

CENTRO UNIVERSITÁRIO FEEVALE

CLÉBER ROBERTO FERREIRA DE MELLO

APLICAÇÃO DE FERRAMENTAS OLAP EM UM *DATA MART*
DE UM CLUBE DESPORTIVO

Novo Hamburgo, julho de 2008.

CLÉBER ROBERTO FERREIRA DE MELLO

APLICAÇÃO DE FERRAMENTAS OLAP EM UM *DATA MART*
DE UM CLUBE DESPORTIVO

Centro Universitário Feevale
Instituto de Ciências Exatas e Tecnológicas
Curso de Ciência da Computação
Trabalho de Conclusão de Curso

Professor Orientador: Ms. Juliano Varella de Carvalho

Novo Hamburgo, julho de 2008.

AGRADECIMENTOS

Em primeiro lugar a Deus, pelo dom da vida e a oportunidade que me oferece.

Agradecimentos ao professor orientador, Ms. Juliano Varella de Carvalho que, na rigidez de seus ensinamentos, fez aprimorar meus conhecimentos.

A empresa Birô Processamento de dados, diretores e colegas que, em vários momentos, estiveram presentes no meu crescimento profissional.

A minha esposa Larissa Nericke, pela paciência de cada momento de ausência para que este trabalho pudesse ser realizado.

Aos meus familiares e amigos, pelo companheirismo e os muitos momentos difíceis e de alegrias compartilhados.

RESUMO

A concorrência e a alta competitividade ocasionadas pela globalização motivam as corporações a dedicarem esforços a fim de conhecerem melhor o seu negócio, seus clientes, seus produtos e o mercado. Os sistemas corporativos desenvolvidos para a utilização diária (OLTP – *Online transaction processor*) visam principalmente automatizar processos. Eles atendem às necessidades operacionais, mas são pobres em responder às necessidades informacionais e analíticas. Conceitos e ferramentas de DW (*Data Warehouse*), DM (*Data Mart*) e OLAP (*Online Analytical Processing*), BI (*Business Intelligence*) são recursos relevantes para desenvolver um ambiente que possibilite fazer análises, definindo agregações e cruzamentos, e que permitem visualizar os dados através de múltiplos níveis de hierarquias e de diferentes perspectivas. Este trabalho tem por finalidade apresentar uma solução de apoio à tomada de decisão gerencial (SAD), utilizando ferramentas proprietárias. Será apresentada uma proposta de utilização de ferramentas OLAP em um *Data Mart*, a fim de auxiliar os departamentos de *marketing* e estratégico de um clube desportivo.

Palavras-chave: *Data Warehouse*. *Data Mart*. OLAP. *Business Intelligence*. Administração de clube desportivo.

ABSTRACT

The competition and the high competitiveness caused by the globalization motivate the corporations to dedicate efforts to better know their business, their customers, their products and the market. The corporate systems developed for daily use (OLTP – Online transaction processor) aim for mainly automate process. They meet the operational needs, but they are poor in meeting the informational and analytical needs. The Concepts and tools for DW (Data Warehouse), DM (Data Mart) and OLAP (Online Analytical Processing), BI (Business Intelligence) are relevant resources to develop an environment that enables analysis done by defining aggregations and junctions, and that allows you to see the data through multiple levels of hierarchies and different perspectives. This work aims to present a solution to support of the decision support system (DSS), using proprietary tools. It will be presented a proposal for use of OLAP tools in a Data Mart, to help the departments of marketing and strategy of a sports club.

Key words: Data Warehouse. Data Mart. OLAP. Business Intelligence. Administration of sports club.

LISTA DE FIGURAS

Figura 1.1 - Níveis de uso dos dados.....	16
Figura 1.2 - Ambiente de <i>Data Warehouse</i>	18
Figura 1.3 - Os conjuntos de dados que pertencem à mesma área de interesse são relacionadas através de uma chave comum entre as tabelas	19
Figura 1.4 - A questão da não volatilidade.....	20
Figura 1.5 - Níveis de granularidade em uma movimentação bancária	21
Figura 1.6 – Processo de ETC	22
Figura 2.1 – Exemplo de tabela de fatos	29
Figura 2.2 – Exemplo de tabela de dimensões	31
Figura 2.3 – Exemplo da modelagem <i>Star Schema</i>	32
Figura 2.6 – Exemplo de um cubo.....	33
Figura 2.7 – Arquitetura MOLAP	37
Figura 2.8 – Arquitetura ROLAP	38
Figura 2.9 – Arquitetura HOLAP	39
Figura 2.10 – Arquitetura DOLAP	40
Figura 3.1 – Diagrama SLQ <i>Server</i> com questões da avaliação	42
Figura 3.2 – Diagrama SLQ <i>Server</i> com repostas da avaliação	43
Figura 3.3 – Representação do banco criado para BI.....	44
Figura 3.4 – Representação do cubo.....	48
Figura 3.5 – Diagrama MS <i>Access</i>	50
Figura 3.6 – Modelagem do cubo Desempenho	51
Figura 3.7 – Cubo Desempenho	53
Figura 3.8 – Cubo Eficiência.....	54
Figura 4.1 – Cubo Clube desportivo.....	65
Figura 5.1 – <i>Management Studio</i> – Recursos visuais facilitam a administração.	68

Figura 5.2 – <i>Business Intelligence Development Studio</i>	70
Figura 5.3 – Definição do <i>Data Source</i> e <i>Data Source View</i>	71
Figura 5.4 – Cubo Acesso.	72
Figura 5.5 – Cubo Torcedor.	72
Figura 5.6 – Medidas criadas no cubo Acesso.	73
Figura 5.7 – Medidas criadas no cubo Torcedor.	74
Figura 5.7 – Medidas criadas na tabela dimensão Tempo.....	75
Figura 5.8 – Medidas criadas no DM ClubeCampeao.	75
Figura 5.9 – Construção de uma consulta no cubo Torcedor.	76
Figura 5.10 – Sócios ativos entre os anos de 2005 e 2006.	77
Figura 5.11 – Faixa etária dos sócios entre o ano de 2004 e 2006.	77
Figura 5.12 – Percentual de sócios e cônsoles por região.	78
Figura 5.13 – Percentual de acessos aos eventos por sexo.....	78
Figura 5.13 – Percentual de acessos aos eventos em 2004 por região.	79
Figura 5.14 – Percentual de acessos aos eventos em 2004 por UF.	79
Figura 5.15 – Quantidade de acessos por evento em 2004.....	80
Figura 5.16 – Quantidade de acesso por faixa de hora.....	80

LISTA DE TABELAS

Tabela 1.1 - Diferenças-chave entre <i>Data Marts</i> e <i>Data Warehouse</i>	25
Tabela 3.1 - Relacionamento entre professores e alunos.....	45
Tabela 3.2 - Professores que esclarecem as dúvidas de alunos em dificuldades.....	45
Tabela 3.3 - Distribuição de clientes segundo índice de massa corporal (IMC) e sexo (M ou F), desconsiderando os indivíduos que não informaram o sexo e/ou não existiu este dado. ...	49

LISTA DE ABREVIATURAS E SIGLAS

BI	<i>Business Intelligence</i>
DM	<i>Data Mart</i>
DOLAP	<i>Desktop On-Line Analytical Processing</i>
DSS	<i>Decision Support System</i>
DTS	<i>Data Transformation Services</i>
DW	<i>Data Warehouse</i>
DWA	<i>Data Warehouse Administrator</i>
EIS	<i>Execute Information System</i>
ETC	Extração, Transformação e Carga de Dados
HOLAP	<i>Hybrid On-Line Analytical Processing</i>
MOLAP	<i>Multidimensional On-Line Analytical Processing</i>
MS	<i>Microsoft</i>
OLAP	<i>Online Analytical Processing</i>
OLTP	<i>Online transaction processor</i>
ROLAP	<i>Relational On-Line Analytical Processing</i>
SGBDR	Sistema de Gerenciamento de Banco de Dados Relacional
SQL	<i>Structured Query Language</i>
UDM	<i>Unified Dimensional Model</i>
URL	<i>Universal Resource Locator</i>

SUMÁRIO

INTRODUÇÃO	12
1 BUSINESS INTELIGENCE E DATA WAREHOUSE	15
1.1 <i>Business intelligence</i>	15
1.2 <i>Data Warehouse</i>	17
1.2.1 Funcionamento de um DW	19
1.2.2 Granularidade	20
1.3 ETC - Extração Transformação e Carga de Dados	22
1.4 <i>Data Mart</i>	24
1.5 <i>Data Warehouse</i> versus <i>Data Mart</i>	24
2 MODELAGEM DIMENSIONAL	27
2.1 Tabela de Fatos	28
2.2 Tabela de Dimensões.....	30
2.3 Técnicas de Modelagem	31
2.3.1 <i>Star Schema</i>	31
2.4 Cubos	32
2.5 OLAP.....	33
2.5.1 Tipos de operações OLAP	35
2.5.2 Arquiteturas OLAP	36
3 TRABALHOS RELACIONADOS	41
3.1 Centro Universitário	41
3.1.1 Etapas do Projeto	43
3.1.2 Considerações do autor	45
3.2 Detecção de fatores de risco para doença arterial coronariana.....	46
3.2.1 Etapas do Projeto	46
3.2.2 Considerações do autor	49
3.3 Proposta de aplicação de BI em chão de fábrica	49
3.3.1 Etapas do projeto	50
3.3.2 Considerações do Autor.....	54
4 CLUBE DESPORTIVO.....	55
4.1 Controle e atendimento ao associado	56
4.2 Acesso ao estádio	57
4.3 Acesso ao estacionamento	58
4.4 Tabelas de fatos	59
4.4.1 Fato TORCEDOR.....	59
4.4.2 Fato ACESSO	60

4.5	Tabelas de dimensões	61
4.5.1	Dimensão TEMPO.....	61
4.5.2	Dimensão SSOCIAL	61
4.5.3	Dimensão SEXO.....	62
4.5.4	Dimensão PORTAO	62
4.5.5	Dimensão MUNICIPIO	62
4.5.6	Dimensão FAIXAHORA.....	63
4.5.7	Dimensão FAIXAETARIA	63
4.5.8	Dimensão COBRADOR.....	63
4.5.9	Dimensão CATEGORIA	64
4.5.10	Dimensão EVENTO	64
4.6	Modelagem dos dados (cubos)	64
4.7	ETC – Extração, Transformação e Carga de Dados.....	65
5	APLICANDO A FERRAMENTA OLAP DO MICROSOFT ANALYSIS SERVICES.	67
5.1	<i>Analysis services</i>	68
5.2	Plataforma e configuração do ambiente	69
5.3	Criando os cubos	71
5.3.1	Criando as medidas.....	73
5.3.2	Criando as hierarquias	74
5.4	Verificação dos resultados.....	76
	CONCLUSÃO.....	81
	Melhorias e trabalhos futuros.	82
	REFERÊNCIAS BIBLIOGRÁFICAS	84
	ANEXO I – SCRIPTS VISUAL FOX PRO PARA FAZER A EXTRAÇÃO DOS DADOS	87

INTRODUÇÃO

Ambientes corporativos apresentam um grande salto tecnológico quando o gerente de tecnologia resolve informatizar seus departamentos. Entretanto, à medida que as empresas estendem a informatização em seu negócio, cria-se um aglomerado de dados em seus servidores.

O avanço das tecnologias de informação e o aumento da capacidade de armazenamento e de processamento dos sistemas computacionais, tem se acentuado significativamente. Com efeito, cada vez mais empresas armazenam em suas bases de dados informações que possibilitam criar interações computacionais com seus clientes e fornecedores.

A necessidade de obter informações interligadas entre os sistemas transacionais corporativos, de tal forma que suas informações sejam concisas e sumarizadas, tem sido relevante do ponto de vista competitivo.

Diante deste contexto, a necessidade de administrar estes bancos de maneira integrada e concisa, estimula a busca por alternativas que auxiliem no processo de tomada de decisões. Os conceitos de *Data Warehouse*, *Data Mart* e ferramentas OLAP, que são os focos principais deste trabalho, formam um conjunto de tecnologias orientadas a disponibilizar informação e conhecimento em uma empresa.

De acordo com Inmon, Welch e Glasey (1999), A introdução dos modernos *Data Warehouses* e suas ferramentas de suporte à decisão, alterará a mundo do acesso a informações corporativas tão profundamente quanto o telefone de discagem rotatória alterou a comunicação. E finalmente o poder da informação estará nas mãos do usuário.

Segundo Inmon (1997, p. 33), define-se: “Um *Data Warehouse* é um conjunto de dados baseado em assuntos, integrado, não-volátil, e variável em relação ao tempo, de apoio às decisões gerenciais”.

Pode-se considerar um *Data Mart* como uma base de dados específica de um departamento que compõe o DW. Conforme Machado (2000, p.27), *Data Mart* “representa um subconjunto de dados do *Data Warehouse*. Os dados do *Data Mart* são direcionados a um departamento ou a uma área específica do negócio.”

Dentre estas ferramentas de análise de dados multidimensionais, destaca-se a tecnologia denominada OLAP que traduzindo do termo inglês, significa Processamento Analítico *On-line*. Conforme definição do Conselho OLAP, Inmon; Welch e Glassey (1999, p.175), a tecnologia define-se como:

Categoria da tecnologia de *software* que permite que analistas, gerentes e executivos obtenham, de maneira rápida, consistente e interativa, acesso a uma variedade de visualizações possíveis de informações que foi transformada de dados puros para refletir a dimensão real do empreendimento do ponto de vista do usuário.

Pode-se ainda definir como um conjunto de processos que possibilitam manipular e gerenciar dados multidimensionais possibilitando ao usuário efetuar visualizações e análises em busca de uma maior compreensão destes fatos.

Considerando o contexto apresentado, este trabalho tem por finalidade desenvolver um SAD em um ambiente de DM, utilizando ferramentas OLAP proprietárias em um clube desportivo. Desta forma, espera-se aplicar estas ferramentas para apoio de tomada de decisão nos departamentos de *Marketing* e estratégico do clube, agregando valor à organização.

Atualmente o clube possui vários pacotes de aplicativos que compõem o conjunto de ferramentas que gerenciam toda a infra-estrutura desta entidade. Dentre os pacotes, destacam-se:

- Controle de acesso de veículos,
- Controle de acesso ao estádio;
- Administração social – Society Software para Clubes,
- Site corporativo.

O Pacote de administração social (Society) é responsável por manter e controlar as regras de negócio do clube. No banco de dados deste aplicativo, concentram-se todos os dados cadastrais do sócio, tais como endereços, dados pessoais, dados de cobrança, acessos aos eventos, como também as mensalidades pagas ou aquelas ainda pendentes. Realizam-se também as manutenções diárias de compra de serviços (cadeiras, esportes, escolinhas...) e admissão, afastamento ou demissão do quadro social.

O sistema de controle de acesso de torcedores às dependências do clube é realizado pelo sistema de controle de acesso ao estádio. Este sistema auxiliar não está sincronizado com o sistema de administração social (Society). A lista dos sócios torcedores com acesso permitido é atualizada através de arquivos de intercâmbio que o sistema de administração social gera e envia para este sistema. Após a finalização de cada evento, realiza-se novamente um intercâmbio de dados a fim de enviar ao sistema principal os acessos que ocorreram.

Diante desta realidade, o clube desportivo encontra dificuldades em conhecer aspectos importantes no seu negócio. Os perfis de seus sócios, tais como preferências de campeonatos, datas dos jogos em que ocorrem maiores procuras, a utilização de estacionamento, a procedência das caravanas, entre outros, tem-se apresentado como fatores críticos dentre a atual demanda. Algumas destas informações não estão dispostas no repositório de dados do sistema de sócios (Society). O sistema de sócios não dispõe de informações gerenciais integradas entre as diferentes bases de dados.

Pretende-se com este trabalho, desenvolver uma aplicação que possibilite a obtenção de informações gerenciais que auxilie a obtenção de conhecimento de perfis e padrões, possibilitando promover ações que estimulem o crescimento e elevem a satisfação do associado.

Esse trabalho é composto de cinco capítulos. No primeiro e segundo capítulos, são abordados os conceitos de *Data Warehouse*, OLAP e banco de dados multidimensional. No terceiro capítulo são descritas sínteses de trabalhos relacionados e as experiências adquiridas de cada autor. No quarto capítulo é feito o estudo de caso e a modelagem do DM e no quinto capítulo é apresentada a solução proposta.

1 *BUSINESS INTELLIGENCE E DATA WAREHOUSE*

O objetivo deste capítulo é caracterizar em linhas gerais o conceito de *Business Intelligence*, como também oferecer explicações tanto sobre *Data Warehouse*, *Data Mart* e OLAP, além das suas contribuições como ferramentas de apoio a tomada de decisões. São apresentados os principais aspectos que caracterizam um DW como: funcionamento, agregação de dados e granularidade. Por fim, pretende-se fazer uma breve verificação das diferenças entre um DW e um DM.

1.1 *Business intelligence*

Empresas que buscam um diferencial competitivo necessitam de recursos que apoiem nas suas tomadas de decisões. Faz-se necessário contar com alguma sistemática que possibilite aos responsáveis por tomada de decisões (os *Decison Makers*), respostas confiáveis, integradas e imediatas.

Isto ocorre porque os sistemas focados em OLTP, disponíveis aos usuários, não fornecem as informações necessárias para este tipo de operação. As tecnologias vigentes estão concentradas em extrair informações altamente pontuais e detalhadas, geralmente destinada ao controle e armazenamento.

Verificamos que nas grandes corporações, existem vários sistemas OLTP que formam as diversas bases de dados existentes em seus servidores. Estes sistemas garantem o controle o armazenamento diário das informações pertinentes as suas regras de negócio.

A Microsoft (2007a), afirma que dentro da organização, podem ser vistos diferentes níveis de uso dos dados.

Nível operacional: São utilizados sistemas de informação que monitoram as atividades e transações elementares da organização.

Nível de administração: São realizadas tarefas de administradores de nível intermediário apoiando as atividades de análise, de acompanhamento, de controle e tomada de decisões, realizando consultas sobre informações armazenadas no sistema, proporcionando relatórios e facilitando a administração da informação por parte dos níveis intermediários.

Nível de conhecimento: Neste nível encontramos os trabalhadores de conhecimento e de dados. Entre suas atribuições, destacam-se as análises e acompanhamentos comportamentais de seus produtos ou serviços como também as tomadas de decisões.

Nível estratégico: Seu objetivo é realizar as atividades de planejamento em longo prazo, tanto do nível de administração quanto dos objetivos apresentados pela empresa.

Observamos mais claramente a composição destes diferentes níveis através da figura 1.1 .

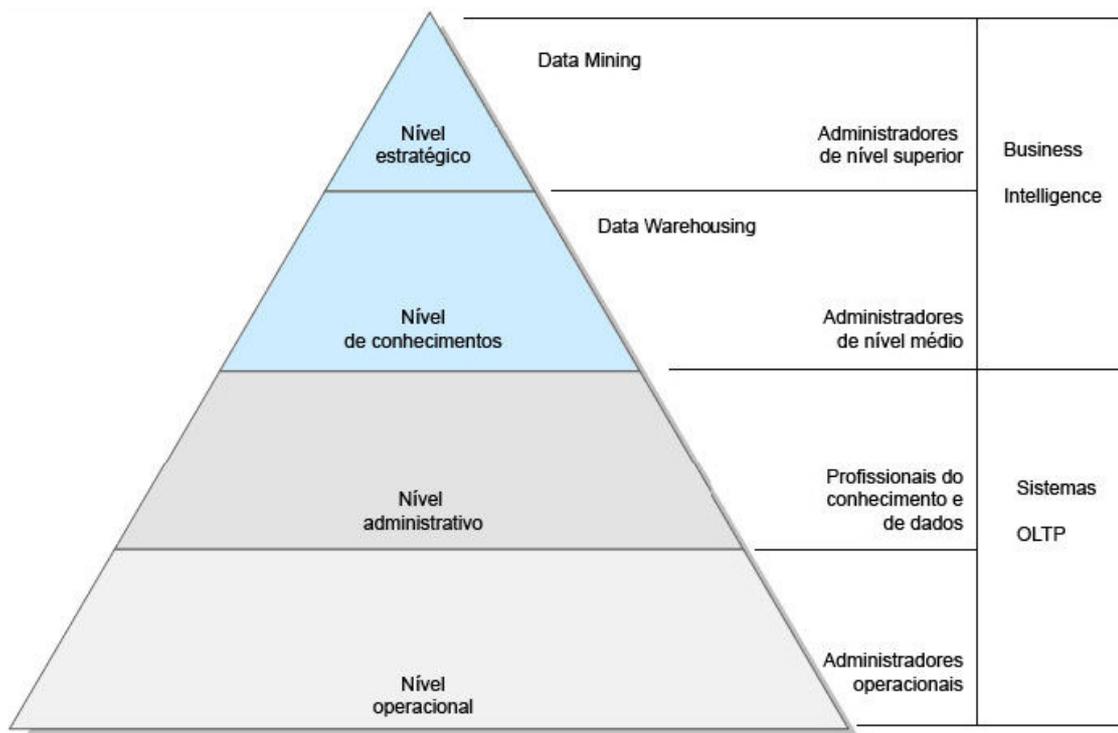


Figura 1.1 - Níveis de uso dos dados
(Fonte: MICROSOFT, 2007a).

Os SAD tem o objetivo de contribuir para os executivos que definem as ações estratégicas, consultar, interpretar e até prever o comportamento do negócio conforme o seu grau de interesse. Os variados sistemas orientados a tomada de decisões são definidos pelo

termo *Business Intelligence* que traduzindo para o português, significa inteligência de negócios (MICROSOFT, 2007a). Percebe-se atualmente que cada vez mais corporações investem em BI, todas possuem como objetivo principal, conhecer mais o cliente e o seu mercado.

Machado (2000) afirma que para realizar BI, deve-se basicamente abranger os seguintes tópicos:

- Recuperar os dados históricos da empresa. Isto significa buscar cópias das bases de dados superiores, há dois anos inclusive;
- Descobrir as necessidades de informação, indicadores de negócio da empresa;
- Projetar banco de dados para armazenar estas informações;
- Disponibilizar os dados históricos para um processo de extração, limpeza, transformação e carga em um *Data Warehouse*;
- Executar estes processos,
- Utilizar uma ferramenta de EIS (*Execute Information System*) ou DSS (*Decision Support System*).

As principais disciplinas e técnicas que oferecem suporte à decisão são *Data Warehouse*, *Data Mart*, *Data Mining* e *OLAP*. Será apresentado no próximo tópico o conceito de *Data Warehouse*.

1.2 *Data Warehouse*

A necessidade de descobrir informações de forte cunho informativo e não somente operacional, têm sido perseguida há bastante tempo e tomou força nos últimos anos. O conceito de *Data Warehouse* se popularizou a partir da década de 80. Conhecido inicialmente como processamento de “extração”, os primeiros processos de BI começaram a surgir e logo tornaram-se muito populares (INMON, 1997). O mesmo autor destaca duas grandes razões que causaram sua popularidade:

- Possibilidade de retirar dados do caminho do processamento *on-line* de alto desempenho com um programa de extração, não havendo conflito em termos de desempenho quando os dados precisam ser analisados coletivamente.
- Quando extraídos os dados do repositório de dados *on-line* utilizando um programa de extração, ocorre uma mudança no controle dos dados. O usuário final acaba “tendo posse” dos dados quando ele ou ela assume o controle dos mesmos.

Inmon (1997) descreve *Data Warehouse* como uma coleção de dados orientada por assunto, integrada, variante no tempo e não volátil que tem por objetivo dar suporte aos processos de tomada de decisão.

Machado (2000, p. 26) complementa a definição, afirmando que DW:

[...] representa uma grande base de dados capaz de integrar, de forma concisa e confiável, as informações de interesse para a empresa, que se encontram espalhadas pelos sistemas operacionais e em fontes externas, para posterior utilização nos Sistemas de Apoio à Decisão.

Pode-se compreender, portanto, como DW sendo um armazém contendo dados históricos sobre as operações da empresa, extraídos de fontes únicas ou múltiplas. Sua finalidade é oferecer uma base de dados unificada para possibilitar a aplicação de recursos, processos e ferramentas que permita um suporte efetivo à tomada de decisões. A Figura 1.2 ajuda a compreender esta estrutura.

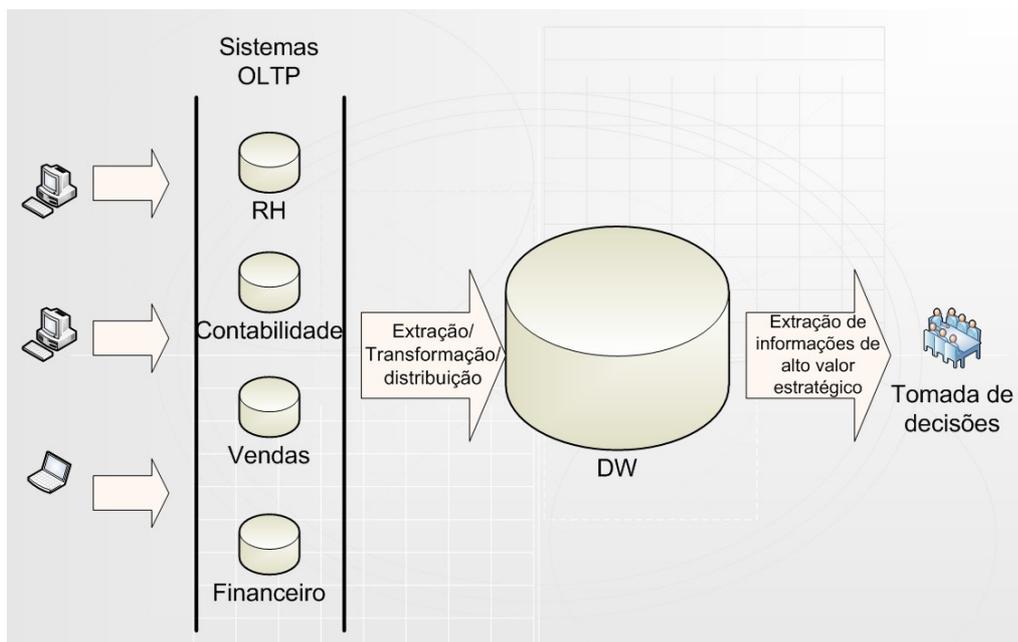


Figura 1.2 - Ambiente de *Data Warehouse*

1.2.1 Funcionamento de um DW

Um *Data Warehouse* baseia-se nos principais assuntos de interesse da empresa. Ele pode estar organizado conforme o modelo de dados projetado anteriormente focado de acordo com os resultados que se deseja buscar. A principal área de interesse termina sendo fisicamente implementada como uma série de tabelas relacionadas entre si, inseridas em um *Data Warehouse* (INMON, 1997).

Como já foi dito, estas tabelas são relacionadas entre si através de chaves comuns localizadas entre as tabelas. A figura 1.3 mostra que a chave chamada *cod_Cliente* conecta todos os dados encontrados na área de interesse cliente.



Figura 1.3 - Os conjuntos de dados que pertencem à mesma área de interesse são relacionadas através de uma chave comum entre as tabelas

Fonte: Adaptação do autor segundo (Inmon, 1997)

Diferentemente de sistemas OLTP, que armazenam dados e que sofrem continuamente modificações, um DM é não-volátil. Ou seja, não permite que sejam modificados os registros que já constam em seu repositório. A figura 1.4 ilustra esta característica.

