



**FACULDADE DE ADMINISTRAÇÃO E NEGÓCIOS DE SERGIPE -
FANESE
NÚCLEO DE PÓS-GRADUAÇÃO E EXTENSÃO – NPGE
CURSO DE PÓS-GRADUAÇÃO “LATO SENSU”
ESPECIALIZAÇÃO EM MBA EM ADMINISTRAÇÃO DE BANCO DE
DADOS**

PAULO ROBERTO DE SOUZA

**DATA WAREHOUSE:
UMA FERRAMENTA DE TECNOLOGIA DA INFORMAÇÃO PARA AMBIENTES
CORPORATIVOS**

ARACAJU/SE
2016

PAULO ROBERTO DE SOUZA

**DATA WAREHOUSE:
UMA FERRAMENTA DE TECNOLOGIA DA INFORMAÇÃO PARA AMBIENTES
CORPORATIVOS**

Artigo científico apresentado junto ao curso de MBA em Administração de Banco de Dados, como requisito à obtenção do título de profissional especialista em administração de banco de dados.

Orientação: Prof. Msc. Felipe Menezes
Cardoso

ARACAJU/SE
2016

**DATA WAREHOUSE:
UMA FERRAMENTA DE TECNOLOGIA DA INFORMAÇÃO PARA AMBIENTES
CORPORATIVOS**

Paulo Roberto de Souza¹

RESUMO:

Apresenta-se uma pesquisa qualitativa do tipo bibliográfico que aborda a temática: *Data Warehouse*, ou em português, Armazém de Dados, pode ser definido como um banco de dados preparado para integrar e gerenciar o fluxo de informações a partir dos bancos de dados corporativos e fontes de dados externas à empresa. Grande parte do trabalho no desenvolvimento de um DW está na análise dos sistemas transacionais e dos dados que neles 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 acaba reduzindo a possibilidade de erros na hora da decisão. Conhecer mais sobre essa tecnologia permitirá aos administradores descobrirem novas e diferentes maneiras de gerenciar 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, oportunidades atuando de diversas maneiras conforme o perfil de seus consumidores.

PALAVRAS CHAVE:

Data Warehouse; Tecnologia da Informação; Organizações.

INTRODUÇÃO

O grande crescimento do ambiente de negócios, médias e grandes empresas armazenam também um alto volume de informações, onde é importante um planejamento estratégico para obter sucesso nos dias de hoje. Com isso é

¹Faculdade de Administração de Negócios de Sergipe (FANESE) - Aracaju - SE – Brasil

necessário que a organização tenha capacidade de analisar, planejar e reagir, rápida e imediatamente às mudanças nas condições de seus negócios atingindo um maior destaque no mercado altamente competitivo, dispondo de mais e melhores informações que constituem a base destes processos.

Este manancial de informação, quando aproveitado de forma eficaz, desempenha um papel fundamental no sucesso das organizações, afinal vivemos numa sociedade tecnológica onde a informação acumulada é muito valiosa, sendo esta podendo determinar a eficiência da empresa e quais as melhores decisões devem ser tomadas.

DESENVOLVIMENTO

Diante da evolução da tecnologia da informação e o grande crescimento do uso de computadores, muitas as empresas de médio e grande porte utilizam sistemas de informação para otimizar seus principais processos. Ao longo do tempo estes sistemas tendem a gerar uma imensa quantidade de dados referentes ao negócio, mas não relacionados entre si, o que é normal que estes dados estejam dispersos por diversos locais e que tenham sido gerados por sistemas desenvolvidos em diferentes ambientes e linguagens.

Os bancos de dados são de fundamental importância para as empresas, pois neles estão contidos os dados armazenados que, de um modo geral, não servem para serem usados como um recurso estratégico em seu estado original.

Os sistemas tradicionais de informática não são projetados para gerar e armazenar as informações estratégicas, o que torna os dados desvalorizados para a tomada de decisões das organizações. Portanto, não se consegue buscar informações que permitam a tomada de decisão embasada em um histórico dos dados. As decisões geralmente são tomadas com base na experiência dos gestores quando poderiam também ser baseadas em fatos históricos, que foram armazenados pelos diversos sistemas de informação utilizados pelas organizações.

[FMC1] Comentário: Não se pode afirmar que praticamente todas, ou todas, ou quase nenhum, etc. sem indicar uma referência em que se prove o que está sendo dito. Por isso, é melhor dizendo muitas, algumas, etc.

O uso dos dados históricos pode identificar e posicionar a empresa estrategicamente para ser mais competitiva e, portanto, maximizar os lucros diminuindo o índice de erros na tomada de decisão.

