

Data webhouse: A Evolução do Data Warehouse para Web e suas Contribuições para o Aperfeiçoamento do Relacionamento com Clientes

Lucyanno Moreira C. de Holanda¹

lucyanno@gmail.com

Isabela Assis Guedes¹

isabelarosas@gmail.com

Ricardo Campos da Rocha¹

ricardo_campos@ciecee.org.br

David Moises Barreto dos Santos²

davidmbs@gmail.com

1- Universidade Federal de Campina Grande (UFCG), - Campina Grande, PB, Brasil.

GEGIT – Grupo de Estudos em gestão, Inovação e Tecnologia – UFCG, Campina Grande – PB, Brasil

2- Universidade Federal de Feira de Santana (UEFS), Feira de Santana, Ba, Brasil

RESUMO

O atual ambiente de negócios e de gestão tem como uma de suas principais implicações a necessidade das organizações serem cada vez mais adaptáveis e flexíveis para conviver em contextos e contingências mutáveis. Para encontrar os mecanismos e soluções mais adequadas para os diversos tipos de problemas enfrentados, as organizações têm utilizado as novas tecnologias de informação (TI's). Este trabalho visa apresentar a ferramenta Data Webhouse, demonstrando suas características, viabilidades de uso e sua importância para o processo de tomada de decisão nas organizações, através da coleta de informações comportamentais dos usuários dos sites da web. Para consecução dos objetivos do trabalho será utilizada como ferramenta metodológica a documentação indireta do tipo pesquisa bibliográfica, visando demonstrar a importância e necessidade da utilização das novas T.I's, a evolução do Data Warehouse e a criação do Data Webhouse, e, conseqüentemente, o alcance de diferenciais competitivos por parte das empresas.

Palavras-Chave: Tecnologias da Informação e Comunicação (TIC); *Data webhouse*; *Data warehouse*.

1 Introdução

O atual ambiente de negócios e de gestão é caracterizado pela necessidade das organizações serem cada vez mais flexíveis e adaptáveis para poderem sobreviver a ambientes instáveis e de ávida concorrência. Para conseguir crescer e se desenvolver nesse ambiente as empresas têm utilizado grande quantidade de Tecnologias de Informação e de Comunicação (TIC) para auxiliar em todos os seus processos, nos seus produtos e principalmente em seus modelos de gestão.

Com isso, o fator determinante para o desenvolvimento de uma organização deixa de ser a questão do tamanho, de proximidade com os recursos necessários para a fabricação de um determinado produto ou a quantidade de capital que ela detém, passando a ter como o maior diferencial a informação e modo que ela é utilizada.

A *Web* é um meio de comunicação proveniente dos avanços da Tecnologia de Informação (TI), entretanto ainda não é bem aproveitada pelas organizações, apesar da rapidez com que foi

difundida e da facilidade com que as informações são disseminadas, bem como das oportunidades comerciais que são amplamente esparzidas. Deste modo, as organizações sentem a necessidade de se ajustarem às práticas e aos conceitos virtuais para que possam adentrar no comércio eletrônico.

O intenso uso da *Internet* criou um ambiente favorável para a expansão dos negócios em rede, e este ainda vem crescendo, trazendo consigo um aumento do comércio eletrônico que pode incluir qualquer empresa, dependendo apenas do modo que estas se inserem e investem nesse meio de comunicação digital.

Tendo em vista esse pressuposto, surgiu a necessidade de se ter uma ferramenta que seja capaz de gerenciar o excesso de dados empresariais gerados digitalmente através da *Web* de forma que os transforme em informação útil para a empresa. A tecnologia *Data Warehouse* (DW) visa solucionar esta problemática, uma vez que seu objetivo central é auxiliar processos decisórios de informações. A evolução do DW que trata dados oriundos da *Web* é denominada *Data Webhouse* (DWh).

Mais especificamente, um DWh acumula dados detalhados da seqüência de clique (comportamento) de usuários ao navegar no *site* da empresa. Através da utilização eficaz desses dados é possível traçar o perfil de cada usuário e personalizar o *website* de acordo as necessidades individuais de cada um. Conseqüentemente, isto implicará no desenvolvimento da empresa virtual proporcionando um aumento das vendas.