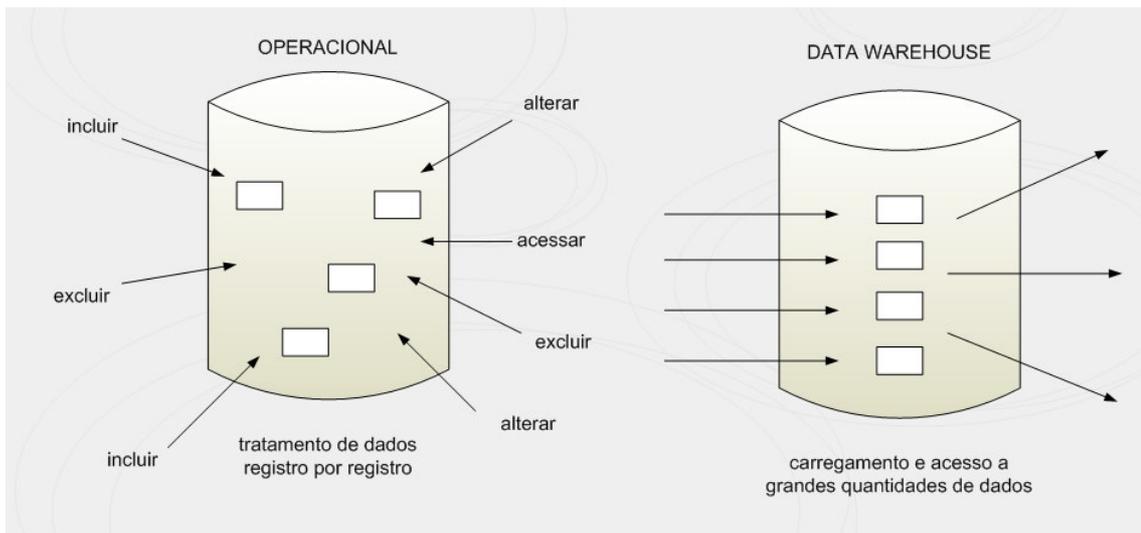


Figura 1.4 - A questão da não volatilidade
 Fonte: Adaptação do autor segundo (Inmon, 1997)

1.2.2 Granularidade

Um conceito muito importante que se deve observar ao fazer o modelo refere-se à granularidade. De acordo com Inmon (1997) a granularidade corresponde ao nível de detalhe ou do resumo que os dados serão sumarizados nas unidades de dados do DW. Quanto mais detalhe deseja-se armazenar, mais baixo será o nível de granularidade.

Machado (2000) destaca a granularidade como um dos fatores mais importantes da modelagem física dos dados, independente da implementação e arquitetura a serem utilizados. O nível de granularidade bem definido é fator principal de sucesso.

O motivo pela qual a granularidade é a principal questão de projeto consiste no fato de que ela afeta profundamente o volume de dados contidos na *data warehouse*. Este fato reflete diretamente ao tipo de consulta que pode ser executada. O volume de dados contidos do *data warehouse* é balanceado de acordo com o nível de detalhe de uma consulta. (INMON, 1997).

De acordo com a Microsoft (2007b), ao definir a granularidade está sendo decidido ao mesmo tempo:

- As análises que são de interesse.
- O grau de detalhe necessário.

Como exemplo de níveis de granularidade, cita-se um resumo de movimentações bancárias de um determinado cliente em um mês. Para obter um nível de detalhe maior, ou seja, os detalhes de cada movimentação, deve-se diminuir o nível de granularidade. A figura 1.5 ajuda a compreender este exemplo.

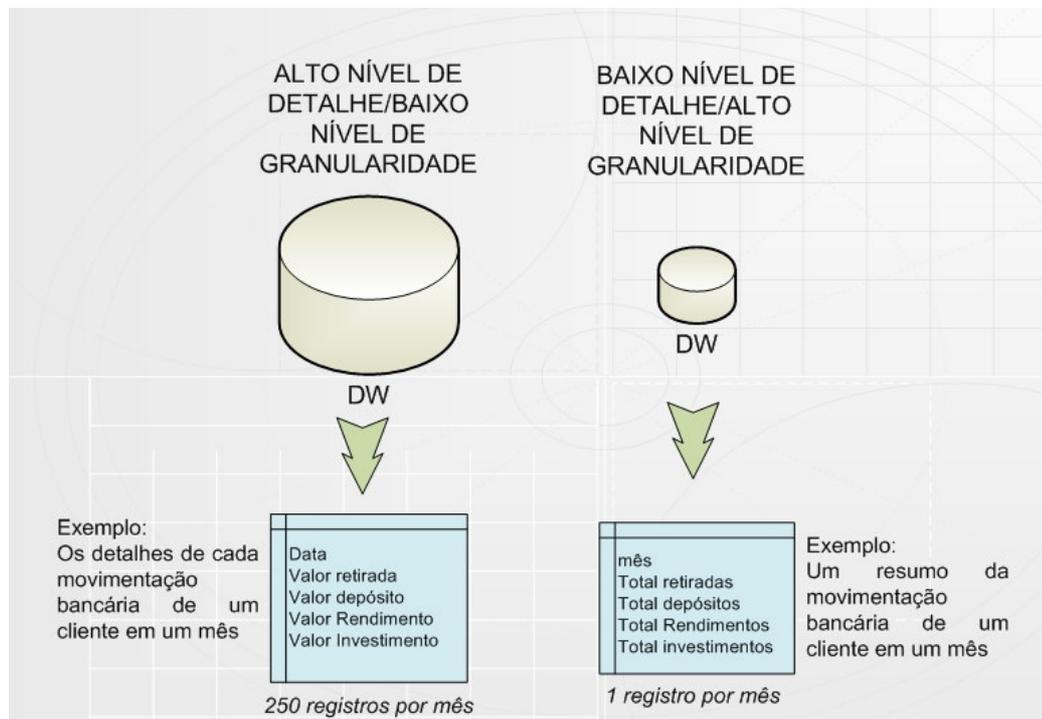


Figura 1.5 - Níveis de granularidade em uma movimentação bancária

Fonte: Adaptação do autor segundo (MACHADO, 2000)

Verifica-se no lado esquerdo da figura o nível de granularidade baixo. Com este modelo, é possível saber quais foram os dias que ocorreram as transações e quais foram os valores de cada depósito ou retirada.

Cabe ressaltar que em um ambiente de DW, dificilmente deseja-se examinar um dado isolado. Consultas como esta, alocariam uma quantidade grande de recursos para varrer uma grande quantidade de registros. Entretanto, com níveis muito elevados de granularidade, lado direito da figura, haverá uma correspondente diminuição da possibilidade da utilização dos dados para atender as consultas.

De acordo com Machado (2000), é necessário avaliar qual a real vantagem estratégica em manter um nível baixo de granularidade no DW. Além do bom senso no momento da modelagem dos dados, faz-se necessário o envolvimento de desenvolvedores e usuários, realizando discussões e análises a fim de estabelecer uma correta granularidade de projeto.

1.3 ETC - Extração Transformação e Carga de Dados

De acordo com Machado (2000) antes dos dados passarem do ambiente operacional para o DW, faz-se necessário uma validação e limpeza dos dados. O local intermediário onde ocorre este processo chama-se *Data Staging Área*.

De acordo com a Microsoft (2007b), os diferentes processos concentrados no conceito de extração, transformação e carga de dados em um *Data Warehouse* denomina-se ETC (Extração - Transformação - Carga de Dados) , em inglês, ETL (*Extract – Transform – Load*).

É comum encontrar em organizações sistemas OLTP que tenham sido desenvolvidos por equipes diferentes de programadores ou *softwarehouses* especializados, e que no seu desenvolvimento, tenham adotado diferentes convenções na codificação de variáveis, nomes dos atributos das tabelas, diferentes tipos de dados ou formatos de datas.

Segundo Nardi e Chiba (2007), o processo de ETC talvez seja o componente de BI que envolva mais riscos em um projeto devido principalmente as diversas fontes heterogêneas e ao grande volume de dados a serem tratados. A figura 1.6 ilustra esta característica.

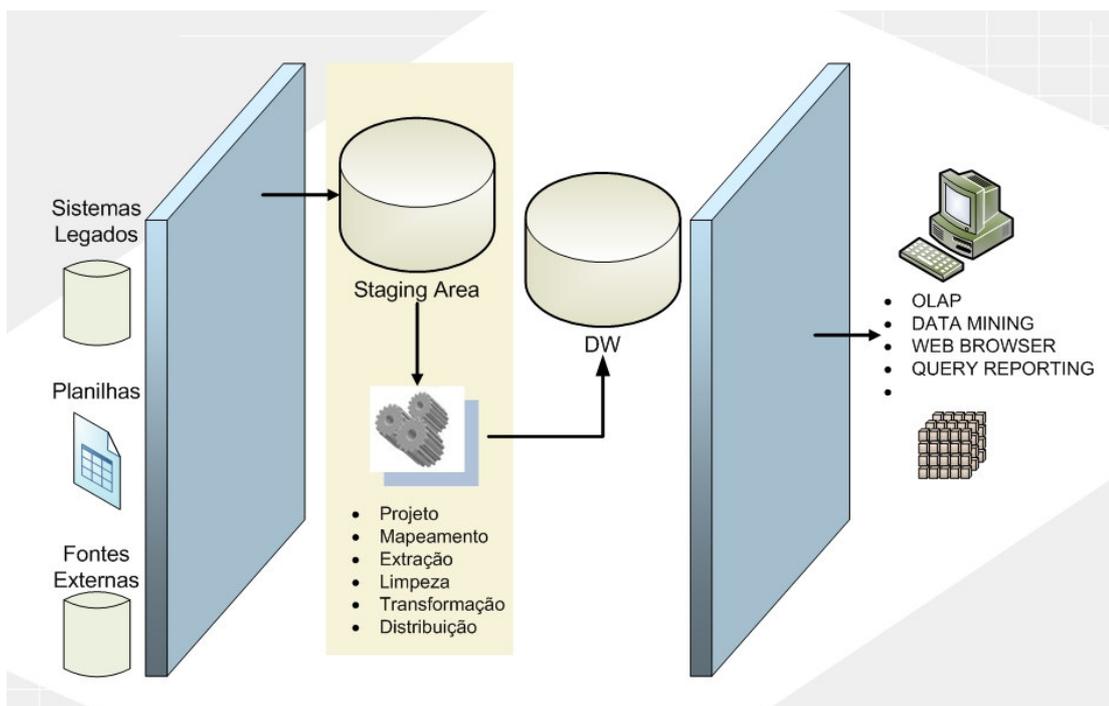


Figura 1.6 – Processo de ETC

Fonte: Adaptação do autor segundo (MACHADO, 2000)

De fato, os dados que alimentam o DW são provenientes de diferentes sistemas transacionais sendo que seus códigos e informações não são integrados. Um claro exemplo ocorre com a representação e descrição do sexo do indivíduo. No sistema transacional A, o sexo masculino e feminino estão identificados, respectivamente como “M” e “F”. No sistema transacional B, os sexos masculino e feminino estão identificados como “1” e “2”, delatando uma incompatibilidade entre as fontes.

No exemplo citado acima, ao aplicar os conceitos de ETC, deverá ser escolhida durante a transformação dos dados, uma convenção única para o DW, que pode ser “M” e “F” e transformar os dados originais em uma tabela de destino padronizada.

De acordo com Cielo ([s.d.]), o processo está dividido em 3 partes distintas: extração, transformação e carga de dados. Suas principais características são:

Extração: consiste na definição das fontes de dados a fazer a extração. As origens destas fontes podem ser várias e também em diferentes formatos, onde se pode encontrar desde os sistemas transacionais das empresas até planilhas, arquivos textos, dados que vêm do SGBD e também arquivos do tipo DBF¹, do antigo Clipper ou Dbase.

Transformação: após a definição das fontes de dados, executa-se o segundo passo, que consiste em transformar e limpar os dados. Esta etapa incide basicamente em limpar e normalizar as redundâncias e heterogeneidades existentes nos diversos sistemas que as tabelas se originam. A limpeza é necessária para corrigir a correta integridade dos dados, tratando dos códigos de caracteres especiais, valores duplicados ou errados. Além da limpeza, a maioria das vezes é necessário realizar o processo de transformação. Frequentemente ocorrem casos que campos que contém a mesma informação, possuem formatação diferente em seus sistemas originais. Como exemplo, citamos o caso do campo data, de algum do sistema transacional, que pode ser do tipo dia, mês e ano (DD/MM/AAAA), e pode estar em outro sistema com tipo ano e mês (AAAA/MM).

Carga de dados: a carga de dados ocorre após o tratamento de extração e transformação dos dados. É a última etapa e seu nível de complexidade é enorme. Esta etapa consiste em carregar os dados tratados e armazenados no *Staging Área* e migrá-los para o

¹ DBF – Formato de arquivo destinado ao armazenamento de dados estruturadamente, utilizado pela família de aplicativos XBase.

DW. Nesta fase ocorrem as combinações e padronizações de chaves estrangeiras e primárias. Ocorre também o processo de criação de índices e particionamento físico das tabelas.

De acordo com Cielo ([s.d.]), existem programas de ETC disponíveis no mercado e são de grande valia principalmente quando os sistemas de OLTP (transacionais) são muitos. Estes programas são uma poderosa fonte de geração de metadados², que contribui muito para a equipe de DWA (*Data Warehouse Administrator*).

1.4 Data Mart

Um DM representa um conjunto de dados do DW. Compreende geralmente em um departamento ou área específica de interesse em uma corporação. Por conter esta característica, pode-se afirmar que um DM é menos genérico e mais específico. Como exemplo, cita-se um departamento de *Marketing* que armazena todos os dados relevantes do *Marketing* em seu próprio DM; o departamento de vendas coloca seus dados de vendas no DM de vendas.

Segundo Spenik e Sledge (2001, p. 648), “um DM é um repositório de dados reunidos a partir de dados operacionais ou outras origens que é projetado para atender um departamento particular ou um grupo funcional”.

Por tratar-se de um conjunto de dados menor, permite que a sua implementação seja compreendida em um espaço de tempo menor, reduzindo também o custo de investimento. Outro aspecto importante é que vários DM podem coexistir em uma mesma corporação. Eles podem ser integrados posteriormente em um único DW. Vejamos a seguir as principais diferenças entre eles.

1.5 Data Warehouse versus Data Mart

Um *Data Warehouse* é uma agregação central dos dados. Um *Data Mart* pode ser derivado de um *Warehouse* ou o *Warehouse* pode ser derivado de *Data Marts* menores especializados. O *Data Mart* enfatiza facilidade de acesso e capacidade de utilização para um propósito particular de projeto. Em geral, um *Data Warehouse* tende a ser estratégico mas freqüentemente incompleto; um *Data Mart* tende a ser tático e dirigido para atender uma necessidade imediata. (SPENIK e SLEDGE, 2001, p. 648)

² Metadados são dados do mais alto nível que descrevem os dados de um nível inferior. Também definido como Dados sobre Dados.

Como já foi dito anteriormente, um DM é um depósito de dados que inclui apenas as informações necessárias para uma área da empresa, normalmente focado em um departamento. Um DW é mais abrangente, incluindo os dados globais da empresa, possibilitando visualizações dentre todos os departamentos. Percebe-se algumas vantagens em um DM em relação ao DW (PETKOVIC, 2001):

- Área de aplicação mais estreita;
- Tempo de desenvolvimento mais curto (e um custo mais baixo de implantação);
- Maior facilidade de manutenção dos dados,
- Desenvolvimento de baixo para cima.

Destacamos na tabela 1.1 as principais diferenças entre um DM e um DW.

<i>Data warehouse</i>	<i>Data mart</i>
Utilização corporativa	Utilizado por departamento ou unidade funcional
Implementação complexa e demorada	Implementação rápida e fácil
Maior volume de dados	Volume de dados menor e mais especializado
Desenvolvido utilizando dados	Desenvolvido a partir das necessidades de dados dos usuários atualmente disponíveis

Tabela 1.1 - Diferenças-chave entre *Data Marts* e *Data Warehouse*

Fonte: (SPENIK e SLEDGE, 2001)

Os DM's apresentam claramente uma resposta mais rápida para as necessidades do departamento. O custo baixo e a facilidade do uso permitem uma implementação rápida e contribuem para que os DM's sejam a preferência das empresas quando estas resolvem desenvolver BI. Deve-se, no entanto, tomar cuidado com a proliferação dos DM's sem que se tenha uma gerência dos diferentes projetos a fim de consolidá-los. Os projetos de DM devem ser acompanhados por um gerente de projetos a fim de evitar que ocorram visões distorcidas dos resultados de cada DM. Divisões ou departamentos que constroem seus próprios DM's

correm o risco de obterem resultados diferentes, por manterem conceitos de negócios diferentes de outros departamentos.

2 MODELAGEM DIMENSIONAL

A Modelagem Dimensional, também conhecida como Modelagem Multidimensional, possui características diferentes daquela habitualmente utilizada nos sistemas OLTP. Os Modelos Relacionais freqüentemente utilizados para compor a estrutura lógica destes sistemas, utilizam uma abordagem normalizada para o projeto de banco de dados. A normalização remove a redundância do esquema, a fim de otimizar o armazenamento tornando-o principalmente mais simples, rápido e organizado. O DW não possui como foco principal a normalização e sim fornecer simplicidade ao usuário final (SPENIK e SLEDGE, 2001).

Tabelas altamente normalizadas não são convenientes para o projeto de um DW, pois o objetivo dos sistemas de DW é significativamente diferente. Se por um lado as tabelas normalizadas são vantajosas para o processamento *on-line*, por outro, os processos de DW são baseados em consultas que operam sobre um enorme volume de dados, tornando proibitivo o uso desta técnica. As consultas em um DW são pouco concorrentes e, cada transação acessa um número muito elevado de registros (PETKOVIC, 2001).

De acordo com Nardi (2007), a principal finalidade de bases de dados dimensionais é fornecer subsídio para realização de análises dos dados. A extração e consulta de informações a partir dos dados coletados para as diferentes áreas de negócio da empresa é a função elementar de um ambiente dimensional.

Crivelini ([s.d.], p. 23) destaca de maneira mais detalhada, o que se pode esperar deste modelo:

- Seja uma representação simples do modelo de negócio estudado;

- Seja um modelo físico de fácil interpretação, de modo que usuários sem treinamento formal em TI possam entendê-lo,
- Facilite a implementação física do modelo, de modo a maximizar performance das consultas aos dados.

Para tanto, sua arquitetura e até mesmo a terminologia empregada são distintas das utilizadas para bancos de dados transacionais. Na Modelagem Dimensional, deixa-se de focar a coleta de dados para se preocupar com a consulta dos dados.

Na Modelagem Dimensional todo modelo é composto de uma tabela ou mais tabelas que armazenam medidas, chamadas de tabelas de fatos, e várias outras tabelas secundárias, chamadas de tabelas de dimensão. A tabela fatos localiza-se no centro do diagrama e as tabelas de dimensões assumem posições em volta da tabela de fatos (PETKOVIC, 2001). Nesta modelagem existem dois elementos básicos que são:

- Fatos: são medidas de desempenho relativas ao assunto de negócio modelado.
- Dimensões: contextos de um fato.

2.1 Tabela de Fatos

Conforme Machado (2000), fato é uma coleção de itens de dados, composta de dados de medidas e de contexto. A tabela de fatos armazena medições quantitativas do negócio.

Um fato é um conceito de interesse primário para o processo de tomada de decisões e corresponde aos eventos que ocorrem de forma dinâmica no negócio da empresa. É a tabela primária do modelo dimensional e nela são armazenados os valores do negócio que se deseja analisar (MICROSOFT, 2007b). É na tabela de fatos que conterà o maior volume de informações no banco de dados dimensional.

A maioria das colunas em uma tabela de fatos é numérica e aditiva, pois tais dados podem ser usados pra executar cálculos necessários. Como exemplo, se pode citar a quantidade de admissões de pacientes em um hospital. Ou ainda, a quantidade de unidades vendidas de produtos em uma loja de calçados.

“Em uma tabela de fatos, uma linha corresponde a uma medição. Uma medição é uma linha em uma tabela de fatos. Todas as medições em uma tabela de fatos devem estar alinhadas na mesma granularidade” (KIMBALL; ROSS apud FELBER, 2006, p. 30).

De acordo com Italiano e Esteves ([s.d.]), a tabela de fatos possui dois grupos de atributos:

- As métricas ou fatos numéricos,
- As chaves para cada uma das dimensões do modelo.

De acordo com Kimball (1998), cada tabela de fatos contém as chaves externas que se relacionam com suas respectivas tabelas de dimensões e as colunas com os valores que serão analisados. As chaves externas são provenientes das tabelas de dimensão. Cada tabela de dimensão normalmente tem uma chave primária, além de possuir vários outros atributos. Em uma tabela de fatos, a chave primária é resultante da combinação das chaves primárias de todas as tabelas de combinação.

É possível compreender que a chave primária da tabela de fatos é constituída de várias chaves estrangeiras. Logicamente o número de chaves estrangeiras na tabela de fatos, especifica a quantidade de tabelas de dimensões. A figura 2.1 ilustra uma tabela de fatos com suas respectivas chaves estrangeira e as colunas de métricas ou fatos.

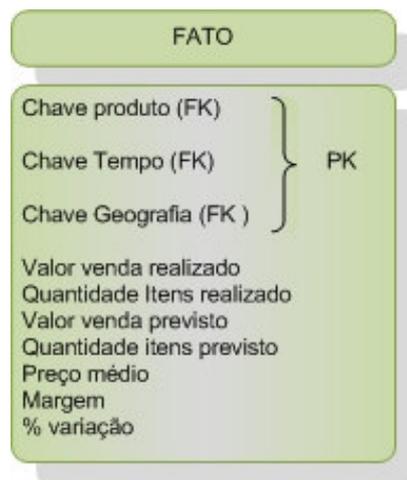


Figura 2.1 – Exemplo de tabela de fatos

Fonte: Adaptação do autor segundo (ITALIANO e ESTEVES, [s.d.])

Observamos nesta figura que a chave primária da tabela é composta pelas chaves estrangeiras: chave produto + chave tempo + chave geografia.

2.2 Tabela de Dimensões

A tabela de dimensão tem como principal função reunir os atributos que serão utilizados para qualificar as consultas e cujos atributos serão utilizados para agrupar e sumarizar as métricas (ou fatos) de acordo com as consultas realizadas pelo usuário (ITALIANO e ESTEVES, [s.d]).

Italiano e Esteves ([s.d.]) fazendo ainda uma analogia entre as diferenças entre tabelas de fato e de dimensões, afirma: “Se por um lado a tabela de fato representa quais informações serão analisadas, as tabelas de dimensões representam como elas serão analisadas”.

Cada tabela de dimensão deve possuir uma chave primária. Esta chave deve ser sempre um atributo único e definido pelo sistema com um valor genérico, inteiros e atribuídos sequencialmente. Esta chave primária é fundamental para garantir a integridade referencial com a tabela fato. Kimbal (1998) recomenda que se utilize chaves genéricas, também chamadas de chaves artificiais, substitutas ou ainda *Surrogate Keys*. Uma das razões deve-se a interdependência que o DW deve manter dos sistemas transacionais a fim de evitar possíveis vulnerabilidades como sobreposição de chaves ou reaproveitamento de códigos. Como um DW manterá informações durante muito tempo, problemas desta magnitude não poderão acontecer.

Kimball (1998) sugere que as tabelas de dimensões possuam o maior número possível de atributos textuais com a finalidade de enriquecer o leque de dados, para conseqüentemente possibilitar análises mais abrangentes. Entretanto, deve-se fazer uma seleção criteriosa a fim de evitar que o DW ou DM tornem-se um grande cadastro de informações, com dados que não fazem sentido ao processo de análise.

A figura 2.2 ilustra alguns exemplos de tabelas de dimensão. Verificamos que cada tabela possui uma chave primária. Estas chaves primárias na tabela dimensão deverão constar na tabela fatos como chaves estrangeiras.



Figura 2.2 – Exemplo de tabela de dimensões

Fonte: Adaptação do autor segundo (ITALIANO e ESTEVES, [s.d.])

2.3 Técnicas de Modelagem

Como já foi dito anteriormente, a construção de um DW requer o desenvolvimento de um modelo dimensional de dados provenientes do SGBD transacional. Através deste modelo obtém-se um conjunto de tabelas relacionadas entre si que compõem uma base de dados de suporte à decisão. Apesar de existir uma série de técnicas para executar esta modelagem será abordada neste trabalho a mais conhecida: *Star schema*.

2.3.1 Star Schema

Star Schema (esquema estrela), é o modelo mais difundido entre os DW existentes. Seu modelo consiste basicamente em um conjunto de tabelas de dimensões relacionadas com uma única tabela de fatos. A tabela de fatos fica localizada no centro do modelo, e suas tabelas de dimensões ficam arranjadas ao redor desta unidade central, originando um formato semelhante ao de uma estrela (MACHADO, 2000).

Machado (2000) ressalta uma característica importante deste esquema referente ao relacionamento das entidades dimensões, que possuem um registro para muitos na tabela de fatos.

Como exemplo, é possível criar um modelo estrela composto pelo fato Vendas, seguido das dimensões Produto, Tempo e Geografia. A figura 2.3 ilustra o exemplo.

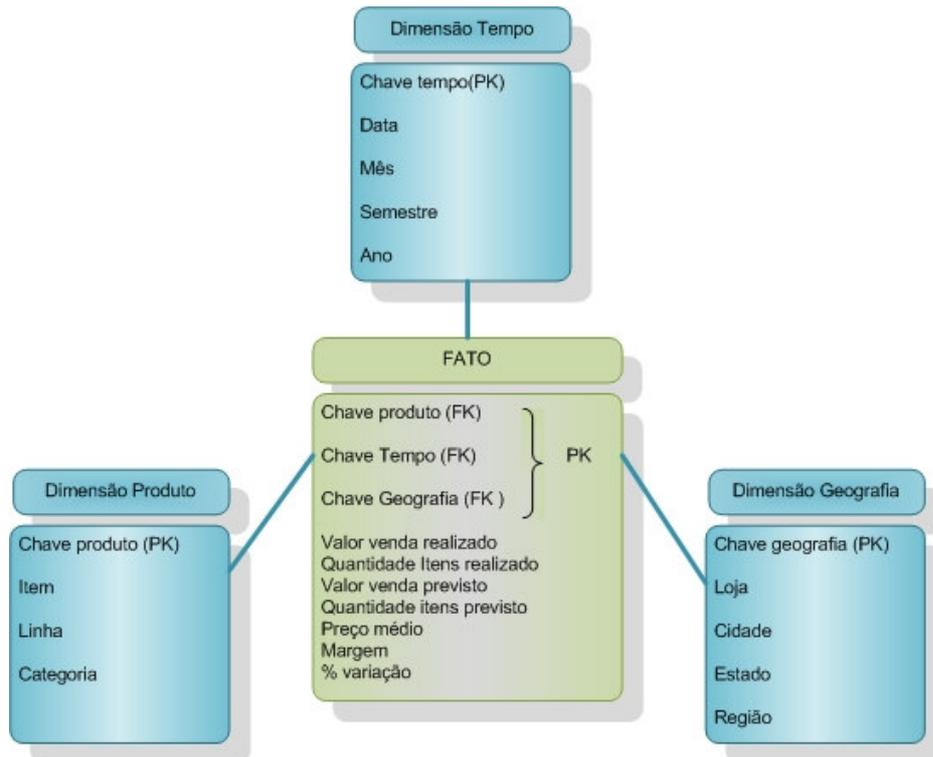


Figura 2.3 – Exemplo da modelagem *Star Schema*
 Fonte: Adaptação do autor segundo (ITALIANO e ESTEVES, [s.d.])

A figura permite uma rápida compreensão do arranjo proposto, devido sua simplicidade, característica deste modelo.