A demanda de novos métodos e tecnologias surgiu da confirmação de que existe uma urgência de informação não atendida pelos sistemas comerciais convencionais, que atuam a nível operacional do negócio e pelo fato de que, a tecnologia de armazenamento de dados utilizada nestes sistemas não atende às necessidades detectadas. Por esse motivo difundiu-se uma nova concepção no mercado, o *Data Warehouse* (DW).

Este é um grupo de técnicas que utilizadas em conjunto geram um sistema de dados que disponibiliza informações para tomada de decisões e funciona tipicamente na arquitetura cliente/servidor. Um *data warehouse* é um grupo de dados orientado por assuntos, integrado, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão (Inmon 1993).

É um banco de dados que contém dados extraídos do ambiente de produção da organização, que foram selecionados e otimizados para processamento de consulta e não para processamento de transações. Em confronto com os bancos 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.

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. Este é construído para que tais dados possam ser armazenados e acessados de forma que não sejam delimitados por tabelas e linhas estritamente relacionais. É um modelo de dados multidimensional que armazena dados de forma integrada e disponibiliza séries no tempo e análises de vertentes através de banco de dados histórico.

É pertinente ressaltar que o DW está separado dos bancos de dados operacionais, portanto as consultas dos usuários não impactam nestes sistemas, que ficam protegidos de alterações indevidas ou perdas de dados. 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 vários 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 completos.

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.

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 um grupo de informações relativas à determinada área estratégica de uma empresa. Um exemplo típico é a área de vendas, como produtos, revendedores, contas e clientes.

Integração

Esta particularidade 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 no desenvolvimento de um DW está na análise dos sistemas transacionais e dos dados

que eles contêm. Esses dados, em geral, 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 clássico exemplo é o que se refere aos gêneros masculino e feminino. Em um 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. Podemos perceber que são as mesmas informações, mas estão em formatos diferentes, e isso em um DW não deverá acontecer nunca. Conseqüentemente é por isso que deverá existir uma integração de dados, padronizando-se uma maneira uniforme de armazenar os mesmos.

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 de o 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.

Entretanto no *data warehouse*, o principal objetivo é analisar o comportamento das mesmas durante um período de tempo maior e embasados nessa variação é que os gerentes tomam as decisões, então em um DW é contínuo manter 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 precisão é válida somente para o momento de acesso. Já os dados existentes num DW são como fotografias que refletem os mesmos em um determinado momento do tempo. Essas fotografias são chamadas de *snapshots*.

A dimensão tempo, sempre estará presente em qualquer fato de um DW, isso ocorre porque, como referido 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. Dessa forma pode-se dizer que os dados armazenados de forma correta no DW não serão facilmente atualizados tendo-se assim uma imagem exata 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 relacionadas ao caminho do dado dentro do DW.

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 blocos para o *data warehouse*, para que estejam disponíveis aos usuários para acesso. Isso pode ser certificado porque a maneira como os dados são carregados e tratados é completamente diferente dos sistemas transacionais. No ambiente operacional, os dados são, em geral, atualizados registro a registro, em múltiplas transações. Esta volatilidade requer um trabalho notável para assegurar a integridade e a consistência. Um *data warehouse* não requer este nível de controle típico dos sistemas orientados a transações, uma vez que o que acontece é somente a leitura dos dados na origem e gravá-los no destino, ou seja, no banco modelado multidimensional.

Tem que se considerar que os dados sempre passam por filtros antes de serem inseridos no DW. Dessa forma muitos deles jamais saem do ambiente transacional, e outros são tão resumidos que não se encontram fora do DW. 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 um por cento de duplicações, essa definição é dada por [Inmon \(1993\)](#).

[FMC2] Comentário: Inmon está referenciado como Inmon, 1996, mas a referência que você adicionou é do ano de 1993. Favor verificar a referência.

Localização

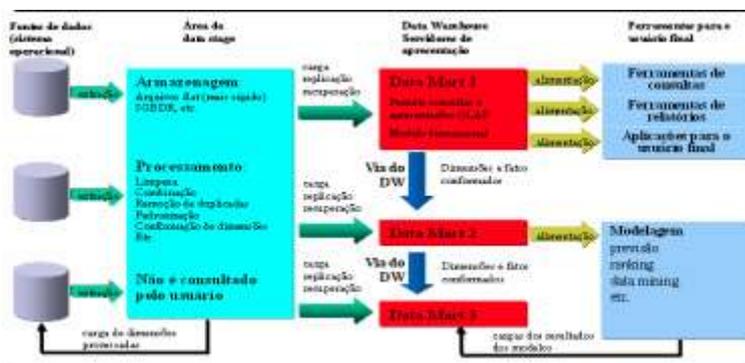
Os dados podem estar fisicamente armazenados de três formas:

- Em um único local centralizado, um banco de dados em um DW integrado, procura-se potencializar o poder de processamento e acelerar a busca dos dados. Esse tipo de armazenagem é bastante utilizado, nada obstante há o inconveniente do investimento em hardware para acomodar a base de dados muito volumosa e o poder de processamento elevado para atender as consultas simultâneas de muitos usuários.
- Distribuídos, em forma de *Data Marts*, armazenados por áreas de interesse. Por exemplo, os dados da gerência financeira em um servidor, dados de marketing em outro e dados da contabilidade em um 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, com um nível de detalhe intermediário no segundo servidor e os dados mais detalhados (atômicos), em um 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.

