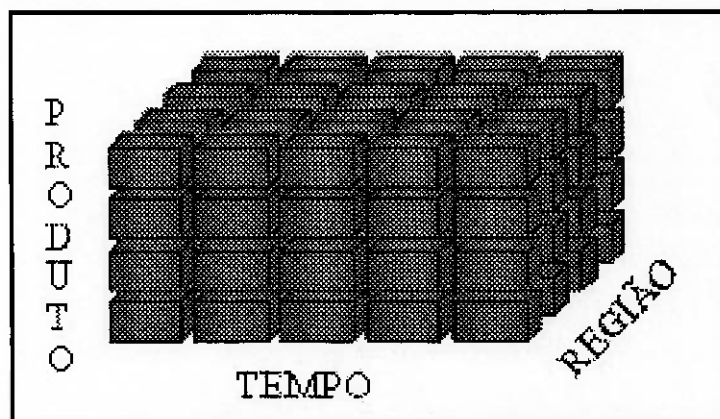


Figura 4 - Cubo de Dados com 3 dimensões
Fonte: Elaborado pelo autor

A utilização destas estruturas definidas acima são apresentadas nos modelos ou esquemas multidimensionais.

— Esquema Estrela (Star Schema)

O Esquema Estrela é um dos esquemas multidimensionais mais comuns. Ele consiste numa tabela de fatos e uma tabela para cada dimensão. A *Figura 5* ilustra um esquema estrela com uma tabela de fatos no centro e dimensões de produto, região e tempo ligados a ela através de suas respectivas chaves[4].

A simplicidade do esquema estrela faz com que ele seja amplamente utilizado no mercado. Vale ressaltar que um Data Warehouse busca simplicidade na sua implementação, visando facilitar o entendimento dos dados por parte dos usuários.

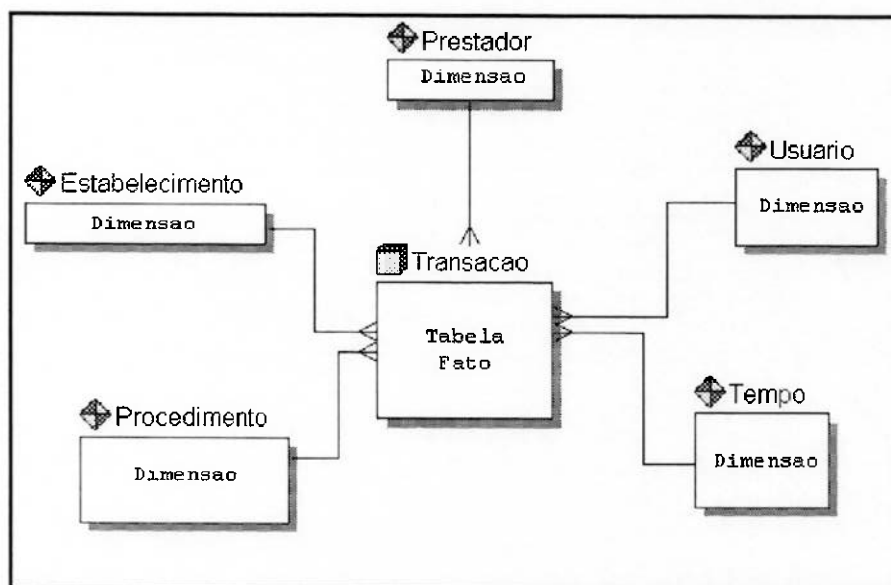


Figura 5 - Representação Esquema Estrela
 Fonte: Adaptado de Kimball[4]

Esquema Floco de Neve (Snowflake Schema)

Uma variação do esquema estrela é conhecida como esquema floco de neve. Ele é criado a partir da normalização de uma ou mais tabelas de dimensão. Atributos dependentes funcionalmente da tabela de dimensão são transportados para uma tabela de dimensão secundária, cuja chave torna-se estrangeira na tabela de dimensão primária[4].

Esta técnica diminui a capacidade de entendimento dos dados por parte do usuário final, uma vez que torna o esquema mais complexo. O esquema floco de neve sofre uma queda de performance em relação ao esquema estrela, mas vale lembrar que em relação aos DW montados sobre bases transacionais, o esquema floco de neve é muito bom.

Porém, em alguns casos, sua utilização pode ser interessante ou mesmo necessária. Um dos casos é a utilização de mini-dimensões. Um atributo de alta cardinalidade em uma tabela de dimensão extremamente grande pode ser

subdividido em mini-dimensões, compostas por um pequeno conjunto de atributos, para conter um numero limitado de valores[11].

A representação deste modelo pode ser observada abaixo na *Figura 6*.

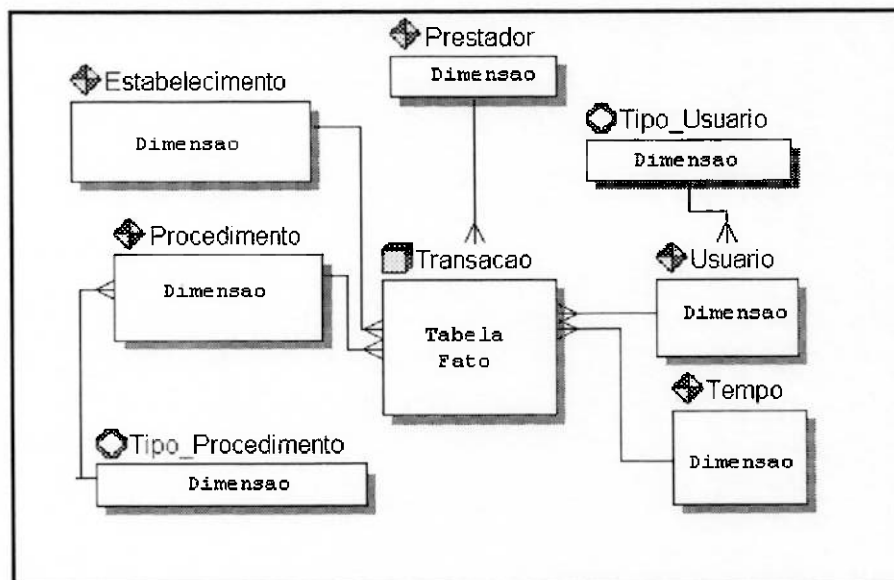


Figura 6 - Representação Esquema Floco de Neve
 Fonte: Adaptado de Kimball[4]

Outro caso em que o esquema floco de neve se faz necessário é quando a modelagem se depara com relacionamentos muitos-para-muitos entre uma tabela de fatos e uma dimensão[11].

ETL – Extração Transformação e Carga

O processo de ETL (Extraction, Transform and Load) é a parte do Data Warehouse responsável pela carga de dados e sua atualização a partir dos dados operacionais. Ele se resume em localizar onde estão os dados a serem extraídos, extraí-los, transformá-los para o formato de dados do DW, integrá-los conforme especificações dos metadados e, finalmente, lançá-los dentro do Data Warehouse. Os metadados do DW são baseados no modelo corporativo de dados.

Segundo Inmon[12], pode parecer um processo simples, dando a impressão que uma programação manual de uma simples ferramenta de extração dará o serviço como feito. Mas não é isso o que acontece geralmente. Ao começar a realizar o serviço de extração, os desenvolvedores do Data Warehouse se dão conta que esta tarefa é muito mais complexa do que parecia no início.

Inmon[12] cita alguns problemas que configuram o ETL como uma tarefa árdua e complexa, como:

- ✓ Há uma mudança na tecnologia de dados – é necessário ler os dados no SGBD legado e gravá-los em um SGBD mais novo e atual, capaz de suportar o Data Warehouse. Isso incorre numa necessidade de ferramentas para a comunicação entre esses dois tipos de base (middleware);
 - ✓ Os dados precisam ser convertidos:
 - Datas, medidas e padrões precisam ter uma única face. Exemplo: as datas em um banco de dados operacional estão no formato AA/MM/DD e em outro BD estão MM/DD/AA. No data warehouse foi adotado o padrão DD/MM/AAAA, então todos os dados precisam ser reformatados para entrar no DW. Essa tarefa é extremamente mecânica e árdua, e poderia ser realizada por conversores automáticos, por exemplo, ao invés de programada manualmente;
 - Os formatos lógicos dos dados, como formatos de texto EBCDIC e ASCII, devem ser convertidos também.
 - ✓ As chaves dos dados precisam ser reestruturadas. Em casos mais simples é somente uma questão de adicionar a dimensão tempo, em outros casos, as chaves nos diversos sistemas legados são diferentes e precisam ser convergidas para um único formato;
-

-
- ✓ A seleção dos dados que irão entrar no DW é extremamente custosa, podendo haver várias fontes para os mesmos dados;
 - ✓ É necessário eliminar a redundância de dados. Um mesmo cliente é cadastrado em várias aplicações diferentes, por exemplo. Ao entrar no DW, esse cliente deverá aparecer somente uma vez;
 - ✓ Os relacionamentos entre os dados em programas herdados não são documentados ou são colocados em uma lógica extremamente otimizada para tais programas. Inmon[12] coloca este problema como um dos piores a serem resolvidos, pois desvendar os dados atrelados dessa forma é um passo muito demorado e custoso;
 - ✓ A grande variedade de fontes de dados. Em consequência, há igual número de formatos e relacionamentos de dados a serem trabalhados;
 - ✓ Correção de dados. Dados não confiáveis de fontes operacionais devem ser corrigidos para não haver dúvidas quanto à veracidade das informações do Data Warehouse. Um simples algoritmo de correção pode resolver, mas talvez sejam necessárias ferramentas de inteligência artificial para tratar tais dados.

É necessário um cuidado especial com as ferramentas de ETL. No caso de pequenos Data Warehouses, com poucas fontes de dados e poucas tabelas, pode-se considerar a possibilidade de programação manual da extração. Mas com um Data Warehouse um pouco maior é praticamente obrigatório o uso de ferramentas automatizadas para a ETL dos dados.

2.2.3. Processo de Construção

O processo de construção que será apresentado a seguir segue a estrutura ilustrada na Figura 7.

