

A UTILIZAÇÃO DO DATA WAREHOUSE NA GERAÇÃO DE INFORMAÇÕES PARA TOMADA DE DECISÃO

Elias Dib Caddah Neto

Roberta Carvalho de Alencar

Resumo:

A necessidade de informações para tomada de decisão é crescente entre as empresas. Os diversos sistemas operacionais existentes em uma empresa produzem uma considerável quantidade de dados acerca das transações efetuadas, e é necessário que se faça a integração destes dados, de forma a fornecer ao tomador da decisão uma forma flexível e eficiente de se obter as informações necessárias aos seus processos decisórios. O presente trabalho tem por objetivo investigar, por meio de pesquisa bibliográfica, a viabilidade e a aplicabilidade do Data Warehouse como instrumento capaz de promover tal integração, com a eficácia desejada. O uso do Data Warehouse, auxiliado por ferramentas inteligentes de recuperação de dados do tipo Data Mining, mostrou-se adequado aos propósitos desejados.

Palavras-chave:

Área temática: *Gestão de Custos e Tecnologia da Informação*

A UTILIZAÇÃO DO *DATA WAREHOUSE* NA GERAÇÃO DE INFORMAÇÕES PARA TOMADA DE DECISÃO

RESUMO

Elias Dib Caddah Neto

Centro de Ensino Superior do Vale do Parnaíba

eliascaddah@uol.com.br

Roberta Carvalho de Alencar

Universidade de Fortaleza

A necessidade de informações para tomada de decisão é crescente entre as empresas. Os diversos sistemas operacionais existentes em uma empresa produzem uma considerável quantidade de dados acerca das transações efetuadas, e é necessário que se faça a integração destes dados, de forma a fornecer ao tomador da decisão uma forma flexível e eficiente de se obter as informações necessárias aos seus processos decisórios. O presente trabalho tem por objetivo investigar, por meio de pesquisa bibliográfica, a viabilidade e a aplicabilidade do *Data Warehouse* como instrumento capaz de promover tal integração, com a eficácia desejada. O uso do *Data Warehouse*, auxiliado por ferramentas inteligentes de recuperação de dados do tipo *Data Mining*, mostrou-se adequado aos propósitos desejados.

ÁREA TEMÁTICA

Gestão de Custos e Tecnologia da Informação

A UTILIZAÇÃO DO DATA WAREHOUSE NA GERAÇÃO DE INFORMAÇÕES PARA TOMADA DE DECISÃO

Introdução

Um fator de grande importância para garantir a continuidade das empresas nos dias atuais é sua capacidade de analisar, planejar e reagir com rapidez às mudanças ocorridas nos seus negócios. Para isso, é fundamental que a organização disponha cada vez mais de informações para o auxílio à tomada de decisões.

Tais informações devem ser adequadas ao tipo de decisão a ser tomada em cada área da organização, bem como nível de decisão em que o gestor se encontra no organograma (estratégico, tático, operacional).

Os sistemas de processamento de transações (sistemas de controle de estoques, de faturamento, de vendas, de contabilidade, de folha de pagamento, etc) geram uma enorme quantidade de dados, e há em tais sistemas toda a sorte de detalhes acerca das transações efetuadas.

Estes sistemas muitas vezes funcionam de forma isolada, com o objetivo de único de dar apoio às atividades específicas dos gestores que trabalham no nível operacional da organização.

Em razão da não integração dos diversos sistemas, muito da riqueza da informação que poderia ser produzida por meio do cruzamento de dados dos vários sistemas, e vista de forma sumarizada e flexível pelos níveis táticos e estratégicos, se perde.

Neste contexto o conceito de *Data Warehouse*, um grande “armazém de dados” onde os dados dos mais diversos sistemas podem ser integrados e requisitados pelos usuários da organização de forma flexível, com a ajuda de ferramentas de busca, apresenta-se como uma interessante opção para melhorar o nível das informações fornecidas aos níveis gerenciais mais altos da organização.

Este trabalho tem por finalidade, sem a intenção de esgotar o assunto, discutir os principais conceitos de sistemas de informação e *Data Warehouse*, trabalhando sua importância, características, arquitetura, desenvolvimento e seu relacionamento com o sistema de informações, com o intuito de despertar o interesse para esta área com enormes perspectivas de desenvolvimento.