Elementos básicos do DW

Na figura 1| abaixo podemos visualizar os elementos básicos que compõem as arquiteturas de um *Data Warehouse*.

[FMC3] Comentário: Adicionar referência à figura abaixo, depois que for adicionada a legenda.



[FMC4] Comentário: Adicionar legenda na Figura

Figura 1 – Elementos DW

Fonte de dados

Sistemas transacionais da empresa, podem ser compostos por diversas formas de dados.

Data Stage

Constituída por uma área de armazenagem e um conjunto de processos. Sua função é extrair os dados dos sistemas transacionais, proceder a limpeza, a transformação, combinação, de duplicação e preparação dos dados para o uso no DW. Estes dados não são apresentados ao usuário final.

Servidor de apresentação

Ambiente onde os dados são organizados e armazenados para consulta direta pelos usuários finais. Normalmente os dados estão disponíveis nestes servidores em bancos de dados relacionais, mas também podem estar armazenados em tecnologia OLAP (*Online Analytical Processing*) já que muitos *data marts* trabalham apenas com dados no modelo dimensional.

Data Mart

Subconjunto lógico do DW, geralmente divididos por departamento ou visões necessárias para os usuários.

Data Mining

Similarmente conhecido como mineração de dados, o *Data Mining* trabalha com grandes massas de dados onde existem muitas correlações entre os dados que não são perceptíveis facilmente. Os *Data Warehouses* são constituídos, normalmente, de imensa quantidade de dados e existe há necessidade de uma ferramenta para varrer automaticamente o DW a fim de pesquisar tendências e padrões através de regras pré-definidas que dificilmente seriam encontrados em uma pesquisa comum.

Ferramentas de acesso aos dados

A maneira em que os dados são extraídos e integrados com cada processo distinto do DW. As funções para a transformação dos dados são:

- Extração: remoção dos dados dos sistemas transacionais e armazenagem na área de *data stage*;
- Carga de dimensões processadas: realimentação do processo para garantir a representação correta dos dados em novo formato.
- Carga, Replicação e Recuperação: quando pronto, o dado é carregado no *data mart* correspondente e são criados (ou atualizados) índices para melhorar a performance das consultas.
- Alimentação: apresenta as visões do *data mart* de acordo com as necessidades dos usuários.
- Carga dos resultados dos modelos: serve para realimentar possíveis modificações no *data mart*, caso este não esteja adequado a aplicação que o utiliza.

Arquitetura do DW

A arquitetura do DW pode variar de acordo com o tipo de assunto abordado, isso ocorre em virtude das necessidades que variam de empresa para empresa.

Arquitetura Genérica

A arquitetura genérica compreende a camada dos dados operacionais que serão acessados pela camada de acesso a dados. As camadas de gerenciamento de processos, transporte e *data warehouse* são responsáveis por distribuir os dados e estão no centro da arquitetura. A camada de acesso à informação possibilita a extração das informações do DW utilizando um conjunto de ferramentas.

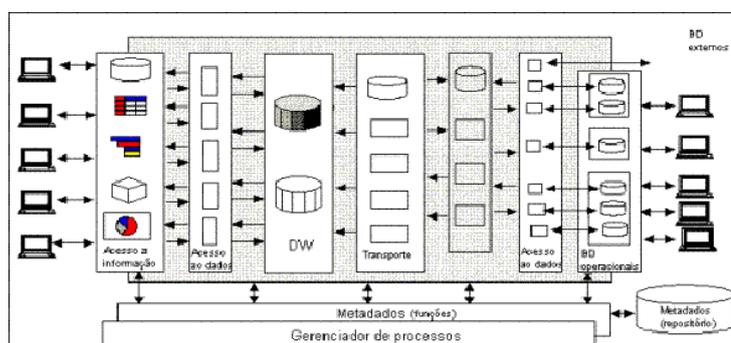


Figura 2 - Arquitetura genérica DW

Arquitetura de Duas Camadas

A arquitetura de duas camadas utiliza um servidor em conjunto com aplicações *front-end*, que são ferramentas que realizam operações sobre os dados consultados e os transformam em informações úteis para os usuários, os componentes *back-end* são ferramentas responsáveis pela extração, limpeza e cargas dos dados, mais conhecidas como ETL e também são utilizadas neste tipo de arquitetura.