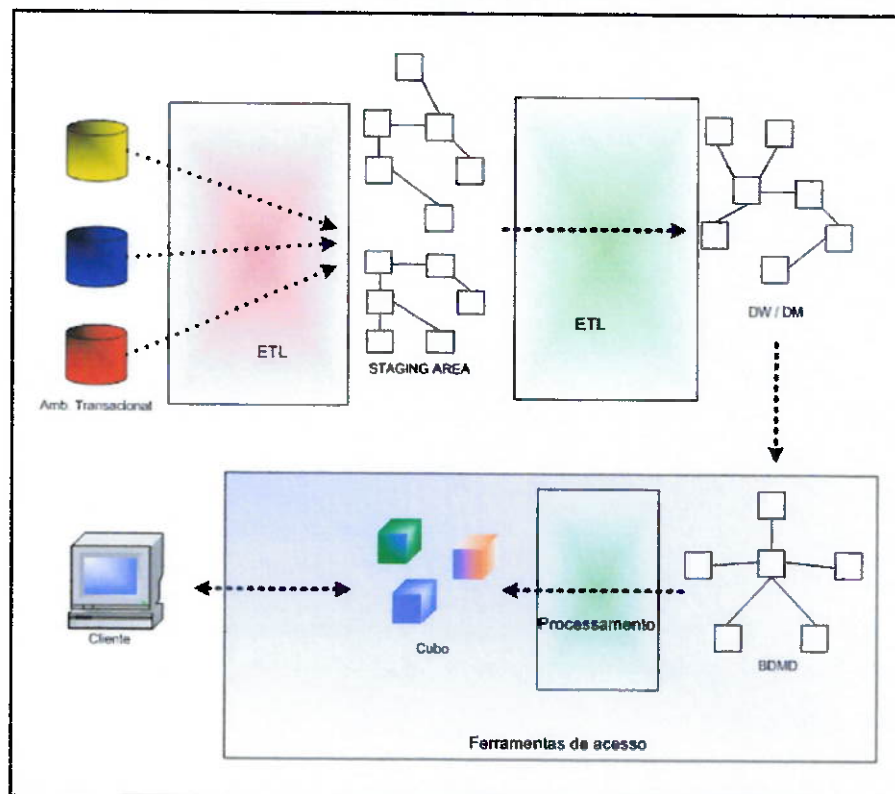


Figura 7 – Arquitetura do Modelo Proposto

Fonte: Elaborado pelo Autor

Como podemos observar, partindo de um ambiente transacional teremos a primeira fase da etapa de ETL já conceituada, integrando, efetuando a limpeza e a padronizando os dados. Estes dados serão armazenados em um modelo relacional denominado Staging Area.

O Staging Area por sua vez, é uma área de dados temporária e poderá ser volátil também, servindo apenas como ponte para o passo seguinte de ETL que realizará a consolidação dos dados que serão armazenados desta vez em um Data Mart/Data Warehouse.

Após serem transferidos para um banco de dados multidimensional, os dados são processados, sumarizados e assim irão compor os cubos, que por sua vez possibilitarão a

aplicação de técnicas e ferramentas de acesso que realizam os cruzamentos e resultam nas informações necessárias para apoiar a tomada de decisão.

3. A ÁREA DE SAÚDE

Iniciaremos apresentando algumas informações relacionadas à área de saúde, mais especificamente a Medicina Suplementar que constitui a base do presente estudo.

A Medicina supletiva como componente da área de saúde, é responsável pelo atendimento de mais de 41 milhões de pessoas com 370 mil leitos em 4.300 hospitais e 120 mil médicos. O mercado brasileiro de atendimento à saúde privada movimenta R\$ 15 bilhões por ano, abrangendo empresas de medicina de grupo, cooperativas, auto-gestão e seguradoras de saúde[13].

Neste mercado 43% estão conveniados à medicina de grupo, capitaneadas pela "Abramge" (Associação Brasileira de Medicina de Grupo), 23% estão sob a tabela da "Ciefas" (Comitê de integração de entidades fechadas de assistência a saúde) e "Abraspe (Associação brasileira de seguros pessoais) e com igual percentual estão as UNIMEDs representadas pela "Fenaseg" (Federação nacional de empresas seguradoras)[13].

Neste contexto a assistência à saúde do indivíduo, os planos de saúde e seguros de saúde são um mercado em expansão vertiginosa. Conseqüentemente, a disputa pelo cliente e sua adesão pela venda de planos requer um atendimento que satisfaça o cliente, cada vez mais exigente. A expectativa de um atendimento por um profissional competente em um serviço que ofereça alta qualidade se constitui hoje no componente obrigatório de rotina oferecida nos serviços privados[14].

Conforme pontua Isoldi[15]:

O período pelo qual estamos passando é caracterizado por mudanças radicais e muito rápidas. As grandes descobertas que ocorriam em séculos passaram a ocorrer em dias. Dentro deste cenário temos por um lado um grande desconforto, pois os padrões demoram a serem estabelecidos e a aparência é de que nada é seguro; por outro lado, temos a chance de mudar significativamente, obtendo importantes

avanços, tendo a possibilidade de resgatar os débitos que a sociedade possui, especialmente no tocante a distribuição da riqueza e do saber.

Já Lovelock e Wright[16], destacam que as novas tecnologias estão alterando radicalmente o modo pelo qual muitas organizações de serviço trabalham com seus clientes, bem como aquilo que acontece na parte interna das empresas. Talvez o fator mais poderoso de mudança atualmente proceda da integração entre computadores e a tecnologia da informação, onde iremos destacar o processo de Business Intelligence.

Diante deste novo cenário, a assistência médica ofertada pelas empresas de medicina supletiva necessita de maior agilidade no decorrer do processo de prestação de serviço e de tomada de decisão para que estes sejam eficientes.

Neste contexto, Leão[17] pontua que o uso da Tecnologia da Informação aplicada à saúde vem aumentando gradativamente. Neste campo do conhecimento podem-se destacar as seguintes atividades: o desenvolvimento dos sistemas de informação hospitalar, as redes de comunicação digital para a saúde, as aplicações voltadas para a saúde comunitária, os *sistemas de apoio à decisão*, o processamento de imagens e sinais biológicos, a avaliação e controle de qualidade dos serviços de saúde e o uso do cartão magnético como *autorizador eletrônico de serviços médicos*, onde este conjunto de tecnologias compõe os Sistemas transacionais na área de saúde abordados mais a seguir.

Mezomo[14] ressalta que a credibilidade dos serviços sofre um incremento com a utilização da Tecnologia de Informação deixando transparecer uma atenção mais moderna e eficiente pela utilização destas novas ferramentas. Atualmente, nossa sociedade depende, cada vez mais, de computadores e novos equipamentos tecnológicos que têm auxiliado os indivíduos em seu dia a dia.

Assim, as empresas de medicina suplementar na sua maioria adotam um Sistema de Cartão Magnético, como autorizador de serviços médicos,

visando oferecer serviços com qualidade, adicionando valor aos clientes, oferecendo comodidade, conforto e rapidez.

Considerando que os clientes deste segmento - serviços de saúde – estão atentos para as instituições que se preocupam com o usuário, quando os concorrentes procuram dinamizar a oferta pelos serviços com preços que se equivalem, o diferencial das empresas de medicina supletiva com este sistema serão os valores agregados aos planos ofertados.

Podemos ressaltar a utilização da tecnologia do Business Intelligence que irá incrementar o ambiente de Sistemas Transacionais, proporcionando maior precisão e apoio nas decisões estratégicas das empresas da área de saúde assim como outros benefícios, que serão abordados no *capítulo 5*.

A seguir no *capítulo 4* será apresentado o modelo inicial que ilustra as arquiteturas de Sistemas Transacionais utilizadas na área de saúde.

4. SISTEMAS TRANSACIONAIS NA ÁREA DE SAÚDE

A descrição abaixo irá ilustrar um modelo de Arquitetura de Sistemas Transacionais utilizado na área de saúde, adotado como referência para o desenvolvimento deste trabalho.

Esta arquitetura é composta por dois ambientes transacionais com característica predominante OLTP, presente nos ambientes que abrigam os dados gerados a partir de procedimentos médicos registrados de forma eletrônica, devido ao volume de transações geradas nos atendimentos a procedimentos médicos. Neste modelo as transações eletrônicas são capturadas através de um POS (Point of Sales), uma URA (Unidade de Resposta Audível) ou via WEB, passando por um processo de autorização eletrônica e por fim são armazenadas em bancos de dados relacionais.

Temos também nestes ambientes acessos provenientes de Sistemas de Informação diversos, a execução de consultas *ad hoc* e ainda a extração de relatórios gerenciais mesmo que limitados no que se refere à quantidade de possíveis cruzamentos de informações.

Podemos observar o impacto negativo e a dificuldade na extração de informações analíticas mais complexas, devido às suas estruturas baseadas em modelos E/R e construídas em SGBD relacionais. Isto faz com que as possibilidades de atender a solicitações de cruzamento de informações sejam inviáveis, devido ao fato de que só podem ser utilizadas duas dimensões. Estas informações são necessárias para as análises, que servem de parâmetro para o processo de tomada de decisão, e até mesmo detecção de possíveis falhas operacionais.

A visualização deste modelo de ambiente inicial está representada na *Figura 8*, onde se pode observar o fluxo das transações desde os meios de captura, até o armazenamento dos dados. Este armazenamento é feito de forma individual em bancos de dados relacionais específicos para cada cliente e simultaneamente replicadas para outra máquina que ficará

disponível para a visualização dos dados em forma de relatórios ou consultas simples provenientes de Sistemas de Informação.

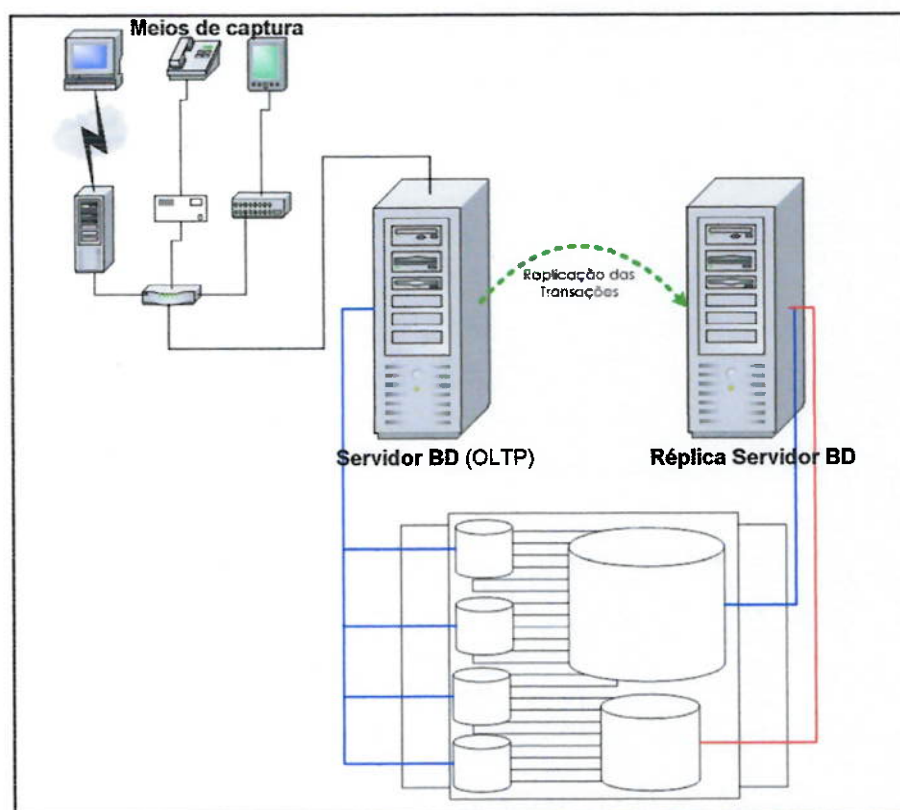


Figura 8 - Composição do Modelo Inicial
Fonte: Elaborado pelo Autor

Este modelo de replicação ilustrado na *Figura 8* atende duas finalidades, onde a primeira realiza o papel de contingência dos dados de transações e a segunda proporciona a distribuição de carga de acessos *ad hoc* e as diversas consultas dos Sistemas de Informações aos dados.