Diante deste contexto, este trabalho visa apresentar a ferramenta de Tecnologia de Informação (TI), *Data Webhouse*, demonstrando suas características, viabilidades de uso e sua importância para a tomada de decisão nas organizações, através da coleta de informações comportamentais dos usuários dos *sites* da *web*.

Para consecução dos objetivos do trabalho será utilizada como ferramenta metodológica a documentação indireta do tipo pesquisa bibliográfica, para demonstrar a importância e necessidade da utilização das novas T.I's; a evolução do *Data Warehouse* e a criação do *Data Webhouse*; e conseqüentemente o alcance de diferenciais competitivos por parte das empresas.

A partir destas considerações, além desta parte introdutória, o artigo está assim estruturado: No item dois é apresentado os conceitos de *Data Warehouse* e sua evolução para *Data Webhouse*. No item três discorre-se sobre o Relacionamento com os Clientes - CRM. No item quatro, trata-se da segurança no *Data Webhouse*. Item cinco faz-se as considerações finais.

2 Data warehouse (DW): Principais conceitos

A necessidade das organizações obterem informações úteis para o processo de tomada de decisão fez surgir a tecnologia *data warehouse* – DW. Para SILBERSCHATZ (1999), um DW é um repositório de informações coletadas em diversas fontes, armazenadas sob um esquema único, em um só local. Uma vez coletados, os dados são armazenados por um período longo, permitindo acesso a dados históricos.

Para MUSSI (2004), um *data warehouse* constitui-se de um conjunto de arquiteturas e/ou sistemas de informação orientados a assunto que existem em plataformas segregadas do ambiente transacional, manipulando grande volume de dados, principalmente históricos, e dão origem a consultas (*read-only*) invariavelmente não previsíveis, que tem por objetivo dar suporte a esses processos.

De acordo com INMON (1997), é uma coleção de dados orientada por assuntos, integrada, variante no tempo e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão.

Em sua essência, o DW apreende dois tipos de dados: a) primitivos ou operacionais, os quais são flexíveis e atualizados e correspondem às atividades funcionais e formam o sistema operacional; b) derivados ou gerenciais, que são dados resumidos ou calculados para atender às necessidades da empresa, não são flexíveis e passíveis de atualizações. Esse dois tipos de dados atendem às atividades gerenciais e formam, assim, os sistemas informacionais.

MUSSI (2004) caracteriza os DW da seguinte maneira: **orientado por assunto**, referindo-se aos assuntos de interesse da empresa; **integrado**, trazendo os dados originais dos sistemas operacionais para que estes tenham um sentido singular; **variante no tempo**, salientando que os dados sempre apresentam uma característica de tempo, ou seja, a cada mudança ocorrida num dado, uma nova entrada é criada e não atualizada; **não volátil**, destacando que os dados não recebem atualizações, eles são carregados uma única vez e, a partir desse momento, só podem ser consultados.

Apesar do excelente desempenho da aplicação do DW no mundo dos negócios, foi necessário que se criasse um sistema semelhante de análise de dados para informações oriundas da *Web*, sendo este hoje o grande referencial de busca de informações. Foi nesse ponto que surgiu o conceito de *Data WebHouse*, um DW para a *internet*.

2.1 DATA WEBHOUSE A EVOLUÇÃO DO DW PARA A WEB

Atualmente, com a popularização da *internet*, os *sites* vêm recebendo milhares de acessos diários, onde se torna possível a coleta de informações valiosas, que podem auxiliar na melhoria dos serviços oferecidos e definição de estratégias competitivas. Deste modo, tornou-se necessário a criação de um DW focado nos usuários da *web*. A união entre a *web* e o DW origina o *Data Webhouse* – DW_h.

A interação de usuários em *sites*, através de seqüência de cliques (*clickstream*), gera grande quantidade de dados comportamentais potencialmente úteis no sentido de fornecer detalhes a respeito de cada ação (chique). Mais especificamente, a seqüência de cliques é uma série cronológica de ações minuciosas de usuários que pode ser agrupada em sessões (KIMBALL e MERZ, 2000).