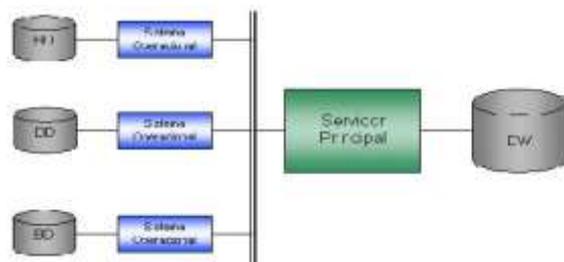


Figura 3 - Arquitetura duas camadas DW

Arquitetura de Três Camadas

A arquitetura de três camadas suporta vários usuários e serviços devido a sua maleabilidade, as informações ficam armazenadas em várias camadas. Na primeira camada estão as interfaces que trabalham com o usuário, onde geralmente são gráficas. Na segunda camada estão os servidores de banco de dados e aplicações e, por isso, têm a necessidade de ter um acesso eficiente e rápido aos dados compartilhados, e na última ficam armazenadas as fontes de dados. A arquitetura de três camadas é a mais utilizada pelos analistas.

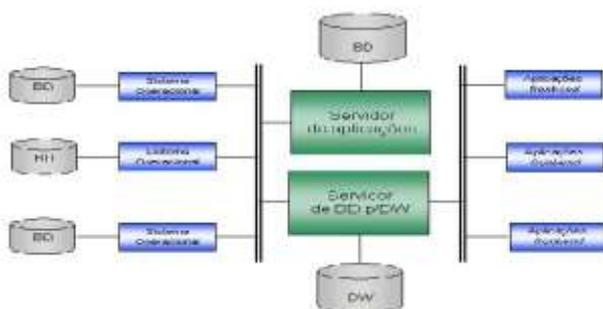


Figura 4 - Arquitetura três camadas DW

Credibilidade dos Dados

A credibilidade dos dados é muito importante para o êxito 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 visivelmente 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. Grupos de dados, processos de entrada, metadados e informações sobre a origem dos dados, são de extrema importância.

Outras questões como a manutenção e a atualização dos dados, as nuances entre os dados para bancos de dados transacionais e para o *Data Warehouse* 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 precisão.

Granularidade

A granularidade é o nível de detalhe ou de resumo dos dados existentes em um 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 proporcional 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 um grupo de dados.

Os dados superficialmente 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 alterações. Por exemplo, se as informações nos dados detalhados atuais são armazenadas por dia, nos dados superficialmente 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 acabam sofrendo 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 um padrão.

O chamado nível duplo de granularidade se enquadra nos requisitos da maioria das organizações. 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.

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 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.

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 banco 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 é onde se deve dar preferência aos 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.

A próxima etapa é fazer as rotinas de extração dos dados, estas 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. Estes 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.

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 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 aquelas 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 detectada 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 estes 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.

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. A tecnologia OLAP permite que o usuário trabalhe com cálculos complexos, através de consultas com maior flexibilidade e funcionalidade.

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 e 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.

CONCLUSÃO

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 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. Um conjunto de ferramentas são necessárias para que se possa desfrutar de suas benefícios 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.

SUMMARY:

It presents a qualitative research of bibliographical that addresses the theme: Data Warehouse, or in Portuguese, Armazém de Dados, can be defined as a database prepared to integrate and manage the flow of information from corporate databases and external data sources to the company. Much of the work in developing a DW is the analysis of the transactional systems and data that contain them. A Data Warehouse allows the generation of integrated and historical data helping the directors to decide grounded in facts, not hunches or speculations, which ends up reducing the possibility of errors at the time of decision. To know more about this technology allows administrators to discover new and different ways to manage your business in a globalized economy, making them safer to define the goals and adopt different strategies in your organization, thereby visualize before your competitors, opportunities working in various ways according to the profile of its consumers.

KEY WORDS:

Data Warehouse; Information Technology; Organizations.

REFERÊNCIAS BIBLIOGRÁFICAS

Data warehouse. Disponível em: <http://www.devmedia.com.br/data-warehouse/12609>. Acesso em: 15 de maio de 2016.

Um estudo sobre as ferramentas OLAP. Disponível em: <http://www.devmedia.com.br/um-estudo-sobre-as-ferramentas-olap/6691>. Acesso em: 30 de junho de 2016.

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

INMON, W.H. **Information System Arquiteure: Development in 90's.** Nova York: John Wiley& Sons Inc., 1993.

INMON, W.H. **Building the Data Warehouse.** Nova York: John Wiley& Sons Inc., 1993.

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 de 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.