A seguir no *capítulo 5* será apresentado o modelo proposto, com a sua arquitetura, componentes e evidenciando as características e especificidades utilizadas na sua implementação.

5. MODELO DE BUSINESS INTELLIGENCE NA ÁREA DE SAÚDE

Após uma breve apresentação de informações relacionadas ao tema deste trabalho, o modelo proposto utilizará como referência o modelo inicial (OLTP) descrito no *item 4* e representado na *Figura 8.*

Este modelo de implementação de Business Intelligence na área de saúde, tem como proposta apresentar uma forma de se aproveitar os dados de transação na geração de benefícios ao dia-a-dia da empresa. Podemos descrever como alguns destes benefícios a identificação de fraudes quanto ao uso do serviço, identificação e mapeamento de pacientes dotados de doenças crônicas, entre outros mais.

A identificação de fraudes poderá ser realizada mapeando o comportamento do cliente quanto ao uso de especialidades as quais ele tem acesso, se estas especialidades são compatíveis com seu perfil de idade, sexo, local de residência entre outros. Em geral a identificação de fraudes é obtida a partir de modelos para Data Mining, que também faz parte de BI, podendo ser criado a partir do BDMD (Banco de Dados Multi Dimensional).

Já o mapeamento de pacientes crônicos segue uma linha parecida, uma vez que se observa que um mesmo paciente tem recorrência freqüente em uma mesma especialidade médica, exames de rotina e receita de medicamentos.

Baseado no mapeamento destas informações a empresa de medicina supletiva poderá realizar campanhas onde ela envia gratuitamente os medicamentos que deverão proporcionar um quadro mais estável ao paciente, faz contatos periódicos para saber sobre a utilização do mesmo e como resultado deste trabalho, deverá ter uma diminuição da utilização dos serviços e procedimentos médicos, por exemplo, uma internação ou exame de ressonância, por parte deste paciente e assim diminuir seus custos com este grupo de pacientes crônicos.

A seguir detalharemos os componentes do ambiente analítico responsável por atender a esta e outras demandas de informação.

5.1. Ambiente Analítico

Partindo de um ambiente OLTP responsável por tratar e responder a transações eletrônicas provenientes de procedimentos médicos, está sendo proposto um modelo de implementação de Business Intelligence na área de saúde, para que uma vez que os dados destas transações foram armazenados em banco de dados possam ser trabalhados para gerar informações.

Para o tratamento destes dados está sendo proposta uma estrutura de Business Intelligence, contida no Ambiente Analítico já definido anteriormente, composta de alguns elementos básicos descritos no item 4. Desta forma detalharemos o uso de cada um destes componentes relacionados à proposta deste trabalho.

Aqui é apresentada uma tabela que exhibe algumas das diferenças entre os ambientes OLTP e OLAP, sob determinados pontos de vista:

Tabela 2 - Comparativo entre OLTP e OLAP
Fonte: Elaborado pelo Autor

	Processamento de transações (OLTP)	Data Warehouse (OLAP)
Objetivo	Operações do dia-a-dia	Recuperação e análise da informação
Estrutura	BD relacional otimizado para processamento de transações	BD relacional otimizado para processamento de consultas
Modelo de Dados	Normalizado	Multidimensional
Acesso	SQL	SQL e ferramentas avançadas de análise
Tipo de Dado	Dados para operar o negócio	Dados para analisar o negócio
Natureza do Dado	Detalhado	Resumido e detalhado

A seguir apresentaremos formas de implementar a Arquitetura do Data Warehouse dentro do modelo proposto por este trabalho.

A escolha da arquitetura é uma decisão de projeto, e leva em consideração os fatores relativos à infra-estrutura disponível, ao ambiente de negócios, juntamente com o escopo de abrangência desejado.

A abordagem de implementação escolhida é uma decisão que pode causar impactos quanto ao sucesso de um DW. Muitas variáveis afetam a escolha da implementação e arquitetura, entre elas o tempo para a execução do projeto, o retorno de investimento a ser realizado, a visualização dos benefícios da utilização das informações, a satisfação dos usuários executivos e os recursos necessários à implementação de uma arquitetura[2].

Antes de apresentar as arquiteturas de DW, é preciso definir um componente importante nesse contexto, o Data Mart (DM). Um Data Mart é um conjunto de tabelas de fatos juntamente a um conjunto de tabelas de dimensão conectadas que atende às necessidades de um grupo particular de negócio, segundo Kimball e Merz[7].

Muitos modelos multidimensionais de aparência semelhante podem ser construídos em torno da empresa, representando fontes diversas de dados e os interesses de grupos diferentes que desejam analisar tais dados. Cada um desses modelos multidimensionais pode ser chamado de um DM ainda segundo Kimball e Merz[7].

As arquiteturas conhecidas na literatura são três: Global, Independente e Integrada.

— Arquitetura Global

A Arquitetura Global é considerada como a que suporta toda ou a maior parte dos requisitos ou necessidades de um DW integrado com grande grau de acesso e utilização das informações para todos os departamentos de uma empresa. Para implementar essa arquitetura é utilizada a técnica *Top-Down*, um exemplo é apresentado na *Figura 9*.

Proposta por Inmon[3], ela requer um maior planejamento e trabalho de definições conceituais de tecnologia completos antes de se iniciar o projeto do DW propriamente dito.

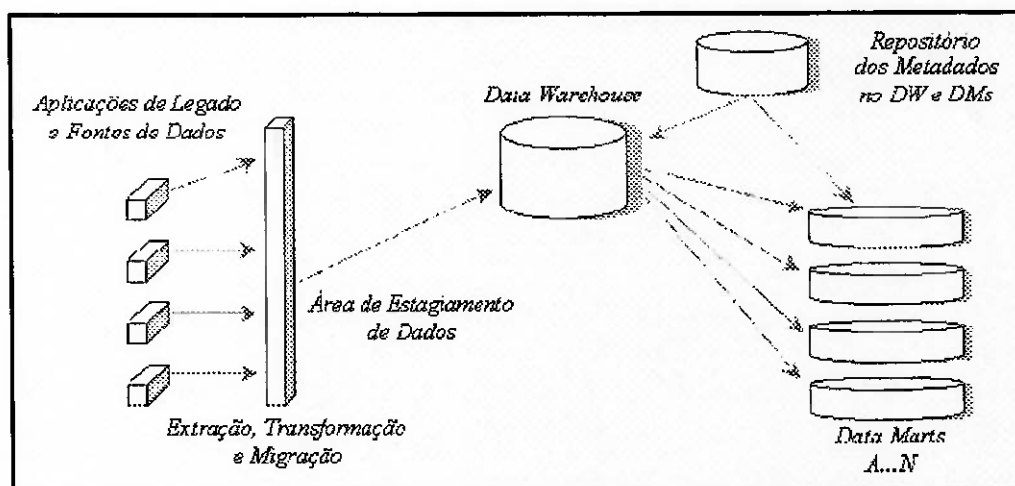


Figura 9 - Implementação Top_Down
Fonte: Adaptado de Machado[2]

A Tabela 3 apresenta uma síntese das vantagens e desvantagens da utilização da implementação Top-Down.

Tabela 3 - Comparativo da Implementação Top-Down

Fonte: Machado[2]

Vantagem	Desvantagem
Herança de arquitetura, permitindo uma fácil manutenção.	Implementação muito longa podendo levar facilmente mais de um ano
Visão global do empreendimento	Alta taxa de risco
Controle e centralização de regras	Herança de cruzamentos funcionais
Repositório de metadados centralizado e simples	Expectativas relacionadas ao ambiente, devido à demora do projeto.

Os dados são extraídos de bases operacionais e fontes de dados externas em horário fora do pico das operações. São filtrados, eliminando-se dados não necessários e tratados para a qualidade dos requisitos levantados para o projeto. Por fim são carregados nas bases de dados apropriadas de Data Warehouse para acesso aos usuários finais[2].

Arquitetura Independente

Já a proposta de Kimball[8] consiste na arquitetura independente que implica em Data Marts *stand-alone*, controlados por um grupo específico de usuários e que atende somente às suas necessidades específicas e departamentais, sem foco corporativo. Esse fato faz com que não exista conectividade entre Data Marts da organização. Esta arquitetura é apresentada na *Figura 10*.

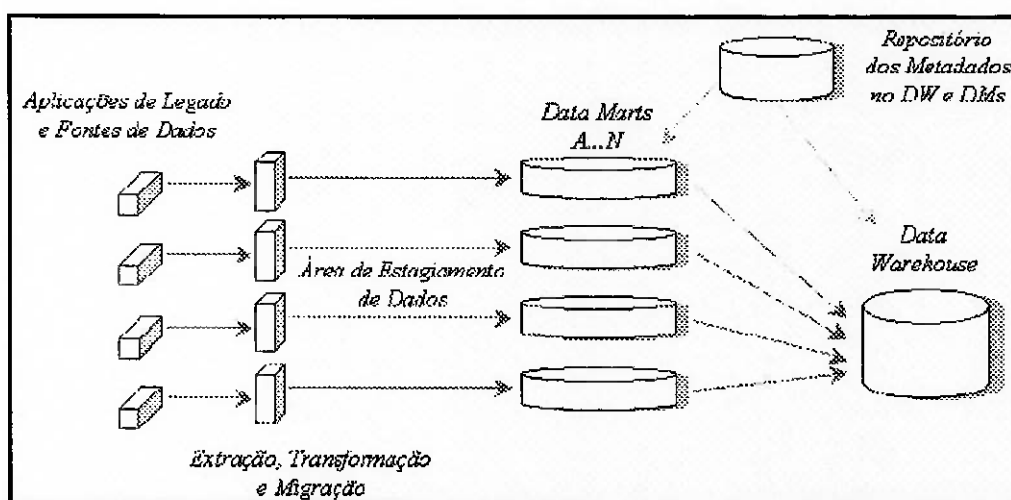


Figura 10 - Implementação Bottom-Up
Fonte: Adaptado de Machado[2]

A *Tabela 4* apresenta um comparativo das vantagens e desvantagens da utilização da implementação Bottom-Up.