O *Star Schema* é simples e de fácil assimilação, além de ser muito familiar mesmo a usuários que não são da área de TI, como por exemplo, usuários da área de negócios, que geralmente participam deste processo. Torna-se possível discutir com os diferentes níveis de usuários esta modelagem. Geralmente é um esquema totalmente não padronizado e pode estar parcialmente padronizado nas tabelas de dimensões.

2.4 Cubos

O fato da arquitetura da modelagem dimensional oferecer subsídios para a realização de análises resulta em uma arquitetura, e até mesmo a terminologia empregada, diferente daquelas utilizadas para banco de dados tradicionais (NARDI, 2007).

A representação mais intuitiva de visualizar um modelo dimensional é através do desenho de um cubo. Petkovic (2001, p. 450), define o cubo como “...um subconjunto de dados do DW que pode ser organizado em estruturas multidimensionais.”

A principal razão de utilizar um cubo como representação gráfica, é de transmitir o conceito de múltiplas dimensões. O objeto cubo é utilizado para fazer referência à dimensionalidade entre as tabelas de dimensões e a tabela fato. Usualmente os modelos dimensionais possuem mais de três dimensões, passando o cubo a denominar-se hipercubo.

De acordo com Parrini (2002), um cubo possui diversas células e cada fato é guardado em uma célula. O fato possui um conjunto de medidas que o representa. Para se referenciar a um fato específico, é necessário especificar qual a sua coordenada ao longo de cada dimensão.

Apresentamos na figura 2.6 um exemplo de cubo contendo as dimensões mês, produto e região, sendo que as células deste cubo representam o fato vendas. É possível extrair, por exemplo, a quantidade vendida do produto leite na região sul no mês de julho.

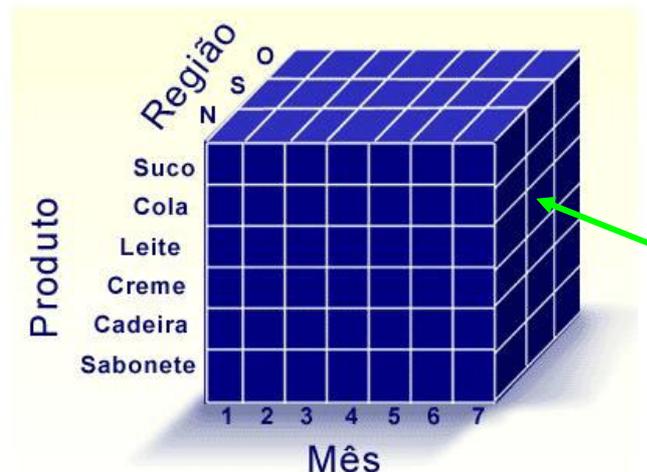


Figura 2.6 – Exemplo de um cubo
Fonte: Adaptação do autor segundo (NARDI, 2007)

2.5 OLAP

Diante do contexto de DW e tomada de decisão, as ferramentas OLAP tem fundamental importância de auxiliar aos usuários na busca de respostas que agreguem valor estratégico. As ferramentas OLAP possuem como alicerce o modelo dimensional. Sua aplicação pode se dar tanto em níveis departamentais (DM) quanto em níveis globais (DW) de uma empresa.

As diferentes ferramentas de OLAP existentes no mercado têm como principal finalidade oferecer suporte aos analistas gerentes e executivos, a fim de oferecer uma visão

mais apurada dos dados de sua empresa através de relatórios, comparativos e também por meio de análise dos dados históricos projetados (INMON; WELCH; GLASSEY, 1999).

Inmon, Welch e Glasey (1999), definem OLAP como um conjunto de funcionalidades que tenta facilitar a análise multidimensional, através de manipulação dos dados que tenham sido agregados em várias categorias ou dimensões.

Ou seja, OLAP é a ferramenta que os *Decison Makers* utilizam para realizar suas pesquisas, comparações e extrações de tendências, oportunizando uma melhor compreensão dos negócios da empresa. Obtem-se através desta ferramenta, a habilidade de apresentar a multidimensionalidade em telas através de uma configuração tridimensional de linhas, colunas e páginas (CAMPOS, 2003).

A Microsoft (2007b), destaca algumas características importantes sobre esta ferramenta:

- Os bancos de dados OLAP apresentam um esquema otimizado para que as perguntas realizadas pelos usuários sejam respondidas rapidamente.
- As perguntas realizadas a um OLAP devem permitir a utilização interativa com os usuários.
- Os cubos OLAP armazenam vários níveis de dados formados por estruturas altamente otimizadas que atendem às expectativas de negócio da empresa.
- Um sistema OLAP está preparado para realizar relatórios complexos de uma forma simples.
- OLAP proporciona uma visão multidimensional dos dados. Os cubos oferecem uma visão multidimensional dos dados que vai além da análise de duas dimensões, oferecida por uma simples planilha de cálculo utilizada como tal.
- Os usuários podem modificar facilmente as filas, as colunas e as páginas nos relatórios do OLAP, sendo possível visualizar a informação da forma mais conveniente para análise.

A seguir, são apresentadas algumas áreas onde o uso de utilizam ferramentas OLAP está difundido:

- Área financeira: Relatórios analíticos, processamento e análise de contas de clientes e fornecedores.
- Vendas e Marketing: Relatórios de análises de produtos, clientes e faturamento.
- Recursos Humanos: Análises de rotatividade, departamentos, folha de pagamento.
- Produção e chão de fábrica: Análise dos produtos, custo de materiais, análise de produção.

De acordo com Parrini (2002), deve-se observar a ferramenta OLAP como apenas um instrumento que um *Decison Maker* pode contar. As visões e a criatividade que o tomador de decisão vai utilizar são elementos inteiramente individuais. As técnicas de análise de problemas são diferentes entre cada pessoa. A possibilidade de desenvolver soluções diferentes é comum de acontecer, tendo em vista que para desenvolver processos decisórios, são envolvidos, além de outras ferramentas, uma série de atividades e abordagens descobertas pelo escopo de ferramentas OLAP.

2.5.1 Tipos de operações OLAP

De acordo com Machado (2000), Kimball (1998) e Nardi (2007), existem diferentes formas de navegar entre as hierarquias a fim de selecionar quais perspectivas deseja-se analisar. Destacamos a seguir as mais difundidas:

Drill Across: está relacionado ao fato de poder movimentar de um esquema para o outro, desde que ambos tenham algumas dimensões em conformidade. Ou seja, as mesmas dimensões estão compartilhadas. Como exemplo, é possível citar os dados de compra de matéria prima, sumarizados por mês, semana e dia. Executando um *Drill Across*, o usuário passaria do mês direto para o dia.

Drill Down: ocorre quando o usuário aumenta o nível de detalhe da informação, diminuindo o grau de granularidade. O usuário pode navegar do mais alto nível até o mais

detalhado. Como exemplo, se pode citar a visualização do total de rotatividade de em um departamento de recursos humanos. Realizando o *Drill Down*, poderíamos visualizar a rotatividade de um determinado cargo.

Roll Up: é o contrário do *Drill Down*. Ocorre quando o usuário aumenta o grau de granularidade, diminuindo o nível de detalhamento da informação. O usuário pode navegar do nível de detalhe até o nível mais alto de sumarização de dados. Como exemplo, se pode citar o total de vendas de um determinado produto, quando utilizado o Roll Up, passaríamos a visualizar a venda de todos os produtos.

Drill Through: este conceito está relacionado ao fato de poder armazenar um nível de detalhe menor do que aquele mencionado na tabela fato, mas permitido pela sua granularidade. Por exemplo, em um nível de granularidade de pessoal, por dia e departamento, deseja-se buscar uma informação que está presente no cartão-ponto. A operação *Drill Through* poderia efetuar esta busca no próprio sistema operacional caso houvesse compatibilidade entre os dois ambientes.

Slice and Dice: é uma das principais características de uma ferramenta OLAP. Este conceito surgiu da necessidade de recuperar uma “fatia” de um microcubo. Ele serve para modificar a posição de uma informação, alterar as linhas por colunas de maneira a facilitar a compreensão dos usuários e girar o cubo sempre que houver necessidade. Como exemplo em um cubo que representa a produção de roupas e calçados, utilizando o *Slice and dice*, poderíamos visualizar somente a produção de calçados.

2.5.2 Arquiteturas OLAP

Existem diferentes métodos de armazenamento de dados em uma aplicação OLAP. Conforme o método de armazenamento utilizado, será elaborada uma arquitetura de aplicação. As arquiteturas existentes são: *Multidimensional On-Line Analytical Processing* (MOLAP), *Relational On-Line Analytical Processing* (ROLAP), *Desktop On-Line Analytical Processing* (DOLAP) e *Hybrid On-Line Analytical Processing* (HOLAP).

2.5.2.1 MOLAP

A arquitetura MOLAP sugere que os dados sejam armazenados de forma multidimensional e que sua consulta ocorra diretamente nesta base de dados. Ou seja, as

consultas ocorrem diretamente no banco de dados dimensional usando-se um cubo multidimensional. (MICROSOFT, 2007d).

De acordo com a Microsoft (2007d), esta arquitetura, apresenta algumas vantagens tais como velocidade nas consultas, excelente rendimento e compressão de dados devido a multidimensionalidade, além de apresentar melhor tempo de resposta, dependendo apenas do percentual de agregações do cubo.

Anzanello (2007) e a Microsoft (2007d) destacam como desvantagem a demora no processo de carga de um banco de dados dimensional devido a complexidade. O processo de carga é rebuscado e complexo, devido à série de cálculos que devem ser realizados durante a agregação dos dados às dimensões. Outra desvantagem deve-se o período de latência que ocorre enquanto os dados de origem que foram modificados ficarem apenas em sua base de dados relacional. Para que ocorra uma nova coerência entre as bases, é necessário realizar uma nova carga no banco de dados dimensional.

Verificamos na figura 2.7 esta arquitetura.

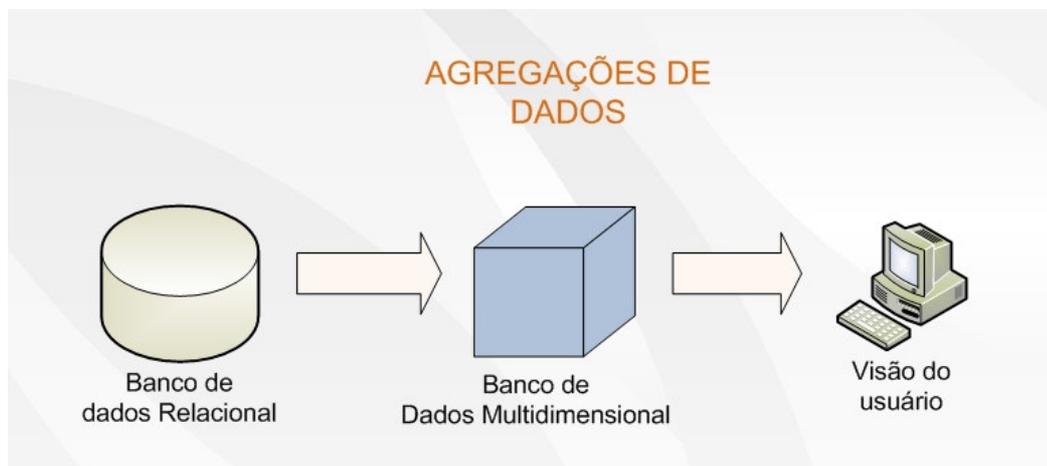


Figura 2.7 – Arquitetura MOLAP

Fonte: Adaptação do autor segundo (MICROSOFT, 2007d)

2.5.2.2 ROLAP

De acordo com Leitão (2000), a principal característica da arquitetura ROLAP está na possibilidade de operar diretamente com banco de dados relacionais. É possível realizar qualquer consulta, não se limitando ao “cubo”, além de maior capacidade em atingir níveis de detalhamento mais baixo.

Diferente do modo de armazenamento MOLAP, o ROLAP não armazena cópia do banco de dados. As tabelas originais são acessadas quando executa-se qualquer tipo de

consulta. Utiliza-se esta arquitetura quando trabalha-se com grande conjunto de dados consultados com pouca frequência e pretende-se economizar espaço de armazenamento. (MICROSOFT, 2007d).

As vantagens desta arquitetura estão nas múltiplas possibilidades em criar novos relacionamentos entre as dimensões existentes, facilitando a rotina de usuários da área de *marketing*, por exemplo, no qual nem sempre possuem seus escopos predefinidos, além de não haver duplicação de dados (LEITÃO, 2000).

Leitão (2000) e a Microsoft (2007d) apontam como principal desvantagem o seu tempo de resposta que é por vezes muito longo devido as complexas e longas pesquisas que são realizadas no banco de dados relacional.

A figura 2.8 exemplifica esta arquitetura.

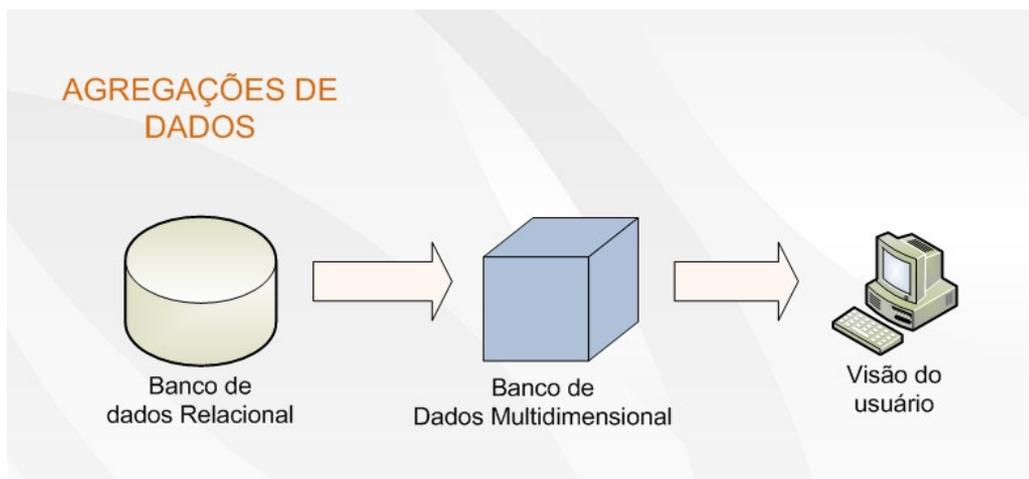


Figura 2.8 – Arquitetura ROLAP

Fonte: Adaptação do autor segundo (MICROSOFT, 2007d)

2.5.2.3 HOLAP

Na arquitetura HOLAP, ocorre combinação entre as arquiteturas ROLAP E MOLAP, citadas anteriormente. É um conceito concebido mais recentemente e tem por principal característica armazenar as agregações em uma estrutura multidimensional e os dados detalhados em um banco de dados relacional. Para procedimentos de busca que acessam dados sumarizados, o HOLAP é equivalente ao MOLAP. Quando os processos de consultas acessam os níveis mais baixos de detalhe, utiliza-se o banco de dados relacional para retirada dos dados (MICROSOFT, 2007d).

A vantagem desta arquitetura está em justamente aproveitar o que tem de melhor em cada uma das estruturas envolvidas: a alta performance do MOLAP e a escalabilidade do ROLAP (ANZANELLO, 2007).

De acordo com a MICROSOFT (2007d) destacam-se os casos mais comuns de uso de HOLAP

- Cubos que requerem resposta rápida.
- Quando existem sumarizações baseadas em uma grande quantidade de dados de origem.
- Solução com o compromisso de reduzir o espaço ocupado sem prejudicar totalmente o rendimento das consultas.

A figura 2.9 ilustra a arquitetura HOLAP.

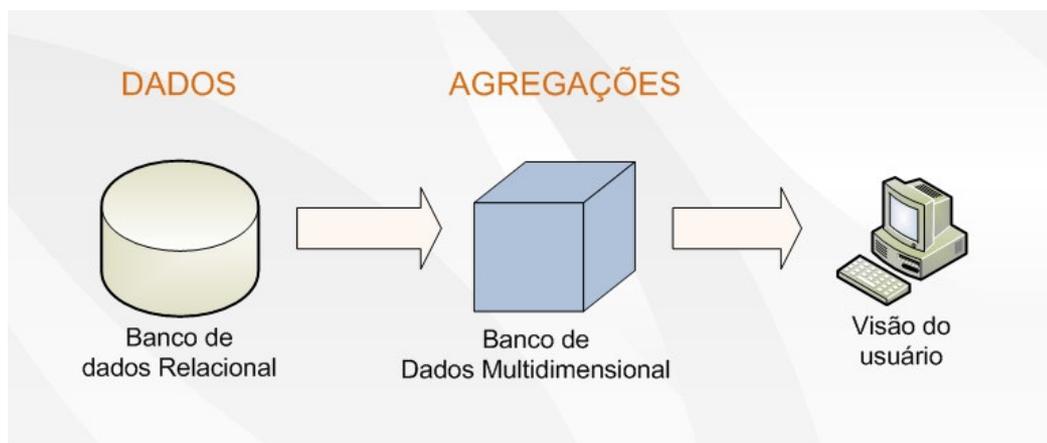


Figura 2.9 – Arquitetura HOLAP
Fonte: Adaptação do autor segundo (MICROSOFT, 2007d)

2.5.2.4 DOLAP

De acordo com Santos (2001) DOLAP é uma arquitetura que acessa seus dados, enviando consulta a partir de uma estação de trabalho para o servidor do banco de dados. Este devolve o “cubo” para ser analisado na própria estação de trabalho. A análise ocorre na máquina do cliente.

A vantagem desta tecnologia está na no trânsito reduzido de tráfego na rede devido à análise do cubo ocorrer na máquina do cliente, além de garantir maior agilidade nas análises

devido ao mesmo fator. A desvantagem deve-se à limitação da capacidade do processamento da estação de trabalho que geralmente não possui uma configuração adequada para tal.

A figura 2.10 ilustra a arquitetura DOLAP.

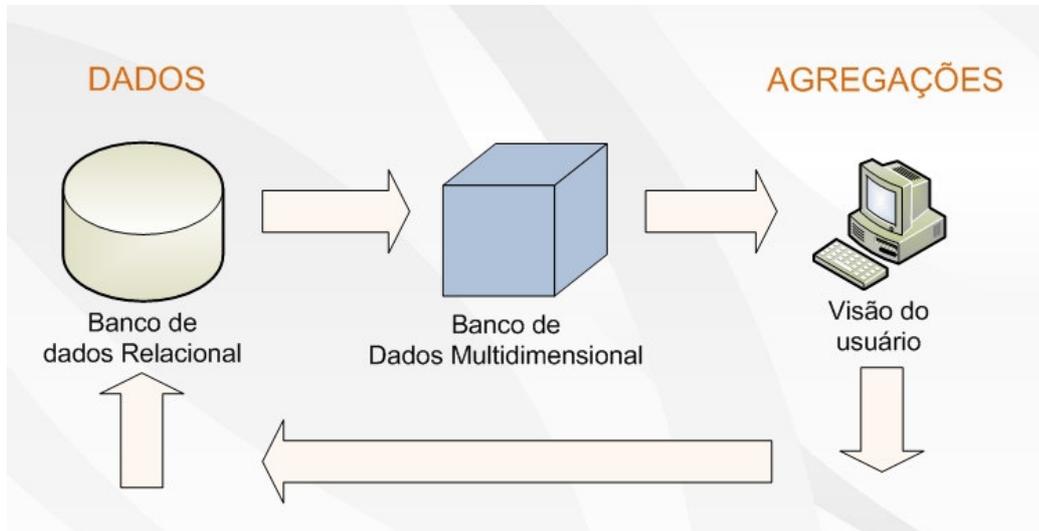


Figura 2.10 – Arquitetura DOLAP

3 TRABALHOS RELACIONADOS

Muito trabalho já tem sido feito em relação à DM's e ferramentas OLAP, como exemplo, pode-se citar os trabalhos de Felber (2006) e Rosa (2004). Na maioria destes trabalhos, o principal foco é buscar informações e combinações com alto valor estratégico que tragam alguma vantagem às corporações. Este capítulo propõe-se a verificar detalhes de outros trabalhos pertinentes à área, a fim de conhecer principalmente quais as experiências adquiridas durante o processo de modelagem, escolha das ferramentas, definição das tabelas de fato e dimensões, além de conhecer os resultados obtidos aplicando ferramentas de OLAP em seus *Data Marts*.

3.1 Centro Universitário

A proposta de Mata (2005) tem o intuito de apresentar uma solução desenvolvida no Centro Universitário de Palmas. Para o desenvolvimento do BI foi utilizado o sistema do Processo de Avaliação Institucional. O principal objeto deste sistema de avaliação é elevar o nível de satisfação do corpo discente e docente da mesma.

Este processo é aplicado pela Comissão Própria de Avaliação e tem por objetivo acompanhar o crescimento do Centro Universitário Luterano de Palmas. O processo ocorre anualmente na instituição com a participação de todas as pessoas que estão diretamente envolvidas com o Centro Universitário (MATA, 2005, p.32).

Para a realização do projeto, foram aplicados os conceitos da tecnologia de DW tais como modelagem dimensional, técnicas de ETC entre outras. Foi desenvolvido uma aplicação em ASP.Net para realizações das consultas no DM com o propósito de facilitar o acesso às informações de modo mais rápido e sumarizado. O SGBD utilizado foi o *Microsoft SQL Server*.

A estrutura básica de origem dos dados contém oito tabelas divididas em duas partes. A primeira parte abrange a distribuição das questões que foram aplicadas, conforme aplicado na figura 3.1.

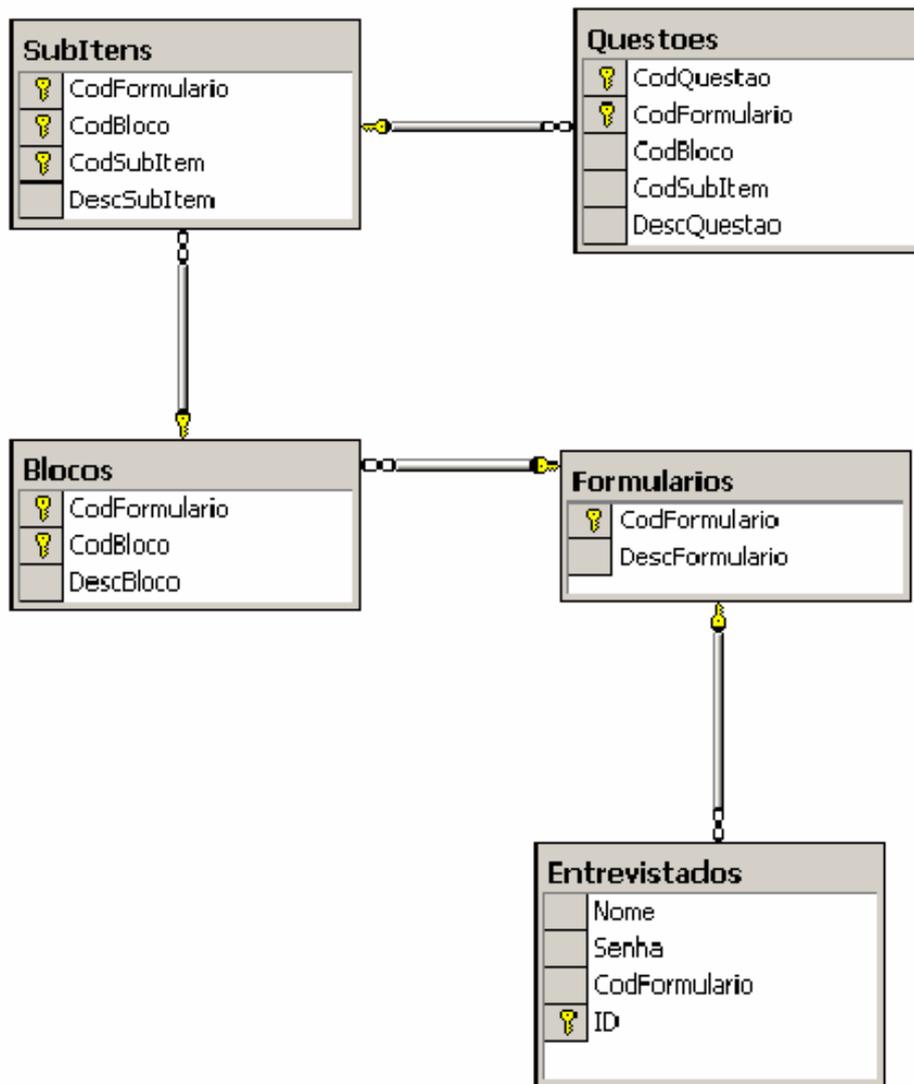


Figura 3.1 – Diagrama SLQ Server com questões da avaliação
Fonte: (MATA, 2005)

A segunda parte abrange a forma como as respostas estão armazenadas, conforme estrutura representada na figura 3.2.

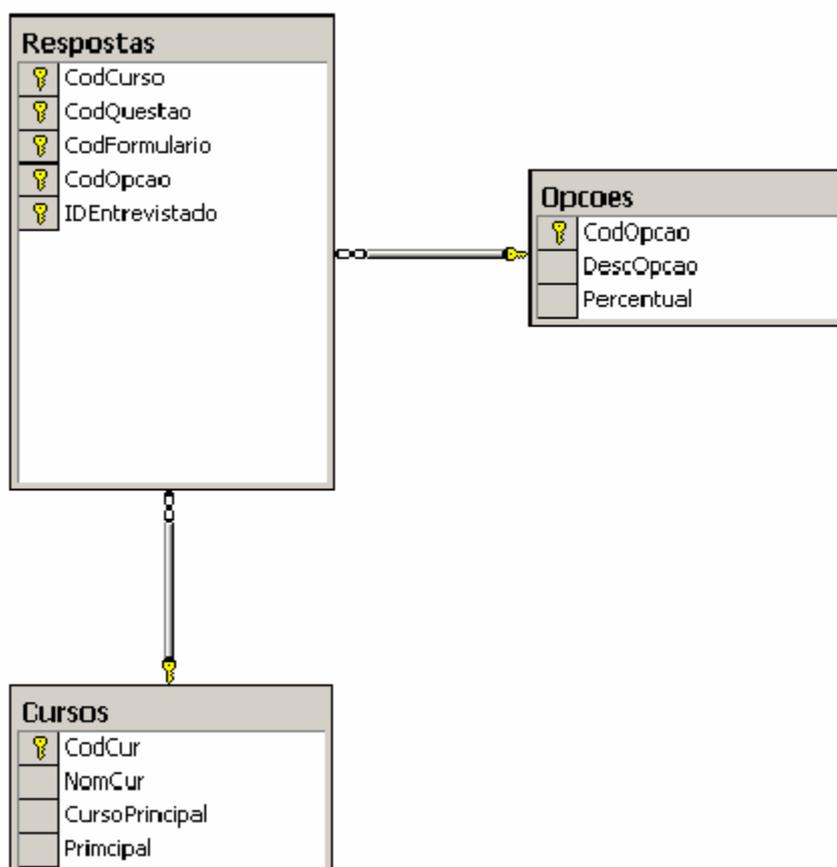


Figura 3.2 – Diagrama SLQ Server com repostas da avaliação
Fonte: (MATA, 2005)

3.1.1 Etapas do Projeto

O projeto pode ser dividido em seis importantes etapas³. As etapas compreendem em: Planejamento, Levantamento das Necessidades, Modelagem Dimensional, Processo de Extração, Transformação e Carga de Dados, Desenvolvimento da Aplicação e Verificação dos Resultados. O desdobramento de cada etapa é apresentado a seguir:

Planejamento: Foi criada uma nova infra-estrutura tecnológica com novo repositório de dados, totalmente separada do sistema de Avaliação Institucional para evitar a perda de desempenho nos sistemas.