Em uma loja de varejo qualquer, o cliente caminha pelos seus corredores, observando produtos, comparando seus preços, selecionando alguns deles, descartando outros, até que, depois de tudo terminado, ele vai até o caixa e registra suas compras. Diferentemente destes tipos de lojas, em um ambiente virtual, é possível rastrear o cliente mensurando o que ele observa, o tempo de observação e quais produtos são rejeitados depois de observados, etc. Igualmente, ainda é possível reestruturar os “corredores” a fim de adequar aos anseios do cliente.

Atualmente, grande parte dos Servidores *web* possui mecanismos para registrar as ações (ou cliques) dos usuários que o visita, cujo objetivo é medir a eficiência do servidor e o impacto provocado por ele. A informação armazenada pode ser utilizada para personalizar dinamicamente a apresentação do *site* ao usuário, aumentando a relevância do conteúdo disponibilizado e o interesse à medida que se navega, ou quando do retorno em outra oportunidade.

Durante muito tempo, os arquivos de *log* (registro) dos servidores eram vistos apenas como um mecanismo para a detecção de atitudes indevidas ou mal intencionadas. Enormes volumes de dados, ricos em informações escondidas, eram jogados fora, todos os dias, pelos administradores dos *sites* em geral.

O reconhecimento da importância das informações armazenadas nos *logs* não apenas fez com que esses arquivos fossem analisados, mas também estimulou o desenvolvimento de novos mecanismos para a captura do comportamento do usuário que visita um determinado *site*.

Os *sites* relacionados às empresas de grande porte podem capturar milhões de ações diariamente, conseqüentemente, os bancos de dados que coletam este tipo de fluxo de informações estão se tornando cada vez maiores. Daí surge a necessidade de utilizar o *Data Warehouse* (DW), visto que é um sistema adequado para armazenar e analisar grande quantidade de dados (ZAIANE, 2005).

Portanto, os dados comportamentais de usuários da empresa gerados pela *Web* alimentam o *Data Warehouse*, mais especificamente, *Data Webhouse* (DWh), permitindo analisar detalhadamente o comportamento do usuário, obtendo informações sobre o cliente a cada clique, gesto e trajetória em um *site*, além da possibilidade de ser capaz de responder perguntas tais como:

- ✚ Qual é o local do *site* mais visitado? E o menos visitado?
- ✚ Quais páginas do *site* são vistas por último, ou seja, onde os usuários encerram a sessão?
- ✚ Qual o perfil de navegação de um usuário cadastrado? E de um usuário rentável?
- ✚ Qual o perfil de navegação de um usuário que cancela o serviço, faz reclamação ou processa a empresa?
- ✚ Quanto tempo os clientes gastam em nosso *site*, e quantas páginas eles visitam?
- ✚ Como induzir um usuário a se registrar no intuito de obter informações valiosas sobre o mesmo?

A partir das considerações explicitadas, é ilustrada a arquitetura que envolve um DWh em uma empresa de comércio eletrônico na Figura 1.