1. Visão sistêmica da empresa e subsistema de informação

A empresa pode ser vista como um sistema, ou seja, um “conjunto de elementos interdependentes, ou um todo organizado, ou partes que interagem formando um todo unitário e complexo” utilizando-se o conceito enunciado por Bio (1985).

O todo, que é a empresa, pode ser decomposto em partes, ou subsistemas, e existem várias formas de efetuar tal decomposição. Peleias (2002) decompõe a empresa nos seguintes subsistemas:

- Subsistema Institucional – composto da missão, crenças e valores da instituição;
- Subsistema de Gestão – compreende o processo de gestão, desdobrado nas etapas de planejamento, execução e controle;
- Subsistema de Organização - reflete a estrutura organizacional adotada e a delegação de autoridade resultante;
- Subsistema Psico-sócio-político-cultural – reflete o comportamento dos indivíduos da organização, considerando seus fatores pessoais e comportamentais;
- Subsistema de produção, físico-operacional ou físico - onde ocorrem às transações; e

- Subsistema de informação – instrumento de suporte ao processo de gestão concebido à luz das diretrizes e políticas emanadas desse processo.

Para Peleias (2002;p.11), o subsistema de informação *“reflete os modelos decisórios dos vários gestores da empresa, e sua concepção leva em conta como e quando a informação deve chegar aos usuários, para quais finalidades é utilizada, quais conceitos e critérios a suportam e os canais utilizados para sua comunicação, buscando a unidade na linguagem organizacional”*

Campos Filho citado por Beuren e Martins (2001;p.9) identifica quatro componentes reunidos no sistema de informação de modo a permitir o melhor atendimento aos objetivos da organização:

- A informação (dados formatados, imagens, sons e textos livres);
- Os recursos humanos (que coletam, armazenam, recuperam, processam, disseminam e utilizam as informações);
- As tecnologias de informação (o hardware e o software);
- As práticas de trabalho (métodos utilizados).

A informação adequada aos objetivos depende, portanto, não somente da disponibilidade de dados, mas também de como estes são coletados, armazenados e interpretados, e neste processo temos tecnologia e pessoas envolvidas.

2. Classificação dos sistemas de informação

Arima (2002;p.83) classifica os sistemas de informações de acordo com o nível gerencial a que se destinam: na base da pirâmide estão os sistemas de informações operacionais, seguidos pelos sistemas de informações gerenciais e no topo da pirâmide os sistemas de informações estratégicas.

Os sistemas de informações operacionais são os sistemas destinados a dar suporte ao próprio funcionamento físico-operacional das diversas áreas da empresa. Como exemplo temos os sistemas de faturamento, estoques, contas a pagar, contas a receber, custos, contabilidade, etc.

São características dos sistemas de informações operacionais, segundo Arima (2002;p.83):

- São estruturados quanto as suas decisões;
- São padronizados e repetitivos quanto ao seu funcionamento;
- Mantém garantia na exatidão dos dados;
- Têm relativa falta de flexibilidade na geração de informações;
- São baseados, normalmente, em dados internos da organização; geram relatórios de controle operacional e, em alguns casos, até gerencial.

Os sistemas de informações gerenciais já apresentam um grau mais elaborado de informação. Como exemplo, Arima (2002;p.84) cita os sistemas de projeção de vendas, controle de produção, análise de custos, etc. e elenca as seguintes características:

- São semi-estruturados em termos de tomadas de decisões;
- São customizados, isto é, ajustados às necessidades das áreas funcionais como vendas, produção, finanças, etc., podendo ser ou não repetitivos;
- Ainda são pouco flexíveis na geração de informações;
- Utilizam projeções, modelos e informações subjetivas;
- Permitem consultas diversas;
- Têm pouca ou nenhuma entrada de dados;
- São integrados às funções do negócio;