Levantamento das necessidades: foi realizado a partir da análise das informações contidas no banco de dados de origem. A partir dos maiores índices de insatisfação e das necessidades estabelecidas pela instituição foi definida uma única tabela de fatos.

³ Estas etapas do projeto foram preservadas nos estudos de casos seguintes para possibilitar maior compreensão entre os casos estudados.

Modelagem dimensional: Foram criadas sete dimensões. Cada dimensão foi modelada de acordo com as informações do modelo fonte com o acréscimo dos campos Data_carga e Seq_Dimensão, respectivamente. Esses campos são utilizados para indicar a data da carga da informação e o número seqüencial, que foi utilizado como campo chave da dimensão, garantindo mais estabilidade do projeto.

Com o auxílio das informações levantadas, foi desenvolvida a modelagem dimensional dos dados. A figura 3.3 apresenta o modelo.

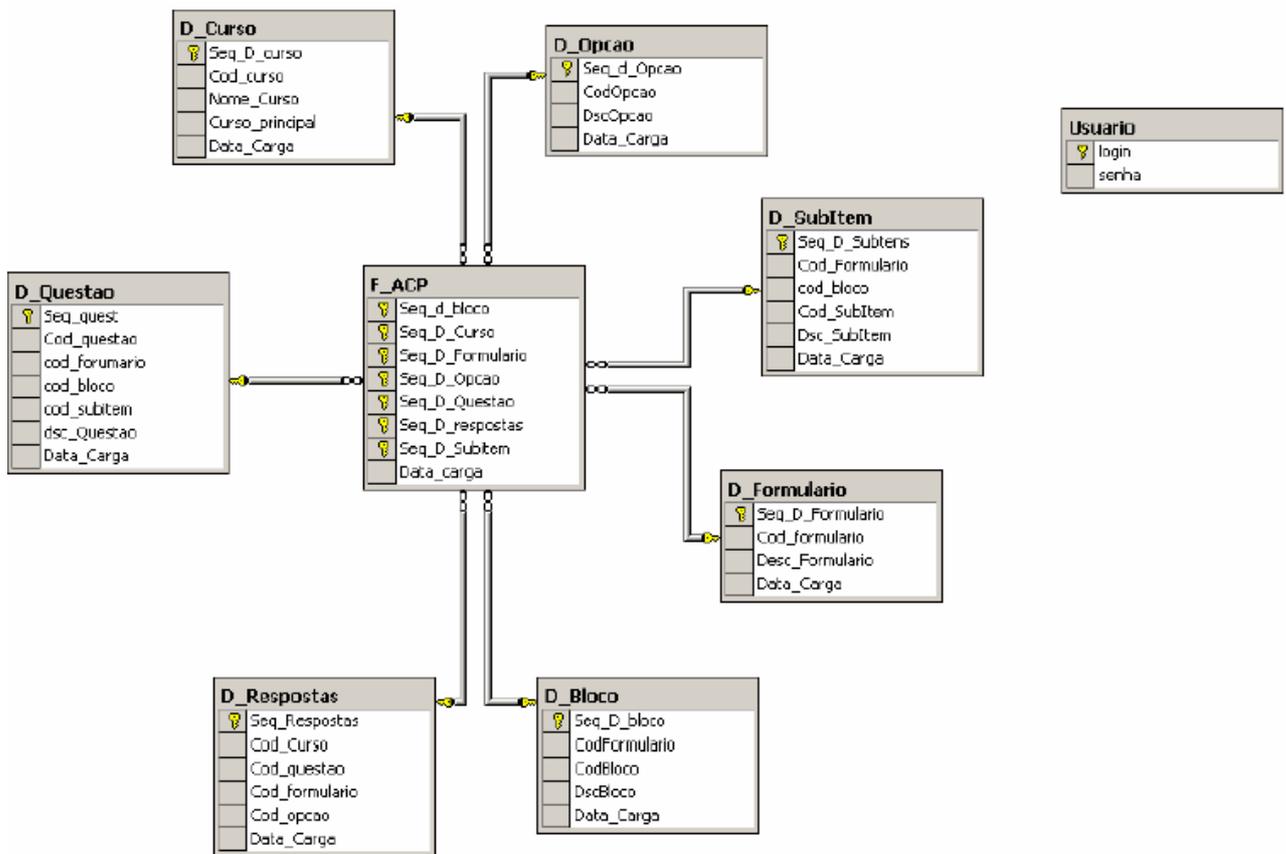


Figura 3.3 – Representação do banco criado para BI

Fonte: (MATA, 2005)

Processo de Extração, Transformação e Carga de Dados: com o apoio do DTS⁴ (*Data Transformation Services*), do *Microsoft SQL SERVER*, foi gerado um procedimento denominado ACP - DTS⁵. Inicialmente, foi gerado um procedimento único para carregar todas as dimensões.

⁴ DTS é uma ferramenta que auxilia na importação, transformação e exportação dos dados.

⁵ ACP – DTS é um método de extração e transformação dos dados de origem, desenvolvido pelo autor.

Desenvolvimento da aplicação: a aplicação foi desenvolvida na linguagem ASP.Net com o apoio da ferramenta *Visual Studio*, foi composta pelo módulo de autenticação e pelo módulo de consulta de aluno e professor.

Resultados: os testes realizados foram aplicados e validados. Foram obtidos vários resultados de valor real para a execução de ações estratégicas. Entre os resultados obtidos, destacam-se o relacionamento entre professores e alunos e a quantidade de professores que esclarecem as dúvidas de alunos que apresentam dificuldades. As tabelas 3.1 e 3.2 apresentam estes resultados.

Opção	Respostas
Indiferente	5
Insatisfeito	4
Não se Aplicada	1
Totalmente Insatisfeito	12

Tabela 3.1 - Relacionamento entre professores e alunos

Fonte: (MATA, 2005)

Opção	Respostas
Indiferente	4
Insatisfeito	6
Não se Aplicada	1
Totalmente Insatisfeito	11

Tabela 3.2 - Professores que esclarecem as dúvidas de alunos em dificuldades

Fonte: (MATA, 2005)

3.1.2 Considerações do autor

Mata (2005) destaca que através do uso das técnicas de BI foi possível compreender melhor o quadro de alunos e professores da instituição e que estes resultados facilitarão o trabalho dos gestores nas tomadas de decisões.

Dentre as etapas mais complexas durante o desenvolvimento de uma tecnologia BI, o autor destaca o levantamento de requisitos junto ao usuário, o estudo do sistema de origem, o estudo dos relatórios além das etapas de ETC.

De acordo com o autor, A grande vantagem de criação deste novo ambiente é a organização das informações. Neste ambiente, os gestores têm possibilidade de executar o cruzamento de informações para a obtenção de resultados, encontrando um diagnóstico de falhas.

3.2 Detecção de fatores de risco para doença arterial coronariana

A proposta de Silva (2005) tem por finalidade apresentar o desenvolvimento de uma solução de SAD para a análise dos dados do evento UNAERP Solidária, da faculdade UNAERP de Ribeirão Preto - SP. O principal objeto desta solução é levantar os fatores de risco para a doença isquêmica cardíaca também conhecida como doença arterial coronariana, na população participante do evento, utilizando o referencial “Campo de Saúde”, nome dado ao questionário eletrônico oferecido aos participantes.

Foi escolhida oportunamente uma ação comunitária e festiva da cidade, em comemoração aos 81 anos da Universidade de Ribeirão Preto e ao aniversário de 149 anos da cidade de Ribeirão Preto. Nesta oportunidade, foi possível desenvolver um ambiente de pesquisa em conjunto com os profissionais dos cursos de Enfermagem, Engenharia de Computação, Análise de Sistemas e do Centro Clínico Electro Bonini.

Para a realização do projeto, foram aplicados os conceitos da tecnologia de DW tais como modelagem dimensional, técnicas de ETC entre outras.

3.2.1 Etapas do Projeto

As seis etapas deste projeto foram desenvolvidas da seguinte forma:

Planejamento: Foi criada uma infra-estrutura tecnológica com repositório de dados limpo. Foi desenvolvida uma aplicação em *Delphi 7*, possibilitando cadastro de dados antropométricos do participante, da história clínica e relacionados ao estilo de vida; além de dados dos exames médicos. Os dados coletados foram armazenados em um SGBD *Microsoft Access 2000*. Para análise dos dados, foi utilizada a ferramenta *OLAP do Microsoft Excel XP*.

O cadastro dos dados antropométricos do participante consiste em: nome, sexo, data de nascimento, peso e altura; dados da história clínica como históricos de hipertensão arterial sistêmica, diabetes *mellitus*, dislipidemias, doença vascular periférica e acidente vascular cerebral; dados relacionados ao estilo de vida, como se o participante consome bebidas alcoólicas, o tipo de bebida a frequência, o uso de tabaco, tempo e quantidade de cigarros por dia; se pratica alguma atividade física e a modalidade; dados do exame médico como valores de pressão arterial, colesterol e glicose.

Levantamento das necessidades: A proposta da realização da pesquisa é de apresentar através dos dados correlacionados, estudos para diagnóstico, prevenção e tratamento da doença arterial coronariana na comunidade. O objetivo é de levantar os fatores de risco para a doença arterial coronariana na população participante do evento, utilizando o referencial teórico de “Campo de Saúde”, por ele proporcionar uma interpretação holística dos problemas dos portadores de coronariopatias.

Modelagem dimensional: Conforme Silva (2005), a implementação do SAD deu-se através da criação de um DM baseado no esquema *Star*. Esta modelagem é composta de um único Cubo. A tabela de fatos corresponde aos dados adquiridos durante a pesquisa. A representação do cubo é verificada na figura 3.4.

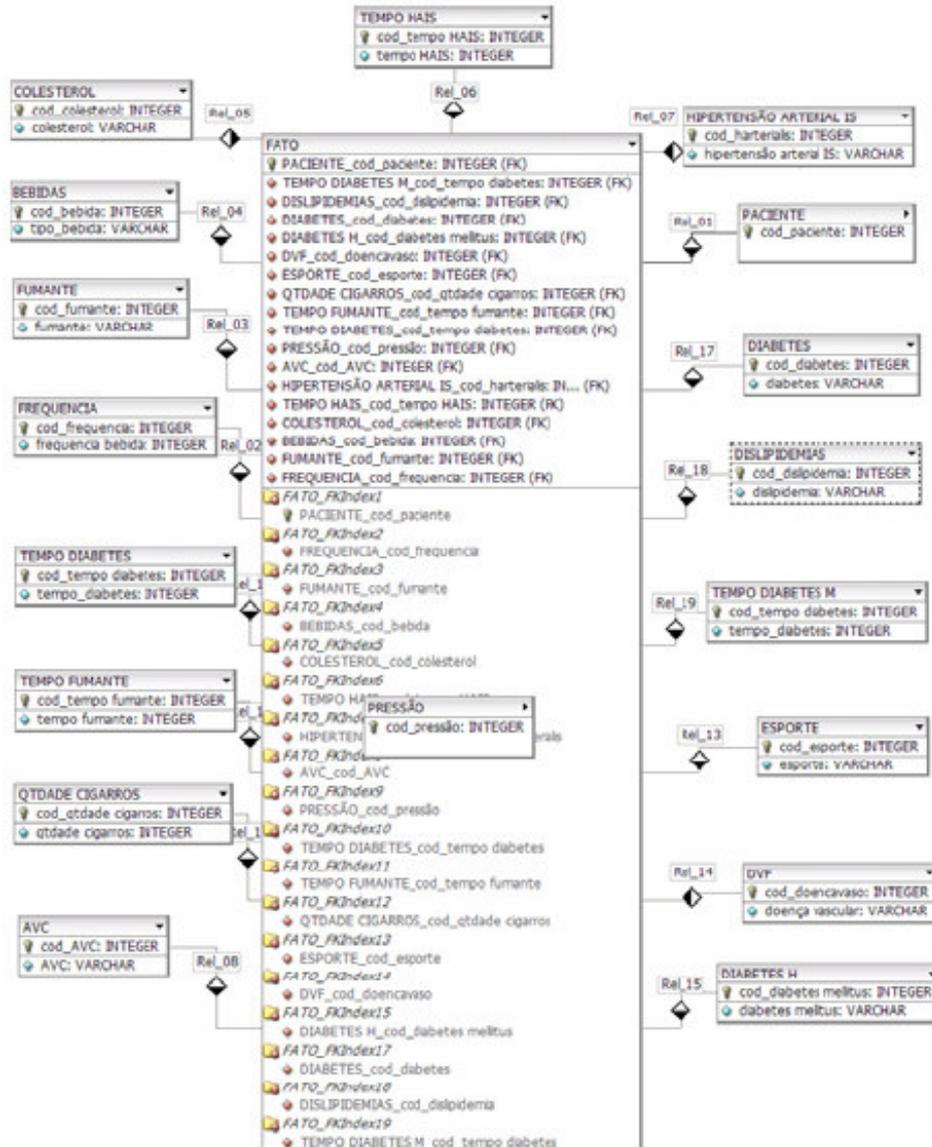


Figura 3.4 – Representação do cubo

Fonte: (SILVA, 2005)

Processo de Extração, Transformação e Carga de Dados: A extração dos dados foi realizada através da ferramenta *Microsoft Query*. Foi gerado um procedimento único para carregar todas as dimensões e a tabela de fatos e exportá-las para a ferramenta *Microsoft Excel*.

Desenvolvimento da aplicação: A aplicação foi desenvolvida na linguagem *Delphi 7* composta por um único módulo disponível aos usuários para o preenchimento da pesquisa.

Resultados: A população que participou da pesquisa, foi constituída por 495 participantes com idades entre 09 e 85 anos, sendo 5% não informaram o sexo, 54% são do

sexo feminino e 41% masculino. Dentre os resultados tabulados, destacam-se aqueles relacionados ao grau de risco, sexo, e IMC⁶. A tabela 3.3 apresenta o índice obtido.

IMC(Kg/m ²)	SEXO %	M	%	F	%	Total
Peso normal (<25)		69	38,8	109	61,2	178
Sobrepeso (25 a 27,4)		61	54	52	46	113
Obeso (27,5 a 39,9)		29	48	31	52	60
Obesidade grave (≥ 40)		0	0	06	100	06
Não medido		42	37	73	63	115

Tabela 3.3 - Distribuição de clientes segundo índice de massa corporal (IMC) e sexo (M ou F), desconsiderando os indivíduos que não informaram o sexo e/ou não existiu este dado.

Fonte: (SILVA, 2005)

3.2.2 Considerações do autor

O Autor ressalta a importância da aplicação dos recursos de BI para o desenvolvimento do seu trabalho de diagnóstico. Dentre os motivos, destaca a estrutura amigável e de baixo custo que foi criada para analisar as informações coletadas.

“...a partir do uso da ferramenta OLAP podem-se fazer análises que antes não eram possíveis, de maneira rápida e flexível garantindo assim, maior acuracidade nas tomadas de decisão, além de interação entre as áreas de diferentes saberes técnico-científicos em busca de qualidade de vida dos cidadãos e adoção de hábitos saudáveis para a população em geral” (SILVA, 2005, p.4).

3.3 Proposta de aplicação de BI em chão de fábrica

Fortulan e Gonçalves Filho (2005) apresentam uma proposta de desenvolvimento de um DW para uma empresa de manufatura. Os autores destacam a grande quantidade de dados que encontram-se dispersos ou desorganizados em todo o ambiente de chão-de-fábrica. Os dados quando organizados e agrupados podem ser transformados em informações úteis, e produzem um bom conjunto de indicadores de desempenho para possíveis análises estratégicas.

“O objetivo é desenvolver um sistema que utilize os dados resultantes do processo produtivo e os transforme em informações que auxiliem o gerente na tomada de decisões, de

⁶ IMC = Índices de Massa corporal

forma a garantir a competitividade da empresa.” (FORTULAN E GONÇALVES FILHO, 2005, p.1).

Neste trabalho foram aplicados os conceitos da tecnologia de DW tais como modelagem dimensional e técnicas de ETC. O SGBD e o módulo para inserção de dados utilizado foi o *Microsoft Access 2000*. A estrutura básica dos dados contém dezesseis tabelas já modeladas para o ambiente de DW. A figura 3.5 apresenta o modelo.

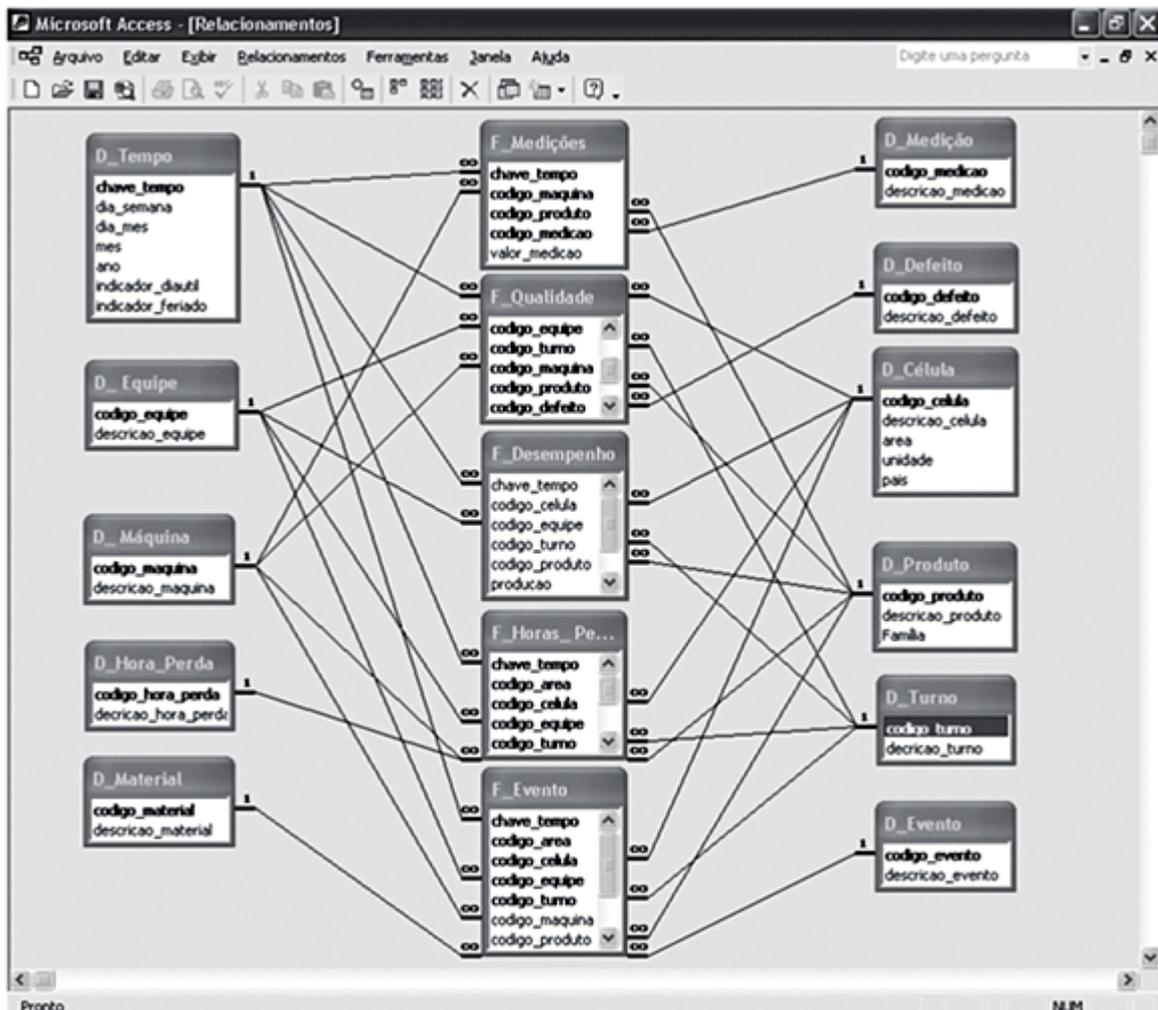


Figura 3.5 – Diagrama MS Access
Fonte: (FORTULAN E GONÇALVES FILHO, 2005)

3.3.1 Etapas do projeto

As seis etapas deste projeto foram desenvolvidas da seguinte forma:

Planejamento: Foi desenvolvido um ambiente de interação com o usuário para o lançamento dos dados do chão de fábrica utilizando SGBD *Microsoft Access 2000*. Os dados foram posteriormente incorporados ao *Analysis Services*, componentes da suíte *Microsoft*

SQL Server 2000 Enterprise. Os autores optaram pelo SGBD Access porque além do repositório de dados, é possível desenvolver os módulos de interação com o usuário final, permitindo maior facilidade para a inserção dos dados.

Levantamento das necessidades: A proposta dos autores é encontrar indicadores que facilite a análise e a tomada de decisão para o ambiente operacional. O foco principal é também o que mais interessa ao cliente, que são: custo, qualidade e prazo. A posse destas informações possibilita a tomada de decisões para ações de melhorias destes quesitos entre outros tais como eficiência e eficácia.

Modelagem dimensional: Os autores desenvolveram cinco tabelas fato e onze tabelas de dimensão que compõem o DM. A tabela de DESEMPENHO explora basicamente a quantidade produzida, a eficiência e a eficácia, sumarizadas por data, célula, equipe, turno e produto. Nesta tabela, a eficiência é relativa ao aproveitamento da mão-de-obra, e a eficácia é relativa ao cumprimento dos prazos de entrega previstos nas ordens de fabricação. A Figura 3.6 ilustra o modelo.

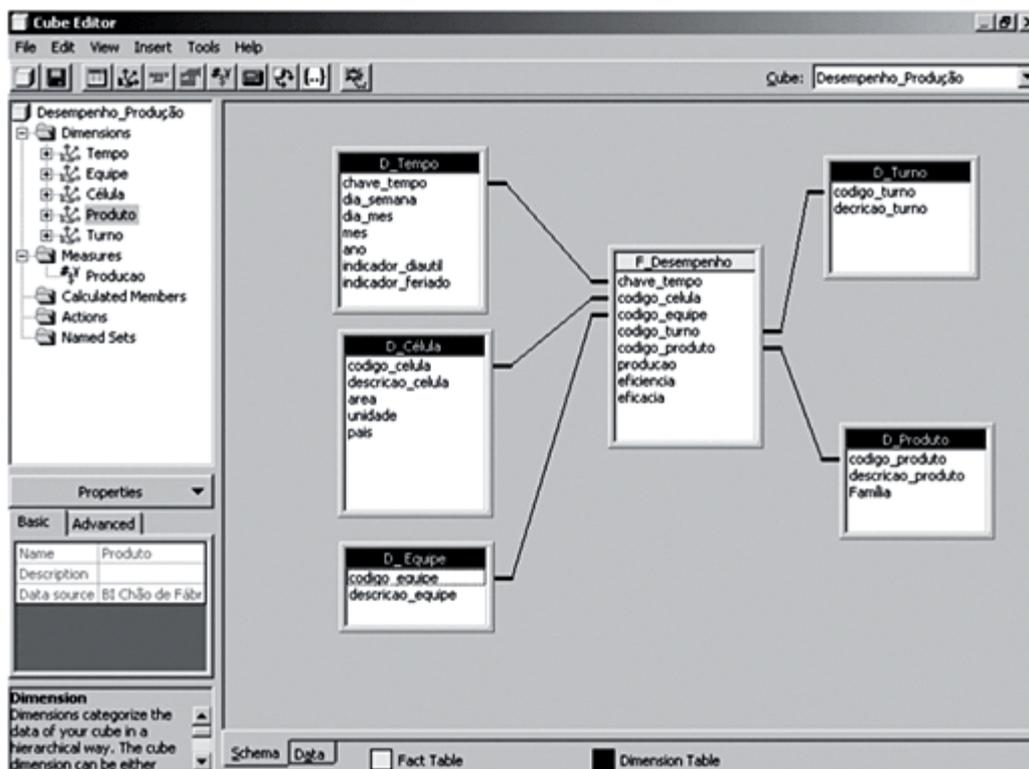


Figura 3.6 – Modelagem do cubo Desempenho
Fonte: (FORTULAN E GONÇALVES FILHO, 2005)

As demais tabelas fato de MEDIÇÕES, QUALIDADE, HORAS-PERDA e EVENTO, embora tenham sido implementadas e analisadas, seus diagramas não foram apresentados no referido trabalho. A seguir será apresentada uma síntese de cada uma.

MEDIÇÕES: Esta tabela tem a finalidade de verificar a conformidade do processo produtivo, e nela são registrados os valores de medições realizadas no processo produtivo, apresentando totalizações por dia de trabalho, máquina, produto e processo de medição.

QUALIDADE: Nesta tabela serão registrados os dados dos defeitos e sucatas verificados no processo produtivo, obtidos pelos testes ou inspeções. São informados os atributos que compõem aquela inspeção. Por exemplo, oxidação, trinca, queima, risco, não liga, etc. Os valores apontados são de quantidade de defeitos e defeitos por milhão (PPM), sumarizados por data, célula, equipe, turno, máquina, produto e tipo de defeito.

HORAS-PERDA: Nesta tabela devem ser apontadas todas as perdas de horas de produção ocorridas por causa de paradas, como manutenções e ajustes em máquinas e equipamentos. Além da hora-perda é apresentado também nesta tabela o tempo perdido.

EVENTO: Esta tabela tem a finalidade de receber o registro de todos os eventos que possam afetar os resultados de Medições, Qualidade, Desempenho ou Horas-Perda.

Processo de Extração, Transformação e Carga de Dados: No referido trabalho, o processo de ETC foi facilitado devido ao sistema de OLTP ter sido planejado com a mesma estrutura de tabelas do DM. A única tarefa nesta etapa foi de extrair os dados e importá-los no *Analysis Services* do MS SQL Server.

Desenvolvimento da aplicação: A aplicação foi desenvolvida no ambiente *Access 2000* e seus módulos de inserção de dados foram modelados de acordo com as tabelas fatos. Conforme citado anteriormente, o SGBD utilizado foi o *Microsoft Access 2000*. A ferramenta de OLAP utilizada foi o *Analysis Services*, parte integrante do produto *Microsoft SQL Server 2000 Enterprise*.

Resultados: Aplicando as ferramentas OLAP do *Analysis Services*, Os autores apresentaram alguns resultados obtidos com os cubos Desempenho e Eficiência.

Através do cubo Desempenho (figura 3.7), foi possível obter informações tais como:

- i. A unidade brasileira do grupo produz 2,79 vezes mais que a americana;
- ii. No Brasil, por sua vez, a unidade do Rio de Janeiro é maior, pelo menos em produção, que a de São Paulo; e
- iii. Dentre as áreas brasileiras, a Área D é a que está sendo mais utilizada, enquanto a Área B é a que está, teoricamente, mais ociosa.

- Pais	- Unidade	- Area	Descricao Célula	MeasuresLevel
All Célula	All Célula Total			307.013,00
	Brasil Total			254.468,00
		Rio de Janeiro Total		146.856,00
			Área D Total	90.298,00
			Célula Y01	21.724,00
			Célula Y02	38.256,00
			Célula Y03	30.318,00
			Área E Total	56.558,00
			Célula Z01	23.116,00
			Célula Z02	33.442,00
	São Paulo Total			107.612,00
			Área B Total	35.195,00
			Célula W01	20.533,00
			Célula W02	14.662,00
			Área C Total	72.417,00
+ USA	USA Total			52.545,00

Figura 3.7 – Cubo Desempenho
Fonte: (FORTULAN E GONÇALVES FILHO, 2005)

Sobre o cubo Eficiência, obteve-se indicadores apresentando o turno “A” como o turno de menor eficiência e o turno “C” apresentando um índice maior (Figura 3.8).

