Universidade Católica de Brasília

PRÓ-REITORIA DE GRADUAÇÃO
TRABALHO DE CONCLUSÃO DE CURSO

Sistemas de Informação

Sistema de suporte à decisão utilizando Data Mart: estudo de caso de um site de empresa de jornais.

AUTORES:
Ailton Ferreira de Medeiros Junior
Leonardo Marques do Nascimento
Natasha de Carvalho França
Orientador: Vilson Carlos Hartmann

BRASÍLIA

2007

Ailton Ferreira de Medeiros Junior
Leonardo Marques do Nascimento
Natasha de Carvalho França

Sistema de suporte à decisão utilizando Data Mart: estudo de caso de site de empresa de jornais.

Monografia submetida à Banca Examinadora do Curso de Bacharelado Em Sistemas de Informação da Universidade Católica de Brasília para obtenção do Título de Bacharel em Sistemas de Informação.

Orientador: Msc. Vilson Carlos Hartmann

Brasília
2007

TERMO DE APROVAÇÃO

Monografia elaborada como requisito para obtenção do título de Bacharel no Curso de Bacharelado em Sistemas de Informação da Universidade Católica de Brasília, defendida e aprovada, em 23 de Junho de 2007, pela banca examinadora constituída por:

Prof. Msc. Vilson Carlos Hartmann
Orientador
Universidade Católica de Brasília

Prof. Mário de Oliveira Braga Filho
Banca
Universidade Católica de Brasília

Brasília
UCB

À Deus, pois ele nos deu a vida, saúde e inteligência para construirmos nossa vida. Aos Nossos Pais, pelo amor e carinho que sempre nos dedicaram e nos apoiaram nos momentos mais difíceis.

Aos Professores que nos passaram seus ensinamentos e suas vivências de vida, ajudando a nos tornar excelentes profissionais.

“Aprender é a única coisa de que a mente nunca se cansa, nunca tem medo e nunca se arrepende.” (Leonardo da Vinci)

Lista de Abreviações/Siglas

DM – Data Mart

DW – Data Warehouse

Web – World Wide Web

PHP – Hypertext PreProcessor

OLAP – Online Analytical Processing

ETL – Extract Transform Load

BI – Business Intelligence

TI – Tecnologia da Informação

SGBD – Sistema de Gerenciamento de Banco de Dados

Lista de Figuras

Figura 1: Elementos básicos do Data Warehouse

Figura 2: Cubo de três dimensões

Figura 3: Modelo Multidimensional

Figura 4: Modelo físico do Banco de Dados

Figura 5: Quantidade de acessos por cidade.

Figura 6: Quantidade de acessos por mês, na cidade de Brasília.

Figura 7: Quantidade de acessos por profissão.

Figura 8: Quantidade de acessos por idade.

Figura 9: Tempo de acesso dos usuários por sexo, em um matéria especifica, por mês.

Figura 10: Quantidade de acesso para o colunista, por idade.

Figura 11: Quantidade de acesso para o colunista, por profissão.

Figura 12: Quantidade de acesso para o colunista, por Cidade.

Figura 13: Quantidade de acesso para o colunista, por sexo.

Figura 14: Quantidade de acesso por o colunista.

Figura 15: Quantidade de acesso no *site*.

Figura 16: Tempo de acesso no *site*.

Figura 17: Duração de acesso nas Matérias.

Figura 18: Duração de acesso nas Colunas.

Figura 19: Duração de acesso nas Promoções.

Figura 20: Duração de acesso nas Edições.

Figura 21: Quantidade de acessos nas Matérias.

Figura 22: Quantidade de acessos por edição.

RESUMO

Muitas empresas possuem dados que às vezes são desperdiçados. Esse projeto visa buscar informações de uma empresa de jornais em sua página na *Internet*, através de um Data Mart, que tratará os dados de forma que deva proporcionar aos gerentes novas oportunidades de negócios e competitividade, através de relatórios e consultas mais específicos e detalhados. Com isso, conseguirão ter em mãos uma ferramenta que lhes permitirá tomar decisões gerenciais importantes para melhorar a administração e o crescimento da corporação.

PALAVRAS CHAVES: Data Mart, Páginas na Internet, decisões gerenciais.

ABSTRACT

Many companies have data that many times they do not use it. This project aims to search information of a periodical company in its page in the Internet, through one Date Mart, that will treat the data, in a form that must provide the new controlling business-oriented chances and competitiveness, through more specific and detailed reports and consultations. With this, they will have in hands a tool that will allow them to take management important decisions to improve the administration and the corporation growth.

KEYWORKS:. Data Mart, Web pages, management decisions.

SUMÁRIO

CAPITULO 1 4

1.1.Introdução 4

1.2. Problemas Diagnosticados 6

1.3. Objetivos Gerais 6

1.4. Objetivos específicos 7

1.5. Resultados Esperados 7

1.6.Metodologia 8

CAPITULO 2 9

2.1 Introdução ao Data Warehouse 9

2.2. Business Intelligence 10

2.3. Abordagens para o projeto de um DataWarehouse/Data Mart 12

2.3.1 Abordagem segundo Bill Inmon 12

2.3.2 Abordagem segundo Ralph Kimball 12

2.4. Fases para o desenvolvimento do Data Mart 13

2.4.1 Planejamento 13

2.4.1.1 Definição da Arquitetura Tecnológica 13

2.4.2 Levantamento das Necessidades 14

2.4.3 Modelagem Dimensional 14

2.4.4 Projeto físico do Banco de Dados 15

2.4.5 Projeto de Extração, Transformação e Carga (Ferramenta ETL) 15

2.4.6 Desenvolvimento de aplicação 15

2.4.7 Validação e teste 15

2.4.8 Treinamento 16

2.4.9 Implantação 16

2.5. Arquitetura do Data Warehouse 16
2.5.1 Sistemas Operacionais de origem 17
2.5.2 Área de Representação de Dados 17
2.5.3 Área de Apresentação dos Dados 18
2.5.4 Ferramentas de Acesso aos dados 19
2.6. Data Marts 20
2.7. Metadados 20
2.7.1 Modelo de Dados 21
2.8. Grão e Granularidade 22
CAPITULO 3 24
3.1. Modelagem Multidimensional 24
3.1.1 Tabela de Fatos 25
3.1.2 Tabela de Dimensões 26
CAPITULO 4 27
4.1. ETL (*Extract, Transform and Load*)_ Extração, Transformação e Carga 27
4.1.1 Extração 28
4.1.2 Transformação 28
4.1.3 Carga 29
4.1.4 Ferramentas de mercado 29
CAPITULO 5 30
5.1. OLAP (Processamento Analítico On-Line) 30
5.1.1 OLAP Multidimensional 31
5.1.2 OLAP Relacional 31
5.1.3 OLAP Híbrido 31
CAPITULO 6 32

6.1Estudo de Caso 32
6.2 Apresentação do Problema 32
6.3 Descrição do Ambiente e suas características 33
6.4 Descrição do Modelo de Dados 34
6.4.1 Fatos 35
6.4.2 Dimensões 35
6.4.3 Modelo Estrela 39
6.5 Solução Business 40
6.6 Modelo Físico dos Dados 41
6.7 Tipos de Análise dos dados 42
6.8 Exemplificações 43
CAPITULO 7 – Conclusão 53
Referências Bibliográficas: 55

CAPITULO 1

1.1. Introdução

Páginas na *Internet* são criados pelas empresas para divulgação de seus produtos, em uma empresa de jornais existe um *site*, que além de apresentar suas edições, também, através do *site*, vende publicidade e faz promoções. Esta empresa tem como seus produtos a venda de jornais avulsos, de assinatura de jornais e de publicidade.

O *site* possui um cadastro de usuário que é gratuito e não é necessário ser assinante, no qual é informado através da digitação dados do internauta, depois de efetuado o cadastro, o usuário terá a sua disposição um *site* sem restrições.

Todos os dias há uma nova edição do jornal, contendo as notícias do dia anterior à data de edição, disponibilizando aos seus clientes informações atualizadas de Brasília, do Brasil e do Mundo sobre vários assuntos divididos em editorias, a página na *Internet* por sua vez é atualizado todos os dias com um resumo desses noticiários.

O *site* contém editorias iguais às contidas no jornal impresso, dentre as quais contém um resumo das matérias publicadas com fotos se houver, pois a matéria completa somente encontra-se no jornal impresso.

Através do *site* o usuário terá a possibilidade de interação com as matérias através de envio de comentários disponibilizado na matéria ou poderá enviá-la a um amigo, adicionar as prediletas onde o usuário poderá deixar as matérias que achar importantes para depois acessá-las mais rapidamente.

A página web *(World Wide Web)* da empresa contém colunistas e suas respectivas notas, há também serviços para auxiliar os usuários em seu dia a dia, exemplo, busca de cep,

de cartórios, telefones úteis, Bancos e outros, nestes são acessados os *links* para os *sites* de acordo com o serviço solicitado.

Também no *site* existe um sistema de busca por palavra chave que procura todas as matérias inclusive de datas anteriores.

Na página da *internet* da empresa são locados *banners* para a publicidade de empresas, também constam promoções lançadas pelo departamento de *marketing,* para divulgação da empresa.

No servidor web onde está armazenada a página é utilizado o sistema operacional Linux com configuração apache permitindo segurança, estabilidade e bom desempenho.

O Banco de Dados utilizado é o *My SQL,* para armazenamento das informações que exige baixa demanda de recursos de *hardware*, é de simples administração, é otimizado para as aplicações típicas da web.

A linguagem de programação na construção do *site* é o *PHP* (*Hypertext Processor)* que fornece controles sobre o uso de memória, arquivos e possuem recursos de orientação a objeto, módulos e bibliotecas prontas disponíveis.

No *design* do *site* é utilizado *Flash* que é um programa gráfico vetorial utilizado para criar animações interativas.

A administração do *site* e alimentação das notícias, é realizado através de uma página de administração na *Internet* implementada pela empresa, em *PHP* a qual uma equipe responsável a utiliza para manutenção.

Há um grande número de acessos à página na *Internet* em média atinge 250.000 acessos diários, muitas informações são armazenadas todos os dias relativos às matérias, aos acessos, às propagandas, às promoções, enfim uma quantidade de dados significativa é adicionada ao banco de dados todos os dias.

Essas informações não são utilizadas pelos administradores para tomadas de decisões. A empresa pode obter um diferencial em seus produtos, trabalhando estes dados, para com isso, os gerentes poderão direcionar as matérias de acordo com o perfil de seus clientes, fazer campanhas de assinatura mais alinhada, verificar quais as regiões que mais lêem o *site*, enfim transformar um simples usuário do *site*, em um cliente personalizado.

1.2. Problemas Diagnosticados

- Não há uma modelagem multidimensional que permite a visão das informações de diferentes ângulos, possibilitando um gerenciamento diferenciado.
- Inexistência de relatórios gerenciais que auxiliem os gerentes a tomarem decisões mais precisas e ter uma análise de seus negócios, direcionando suas vendas e seus produtos à transformar os leitores do *site* em clientes ativos.
- Há falta de consultas ágeis possibilitando informações atualizadas e históricas nos momentos necessários.
- Falta de gráficos gerenciais informando o desempenho do negócio em vários períodos de tempo.
- O Gerente de TI (Tecnologia da Informação) não consegue gerar um perfil para o usuário da página na *internet*, definindo suas necessidades e preferências.

1.3. Objetivos Gerais

Implementar a tecnologia de *Data Mart* com o intuito de gerenciar as informações, criar consultas, relatórios e gráficos gerenciais precisos e ágeis diversificando as combinações

de resultados para a obtenção de informações sobre o perfil e as preferências dos clientes, concedendo aos gerentes uma nova visão dos negócios.

1.4. Objetivos específicos

- Será criada uma modelagem multidimensional, e extraida as informações da fonte de dados para transferir para a base de dados do *Data Mart.*
- Será utilizada uma ferramenta de mercado para apresentação dos dados, que irá fazer as consultas gerenciais e gráficos, realizando vários cruzamentos entre as tabelas e possibilitando consultas complexas.
- Relatórios gerenciais serão criados a partir dos dados contidos no *Data Mart*, que são dados trabalhados, contendo somente as informações necessárias para serem transformados em relatórios para tomadas de decisões.

1.5. Resultados Esperados

Este projeto possibilitará tomadas de decisões na empresa, através de consultas gerenciais e de relatórios, para direcionar o seu produto, o jornal, de acordo com as preferências dos clientes. Proporcionará o aumento da venda de publicidade, de assinaturas, de jornais avulsos e de acessos ao *site*, com isso, alcançando um diferencial competitivo entre os concorrentes.

1.6.Metodologia

Pesquisa sobre arquitetura *Data Warehouse*, ferramentas OLAP (*Online Analytical Processing*), ETL (*Extract Transform Load*), *Data Mart* e todos seus componentes, para utilização dessas tecnologias da forma mais correta.

Análise do negócio da empresa, obtenção das informações importantes para serem adicionadas no *Data Mart*, verificaremos quais as melhores práticas de armazenamento existentes para essa tecnologia facilitando posterior consulta de dados.

Analisaremos as ferramentas ETL, para extração dos dados, e *OLAP* contidas no mercado e, utilizaremos a que considerarmos mais completa, através dessa ferramenta OLAP as informações chave podem ser visualizadas através de qualquer combinação de dimensões, períodos de tempo, regiões, concedendo à equipe gerencial da empresa controle e análise dos dados, geração de relatórios dinâmicos, gráficos, consultas, visualização das informações de diferentes formatos.

CAPITULO 2

2.1 Introdução ao Data Warehouse

As empresas atualmente vêem-se repletas de dados, porém muitas vezes não sabem utilizá-los para tomada de decisões, pois não conhecem uma maneira de agrupar estes dados de modo a apoiá-los na gestão empresarial. (DAL'ALBA, 2006)

Com os dados brutos armazenados em diferentes locais e formatos, não é possível fazer uma profunda análise de negócios, as corporações possuem diversas questões relacionadas no seu âmbito empresarial que necessita de diversas respostas rápidas e eficientes. (SINGH,2001)

Através da proliferação de aplicativos tornou-se mais difícil à integração dos dados provenientes destes, e à medida que a empresa cresce mais informações são geradas dificultando geração de relatórios para se obter informações estratégicas adequadas a certos questionamentos. (INMON, 1997)

O *Data Warehouse* surgiu para solucionar o problema com o histórico dos dados das empresas, informações que muitas vezes eram perdidas são trabalhadas para auxiliar os gerentes a terem uma outra visão de seus negócios. (SINGH,2001)

Ao trabalhar estes dados transformando-os em informações, o DW (*Data Warehouse*) oferece aos administradores pesquisas mais substanciais, precisas e consistentes. Por intermédio dessa tecnologia as empresas podem tomar decisões gerenciais tornando-se mais competitivos, direcionando seus produtos de acordo com as preferências dos seus clientes, através de relatórios estes podem medir o desempenho de seus negócios. (SINGH,2001)

Com o *Data Warehouse* os dados de diversos bancos são armazenados em uma única estrutura, possuindo dados integrados, informações históricas armazenadas por um longo período de tempo, consistentes e orientadas ao assunto, informações não voláteis, ou seja, não podem ser alteradas, diferente de bancos tradicionais os dados no *Data Warehouse* são somente leitura. (SILBERSCHATZ; KORT; SUDARSHAN,1999)

Dados orientados ao assunto, o DW caracteriza-se por unir as informações direcionadas ao negócio da empresa, fornecendo apoio a decisões gerenciais. (SINGH,2001)

É realizada uma análise de todos os sistemas contidos na empresa e de seus dados, para condensar em somente uma base de dados com várias representações. (INMON, 1997)

Os dados são históricos onde uma mesma informação sobre um determinado assunto é armazenada por um longo período, permitindo que os gerentes possam tomar decisões observando seus dados em várias datas, analisando seus comportamentos, seus declínios e sucessos e por ser um banco não volátil, tem-se a certeza que os dados refletem exatamente o que aconteceu em uma determinada época, pois não podem ser alterados. (INMON, 1997)

Os dados ao serem extraídos do banco fonte, são analisados, passam por um processo de transformação e de limpeza eliminando redundâncias, são selecionados permitindo que somente dados importantes para a gestão sejam inseridos no DW, fornecendo informações precisas e seguras. (CIELO4,2006)

2.2. Business Intelligence

Com o surgimento da globalização, o mundo tornou-se mais competitivo e as empresas buscam ter maiores lucros e serem líder de mercado.