- Baseiam-se em dados internos e externos da organização;
 - Geram informações tanto analítica quanto sintéticas, bem como projeções.
- O sistema de informações estratégicas destina-se aos altos escalões da organização. Algumas das características deste sistema segundo Arima (2002;p.85):
- Têm a tendência de não estruturar problemas específicos voltados à gerência de alto nível;
 - Utilizam modelos e técnicas analíticas em conjunto com métodos tradicionais de acesso e recuperação de dados;
 - São orientados para dispositivos que facilitem o uso de computador por pessoas não especializadas e estranhas ao ambiente de tecnologia da informação, permitindo que o processo decisório seja realizado de forma simples e rápida;
 - Dão maior ênfase à flexibilidade e à adaptabilidade a mudanças do ambiente e do processo decisório do usuário, lidando com situações diferentes;
 - Devem ser individualizados para cada tomador de decisão;
 - Devem ser comunicativos, sendo que a qualidade do diálogo passa a ser um fator preponderante no uso do respectivo sistema;
 - Devem ser interativos e responder a questões do tipo “o que – se” (*what if*) e análise de lucratividade (*trade-off*).

Para Beuren e Martins (2001;p.22) os sistemas de informação são classificados em dois grandes blocos: sistemas de apoio às operações e sistemas de apoio à gestão. No primeiro grupo enquadram-se os sistemas de automação (SA), cuja função é automatizar processos onde haja a mínima intervenção do ser humano e os sistemas de processamento de transações (SPT), que processam e registram as transações realizadas na organização.

O segundo bloco, ainda conforme os mesmos autores, contempla os níveis gerenciais e estratégicos da organização, comportando:

- O sistema de informações gerenciais (SIG) que possibilita ao administrador visualizar as atividades realizadas na sua área funcional, por meio de apontamentos resumidos e sumarizados;
- O sistema de apoio à decisão (SAD), o qual oferece subsídios relevantes que permitem a escolha de alternativas para decisões específicas;
- O sistema de informações executivas (SIE) que proporciona ao executivo o monitoramento constante dos fatores críticos de sucesso do negócio, por meio de status indicadores de desempenho, e auxilia na exploração de dados, permitindo identificação de problemas e oportunidades para o cumprimento da missão da organização.

Beuren e Martins (2002; p.23) afirmam ainda que um sistema de informação voltado para os gestores responsáveis por decisões estratégicas deverá ter as seguintes características:

- A partir de dados sumarizados, disponibilizar os detalhes dos mesmos (*drill down*);
- Acesso a informações externas ao ambiente da empresa;
- Utilização de recursos gráficos na apresentação dos resultados;
- Permitir a recuperação rápida da informação;
- Facilitar o uso operacional do sistema;
- Estar voltado ao acompanhamento dos fatores críticos de sucesso através dos indicadores de desempenho, e;
- Realizar a triagem de informações filtrando-as e resumindo-as.

3.Data Warehouse

Traduzindo literalmente para o português, a expressão *Data Warehouse* significa “armazém de dados”.

De acordo com Inmon (1997:p.33), *“um Data Warehouse é uma coleção de dados orientada por assuntos, integrada, variante no tempo e não volátil, que objetiva dar suporte aos processos de tomada de decisão.”*

Para Hcakothon apud Campos e Rocha Filho (1999), o *Data Warehouse* tem como objetivo fornecer uma *“imagem única da realidade do negócio”*.

Gray e Watson apud Pereira (2000:p.14), afirmam que um *“Data Warehouse é tipicamente um sistema de banco de dados dedicado que é separado dos sistemas OLTP da organização.”*

Segundo Kimball apud Pereira (2000:p.15), o *Data Warehouse* é o *“lugar onde as pessoas podem acessar seus dados”*.

Para Gardner apud Pereira (2000:p.15) *Data Warehouse “é um processo, não um produto, para montagem e administração de dados provenientes de várias fontes com o propósito de obter uma simples e detalhada visão da parte do todo do negócio.”*

Na mesma linha,Berson e Smith (1997: p.5) afirmam que o *Data Warehouse* é *“um ambiente, não um produto”*, que fornece aos usuários *“informações correntes e históricas para suporte à decisão que são difíceis de apresentar ou de acessar nos arquivos tradicionais de dados operacionais.”*

Existem diferentes visões sobre o *Data Warehouse*, onde o mesmo é considerado como um conjunto de dados, uma arquitetura, um processo em constante evolução etc.

De uma maneira geral um *Data Warehouse* contém:

- 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 os dados a seus usuários.