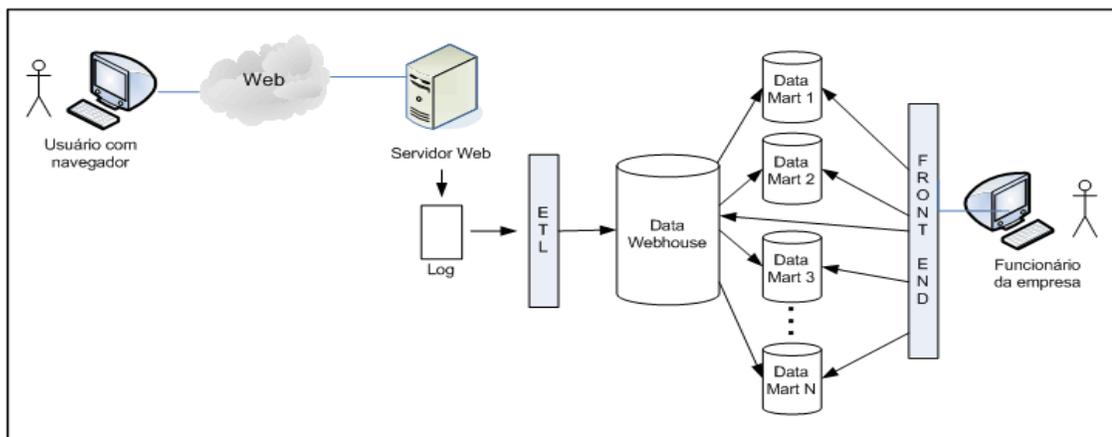


FIGURA 1: Arquitetura Envolvida para Desenvolvimento de um DWh

FONTE: (KIMBALL e MERZ, 2000)

Um usuário, através de um navegador, acessa o *site* hospedado no Servidor *Web* da empresa. As interações do usuário com o *site* são então capturadas pelo Servidor *Web*, e armazenadas em um *log*.

A partir do *log*, inicia-se a limpeza dos dados (remoção de inconsistências nos dados) e em seguida o processo de extração, limpeza, transformação e carga (ETL – *Extract, Transform and Load*) dos dados para o DWh por meio do componente ETL. Este tem como objetivo servir como um pós-processador de seqüência de cliques, preparando os dados capturados para serem carregados no *Data Webhouse*.

Uma vez alimentado o DWh, são criados vários *Data Marts* (DM) a depender da finalidade imediata. *Data Marts* são subconjuntos lógicos do DWh voltados para áreas mais específicas. Kimball (2002) define um DM como sendo uma parte do *Data Warehouse* restrita a um único processo de negócio, que tem como objetivo dar suporte à análise das informações sobre o comportamento dos usuários durante a utilização de *sites* na *web* e sistemas de informação. Por fim, usando um componente *front end*, conjunto de aplicações que permite gerenciar o DWh e acessar seus respectivos DMs, um usuário final da empresa analisa as informações.

Ainda através do *Data Mart*, pode-se agrupar os clientes de forma bastante simples, de acordo com três critérios:

- ✚ Recentidade: quantos dias faz desde a última visita do usuário;
- ✚ Frequência: quantas vezes o usuário esteve no *site*;
- ✚ Intensidade: total das compras do usuário;

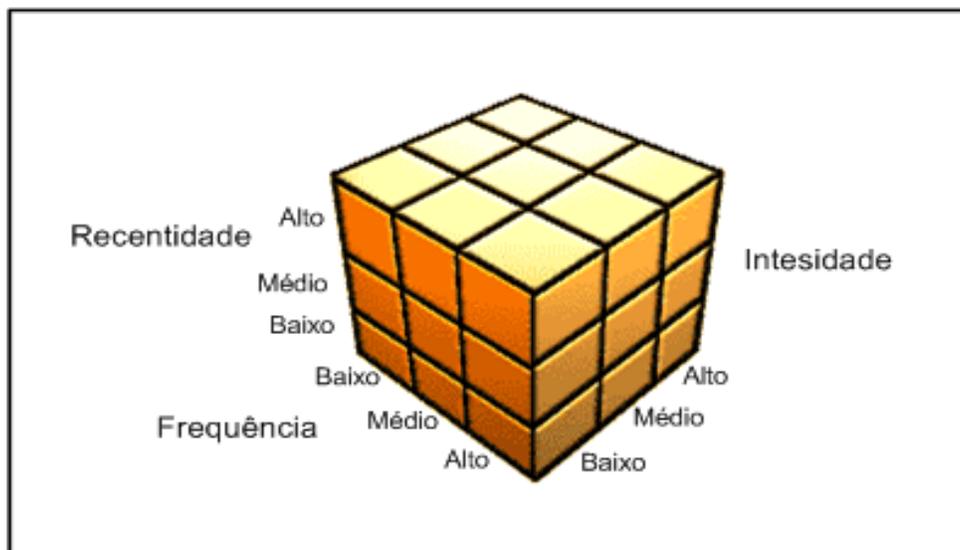


Figura 2: Técnica de Agrupamento de Usuários

Fonte: (KIMBALL e MERZ, 2000)

Em seguida, para cada critério, podem-se atribuir três valores intuitivos: alto, médio e baixo. Assim, cada usuário será encaixado em uma das 27 possíveis células (3 valores de recentidade x 3 valores de frequência x 3 valores de intensidade), como é demonstrado na figura 2.