A inteligência dos negócios visa à busca da utilização dos dados das organizações para análises, transformando-as em informações, dando-lhes competência para concorrer com o mercado dos negócios. (BARBIERI, 2001)

Essa inteligência permite estimação de riscos e custos, pois coletar as necessidades dos clientes, suas preferências de compras, de condições de pagamento, permite também a criação de metas realistas para serem alcançadas pela empresa. (CIELO3,2006)

Com o avanço tecnológico surgiram vários sistemas para auxiliar os administradores nas tomadas de decisões, como por exemplo, o *Data Warehouse*, que proporciona ao usuário uma visão global de seus negócios e planejamento de estratégias com o intuito de crescimento empresarial. (CIELO3,2006)

Esses sistemas utilizam a capacidade analítica das ferramentas que são implementadas para ele, de forma que armazene em um só lugar todas as informações necessárias para o processo decisório, com isso, torna-se uma ferramenta de grande diferencial competitivo entre as empresas. (CIELO3,2006)

O BI (*Business Intelligence)* é uma tecnologia que pode ser adotada para atingir metas, tais como: novas oportunidades, auxiliar com uma visão mais profunda do consumidor, aumentar vendas, reduzir custos, ajustar orçamento e substituir relatórios e procedimentos de relatos estáticos por dinâmicos. Com ela as empresas podem organizar grandes quantidades de dados, de forma rápida, minuciosa e com precisão analítica para a tomada de decisões.

E o *Data Warehouse* vem por trás das estratégias de BI estimulando as vantagens de se ter construído uma base de conhecimento ao longo do tempo. As empresas investem constantemente tempo, dinheiro e recursos de TI (Tecnologia da Informação) para manter suas margens de competitividade. (CIELO3,2006)

Essa integração dos dados permitirá a administração de informações essenciais no negócio e a identificação das mais importantes tendências que acontecem na empresa.

2.3. Abordagens para o projeto de um DataWarehouse/Data Mart

2.3.1 Abordagem segundo Bill Inmon

Segundo a abordagem de Inmon, o *Data Warehouse* deverá ser modelado de maneira atômica e normalizado a todo o negócio da empresa. O DW será o grande repositório de dados, que servirá de fonte para diversos tipos de aplicação, como os *Data Mart*. O *Data Warehouse* e o *Data Mart* que possuem estruturas essencialmente diferentes e um conjunto de *Data Mart* dificilmente será integrado, e mesmo se for, não resultará em um DW. (INMON, 1997)

Inmon defende a idéia da criação em primeiro lugar do *Data Warehouse*, depois as dos *Data Mart*, pois uma vez que os dados da organização já estão integrados no DW os *Data Marts*, que atenderiam a níveis departamentais, gerariam os dados íntegros e corporativos. (INMON, 1997)

2.3.2 Abordagem segundo Ralph Kimball

Kimball tem como metodologia a criação de *Data Mart* não a nível departamental, mas sim orientado aos dados ou fonte de dados. Para ele os DM (*Data Mart*) são um subconjunto lógico do DW, sendo assim a junção de todos os *Data Mart* forma o *Data Warehouse*. (KIMBAL; ROSS, 2002)

Com essa idéia o *Data Warehouse* é projetado em pequenos módulos, pelos quais serão projetados de maneira bem mais reduzida, tendo o enfoque em áreas ou assuntos específicos e serão reunidos de acordo com o tempo. (KIMBAL; ROSS, 2002)

Kimball adotou o esquema estrela (Star Schema), que tem como objetivo transformar os dados em tabelas fatos, onde os dados são reunidos por áreas de interesse, que poderão sofrer manipulações numéricas e estatísticas; e também em tabelas dimensões, tabelas auxiliares, que contém chaves de entrada do modelo, inclusive informações descritivas de cada dimensão. (KIMBAL; ROSS, 2002)

2.4. Fases para o desenvolvimento do Data Mart

Para o desenvolvimento de um *Data Mart* é necessário que sejam levantados todos os dados com relação ao sistema existente e com entrevistas com os usuários atuais do sistema, de forma a verificar como está funcionando atualmente e como eles gostariam que funcionasse no futuro. (BARBIERI, 2001)

Para isso é necessário que o projeto passe por algumas fases, que são:

2.4.1 Planejamento – Deverá ser focado o negócio da empresa, de forma a ser montado o escopo do projeto do DM. Deverá ter prioridade às informações gerenciais urgentes, buscando em especial em área de risco da empresa. (BARBIERI, 2001)

2.4.1.1 Definição da Arquitetura Tecnológica

Este requisito é de grande importância para um projeto de *Data Mart*, pois nele são definidos os níveis de serviços e graus de compromissos nas devidas ações do sistema, com isso serão definidas a performance e disponibilidade do sistema.

Os principais componentes tecnológicos básicos são:

* Ferramenta para processos de Extração, Transformação e Carga (ETL): ferramentas com o objetivo de realizar os processos de transformação do dado, aliado a integridade dos dados fontes se tornam fatores críticos de sucesso no processo de *Data Mart*;

* É necessário um SGBD (Sistema Gerenciador de Banco de Dados), para a implementação de tal sistema, é necessários uma máquina robusta e um ambiente de alta disponibilidade, desempenho e segurança.

2.4.2 Levantamento das Necessidades – será necessário levar em conta o modelo dimensional e o modelo fonte de dados. O modelo dimensional é uma representação de blocos conceituais de dados que são de extrema importância devido ao fato de adquirirem os objetivos do sistema de suporte a decisão. (BARBIERI, 2001)

O modelo de dados tem como objetivo relacionar-se com a fonte de dados existentes, de forma a registrar os blocos de dados, pela qual deverá ser incluída as descrições como também as formas atuais para o armazenamento que são utilizados pelo sistema atual. (BARBIERI, 2001)

Para tal análise, serão necessárias a presença do analista de sistemas, usuários e a equipe de tecnologia da informação da empresa.

2.4.3 Modelagem Dimensional – Deve ser feita uma estrutura dimensional para que possa ser visualizada em cubo, para ter uma estrutura padronizada e para que os acessos aos dados ocorram com um alto desempenho. (KIMBAL; ROSS, 2002)

Para isso, ficam dividas em três etapas o processo de criação do modelo dimensional, que são:

1- Processo de Negócio – Deverá ser verificado todo o processo de negócio a ser modelado, para isso serão feitas reuniões com os usuários com o intuito de capturarem dados que darão origem ao modelo. (KIMBAL; ROSS, 2002)

2- Granularidade – Essa fase deve ser especificada o que cada linha da tabela fato representa e suas especificações. (KIMBAL; ROSS, 2002)

3- Dimensões – Após ter definido bem a granularidade do projeto, as dimensões serão facilmente identificadas e logo após serem identificadas deverão ser listadas em forma de texto. (KIMBAL; ROSS, 2002)

2.4.4 Projeto físico do Banco de Dados – Nessa fase do projeto deverão ser evidenciadas as estruturas lógicas do modelo dimensional, juntamente com as definições de tabelas fato e tabelas dimensões, relacionamentos, indexação, atributos de tabelas e implantação de regras. (KIMBAL; ROSS, 2002)

2.4.5 Projeto de Extração, Transformação e Carga (Ferramenta ETL) – Nessa fase do projeto ele tem um objetivo de extrair os dados do sistema que está sendo utilizado atualmente, e a partir dessa extração deverá ser trabalhada e transformada da melhor maneira para que possam ser armazenados os dados, de forma que, posteriormente possam ser acessados eficazmente pelo *Data Mart*. (KIMBAL; ROSS, 2002)

2.4.6 Desenvolvimento de aplicação – Utiliza-se uma ferramenta com uma interface amigável ao usuário, na qual deverá produzir relatórios, condições para a visualização dos dados de forma multidimensional e também proporcionar a importação desses dados para planilhas e processadores de texto para posterior utilização do usuário.

2.4.7 Validação e teste – Nessa parte do desenvolvimento do projeto, o sistema sofrerá várias simulações de volume e de processamento por grupos de usuários determinados. Com

isso, será analisado, de forma que possa ou não ser aprovado e entregue ao ambiente produtivo.

2.4.8 Treinamento – Deverá ser feito treinamento de gerentes e principais usuários do sistema, de forma que todos os usuários fiquem cientes das funcionalidades do sistema.

2.4.9 Implantação – Para que possa ocorrer à implantação do sistema, deverá ser feito anteriormente um acompanhamento rigoroso das aplicações existentes, de forma que sejam levantadas as informações críticas e sugestões que serão feitas pelos usuários.

2.5. Arquitetura do Data Warehouse

Na Figura 1 está ilustrada a representação dos elementos de um *Data Warehouse*, que são fontes de dados dos sistemas operacionais de origem, área de representação de dados, apresentação dos dados e ferramentas de acesso aos dados. (KIMBAL; ROSS, 2002)

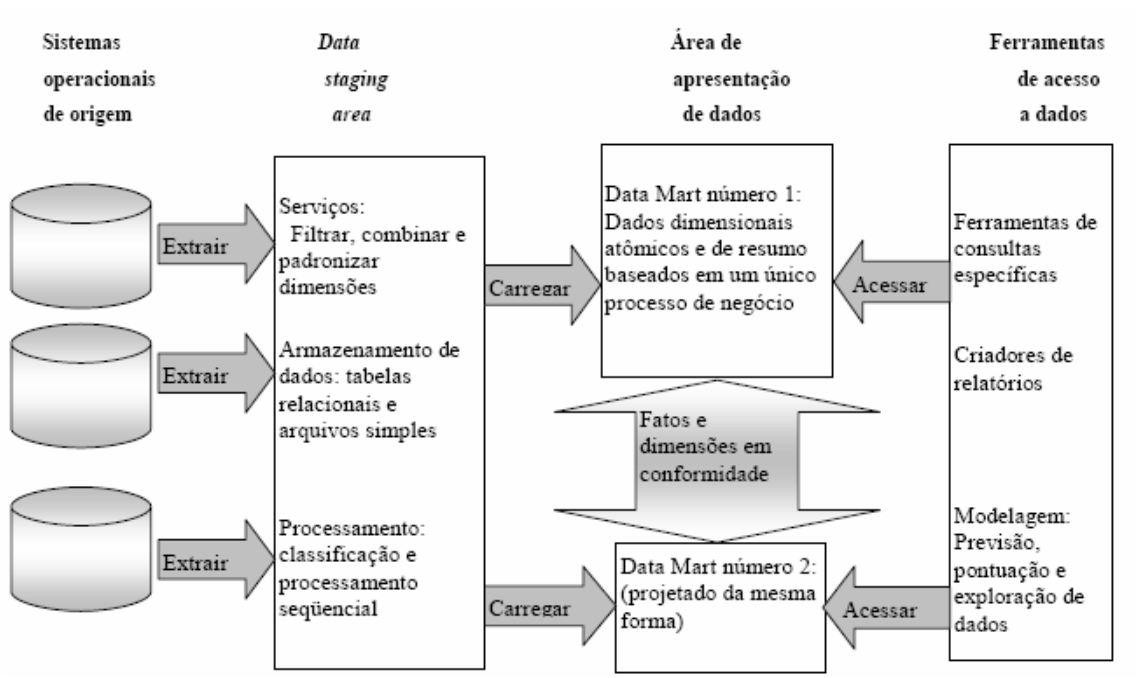


Figura 1-Elementos básicos do *Data Warehouse*

Fonte : (KIMBAL; ROSS, 2002).

2.5.1 Sistemas Operacionais de origem

Os sistemas fonte armazenam dados dos processos de negócios das empresas, provenientes de seus sistemas utilizados para realizar as operações diárias das empresas. A principal prioridade para um sistema fonte é a apresentação e armazenamento de dados dos processos, muitas vezes essas informações não são utilizadas para certos questionamentos devido às restrições dos bancos utilizados. (KIMBAL; ROSS, 2002)

Para criação de relatórios com os sistemas fontes é demorado e talvez não eficiente, pois levaria muito tempo para unir as informações necessárias provindas de várias tabelas, teria que localizar os dados necessários, obter essas informações, o que poderia, dependendo da complexidade, levar dias para ficar pronto. (INMOM,2002)

Esses dados operacionais são baseados em aplicações, detalhados, podem ser alterados, atendem à comunidade funcional, às necessidades cotidianas, pois é voltado às aplicações. (INMOM,2002)

Existem enumeras diferenças entre dados primitivos provenientes de fontes de dados e dados derivados utilizados no *Data Warehouse*, cada um tem uma função diferente, um voltado para o funcional e o outro para o gerencial. (INMOM,2002)

2.5.2 Área de Representação de Dados

A área de representação dos dados é onde os dados são limpos, combinados, padronizados, transformados, arquivados e exportados para os DM, é conhecido como extração, transformação e carga. (KIMBAL; ROSS, 2002)

Primeiramente deve-se conhecer e entender a base de dados fonte para a extração do dado, que é necessário para conter no *Data Warehouse*, suas origens podem ser de várias fontes e formatos. (KIMBAL; ROSS, 2002)

Essas seleções dos dados devem ser bem analisadas, pois é necessário integrar informações de várias tabelas para compor um único campo no *Data Warehouse.* (CIELO4,2006)

Depois de definida as fontes estas são extraídas para manipulação, em seguida deve-se limpar os dados retirando inconsistências, transformar analisando os formatos dos dados, pois as informações vêm de várias tabelas e uma informação poderá estar representada de várias formas, então deve-se padronizar combinação dos dados de várias fontes em uma somente. (CIELO4,2006)

Também é atribuída uma chave aos dados para facilitar as operações do DW, pois os dados são redundantes por ser histórico, necessitando de identificação. (KIMBAL; ROSS, 2002)

Após feita essas etapas, os dados podem ser migrados para o banco de dados do *Data Warehouse*.