Figure 3.8 is a screenshot of the 'Cube Browser - Desempenho Eficiência' application. It features a filter section at the top with dropdown menus for 'Produto' (All Produto), 'Equipe' (All Equipe), 'Tempo_DiaSemana' (All Tempo_DiaSemana), and 'Measures' (Eficiência Média). Below the filters is a pivot table with the following data:

	Descricao Turno	Turno A	Turno B	Turno C
+ Pais	All Turno	79,1	80,7	81,1
All Célula		80,3		
+ Brasil		78,8	81,1	81,6
+ USA		81,0	79,2	78,6

Annotations in the image include circles around the 'All Célula' row, the 'Brasil' and 'USA' rows, and the 'Turno A', 'Turno B', and 'Turno C' columns. Two callout boxes at the bottom point to the data: 'Comparação de turnos entre países' points to the 'All Célula' row, and 'Comparação entre turnos' points to the 'Turno A', 'Turno B', and 'Turno C' columns. The interface also includes a 'Close' button and a 'Help' button.

Figura 3.8 – Cubo Eficiência
Fonte: (FORTULAN E GONÇALVES FILHO, 2005)

Os demais resultados não foram apresentados no referido trabalho, mas, de acordo com os autores todos foram implementados e experimentados.

3.3.2 Considerações do Autor

Fortulan e Gonçalves Filho destacam o grande volume de dados que um chão-de-fábrica proporciona, inviabilizando a capacidade do gerente obter informações relevantes e em tempo hábil.

“Por meio dos exemplos dos cubos montados, é possível verificar a facilidade e flexibilidade de tais sistemas de BI, conforme prometido na literatura. Sem dúvida, este sistema trará ao gerente um bom suporte para a tomada de direções em busca de melhores resultados para a sua área.” (FORTULAN E GONÇALVES FILHO, 2005, p. 11)

Até o presente capítulo, foi verificado o conceito teórico de BI e algumas aplicações práticas relatadas por diversos autores. Pretende-se a partir do próximo capítulo apresentar o estudo de caso de um clube desportivo, que é o principal foco deste trabalho, como também apresentar as etapas mais importantes de desenvolvimento de um DM e a aplicação de ferramentas OLAP, utilizando a suíte *Microsoft SQL Server 2005*.

4 CLUBE DESPORTIVO

O desenvolvimento de um ambiente DM juntamente com a aplicação de ferramentas OLAP, possibilitam criar repositórios de grande volume de dados que servem como fontes de consulta para realização de cálculos, cruzamentos e análises complexas, de maneira sintética, rápida e concisa. A possibilidade de criar nichos de consultas com bases de dados interligadas atrai administradores e líderes empresariais, que procuram através do conhecimento de seus produtos e clientes, vantagem competitiva.

A principal motivação deste trabalho compreende em apresentar uma proposta que facilite aos departamentos de *marketing* e estratégico do clube, à realização de análises dos dados de seu quadro de sócios utilizando recursos que auxiliem este processo e que contribua para a tomada de decisões. O desenvolvimento de um DM e a utilização de ferramentas OLAP visam facilitar a compreensão destes dados históricos, buscando traçar perfis de seus associados.

Este capítulo tem por finalidade apresentar, de maneira simplificada, o funcionamento do departamento administrativo de um clube desportivo da área futebolística, descrevendo os principais processos que ocorrem diariamente neste setor, baseado no *software* de administração de sócios Society, desenvolvido pela empresa Birô Processamento de dados Ltda⁷. Os demais *softwares* utilizados por estes setores possuem funções auxiliares e serão referenciados superficialmente por não tratarem-se do foco deste trabalho. Destaca-se neste quesito o *software* de controle de acessos e o *software* de controle de estacionamento, ficticiamente denominados Acesso e Estaciona.

⁷ Birô Processamento de dados Ltda é uma *softwarehouse* que desenvolve sistemas de ERP para clubes sociais, desportivos e náuticos. Localizada à Rua Júlio de Castilhos, 405 – sala 405 – Centro – Novo Hamburgo/RS 93510-130.

Pretende-se também apresentar uma proposta de construção de um DM com aplicação das ferramentas OLAP utilizando a suíte *Microsoft Analysis Server 2005*.

O Clube Campeão (nome fictício), é um clube de futebol que possui uma quantidade muito elevada de sócios torcedores e por conseqüência, possui uma estrutura bastante grande e complexa para atender as expectativas que um sócio espera de seu clube futebolístico.

Ao longo dos anos, o Clube Campeão acumulou títulos e vitórias elevando ainda mais o seu status fazendo com que vários torcedores aderissem ao seu quadro social. Tal fato estimulou os dirigentes e administradores à adesão de iniciativas de maior rigor frente ao controle e manutenção de seu quadro social. Atualmente o clube possui sistemas semi-integrados que realizam e automatizam estes controles atendendo plenamente as demandas do clube.

As rotinas de manutenção, controle dos cadastros dos sócios, assim como cobranças de mensalidades, acesso ao estádio e estacionamento são realizadas utilizando sistemas automatizados que auxiliam nestes processos.

Existem basicamente três áreas críticas de controle dos sócios e torcedores em geral (não-sócios). Estas áreas são: Controle e atendimento ao associado, Acesso ao estádio e o Estacionamento. Serão apresentadas as principais características de cada um:

4.1 Controle e atendimento ao associado

Para o torcedor associar-se, é necessário dirigir-se ao Departamento de Atendimentos ao Sócio para que um atendente realize o seu cadastro em um sistema comercial sob medida para controle de sócios. Neste aplicativo são informados os dados pessoais do sócio, modo de cobrança e uma categoria social. Cada categoria possui um valor de mensalidade específico. De acordo com a categoria escolhida, o sócio recebe uma carteirinha de sócio que lhe concede o direito a uma acomodação em diferentes partes do estádio para assistir aos jogos. O Sócio permanecendo em atividade no quadro social terá uma tarifa mensal a pagar de acordo com a categoria aderida. Quando o sócio possui ativo um serviço de cadeira, a mesma também gerará um custo mensal que será cobrado juntamente com a mensalidade social.

Além dos procedimentos de cadastro social, ocorrem neste departamento a venda de ingressos dos jogos (para sócio e não-sócios), 2ª via de carteirinha, alterações cadastrais, troca

de situação social, entre outros eventos administrativos. Todos estes procedimentos são registrados no cadastro do sócio.

Cada sócio e não-sócio cadastrado no clube, pode permanecer em quatro situações distintas referente a situação social do associado e que tornam-se importante destacá-las. Estas situações definem qual o status social de cada sócio. As situações existentes são:

- Atividade: situação social que define como sócio ativo. Sócio nesta situação paga mensalidade normal e tem direito a usufruir de todos os benefícios que o clube oferece de acordo com a sua categoria. São sócios em atividade plena no clube.
- Isenção: situação social que define como sócio isento de mensalidade. Nesta situação social, subdivididas por código, é possível identificar qual o motivo da isenção. As situações de isenção mais utilizadas são de licença militar, licença hospitalar, licença de viagem entre outras. Sócios nesta situação não têm o direito de acesso aos jogos e estão livres de cobrança de mensalidade social.
- Afastamento: situação social que define afastamento do quadro social. Sócios afastados estão livres de cobrança de mensalidade social e não tem acesso aos jogos. As situações de afastamento mais utilizadas são para os casos de afastamento por fim de prazo, afastamento por infringir o estatuto, entre outros.
- Demissão: situação que define demissão do quadro social. Através desta ocorrência, é possível informar qual o motivo da demissão. As situações de demissão mais utilizadas são demissão por falta de pagamento, por solicitação e por falecimento.

4.2 Acesso ao estádio

Em dias que antecedem o evento, as bilheterias do clube são abertas antecipadamente ao jogo para a venda de ingressos. Existem guichês separados para vendas de ingressos entre sócios e não-sócios. Os ingressos vendidos para os sócios são adicionados em seu histórico

cadastral. Sócios que possuem acomodações de cadeira não precisam comprar ingresso porque sua modalidade garante o acesso sem ingresso.

Torcedores sócios e não-sócios portadores de ingressos ou que possuem cadeiras cativas acessam o estádio por portões espalhados ao redor do mesmo. Todos os portões são identificados e cada um dá acesso a um local específico do estádio. Para validar o ingresso, existem catracas que fazem a validação dos ingressos e carteiras. Carteiras com dívidas ou outro impedimento social (afastamento ou demissão) além de ingressos inválidos são barrados e seu acesso não é liberado. Todos os registros de acessos são computados e armazenados em um SGBD.

4.3 Acesso ao estacionamento

O controle do estacionamento é realizado validando a carteirinha do sócio através do sistema de controle de acesso ou emitindo tickets que valem como comprovantes para aqueles não-sócios. Qualquer torcedor tem direito ao estacionamento, desde que ainda hajam vagas.

Diante desta realidade, ocorre uma lacuna muito importante do ponto de vista estratégico ocasionando dificuldades ao administrador. As dificuldades atualmente encontradas concentram-se basicamente na falta de informações sumarizadas e condizentes no que diz respeito ao perfil dos seus sócios. Como exemplo pode-se citar a inexistência de índices que apontam para inadimplência dos sócios, agrupadas por região, categoria ou por faixa etária. Os relatórios estáticos atualmente disponíveis atendem parcialmente estas necessidades. A manipulação e análises de dados gerenciais utilizando planilhas eletrônicas tornam-se pouco úteis devido a fragilidade e limitação recursiva destes produtos.

Outra dificuldade encontrada e que possui grande importância para as áreas administrativas e de Marketing, apontam para a sazonalidade de seus associados. Também pretende-se descobrir qual a quantidade de sócios cômputos espalhados pelo mundo e sua importância no quadro social referente aos sócios provenientes da mesma região. Sobre os acessos, pretende-se descobrir quais as faixas de tempo nas quais os sócios acessam o estádio antes do início da partida.

A proposta de desenvolvimento de um DM vem ao encontro destas dificuldades devido a sua enorme capacidade de prover recursos e ferramentas de apoio à tomada de decisões. Ferramentas OLAP disponíveis no mercado oferecem capacidades que poderão

contribuir dispondo de dados que possibilite ao *Decision Maker* traçar um perfil mais qualificado dos sócios

Com base neste cenário, é proposta a construção de um DM focado inicialmente no cadastro dos torcedores e nos acessos aos eventos. Estes dados são extraídos dos SGBD's existentes no clube e devem auxiliar na descoberta das necessidades apontadas anteriormente. O DM deverá ser composto de 2 tabelas fato, denominadas Torcedor e Acesso. As tabelas de dimensão serão compreendidas em: Tempo, FaixaEtaria, FaixaHora, Sexo, Categoria, SSocial, Município, Cobrador, Evento e Portão.

4.4 Tabelas de fatos

O modelo é composto de duas tabelas de fatos: TORCEDOR e ACESSO. Estas tabelas representam os dados cadastrais e financeiros do torcedor e seus acessos aos eventos, respectivamente. É detalhado a seguir a estrutura de cada tabela.

4.4.1 Fato TORCEDOR

A tabela de fatos TORCEDOR contém informações sobre os dados cadastrais tais como categoria, situação social, situação financeira, além de informações de localização e distância entre o clube e sua residência. Sua principal composição é de chaves estrangeiras que permitem relacionar com as dimensões descritas na próxima seção. Segue abaixo explicação dos atributos que compõem esta tabela.

ID_TORCE – Chave primária (*Primary Key*) da tabela de fato Torcedor.

ID_TEMPO – Código da data que foi verificado os dados do torcedor.

ID_SEXO – Código do sexo do torcedor. Este código quando apontado para o sexo Jurídico, representa as empresas e parceiras de pessoa jurídica que encontram-se cadastradas.

ID_FETARIA – Código da faixa etária do torcedor.

ID_CATEGO – Código da categoria do torcedor.

ID_SSOCIAL – Código da situação social do torcedor.

ID_MUNIC – Código do município do torcedor.

ID_COBRA – Código do cobrador das mensalidades definido pelo torcedor (o torcedor define qual o banco ele deseja que suas mensalidades sejam cobradas).

CONSUL – Atributo lógico que define se o torcedor é CÔNSUL.

VLRPREVI – Medida que determina o valor previsto para pagamento no referido mês pelo torcedor.

VLRREALI – Medida que determina o valor realizado pelo torcedor no referido mês.

ADIMPLENTE – Medida que determina se o torcedor é adimplente no referido mês. É considerado adimplente aquele torcedor que não possuir nenhuma prestação em aberto ao final do referido mês.

4.4.2 Fato ACESSO

A tabela de fatos ACESSO contém informações sobre os acessos aos eventos realizados no estádio. Constituem ainda esta tabela, informações de localização de portão, situação social e município do torcedor. A seguir é apresentada uma explicação sobre seus atributos:

ID_ACESSO – Chave primária da tabela de fato Acesso.

ID_EVENTO – Código do evento.

ID_TEMPO – Código do tempo que ocorreu o acesso.

ID_SEXO – Código do Sexo do torcedor.

ID_FETARIA – Código da faixa etária do torcedor.

ID_CATEGO – Código da categoria do torcedor.

ID_SSOCIAL – Código da situação social do torcedor.

ID_MUNIC – Código do município do torcedor.

ID_FHORA – Código da faixa de hora que o torcedor acessou o estádio em relação ao início do evento.

ID_PORTAO – Código do portão que o sócio realizou o acesso.

CONSUL – Atributo lógico que define se o torcedor é Cônsul.

4.5 Tabelas de dimensões

Conforme descrito na seção 2.2 a função das tabelas de dimensão é reunir atributos que serão utilizados para qualificar as consultas. As tabelas fatos estão diretamente ligadas às dimensões descritas a seguir. Estas dimensões compreendem dez tabelas assim denominadas: TEMPO, SSOCIAL, SEXO, PORTAO, MUNICIPIO, FAIXAHORA, FAIXAETARIA, COBRADOR, CATEGORIA e EVENTO.

4.5.1 Dimensão TEMPO

O tempo é uma das dimensões mais importantes de uma modelagem dimensional por permitir que as análises dos dados obtidos sejam feitas em uma linha do tempo. A dimensão tempo criada neste modelo foi baseada num período verificação dos dados do cadastro dos sócios (mensais) e nas datas que ocorreram os acessos aos eventos. A seguir é apresentada uma explicação sobre seus atributos:

ID_TEMPO - Chave primária da tabela de dimensão Tempo.

DATA – Data armazenada no formato dd/mm/aaaa (dia/mês, ano)

MÊS – Descrição do mês.

BIMESTRE – Descrição do bimestre do ano.

TRIMESTRE – Descrição do trimestre do ano.

SEMESTRE – Descrição do semestre do ano.

ANO – Número do ano.

4.5.2 Dimensão SSOCIAL

Esta dimensão foi criada para atribuir os códigos da situação social de cada sócio. A explicação sobre seus atributos é apresentada a seguir:

ID_SSOCIAL - Chave primária da tabela de dimensão SSOCIAL.

DESCRICA0 – Descrição da situação social.

4.5.3 Dimensão SEXO

A dimensão Sexo foi criada para permitir identificar o sexo dos torcedores. A Estrutura da tabela dimensão Sexo é apresentada abaixo:

ID_SEXO – Chave primária da tabela de dimensão Sexo.

DESCRICA0 – Descrição do sexo.

4.5.4 Dimensão PORTAO

A dimensão Portao foi criada para permitir identificar qual portão do estádio ocorreu o acesso do torcedor. A estrutura da tabela dimensão Portão é apresentada abaixo:

ID-PORTAO – Chave primária da tabela de dimensão Portao.

DESCRICA0 – Descrição do portão de acesso do estádio.

4.5.5 Dimensão MUNICIPIO

A Dimensão Município foi criada para permitir identificar o município residencial do torcedor como também identificar qual a distância entre a cidade de origem e o clube. A lista dos municípios abrange todo o Brasil, além de algumas cidades do exterior onde existem sócios torcedores. A estrutura da tabela dimensão Município é apresentada abaixo:

ID_MUNIC - Chave primária da tabela de dimensão Município.

MUNICIPIO – CEP do município.

NOME – Descrição do município.

UF – Unidade Federativa do município.

DISTANCIA – Distância em KM entre o município do sócio torcedor e a sede do clube.

REG_FUNCIO – Grupamento da região funcional que pertence o município.

REGIAO – Região funcional que pertence o município.

CORED – Número do CORED⁸ do município.

POPULACAO – população estimada do município.

4.5.6 Dimensão FAIXAHORA

Esta tabela tem a finalidade de identificar qual a faixa de tempo (em minutos) que o sócio ingressou no portão que dá acesso ao estádio. A seguir é apresentada uma explicação sobre seus atributos:

ID-FHORA - Chave primária da tabela de dimensão Faixahora.

DESCRICA0 – Descrição dos intervalos de faixas de horários.

4.5.7 Dimensão FAIXAETARIA

A Dimensão Faixaetaria foi criada para permitir identificar a faixa etária (em anos) do sócio torcedor. A estrutura da tabela dimensão FaixaEtaria é apresentada abaixo:

ID_FETARIA – Chave primária da tabela de dimensão Faixahora.

DESCRICA0 – Descrição dos intervalos de faixas de idades.

4.5.8 Dimensão COBRADOR

Esta dimensão é onde poderá ser consultado o cobrador de cada sócio torcedor. A estrutura da tabela dimensão Cobrador é exibida abaixo:

ID_COBRA – Chave primária da tabela de dimensão Faixahora

COBRADOR – Código do cobrador.

NOME – Nome do cobrador.

⁸ CORED = Conselhos Regionais de Desenvolvimento

4.5.9 Dimensão CATEGORIA

A Dimensão Categoria tem a finalidade de apresentar a categoria social o sócio torcedor. A estrutura da tabela dimensão Categoria é apresentada abaixo:

ID_CATEGO – Chave primária da tabela de dimensão Categoria.

CATEGORIA – Código da categoria.

DESCRICA0 – Descrição da categoria.

PAGANTE – Descrição se categoria é pagante.

TIPOSOCIO – Descrição do tipo de sócio.

4.5.10 Dimensão EVENTO

Esta dimensão tem como objetivo apresentar os eventos ocorridos na sede do clube. Todos os eventos analisados são partidas de futebol que ocorreram com times adversários nos mais diversos campeonatos, torneios e amistosos que o clube participou. A estrutura da tabela dimensão Evento é descrita abaixo:

ID_EVENTO – Chave primária da tabela de dimensão Evento.

EVENTO- Descrição do Evento.

4.6 Modelagem dos dados (cubos)

As sugestões de criação das tabelas de fatos e dimensões apresentadas anteriormente basearam-se nas necessidades apontadas pelos departamentos de Marketing e administrativo do clube. A partir dos requisitos levantados, descobriu-se quais tabelas iriam compor o DM. Elas foram criadas e populadas especialmente para utilização no DM. O processo de ETC será explicado a partir da próxima seção.

O tipo de modelagem utilizada foi o *Star Schema*, também conhecido por modelo estrela. Como já visto na seção 2.3, este modelo possui semelhança à uma estrela, sendo que a tabela de fatos fica no centro, rodeada de suas dimensões.

No modelo proposto, agregadas às duas tabelas de fatos estão as tabelas de dimensão conforme Figura 4.1.

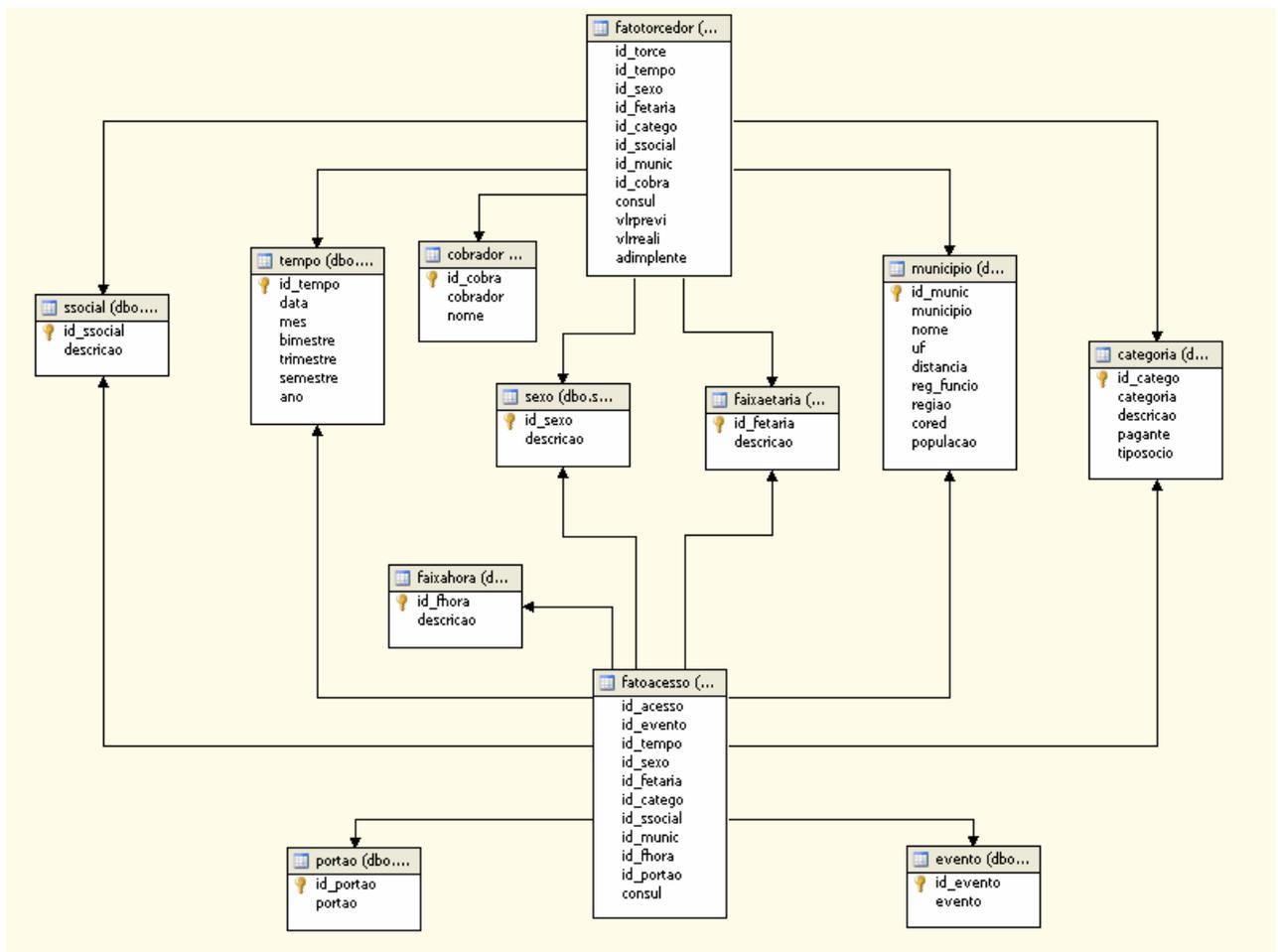


Figura 4.1 – Cubo Clube desportivo

4.7 ETC – Extração, Transformação e Carga de Dados.

Conforme descrito na seção 1.3, é necessário que durante a construção de um DW (neste caso um DM), ocorra o processo de ETC. Sua finalidade é fazer uma correta extração dos dados de seus locais de origem (sistemas OLTP) para posteriormente criar uma única convenção, transformando os dados originais em uma tabela de destino padronizada.

De acordo com Machado, (2000) o ETC representa um armazenamento intermediário dos dados, oferecendo facilidade na integração dos dados dos sistemas legados antes da sua atualização no DW. Este processo de ETC também é conhecido por ODS (*Operational Data Storage* ou *staging Area*).

O processo de montagem do DM, parte integrante deste trabalho, foi desenvolvido baseado nas tabelas dos sistemas OLTP utilizados pelo clube citados na seção 4. Os dados

destes sistemas estão armazenados no banco de dados nativo da linguagem *Visual Fox Pro 9.0*. Esta linguagem disponibiliza seu banco de dados em arquivos do tipo DBF (*database file*).

A extração foi feita através de comandos da linguagem SQL juntamente com rotinas e comandos da linguagem *Visual Fox Pro 9.0*. As funções de apoio utilizadas nos scripts são partes integrantes da biblioteca de funções do ERP Society. Os dados coletados sofreram modificação em seus conteúdos para que não houvesse revelação de informações confidenciais ou proprietárias relativas ao clube estudado.

A carga de dados no DM será sempre total, ou seja, a cada nova carga, todos os dados existentes do DM são excluídos, sendo populados novamente na próxima carga. Para realização das primeiras análises, a extração dos dados foi realizada entre 01/01/2004 a 31/12/2006. O período proposto para extração é específico para aplicação neste trabalho. Pretende-se quando implantado, utilizar um período mais extenso para que o conjunto de dados coletados esteja mais amplo e conseqüentemente mais próximo do real.

Os dados coletados foram tratados e armazenados em tabelas do tipo DBF. Os *scripts* utilizados para extração e transformação estão disponíveis no capítulo 6. Todas as tabelas que foram criadas para utilização no DM estão disponibilizadas em uma pasta do disco rígido denominada C:\ETC. Após a extração, foi criado um SGBD no *SQL Server 2005*. Todas as tabelas foram importadas para este banco de dados, preservando a mesma estrutura e conteúdo.

Pretende-se apresentar uma breve explicação do produto *Microsoft SQL Server 2005 - Analysis Services*, como também a carga dos dados e a construção do DM neste ambiente. Por fim será apresentada a verificação dos resultados obtidos.

5 APLICANDO A FERRAMENTA OLAP DO *MICROSOFT ANALYSIS SERVICES*

Este capítulo tem por finalidade, apresentar uma aplicação do DM utilizando como base a estrutura do DM descrito no capítulo anterior. A ferramenta proposta para aplicação é o *Analysis Services* recurso integrante do produto *Microsoft SQL Server 2005*. A escolha desta ferramenta deve-se ao fato da empresa mantenedora do ERP Society ser empresa parceira e consumidora dos produtos *Microsoft* e também pela preferência do próprio cliente. Será apresentada uma breve explicação de suas características seguida das principais etapas para a construção do DM.

A *Microsoft* lançou em novembro de 2005 o SGBD *SQL Server 2005* com a finalidade de suceder o *SQL Server 2000*. Nesta nova versão, ocorreram mudanças significativas em sua estrutura, sendo agregadas mais funcionalidades e aprimoradas outras já existentes na versão 2000. Com estes novos recursos, é possível construir e implementar aplicações corporativas mais confiáveis e seguras. Além disso, o *SQL Server 2005* permite compartilhar dados entre múltiplas plataformas para facilitar a conexão entre sistemas internos e externos.

“O *Microsoft SQL Server 2005* redefine completamente a plataforma de banco de dados *SQL Server* e fornece a fundação sobre as quais organizações grandes, médias e pequenas podem construir sua próxima geração de infra-estrutura de TI” (STANEK, 2006, p.29). Agregado a este produto são oferecidos os seguintes serviços e ferramentas: *SQL Server Database Services*, *Data Integration Services*, *Notification Services*, *Reporting Services*, *Service Broker* e o *Analysis Services* que será detalhado na seção 5.1.

O *SQL Server 2005* oferece vários tipos de ferramentas gráficas para a administração. O *SQL Server Management Studio* (Figura 5.1) é uma das principais ferramentas. Ela apresenta várias visões que podem ser acessadas tais como o *object Explorer*,

que permite visualizar e conectar-se aos diferentes serviços como também o *Solution Explorer* que permite acesso rápido aos projetos existentes dos diferentes serviços.

