

Data Warehouse: uma ferramenta de tecnologia de informação para as organizações.

Gabriela Góis Cavalcanti (UFPE/PROPAD) gabrielagoisc@yahoo.com.br

André Felipe de Albuquerque Fell (UFPE/PROPAD) anfaf_05@yahoo.com.br

Jairo Simião Dornelas (UFPE/PROPAD) jairo@ufpe.br

Resumo

Em termos simples, um Data Warehouse, ou em português, Armazém de Dados, pode ser definido como um banco de dados especializado, o qual integra e gerencia o fluxo de informações a partir dos bancos de dados corporativos e fontes de dados externas à empresa. Grande parte do trabalho na construção de um DW está na análise dos sistemas transacionais e dos dados que eles contêm.

Um Data Warehouse permite a geração de dados integrados e históricos auxiliando os diretores a decidirem embasados em fatos e não em intuições ou especulações, o que reduz a probabilidade de erros aumentados à velocidade na hora da decisão. Conhecer mais sobre essa tecnologia permitirá aos administradores descobrir novas maneiras de diferenciar sua empresa numa economia globalizada, deixando-os mais seguros para definirem as metas e adotarem diferentes estratégias em sua organização, conseguindo assim visualizar antes de seus concorrentes novos mercados e oportunidades atuando de maneiras diferentes conforme o perfil de seus consumidores.

O artigo contempla uma apresentação da importância do Data Warehouse nas organizações e algumas definições e características dessa tecnologia. Disserta-se também sobre sua implementação e alguns problemas que podem ocorrer. Para finalizar são apresentadas algumas ferramentas de extração de informações e considerações finais.

Palavras-Chave: Data Warehouse; Organizações; Ferramentas.

1 Introdução

Uma importante questão estratégica para o sucesso de qualquer organização nos dias de hoje é a sua capacidade de analisar, planejar e reagir, rápida e imediatamente, às mudanças nas condições de seus negócios. Para que isso aconteça, é necessário que a organização disponha de mais e melhores informações, que constituem, reconhecidamente, a base destes processos.

Os avanços da tecnologia de informação vieram garantir a possibilidade das empresas manipularem grandes volumes de dados e atingirem um alto índice de troca de informações,

com o uso das redes viabilizando operações em nível mundial. Com a globalização, os negócios não têm mais fronteiras, ter a informação correta no menor tempo possível e utilizar-se de sistemas de apoio à decisão tornou-se o grande diferencial para as empresas.

Diariamente, dados sobre os mais variados aspectos dos negócios da organização são gerados e armazenados, e passam a fazer parte dos recursos de informação da mesma. Entretanto, essas informações encontram-se, em geral, espalhadas em diversos sistemas e exigem um esforço considerável de integração para que possam dar suporte efetivo à tomada de decisão de gerentes e executivos. Isso significa que, em que pese os avanços da tecnologia de armazenamento e manipulação de dados, ainda se verifica, nos dias de hoje, uma grande deficiência na obtenção ágil de informações estratégicas (datawarehouse, 2005).

Por esse motivo, um novo conjunto de conceitos e ferramentas vem ganhando enorme destaque nos últimos anos como a tecnologia de Data Warehouse, que oferece às organizações uma maneira flexível e eficiente de obter as informações necessárias a seus processos decisórios. De acordo com Elmasri e Navathe (1994) o Data Warehouse (DW) provê armazenamento, funcionalidade e responsividade às perguntas que exigem capacidade de armazenamento além dos bancos de dados orientados para transações. Provê também acesso a análise de dados complexos, obtenção de conhecimento e, principalmente, está direcionada para aplicações de suporte à decisão.

2 Data Warehouse como uma Necessidade Organizacional

Com a evolução da tecnologia de informação e o crescimento do uso de computadores interconectados, praticamente todas as empresas de médio e grande porte estão utilizando sistemas informatizados para realizar seus processos mais importantes. Com o passar do tempo esses sistemas acabam gerando uma enorme quantidade de dados relacionados ao negócio, mas não relacionados entre si, pois é normal que estes dados estejam espalhados por diversos locais e que tenham sido gerados por sistemas desenvolvidos em diferentes ambientes e linguagens.