2.5.3 Área de Apresentação dos Dados

A área de apresentação dos dados é o lugar onde os dados estão armazenados, organizados após extração, essas informações ficam disponíveis para criação de relatórios, consultas gerenciais. (KIMBAL; ROSS, 2002)

Nela estão os dados importantes para tomadas de decisões para que os gerentes possam resolver questionamentos. Esses dados são acessados através de ferramentas de acesso aos dados. (KIMBAL; ROSS, 2002)

A área de apresentação é uma série de dados integrados representando os processos de negócio da empresa, sua modelagem é multidimensional onde os dados são armazenados em forma de um cubo, possibilitando várias visões e combinações de consultas. (KIMBAL; ROSS, 2002)

2.5.4 Ferramentas de Acesso aos dados

Ferramentas de acesso aos dados ou ferramentas OLAP são aplicações utilizadas para que os usuários possam ter acesso às informações do Data Warehouse, estes geram relatórios gerenciais para apoio à tomada de decisões.(CIELO2 ,2006)

As ferramentas OLAP permitem aos gerentes analisar os dados corporativos de forma rápida, agilizando os processos de decisões. É feita uma análise multidimensional dos dados permitindo uma nova visão da realidade da empresa. Essa tecnologia pode ser tanto cliente-servidor quanto web oferecendo analises históricas, perfil dos clientes e visões personalizadas. (SOUZA1,2003)

Uma das características dessa ferramenta é que as consultas podem ser feitas pelo próprio usuário de acordo com suas necessidades, através do cruzamento dos dados tem-se uma visão em diferentes níveis de detalhe em vários períodos, em diferentes regiões demográficas, por exemplo. (CIELO2, 2006)

2.6. Data Marts

O Data Mart pode ser definido como uma solução apropriada às grandes e médias corporações, uma vez que representa um tipo de DW menos complexo em termos de implementação e mais simples de ser gerenciado, pois têm requisitos menos complexos em termos de infra-estrutura e de abrangência funcional. Os Data Marts são geralmente utilizados nas áreas principais da empresa como recursos humanos, finanças e marketing, são mais rápidos de se implementar e custam menos. (SASSI, 2006)

O Data Mart trata de uma área local, departamental, e um Data Warehouse envolve toda a companhia permitindo suporte à decisões em todos os níveis da organização.

Os Data Marts tem uma característica importante que é o curto tempo de desenvolvimento e implementação, pois o escopo é menor que em um DW, no DM os dados estão mais próximos do usuário, ou seja, os dados estarão de fácil acesso aos usuários, de forma rápida e consistente.

Com o Data Mart é possível que uma organização analise os seus dados históricos em banco de dados multidimensionais e com isso obter suas informações de maneira mais coerente com a realidade da empresa. Com isso a empresa terá um diferencial, pois a tornará mais competitiva. (PEREIRA, 2000)

2.7. Metadados

A tecnologia de metadados surgiu devido às organizações necessitarem conhecer melhor os dados que elas mantêm.

Os metadados são dados que descrevem dados, provém uma descrição concisa a respeito dos dados. Os dados podem ser documentos, coleção de documentos, gráficos, tabelas, imagens, vídeos, entre tantos outros. (VAZ ; MARCON, 2004)

Metadados podem ser definidos também como uma abstração dos dados ou dados de mais alto nível que descrevem dados de um nível inferior, tem um papel muito importante na administração de dados, mas podem ser considerados de suma importância, pois é a partir deles que as informações serão processadas, atualizadas e consultadas. (SOUZA2, 2003)

As informações de como os dados foram criados ou derivados, ambiente em que reside ou residiu, alterações feitas, entre outras são obtidas de metadados fornecendo os recursos necessários para entender os dados através do tempo. (VAZ ; MARCON, 2004)

Geralmente os metadados podem ser apresentados em três camadas diferentes, que são:

- **Metadados Operacionais:** Definem a estrutura dos dados mantidos pelos bancos operacionais, usados pelas aplicações de produção da empresa.

- **Metadados Centrais para um DM:** Devem ser orientados por assunto e definirem como os dados transformados devem ser interpretados, incluindo definições de agregação e campos calculados, assim como visões sobre cruzamentos de assuntos.

- **Metadados do Nível do Usuário:** Organizam os metadados do DM para conceitos que sejam familiares e adequados aos usuários finais.

2.7.1 Modelo de Dados

O modelo de dados tem um papel fundamental para o desenvolvimento interativo tanto do DM como para todo o DW. Quando os esforços de desenvolvimentos são baseados em um único modelo de dados, sempre que for necessário unir estes esforços, os níveis de sobreposição de trabalho e desenvolvimento desconexo serão muito baixos, pois todos os componentes do sistema estarão utilizando a mesma estrutura de dados (SOUZA2, 2003).

2.8. Grão e Granularidade

Antes de conceituarmos grão é importante ter o conceito de modelagem multidimensional, dimensão e fato, pois ao se tratar de Data Warehouse será necessário entender cada item deste a fim de se desenvolver uma estrutura sólida e confiável.

A modelagem multidimensional consiste em uma abordagem lógica que dispõe de dados de forma intuitiva e que possibilite um acesso com alto desempenho. (KIMBAL; ROSS, 2002) Utiliza-se de princípios do modelo relacional, porém cada representação é composta por uma tabela com múltiplas chaves (Fatos) e um conjunto de tabelas que interagem com o fato (Dimensão).

Os fatos são medições numéricas do negócio a partir da intenção das dimensões a ela vinculadas podendo ser armazenadas em uma ou mais tabelas de fatos. Os fatos podem ser classificados como básicos ou derivados. Os básicos são aqueles obtidos diretamente da base de dados e os derivados são aqueles oriundos de cálculos e transformações sobre os dados básicos. A tabela de fatos sozinha não tem nenhuma informação que possa ser aproveitada pelo usuário. Nela estão armazenados os valores de maior relevância para o negócio. (KIMBAL; ROSS, 2002)

A dimensão consiste em uma coleção de atributos, utilizadas para caracterizar as medidas presentes no fato. Os atributos representam as fontes das restrições mais interessantes e os cabeçalhos de linhas no conjunto de respostas ao usuário. Os atributos estão relacionados diretamente a qualidade do DW, pois quanto melhor a descrição dos atributos melhor será o resultado do DW. (ANDRADE ; SANTIAGO, 2003)

O grão é o nível fundamental de dados que será analisado, ou seja, é o menor nível de detalhe da informação contida no fato ou na dimensão. (KIMBAL; ROSS, 2002)

A granularidade é a combinação dos grãos existentes em um fato ou em uma dimensão, ela na tabela de fatos é um fator de influência no desempenho do *Data Warehouse*. Diz respeito a um total calculado a partir de um determinado fator, seja ele temporal ou qualquer outra forma de dimensão mensurável que irá influenciar na quantidade de informações armazenadas na tabelas de fatos, sendo assim, quanto maior o nível de detalhes menor o nível de granularidade e maior o espaço em disco para armazenamento. (ANDRADE; SANTIAGO, 2003)

Cada dimensão possui sua própria granularidade sendo que ela não pode ser menor que a granularidade da tabela fato a ela associada.

CAPITULO 3

3.1. Modelagem Multidimensional

A modelagem multidimensional é utilizada para facilitar e criar as consultas para os padrões em um *Data Warehouse*.

No modelo é utilizado uma tabela fato e várias tabelas de dimensões interligadas, que associam o fato a uma dimensão do cubo. (ANDRADE; SANTIAGO, 2003)

Na modelagem multidimensional existem medidas numéricas para análises gerenciais, essas medidas encontram-se na tabela fato ligado a uma dimensão, possibilitando uma visão em cubo. (RIANE,2002)

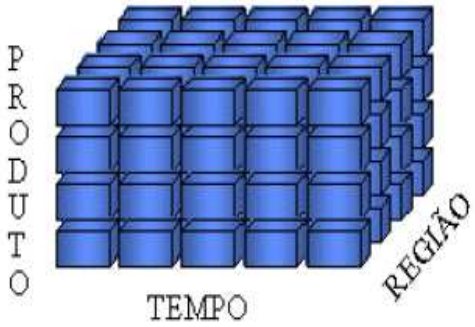


Figura2 -Cubo de três dimensões

(ANDRADE; SANTIAGO, 2003)

No Exemplo acima temos um cubo de três Dimensões, produto, tempo e região possibilitando diversidade de consultas, combinando essas dimensões para consultas gerenciais, possibilitando fatiar esse cubo e aprofundar cada dimensão para obtenção de informações mais detalhadas.

“O modelo dimensional permite visualizar dados abstratos de forma simples e relacionar informações de diferentes setores da empresa de forma muito eficaz.” (MOREIRA, 2006) o DW reúne as informações da empresa em um só banco de forma dimensional que permite consultas rápidas e eficientes.

As informações no modelo dimensional são relacionadas de forma a serem representadas por um cubo, assim pode-se aprofundar nas consultas, criando diferentes combinações de informações e extrair mais detalhes para tomadas de decisões. (MOREIRA,2006)

“Sistemas de gerenciamento de banco de dados multidimensionais (às vezes chamados de Datamart) proporcionam uma estrutura de sistemas de informação que permite que uma empresa disponha de acesso muito flexível à dados(...)” (INMON, 1997 p. 173), no modelo multidimensional o usuário tem uma disponibilidade maior de consultas, através da combinação das informações, agrupando os dados, criando consultas de diferentes formatos, possibilitando acesso à informações gerenciais, permitindo uma visão diferenciada dos negócios da empresa.

3.1.1 Tabela de Fatos

As tabelas de fatos contêm medidas resultantes dos processos de negócios da empresa a partir da interseção das dimensões. As medidas mais utilizadas são numéricas, pois assim pode-se representar a medida dos negócios, como total de vendas, essas medidas podem ser um cálculo realizado entre mais de uma informação e armazenado na fato. (KIMBAL; ROSS, 2002)

As fatos são constituídas das chaves de dimensões e valores relacionados ao negócio, ela é a principal no modelo multidimensional. (ANDRADE ; SANTIAGO, 2003)

Normalmente o contexto dos atributos da tabela fato está nas várias tabelas de dimensões, que estão ligadas pela chave estrangeira. (KIMBALL ; CASSERTA, 2004)

3.1.2 Tabela de Dimensões

Nas tabelas de dimensão estão contidas descrições textuais provindas do banco de dados fonte, contendo as características de algo que será utilizado para pesquisas no DM. (MOREIRA, 2006)

As informações das tabelas de dimensões devem estar bem descritas e completas para gerar melhores resultados nas consultas, quanto mais detalhadas, mais informações os gerentes terão para tomar decisões. (ANDRADE; SANTIAGO,2003)

Os atributos da dimensão desempenham um papel vital, pois nelas estão as fontes utilizadas para compreensão das consultas e relatórios no DM, contém as descrições dos negócios. (KIMBAL; ROSS, 2002)

A qualidade do DM depende da dimensão, deve-se fornecer dados completos, povoar a tabela com informações precisas, deixando as dimensões com as informações necessárias para resultar em um melhor DM. (KIMBAL; ROSS, 2002)

CAPITULO 4

4.1. ETL (*Extract, Transform and Load*)_ Extração, Transformação e Carga

Esta etapa é uma das fases mais críticas de um *Data Warehouse*, pois envolve a fase de movimentação dos dados. A mesma se dá basicamente em três passos: extração, transformação e carga dos dados, essa etapa é trabalhosa, complexa e também muito detalhada, embora existam várias ferramentas que nos auxiliam na execução desse trabalho.

Nesta etapa serão extraídas as informações da base de dados fonte para o *Data Mart*, fazendo a extração, transformando através da manipulação das informações, enviando ao DM apenas o que for necessário e por fim a carga, enviando os dados previamente filtrados.(CIELO1,2006)

A extração consiste no primeiro passo a ser executada para um DM, é a definição das fontes de dados, as fontes podem ser várias e de diferentes formatos, desde planilhas até arquivos do Dbase.

Ao selecionar os dados a serem extraídos, estes muitas vezes poderão incluir lixo e inconsistências, sendo o papel da transformação limpar estes dados e padronizá-los para que qualquer dado tenha um formato uniforme. (SINGH,2001)

Após a limpeza e a padronização é realizada a carga dos dados devidamente consistentes para o DW.

4.1.1 Extração

Na extração são verificadas as fontes de dados da empresa que irá extrair campos específicos de dados, esta extração poderá ser complexa, pois muitas vezes será necessário extrair vários campos de um sistema transacional para compor um único campo no *Data Warehouse*, podendo extrair arquivos inteiros dependendo das necessidades ao suporte à decisão. (SINGH,2001)

A extração de dados do ambiente operacional para o ambiente de *Data Warehouse* demanda uma mudança na tecnologia. Os dados são transferidos de bancos de dados hierárquicos ou de bases do grande porte, como o SQL Server, para uma nova tecnologia de SGBD para *Data Warehouse*, tal como do SQL Server para DW.

Neste ponto serão extraídos os dados do *site* que estão armazenados em apenas um banco, na qual irá se conectar ao banco da empresa e coletará as informações levando-as ao *Data Mart* para possibilitar as consultas e relatórios gerenciais.

4.1.2 Transformação

Os dados são transformados para se adequar às necessidades e retirar possíveis inconsistências das informações para serem enviadas ao DM, para isto, é necessário planejamento detalhado. (SINGH,2001)

Uma parte importante da transformação é trabalhar com o conteúdo dos dados de tal forma que retire as redundâncias dos dados, para isto é necessário estabelecer regras e padrões com intuito de criar uma fonte de dados verdadeira. (SINGH,2001)

As transformações feitas são integrações que reúnem as diversas informações de vários bancos e une em apenas um, sendo a limpeza e a validação os processos mais importantes, pois estão diretamente relacionados com a qualidade dos dados, derivação e sumarização, atualização do histórico e ordenação. (COSTA; ANCIÃES, 2001)

4.1.3 Carga

Refere-se ao povoamento das informações no *Data Mart*, ou seja, depois dos dados selecionados e filtrados os dados são incluídos no DM, para esta etapa deve-se estar certificado que os processos anteriores estejam de acordo com os requisitos do DM.(COSTA; ANCIÃES, 2001)

4.1.4 Ferramentas de mercado

As ferramentas de ETL mais utilizadas no mercado são o Data Stage da Ardent, agora adquirido pela Informix, o ETI da IBM, Sagent da própria Sagent, Informática Power Conect da Informática, o DTS da Microsoft, neste projeto será utilizada a ferramenta do SQL Server para extrair os dados para o *Data Mart*.

Algumas ferramentas têm a curva de aprendizado mais suave, outras um pouco mais íngreme, mas em certos casos mesmo sendo uma ferramenta de difícil aprendizado exigindo maiores investimentos em pessoal, serão compensados com o desempenho e flexibilidade da mesma.

Há outras ferramentas que tem custo zero de aquisição, pois vem embutida junto com um SGBD, Sistema Gerenciador de Banco de Dados, são um pouco limitadas no que se refere às extrações de dados e exigem uma maior codificação dos processos de ETL.