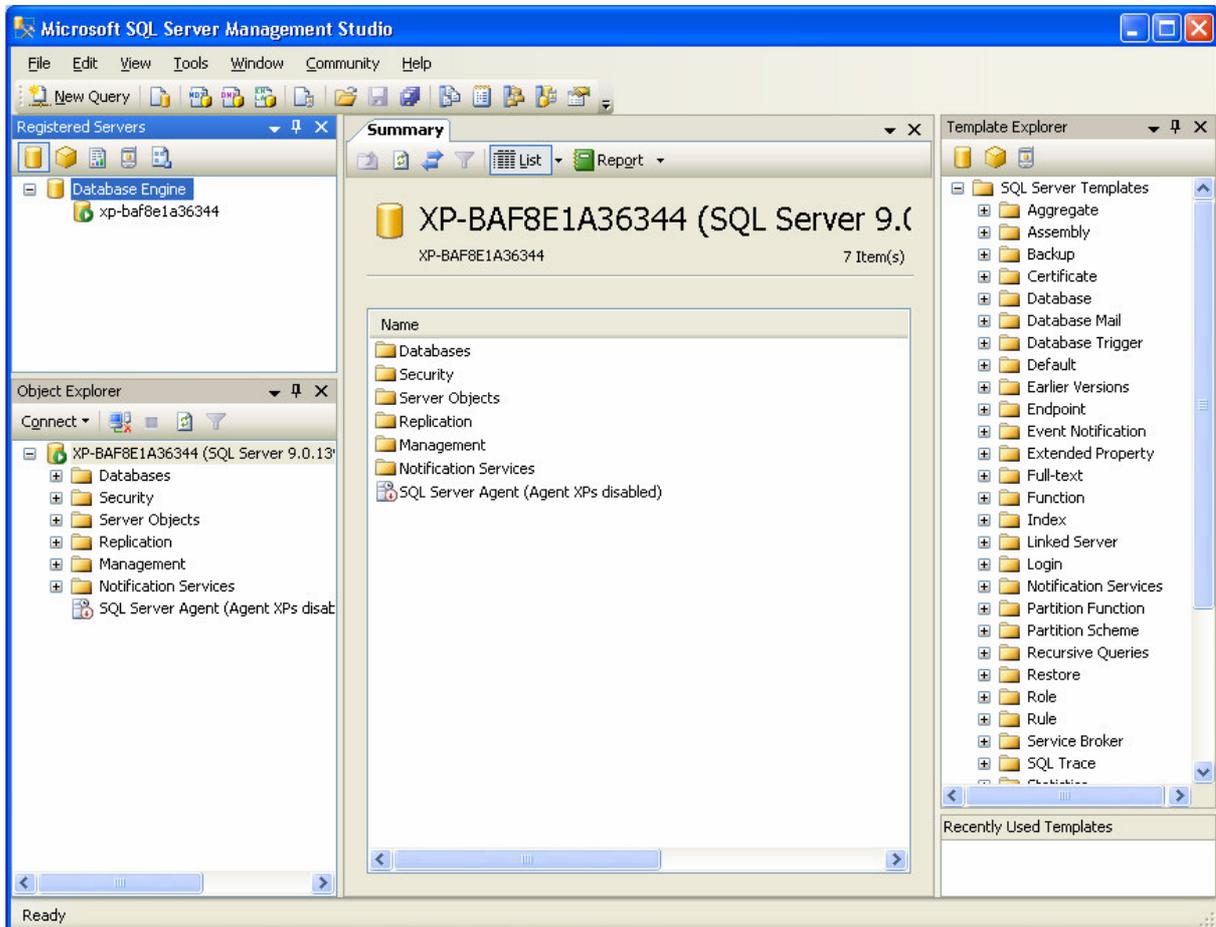


Figura 5.1 – *Management Studio* – Recursos visuais facilitam a administração.

5.1 *Analysis services*

Como visto anteriormente, o *Analysis Services* é uma ferramenta integrante do SQL Server e tem a finalidade de auxílio ao desenvolvimento de soluções de BI. De acordo com Jacobson, Misner e Hitachi Consulting (2007), ela fornece suporte a banco de dados multidimensionais e exploração de dados além dos recursos especiais de OLAP que tornam os dados do DW facilmente acessíveis.

De acordo com Rittman, ([s.d.], p. 63), o servidor OLAP versão 2000 da Microsoft já era líder de mercado com 23% do total de soluções implementadas. Ainda segundo o mesmo autor, este número certamente foi superado logo no lançamento da versão 2005.

Para Rittman, uma das principais novidades do *Analysis Services* é o conceito do UDM (*Unified Dimensional Model*), que traduzido significa: Modelo Dimensional Unificado.

O UDM combina os melhores recursos dos modelos de dados relacionais e OLAP, servindo para estreitar a linha de separação entre os bancos de dados relacionais e multidimensionais. Ele permite que DBAs e desenvolvedores definam dimensões, cubos OLAP e hierarquias de forma “abstrata” que possam subsequencialmente ser armazenados em bases relacionais e multidimensionais.

O *Analysis Services* fornece todos os recursos para o desenvolvimento de projetos grandes, médios e pequenos, sejam eles um único DM ou um complexo DW. Através de recursos gráficos, é possível criar a conexão com os mais variados tipos de SGBD, não necessariamente provenientes de versões do *SQL Server*. Com um *Data Source*⁹ definido, é possível a construção de tabelas de fatos e dimensões, além de desenvolvimento de medidas hierarquias.

Baseado nestas informações será apresentado na próxima seção os principais passos para a construção do DM do clube desportivo.

5.2 Plataforma e configuração do ambiente

Para utilização do *Analysis Services*, é necessário que ocorra a instalação do componente denominado *Analysis Server* que acompanha o pacote de instalação do *SQL Server*. A parte de instalação e configuração do ambiente não faz parte do escopo deste trabalho, entretanto sua execução é necessária.

Com o *Analysis Services* devidamente instalado e configurado, é possível invocar o gerenciador gráfico de BI, denominado *Business Intelligence Development Studio* (Figura 5.2). Conforme verificado na figura, este ambiente é uma cópia do *Visual Studio 2005*, com os *designers* de BI instalados em vez de *designers* específicos para desenvolvimento de aplicativos tais como C#.NET e VB.NET. Para instalação do *SQL Server* e do *Analysis Server*, foram seguidos os passos contidos em (STANEK, 2006).

A integração entre diferentes ambientes de desenvolvimento ocorre graças aos recursos que o *.NET Framework* oferece. Através destas funcionalidades, é possível a construção de um único projeto, mesmo tratando-se de um ambiente de grande escala e que

⁹ *Data Source* é o parâmetro responsável pelo armazenamento das informações sobre a conexão e o caminho lógico com sua base de dados.

neste projeto possibilite agregar o trabalho de vários desenvolvedores de forma integrada e produtiva.

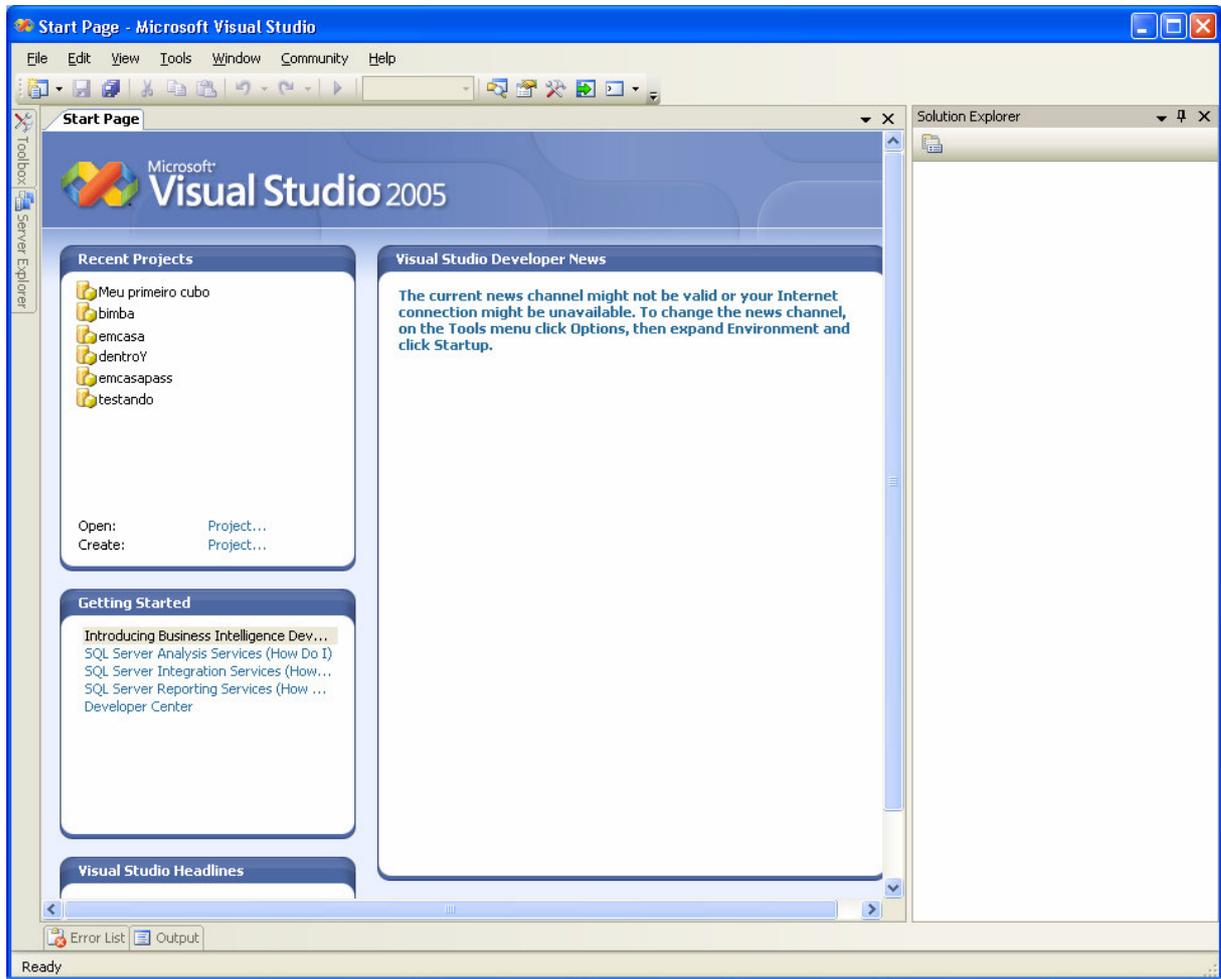


Figura 5.2 – *Business Intelligence Development Studio.*

Foi criado um projeto denominado ClubeCampeao. Após a sua criação foi realizada a conexão com o SGBD. Na propriedade *Data Source* do *solution explorer* foi criada a conexão com o banco de dados utilizando o Wizard. O processo de conexão com o banco de dados foi realizado apontando para o SGBD *SQL Server 2005*, onde está a base de dados denominada ClubeCampeao. Para criação do projeto no *Analysis Services*, foram seguidos os passos contidos em (JACOBSON, MISNER E HITACHI CONSULTING 2007).

Após a configuração do *Data Source*, foi realizada a criação do *Data Source Views*. Esta propriedade tem a finalidade de “apresentar” a estrutura das tabelas provenientes do *Data Source*. Neste ambiente é possível realizar as definições de parâmetros de campos e demais relacionamentos entre as chaves que possivelmente não tenham ocorrido na fase da modelagem do banco de dados. Durante a configuração deste parâmetro, o *Wizard* solicita a confirmação dos relacionamentos caso os mesmos devam ser mantidos. Para este modelo de

DM, foi selecionada a opção de manter os relacionamentos e os nomes das chaves estrangeira e primária. A figura 5.3 apresenta a definição dos dois parâmetros comentados.

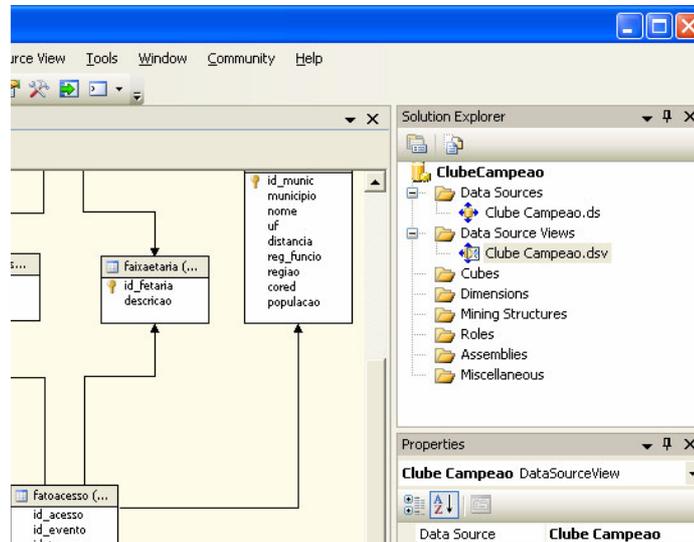


Figura 5.3 – Definição do *Data Source* e *Data Source View*.

Finalizada a definição do *Data Source* e *Data Source Views*, é possível criar os cubos e definir suas medidas e hierarquias.

5.3 Criando os cubos

A criação dos cubos ocorre na propriedade *CUBES* do *Solution Explorer*. Para este processo foi utilizado o *Wizard* disponível. Foram criados dois cubos utilizando as tabelas *FatoTorcedor* e *FatoAcesso*.

Os cubos criados receberam os nomes de *TORCEDOR* e *ACESSO*. Finalizada as criações, o editor insere as tabelas de cada cubo dentro de um *schema*, representando graficamente o ambiente do cubo montado, contendo suas tabelas de fato e dimensões.

Os cubos *TORCEDOR* e *ACESSO* foram criados em etapas separadas. Para criação dos cubos foram seguidos os passos contidos em (JACOBSON, MISNER E HITACHI CONSULTING 2007). As figura 5.4 e 5.5 apresentam os cubos criados.

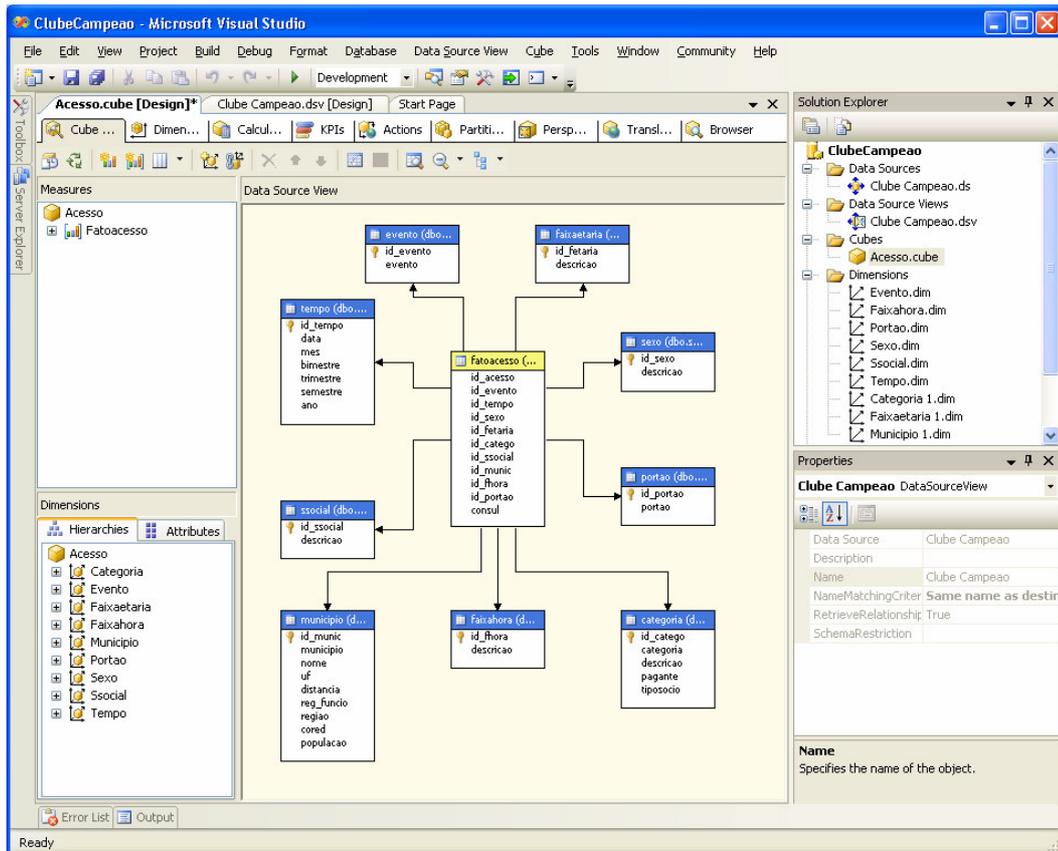


Figura 5.4 – Cubo Acesso.

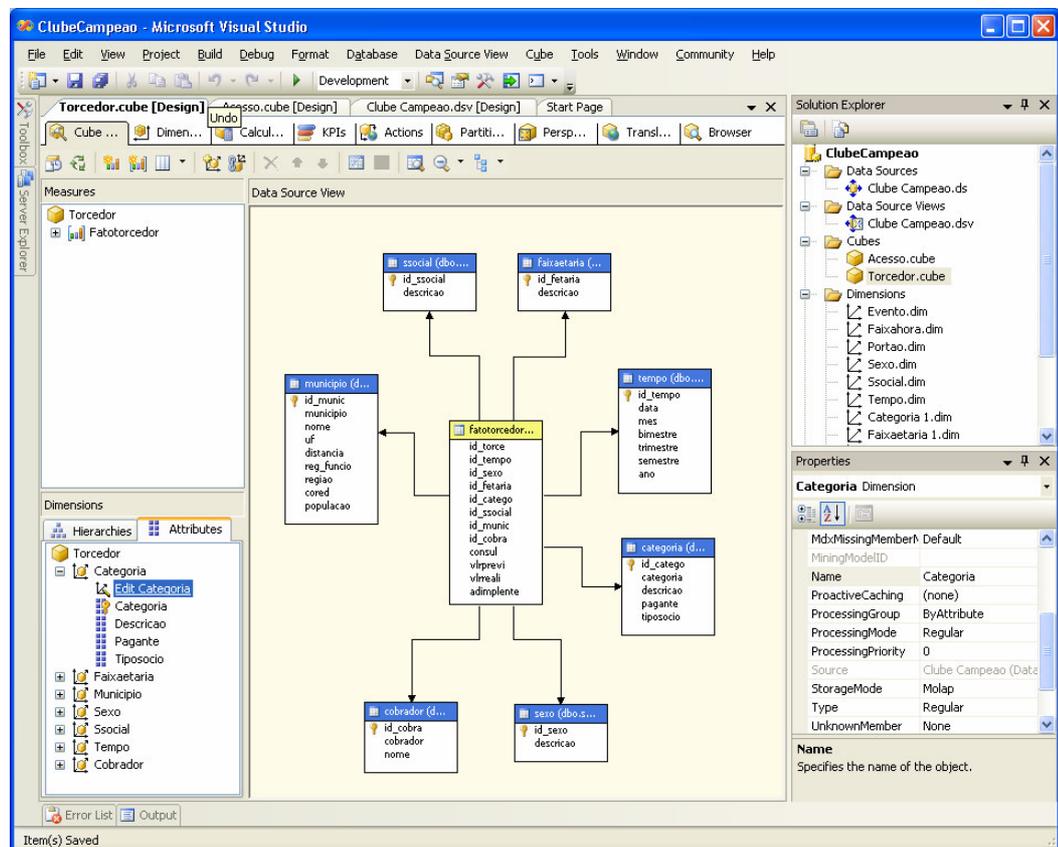


Figura 5.5 – Cubo Torcedor.

Os cubos TORCEDOR e ACESSO receberam definições de medidas e hierarquias descritas a seguir:

5.3.1 Criando as medidas

De acordo com Jacobson, Misner e Hitachi Consulting (2007), medida é a definição para qualquer valor numérico da tabela de fatos que possa ser resumido e que possa ser usado para monitorar o negócio.

A definição das medidas em um cubo é um fato importante porque é através delas que teremos os resumos das informações resultantes dos cubos criados.

Para o Cubo Acesso, foram criadas duas medidas onde uma delas quantifica os acessos de sócios aos eventos, e a outra quantifica os acessos dos cônsoles aos eventos, conforme apresentada na figura 5.6.

Para criação das medidas foram seguidos os passos contidos em (JACOBSON, MISNER E HITACHI CONSULTING 2007)

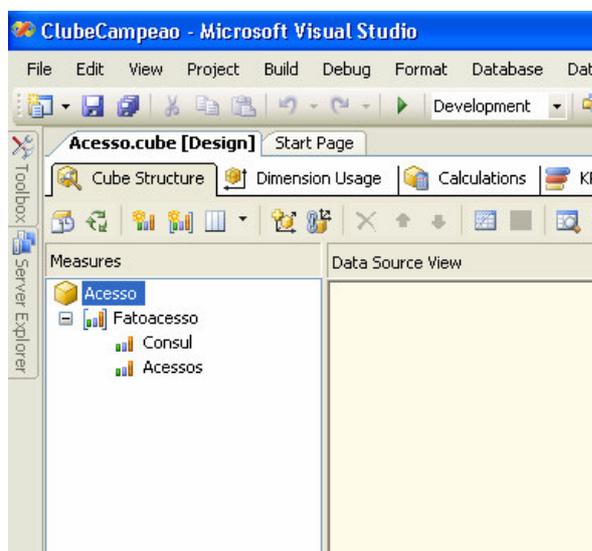


Figura 5.6 – Medidas criadas no cubo Acesso.

Para o cubo Torcedor, foram criadas cinco medidas conforme listadas a seguir:

SOCIOS – quantidade de sócios ativos no clube.

VALORPREVISTO – valor em moeda das prestações previstas para realização.

VALORREALIZADO – valor em moeda das prestações realizadas (pagas).

CONSULES – quantidade de cômsoles.

ADIMPLENTE – quantidade de sócios adimplentes.

A figura 5.7 apresenta as medidas criadas no cubo Torcedor.

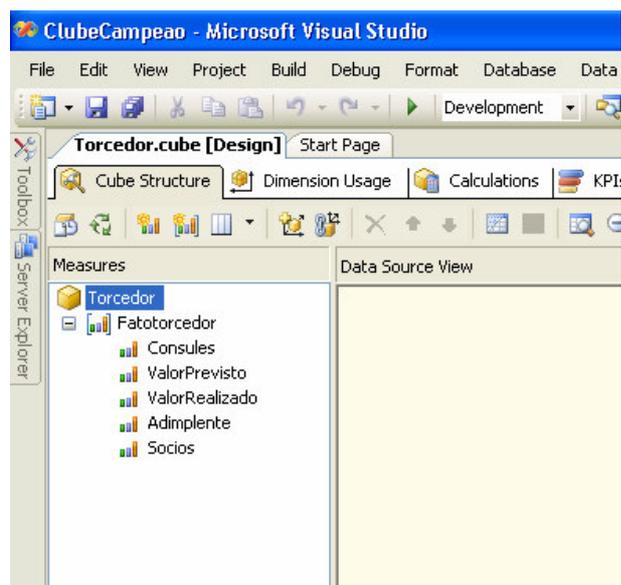


Figura 5.7 – Medidas criadas no cubo Torcedor.

5.3.2 Criando as hierarquias

As hierarquias são importantes para que as visualizações dos cubos tenham coerência lógica de cada dimensão existente. As definições das hierarquias consistem em definir os membros superiores considerados pais dentro da estrutura hierárquica.

Todas as hierarquias criadas nas tabelas de dimensões têm efeito em todas as suas tabelas de fato relacionadas. No DM do clube desportivo, por exemplo, as hierarquias construídas, e que a seguir serão descritas, foram utilizadas nos cubos Acesso e Torcedor. Para criação das hierarquias foram seguidos os passos contidos em (JACOBSON, MISNER E HITACHI CONSULTING 2007). A figura 5.8 apresenta um exemplo das hierarquias criadas na tabela de dimensão Tempo.

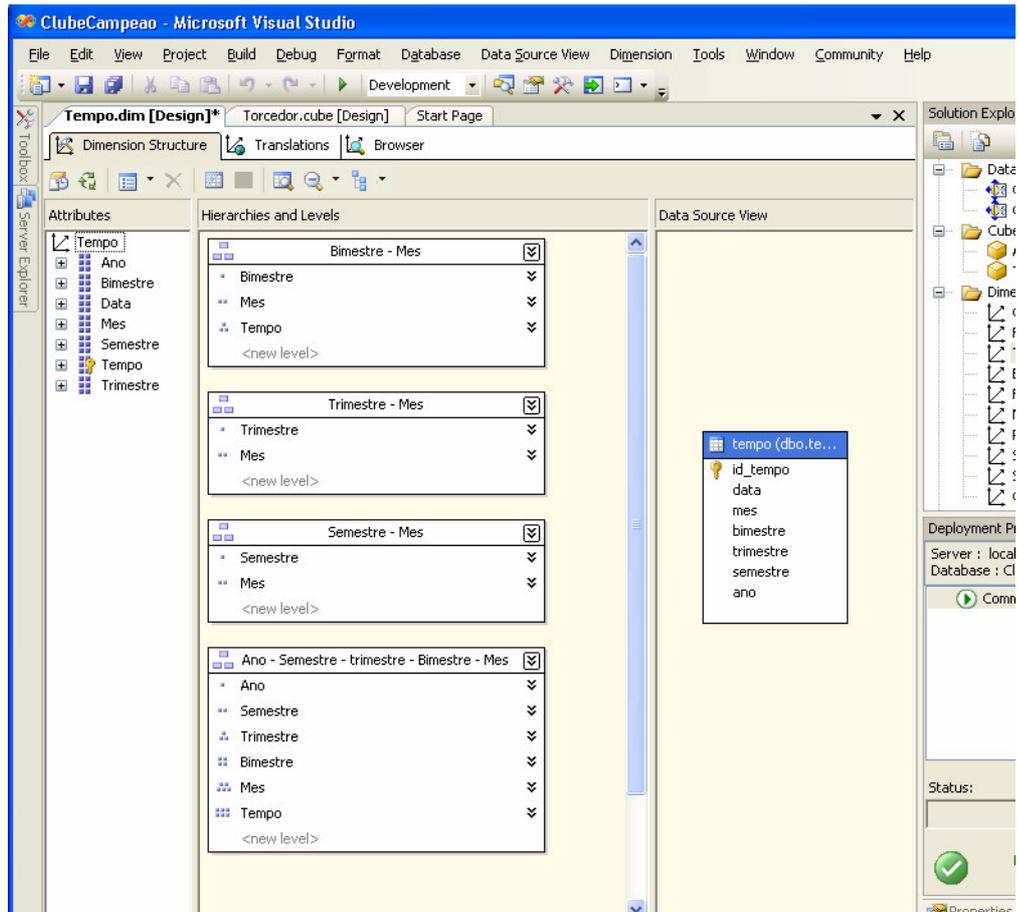


Figura 5.7 – Medidas criadas na tabela dimensão Tempo.

As hierarquias criadas para as demais tabelas de dimensões são apresentadas resumidamente na figura 5.8.



Figura 5.8 – Medidas criadas no DM ClubeCampeao.

Com as etapas descritas até o momento concluídas, é possível realizar a verificação dos resultados no DM construído. O próximo tópico abordará o modo que estas verificações são realizadas e apresentará alguns resultados obtidos.