Sabe-se que os bancos de dados são de vital importância para as empresas e também que sempre foi difícil analisar os dados neles existentes. Ou seja, estes dados armazenados em um ou mais sistemas operacionais de uma empresa são um recurso, mas de modo geral, raramente servem como recurso estratégico no seu estado original.

Os sistemas convencionais de informática não são projetados para gerar e armazenar as informações estratégicas, o que torna os dados vagos e sem valor para o apoio ao processo de tomada de decisões das organizações. Além disso, não se consegue buscar informações que permitam à tomada de decisão embasada num histórico dos dados. As decisões normalmente são tomadas com base na experiência dos administradores quando poderiam também ser baseadas em fatos históricos, que foram armazenados pelos diversos sistemas de informação utilizados pelas organizações. O uso dos dados históricos poderia identificar as tendências e posicionar a empresa estrategicamente para ser mais competitiva e conseqüentemente maximizar os lucros diminuindo o índice de erros na tomada de decisão. Cerca de 88% dos diretores admitem que dedicam quase 75% do tempo às tomadas de decisão apoiadas em análises subjetivas (Aspect International Consulting, 1997, *apud* datawarehouse, 2005), menosprezando o fato de que por volta de 100% deles tem acesso a computadores.

A necessidade de novos métodos e tecnologias surgiu da constatação, primeiro de que existe uma necessidade de informação não atendida pelos aplicativos comerciais convencionais, que

atuam a nível operacional do negócio, e segundo, pelo fato de que a tecnologia de armazenamento de dados utilizada nestes aplicativos não atende às necessidades detectadas. Pensando nisso, introduziu-se um novo conceito no mercado, o Data Warehouse (DW). Turban, Rainer e Potter (2003) corroboram com as afirmações acima ao descrever o DW como um transformador de dados em recursos mais útil, agrupando-os de modo mais adequado para os usuários finais, colocando-os em formatos mais úteis, permitindo a sua análise e distribuindo-os para os grupos de trabalho pertinentes, para aumentar a disponibilidade e acessibilidade. Esse consiste em organizar os dados corporativos da melhor maneira, para dar subsídio de informações aos gerentes e diretores das empresas para tomada de decisão. Stair e Reynolds (2002) creditam que ele seja a atual evolução dos sistemas de suporte à decisão.

3 Definições de Data Warehouse

O Data Warehouse (DW) é um conjunto de técnicas que aplicadas em conjunto geram um sistema de dados que proporcionam informações para tomada de decisões. Ele funciona tipicamente na arquitetura cliente/servidor. Segundo Inmon (1993), considerado um pioneiro no tema, um data warehouse é uma coleção de dados orientada por assuntos, integrado, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão.

O data warehouse é um banco de dados contendo dados extraídos do ambiente de produção da empresa, que foram selecionados e depurados, tendo sido otimizados para processamento de consulta e não para processamento de transações. Em comparação com os banco de dados transacionais, o data warehouse possui enormes quantidades de dados de múltiplas fontes, as quais podem incluir banco de dados de diferentes modelos e algumas vezes arquivos adquiridos de sistemas independentes e plataformas (Elmasri e Navathe, 1994).

De acordo com Richard Hackathorn (Data Warehouse, 2005), presidente e fundador da Bolder Technology Inc. – consultoria com 12 anos de mercado - o objetivo de um data warehouse é fornecer uma "imagem única da realidade do negócio". De uma forma geral, sistemas de data warehouse compreendem um conjunto de programas que extraem dados do ambiente de dados operacionais da empresa, um banco de dados que os mantém, e sistemas que fornecem estes dados aos seus usuários.

Sistemas de Data Warehouse revitalizam os sistemas da empresa, pois de acordo com WG Systems (2005):

- Permitem que sistemas mais antigos continuem em operação;
- Consolidam dados inconsistentes dos sistemas mais antigos em conjuntos coerentes;
- Extraem benefícios de novas informações oriundas das operações correntes;
- Provêm ambientes para o planejamento e arquitetura de novos sistemas de cunho operacional.

Em termos simples, um Data Warehouse, ou em português, Armazém de Dados, pode ser definido como um banco de dados especializado, o qual integra e gerencia o fluxo de informações a partir dos bancos de dados corporativos e fontes de dados externas à empresa.