Um *Data Warehouse* é, portanto um sistema que coleta dados existentes nos mais diversos formatos, junto a diferentes sistemas dentro da organização, tornando-os compatíveis entre si, e fornecendo-os aos diversos usuários interessados nestes dados.

2.1 Características

Como características do *Data Warehouse* podemos citar as seguintes:

2.1.1 Orientado por assuntos

O *Data Warehouse* armazena informações sobre assuntos/temas específicos. Assuntos são um conjunto de informações relativas a determinada área estratégica de uma empresa. Os dados são organizados na maneira como são referidos pelos usuários. Como exemplo, em uma revendedora de carros, temos áreas de marketing, financeiras etc, e dentro destas áreas existem vários assuntos como vendas e serviços.

2.1.2 Integrado

Refere-se à consistência dos dados, que são convertidos para um estado uniforme, desfazendo-se as inconsistências. É uma das características mais importantes do *Data Warehouse*, pois através dela se padroniza uma única representação para os

dados de todos os sistemas que irão fazer parte da base de dados do *Data Warehouse*.

Por exemplo, ao considerar sexo como um elemento de dado, em uma aplicação ele pode ser codificado como M/F, em outra como H/M, ou em 1/0. Ao se encaminhar estes dados para o *Data Warehouse*, são convertidos para um estado uniforme, sendo o dado sexo codificado apenas de uma forma.

3.1.3. Não Volátil

Significa que os dados do *Data Warehouse* são somente de leitura, não podendo ser atualizados pelos usuários.

A atualização dos dados é feita no ambiente operacional. Estas operações requerem um considerável trabalho no que diz respeito a manter a integridade e consistência, aspecto que não ocorre em um *Data Warehouse*.

Como exemplo, em um sistema de contabilidade, as alterações feitas geram atualizações em dados relativos aos saldos das contas. O mesmo ocorre com um sistema de Contas e Receber, onde os saldos de valores a receber de cada cliente sofrem mudanças à medida que novas transações de vendas e de recebimentos de duplicatas são processadas. Já no *Data Warehouse*, os dados são lidos na origem e gravados no destino, de forma cumulativa, ou seja, não sofrem mudanças decorrentes da inclusão de dados oriundos de novas transações.

3.1.4 Variante no tempo

Diz respeito ao fato do dado em um *Data Warehouse* ser histórico, isto é, refere-se a um momento específico, não sendo, portanto, atualizável. Num ambiente operacional, as informações são fornecidas em um dado momento, já no *Data Warehouse* o objetivo é analisar o comportamento das mesmas durante um período de tempo maior. A atualização do dado de produção ocorre de acordo com a mudança de estado do objeto em questão, em um momento de acesso. A cada ocorrência de mudança, uma nova entrada para marcar esta mudança é criada em um *Data Warehouse*.

3.2 Data Mart

Um *Data Warehouse* como já foi visto, engloba os dados da organização como um todo. Quando estes dados estão restritos a uma área de negócio ou unidade administrativa da organização em particular, como, por exemplo, marketing, vendas, finanças etc., chama-se a este subconjunto do *Data Warehouse* de *Data Mart*.

Um *Data Mart* tem todas as características de um *Data Warehouse*, mas armazena dados sobre apenas parte da empresa. Muitas vezes os *Data Mart* são utilizadas como estratégia para se alcançar um objetivo maior, que é chegar a um *Data Warehouse* corporativo. Dentro desta estratégia, são implantados diversos *Data Mart* que são depois interligados para formar o *Data Warehouse*.

Berson e Smith (1997: p.125) alertam para o fato de que estes *Data Mart* tem que ser cuidadosamente definidos, de modo a preservar a integração de dados, que é o coração do *Data Warehouse*. Para tanto, a estratégia de desenvolvimento deve obedecer a regras definidas para a organização como um todo.

3.3 Arquitetura

O estudo da arquitetura de um *Data Warehouse* compreende a arquitetura funcional e a arquitetura dos dados.

3.3.1 Arquitetura Funcional do *Data Warehouse*

Conforme já mencionado anteriormente, de uma forma genérica pode-se dizer que um *Data Warehouse* é composto de estruturas de captação e tratamento dos dados operacionais da empresa; armazenamento de dados e apresentação destes aos usuários.