Tabela 4 - Comparativo da Implementação Bottom-Up
Fonte: Machado [2]

Vantagem	Desvantagem
Implementação rápida	Perigo de <i>legamarts</i> : DM isolados
Retorno rápido	Desafio de possuir visão do empreendimento
Manutenção no enfoque da equipe	Administrar e coordenar múltiplas equipes e iniciativas
Herança incremental	Maldição do sucesso: enquanto um grupo fica feliz com os resultados, outros grupos pressionam por resultados.

Ao contrário da abordagem anterior, esta parte da implementação é chamada de Bottom-Up. Dessa forma serão implementados Data Marts individualmente, sem partir da análise de requisitos para toda a empresa[6].

Arquitetura Integrada ou BUS

Essa forma de implementação tem o propósito de integrar a implementação Top-Down com a Bottom-up. Essa abordagem é conhecida como desenvolvimento baseado em Data Marts incrementais e foi proposta por Kimball[8].

A solução proposta por Kimball e Merz[11], é utilizar-se de dimensões e fatos adaptados que estão distribuídos por todas as partes do DW. Uma dimensão adaptada é uma dimensão que significa a mesma coisa e que tem a mesma estrutura através de diversos Data Marts. Já um fato adaptado é uma definição de um fato medido que é consistente através de diversos Data Marts. Geralmente, uma instância específica de um fato existe em somente uma localização do Data Warehouse. É a definição do fato que é consistente através dos Data Marts.

Conforme citam Kimball e Merz[11]: *Dimensões adaptadas e fatos adaptados implementam uma arquitetura Data Warehouse Bus. São somente os dados de dimensão que precisam e devem ser duplicados.*

Com base nas informações detalhadas até aqui definimos que a arquitetura mais adequada para a proposta deste trabalho é a Arquitetura Integrada ou BUS.

A implementação da arquitetura definida diz respeito ao modelo que combina *Top Down* e *Bottom Up*, que se baseia no planejamento *Top Down* de um *Data Warehouse* com desenvolvimento *Bottom Up*, construindo um Data Mart que poderá depois ser agregado a outros, formando assim o Data Warehouse.

Esta definição leva em consideração que, na construção de um DW há uma gama de fatores que afetam a complexidade como, por exemplo, a construção do projeto que é lenta e cara. Com o objetivo de equilibrar os gastos e oferecer resultados em prazos mais curtos, optou-se por construir Data Marts incrementais.

Esta definição também está baseada na leitura onde segundo Kimball a área de saúde apresenta um cenário bastante interessante no que diz respeito ao desenvolvimento de um Data Warehouse, trazendo uma grande criticidade e complexidade no desenvolvimento destes modelos.

A criticidade é proveniente do grande número de dados críticos compartilhados entre os participantes chamados de consórcio de saúde. Este por sua vez é formado por clínicas, hospitais, laboratórios, planos de saúde e seguradoras de saúde.

Já a complexidade citada vem do fato das transações da área de saúde trafegarem informações do paciente, do procedimento médico, e ainda informações financeiras relacionadas ao(s) procedimento(s) efetuados, e/ou materiais utilizados nos procedimentos.

Devido a esta quantidade de informações contidas numa mesma transação deveremos, ao iniciar o processo de tratamento dos dados para alimentar o Data Warehouse, segmentar esta informações em diferentes sessões de dados.

Como ilustração da arquitetura proposta, segue a figura com uma visão macro de todo o ambiente analítico que irá conter os dados gerados nos sistemas transacionais e que agora poderão ser tratadas e utilizadas na geração das informações analíticas.

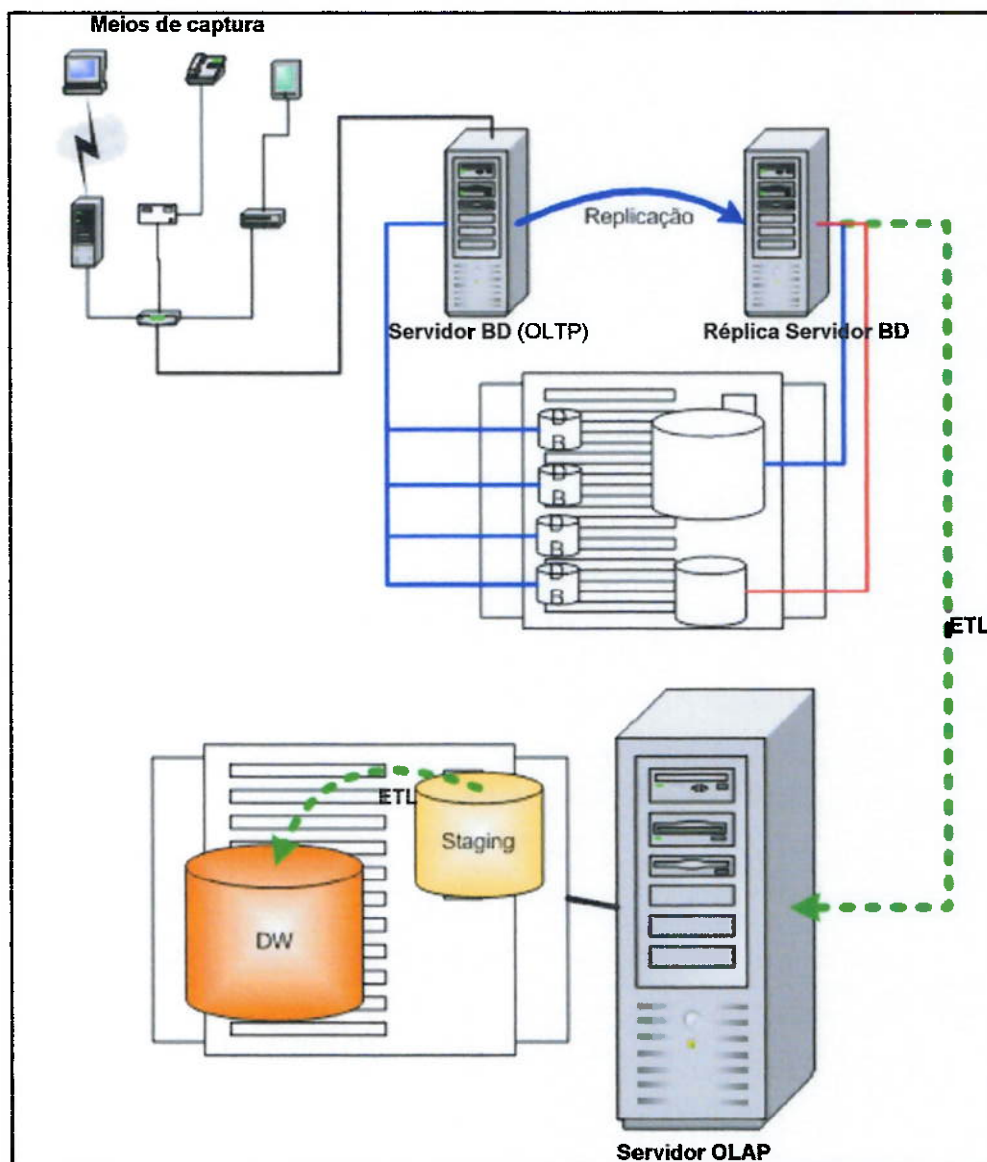


Figura 11 - Composição do Modelo Analítico Proposto
 Fonte: Elaborado pelo Autor

Podemos destacar o ambiente descrito na base da *Figura 11* dotada de link tracejado indicando o fluxo e de que forma ele estará interligado ao *modelo de referência* já abordado anteriormente.

A seguir iremos detalhar os componentes da arquitetura do Modelo de Implementação de Business Intelligence na área de saúde e seu processo de implementação.

5.2. Staging Area - ODS

Iniciaremos com a modelagem de uma Staging Area, a qual servirá de ambiente intermediário entre o ambiente inicial (OLTP) e os Data Marts, com o propósito de armazenar e tratar os dados.

A seguir será apresentada a proposta de duas Staging Area, sendo a primeira denominada "Staging Transações" e a segunda "Staging Financeiro".

— Staging Transações

A *Staging Transações* irá armazenar os dados gerados por transações eletrônicas referentes a execução de procedimentos médicos, como por exemplo, que tipo de procedimento foi executado (consulta, internação, exame), em quais quantidades, em que data, local e assim por diante. Estes dados estarão compondo a "Staging Transações", ilustrado na *Figura 12*.

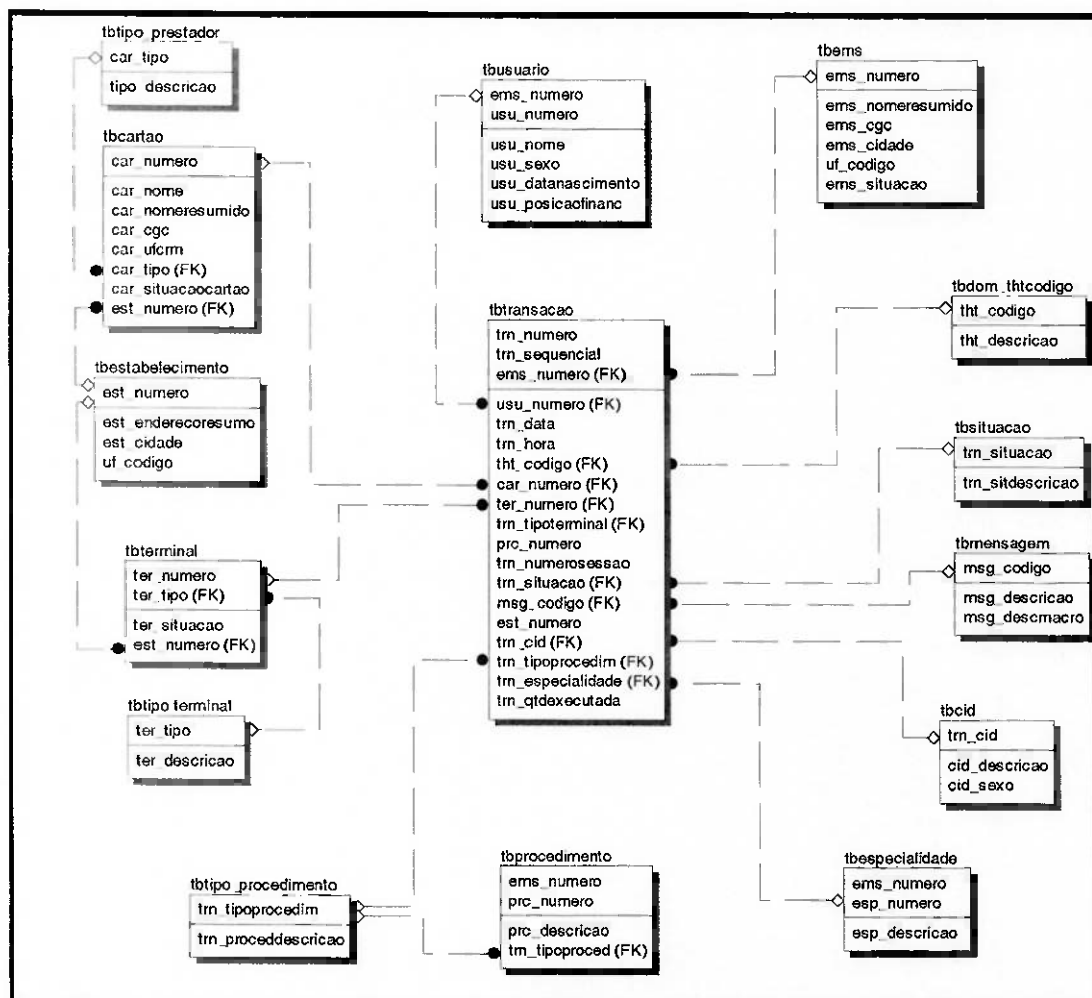


Figura 12 - Modelo de Staging Transações

Fonte: Elaborado pelo Autor

Staging Financeiro

Os dados referentes a valores (custos) relacionados a estas transações de procedimentos médicos e/ou materiais utilizados estarão compondo o *Staging Financeiro*.