Um Data Warehouse é construído para que tais dados possam ser armazenados e acessados de forma que não sejam limitados por tabelas e linhas estritamente relacionais. Elmasri e Navathe (1994) complementam ao afirmando que ele é um modelo de dados multidimensional que armazena dados de forma integrada e disponibiliza séries no tempo e análises de tendências através de banco de dados histórico e não-volatilidade. É interessante salientar que o DW está separado dos bancos de dados operacionais, portanto as consultas dos usuários não impactam nestes sistemas, que ficam resguardados de alterações indevidas ou perdas de dados.

Stair e Reynolds (2002) atestam que como o propósito do DW é suportar requisições de informações, esse tem como características: fonte de dados de arquivos múltiplos, de banco de dados internos e externos; os acessos dos usuários são para somente leitura; o modo de acesso primário aos dados se dá por consultas simples ou complexas, com uso crescente do data mining; utilização de um modelo de banco de dados relacional e multidimensional; os níveis de detalhes são frequentemente resumidos; dados históricos de múltiplos anos; há um processo periódico de atualização, mas esses são complexos, já que se combinam diversas fontes; e é necessário um grande esforço de limpeza dos dados para que esses fiquem íntegros.

4 Características do Data Warehouse

Segundo Inmon (1996), um DW deve ser orientado por assuntos, integrado, variável no tempo e não volátil. Essas seriam as principais características de um DW, porém existem outras também importantes como a localização, credibilidade dos dados e granularidade (Cazella, 2005).

4.1 Orientação por Assunto

A orientação por assunto é uma característica marcante de um DW, pois toda modelagem é voltada em torno dos principais assuntos da empresa. Enquanto todos os sistemas transacionais estão voltados para processos e aplicações específicas, os DWs objetivam assuntos. Os assuntos são o conjunto de informações relativas à determinada área estratégica de uma empresa. Um exemplo típico pode ser ilustrado da área de vendas como produtos, revendedores, contas e clientes.

4.2 Integração

Esta característica talvez seja a mais importante do DW. É através dela que se padroniza uma representação única para os dados de todos os sistemas que formarão a base de dados do DW. Por isso, grande parte do trabalho na construção de um DW está na análise dos sistemas transacionais e dos dados que eles contêm. Esses dados geralmente encontram-se armazenados em vários padrões de codificação, isso se deve aos inúmeros sistemas existentes nas empresas, e que eles tenham sido codificados por diferentes analistas. Isso quer dizer que os mesmos dados podem estar em formatos diferentes.

Um exemplo clássico é o que se refere aos gêneros masculino e feminino. Num sistema OLTP, o analista convencionou que o sexo seria 1 para masculino e 0 para feminino, já em outro sistema outro analista usou para guardar a mesma informação a seguinte definição, M para masculino e F para feminino, e por fim outro programador achou melhor colocar H para masculino e M para feminino. Como podemos ver, são as mesmas informações, mas estão em formatos diferentes, e isso num DW jamais poderá acontecer. Portanto é por isso que deverá

existir uma integração de dados, convencionando-se uma maneira uniforme se armazenar os mesmos.

4.3 Variação no Tempo

Segundo Inmon (1996), os Data Warehouses são variáveis em relação ao tempo, isso nada mais é do que manter o histórico dos dados durante um período de tempo muito superior ao dos sistemas transacionais. Diz respeito ao fato do dado em um data warehouse referir-se a algum momento específico, significando que não é atualizável, a cada ocorrência de uma mudança, uma nova entrada é criada, para marcar esta mudança.

Já que no Data Warehouse, o principal objetivo é analisar o comportamento das mesmas durante um período de tempo maior e fundamentados nessa variação é que os gerentes tomam as decisões, então num DW é freqüente manter-se um horizonte de tempo bem superior ao dos sistemas transacionais. Deste modo é válido dizer que os dados nos sistemas transacionais estão sendo atualizados constantemente, cuja exatidão é válida somente para o momento de acesso. Já os dados existentes num DW são como fotografias que refletem os mesmos num determinado momento do tempo. Essas fotografias são chamadas de *snapshots* (Cazella, 2005).

A dimensão tempo, sempre estará presente em qualquer fato de um DW, isso ocorre porque, como citado anteriormente, sempre os dados refletirão um determinado momento de tempo, e obrigatoriamente deverá conter uma chave de tempo para expressar a data em que os dados foram extraídos. Portanto pode-se dizer que os dados armazenados corretamente no DW não serão facilmente atualizados tendo-se assim uma imagem fiel da época em que foram gerados.