Inicialmente, têm-se os dados “brutos” (*log do servidor Web*) que são convertidos em informação ao exportá-los para o DWh. Esta, por sua vez, é transformada em conhecimento através dos diversos questionamentos que podem ser feitos. Por fim, as informações contidas no DWh tem como objetivo dar apoio ao processo de tomada de decisões da empresa.

Outra ferramenta de TI que utiliza as informações obtidas e processadas pelo DWh é o CRM. A ligação entre essas ferramentas acontece devido ao DWh gerar informações para o processo de tomada de decisão, e CRM utilizar essas informações na busca da melhoria do relacionamento com os clientes.

3 Relacionamentos com os clientes – CRM

A ferramenta CRM (Customer Relationship Management) procura capturar os dados do cliente ao longo de toda a empresa, consolidar todos os dados capturados interna e externamente em um banco de dados central, analisa-los, distribuir os resultados dessa análise aos vários pontos de contato com o cliente e usar essa informação ao interagir com o cliente através de qualquer ponto de contato com a empresa.

Com o CRM, as companhias devem reconhecer que os clientes fazem as decisões de compra baseadas mais do que apenas no produto ou no preço. Os clientes compram baseados em sua experiência geral que inclui o produto e o preço, mas incluem também vendas, serviço, imagem percebida, suporte e atendimento. Se as companhias puderem obter todos aqueles fatores correta e consistentemente estarão recompensadas com um cliente fiel e rentável.

Uma das principais formas de viabilização para a implementação do *e-commerce* (comércio eletrônico) tem sido a criação de empresas virtuais que, com o uso de Tecnologias de Informação, oferecem um mix de produtos e/ou serviços para o consumidor.

Mesmo sabendo das vantagens e benefícios oriundos da prática do *e-commerce*, a atitude e o comportamento do consumidor em relação a este novo meio de compra tem sido numa direção inversa, dando por preferência, em alguns casos, a compras em empresas tradicionais. Muitos consumidores ainda não conduzem esta transação pela *Internet* por conta dos riscos percebidos neste processo, o que influencia a decisão de compra de um produto transacionado virtualmente.

Muitas empresas têm utilizado largamente a *Internet* a fim de manter um relacionamento mais próximo com seus fornecedores e clientes, podendo assim aumentar o número de transações, com intuito de atender um maior número de clientes de forma mais rápida e personalizada.

Um novo termo, e-CRM, engloba as funções de *e-commerce* com CRM. O “e” indica personalização do relacionamento com o cliente com reconhecimento (identificação) automático das oportunidades de venda; significa que o sistema captura eletronicamente as regras de comércio desempenhadas no mundo de negócios convencional de *marketing* e vendas. O e-CRM provê a reintermediação da *Internet*, revertendo a tendência de desintermediação surgida junto com a *Web*, que eliminou o componente humano no relacionamento com o cliente sem substituição adequada.

A implementação das funções de CRM para personalização e automação de forma efetiva, apóia-se em ferramentas robustas para análise de dados. Acoplando-se e-CRM e as facilidades de *Data Warehouse*, pode-se chegar à integração e automação de todas as etapas do relacionamento

com o cliente desde o *marketing*, passando pela produção e distribuição, até a venda e o suporte ao cliente.

4 Segurança no data webhouse

Um bom programa de segurança é essencial para os interesses de *CRM (Customer Relationship Management)*. O cliente precisa estar seguro de que informações confidenciais sobre ele não estarão disponíveis para pessoas não autorizadas.

Segurança da Conexão - Quando o usuário conecta-se ao *Data WebHouse* através da *Web*, deve-se assumir que a conexão é feita através de uma rede insegura e que pode ocorrer espionagem. Nesse caso, uma conexão de rede privada virtual deve ser estabelecida. Uma conexão de *VPN (Virtual Private Network)* criptografa toda a comunicação entre o usuário e o servidor de autenticação.