CAPITULO 5

5.1. OLAP (Processamento Analítico On-Line)

OLAP é um software que possui uma tecnologia de construção, na qual os analistas de negócios, gerentes e executivos conseguem analisar e visualizar os dados corporativos de forma rápida, consistente e interativa. A sua principal característica é fazer uma análise dinâmica e multidimensional dos dados consolidados, permitindo ao usuário final que desenvolva atividades tanto analíticas quanto navegacionais. (SINGH, 2001)

A tecnologia OLAP é geralmente implementada em ambiente multiusuário e cliente/servidor, e recentemente vem sendo usada em ambiente Web, com isso, ela é capaz de oferecer respostas rápidas às consultas com cruzamento de informações sem se preocupar com tamanho do banco e sua complexidade. (SINGH, 2001)

Por essa razão, a ferramenta é essencial para a implementação de um sistema de suporte a decisão, na qual, define um conjunto de requisitos funcionais que são utilizados por usuários para aplicações de planejamento, relatórios e análise da empresa. (SINGH, 2001)

Os bancos de dados relacionais que possuem estrutura bidimensional (linhas e colunas), não possuem uma ferramenta de visualização multidimensional dos dados, com isso, as ferramentas OLAP têm ganhado uma grande importância, pois possuem sofisticadas funções analíticas e representam os dados multidimensionalmente ao usuário. Por ela obter essas características está sendo introduzido em sistemas que possuam banco de dados relacional, na qual elas possam montar relatórios gerenciais obtidos de análises complexas retiradas do banco de dados. (SINGH, 2001)

As ferramentas OLAP são classificadas como:

5.1.1 OLAP Multidimensional

MOLAP é a forma clássica do OLAP e algumas vezes é referida como simplesmente OLAP. MOLAP usa as estruturas de bancos de dados que são atributos genéricos otimizados como período de tempo, localidade, produto ou código de conta. A forma como cada dimensão vai ser agregada é definida a princípio por uma ou mais hierarquias. (SINGH, 2001)

5.1.2 OLAP Relacional

ROLAP trabalha diretamente com bancos de dados relacionais. Os dados e as tabelas de dimensões são armazenados como tabelas relacionais e novas tabelas são criadas para receber a informação agregada. (SINGH, 2001)

5.1.3 OLAP Híbrido

Não há consenso claro na indústria sobre o que é "OLAP Híbrido", exceto que um banco de dados vai dividir dados entre armazenamento relacional e especial. Para alguns fornecedores, por exemplo, um banco de dados HOLAP usa tabelas para manter as grandes quantidades de dados detalhados, e um armazenamento especializado para pelo menos alguns aspectos dos dados das quantidades menores que são menos agregadas ou detalhadas. (SINGH, 2001)

CAPITULO 6

6.1Estudo de Caso

O caso de uso que será trabalhado nesse capítulo é focado no desenvolvimento de um *Data Mart*, a partir dos dados armazenados da página na *Internet* de uma empresa, com o intuito de possibilitar o gerenciamento das informações vindas do *site* do jornal através de gráficos e relatórios e com grandes possibilidades de combinações de consultas.

6.2 Apresentação do Problema

O *site* da empresa de edição de jornais é constituído de matérias publicadas no jornal impresso, que contém um resumo com as principais notícias e, que são atualizadas diariamente. Para o cliente acessar estas matérias, ele deverá cadastrar-se no *site* para assim, poder ler as matérias e colunas de uma certa edição do jornal. No *site* também são inseridas promoções, que são realizadas pelo departamento de marketing, com o intuito de divulgar a corporação e também para fazer com que o cliente conheça melhor seu produto final.

Devido o grande número de acessos ao *site*, e da quantidade de informações que são geradas, verifica-se que estas informações podem ser aprimoradas, de forma que, possa esclarecer dúvidas e servir como apoio nas tomadas de decisões. Portanto, constatou-se que a necessidade dos gerentes da empresa é justamente a de possuir uma ferramenta que facilite as consultas mensais, trimestrais, anuais, semanais, em relação às notícias, promoções, perfis de cliente, quantidade e duração de acessos ao *site*.

O DM será uma ferramenta que auxiliará os gerentes, concedendo-lhes informações de seus clientes, na qual ele poderá identificar quais são as matérias mais acessadas, se uma

promoção que foi colocada no *site* foi explorada pelos usuários, verificando qual a faixa etária de idade das pessoas que lêem uma coluna, dentre outras questões.

O projeto do DM, é direcionado de acordo com o negócio do Jornal, para identificar as preferências de seus usuários e possibilitar uma visão diferenciada do perfil dos clientes que acessam o *site*, e assim auxiliar os gestores a direcionar seus produtos de acordo com as preferências dos usuários.

6.3 Descrição do Ambiente e suas características

No servidor onde está armazenado a página da empresa na *Internet* é utilizado o sistema operacional Linux, com banco de dados MySQL, a linguagem de programação utilizada é o PHP com auxílio da ferramenta *Dreamweaver*.

Para este projeto iremos utilizar o ProClarity que é uma ferramenta OLAP para análise e visualização de informações em associação com a plataforma de *business intelligence*, para que utilização desta ferramenta, transferimos a base de dados original do MySQL para o SQL Server 2000, pois o Pro Clarity somente possue compatibilidade com esse SGBD.

A ferramenta ETL utilizada é o *SQL SERVER*, o próprio banco de dados onde está armazenada a fonte de dados, devido este SGBD possuir recursos de ETL, que irá selecionar os dados na fonte e armazenar, estas informações trabalhadas, nas tabelas criadas dentro do banco desenvolvido para o DM.

6.4 Descrição do Modelo de Dados

O modelo é constituído de uma tabela fato e seis dimensões para gerenciamento das informações da página na *Internet* da empresa de jornais.

Na figura 3 consta o modelo multidimensional para o *Data Mart (Modelo Estrela)* utilizado para o desenvolvimento do DM. A partir do modelo desenvolvido, a empresa terá condições obterem informações gerenciais, auxiliando-os nas decisões através dos dados extraídos e trabalhados provenientes de sua página na *internet*.

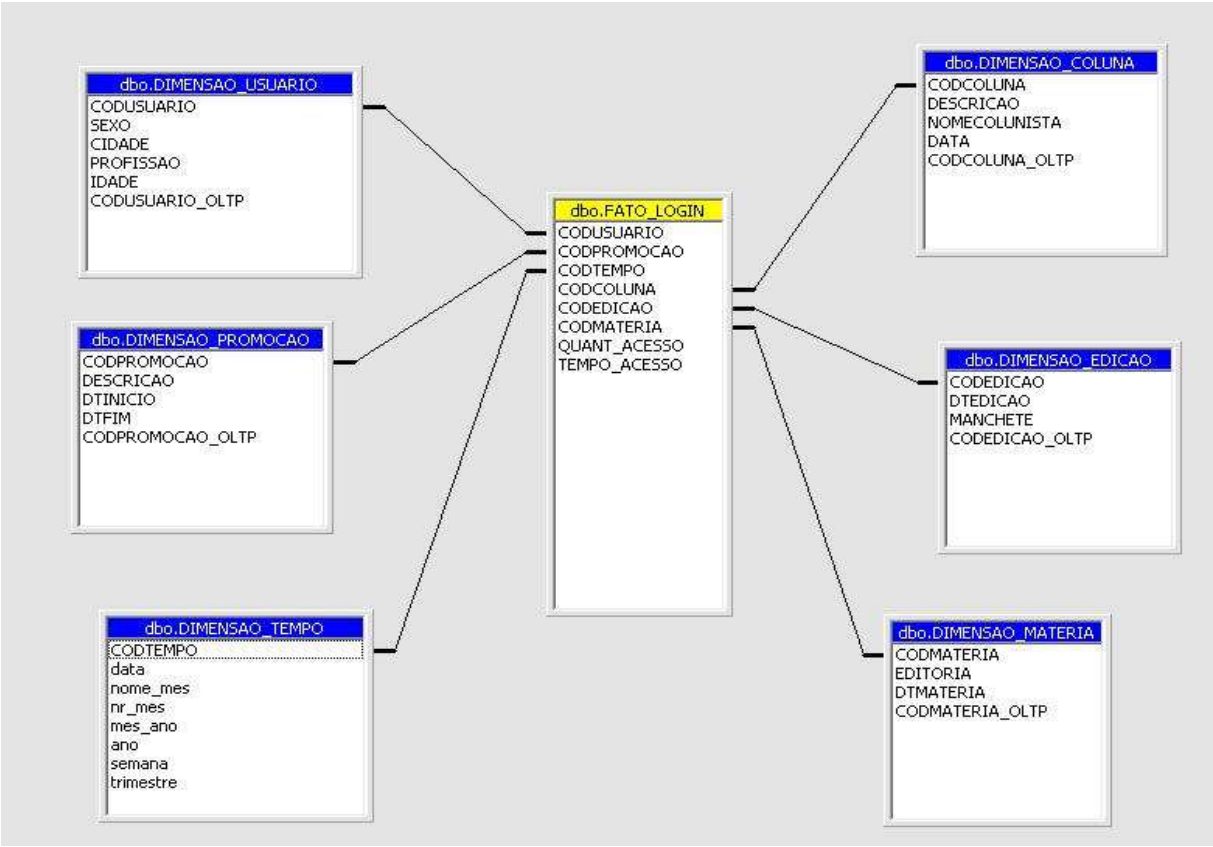


Figura 3: Modelo Multidimensional

6.4.1 Fatos

A tabela fato foi baseada nas divisões que a empresa faz em sua página da *Internet* também utilizada no seu jornal impresso, evidencia as páginas que os leitores do *site* visitaram a data e o tempo de permanência do mesmo no *site*.

I. FATO_LOGIN

É encarregada de armazenar informações gerais de todo o *site*, está relacionada às dimensões usuário através da chave estrangeira CODUSUARIO, da Tempo através da chave estrangeira CODTEMPO, das promoções proporcionadas no *site* pela equipe de Marketing da empresa através da chave estrangeira CODPROMOCAO.

A tabela fato também traz referência à dimensão COLUNA através da chave CODCOLUNA, a dimensão Edição que consta as informações das edições diárias publicadas pelo jornal, esta se liga a fato pela chave CODEDICAO, as matérias do *site* é um resumo do jornal impresso, elas dividem-se em Editorias, são feitas diariamente, estão referenciadas na fato através da chave estrangeira CODMATERIA.

As métricas constituí-se de qtde_acesso que é a quantidade de acessos dos usuários ao *site*; tempo_acesso é a duração que o usuário permaneceu no *site*, esta duração é realizada em minutos.

6.4.2 Dimensões

I. Usuario

Essa dimensão tem contém informações pessoais do usuário do *site*. Com está dimensão temos informações sobre a idade, sexo e profissão do usuário, afim de

identificarmos suas preferências e podermos oferecer consultas direcionadas e identificar qual a classificação das pessoas que acessam o *site*.

Atributos da dimensão:

- codusuario: código interno que é inserido automaticamente e sequencialmente quando o usuário se cadastra no *site* para acesso ao mesmo.
- Sexo: O sexo do usuário.
- Cidade: A cidade da residência do usuário.
- Profissao: A profissão exercida pelo usuário.
- Idade: A idade do usuário esta data é calculada internamente através da data de nascimento no ato do cadastro do mesmo no *site*.
- CodUsuario_OLTP: é a chave que interliga a dimensão Usuário com tabela, da base de dados origem, Usuário.

II. Materia

Esta dimensão contém informações a respeito de cada matéria, estas são escritas por repórteres e são realizadas diariamente, possue uma divisão por editorias para facilitar a busca pelo usuário de acordo com sua preferência, por exemplo, editorias Brasil, mundo, cidades, esportes, em cada uma delas contém as matérias.

Atributos da dimensão:

- CodMateria: O código da matéria publicada .
- Editoria: O nome da editoria responsável pela matéria.
- DtMateria: A data em que a matéria foi publicada.

- CodMateria_OLTP: é a chave que interliga a dimensão Materia com a tabela, da base de dados origem, Materia.

III. Promoções

A dimensão promoções consta informações que se referem às promoções publicadas no *site* desenvolvidas pelo departamento de marketing da empresa, para divulgação de seus produtos e para atrair os usuários a visitar o *site*. Seus atributos são:

- CodPromocao: Código da promoção.
- Descricao: Uma breve descrição referente à promoção.
- DtInicio: Data de inicio da promoção.
- DtFim: Data do final da promoção.
- CodPromocao_OLTP: é a chave que interliga a dimensão Promocao com a tabela, da base de dados origem, Promocao.

IV. Tempo

Na dimensão tempo é definido o período para auxilio na medição da fato, para identificar se a consulta vai ser mensal, diária, anual, semanal, de acordo com às necessidades da empresa. Seus atributos são:

- Codigo: Código do tempo.
- Data: Data composta por dia, mês e ano.
- Nome_mes: os nomes dos meses.

- Nr_mes: o número correspondente ao mês.
- Mês_ano: Medição composta por mês e ano.
- Ano: Medição de tempo anual.
- Semana: Medição de tempo semanal.
- Trimestre: Medição de tempo trimestral.

V. Coluna

Esta dimensão contém informações sobre as colunas publicadas no *site*, as colunas são feitas por colunistas, essas matérias nem sempre são feitas diariamente e são uma crítica escrita de assuntos variados, reflete a opinião do colunista.

- Codigo: Código da coluna
- Descricao: Uma breve descrição referente à coluna.
- NomeColunista: O nome da pessoa que escreveu a coluna.
- Data: A data em que a coluna foi publicada.
- CodColuna_OLTP: é a chave que interliga a dimensão Coluna com a tabela Coluna da base de dados origem.

VI. Edicao

Nas dimensões de edição contem informações do jornal, que no *site* é um resumo do jornal impresso, onde possue as matérias com maiores destaques, ou seja, é a união de todos os componentes do jornal impresso que compõe o *site*, essas edições são publicações feitas diariamente, todos os conteúdos do *site* estão interligados a uma edição.

- Codigo: Código da edição.
- Data_Edicao: Data da edição.
- Manchete: A manchete da Edição.
- CodEdicao_OLTP: é a chave que interliga a dimensão Edicao com a tabela Edição da base de dados origem.

6.4.3 Modelo Estrela

No modelo Estrela existe uma tabela dominante no centro, chamada tabela de fatos, com múltiplas junções conectando a outras tabelas, sendo estas chamadas de tabelas de dimensão. Cada uma das tabelas secundárias possui apenas uma junção com a tabela central. O modelo Estrela tem a vantagem de ser simples e intuitivo, mas também faz uso de novos enfoques de indexação e união de tabelas. (KIMBAL; ROSS, 2002)

A tabela de fatos contém valores e medidas do negócio do *site*, como período de permanência no *site* e quantidade de acesso ao *site*. Cada uma destas medidas é tomada segundo a interseção de todas as dimensões. (KIMBAL; ROSS, 2002)

As tabelas de dimensão armazenam as descrições textuais das dimensões do *site*. Cada uma dessas descrições textuais ajuda a definir um componente da respectiva dimensão. Uma das principais funções dos atributos das tabelas de dimensão é servir como fonte para restrições em uma consulta ou como cabeçalhos de linha no conjunto de resposta do usuário. (KIMBAL; ROSS, 2002)

Tabelas dimensões tendem a utilizar tipos caracteres ao invés de numéricos, de forma que suas linhas são muito mais longas, mas em pouca quantidade ocupando uma pequena percentagem de espaço em disco. (KIMBAL; ROSS, 2002):

Segundo Kimball, o modelo dimensional apresenta várias vantagens no que diz respeito a sua utilização para DW, dentre estas estão:

- O modelo Estrela tem uma arquitetura padrão e previsível. As ferramentas de consulta e interfaces do usuário podem se valer disso para fazer suas interfaces mais amigáveis e fazer um processamento mais eficiente;

Todas as dimensões do modelo são equivalentes, ou seja, podem ser vistas como pontos de entrada simétricos para a tabela de fatos. As interfaces do usuário são simétricas, as estratégias de consulta são simétricas, e o SQL gerado, baseado no modelo, é simétrico.

6.5 Solução Business

O objetivo deste sistema é manter informações atualizadas sobre usuário do *site* jornal, as publicações de matérias, colunas e promoções realizadas na página da *internet* do jornal. As informações serão disponibilizadas aos gestores de forma a subsidiá-los em suas decisões, através de consultas e relatórios que estes irão realizar de acordo com a necessidade.