5.4 Verificação dos resultados

No *Analysis Services*, a manipulação do cubo e a verificação dos resultados, ocorrem no ambiente denominado BROWSER. Neste ambiente é possível selecionar exatamente as dimensões e medidas desejadas. Esta operação é realizada com o arrasta-e-solta dos objetos sobre a planilha disponível. Nas colunas e linhas são inseridos os membros das tabelas de dimensões. No detalhe são inseridas as medidas da tabela de fato. A figura 5.9 ilustra a construção de uma consulta utilizando as dimensões de situação social e tempo com a medida sócios.

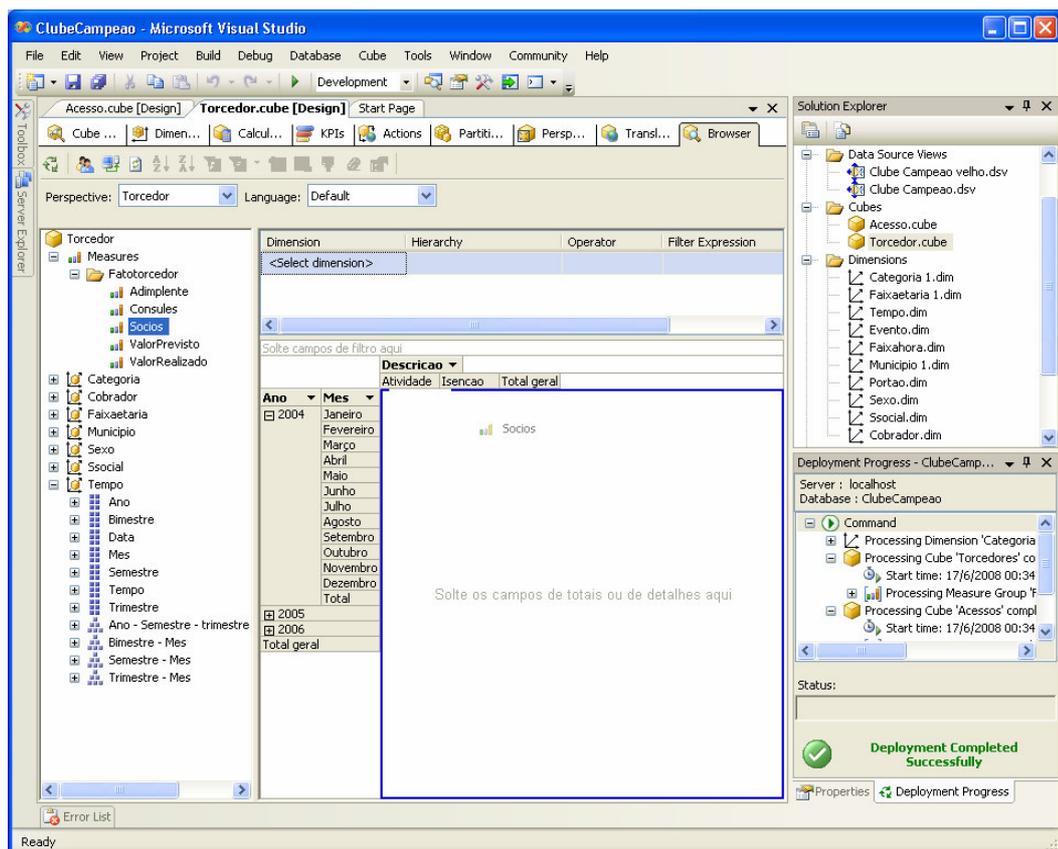


Figura 5.9 – Construção de uma consulta no cubo Torcedor.

De acordo com as consultas realizadas, foi possível verificar índices de grande valor estratégico como previsto no início deste trabalho. Como exemplo, podemos citar a quantidade de sócios em atividade social no ano de 2006, como apresentado na figura 5.10.

Ano	Semestre	Mes	Socios	Adimplente
2005	Primeiro	Janeiro	29113	20841
		Fevereiro	29489	21898
		Março	29883	22297
		Abril	31286	23929
		Maio	31604	23691
		Junho	31664	23404
	Segundo	Julho	32379	24820
		Agosto	33423	25644
		Setembro	33330	27205
		Outubro	38369	29876
		Novembro	36748	29770
		Dezembro	36761	28546
2006	Primeiro	Janeiro	33981	20886
		Fevereiro	34778	22243
		Março	39284	27309
		Abril	42326	32077
		Maio	43501	33710
		Junho	43798	32183
	Segundo	Julho	49290	40433
		Agosto	50725	41536
		Setembro	53013	42107
		Outubro	52799	42015
		Novembro	52945	41657
		Dezembro	54457	43288

Figura 5.10 – Sócios ativos entre os anos de 2005 e 2006.

Comparando o número de sócios ativos entre janeiro/2005 e dez/2006, é possível constatar um aumento de 87,05% do quadro social no mesmo período. O número de sócios adimplentes no mesmo período também aumentou de 72% para 79%.

Foi possível verificar também que a maioria dos sócios analisados estão com faixa etária entre 41 e 50 anos (15,57%) seguido da faixa etária de sócios acima de 60 anos (13,43%), de acordo com a figura 5.11.

Ano	Descricao												Total geral
	Socios	Socios	Socios	Socios	Socios	Socios	Socios	Socios	Socios	Socios	Socios	Socios	
2004	7,20%	10,45%	7,68%	6,58%	4,97%	6,59%	7,75%	8,44%	15,18%	9,16%	16,01%	100,00%	
2005	6,36%	9,57%	8,59%	6,09%	5,48%	7,92%	8,37%	8,40%	15,77%	8,90%	14,55%	100,00%	
2006	4,98%	6,69%	5,26%	8,19%	8,52%	12,34%	10,09%	9,28%	15,69%	8,03%	10,91%	100,00%	
Total geral	6,01%	8,61%	6,94%	7,11%	6,62%	9,41%	8,92%	8,78%	15,57%	8,61%	13,43%	100,00%	

Figura 5.11 – Faixa etária dos sócios entre o ano de 2004 e 2006.

Em outra combinação de dimensões no cubo Torcedor foi possível verificar a quantidade de Cónsules de uma determinada região e a importância deste cónsul dado o número de sócios que o mesmo delega. Esta consulta possibilita ao administrador uma melhor verificação do desempenho e também do potencial de adesão de novos sócios. É possível também verificar quais regiões estão com número de cónsul abaixo ou acima do ideal. A Figura 5.12 apresenta uma síntese desta consulta.

Coreid	Regiao	Nome	Consules	Socios
9			1,80%	0,14%
10			2,72%	0,57%
11			0,44%	0,06%
12			2,35%	0,34%
13			1,60%	0,24%
14			2,91%	0,42%
15			1,27%	0,56%
16			3,48%	1,05%
17			1,32%	1,31%
18			3,16%	0,45%
19	Vale do Rio dos Sinos	ARARICA	0,00%	0,00%
		CAMPO BOM	0,34%	0,20%
		CANOAS	0,34%	3,58%
		DOIS IRMAOS	0,34%	0,07%
		ESTANCIA VELHA	0,34%	0,12%
		ESTEIO	0,21%	0,76%
		IVOTI	0,29%	0,13%
		NOVA HARTZ	0,00%	0,01%
		NOVA SANTA RITA	0,34%	0,07%
		NOVO HAMBURGO	1,02%	1,17%
		PORTAO	0,34%	0,09%
		SAO LEOPOLDO	0,54%	1,38%
		SAPIRANGA	0,39%	0,30%
		SAPUCAIA DO SUL	0,34%	0,58%
		Total	4,86%	8,45%
	Total		4,86%	8,45%
20			1,81%	0,43%
21			3,27%	0,50%
22			19,09%	65,97%
23			0,69%	0,08%
24			0,79%	0,20%
25			0,67%	0,13%
26			1,37%	0,19%
Total geral			100,00%	100,00%

Figura 5.12 – Percentual de sócios e consules por região.

As consultas realizadas no cubo Acesso permitiram verificar informações importantes quanto aos tipos de sócios que frequentam os eventos como também descobrir a procedência das localidades dos sócios torcedores. A figura 5.13 apresenta o percentual de acessos aos eventos realizados pelos sócios no ano de 2004 separados por sexo.

Ano	Evento - Evento	Descricao				
		Feminino	Juridico	Masculino	nao infor	Total geral
		Acessos	Acessos	Acessos	Acessos	Acessos
2004	Campeao FC x Atlético Jr.(23)	9,82%	3,67%	86,51%		100,00%
	Campeao FC x Atlético MG	8,93%	4,02%	87,04%		100,00%
	Campeao FC x Atlético PR (10)	9,90%	2,34%	87,73%	0,03%	100,00%
	Campeao FC x Boca Juniors (26)	11,81%	3,54%	84,64%	0,02%	100,00%
	Campeao FC x Botafogo (21)	10,04%	3,21%	86,75%		100,00%
	Campeao FC x Corinthians (24)	11,08%	2,87%	86,04%	0,01%	100,00%
	Campeao FC x Coritiba (15)	9,97%	3,59%	86,40%	0,04%	100,00%
	Campeao FC x Criciúma (08)	9,57%	3,15%	87,28%		100,00%
	Campeao FC x Cruzeiro (27)	9,71%	2,96%	87,31%	0,02%	100,00%
	Campeao FC x Figueirense (14)	9,40%	4,56%	86,02%	0,03%	100,00%
	Campeao FC x Flamengo (16)	10,07%	3,65%	86,27%	0,02%	100,00%
	Campeao FC x Fluminense	11,82%	1,92%	86,25%		100,00%
	Campeao FC x Glória	10,04%	2,38%	87,56%	0,02%	100,00%
	Campeao FC x Goiás	10,94%	1,20%	87,86%		100,00%
	Campeao FC x Grêmio (09)	9,15%	3,50%	87,34%		100,00%
	Campeao FC x Grêmio (17)	9,78%	3,65%	86,56%	0,01%	100,00%
	Campeao FC x Guarani (22)	11,16%	3,35%	85,47%	0,02%	100,00%
	Campeao FC x Juventude (11)	10,11%	2,94%	86,92%	0,03%	100,00%
	Campeao FC x Paraná (28)	10,21%	2,39%	87,39%	0,01%	100,00%
	Campeao FC x Ponte Preta (13)	10,20%	4,69%	85,08%	0,03%	100,00%
	Campeao FC x Santos (19)	9,17%	3,76%	87,06%	0,01%	100,00%
	Campeao FC x São Caetano (25)	10,68%	3,86%	85,46%		100,00%
	Campeao FC x São Paulo (12)	9,53%	3,21%	87,22%	0,03%	100,00%
	Campeao FC x Vasco	8,92%	3,29%	87,79%		100,00%
	Campeao FC x Vitória (18)	11,20%	3,97%	84,80%	0,03%	100,00%
	Total	10,20%	3,36%	86,43%	0,01%	100,00%
Total geral		10,20%	3,36%	86,43%	0,01%	100,00%

Figura 5.13 – Percentual de acessos aos eventos por sexo.

É possível constatar que a maioria dos acessos dos sócios torcedores é do sexo masculino com média de 86,43%, seguidos dos sócios do sexo feminino média de 10,20%. Verifica-se ainda nesta consulta a média de 3,33% de sócios Jurídicos, representados por empresas sócias ou parceiras do clube, além do tipo não informado, representados por sócios que não possuem o sexo definido em seu cadastro. Para este tipo de sócio, a média de acessos no ano de 2004 foi de 0,01%.

Criando a consulta de acessos por região, foi possível constatar que no ano de 2004 a grande maioria dos acessos ocorreu de sócios residentes na região metropolitana do estado com 77,93% dos acessos, seguido da região do Vale do Rio dos Sinos, com uma média de 7,95% do total no período. Os acessos realizados por sócios que não possuem o município informado em seu cadastro somado aos sócios provenientes de outros estados e também do exterior, perfazem um índice de 6,02% do total de acessos no período analisado. As figuras 5.13 e 5.14 apresentam os acessos aos eventos ocorridos no ano de 2004, separados por região e por unidade federativa, respectivamente.

Ano	Regiao	Acessos
2004	Metropolitano Delta do Jacui	77,93%
	Vale do Rio dos Sinos	7,95%
	Nao informado	3,64%
	Demais estados	2,38%
	Fronteira Oeste	1,27%
	Serra	0,79%
	Sul	0,64%
	Vale do Cai	0,57%
	Paranhana-Encosta da Serra	0,43%
	centro-Sul	0,43%
	Central	0,42%
	Litoral	0,41%
	Vale do Rio Pardo	0,34%
	Missoes	0,33%
	Fronteira Noroeste	0,29%
	Vale do Taquari	0,29%
	Producao	0,27%
	Hortensias	0,25%
	Campanha	0,24%
	Jacui-Centro	0,19%
	Campos de Cima da Serra	0,19%
	Medio Alto Uruguai	0,17%
	Norte	0,17%
	Noroeste Colonial	0,16%
	Alto Jacui	0,08%
	Rio da Varzea	0,06%
	Nordeste	0,04%
	Exterior	0,04%
	Alto da Serra do Botucara (Em branco)	0,02%
	Total	100,00%
Total geral		100,00%

Figura 5.13 – Percentual de acessos aos eventos em 2004 por região.

Ano	Uf	Acessos
2004	RS	93,94%
	NI	3,64%
	SC	1,11%
	PR	0,48%
	SP	0,22%
	DF	0,19%
	RJ	0,17%
	BA	0,06%
	EX	0,04%
	GO	0,03%
	AC	0,02%
	PE	0,02%
	RN	0,02%
	MS	0,01%
	MG	0,01%
	CE	0,01%
	MT	0,01%
	AM	0,00%
	ES	0,00%
	PB	0,00%
	AL	0,00%
	SE	0,00%
	TO	0,00%
	AP	0,00%
	PI	0,00%
	RO	0,00%
	Total	100,00%
Total geral		100,00%

Figura 5.14 – Percentual de acessos aos eventos em 2004 por UF.

Através da consulta de acessos aos eventos, foi possível constatar que o evento que obteve maior índice de acesso no ano de 2004 foi a partida de futebol entre Clube Campeão FC X Boca Junior, com total de 22301 acessos (figura 5.15). A figura 5.16 apresenta a faixa

de horários que ocorreram os acessos. Nesta consulta é possível constatar que as faixas de horários que acontecem o maior volume de acessos ocorrem entre 45 minutos até 1 minuto antes do início da partida. A soma destes intervalos perfaz um total de 56,62% do total dos acessos no período. Um dado interessante é o percentual de 6,95% dos acessos ocorrerem somente após o segundo tempo da partida, ou seja, acima de 45 minutos depois do início.

Ano	Evento - Evento	Acessos
2004	Campeao FC x Boca Juniors (26)	22301
	Campeao FC x Grêmio (09)	12072
	Campeao FC x Atlético Jr.(23)	11681
	Campeao FC x Juventude (11)	11487
	Campeao FC x Santos (19)	9306
	Campeao FC x Paraná (28)	8953
	Campeao FC x Grêmio (17)	8703
	Campeao FC x Botafogo (21)	8693
	Campeao FC x Ponte Preta (13)	7351
	Campeao FC x Figueirense (14)	7309
	Campeao FC x Atlético PR (10)	7109
	Campeao FC x Corinthians (24)	7011
	Campeao FC x São Caetano (25)	6754
	Campeao FC x São Paulo (12)	6136
	Campeao FC x Cruzeiro (27)	5943
	Campeao FC x Glória	5884
	Campeao FC x Vitória (18)	5770
	Campeao FC x Guarani (22)	5760
	Campeao FC x Vasco	5685
	Campeao FC x Coritiba (15)	5096
	Campeao FC x Flamengo (16)	5047
	Campeao FC x Fluminense	4161
	Campeao FC x Criciúma (08)	4160
	Campeao FC x Atlético MG	4075
	Campeao FC x Goiás	2495
	Total	188942
	Total geral	188942

Figura 5.15 – Quantidade de acessos por evento em 2004.

Descricao	Acessos
acima de 4:00 antes	0,00%
de 04:00 a 03:46 antes	0,00%
de 03:45 a 03:31 antes	0,32%
de 03:30 a 03:16 antes	0,28%
de 03:15 a 03:01 antes	0,36%
de 03:00 a 02:46 antes	0,73%
de 02:45 a 02:31 antes	0,76%
de 02:30 a 02:16 antes	1,01%
de 02:15 a 02:01 antes	1,75%
de 02:00 a 01:46 antes	2,17%
de 01:45 a 01:31 antes	3,37%
de 01:30 a 01:16 antes	5,08%
de 01:15 a 01:01 antes	6,28%
de 01:00 a 00:46 antes	9,71%
de 00:45 a 00:31 antes	13,51%
de 00:30 a 00:16 antes	14,70%
de 00:15 a 00:01 antes	14,85%
de 00:00 a 00:15 depois	13,56%
de 00:16 a 00:30 depois	3,59%
de 00:31 a 00:45 depois	1,03%
acima de 00:45 depois	6,95%
Total geral	100,00%

Figura 5.16 – Quantidade de acesso por faixa de hora.

As consultas apresentadas neste trabalho são algumas amostras que podem ser realizadas utilizando os cubos Torcedor e Acesso. As possibilidades e opções de combinações são bastante amplas, possibilitando ao usuário, condições de realizar consultas em diversos níveis, hierarquias e também direcionadas por assuntos.

CONCLUSÃO

Os estudos de BI, como também DW e demais tecnologias relacionadas, contribuíram para compreender a importância destas ferramentas, do ponto de vista estratégico em um ambiente corporativo. Estas tecnologias tornam-se significantes para aquelas organizações que necessitam de ferramentas de auxílio nas tomadas de decisões de forma rápida e segura.

Através deste trabalho foi possível adquirir uma fundamentação teórica sobre as tecnologias de BI permitindo compreender, planejar e aplicar os conhecimentos adquiridos em um ambiente real.

Foi interessante observar que o conceito de BI é uma tecnologia, e não necessariamente uma solução. Esta premissa indica o fato de que o desenvolvimento de um DW mal concebido pode resultar em fracasso. Faz-se necessário que o DWA tenha conhecimentos aprofundados sobre o projeto para que consiga desenvolver junto com os demais colaboradores, um DW que contenha uma estrutura que garanta resultados coerentes e satisfatórios.

Cabe ressaltar que as ferramentas de OLAP, são ferramentas de apoio às decisões. Estas ferramentas não dão as respostas aos problemas, mas oferecem informações que facilitam o trabalho daqueles que têm o poder de decidir.

A proposta de construção de um DM no clube desportivo estudado torna-se interessante devido a dois principais motivos: necessidade prioritária de implantar uma ferramenta de tomada de decisões e por tratar-se de uma área de dados específica do clube. A criação de um DW exigiria um custo maior de tempo além de necessitar de uma análise mais aprofundada de todas as bases de dados existentes.

O referido trabalho cria a possibilidade de continuidade de projetos desta natureza, agregando valor ao produto Society. O mesmo poderá ser oferecido aos clientes que poderão obter seus DM sob medida.

O trabalho trouxe um bom embasamento para que a execução efetiva do DM no Clube desportivo em questão seja realizada com maior propriedade e discernimento. Possibilitou-se também conhecer a nova ferramenta da Microsoft (SQL Server 2005) proporcionando um avanço nos conhecimentos fundamentais e técnicos.

A flexibilidade e a facilidade de utilização da ferramenta *Analysis Services* da *Microsoft* contribuíram para o sucesso na realização deste trabalho, comprovando ser uma ótima ferramenta de BI, aliado à farta documentação disponível, garantindo a seus usuários a realização de trabalhos de alto nível e produtividade.

Melhorias e trabalhos futuros.

Baseado no estudo apresentado abre-se a possibilidade de melhorias e trabalhos futuros na área. São apresentadas a seguir algumas sugestões:

- Pode-se criar outros cubos utilizando as mesmas tabelas de fatos, separados por assunto, ao invés dos dois cubos criados contendo todas as dimensões disponíveis. Desta forma os cubos poderiam ficar organizados de forma mais específica de acordo com o assunto desejado.
- Proposta de desenvolvimento de scripts que façam o processo de extração de dados dos sistemas OLTP existentes para dentro do SGBD já criado.
- Criação de novos cubos pertinentes á área financeira dos sócios, explorando com maiores detalhes as dívidas dos associados, separadas por contas agrupadoras. Outros cubos possíveis de criações seriam aqueles voltados ao motivo dos desligamentos dos sócios do quadro social.
- Possibilitar aos cônsoles a realização de consultas aos cubos de sócios e acessos pertinentes a sua jurisprudência e área de atuação.
- Aprimoramento da tabela fato de Acesso, agregando aos eventos informações tais como resultados da partida, desempenho do time diante da tabela

classificatória de cada campeonato, além de associar o evento ao campeonato que ele pertence.

- Desenvolver cubos para análise de perfis dos não-sócios que utilizam os serviços do clube como também os acessos aos eventos esportivos.
- Desenvolvimento de um ambiente com interface mais amigável ao usuário final. O recurso que poderá ser explorado é o *Reporting Services*, que permite o desenvolvimento de relatórios dos cubos do DM para posteriormente serem disponibilizados à URL¹⁰.

Enfim, existem inúmeras possibilidades para pesquisas pertinentes à área, objetivando o refinamento da solução apresentada e a maior exploração dos recursos oferecidos pela ferramenta utilizada, tornando a solução mais robusta e versátil.

¹⁰ Do Inglês *Universal Resource Locator*. Sequência de caracteres que indica a localização de um recurso na Internet, como por exemplo, <http://www.society.com.br/>.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDREI, Alencastro; SEMAA, Gustavo; DIAS, Ricardo Rodrigo. Data Mining e a descoberta de associações em dados. **SQL Magazine**, Grajaú, ano 2, Ed. 26, p. 22-29, [s.d.]

ANZANELLO, Cynthia A. **OLAP Conceitos e Utilização**. Instituto de Informática – Universidade Federal do Rio Grande do Sul. Porto Alegre. Disponível em: <http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_cynthia.pdf>. Acesso em: 16 nov. 2007

CAMPOS, Maria L. M. **Tópicos Especiais em Banco de Dados I**. Apresentação de slides. Rio de Janeiro, 2003. 47 slides. Disponível em: <<http://dataware.nce.ufrj.br:8080/dataware/areas/dw>> Acesso em: 07 out. 2007.

CIELO, Ivã. **ETL - Extração, Transformação e Carga de Dados**. [s.d]. Disponível em: <<http://www.datawarehouse.inf.br/>>. Acesso em: 18 nov. 2007.

CRIVELINI, Wagner. Mitos e Verdades sobre Modelagem Multidimensional. **SQL Magazine**, Grajaú, ano 3, Ed. 39, p. 6-14, [s.d.]

FELBER, Edmilson J. W. **Proposta de Uma Ferramenta OLAP em um Data Mart Comercial: Uma Aplicação prática na Indústria Calçadista**. Novo Hamburgo: 2006. 102 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Instituto de Ciência Exatas e Tecnológicas, Centro Universitário Feevale, Novo Hamburgo.

FORTULAN, Marcos R.; GONÇALVES FILHO, Eduardo Vila. Uma proposta de aplicação de *Business Intelligence* no chão-de-fábrica. **Gestão da Produção**. V.12 n..1 ISSN 0104-530X São Carlos. Jan./Apr. 2005. Disponível em: <http://www.scielo.br/scielo.php?pid=S0104-530X2005000100006&script=sci_arttext&tlng=pt>. Acesso em: 10 abr. 2008.

INMON, W. H. **Como construir o Data Warehouse**. Rio de Janeiro: Campus, 1997. 388 p.

INMON, W. H.; WELCH, J. D.; GLASSEY, K. L. **Gerenciando Data Warehouse**. São Paulo: Makron Books, 1999. 375 p.

ITALIANO, Isabel Cristina; ESTEVES, Luiz Antônio. Modelagem de Data Warehouse e Data Mart. **SQL Magazine**, Grajaú, ano 2, Ed. 14, p. 37-43, [s.d.]

JACOBSON, Reed; MISNER, Stacia; HITACHI CONSULTING. **Microsoft SQL Server 2005 Analysis Services: passo a passo**. Porto Alegre? Bookman, 2007. 352 p.

KIMBALL, Ralph. *Data Warehouse Toolkit*. São Paulo: Makron Books, 1998. 388 p.

MACHADO, Felipe N. R. **Projeto de Data Warehouse: Uma Visão Multidimensional**. São Paulo: Érica, 2000. 248 p.

MATA, Leonardo Santos da. **Aplicação da Tecnologia de Business Intelligence no Processo de Avaliação Institucional**. Palmas: 2005. 70 p. Trabalho de Conclusão de Curso (Sistemas de Informação) – Centro Universitário Luterano de Palmas, Palmas.

MICROSOFT. **Módulo I - O que é Business Intelligence**. 2007. 15 p. Disponível em: <<http://www.microsoft.com/brasil/servidores/bi/default.aspx>> Acesso em: 13 set. 2007.

MICROSOFT. **Módulo 2 - Definindo Soluções OLAP**. 2007. 16 p. Disponível em: <<http://www.microsoft.com/brasil/servidores/bi/default.aspx>> Acesso em: 13 set. 2007.

MICROSOFT. **Módulo 3 - Desenhando uma solução OLAP**. 2007. 22 p. Disponível em: <<http://www.microsoft.com/brasil/servidores/bi/default.aspx>> Acesso em: 13 set. 2007.

MICROSOFT. **Módulo 4 - Construindo uma solução OLAP**. 2007. 15 p. Disponível em: <<http://www.microsoft.com/brasil/servidores/bi/default.aspx>> Acesso em: 13 set. 2007.

MICROSOFT. **Módulo 5 - Implementando Cubos OLAP**. 2007. 22 p. Disponível em: <<http://www.microsoft.com/brasil/servidores/bi/default.aspx>> Acesso em: 13 set. 2007.

NARDI, Alexandre Ricardo. **Fundamentos e Modelagem de Bancos de Dados Multidimensionais**. 2007. Disponível em: <http://www.microsoft.com/brasil/msdn/tecnologias/arquitetura/Dados_Multidimensionais.msp#ERDAE>. Acesso em: 19 out. 2007.

NARDI, Alexandre Ricardo; CHIBA, Cláudio. **Microsoft Business Intelligence de Ponta-a-Ponta**. 2007. Disponível em: <http://www.microsoft.com/brasil/msdn/tecnologias/arquitetura/BI_Ponta_a_Ponta.msp#EJC>. Acesso em: 18 nov. 2007.

PARRINI, Esther. **Gestão do Conhecimento no Suporte à Decisão em Ambiente OLAP**. Rio de Janeiro: 2002. 157 p. Dissertação (Mestrado em Informática) – Instituto de Matemática, Núcleo de Computação Eletrônica – NCE, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

PETKOVIC, Dusan. **SQL Server 2000 – Guia Prático**. São Paulo: Makron Books, 2001. 555 p.

RITTMAN, Mark. O que esperar das novas tecnologias OLAP da *Microsoft* e *Oracle*. **SQL Magazine**, Grajaú, ano 2, Ed. 14, p. 63-66, [s.d.]

ROSA, Kátia Liane da. **Análise de ferramentas OLAP com acesso à banco de dados**. Novo Hamburgo: 2004. 90 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Instituto de Ciências Exatas e Tecnológicas, Centro Universitário Feevale, Novo Hamburgo.

SANTOS, José Geraldo dos. **Oferta de produtos com configurações customizadas**. Belo Horizonte, 2001. 103 p. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. 2001 Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/9068.dpf>>. Acesso em: 17 nov. 2007.

SILVA, Sílvia Sidnéia da; CARITÁ, Edílson Carlos; FELICIO, Aline Cazarini. **Ferramenta de Apoio à Decisão para Detecção de Fatores de Risco para Doença Arterial Coronariana**. Ribeirão Preto, 2005. Artigo do curso de (Enfermagem). Centro de Ciências Exatas, Naturais e Tecnológicas - Universidade de Ribeirão Preto (UNAERP).