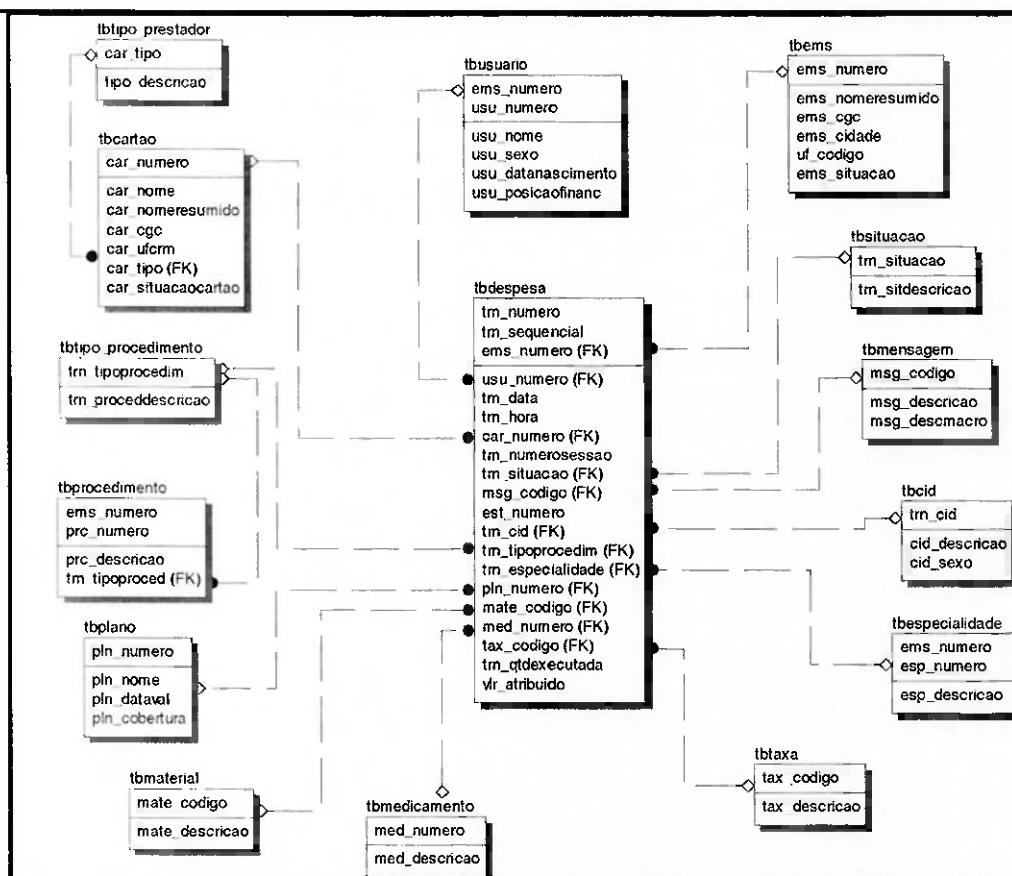


Figura 13 - Modelo de Staging Financeiro
 Fonte: Elaborado pelo Autor

Este segundo modelo "Staging Financeiro" será sugerido apenas para demonstrar a arquitetura BUS definida para este estudo, uma vez que será dada maior ênfase ao "Staging Transações".

5.3. ETL – Extração Transformação e Carga

O processo de ETL(Extraction, Transform and Load) ou extração, transformação e carga estará presente em dois momentos do modelo proposto: entre o ambiente OLTP e o Staging Transações e posteriormente entre o Staging Transações e o Data Mart Transações.

O Staging receberá os dados referentes às transações através de um processo denominado ETL, descrito na seção 2.2.2. Tal processo executará a extração de dados relevantes já pré-definidos, seguida de um processo de transformação e/ou adequação dos dados e por fim efetuará a carga destes no modelo definido para o Staging Transação.

Os dados contidos neste ambiente de Staging Transação serão do tipo volátil, sendo excluídos ou até mesmo atualizados a cada período de tempo.

Após a carga finalizada no ambiente do Staging Transações, iniciaremos uma nova etapa do processo de ETL destinada a dar carga das informações no Data Mart Transações. Este irá conter informações não voláteis, ou seja, acumulará informações com frequência a ser definida formando assim a base de informações históricas, que por sua vez possibilitará a pesquisa de informações de comportamentos, estatística, entre outras.

O processo de ETL está ilustrado a seguir demonstrando no retângulo não tracejado o detalhe de cada uma das três etapas (extração, transformação e carga).

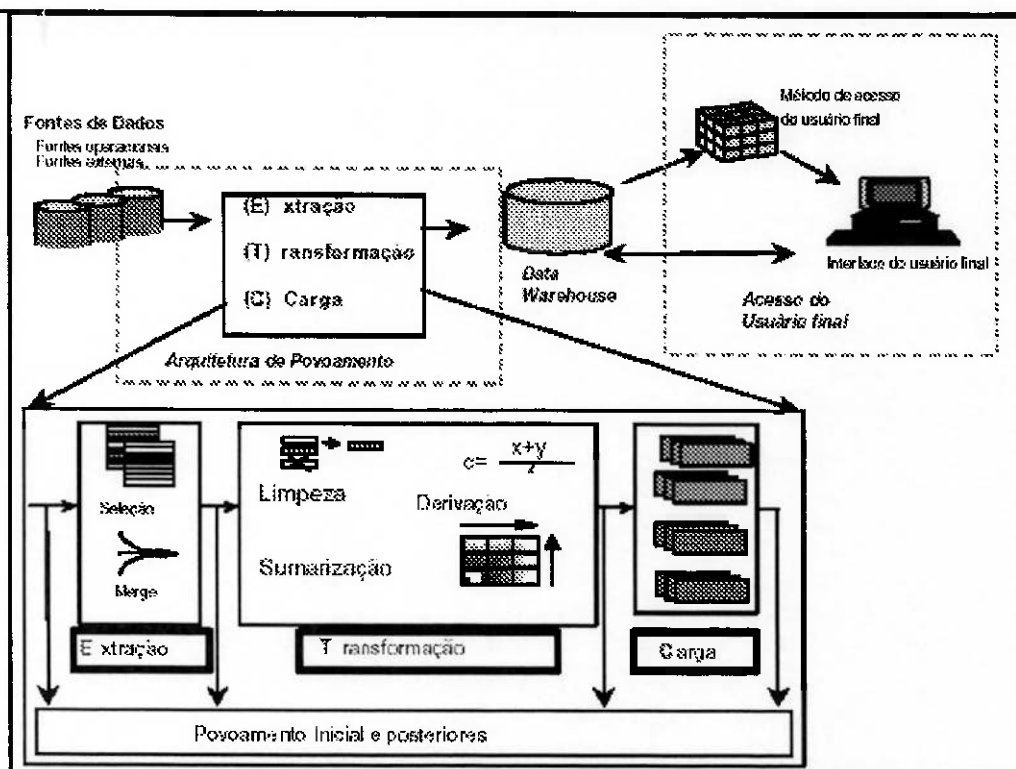


Figura 14- Processo de ETL de um Data Warehouse
Fonte: Goodyear[18]

Várias empresas sejam por falta de conhecimento, seja por falta de consideração, menosprezam a fase da Staging Area justamente por pensarem nela como um processo que apenas realiza conversões de dados de bancos transacionais para os ambientes de Data Warehouses. Este engano fatalmente acarreta desperdício de tempo e, conseqüentemente, dinheiro, por terem que refazer todo o escopo do projeto de migração implementado[9].

A Staging Area pode ser considerada uma das fases mais críticas e importantes do Data Warehouse; constitui uma área de armazenamento intermediário entre as fontes de dados dos sistemas transacionais (fontes OLTP) e os DWs, e é nela que está implementado o processo conhecido por ETL.

A fase de extração consiste na captura dos dados, desde planilhas e documentos textos até sistemas transacionais, e posterior alocação no Data Warehouse. Porém, os dados migrados nem sempre são

íntegros e podem apresentar diversos problemas, como inconsistências e duplicidades.

Após todo o processo anterior, a fase de transformação começa a atuar, tornando os dados íntegros e convertendo-os para formatos consistentes, além da realização do processo de limpeza, onde os erros encontrados nos dados são removidos.

A fase de carga dos dados age a partir do momento em que se tem a garantia da consistência dos dados, capturando-os e carregando-os nas tabelas do Data Warehouse.

EXTRAÇÃO

O objetivo dos métodos de extração é isolar os dados que serão empregados pelos sistemas de suporte à decisão. Porém, uma das preocupações destes métodos é a captura apenas das porções de dados modificadas, a fim de se evitar desperdício de desempenho e espaço.

Existem diversas técnicas de captura dos dados no processo de extração, sendo classificadas em Estática e Incremental.

A Técnica de Captura Estática de dados consiste numa aquisição periódica do dado, ficando a cargo do usuário definir a periodicidade do processo. Assim que é inicializado, o processo adquire as modificações realizadas desde o período da última atualização e as atualiza no DW. Todavia, este processo não adquire as versões intermediárias dos registros, mas a sua versão corrente. Uma vantagem desta aplicação é a possibilidade de execução em horários fora de pico de trabalho, não comprometendo a performance dos sistemas transacionais.