Assim como os dados, é importante frisar que os metadados, também possuem elementos temporais, porque mantêm um histórico das mudanças nas regras de negócio da empresa. Os metadados são responsáveis pelas informações referentes ao caminho do dado dentro do DW.

4.4 Não Volatilidade

No DW existem somente duas operações, a carga inicial e as consultas dos *front-ends* aos dados. Após serem integrados e transformados, os dados são carregados em bloco para o data warehouse, para que estejam disponíveis aos usuários para acesso. Isso pode ser afirmado porque a maneira como os dados são carregados e tratados é completamente diferente dos sistemas transacionais. No ambiente operacional, ao contrário, os dados são, em geral, atualizados registro a registro, em múltiplas transações. Esta volatilidade requer um trabalho considerável para assegurar integridade e consistência. Um data warehouse não requer este grau de controle típico dos sistemas orientados a transações, pois, no DW o que acontece é somente ler os dados na origem e gravá-los no destino, ou seja, no banco modelado multidimensional.

Deve-se considerar que os dados sempre passam por filtros antes de serem inseridos no DW. Com isso muitos deles jamais saem do ambiente transacional, e outros são tão resumidos que não se encontram fora do DW. Em outras palavras, a maior parte dos dados é física e radicalmente alterada quando passam a fazer parte do DW. Do ponto de vista de integração, não são mais os mesmos dados do ambiente operacional. À luz destes fatores, a redundância de dados entre os dois ambientes raramente ocorre, resultando em menos de 1 por cento de duplicações, essa definição é dada por Inmon (1996), e é muito válida.

4.5 Localização

Os dados podem estar fisicamente armazenados de três formas:

- Num único local centralizado, com banco de dados em um DW integrado procura-se nessa forma maximizar o poder de processamento e agilizando a busca dos dados. Esse tipo de armazenagem é bastante utilizada, porém há o inconveniente do investimento em hardware para comportar a base de dados muito volumosa e o poderio de processamento elevado para atender satisfatoriamente as consultas simultâneas de muitos usuários.
- Os distribuídos, em forma de Data Marts, armazenados por áreas de interesse. Por exemplo, os dados da gerência financeira num servidor, dados de marketing noutra e dados da contabilidade num terceiro lugar. Tem por benefício bastante performance, pois não sobrecarrega um único servidor e as consultas serão sempre atendidas em tempo satisfatório.
- Armazenados por níveis de detalhes, em que as unidades de dados são mantidas no DW. Pode-se armazenar dados altamente resumidos num servidor, dados resumidos com nível de detalhe intermediário no segundo servidor e os dados mais detalhados (atômicos), num terceiro servidor. Os servidores da primeira camada podem ser otimizados para suportar um grande número de acessos e um baixo volume de dados, enquanto alguns servidores nas outras camadas podem ser adequados para processar grandes volumes de dados, mas baixo número de acesso.

4.6 Credibilidade dos Dados

A credibilidade dos dados é o muito importante para o sucesso de qualquer projeto. Discrepâncias simples de todo tipo podem causar sérios problemas quando se quer extrair dados para suportar decisões estratégicas para o negócio das empresas. Dados não dignos de confiança podem resultar em relatório inúteis, que não têm importância alguma. "Se você tem dados de má qualidade e os disponibiliza em um DW, o seu resultado final será um suporte à decisão de baixo nível com altos riscos para o seu negócio", afirma Robert Craig, analista do Hurwitz Group (Data Warehouse, 2005).

Coisas aparentemente simples, como um CEP errado, podem não ter nenhum impacto em uma transação de compra e venda, mas podem influir nas informações referentes a cobertura geográfica, por exemplo. "Não é apenas a escolha da ferramenta certa que influi na qualidade dos dados", afirma Richard Rist, vice-presidente Data Warehousing Institute (Data Warehouse, 2005). Segundo ele, conjuntos de coleções de dados, processos de entrada, metadados e informações sobre a origem dos dados, são importantíssimos. Outras questões como a manutenção e atualização dos dados e as diferenças entre dados para bancos transacionais e para uso em Data Warehousing também são cruciais para o sucesso dos projetos. Além das camadas do DW propriamente dito, tem-se a camada dos dados operacionais, de onde os dados mais detalhados são coletados. Antes de fazer parte do DW estes dados passam por diversos processos de transformação para fins de integração, consistência e acurácia.