6.6 Modelo Físico dos Dados

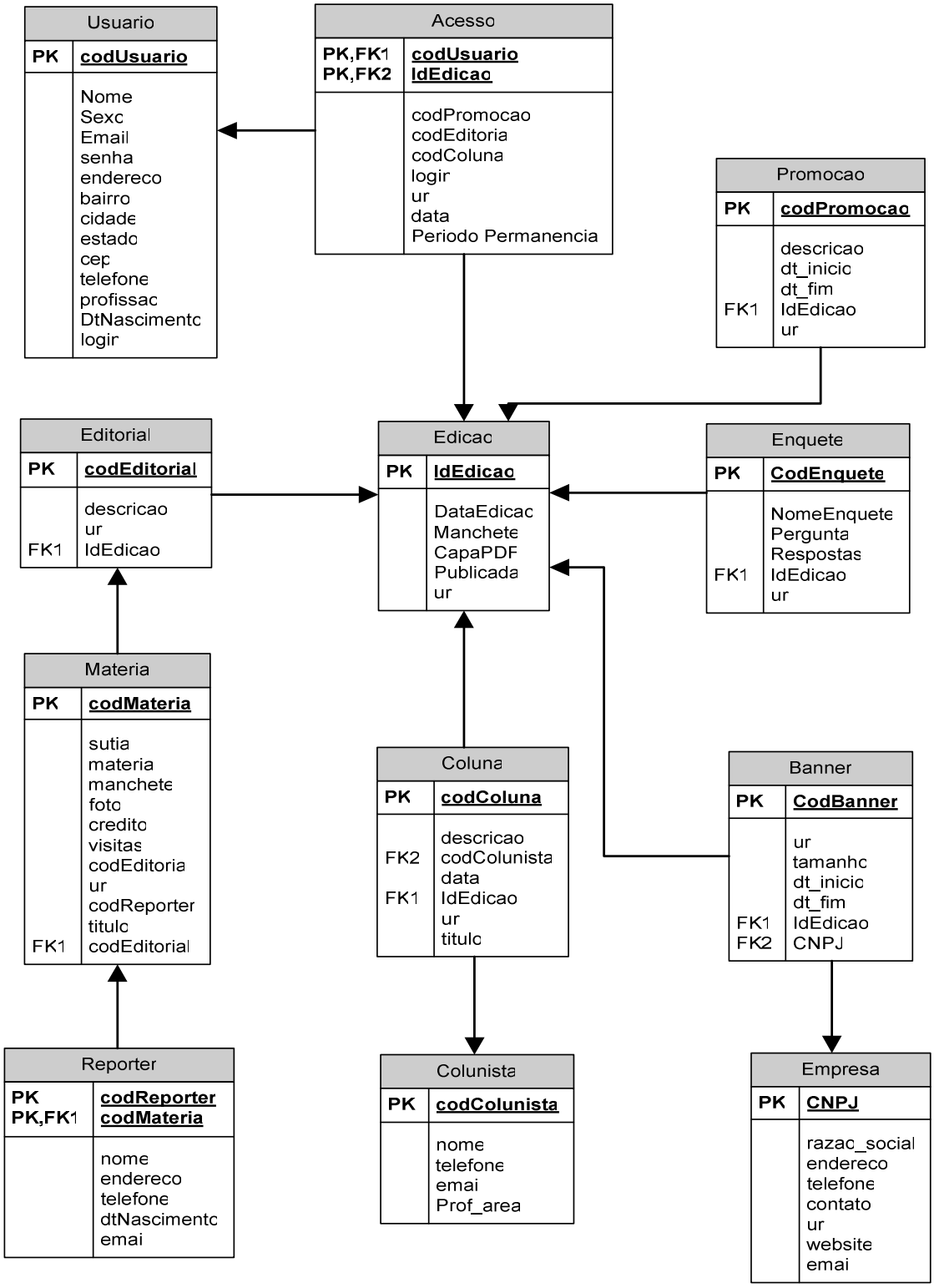


Figura 4: Modelo físico do Banco de Dados

6.7 Tipos de Análise dos dados

Partindo dos primórdios da informatização, quando um sistema que gerava relatórios era a principal fonte de dados residentes na empresa, toda vez que fosse necessário realizar uma análise, era necessário produzir novos relatórios. Estes relatórios tinham que ser produzidos pela área de informática e normalmente necessitava de muito tempo para ficar pronto. Apresentando alguns problemas como: relatórios estáticos e o acúmulo de diferentes tipos de relatórios num sistema geravam um problema de manutenção. (António, 2004)

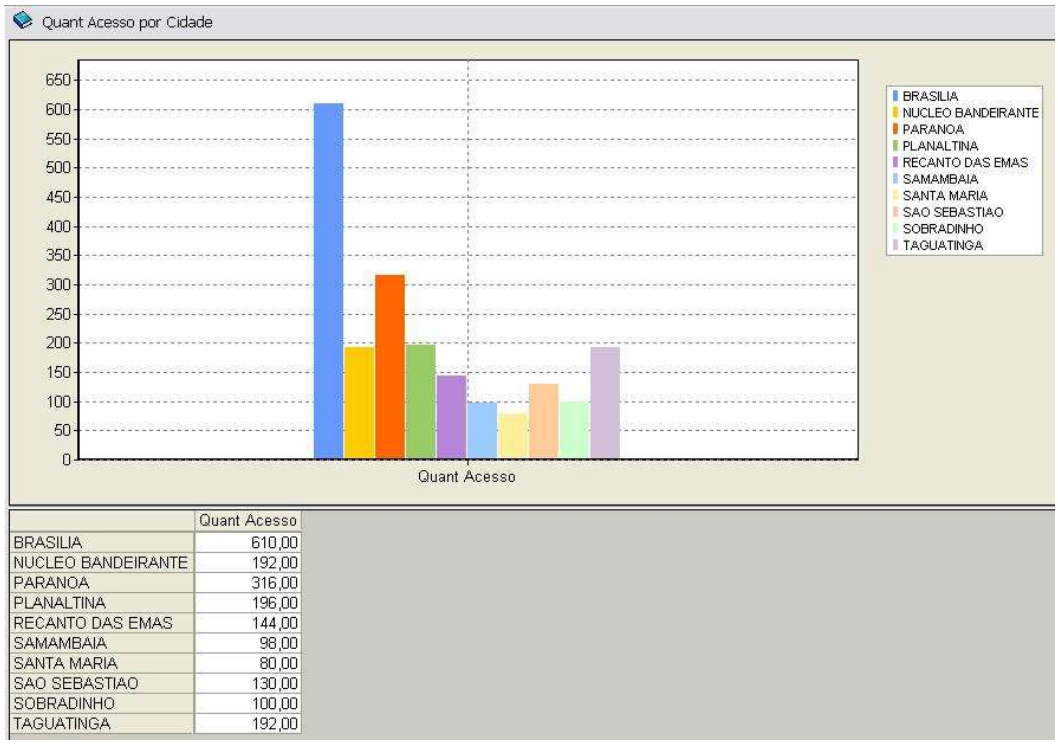
Para que a análise dos dados seja feita de forma *on-line* e dinamicamente, ou seja, permitir que os usuários finais montem seus próprios relatórios, será utilizada uma ferramenta OLAP que será responsável por analisar os dados corporativos de forma rápida, agilizando os processos de decisões permitindo uma nova visão da realidade da empresa. Essa tecnologia pode ser tanto cliente-servidor quanto *web* oferecendo analises históricas do perfil dos clientes, visões personalizadas.

Analisar os dados representa uma forma de classificar os dados do DM, por exemplo contar o número de clientes distribuídos pela dimensão sexo, ou profissão, ou região em relatórios criados pelos próprios usuários finais do sistema.

6.8 Exemplificações

O Data Mart utilizará uma ferramenta para consultas OLAP, onde com base nas necessidades levantadas e nas modelagens multidimensionais feitas anteriormente, pode-se efetuar várias combinações de pesquisas para emissão de relatórios na base de dados do DM.

- Quantidade de acessos por cidade (Figura 5), por mês (Figura 6), por profissão (Figura 7), por idade dos usuários *site* (Figura 8).



	Quant Acesso
BRASILIA	610,00
NUCLEO BANDEIRANTE	192,00
PARANOA	316,00
PLANALTINA	196,00
RECANTO DAS EMAS	144,00
SAMAMBAIA	98,00
SANTA MARIA	80,00
SAO SEBASTIAO	130,00
SOBRADINHO	100,00
TAGUATINGA	192,00

Figura 5 – Quantidade de acessos por cidade.

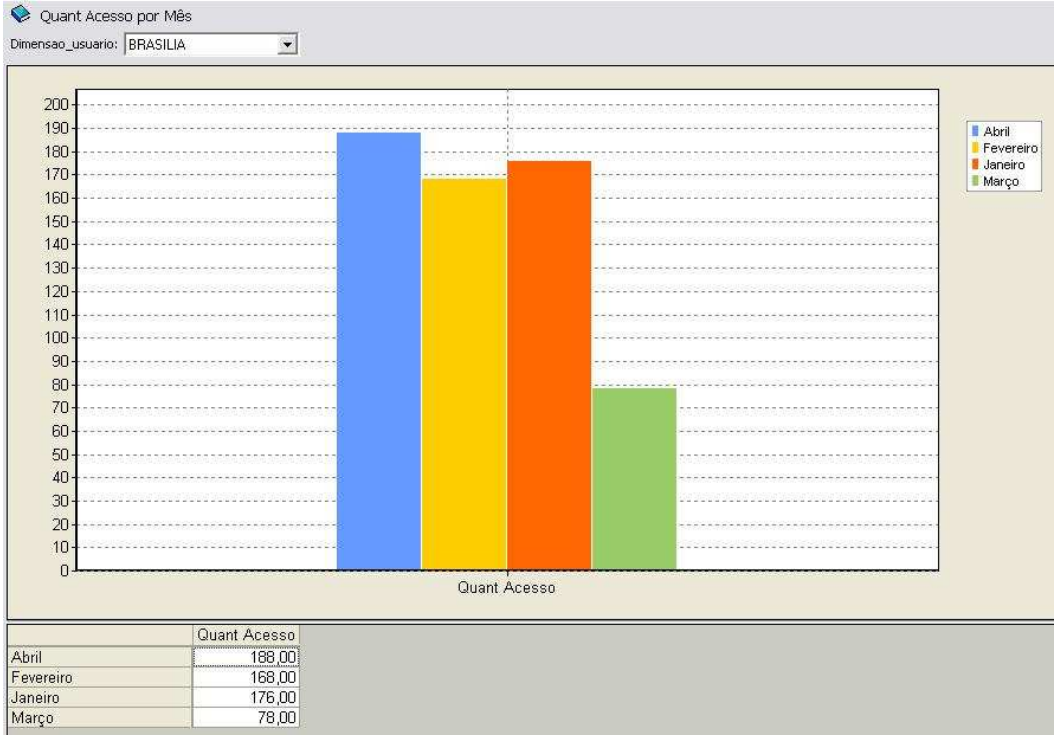


Figura 6 – Quantidade de acessos por mês, na cidade de Brasília.

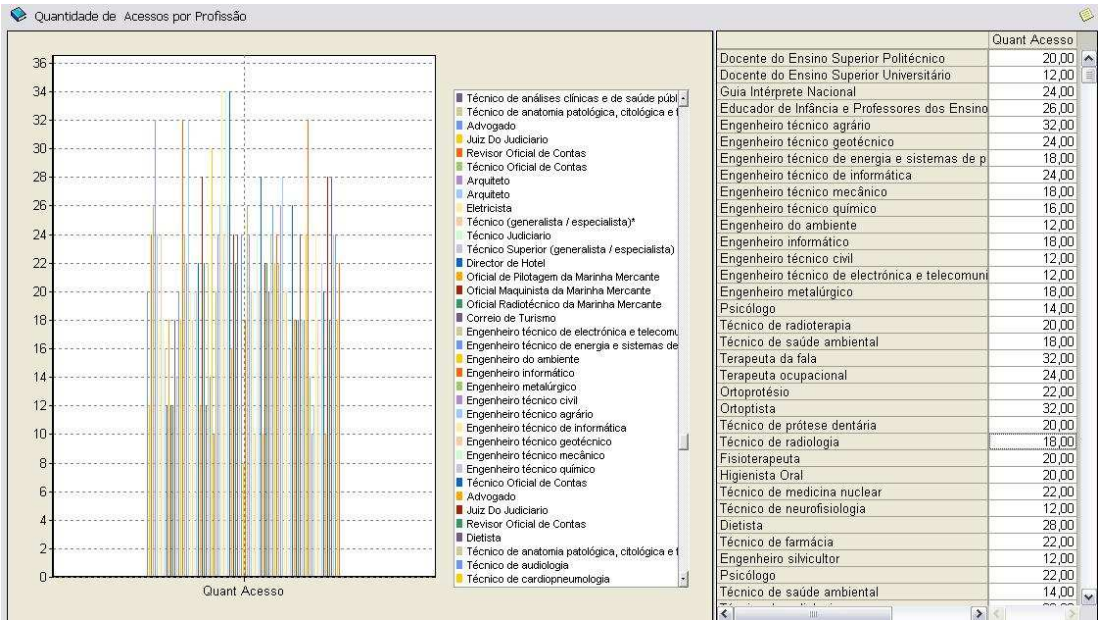
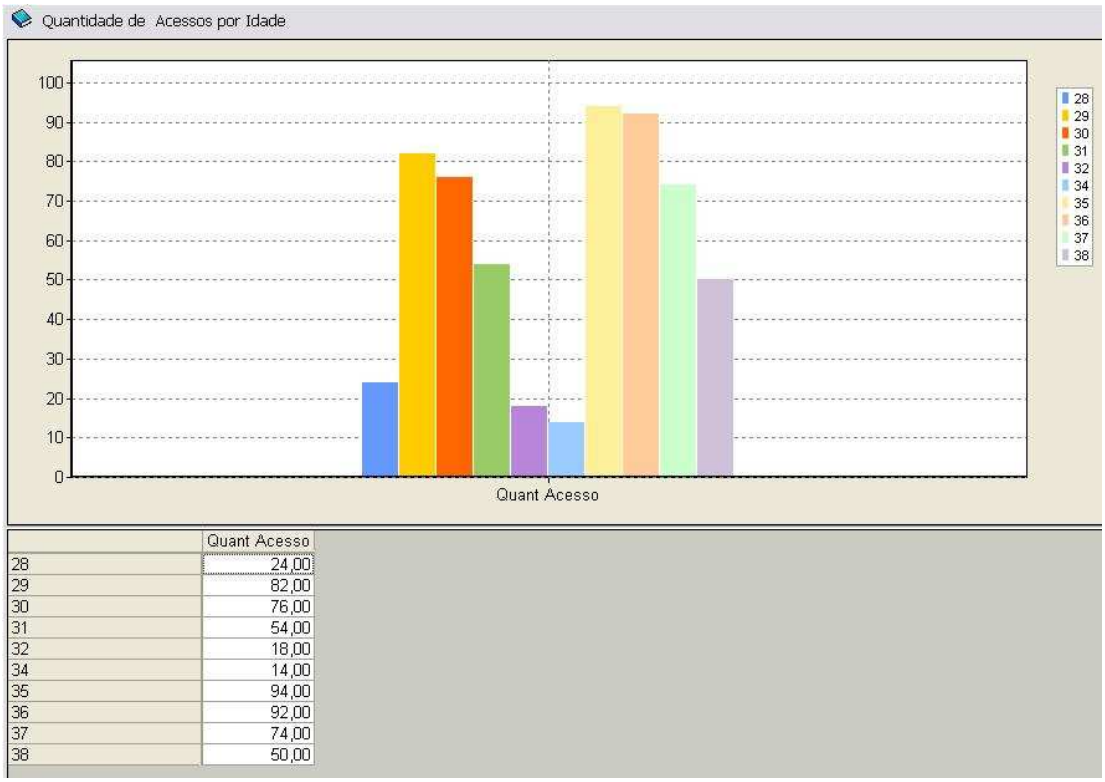