Snapshot dos dados, timestamp dos dados e comparação de arquivos são exemplos de técnicas enquadradas na captura estática.

O princípio da técnica Snapshot dos dados é capturar os dados dos sistemas e carregá-los no DW. Esta carga pode ser completa ou, então, através da adição de dados alterados.

A técnica de Timestamp captura os dados com data de atualização posterior à data da última execução do processo; sua vantagem é a captura de um volume menor de dados.

Já a Comparação de Arquivos se baseia na captura de registros que não constam no escopo atual do sistema através da comparação das suas chaves, podendo identificar as inclusões e as exclusões dos dados.

A técnica de Captura Incremental dos dados, diferentemente da técnica estática, tem como meta capturar todos os estados pelos quais os registros passam, ou seja, qualquer mudança de valor é percebida e computada no sistema, atualizando assim o DW. Esta técnica, apesar de armazenar todos os históricos dos registros, acaba ocasionando *overhead* para o sistema.

Entre as técnicas que se encaixam nesta categoria estão os gatilhos, os arquivos de logs e o gerenciador de aplicações.

Gatilhos também conhecidos por *Triggers* são procedimentos disparados quando um evento específico para eles acontece. Assim, ao constatar a modificação de um registro na base de dados, uma determinada trigger é disparada e instantaneamente este registro é atualizado na base do DW.

Os arquivos de logs são utilizados para armazenar informações das alterações ocorridas na base de dados de um sistema. Portanto, periodicamente este log é consultado e a base do DW é atualizada a partir dele, desde o ponto da última atualização efetuada.

A técnica de Gerenciamento de Aplicações consiste na mudança da lógica da aplicação, de forma a duplicar uma alteração na base transacional para uma área temporária, para sua posterior recuperação.

TRANSFORMAÇÃO

A começar do momento em que os dados já tenham sido extraídos dos sistemas transacionais pelo processo de extração e alocados na Staging Area, a fase de transformação é iniciada. E a partir dos dados

desta área, nesta fase, um conjunto de processos é executado, dentre eles: limpeza, geração de novas chaves e obtenção do modelo multidimensional.

A construção do modelo multidimensional propõe a criação de um modelo não normalizado na Staging Area para, em seguida, realizar a derivação dos modelos multidimensionais através de transformações a serem aplicadas sobre o modelo de dados da Staging Area.

As rotinas de limpeza e integração dos dados têm como objetivo assegurar a consistência no DW. Assim, existem rotinas de padronização de campos, como definição de codificações e formatos de campos e dados, além de compatibilizar suas chaves e verificar a veracidade das informações.

Existe um conjunto de transformações e filtrações que são realizadas sobre a Staging Area.

São realizadas exclusões de informações desnecessárias, excluindo entidades e atributos que não estejam no escopo a ser tratado do DW, ou seja, não apresentam informações interessantes para o Data Warehouse; desnormalização das relações de tabelas, influenciando o desempenho das consultas, pois diminui a quantidade de junções de tabelas; processo de *merging* de tabelas, propondo a junção de duas ou mais tabelas semelhantes numa única; definição de categorias, atribuindo categorias a determinados tipos de informações, de tal forma a separá-los de maneira hierárquica, classificando-os para uma futura análise. Normalmente estas informações são divididas em faixas, de modo a agrupá-las em intervalos e serem melhor analisadas.

É realizada também a inserção do elemento Tempo à dimensão, desde que haja a necessidade de rastrear modificações em dados e atributos no decorrer do tempo.

A geração de novas chaves é outra técnica utilizada na fase de transformação e filtração.

Segundo Kimball[4], as chaves identificadoras das dimensões do Data Warehouse não devem ser as mesmas existentes nos sistemas transacionais, a fim de acompanhar as atualizações nas dimensões.

A criação da dimensão Tempo no DW vai possibilitar a realização de análises comparativas sobre os fatos, com o intuito de manter armazenado o histórico dos dados no decorrer do tempo. É justamente esta dimensão que vai determinar o grau de granularidade do DW, como, por exemplo, dados relativos a um dia, uma semana ou, até mesmo, um ano.

CARGA

O processo de carga é realizado após todos os tratamentos feitos nos dados nos processos de extração e transformação. Esta etapa consiste em carregar os dados tratados e armazenados na Staging Area e migrá-los para o Data Warehouse.

Esta migração, porém, não é simples. É necessário verificar a integridade dos dados de acordo com as chaves existentes na Staging Area e no Data Warehouse, tornando-as compatíveis.

A carga pode ser feita de forma incremental ou com base nos dados. A forma incremental normalmente é feita para tabela de fatos, incrementando e atualizando relacionamentos e dimensões. A segunda forma é realizada nas tabelas dimensões, apagando todas estas tabelas e reconstruindo-as, incluindo assim seus novos dados.

Dessa forma, vale ressaltar que a escolha das técnicas de implementação dos processos de extração, transformação e carga e sua ligação com a Staging Area vai incidir diretamente no funcionamento do Data Warehouse, atribuindo-lhe características específicas, como grau de confiabilidade, desempenho, versatilidade, robustez e escalabilidade.

Assim, a implementação dos processos de extração, transformação e carga e a sua comunicação com a Staging Area vai ser de vital importância para o bom funcionamento do DW e sua

integração com a corporação, extraindo o máximo de informações possíveis e proveitosas.

5.4. Data Mart / Data Warehouse

Os Data Marts idealizados para a composição de um Data Warehouse na área de saúde estão listados abaixo:

- ✓ *DM Transação*: contém informações sobre a execução procedimentos médicos;
- ✓ *DM Financeiro*: contém informações sobre contas a pagar;
- ✓ *DM Cobrança*: contém informações sobre contas a receber;
- ✓ *DM Participação Prestador*: contém informações sobre o acompanhamento da utilização do sistema de transações;
- ✓ *DM Gerenciamento de Doenças*: contém informações que possibilitam a identificação de indivíduos doentes ou com potencial de contrair doenças, com objetivo de aplicar programas preventivos;
- ✓ *DM Análise Glosa*: contém informações sobre análise dos processos de glosa - termo usado quando a operadora de saúde se recusa a pagar determinado procedimento executado.

Neste estudo serão detalhados apenas a arquitetura e o modelo do Data Mart Transação e Data Mart Financeiro.

Os Data Marts surgirão após todo o processo inicial de ETL que armazenou dados provenientes do ambiente transacional no Staging Area. Executa-se então um novo processo de ETL que armazena os dados no Data Mart.

Como podemos observar, a figura abaixo traz o modelo *floco de neve* desenvolvido para o propósito deste trabalho. Nele encontraremos diversas informações como, por exemplo: para cada conjunto de transação analisada podemos determinar em que região (país, região, estado) ela foi realizada, em que data (ano, mês, semana), em qual estabelecimento (clínica, hospital, laboratório), quem foi o prestador (médico, técnico laboratorial), quais especialidades este prestador está habilitado para atender (cardiologia, clínica geral, neurologia), qual foi o procedimento executado (consulta, internação, exame), situação deste

procedimento (aprovado, negado, justificado) e assim os diversos cruzamentos de informações que este modelo possibilita.

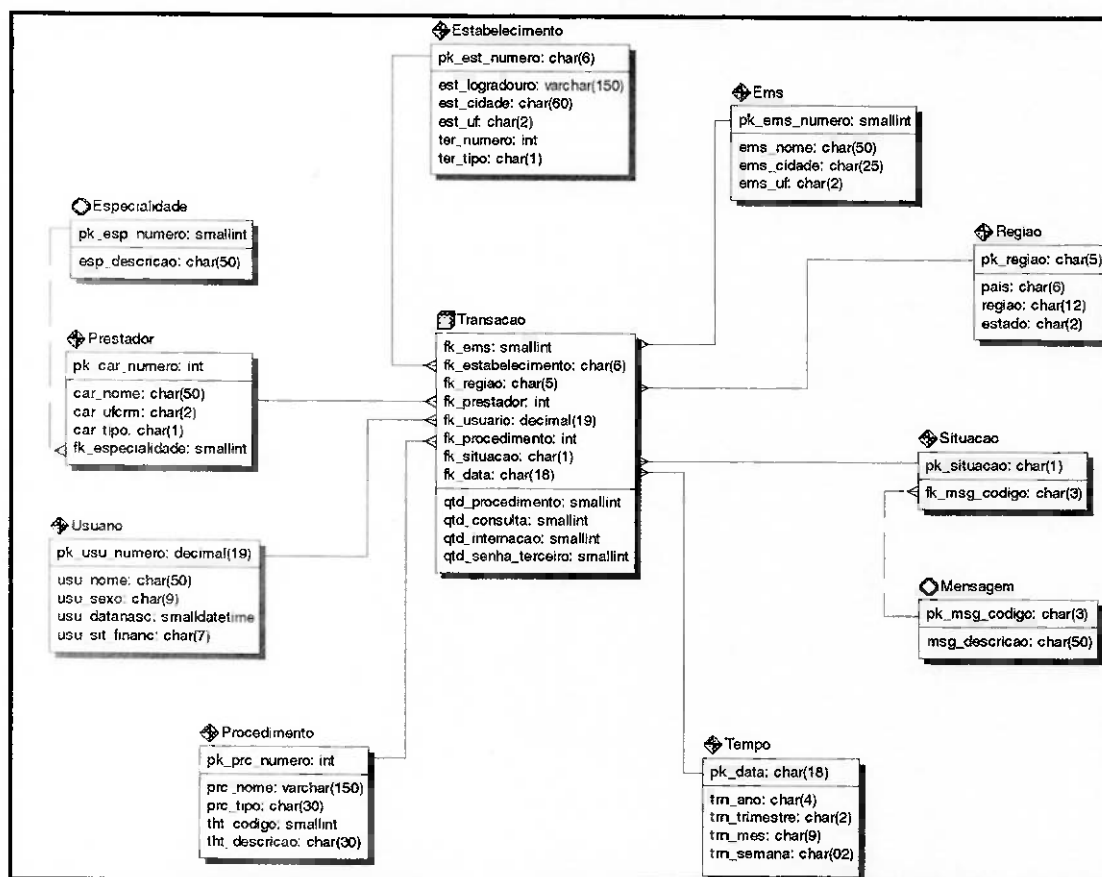


Figura 15 – Modelo do Data Mart Transações
Fonte: Elaborado pelo Autor

Neste Data Mart Transações podemos observar ao centro a tabela de fatos denominada *Transação* e suas diversas *Dimensões*: EMS (Empresa de Medicina Suplementar), Região, Situação, Tempo, Procedimento, Usuário, Prestador/Especialidade e Estabelecimento.

Aqui será apresentado, porém não será detalhado a estrutura do Data Mart Financeiro.