STANEK, William R. **Microsoft SQL Server 2005: Guia de bolso do administrador**. Porto Alegre: Bookman, 2006. 576 p.

SPENIK, MARK; SLEDGE, ORRYN. **Microsoft SQL Server 2000 DBA: Guia de sobrevivência**. Rio de Janeiro: Campus, 2001. 773 p.

ANEXO I – *SCRIPTS VISUAL FOX PRO* PARA FAZER A EXTRAÇÃO DOS DADOS

```
*Cria a tabela fato torcedor

*--Cria cursor fato
CREATE CURSOR crsFatoNome (;
  ID_Torce      N(10),;
  ID_Tempo      N(8),;
  ID_Sexo       N(3),;
  ID_FEtaria    N(3),;
  ID_Catego     N(5),;
  ID_SSocial    N(2),;
  ID_Munic      N(10),;
  ID_Cobra      N(5),;
  Consul        N(1),;
  VlrPrevi     N(10,2),;
  VlrReali     N(10,2),;
  Adimplente    N(1))

*--Grava Tabela
COPY TO C:\ETC\FatoTorcedor.DBF FOX2X

*--Abre as dimensões para ligar as keys
USE C:\ETC\Tempo      IN 0
USE C:\ETC\categoria IN 0
USE C:\ETC\cobrador  IN 0
USE C:\ETC\municipio IN 0
USE C:\ETC\FatoTorcedor IN 0

AbreTempo('Prest')

*--Cursor contendo nomes
SELECT Matricula, Cobrador, Categoria, DtNascim, Sexo ;
FROM Nomes ;
INTO CURSOR CrsNomes ORDER BY Matricula

LOCAL ldData, lnConta, ln_ID, lnVlrPrevisto, lnVlrRealizado, lcConsul, lnConsul, ;
lnSexo, lnCategoria, lnMunicipio, lnCobrador, lnFaixaEtaria, lnAdimplente, ;
lnTempo, lnCodSSocial
STORE 0 TO lnConta, ln_ID, lnVlrPrevisto, lnVlrRealizado, lnSexo, lnCategoria, ;
lnMunicipio, lnCobrador, lnFaixaEtaria, lnTempo, lnCodSSocial, lnConsul, lnAdimplente

ldData = CTOD('31/01/2004')

*--Extrai 3 anos da base de dados a partir de 31/01/2004
FOR lnConta = 1 TO 36

  SELECT crsNomes
  SCAN
    IF Situacao(ldData, crsNomes.Matricula) $ "AI"
      ln_ID = ln_ID + 1

      *--Procura ID Tempo
      =SEEK(ldData, 'TEMPO', 'DAT')
      lnTempo = Tempo.ID_Tempo

      *--É consul
      lnConsul = IIF(SituacaoSv(ldData, crsNomes.matricula, "CONSUL ")$"AI", 1, 0)
```

```

*--Verifica o que era previsto receber no mês
*--Define o valor previsto das parcelas no mes analisado

lnVlrPrevisto = 0
=SEEK(crsNomes.Matricula, 'Prest', 'CCS')
DO WHILE Prest.Matricula + Prest.Conta + Prest.Chave = crsNomes.Matricula

    *-- Soma
    IF EOM(Prest.Vencimento)= ldData

        lnVlrPrevisto = lnVlrPrevisto + ;
            CalcTarifa(Prest.Matricula, Prest.Conta, Prest.Chave,;
                Prest.Sequencia, Prest.Vencimento, gdDataSis, Prest.NrPrest,;
                Prest.CobPosse)
        ENDIF

    SKIP IN Prest

ENDDO

*--Volta para o cursor principal
SELECT crsNomes

*--Verifica o que foi recebido no mês
*--seleciona todas as prestações que o pagamento
*--tenha acontecido dentro do período

lnVlrRealizado = 0
=SEEK(crsNomes.Matricula, 'Prest', 'CCS')
DO WHILE Prest.Matricula + Prest.Conta + Prest.Chave = crsNomes.Matricula

    *-- Soma
    IF EOM(Prest.DtCobr) = ldData AND Prest.Situacao="P"

        lnVlrRealizado = lnVlrRealizado + ;
            CalcTarifa(Prest.Matricula, Prest.Conta, Prest.Chave,;
                Prest.Sequencia, Prest.Vencimento, gdDataSis, Prest.NrPrest,;
                Prest.CobPosse)
        ENDIF

    SKIP IN Prest

ENDDO

*--Volta para o cursor principal
SELECT crsNomes

*--Define a adimplencia
*--se houver alguma prestacao com vencimento igual ou menor ao mes analisado e
*--que esteja aberta ou
*--paga depois do mes analisado, ele é inadimplente.

SELECT Matricula ;
    FROM Prest;
    WHERE Matricula + Conta + Chave + Sequencia = crsNomes.Matricula AND ;
        EOM(Prest.vencimento) <= ldData AND !EMPTY(Prest.Vencimento) AND ;
        ((Prest.DtCobr > ldData AND Prest.Situacao="P") OR Prest.Situacao$"GI") ;
    ORDER BY Vencimento, Matricula, Conta, Chave, Sequencia ;
INTO CURSOR crsPrestAdim

IF _TALLY = 0
    lnAdimplente = 1
ELSE
    lnAdimplente = 0
ENDIF

*--Volta para o cursor principal
SELECT crsNomes

*--Define Sexo
DO CASE
CASE crsNomes.Sexo = "M"
    lnSexo = 1
CASE crsNomes.Sexo = "F"
    lnSexo = 2
OTHERWISE

```

```

        lnSexo = 3
    ENDCASE

    *--Define FaixaEtaria
    DO CASE
    CASE VariAno(ldData, crsNomes.DtNascim) < 12
        lnFaixaEtaria = 1
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 12, 15)
        lnFaixaEtaria = 2
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 16, 18)
        lnFaixaEtaria = 3
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 19, 22)
        lnFaixaEtaria = 4
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 23, 25)
        lnFaixaEtaria = 5
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 26, 30)
        lnFaixaEtaria = 6
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 31, 35)
        lnFaixaEtaria = 7
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 36, 40)
        lnFaixaEtaria = 8
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 41, 50)
        lnFaixaEtaria = 9
    CASE BETWEEN(VariAno(ldData, crsNomes.DtNascim), 51, 60)
        lnFaixaEtaria = 10
    OTHERWISE
        lnFaixaEtaria = 11
    ENDCASE

    *--Procura Categoria
    =SEEK(crsNomes.Categoria, 'CATEGORIA', 'CAT')
    lnCategoria = Categoria.id_Catego

    *--Define Situacao Social
    DO CASE
    CASE Situacao(ldData, crsNomes.Matricula) = 'A'
        lnCodSSocial = 1
    CASE Situacao(ldData, crsNomes.Matricula) = 'I'
        lnCodSSocial = 2
    CASE Situacao(ldData, crsNomes.Matricula) = 'F'
        lnCodSSocial = 3
    CASE Situacao(ldData, crsNomes.Matricula) = 'D'
        lnCodSSocial = 4
    ENDCASE

    *--Procura Municipio
    =SEEK(Correio('MUNICIPIO', crsNomes.Matricula), 'MUNICIPIO', 'MUN')
    lnMunicipio = Municipio.id_Munic

    *--Procura cobrador
    =SEEK(crsNomes.Cobrador, 'COBRADOR', 'COB')
    lnCobrador = Cobrador.id_Cobra

    *--Adiciona dados na tabela
    APPEND BLANK IN FatoTorcedor
    REPLACE id_Torce WITH ln_ID,;
        ID_Tempo WITH lnTempo,;
        ID_Sexo WITH lnSexo,;
        ID_FEtaria WITH lnFaixaEtaria,;
        ID_Catego WITH lnCategoria,;
        ID_SSocial WITH lnCodSSocial,;
        ID_Munic WITH lnMunicipio,;
        ID_Cobra WITH lncobrador,;
        Consul WITH lnConsul,;
        VlrPrevi WITH lnVlrPrevisto,;
        VlrReali WITH lnVlrRealizado,;
        Adimplente WITH lnAdimplente IN FatoTorcedor

    SELECT crsNomes

    ENDIF

    ENDSKAN

    *--Incrementa o mês de referência
    ldData = EOM(GOMONTH(ldData, 1))

```

ENDFOR

SELECT crsFatoNome

CLOSE TABLES all

*Cria a tabela fato acesso

```
*--Cria cursor fato
CREATE CURSOR crsFatoAcesso (;
  ID_Acesso      N(10),;
  ID_Evento      N(10),;
  ID_Tempo       N(8),;
  ID_Sexo        N(3),;
  ID_FEtaria     N(3),;
  ID_Catego      N(5),;
  ID_SSocial     N(2),;
  ID_Munic       N(10),;
  ID_FHora       N(3),;
  ID_Portao      n(3),;
  Consul         N(1))
```

*--Abre as dimensões para ligar as keys

```
USE C:\ETC\Tempo      IN 0
USE C:\ETC\categoria  IN 0
USE C:\ETC\Evento     IN 0
USE C:\ETC\municipio IN 0
USE C:\139\SSO\InterAce IN 0
```

SELECT * FROM InterAce ORDER BY MATRICULA INTO CURSOR crsInterace READWRITE

*--Arruma as matrículas de sócios que estão erradas na tabela
 REPLACE ALL Matricula WITH RIGHT(Matricula, 7) FOR LEFT(Matricula, 2) # "99" IN crsInterace

*--Arruma as matrículas de não-sócios que estão erradas na tabela
 REPLACE ALL Matricula WITH "99" + RIGHT(Matricula, 5) FOR LEFT(Matricula, 2) = "99" IN crsInterace

*--Cursor contendo nomes

```
SELECT crsInterAce.*, Nomes.Matricula AS MatriculaOK, Nomes.Categoria, Nomes.DtNascim,
Nomes.Sexo ;
FROM crsInterAce ;
LEFT JOIN Nomes ON crsInterAce.Matricula = LEFT(Nomes.Matricula, 7) ;
ORDER BY Nomes.Matricula ;
INTO CURSOR CrsNomes READWRITE
```

*--Atribui ao campo matrícula, seu número completo

REPLACE ALL Matricula WITH MatriculaOK FOR !isnull(MatriculaOk) IN crsNomes

```
LOCAL lnConta, ln_ID, lnConsul, lnCodSexo, lnCodCategoria, lnCodSSocial, lnCodEvento, ;
  lnCodMunicipio, lnCodFaixaEtaria, lnFaixaHora, lnCodTempo, lnHoraEntrada, lnCodFHora
```

```
STORE 0 TO lnConta, ln_ID, lnCodSexo, lnCodCategoria, lnCodMunicipio, lnCodSSocial, ;
  lnCodFaixaEtaria, lnFaixaHora, lnCodTempo, lnCodFHora, lnCodEvento, lnConsul
```

SELECT crsNomes

SCAN

IF crsNomes.Acesso = "Acesso normal"

ln_ID = ln_ID + 1

```
*--Procura ID Evento
=SEEK(crsNomes.Evento, 'EVENTO', 'EVE')
lnCodEvento = Evento.ID_Evento
```

```
*--Procura ID Tempo
=SEEK(crsNomes.Data, 'TEMPO', 'DAT')
lnCodTempo = Tempo.ID_Tempo
```

*--É consul

lnConsul = IIF(SituacaoSv(crsNomes.Data, crsNomes.matricula, "CONSUL ")\$"AI", 1, 0)

*--Define Sexo

DO CASE

CASE crsNomes.Sexo = "M"

lnCodSexo = 1

```

CASE crsNomes.Sexo = "F"
  lnCodSexo = 2
CASE crsNomes.Sexo = "J"
  lnCodSexo = 3
OTHERWISE
  lnCodSexo = 4 && quando nao informado ou sócio excluído do cadastro
ENDCASE

*--Define FaixaEtaria
DO CASE
CASE VariAno(crsNomes.Data, crsNomes.DtNascim) < 12
  lnCodFaixaEtaria = 1
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 12, 15)
  lnCodFaixaEtaria = 2
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 16, 18)
  lnCodFaixaEtaria = 3
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 19, 22)
  lnCodFaixaEtaria = 4
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 23, 25)
  lnCodFaixaEtaria = 5
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 26, 30)
  lnCodFaixaEtaria = 6
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 31, 35)
  lnCodFaixaEtaria = 7
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 36, 40)
  lnCodFaixaEtaria = 8
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 41, 50)
  lnCodFaixaEtaria = 9
CASE BETWEEN (VariAno(crsNomes.Data, crsNomes.DtNascim), 51, 60)
  lnCodFaixaEtaria = 10
CASE VariAno(crsNomes.Data, crsNomes.DtNascim) > 60
  lnCodFaixaEtaria = 11
OTHERWISE
  lnCodFaixaEtaria = 12 && socio excluído ou idade nao informada
ENDCASE

*--Procura Categoria
=SEEK(crsNomes.Categoria, 'CATEGORIA', 'CAT')
lnCodCategoria = Categoria.id_Catego

*--Define Situacao Social
DO CASE
CASE Situacao(crsNomes.Data, crsNomes.Matricula) = 'A'
  lnCodSSocial = 1
CASE Situacao(crsNomes.Data, crsNomes.Matricula) = 'I'
  lnCodSSocial = 2
CASE Situacao(crsNomes.Data, crsNomes.Matricula) = 'F'
  lnCodSSocial = 3
CASE Situacao(crsNomes.Data, crsNomes.Matricula) = 'D'
  lnCodSSocial = 4
OTHERWISE
  lnCodSSocial = 5 && Sócio excluído do cadastro
ENDCASE

*--Procura Municipio
=SEEK(Correio('MUNICIPIO', crsNomes.Matricula), 'MUNICIPIO', 'MUN')
lnCodMunicipio = Municipio.id_Munic

*--Pega Hora entrada
ltHoraEntrada = CTOT(RIGHT(GETWORDNUM(crsNomes.hora, 1, ':'), 2) + ':' +
GETWORDNUM(crsNomes.hora, 2, ':'))

*--Verifica tempo antecedencia entrada em minutos
lnFaixaHora = (CTOT(crsNomes.HInicio) - ltHoraEntrada)/60

*--Define FaixaHora
DO CASE
CASE lnFaixaHora > 240
  lnCodFHora = 1
CASE BETWEEN(lnFaixaHora, 226, 240)
  lnCodFHora = 2
CASE BETWEEN(lnFaixaHora, 211, 225)
  lnCodFHora = 3
CASE BETWEEN(lnFaixaHora, 196, 210)
  lnCodFHora = 4
CASE BETWEEN(lnFaixaHora, 181, 195)
  lnCodFHora = 5

```

```

CASE BETWEEN(lnFaixaHora, 166, 180)
  lnCodFHora = 6
CASE BETWEEN(lnFaixaHora, 151, 165)
  lnCodFHora = 7
CASE BETWEEN(lnFaixaHora, 136, 150)
  lnCodFHora = 8
CASE BETWEEN(lnFaixaHora, 121, 135)
  lnCodFHora = 9
CASE BETWEEN(lnFaixaHora, 106, 120)
  lnCodFHora = 10
CASE BETWEEN(lnFaixaHora, 91, 105)
  lnCodFHora = 11
CASE BETWEEN(lnFaixaHora, 76, 90)
  lnCodFHora = 12
CASE BETWEEN(lnFaixaHora, 61, 75)
  lnCodFHora = 13
CASE BETWEEN(lnFaixaHora, 46, 60)
  lnCodFHora = 14
CASE BETWEEN(lnFaixaHora, 31, 45)
  lnCodFHora = 15
CASE BETWEEN(lnFaixaHora, 16, 30)
  lnCodFHora = 16
CASE BETWEEN(lnFaixaHora, 01, 15)
  lnCodFHora = 17
CASE BETWEEN(lnFaixaHora, -15, 00)
  lnCodFHora = 18
CASE BETWEEN(lnFaixaHora, -30, -16)
  lnCodFHora = 19
CASE BETWEEN(lnFaixaHora, -45, -31)
  lnCodFHora = 20
CASE lnFaixaHora > -45
  lnCodFHora = 21
ENDCASE

*--Adiciona dados na tabela
APPEND BLANK IN crsFatoAcesso
REPLACE ID_Acesso WITH ln_ID,;
  ID_Evento WITH lnCodEvento,;
  ID_Tempo WITH lnCodTempo,;
  ID_Sexo WITH lnCodSexo,;
  ID_FEtaria WITH lnCodFaixaEtaria,;
  ID_Catego WITH lnCodCategoria,;
  ID_SSocial WITH lnCodSSocial,;
  ID_Munic WITH lnCodMunicipio,;
  ID_FHora WITH lnCodFHora,;
  ID_Portao WITH VAL(crsNomes.Portao),;
  Consul WITH lnConsul IN crsFatoAcesso

SELECT crsNomes

ENDIF

ENDSCAN

SELECT crsFatoAcesso

*--Grava Tabela
COPY TO C:\ETC\FatoAcesso.DBF FOX2X

CLOSE TABLES all

*--Cria a dimensão categoria
SELECT 00000 AS id_Catego, chave AS Categoria, Descricao, ;
  Mensalid + SPACE(10) AS Pagante, TipoSocio + SPACE(9) AS TipoSocio ;
FROM Chave ;
WHERE Tipo=="01" ;
INTO CURSOR crsCategoria READWRITE

LOCAL lnCodCategoria, lcTipoSocio
lnCodCategoria = 0

SCAN
DO CASE
CASE crsCategorias.TipoSocio = "T"
  lcTipoSocio = "TITULAR"
CASE crsCategorias.TipoSocio = "D"
  lcTipoSocio = "DEPENDENTE"

```

```

CASE crsCategorias.TipoSocio = "N"
    lcTipoSocio = "NAO SOCIO"
ENDCASE

lnCodCategoria = lnCodCategoria + 1

REPLACE id_catego WITH lnCodCategoria, ;
        Pagante WITH IIF(Pagante = "S", "PAGANTE", "NAO PAGANTE"), ;
        TipoSocio WITH lcTipoSocio IN crsCategoria

ENDSCAN

*--cria categoria excluído

REPLACE id_Catego WITH 0, ;
        Categoria WITH "DESC", ;
        Descricao WITH "Desconhecida", ;
        Pagante WITH "INDEFINIDO", ;
        TipoSocio WITH "INDEFINIDO" ;
IN crsCategoria

COPY TO C:\ETC\CATEGORIA.DBF FOX2X



---


*--Cria a dimensão cobradores

SELECT 00000 As id_Cobra, Cobrador, Nome ;
FROM Cobra ;
WHERE Tipo=="E" ;
INTO CURSOR crsCobrador READWRITE

LOCAL lnCodCobrador
lnCodCobrador = 0

SCAN

    lnCodCobrador = lnCodCobrador + 1

    REPLACE id_Cobra WITH lnCodCobrador IN crsCobrador

ENDSCAN

*--Cria o Cobrador -não informando-
APPEND BLANK IN crsCobrador
REPLACE id_Cobra WITH 0, ;
        Cobrador WITH "NI", ;
        Nome WITH "Nao informado" ;
IN crsCobrador

COPY TO C:\ETC\Cobrador.DBF FOX2X



---


*--cria a dimensão eventos

LOCAL lnCodEvento
*--abre tabela de acesso
USE C:\139\SSO\INTERACE IN 0

SELECT 00000 AS Id_Evento, Evento FROM InterAce GROUP BY Evento INTO CURSOR crsEvento
READWRITE

lnCodEvento = 1

SCAN
    *--Grava ID do Evento
    REPLACE ID_Evento WITH lnCodEvento IN crsEvento

    lnCodEvento = lnCodEvento + 1

ENDSCAN

*--grava tabela
COPY TO C:\ETC\Evento FOX2X

CLOSE TABLES all



---


*--Cria a dimensão municípios

```

```

USE C:\139\SSO\MUNICIPIOSRS IN 0
USE C:\139\SSO\INTERCID

SELECT 0000000000 As id_Munic, *, SPACE(14) AS Reg_Funcio, SPACE(30) AS Regiao, ;
    SPACE(2) AS Cored, 0000000000 AS Populacao ;
FROM InterCid ;
INTO CURSOR crsMunicipio READWRITE

LOCAL lnCodMunicipio
lnCodMunicipio = 0

SCAN
    lnCodMunicipio = lnCodMunicipio + 1
    REPLACE id_Munic WITH lnCodMunicipio IN crsMunicipio

    =SEEK(ALLTRIM(crsMunicipio.Nome), 'MunicipiosRS', 'MUN')

    REPLACE Reg_Funcio WITH MunicipiosRS.Reg_Funcio, ;
        Regiao WITH MunicipiosRS.Regiao, ;
        Cored WITH ALLTRIM(STR(MunicipiosRS.cored)), ;
        Populacao WITH MunicipiosRS.Populacao IN crsMunicipio

ENDSCAN

*--Cria município -não informando-
APPEND BLANK IN crsCobrador
REPLACE id_Munic WITH 0, ;
    MUNICIPIO WITH 0, ;
    Nome WITH "NAO INFORMADO", ;
    Reg_Funcio WITH "Nao informado", ;
    Regiao WITH "Nao informado", ;
    UF WITH "NI", ;
    Distancia WITH 99999, ;
    Cored WITH "0", ;
    Populacao WITH 0 ;
IN crsMunicipio

*--Demais Estados
REPLACE ALL Reg_Funcio WITH "Demais estados" FOR !UF$"RS,EX" IN crsMunicipio
REPLACE ALL Regiao WITH "Demais estados" FOR !UF$"RS,EX" IN crsMunicipio

*--Exterior
REPLACE ALL UF WITH "EX" FOR EMPTY(UF) IN crsMunicipio
REPLACE ALL Reg_Funcio WITH "Exterior" FOR UF="EX" IN crsMunicipio
REPLACE ALL Regiao WITH "Exterior" FOR UF="EX" IN crsMunicipio

COPY TO C:\ETC\MUNICIPIO.DBF FOX2X

CLOSE TABLES all

*--Cria a dimensao tempo

CREATE CURSOR crsTempo (
    ID_tempo N(8),;
    Data D,;
    Mes C(12),;
    Bimestre C(12),;
    Trimestre C(12),;
    Semestre C(12),;
    Ano n(4))

LOCAL ldData, lnConta, ln_CodTempo, lcBimestre, lcTrimestre
STORE 0 TO lnConta, lnCodTempo

ldData = CTOD('31/01/2004')

*--Extrai 3 anos da base de dados a partir de 31/01/2004
FOR lnConta = 1 TO 36

    *--Bimestre
    DO CASE
    CASE BETWEEN (MONTH(ldData),1,2)
        lcBimestre = "Primeiro"
    CASE BETWEEN (MONTH(ldData),3,4)
        lcBimestre = "Segundo"
    CASE BETWEEN (MONTH(ldData),5,6)

```

```

        lcBimestre = "Terceiro"
    CASE BETWEEN(MONTH(ldData),7,8)
        lcBimestre = "Quarto"
    CASE BETWEEN(MONTH(ldData),9,10)
        lcBimestre = "Quinto"
    CASE BETWEEN(MONTH(ldData),11,12)
        lcBimestre = "Sexto"
    OTHERWISE
        lcBimestre = "Setimo"
    ENDCASE

    *--trimestre
    DO CASE
    CASE BETWEEN(MONTH(ldData),1,3)
        lcTrimestre = "Primeiro"
    CASE BETWEEN(MONTH(ldData),4,6)
        lcTrimestre = "Segundo"
    CASE BETWEEN(MONTH(ldData),7,9)
        lcTrimestre = "Terceiro"
    OTHERWISE
        lcTrimestre = "Quarto"
    ENDCASE

    *--Grava o registro
    APPEND BLANK
    REPLACE ID_Tempo WITH lnConta,;
        Data WITH ldData,;
        Mes WITH BCMonth(ldData),;
        Bimestre WITH lcBimestre,;
        Trimestre WITH lcTrimestre,;
        Semestre WITH IIF(MONTH(ldData)<7,'Primeiro', 'Segundo'),;
        Ano WITH YEAR(ldData) IN crsTempo

    *--Incrementa o mês de referência
    ldData = EOM(GOMONTH(ldData, 1))

ENDFOR

*--Grava Tabela
COPY TO C:\ETC\Tempo.DBF FOX2X
-----
*--Incrementa a dimensão tempo
USE C:\ETC\TEMPO IN 0

LOCAL lnCodTempo, ldData, lcBimestre, lcTrimestre, ldDataI, ldDataF

*--Pega último número do ID_Tempo
SELECT TOP 1 ID_Tempo FROM Tempo ORDER BY ID_Tempo DESC INTO CURSOR crsTempoTop

*--Cria cursor tempo para percorrer
SELECT * FROM Tempo ORDER BY Data INTO CURSOR crsTempo
INDEX ON Data TO DAT

lnCodTempo = crsTempoTop.ID_Tempo + 1

*--Pega a data inicial para geração
DO WHILE EMPTY(ldDataI)
    ldDataI = CTOD((INPUTBOX("Geração de Datas", "Informe a Data inicial",'01/01/2004')))

    IF EMPTY(ldDataI)
        IF MESSAGEBOX('Data inválida. Tentar novamente?', 1, "Atenção") = 2
            CLOSE TABLES ALL
            RETURN .F.
        ENDIF
    ENDIF
ENDDO

*--Pega a data final para geração
DO WHILE EMPTY(ldDataF)
    ldDataF = CTOD((INPUTBOX("Geração de Datas", "Informe a Data final",'31/12/2006')))

    IF EMPTY(ldDataF)
        IF MESSAGEBOX('Data inválida. Tentar novamente?', 1, "Atenção") = 2
            CLOSE TABLES ALL
            RETURN .F.
        ENDIF
    ENDIF
ENDDO

```

```

ldData = ldDataI
DO WHILE ldData <= ldDataF

  IF !SEEK(ldData, 'CRSTEMPO', 'DAT')

    *--Bimestre
    DO CASE
    CASE BETWEEN (MONTH(ldData),1,2)
      lcBimestre = "Primeiro"
    CASE BETWEEN (MONTH(ldData),3,4)
      lcBimestre = "Segundo"
    CASE BETWEEN (MONTH(ldData),5,6)
      lcBimestre = "Terceiro"
    CASE BETWEEN (MONTH(ldData),7,8)
      lcBimestre = "Quarto"
    CASE BETWEEN (MONTH(ldData),9,10)
      lcBimestre = "Quinto"
    CASE BETWEEN (MONTH(ldData),11,12)
      lcBimestre = "Sexto"
    OTHERWISE
      lcBimestre = "Setimo"
    ENDCASE

    *--trimestre
    DO CASE
    CASE BETWEEN (MONTH(ldData),1,3)
      lcTrimestre = "Primeiro"
    CASE BETWEEN (MONTH(ldData),4,6)
      lcTrimestre = "Segundo"
    CASE BETWEEN (MONTH(ldData),7,9)
      lcTrimestre = "Terceiro"
    OTHERWISE
      lcTrimestre = "Quarto"
    ENDCASE

    *--Grava o registro
    APPEND BLANK IN Tempo
    REPLACE ID_Tempo WITH lnCodTempo,;
      Data WITH ldData,;
      Mes WITH BCMonth(ldData),;
      Bimestre WITH lcBimestre,;
      Trimestre WITH lcTrimestre,;
      Semestre WITH IIF(MONTH(ldData)<7,'Primeiro', 'Segundo'),;
      Ano WITH YEAR(ldData) IN Tempo

    *--Incrementa lnCodTempo
    lnCodTempo = lnCodTempo + 1

  ENDIF

  *--Incrementa Data
  ldData = ldData + 1

ENDDO

CLOSE TABLES all

```