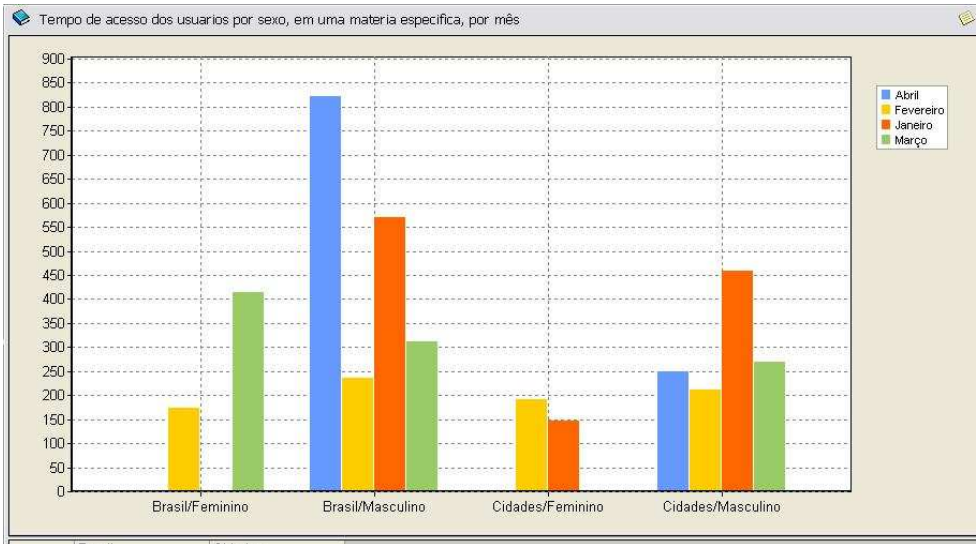
Figura 7 – Quantidade de acessos por profissão.



	Quant Acesso
28	24,00
29	82,00
30	76,00
31	54,00
32	18,00
34	14,00
35	94,00
36	92,00
37	74,00
38	50,00

Figura 8 – Quantidade de acessos por idade.

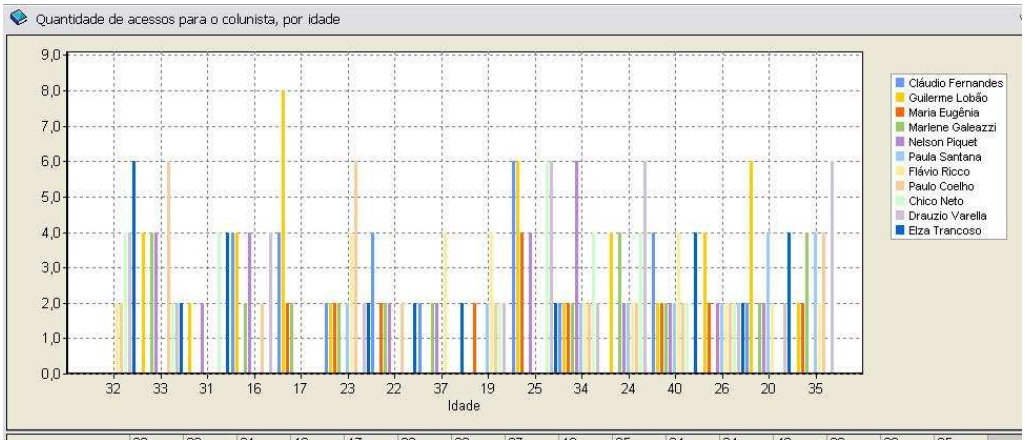
- Por quanto tempo os usuários do sexo masculino permaneceram no *site*, ou em uma matéria especifica, por mês (Figura 9).



	Brasil		Cidades	
	Feminino	Masculino	Feminino	Masculino
Abril		822,00		250,00
Fevereiro	174,00	236,00	192,00	212,00
Janeiro		570,00	146,00	458,00
Março	414,00	312,00		270,00

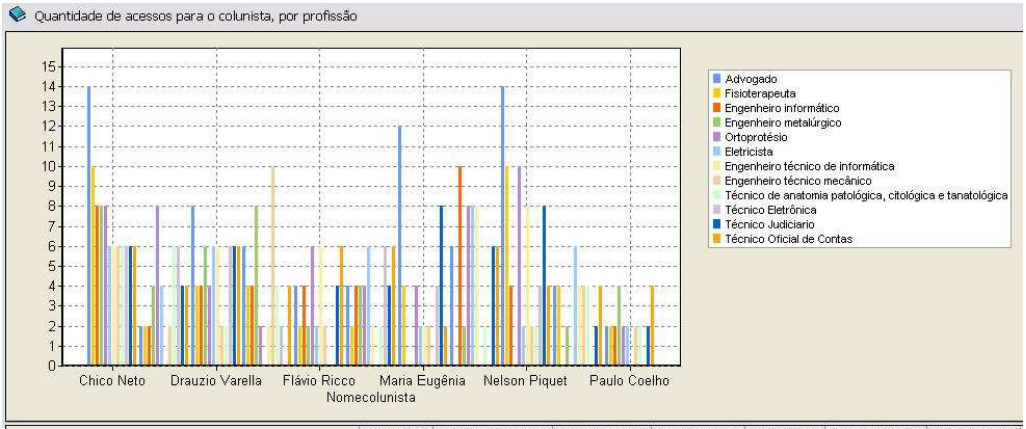
Figura 9 – Tempo de acesso dos usuários por sexo, em uma matéria especifica, por mês.

- Qual colunista teve mais acesso, por idade (Figura 10), por profissão (Figura 11), por cidades (Figura 12) e por sexo (Figura 13).



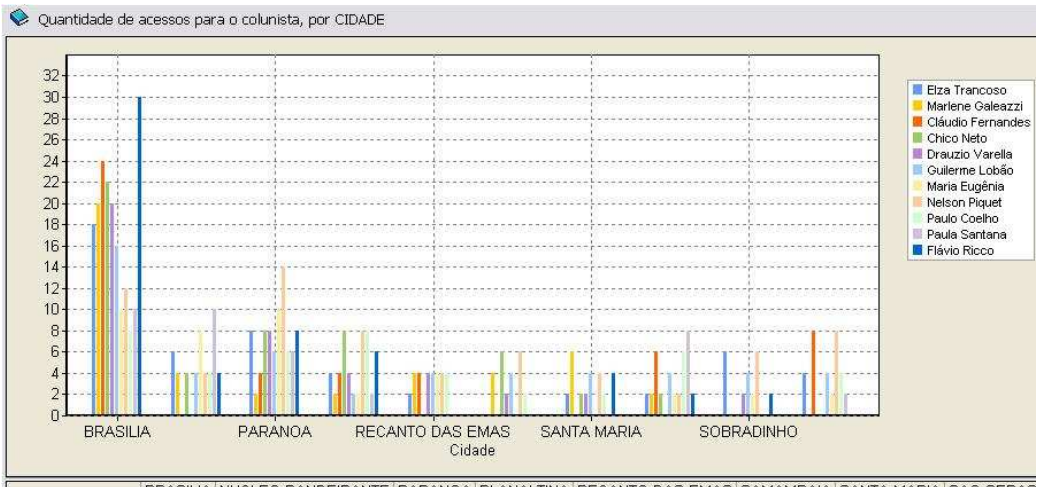
	32	33	31	16	17	23	22	37	19	25	34	24	40	26	20	35
Cláudio Fernandes				4,00	4,00	2,00	4,00	2,00		6,00	2,00		4,00		2,00	
Guilerme Lobão		4,00	2,00	4,00	8,00	2,00				6,00	2,00	4,00	2,00	4,00	6,00	2,00
Maria Eugênia					2,00	2,00	2,00		2,00	4,00	2,00		2,00	2,00		2,00
Marlene Galeazzi		4,00		2,00	2,00	2,00	2,00	2,00			2,00	4,00	2,00		2,00	4,00
Nelson Piquet		4,00	2,00	4,00			2,00	2,00		4,00	6,00	2,00	2,00	2,00	2,00	
Paula Santana						2,00			2,00		2,00	2,00	2,00	2,00	4,00	4,00
Flávio Ricco	2,00					4,00		4,00	4,00		2,00	2,00	4,00	2,00	2,00	2,00
Paulo Coelho	2,00	6,00		2,00		6,00	2,00		2,00		2,00	2,00	2,00	2,00		4,00
Chico Neto	4,00	2,00	4,00						2,00	6,00	4,00	4,00	2,00	2,00		
Drauzio Varella	4,00	2,00		4,00		2,00			2,00	6,00	2,00	6,00		2,00	2,00	6,00
Elza Trancoso	6,00	2,00	4,00			2,00	2,00	2,00		2,00			4,00	2,00	4,00	

Figura 10 – Quantidade de acesso para o colunista, por idade.



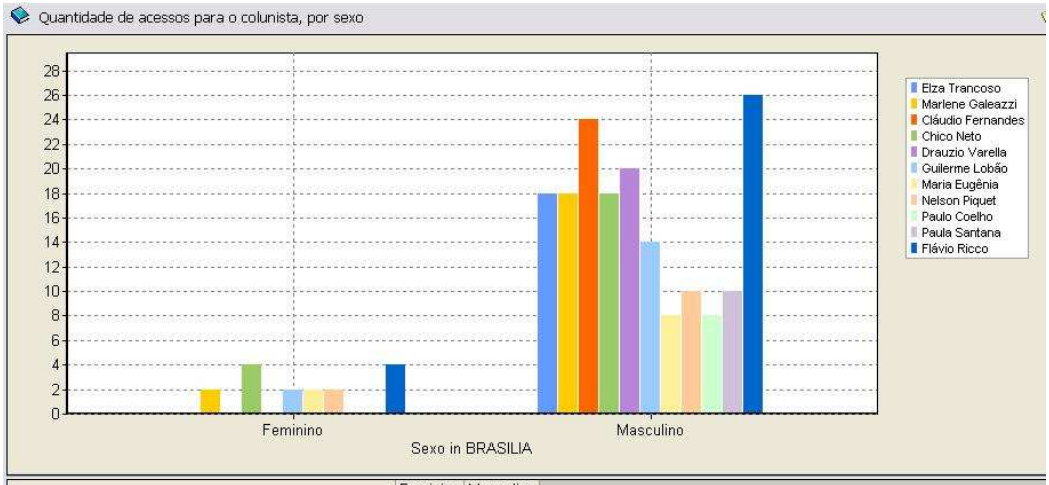
	Chico Neto	Cláudio Fernandes	Drauzio Varella	Elza Trancoso	Flávio Ricco	Guilerme Lobão	Maria Eugênia
Advogado	14,00	2,00	8,00	6,00	4,00	4,00	12,00
Fisioterapeuta	10,00	2,00	4,00	4,00	2,00	2,00	4,00
Engenheiro informático	8,00	2,00	4,00	4,00	4,00	4,00	
Engenheiro metalúrgico	8,00	4,00	6,00	8,00	2,00	4,00	
Ortoprotésio	8,00	8,00	4,00	2,00	6,00	4,00	4,00
Eletricista	6,00	4,00	6,00		2,00	6,00	2,00
Engenheiro técnico de informática	6,00		6,00	2,00	6,00	2,00	2,00
Engenheiro técnico mecânico	6,00	2,00	2,00	10,00	2,00		2,00
Técnico de anatomia patológica, citológica e tanatológica	6,00	6,00	2,00	4,00		2,00	
Técnico Eletrônica	6,00	6,00	6,00	2,00		6,00	4,00
Técnico Judiciario	6,00	4,00	6,00		4,00	4,00	8,00

Figura 11 – Quantidade de acesso para o colunista, por profissão.



	BRASILIA	NUCLEO BANDEIRANTE	PARANOA	PLANALTINA	RECANTO DAS EMAS	SAMAMBAIA	SANTA MARIA	SAO SEBAS
Elza Trancoso	18,00	6,00	8,00	4,00	2,00		2,00	
Marlene Galeazzi	20,00	4,00	2,00	2,00	4,00	4,00	6,00	
Cláudio Fernandes	24,00		4,00	4,00	4,00			
Chico Neto	22,00	4,00	8,00	8,00		6,00	2,00	
Drauzio Varella	20,00		8,00	4,00	4,00	2,00	2,00	
Guilerme Lobão	16,00	4,00	6,00	2,00	4,00	4,00	4,00	
Maria Eugênia	10,00	8,00	10,00	2,00	4,00			
Nelson Piquet	12,00	4,00	14,00	8,00	4,00	6,00	4,00	
Paulo Coelho	8,00	4,00	6,00	8,00	4,00	2,00	2,00	
Paula Santana	10,00	10,00	6,00	2,00				

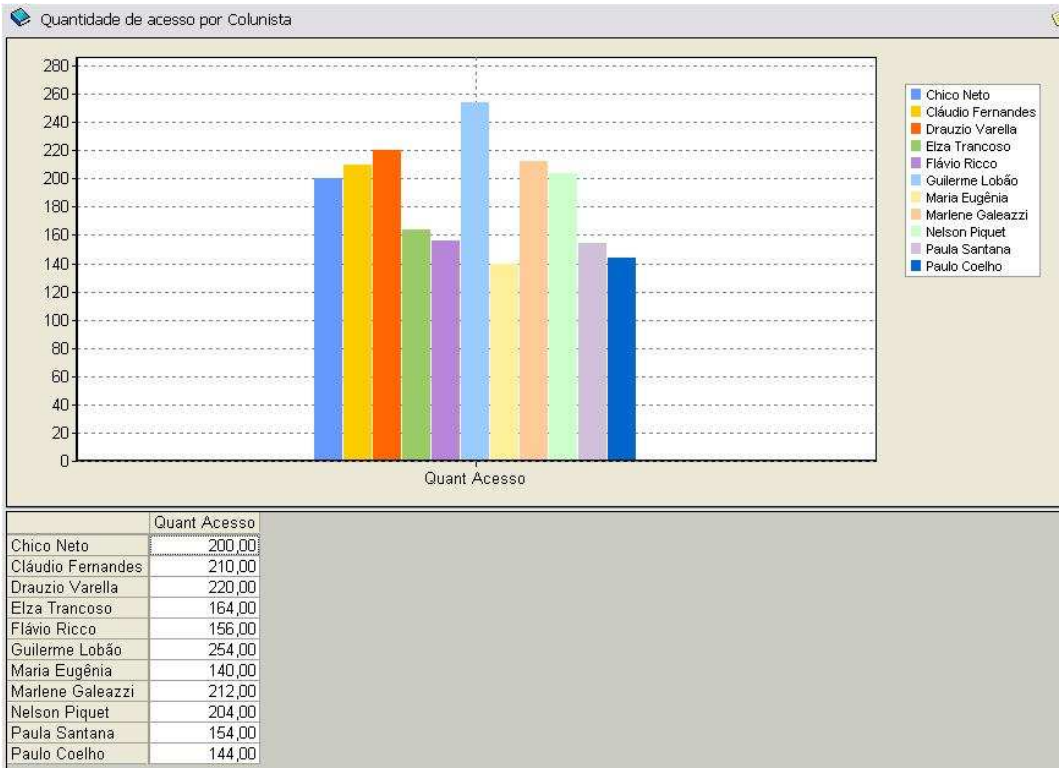
Figura 12 – Quantidade de acesso para o colunista, por Cidade.