Qualquer usuário final conectado à rede local pode instalar esse *Sniffer* e rapidamente coletar senhas e outras comunicações sensíveis que fluem através da rede. Por essa razão, recomenda-se que o *HandShaking* de autenticação entre o usuário final e o servidor de autenticação esteja criptografado, mesmo em uma *Intranet* corporativa supostamente segura.

Definir Perfil do Usuário Autenticado - A cada usuário deverá ser atribuído um ou mais perfis. Uma organização grande pode possuir dúzias ou centenas de perfis. Mas o objetivo dos perfis é reduzir o número de decisões complexas que devem ser tomadas para submeter os usuários a utilização adequada dos dados.

Funcionários (técnicos, gerentes, executivos e diretores), fornecedores e clientes devem ter perfis separados. Dessa maneira, certifica-se de que cada perfil pode ser isolado, de modo que os dados não ultrapassem os limites estabelecidos.

5 Considerações Finais

Através da análise dos dados obtidos dos arquivos de *log* por meio de ferramentas de extração de informação e utilizando-se técnicas apropriadas é possível se obter, por área de interesse, informações muito relevantes para a tomada de decisão. Tais informações dizem respeito às características e ao perfil das pessoas que navegam em um portal.

O uso da tecnologia *Webhouse* levará ao empreendedor explorar novas possibilidades, a conhecer melhor seus clientes (usuários) e a oferecer-lhes novos produtos e serviços. Desta forma, não só seria beneficiada a estrutura do *site* como também permitiria uma maior divulgação dos produtos oferecidos aos clientes na *Web*.

Diante do exposto, percebe-se que existem muitos desafios a serem transpostos e que a *Web* é uma forma de mercado diferente do mercado tradicional. O *Data Webhouse* é a fonte de informações que apoiará o relacionamento com clientes na *Internet*. Não um serviço que simplesmente manda *e-mails* sem critério específico, mas sim a automação que suporta uma série de campanhas segmentadas para cada tipo de cliente, podendo atender melhor às suas necessidades específicas e utilizando-se de um investimento de *marketing* mais preciso e eficiente.

Referências Bibliográficas

BARBOSA, D. M.; SELL, D.; PACHECO, R. C. S.; JUNIOR, O. G. F. Date, C. J. Metodologia para desenvolvimento de *data webhouse* voltado para o monitoramento de portais corporativos. In: Workshop Brasileiro de Inteligência Competitiva e Gestão do Conhecimento, São Paulo, 2002.

DATE, C. J. *Introdução ao Sistema de Banco de Dados*, Rio de Janeiro: Campus, 2000.

INMON, W.H. *Como Construir Data Warehouse*, Campus, 1997.

KIMBALL, R.; MERZ, R. *Data Webhouse: construindo o Data Warehouse para a Web*. Rio de Janeiro: Campus, 2000.

KIMBALL, R. *Data Warehouse Toolkit*, São Paulo: Campus, 2002.

KIMBALL, R. *Clicking with your Customer. On-line.* Disponível em: http://www.intelligententerprise.com/db_area/archives/1999/990501/warehouse.jhtml?requestid=48688. Acesso em 16 Março 2005.

KIMBALL, R. *The special dimensions of the Clickstream, Intelligent Enterprise. On-line, em* <http://www.intelligententerprise.com/00120/webhouse.jhtml?requestid=543846>. Acesso em 16 Março de 2005.

KIMBALL, R. *Clickstream Data Mart, Intelligent Enterprise. On-line.* Disponível em: http://www.intelligententerprise.com/011205/418warehouse1_1.jhtml?requestid=544002. Acesso em 16 março 2005.

MUSSI, C. *Data Warehouse - Da Modelagem a Implantação*, Anais do Intempres, V *Workshop Internacional de Inteligência Empresarial*, 2004.

ZAIANE, O. *Glossary of Data Mining Terms. University of Alberta*, 2005. Disponível: <http://www.cs.ualberta.ca/~zaiane/courses/cmp690/glossary.html> [capturado em 15 de mai. 2005].

SILBERSCHATZ, ABRAHAM; KORTH, HENRY; SUDARSHAN, S. *Sistema de Banco de Dados*, São Paulo: Makron Books, 1999.