4.7 Granularidade

Granularidade nada mais é do que o nível de detalhe ou de resumo dos dados existentes num DW. Quanto maior for o nível de detalhes, menor será o nível de granularidade. O nível de

granularidade afeta diretamente o volume de dados armazenados no DW, e ao mesmo tempo o tipo de consulta que pode ser respondida.

Quando se tem um nível de granularidade muito alto o espaço em disco e o número de índices necessários, tornam-se bem menores, porém há uma correspondente diminuição da possibilidade de utilização dos dados para atender a consultas detalhadas.

Com o nível de granularidade muito baixo, é possível responder a praticamente qualquer consulta, mas uma grande quantidade de recursos computacionais é necessária para responder perguntas muito específicas. No entanto, no ambiente de DW, dificilmente um evento isolado é examinado, é mais provável que ocorra a utilização da visão de conjunto dos dados.

Os dados levemente resumidos compreendem um nível intermediário na estrutura do DW, são derivados do detalhe de baixo nível encontrado nos dados detalhados atuais. Este nível do DW é quase sempre armazenado em disco. Na passagem para este nível os dados sofrem modificações. Por exemplo, se as informações nos dados detalhados atuais são armazenadas por dia, nos dados levemente resumidos estas informações podem estar armazenadas por semanas. Neste nível o horizonte de tempo de armazenamento normalmente fica em cinco anos e após este tempo os dados sofrem um processo de envelhecimento e podem passar para um meio de armazenamento alternativo.

Os dados altamente resumidos são compactos e devem ser de fácil acesso, pois fornecem informações estatísticas valiosas para os Sistemas de Informações Executivas (EIS), enquanto que nos níveis anteriores ficam as informações destinadas aos Sistemas de Apoio a Decisão (SAD), que trabalham com dados mais analíticos procurando analisar as informações de forma mais ampla.

O balanceamento do nível de granularidade é um dos aspectos mais críticos no planejamento de um DW, pois na maior parte do tempo há uma grande demanda por eficiência no armazenamento e no acesso aos dados, bem como pela possibilidade de analisar dados em maior nível de detalhes. Quando uma organização possui grandes quantidades de dados no DW, faz sentido pensar em dois ou mais níveis de granularidade, na parte detalhada dos dados. Na realidade, a necessidade de existência de mais de um nível de granularidade é tão grande, que a opção do projeto que consiste em duplos níveis de granularidade deveria ser o padrão para quase todas as empresas.

O chamado nível duplo de granularidade se enquadra nos requisitos da maioria das empresas. Na primeira camada de dados ficam os dados que fluem do armazenamento operacional e são resumidos na forma de campos apropriados para a utilização de analistas e gerentes. Na segunda camada, ou nível de dados históricos, ficam todos os detalhes vindos do ambiente operacional. Como há uma verdadeira montanha de dados neste nível, faz sentido armazenar os dados em um meio alternativo como fitas magnéticas.

Com a criação de dois níveis de granularidade no nível detalhado do DW, é possível atender a todos os tipos de consultas, pois a maior parte do processamento analítico dirige-se aos dados levemente resumidos que são compactos e de fácil acesso. E para ocasiões em que um maior nível de detalhe deve ser investigado existe o nível de dados históricos. O acesso aos dados do nível histórico de granularidade é caro, incômodo e complexo, mas caso haja necessidade de alcançar esse nível de detalhe, lá estará ele.

5 Implantação de um Data Warehouse

O DW não é como um *software*, que pode ser comprado e instalado em todos os computadores da empresa em algumas horas, na realidade sua implantação exige a integração de vários produtos e processos. O Data Warehouse (DW) apresenta um processo complexo composto por vários itens como metodologias, técnicas, máquinas, bancos de dados, ferramentas de *front-end*, extração, metadados, refinamento de dados, replicação, e principalmente, recursos humanos. Cada elo desta corrente está sujeito a falhas que podem transformar um projeto de milhões de reais, que era tido pela empresa como a tábua de salvação para seus problemas, num grande pesadelo (Flores, 2005).