	Feminino	Masculino
Elza Trancoso		18,00
Marlene Galeazzi	2,00	18,00
Cláudio Fernandes		24,00
Chico Neto	4,00	18,00
Drauzio Varella		20,00
Guilerme Lobão	2,00	14,00
Maria Eugênia	2,00	8,00
Nelson Piquet	2,00	10,00
Paulo Coelho		8,00
Paula Santana		10,00
Flávio Ricco	4,00	26,00

Figura 13 – Quantidade de acesso para o colunista, por sexo.

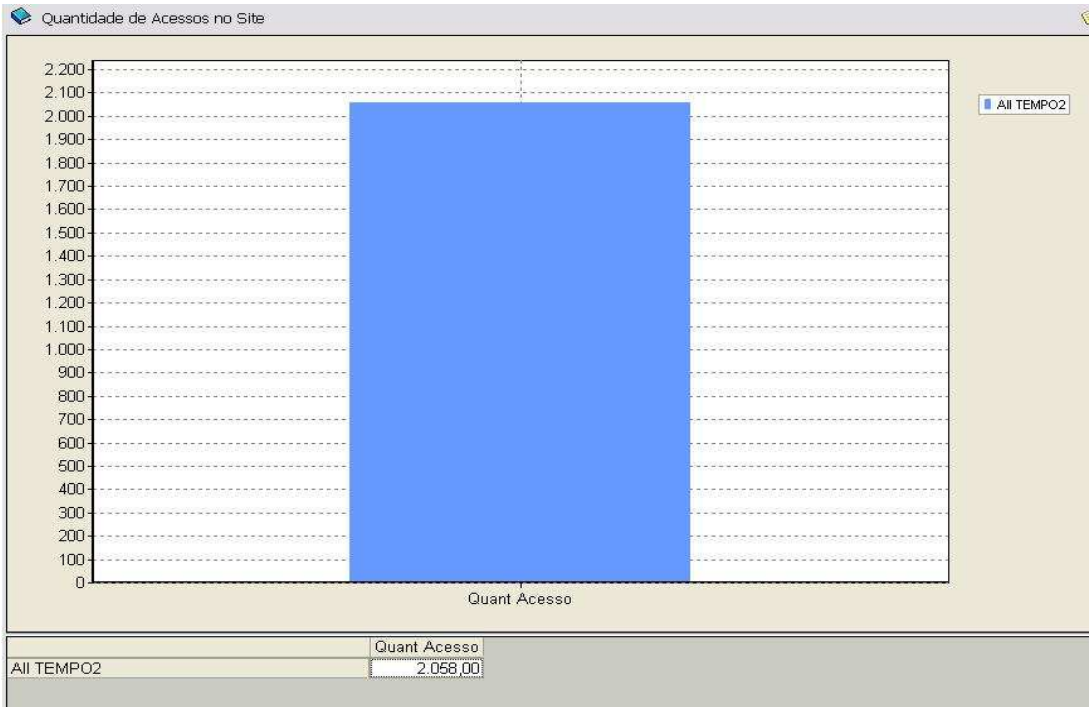
- Quantidade de acessos por colunista (Figura 14).



	Quant Acesso
Chico Neto	200,00
Cláudio Fernandes	210,00
Drauzio Varella	220,00
Elza Trancoso	164,00
Flávio Ricco	156,00
Guilerme Lobão	254,00
Maria Eugênia	140,00
Marlene Galeazzi	212,00
Nelson Piquet	204,00
Paula Santana	154,00
Paulo Coelho	144,00

Figura 14 – Quantidade de acesso por o colunista.

- Quantidade de acessos do *site* (Figura 15) e tempo de acesso total no *site* (Figura 16).



	Quant Acesso
All TEMPO2	2.058,00

Figura 15 – Quantidade de acesso no *site*.

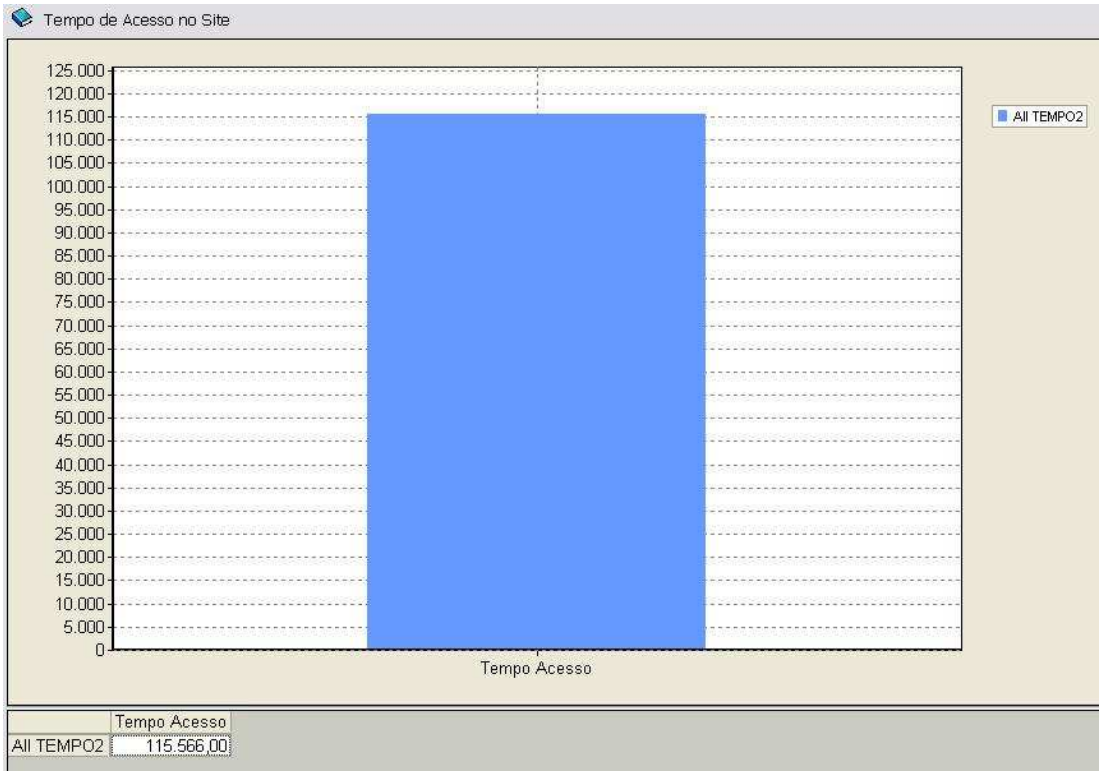


Figura 16 – Tempo de acesso no *site*.

- Duração de acessos da matéria (Figura 17), nas colunas (Figura 18), nas promoções (Figura 19) e nas edições (Figura 20).

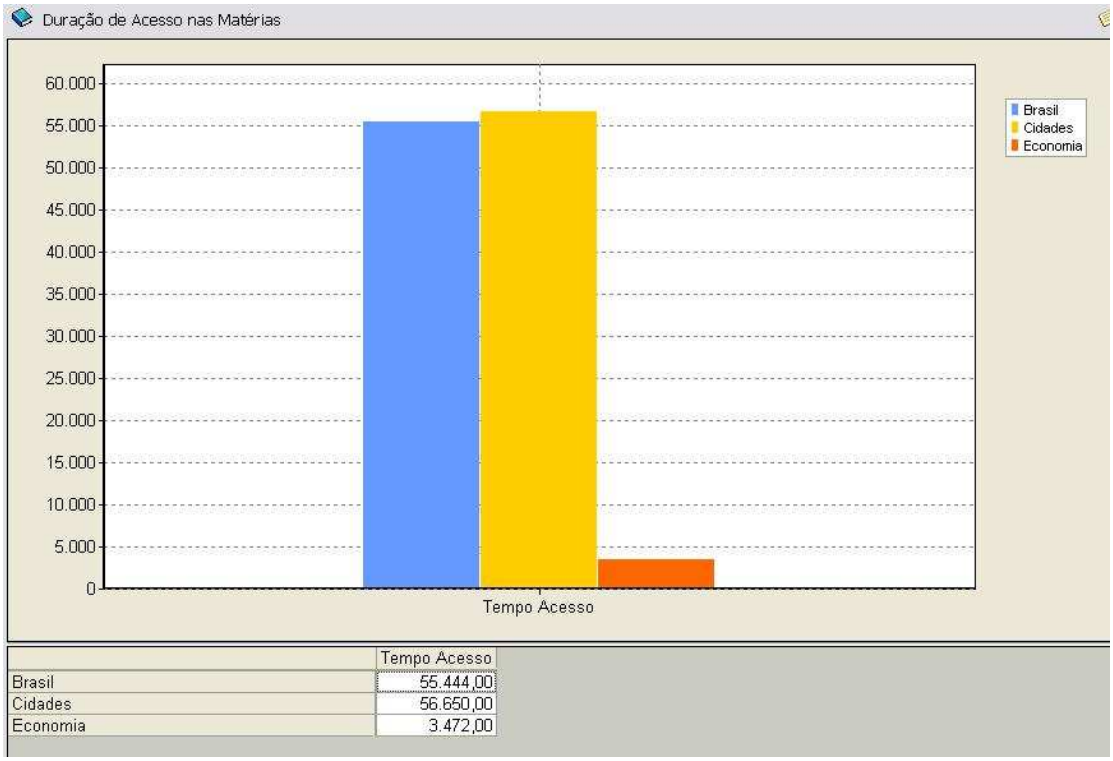


Figura 17 – Duração de acesso nas Matérias.

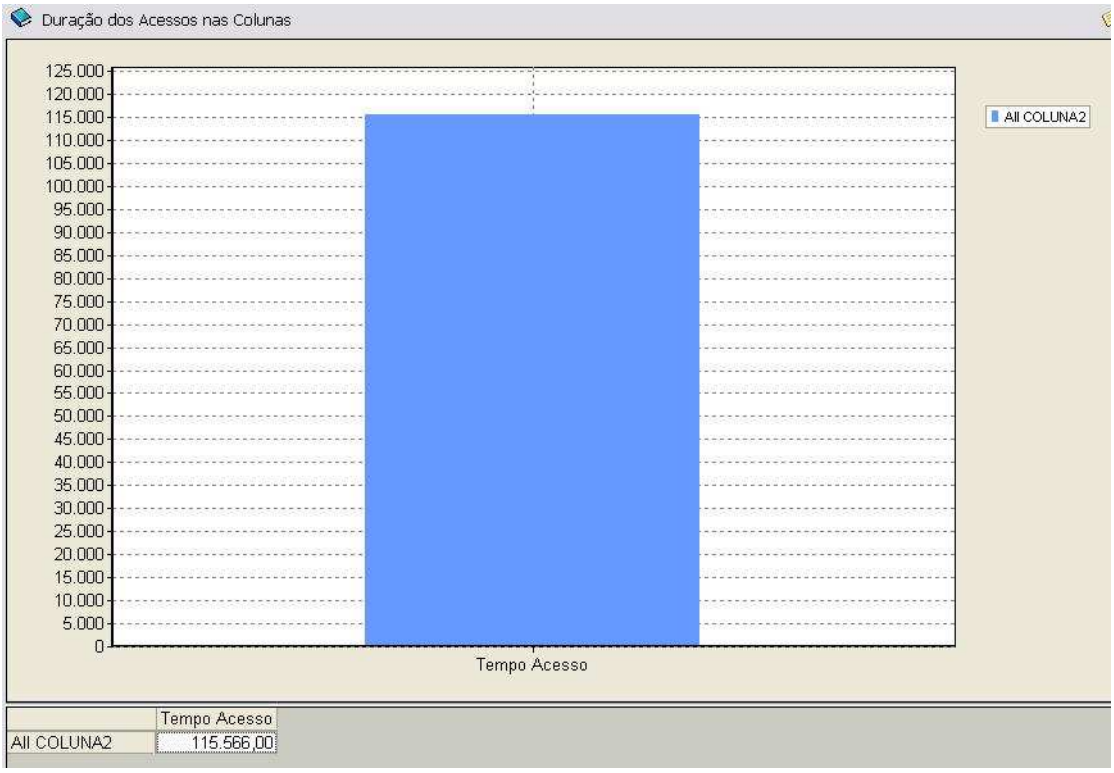


Figura 18 – Duração de acesso nas Colunas.

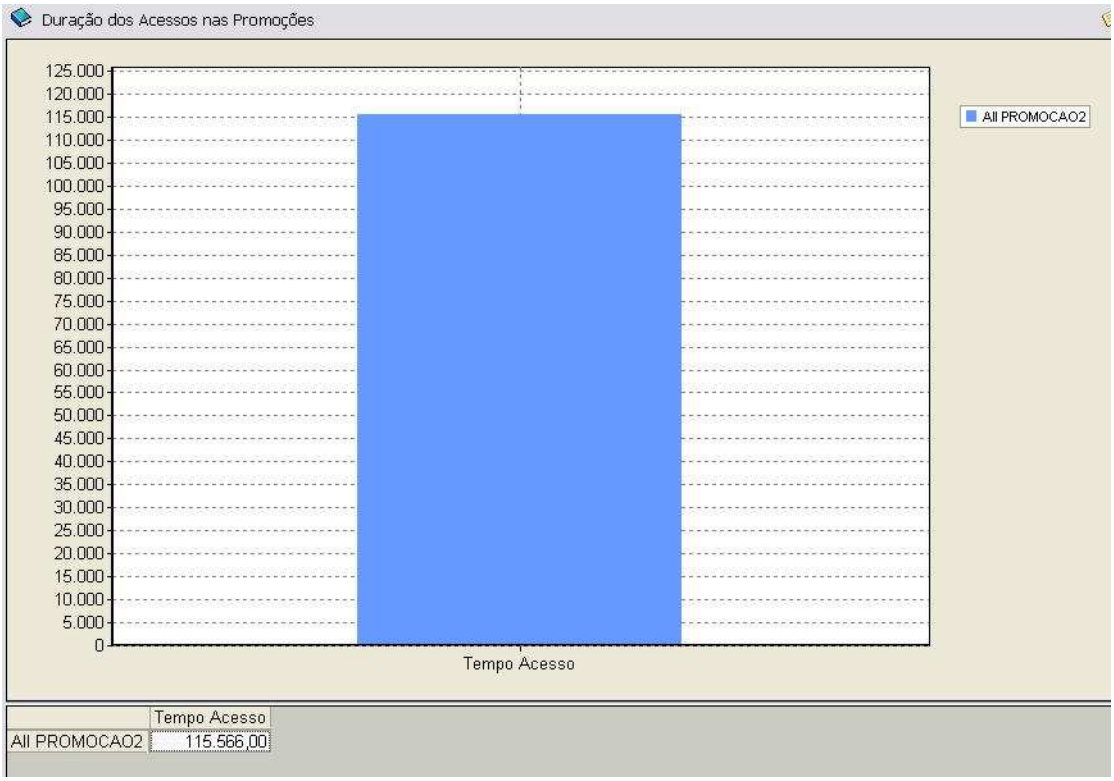


Figura 19 – Duração de acesso nas Promoções.