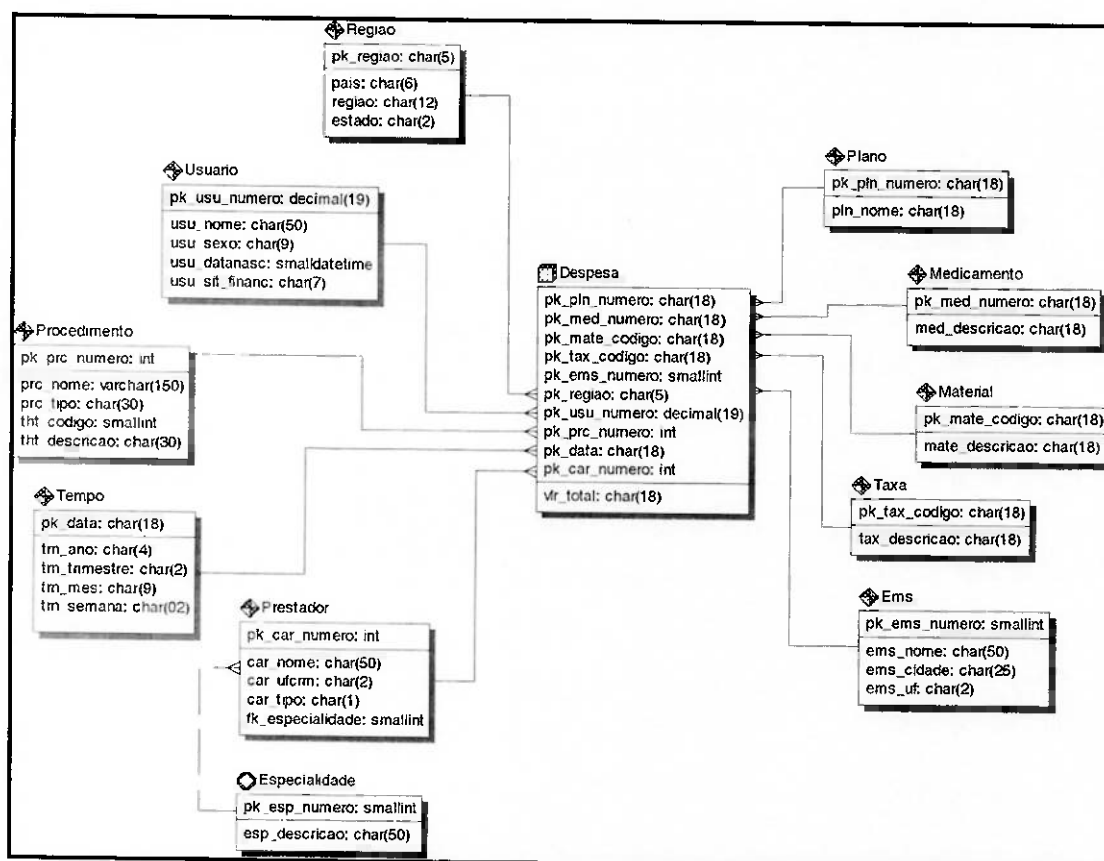


Figura 16 - Modelo do Data Mart Financeiro
 Fonte: Elaborado pelo Autor

Considerando que todos os Data Marts relacionados no início deste item foram devidamente populados com as informações definidas pelo seu modelo, teremos com a “união” deles a representação do Data Warehouse.

A seguir na *Figura 17* temos a ilustração do Data Warehouse representado aqui pelo DM Transação e DM Financeiro, proposto para armazenar e possibilitar a exploração das informações de forma analítica.

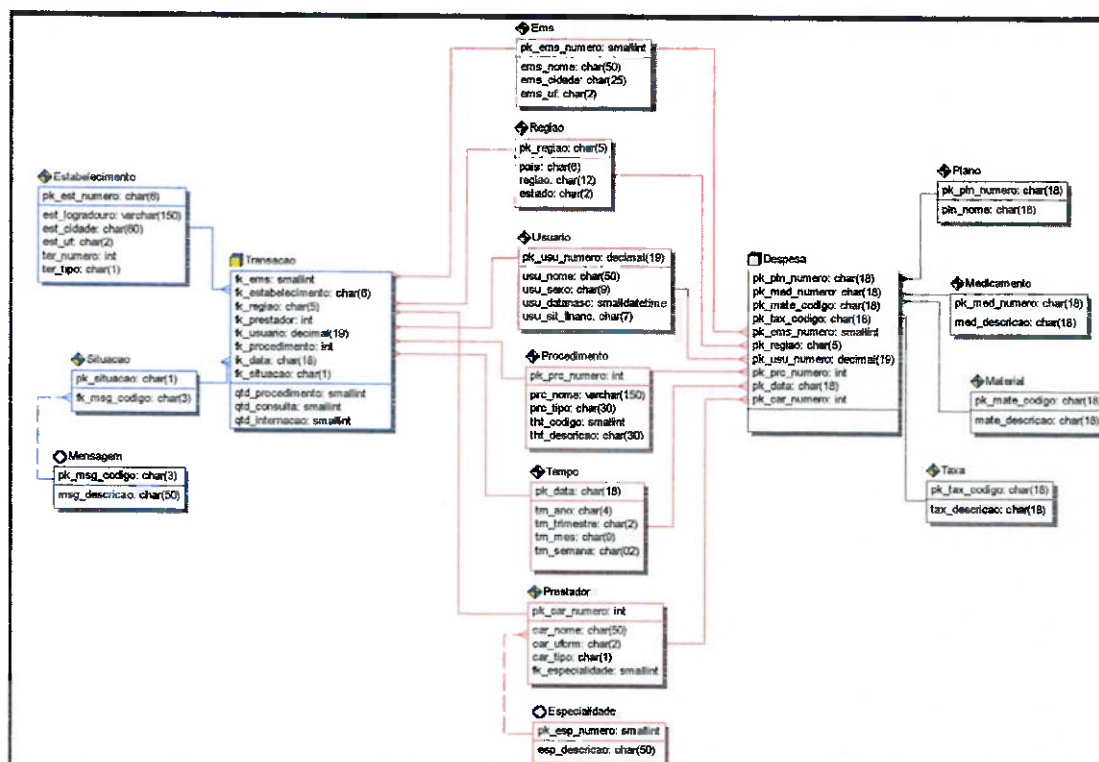


Figura 17 - Modelo do Data Warehouse parcial para Área de Saúde

Fonte: Elaborado pelo Autor

Uma vez criado o Data Warehouse poderá ser iniciado o uso das ferramentas OLAP de extração e manipulação das informações, como poderemos ver na sessão 5.5.

5.5. Modelos de Representação Física

Para falar do Modelo de Representação Física, iniciaremos com as formas de Analisar os dados:

É importante definir as principais categorias de exploração de dados e mostrar quais as transformações que precisam ser feitas nos dados do Data Warehouse para deixá-los prontos para a exploração de dados. A seguir é apresentado um conjunto de características para análise dos dados.

- ✓ A exploração de dados é um conjunto de técnicas de análise poderosas para dar sentido a conjuntos de dados muito grandes;

-
- ✓ Não há uma abordagem de exploração de dados, mas antes, um conjunto de técnicas que podem ser freqüentemente utilizadas em combinação umas com as outras para extrair mais *insights* ou idéias dos dados explorados;
 - ✓ Cada ferramenta de exploração de dados pode ser visualizada logicamente como um aplicativo que é um cliente do Data Warehouse;
 - ✓ O objetivo do Data Warehouse é fornecer conjuntos de observação prontos para utilização para a exploração de dados.

O propósito do Data Warehouse é coletar, armazenar e apresentar dados da melhor maneira possível a ferramentas de exploração de dados. Não é seu propósito realizar realmente a exploração de dados. Essa função cabe mais para um aplicativo analítico do que para um banco de dados.

A seguir estaremos descrevendo algumas das formas de acesso aos dados:

A *consulta ad-hoc* é um tipo específico de ferramenta de acesso a dados do DW por usuários finais que solicita ao usuário para formar suas próprias questões por manipulação direta de tabelas relacionais e suas junções[8].

A *análise estatística* é projetada para reduzir uma grande quantidade de dados a uma simples relação ou fórmula, com cálculos de média.

Segundo Harrison[19], *Data Mining* é a exploração e análise, por meios analíticos ou semi-analíticos de grandes quantidades de dados para descobrir modelos e regras significativas. É particularmente útil para problemas de modelagem não-linear com grande número de

variáveis. O conceito de Data Mining além de outros propósitos, é utilizado na detecção de fraude, como as fraudes exemplificadas no item 5.

Ferramentas OLAP provêm análises multidimensionais, com operações dinâmicas com as seguintes características que são necessárias para capacitar a análise das informações: *slice-and-dice* (operações para realizar navegação por meio dos dados na visualização de um cubo), *drill-down* (aumento do nível de detalhamento) e *drill-up* (saindo do nível de detalhamento).

Existem três arquiteturas básicas OLAP: A Multidimensional pura (*MOLAP*), a Relacional pura (*ROLAP*) e uma híbrida (*HOLAP*) que combina as características de MOLAP e ROLAP. Cada arquitetura tem seu conjunto de benefícios e desvantagens apresentadas na *Tabela 5*.

O *ROLAP* (Relational On Line Analytical Processing) fornece uma análise multidimensional de dados armazenados em uma base de dados relacional. Existem algumas maneiras de se fazer este trabalho:

- ✓ fazer todo o processamento dos dados no servidor da base de dados, neste caso o servidor OLAP gera os comandos SQL em múltiplos passos e as tabelas temporárias necessárias para o devido processamento das consultas;
- ✓ executar comandos SQL para recuperar os dados, mas fazer todo o processamento (incluindo junções e agregações) no servidor OLAP;
- ✓ um servidor analítico é responsável por receber os pedidos de consulta e processá-los. Esses pedidos vêm na forma de queries MDX (multidimensional expressions).

No *ROLAP* a sumarização é executada por uma ferramenta externa, os dados são mapeados nos registros das tabelas do modelo

estrela, agregações são pré-calculadas nas tabelas sumário e metadados são armazenados em tabelas relacionais (índices).

MOLAP (Multidimensional On Line Analytical Processing) permite a execução de análises bastante sofisticadas usando como gerenciador de banco de dados, um banco multidimensional[20].

A tecnologia MOLAP não é somente uma ferramenta de visualização, e sim, uma tecnologia complexa de armazenamento e visualização de dados onde devemos nos preocupar com todo o processo de engenharia de banco de dados, tamanho, estruturação, processo de carga, indexação, otimização, etc.

No MOLAP os dados são armazenados em matrizes proprietárias concebidas especificamente para análises multidimensionais e indexados de maneira a prover um ótimo desempenho no acesso a qualquer elemento. Os índices são necessários apenas para o gerenciamento de dados esparsos e cabem na memória principal.