O momento mais crucial de todo processo, e o causador da maior parte dos problemas, é o da escolha das ferramentas, dos bancos de dados, das consultorias, e da definição do escopo do projeto e da seleção dos indivíduos que farão parte do staff de DW. Mais do que um bancos de dados de última geração é importante selecionar com extremo critério os profissionais que farão parte de um projeto.

Depois da escolha dos profissionais responsáveis, deve-se fazer um levantamento e definir os objetos de negócios e, por conseguinte, as questões gerenciais que os respondem. Essa etapa é de extrema importância, pois é ela quem irá determinar quais serão os dados a serem armazenados no Data Warehouse. Exigirá, também, uma metodologia específica e diferente dos sistemas transacionais. Terminada esta etapa, partiremos então para a segunda fase da implantação que consiste em fazer a modelagem dimensional, ou seja, partindo dos objetos gerenciais levantados faz-se a modelagem do novo banco gerando os fatos e as dimensões. Logo em seguida, parte-se para a criação física do modelo, onde as especificidades de um SGBD e da ferramenta OLAP escolhidos são levadas em consideração para otimizar as futuras consultas ao banco e onde se deve dar preferência os esquemas estrela.

Feito isso, o passo seguinte é a carga dos dados no DW. Para isso, é necessário que se definam as origens dos mesmos nos sistemas legados, ou seja, identificar em quais sistemas e onde estão armazenados. Nessa etapa é imprescindível a presença de um analista de sistemas da empresa que conheça profundamente os sistemas transacionais, pois isso facilitará muito o trabalho de localização e identificação dos dados.

Seguindo o roteiro descrito por Flores (2005) a próxima etapa é fazer as rotinas de extração dos dados. Essas rotinas podem ser desenvolvidas por programadores em qualquer linguagem de programação. Após a extração resta dar a carga dos dados no banco dimensional criado no segundo passo. Essa carga também pode ser feita através de rotinas desenvolvidas pelos programadores. Concluída a carga deve-se fazer uma checagem profunda da consistência dos dados, isso é muito importante, pois está se trabalhando com informações para dar suporte a decisão e, qualquer dado errado poderá determinar o fracasso da análise do negócio em questão.

Outro passo importantíssimo na elaboração é a confecção e armazenamento dos metadados. Os metadados são os dados de controle do Data Warehouse.

Para fazermos a análise dos dados temos as ferramentas OLAP, que permitem a visualização dos dados da base dimensional e sua análise de acordo com a imaginação do executivo. Com esses softwares, também denominados ferramentas de *front-end*, o usuário poderá analisar as informações que estão contidas no banco multidimensional.

Além dos passos anteriores, os usuários devem ser treinados de forma a aprender como manipular as informações existentes seja na criação das estatísticas, seja na criação de gráficos e relatórios.

7 Problemas que podem existir na Implantação de um Data Warehouse

Existem diversos problemas que podem ocorrer durante o desenvolvimento de um sistema de DW. Dentre estes problemas, segundo Bar (1996, *apud* Data Warehouse, 2005), os mais comuns são:

a) Não envolver a alta direção da empresa no projeto: O projeto de um DW só terá sucesso se os futuros usuários se envolverem diretamente desde o início nas atividades, pois isto facilitará a destinação das verbas necessárias nos momentos oportunos além de direcionar os trabalhos para que os reais objetivos do DW para o negócio da empresa sejam alcançados no momento da implantação.

b) Gerar falsas expectativas com promessas que não poderão ser cumpridas: citar frases do tipo "O DW mostrará aos gerentes as melhores decisões" pode causar tanto desconfiança no projeto quanto desprezo. O DW não mostrará as melhores decisões, mas sim respostas às consultas efetuadas. Cabe aos usuários elaborar consultas inteligentes e analisar as respostas obtidas.

c) Carregar no DW informações somente porque elas estão disponíveis nos sistemas transacionais: Nem todos os dados disponíveis nos sistemas operacionais da empresa são necessariamente úteis para o DW. Cabe ao arquiteto dos dados analisar junto com os usuários quais os dados que realmente contêm informações necessárias e desprezar aqueles que não fazem parte dos objetivos do DW.