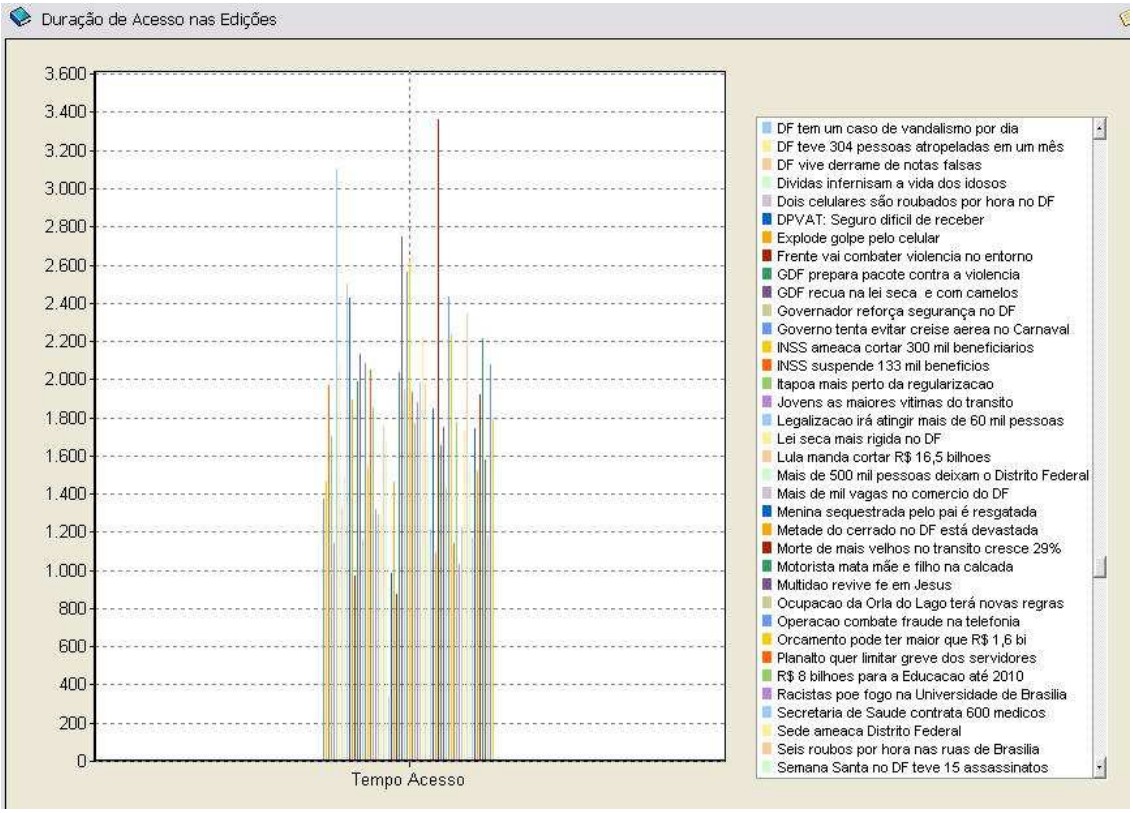


Figura 20 – Duração de acesso nas Edições.

- Quantidade de acessos da matéria (Figura 21).

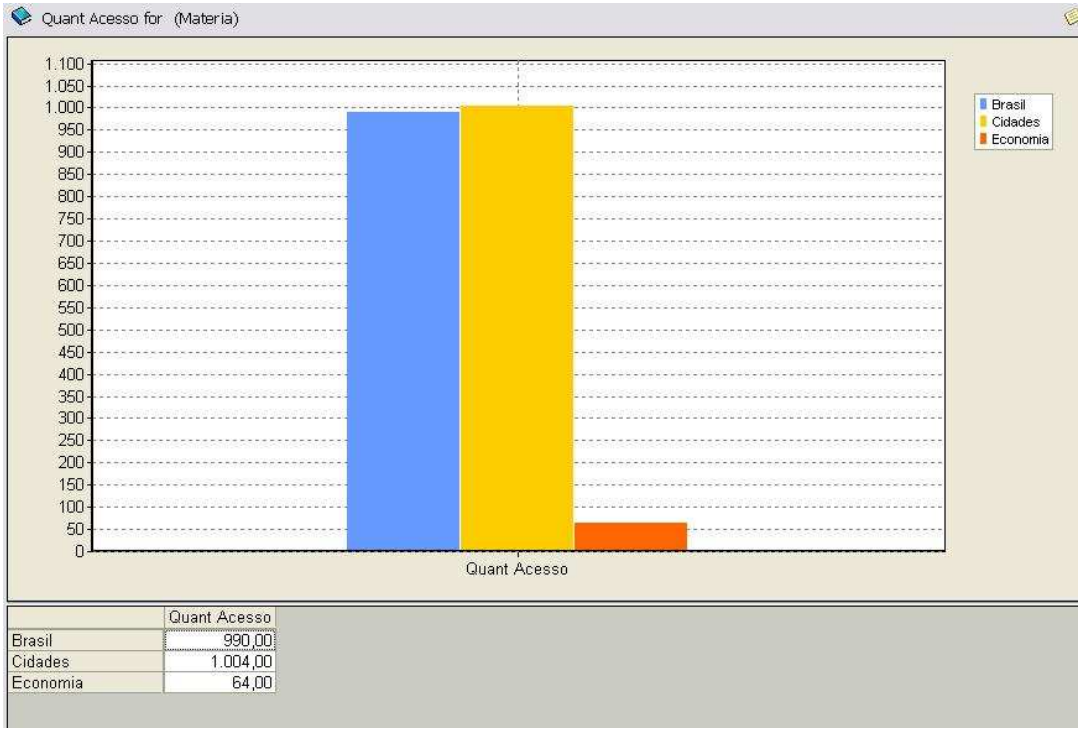


Figura 21 – Quantidade de acessos nas Matérias.

- Quantidade de acessos por Edição (Figura 22).

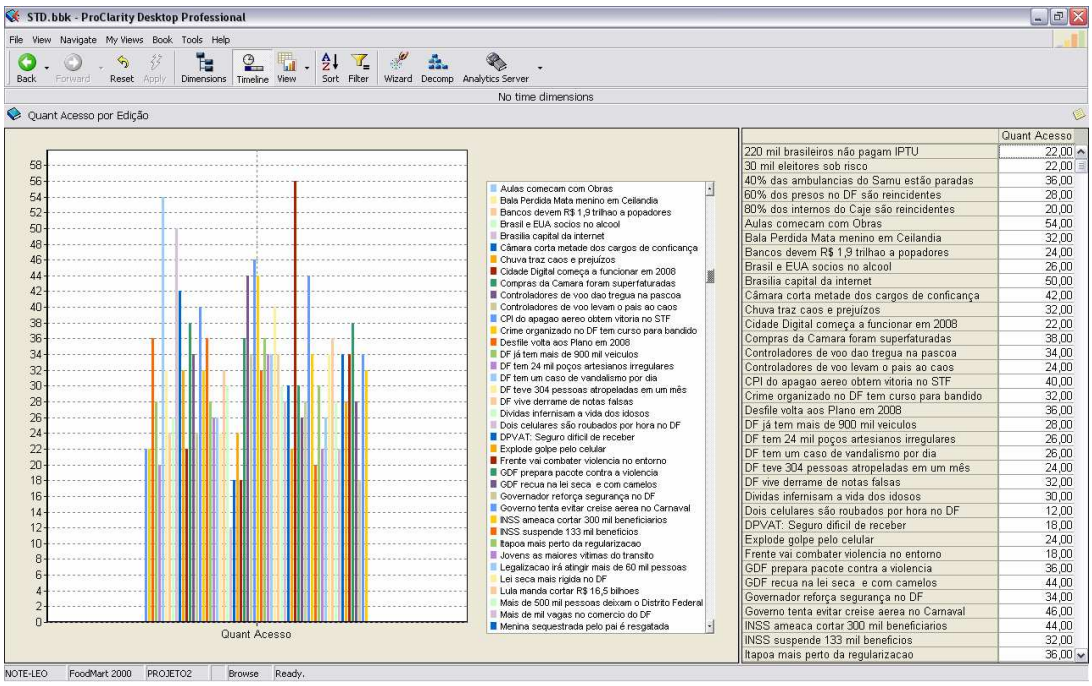


Figura 22 – Quantidade de acessos por edição.

CAPITULO 7 – Conclusão

Dados de uma empresa podem ser transformados em informação para auxiliar os gerentes em decisões importantes para contribuir com crescimento da empresa, tornando uma empresa competitiva.

Na empresa de jornais, os dados foram transformados em informações através do *Data Mart*, isto possibilitou aos tomadores de decisões a direcionarem seu produto de acordo com as preferências do cliente e a conquistarem maior quantidade de usuários, com transformações realizadas tendo como base relatórios e consultas gerenciais disponibilizadas no sistema.

Foi desenvolvida uma modelagem multidimensional, pode-se verificar no capitulo 6 item 6.4 Descrição do Modelo de Dados, onde foi definida as fatos e dimensões de acordo com o negócio da empresa, o que possibilitou a geração de relatórios utilizando a teoria do cubo, para realização de consultas dinâmicas com facilidade, agilidade e com dados históricos, possibilitando que o próprio usuário realize essas consultas e gere relatórios de acordo com a realidade que necessita.

Torna-se importante que as empresas utilizem informações concretas para ter seu diferencial competitivo, ter algo que chame atenção do seu cliente e que torne sua empresa mais lucrativa, sempre pensando em agradar seus clientes e conquistá-los para que a corporação possa competir em alto grau com seus concorrentes.

Observa-se que as empresas às vezes não dão valor às informações obtidas de seus *sites*, e não percebem que estas podem auxiliá-los para análises de negócios. Nesta empresa em questão as informações armazenadas em banco de dados não era utilizada, e visto que seus usuários do *site* são maiores em quantidade que seus assinantes do jornal e, havia uma necessidade de identificar quais promoções foram bem sucedidas, quais são as faixas etárias dos leitores de uma coluna ou de uma matéria, dentre outras questões. Este projeto supre não

só essas necessidades como outras, exemplos no capitulo 6 item 6.8 Exemplificações, permitindo uma visão diferenciada de seu negócio, através dos gráficos e relatórios gerenciais.

O projeto contempla que as informações do *site*, seja avaliada, transformada e armazenada no *Data Mart* através das melhores práticas de ETL, e através de uma ferramenta OLAP, pesquisada no mercado, pode-se realizar o cruzamento dessas informações para geração de relatórios e consultas de acordo com as necessidades da empresa.

O mundo competitivo exige que as empresas pensem no futuro com referência no passado, para que suas metas sejam atingidas e para evitar trabalhar com tentativas evitando erro, mas sim trabalhando com informações concretas que o auxiliem a realizarem atitudes, traçarem metas e ter total sucesso nos seus objetivos.

Referências Bibliográficas:

(BARBIERI, 2001): BARBIERI, Carlos. *BI – Bussiness Intelligence: Modelagem ; Tecnologia*. Rio de Janeiro. Axcel Books Brasil, 2001.

(SILBERSCHATZ, KORT, SUDARSHAN,1999): SILBERSCHATZ, Abraham ; KORTH, Henry F. ; SUDARSHAN, S.*Sistema de Banco de Dados*. São Paulo: PearsonMakron Books, 1999. pág 712-714.

(INMON, 1997): INMON, William H. *Como construir o Data Warehouse tradução de Ana Maria Netto Guz.*Rio de Janeiro: Editora Campos, 1997.

(SINGH,2001): SINGH, Harry S.*Data Warehouse conceitos, tecnologias, implementação e gerenciamento*.São Paulo: Makron Books, 2001.

(DAL'ALBA, 2006): DAL'ALBA ,Adriano. *Características do Data Warehouse*. Disponível em: <http://www.geocities.com/SiliconValley/ Port/5072/ Introducao.htm> . Acessado em: 18 ago. 2006.

(KIMBALL; CASSERTA, 2004): KIMBALL, Ralph ; CASSERTA, Joe. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning,Conforming, and Delivering Data.* Wiley. Canadá, 2004.

(KIMBAL; ROSS, 2002): KIMBALL, Ralph; ROSS, Margy. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley. Canadá: 2002.

(CIELO1 ,2006) : CIELO, Ivan. *ETL - Extração, Transformação e Carga de Dados*. <http://www.datawarehouse.inf.br/artigos/etl.asp>. Acessado em: 14 set. 2006.

(CIELO2,2006): CIELO, Ivan. *OLAP*. Disponível em: <http://www.datawarehouse.inf.br/artigos/olap.asp>. Acessado em: 14 set. 2006.

(CIELO3,2006) : CIELO, Ivan. *Data Warehouse como diferencial competitivo*. Disponível em: <http://www.datawarehouse.inf.br/>. Acessado em: 1 set. 2006.

(CIELO4, 2006): CIELO, Ivan. *Construindo um sistema de apoio à decisão*. Disponível em: <http://www.datawarehouse.inf.br/>. Acessado em: 1 set. 2006.

(PEREIRA, 2000): PEREIRA, Walter Adel Leite. *Uma metodologia de inserção de tecnologia de Data Warehouse em Organizações*. Disponível em: <http://www.pucrs.br/uni/poa/info/pos/dissertacoes/arquivos/walter.pdf>. Rio Grande do Sul, 2000. Acessado em: 28 set. 2006.

(SASSI,2006): SASSI , Renato José. *Análise das Modalidade de um Data Warehouse (DW)*. Disponível em: <http://www.cefetsp.br/edu/sinergia/6p9c.html>. Acessado em 28 set. 2006.

(SOUZA1, 2003): SOUZA, Michel. *Ferramentas OLAP*. Disponível em: <http://www.imasters.com.br/artigo/1498/gerencia/ferramentas_olap/>. Acessado em: 14 set. 2006.

(SOUZA2, 2003): SOUZA, Michel de. *Metadados*. Disponível em: <http://www.imasters.com.br/artigo/1569/bi/metadados/> . 28 de novembro de 2003. Acessado em: 02 out. 2006.

(VAZ; MARCON, 2004): VAZ ,Gomes; MARCON , Maria Salete. *Metadados*. Disponível em: <http://www.sqlmagazine.com.br/Colunistas/MariaSalete/02_Metadados.asp>. Acessado em: 06 out. 2006.

(ANDRADE; SANTIAGO, 2003) : ANDRADE, Fábio A. ; SANTIAGO, Luciano D. *Introdução aos conceitos de modelagem multidimensional* . 2003. Disponível em: <http://www.cientefico.frb.br/TextosCienteFico.2/INFO/>. Acesso em: 20 out. 2006.

(RIANE,2002): RIANE, Alexandre Paulo da Costa. *A questão do uso de banco de dados Relacional e Multidimensional em consultas em Data Warehouses.* Monografia de Graduação, 2002.

(MOREIRA, 2006): MOREIRA, Eduardo. *Modelo Dimensional para Data Warehouse*. Disponível em: <http://www.imasters.com.br/artigo/3836/bi/modelo_dimensional_para_data_warehouse/>. Acessado em: 6 out. 2006.