Ligados a todas estas estruturas estão os metadados, que são informações acerca dos dados trabalhados no *Data Warehouse*, e o processo de gerenciamento e controle.

Campos e Rocha Filho (1999) citam Orr, que propõe uma arquitetura genérica que sistematiza os diversos papéis no ambiente de *Data Warehouse*, chamando-os de camadas:

- **Camada de Bancos de Dados Operacionais e Fontes Externas** – São os dados oriundos dos sistemas operacionais da empresa e de outras fontes externas, que serão tratados e integrados para que possam compor o *Data Warehouse*.
- **Camada de Acesso à Informação** – Corresponde à camada de interação com os usuários finais.
- **Camada de Acesso aos Dados** – Esta camada efetua a ligação entre as ferramentas de acesso à informação e os bancos de dados operacionais.
- **Camada de Metadados (Dicionário de Dados)** – Os Metadados são informações sobre os dados mantidos pela empresa, ou “dados sobre os dados”.
- **Camada de Gerenciamento de Processos** – Esta camada é a responsável pelo gerenciamento do processo de construção e manutenção das informações do dicionário de dados e do *Data Warehouse*.
- **Camada de Transporte ou *Middleware*** - Esta camada gerencia o transporte de informações pelo ambiente de redes.
- **Camada do *Data Warehouse*** – Os dados utilizados para produção de informações, compõem a camada do *Data Warehouse* propriamente dito. Pode envolver ou não o armazenamento físico de dados.
- **Camada de Gerenciamento de Replicação** - Esta camada refere-se aos processos realizados para selecionar, editar, resumir, combinar e carregar o *Data Warehouse* e as correspondentes informações de acesso a partir das bases operacionais e fontes externas.

3.3.2 Arquitetura dos Dados

A arquitetura dos dados em um *Data Warehouse* refere-se à forma como estes dados são armazenados e recuperados/apresentados ao usuário final.

Campos e Rocha Filho (1999) apresentam três tipos de arquitetura. O primeiro tipo de arquitetura consolida os dados em um *Data Warehouse* único e centralizado. Este tipo de arquitetura procura otimizar a capacidade de processamento disponível. Uma segunda opção de arquitetura distribui os dados setorialmente, em diversos *Data Marts*, integrados ao *Data Warehouse*.

Por fim, os autores apresentam uma arquitetura de dados por camadas, em que dados extremamente resumidos são disponibilizados em um servidor, dados em nível de detalhe intermediário são armazenados em um outro servidor, e por último, os dados no menor nível de detalhe (dados atômicos) são inseridos em um terceiro servidor. Naturalmente, os dados aqui referidos são os dados já extraídos e tratados a partir dos dados operacionais da empresa.

3.3.3. Os Metadados

Os metadados são de grande importância para o processo de controle das operações em um *Data Warehouse*. Os metadados são dados acerca dos dados constantes no *Data Warehouse*. Durante todas as fases do projeto de um *Data Warehouse*, e também após o início de sua operacionalização, metadados devem ser armazenados. Existem no mercado ferramentas próprias para armazenar e gerenciar metadados, dos quais o *Sybase Warehouse Control Center* e o *Prism Warehouse Directory* são exemplos.

Segundo Inmon (1997), os metadados são informações sobre “o que está aonde” no *Data Warehouse*. Os aspectos sobre os quais os metadados mantêm informações são:

- “A estrutura dos dados, segundo a visão do programador;
- A estrutura dos dados segundo a visão dos analistas de sistemas de apoio à decisão;
- A fonte de dados que alimenta o *Data Warehouse*;
- A transformação sofrida pelos dados no momento de sua migração para o *Data Warehouse*;
- O modelo de dados;
- O relacionamento entre o modelo de dados e o *Data Warehouse*;
- O histórico das extrações de dados.”

3.4 Bancos de Dados

Alguns dos aspectos que devem ser levados em conta ao se modelar um banco de dados para um *Data Warehouse* são descritos neste tópico.

3.4.1 .Modelagem Dimensional

Um *Data Warehouse* requer um banco de dados que facilite a análise multidimensional, ou seja, um mesmo assunto pode ser explorado de diferentes pontos de vista.

Informações sobre vendas, por exemplo, pode ser analisada distribuindo-se os dados por período, tipo de produto, região, vendedor, porte do cliente etc. Cada aspecto considerado é uma dimensão.