d) Imaginar que o projeto do banco de dados do DW é o mesmo que o projeto de um sistema transacional: num processo transacional devem ser dimensionados os recursos para que se atinja uma alta velocidade de acesso e grandes facilidades na atualização de registros. Nos sistemas de apoio à decisão a realidade é totalmente outra. O objetivo destes sistemas é fornecer acessos agregados, ou seja, somas, médias, tendências, etc. Outra diferença entre os dois tipos de sistemas pode ser detectado no tipo de usuários. Nos sistemas transacionais um programador desenvolve uma consulta que poderá ser utilizada milhares de vezes. No DW o usuário final desenvolve suas próprias consultas que podem ser utilizadas somente uma vez.

e) Na seleção do pessoal, escolher um gerente para o DW com orientação essencialmente técnica: os sistemas de apoio à decisão são na verdade uma prestação de serviços e não um serviço de armazenamento de dados. Por isso, é fundamental que o gerente do DW seja uma pessoa voltada aos interesses dos usuários e principalmente que, saiba dos termos utilizados diariamente pelos altos gerentes e outros tomadores de decisões.

f) Dedicar-se ao tratamento de dados do tipo registros numéricos e *string*: Muitos poderiam imaginar que as informações que serão utilizadas em um DW seriam oriundas especificamente dos registros das bases de dados transacionais, e que estas informações seriam apenas números ou palavras. Porém a inclusão de textos, imagens, sons e vídeos podem ser bastante úteis no momento da análise de determinadas situações da empresa e do negócio.

g) Projetar um sistema com base em um *hardware* que não poderá comportar o crescimento da demanda do DW: A capacidade dos servidores está em constante crescimento, porém pode ser dimensionando um ou mais equipamentos para trabalharem por um ou dois anos, mas as características destes sistemas indicam que a quantidade de informações armazenadas pode atingir até mesmo alguns terabytes. É importante que o servidor do banco de dados do DW seja fornecido por uma empresa confiável e que garanta a possibilidade de serem realizadas expansões a valores e prazos compatíveis com os de mercado.

h) Imaginar que após a implantação do DW os problemas estarão terminados: Muito esforço deve ser despendido para que um sistema de DW saia da prancheta. Porém, após a sua implantação os usuários começarão a criar mais e mais consultas, e estas consultas necessitarão de novos dados que resultarão em novas consultas. Desta forma, o projeto de um sistema de apoio à decisão precisa ser revisado e atualizado constantemente, não apenas com novos dados, mas também com novas tecnologias.

8 Ferramentas que Permitem Extrair Informações do Data Warehouse

Mesmo sabendo que a informação sobre o perfil do cliente típico ou do produto de sucesso de uma empresa encontra-se de alguma forma entre os muitos gigabytes de dados de marketing e de vendas armazenados nos bancos de dados da empresa, ainda pode existir um longo caminho a ser percorrido até que esta informação esteja de fato disponível. A sua extração eficaz, de modo a poder subsidiar decisões, depende da existência de ferramentas especializadas que permitam a captura de dados relevantes mais rapidamente e a sua visualização através de várias dimensões. O termo extração neste contexto não deve ser confundido com a extração dos dados das fontes para posterior alimentação do data warehouse.

As ferramentas não devem apenas permitir o acesso aos dados, mas também permitir análises de dados significativas, de tal maneira a transformar dados brutos em informação útil para os processos estratégicos da empresa. O sucesso de um data warehouse pode depender da disponibilidade da ferramenta certa para as necessidades de seus usuários.

As ferramentas mais simples são os produtos para consultas e geradores de relatórios básicos. Estes geradores de relatório não atendem a usuários que precisem mais do que uma visão estática dos dados e que não pode mais ser manipulada. Ferramentas OLAP podem oferecer a este tipo de usuário maior capacidade de manipulação, permitindo analisar o porque dos resultados obtidos. Estas ferramentas, muitas vezes, são baseadas em bancos de dados multidimensionais, o que significa que os dados precisam ser extraídos e carregados para as estruturas proprietárias do sistema, já que não há padrões abertos para o acesso de dados multidimensionais - outra solução oferecida por fornecedores nesta área é o OLAP relacional (ROLAP) e o OLAP multidimensional (MOLAP). O OLAP não é uma solução imediata, configurar o programa de OLAP e ter acesso aos dados requer uma clara compreensão dos modelos de dados da empresa e das funções analíticas necessárias aos executivos e outros analistas de dados.