Usando indexação, antecipação da maneira como os dados são acessados e alto grau de agregação dos dados no processamento das agregações, o MOLAP é capaz de apresentar uma melhor performance em relação ao ROLAP no tempo de resposta ao usuário, uma vez que no processo de carga dos dados ele pode pré-calcular diversas informações.

Servidores MOLAP aplicam estratégias de compressão para manipular cubos esparsos (cubos em que a maioria das células não tem valor), fazendo um uso eficiente do '*caching*'. Por outro lado, ele consome muito espaço. Como consequência, o MOLAP é geralmente usado em Data Marts.

HOLAP (Hybrid Multidimensional On Line Analytical Processing) constitui-se de uma tecnologia que procura combinar as melhores características de MOLAP e ROLAP, de modo a apresentar um melhor desempenho e extensiva escalabilidade. A agregação de dados é armazenada no MOLAP, enquanto os dados base são deixados no

banco de dados relacional, com a vantagem em que os dados de base não são duplicados.

Características	Tipos de Software OLAP		
	MOLAP	HOLAP	ROLAP
Tamanho potencial do DW	Pequeno (Cubo de dados somente)	Grande (Cubos de dados e base de dados relacionais)	Grande (Base de dados relacionais)
Performance de consultas iniciais	O mais rápido	Moderado	Sem prognóstico
Tempo de resposta em consultas <i>Drill</i>	Consistente	Relativamente consistente	Inconsistente (depende da consulta)
Capacidade analítica avançada (análise de séries temporais, ranking, exceções estatísticas, etc).	Alto	Alto	Baixo (limitado pelo SQL)
Eficiência no tempo de carga	Muito baixo (Simples carga de dados do cubo)	Médio (Impacto pelos dados do cubo)	Muito alto (Usa paralelismo SQL)

Tabela 5 - Mapa de Comparação das Características da Arquitetura OLAP
 Fonte: Adaptado de Sweiger[21]

Além destas três arquiteturas básicas OLAP comparadas na tabela acima, existem outras duas extensões que apareceram posteriormente: A DOLAP e a WOLAP.

A DOLAP é uma OLAP que se baseia numa arquitetura DESKTOP, ou seja, é uma ferramenta para usuários que possuam uma cópia de sua base multidimensional ou de um subconjunto dela ou que queiram acessar um repositório de dados central. Basicamente acessa os cubos já existentes do banco de dados ou um conjunto de cubos selecionados pelo usuário.

A WOLAP ou WebOLAP, ferramenta OLAP que permite algum nível de acesso via browser Web. As facilidades dessa arquitetura são: a possibilidade de plataformas independentes para dar suporte a usuários distantes e aplicações de groupware.

Após a apresentação das opções de forma de acesso aos dados, podemos definir que para a extração das informações provenientes

deste modelo de Business Intelligence a opção mais adequada é o MOLAP, considerando o volume de transações que são geradas e serão analisadas, e uma vez que o MOLAP apresenta algumas características importantes como:

- ✓ Tempo de acesso muito rápido referente às consultas iniciais;
- ✓ Apresenta um bom desempenho e extensiva escalabilidade.
- ✓ Alta capacidade analítica avançada (análise de séries temporais, exceções estatísticas, ranking, entre outras);
- ✓ Aplica estratégias de compressão para manipular cubos esparsos (cubos em que a maioria das células não tem valor), fazendo um uso eficiente do 'caching';
- ✓ Eficiência alta na carga dos dados (Simples carga de dados do Cubo);

Por fim podemos dizer que usando a modelagem de dados, mecanismos de busca em banco de dados, técnicas analíticas em conjunto com ferramentas OLAP que tenham uma grande quantidade de características, a análise de um Data Warehouse pode melhorar enormemente na performance das consultas de usuários[21].

Estas consultas possibilitam um enorme número de cruzamentos de informações que estarão incrementando o processo de tomada de decisão e direcionamento estratégico das empresas da área de saúde.

A seguir apresentamos uma lista de informações desejadas, atendidas pelo modelo proposto:

- ✓ Quantidade de Consultas Médicas por especialidade, por localização;
 - ✓ Índices de produtividade por profissional, por região ou por serviço;
 - ✓ Índices de Consultas (por internação, por retornos, por encaminhamentos para especialistas, por exames complementares);
 - ✓ Índices de sinistralidade por faixa etária;
-

-
- ✓ Custos médios por atendimento;
 - ✓ Informações de pagamentos aos Prestadores de Serviço;
 - ✓ Relação de Prestadores por especialidade, por região;
 - ✓ Relação de Pacientes crônicos por especialidade, por região;
 - ✓ Entre outros que se façam necessários e sejam possíveis de serem construídos, baseados no modelo proposto.
-

6. CONCLUSÃO

O projeto de desenvolvimento deste trabalho como um todo foi de grande importância para a complementação da formação do aluno.

Dentre vários aspectos, serviu como uma primeira experiência no processo de coleta e ampliação de conhecimento referente ao ambiente analítico. Além disso, permitiu ao aluno desenvolver a escrita de textos acadêmicos e formais também importante na sua formação.

Algumas dificuldades foram encontradas durante o desenvolvimento do material, porém estas dificuldades foram de grande valia impulsionando ainda mais a pesquisa de informações para que o modelo proposto fosse desenvolvido de forma consistente baseado em conceitos dos mais renomados autores relacionados ao tema. Estas dificuldades também estão relacionadas com a pequena quantidade de material disponível com enfoque na área de saúde.

Um ponto importante para salientar, observado no processo de construção deste trabalho, está na congruência existente entre as estratégias utilizadas no Data Warehouse e nos negócios. Concluindo-se assim que a utilização do Data Warehouse, por si só, não garante vantagem competitiva para a empresa, mas sim na sua utilização de maneira eficiente.

Foram discutidas ainda as vantagens e desvantagens das estruturas envolvidas, dando assim uma visão mais ampla ao leitor e até informações que poderão ser utilizadas em implementações baseadas neste estudo.

Por fim, conclui-se, com base nos textos estudados até aqui, que a construção de um ambiente de Business Intelligence na área de saúde amplia consideravelmente os novos modelos de consultas e conseqüentemente as informações geradas, ajudando as empresas de medicina supletiva tornarem-se mais competitivas.

7. TRABALHOS FUTUROS

São sugestões para dar continuidade a este trabalho:

- ✓ Desenvolver e complementar o Data Warehouse com os outros Data Marts não detalhados no *item 5.4*;
 - ✓ Desenvolver um estudo contábil do valor da implementação versus benefícios do Data Warehouse em médio e longo prazo;
 - ✓ O estudo do impacto que a implementação do Data Warehouse pode causar nas dimensões Gerenciais, Tecnológicas e Humana;
 - ✓ Implementar um estudo de utilização das técnicas de mineração de dados com a finalidade de ampliar a gama de informações estratégicas;
-

REFERÊNCIAS

- [1] LEITE, Edson. *Tivit reforça área de saúde com aquisição da Solutia*, <http://www.decisionreport.com.br/publique/cgi/cgilua.exe/sys/start.htm?infoid=266&sid=14>, acesso em 06/12/2006.
 - [2] MACHADO, Felipe Nery Rodrigues. : *Tecnologia e Projeto de Data Warehouse: uma visão multidimensional* – 1. ed. – 2004
 - [3] INMON, Willian H.: *Building the Data Warehouse* 4th Edition – 2005
 - [4] KIMBALL, Ralph. : *The data warehouse toolkit: the complete guide to dimensional modeling* / Ralph Kimball, Margy Ross. - 2nd ed - 2002.
 - [5] HARJINDER, G. e Rao, P. C. – *The Official Guide to Data Warehousing*, Que Corporation, 1996.
 - [6] SELL, D. - *Uma Arquitetura para Distribuição de Componentes Tecnológicos de Sistemas de Informações Baseados em Data Warehouse.*, Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina, 2001.
 - [7] KIMBALL, Ralph; MERZ, Richard. *Data Webhouse: Construindo o Data Warehouse para a Web*. Tradução: Edson Furmankiewicz, Joana Figueiredo. 1. ed. Rio de Janeiro: Campus, 2000.
 - [8] KIMBALL, R., Reeves, L., Ross, M., Thornthwaite, W. *The Data Warehouse Lifecycle Toolkit*. New York: John Wiley & Sons, Inc., 1a. ed. -1998.
 - [9] COME, G. *Contribuição ao Estudo da Implementação de Data Warehousing: Um Caso no Setor de Telecomunicações*. Tese (Mestrado), USP. São Paulo 2001.
 - [10] DW Brasil, *Características do Data Warehouse – Por Time DWBrasil*, http://www.dwbrasil.com.br/html/artdw_carac.html, acesso em 06/11/2006.
 - [11] PEREIRA, Denise Maciel. *Uso do Padrão OIM de Metadados no Suporte às Transformações de Dados em Ambiente de Data Warehouse*. 2000. Mestrado – Universidade Federal do Rio de Janeiro – UFRJ, Rio de Janeiro, 2000.
 - [12] INMOM, W.H.: Hackathorn, Richard D. *Como Usar o Data Warehouse*. Rio de Janeiro: Editora Infobook, 1997.
-

-
- [13] OTTA, Hakaru - O impacto da utilização do cartão magnético na cooperativa de trabalho médico: Um estudo de caso, Dissertação - Universidade Federal de Santa Catarina – 2003
- [14] MEZOMO, João Catarin. *Gestão da Qualidade na Saúde*. Barueri – São Paulo: Ed. Manole, 2001.
- [15] ISOLDI, Sergio, *Utilização de tecnologia de informação nas organizações públicas*. Campinas: PUC-Campinas, 2000.
- [16] LOVELOCK, Cristopher. WRIGHT, Lauren. *Serviços: Marketing e Gestão*. São Paulo: Saraiva, 2001.
- [17] LEÃO, Beatriz de Faria. *O prontuário eletrônico: como chegar lá?* Informática Brasileira em Análise. Ano II. n.16 - 2002. Disponível em www.cesar.org.br
- [18] GOODYEAR, Mark, RYAN, Hugh W., SARGENT, Scott R., et al. *Netcentric and Client/Server Computing – a practical guide*. Andersen Consulting. Boca Raton: Auerbach, 1999.
- [19] HARRISON, Thomas H., *Intranet data warehouse*. São Paulo: Berkeley, 1998.
- [20] FIGUEIREDO, Adriana Maria C.M. *MOLAP x ROLAP: Embate de Tecnologias para Data warehouse*. Developer's Magazine. Rio de Janeiro: Axcel Books do Brasil Editora Ltda. 1998, nº 18 p.24
- [21] SWEIGER, Mark et al. *Clickstream Data Warehousing*. New York: J. Wiley & Sons Inc, 2002.
-