Comparativamente ao OLAP, Sistemas de Informações Executivas (SIE) apresentam uma visualização de dados mais simplificada, altamente consolidada e, na maior parte das vezes estática. Até porque, em geral, os executivos não dispõem do tempo e da experiência para executar uma análise OLAP.

O data mining ou mineração de dados é uma categoria de ferramentas de análise *open-end*. Ao invés de fazerem perguntas, os usuários entregam para a ferramenta grandes quantidades de dados em busca de tendências ou agrupamentos dos dados. A diferença entre sistemas do tipo SIE e a tecnologia de data mining pode ser vista da seguinte forma: se você tem perguntas específicas e sabe os dados de que necessita, utilize um SIE; quando você não sabe qual a pergunta, mas mesmo assim precisa de respostas, use data mining.

9 Considerações Finais

Um Data Warehouse permite a geração de dados integrados e históricos auxiliando os diretores a decidirem embasados em fatos e não em intuições ou especulações, o que reduz a probabilidade de erros aumentados à velocidade na hora da decisão.

Esse instrumento estudado mostra o quanto ele é útil para os gerentes das empresas para a obtenção e análise da informação se realizado da forma correta. Também é interessante ressaltar que não é um mecanismo de fácil implementação, pois exige de recursos dos mais diversos e dentre eles o recurso humano é que se destaca. Umás séries de ferramentas são necessárias para que se possa desfrutar de suas benéficas e vários aspectos relacionados aos modelos dimensionais que permitem uma organização dos processos envolvidos.

Conhecer mais sobre essa tecnologia permitirá aos administradores descobrir novas maneiras de diferenciar sua empresa numa economia globalizada, deixando-os mais seguros para definirem as metas e adotarem diferentes estratégias em sua organização, conseguindo assim visualizarem antes de seus concorrentes novos mercados e oportunidades atuando de maneiras diferentes conforme o perfil de seus consumidores.

Referências

CAMPOS, Maria Luiza; ROCHA, Arnaldo V. Filho. Data Warehouse. Disponível em: <http://tartaruga.nce.ufrj.br/dataware/tutorial/home.html>. Acesso em: 18 maio de 2005.

CAZELLA, Sílvio César. Slides de aula sobre sistema de apoio à decisão. Disponível em: http://www.inf.unisinos.br/~cazella/dss/sad_aula_11.pdf. Acesso em: 18 maio de 2005. _____
Slides de aula sobre sistema de apoio à decisão. Disponível em: http://www.inf.unisinos.br/~cazella/dss/sad_aula_12.pdf. Acesso em: 18 maio de 2005.

DATAWAREHOUSE. Disponível em: www.datawarehouse.inf.br. Acesso em : 18 maio 2005.

DATA WAREHOUSE. Disponível em: www.datawarehouse.com. Acesso em: 18 de maio de 2005.

ELMASRI, Ramez.; NAVATHE, Sham.. **Fundamentals of database systems**. 2nd ed. - Redwood City Calif. : Benjamin/Cummings, c1994

FLORES, Christian Feltrin. Projeto e Desenvolvimento de Datawarehouse Hospitalar. Disponível em: <http://www.hcaa.com.br/antiga/dw/index.htm>. Acesso em: 18 de maio de 2005.

INMON, W.H. **Information System Arqitetura: Development in 90's**. Nova York: John Wiley & Sons Inc., 1993. Partes do livro disponível em: www.billinmon.com.

_____. **Building the Data Warehouse**. Nova York: John Wiley & Sons Inc., 1996. Partes do livro disponível em: <http://www.billinmon.com/>.

JARKE, Matthias; LENZERINI, Maurizio; VASSILIOU, Yannis; VASSILIADIS, Panos. **Fundamentals of Data Warehouse**. Nova York: Springer, 2003, 2 ed.

KIMBALL, RALPH **Is E-R Modeling Hazardous to DSS?** DBMS, Outubro 1995.

STAIR, Ralph M.; REYNOLDS, George W. **Princípios de Sistemas de Informações**. Rio de Janeiro: LTC, 2002, 4. ed.

TURBAN, Efraim; RAINER, R. Kelly Jr.; POTTER, Richard E. **Administração de Tecnologia de Informação: Teoria e prática**. Rio de Janeiro: Campus, 2003.

WG SYSTEMS. Disponível em: <http://www.wgsystems.com.br/bi/dw.htm>. Acesso em: 18 de maio de 2005.