Em bancos de dados relacionais, as principais modalidades de representação e armazenamento de dados multidimensionais são os esquemas do tipo “estrela” (*star schema*) e o esquema “flocos de neve” (*snowflake schema*).

No esquema estrela há uma tabela principal (denominada Tabela Fato) ao qual ligam-se outras tabelas, referentes a dimensões.

No esquema flocos de neve, ocorre uma extensão do esquema estrela, onde as tabelas de dimensão são desdobradas em outras dimensões.

3.4.2 . Granularidade

Granularidade é o nível de resumo ou de detalhe dos dados existentes em um *Data Warehouse*. Quanto maior a granularidade, menor o nível de detalhe.

Por exemplo, os dados operacionais de vendas podem ser armazenados no *Data Warehouse* transação a transação, ou serem submetidos a algum tipo de sumarização (por cliente, data, praça, vendedor etc.).

O nível de granularidade afeta o volume de dados armazenados no *Data Warehouse* bem como o tipo de consulta que pode ser respondida. Com um nível de granularidade muito baixo, o *Data Warehouse* tem a capacidade de responder a

grande maioria das consultas, mas irá requerer uma grande capacidade de armazenamento, e tornará as consultas mais lentas. Dificilmente, no entanto, são examinados eventos isolados, de forma que os dados operacionais podem, e devem, ser submetidos a algum tipo de sumarização.

Um nível de granularidade muito alto permite rapidez na resposta e consumirá pouco espaço, mas deixará muitas questões sem resposta.

É preciso encontrar o equilíbrio entre os dois extremos, e esta questão deve receber especial atenção durante o desenvolvimento de um *Data Warehouse*.

Inmon (1997:p.49), destaca que na maioria das vezes os usuários desejam as duas coisas: eficiência no armazenamento de dados e no acesso a eles, bem como possibilidade de analisar dados em maior detalhe.

A solução proposta por Inmon (1997:p.50) é ter dois níveis de granularidade no *Data Warehouse*: uma camada de dados para pesquisas mais imediatas com dados levemente resumidos, e uma outra camada com dados históricos detalhados.

3.4.3 . Particionamento dos Dados

O particionamento de dados é uma repartição dos dados em unidades físicas separadas. Estas unidades físicas podem ser tratadas independentemente. No *Data Warehouse*, na visão de Inmon (1997:p.57) a questão não é se é necessário fazer o particionamento dos dados, mas como ele deve ser feito.

A afirmativa de Inmon é coerente, pois em um *Data Warehouse* um grande volume de dados é processado. O tratamento destes dados em servidores separados permite uma maior eficiência no gerenciamento do banco de dados.

Os dados podem ser particionados de várias maneiras, como por exemplo: por data, área de negócio, unidade organizacional etc.

É bastante comum se pensar que caso a granularidade e o particionamento forem corretamente executados, todos os outros aspectos do projeto e implementação do *Data Warehouse* irão surgir naturalmente. Porém, caso a granularidade não seja apropriadamente conduzida e o particionamento não for projetado e implementado com bastante cuidado, nenhum outro aspecto do projeto terá importância.

3.5 Desenvolvendo um Data Warehouse

Com relação ao desenvolvimento de um *Data Warehouse*, é difícil encontrarmos uma metodologia consolidada e aceita. Na literatura e na prática o que se vê são sucessos na implementação do *Data Warehouse* em empresas.

Uma importante visão do projeto de desenvolvimento de um *Data Warehouse* é o fato de que ele deve ser iniciado a pedido de outras áreas que não a de informática, como a controladoria, finanças, marketing etc, pois estas justificam o projeto, tendo condições de avaliar seus benefícios, tais como o aumento da rentabilidade com sua implantação. Um projeto de iniciativa exclusiva da área de informática tem grandes chances de ser fadado ao fracasso.

Por ser considerado um dos trabalhos mais completos, detalhando as diversas fases da metodologia do desenvolvimento, a referência utilizada neste trabalho será a de Kimball.

Na metodologia de desenvolvimento de um *Data Warehouse*, representada na figura 1, observa-se à existência de várias fases, sendo algumas delas executadas concomitantemente com outras.

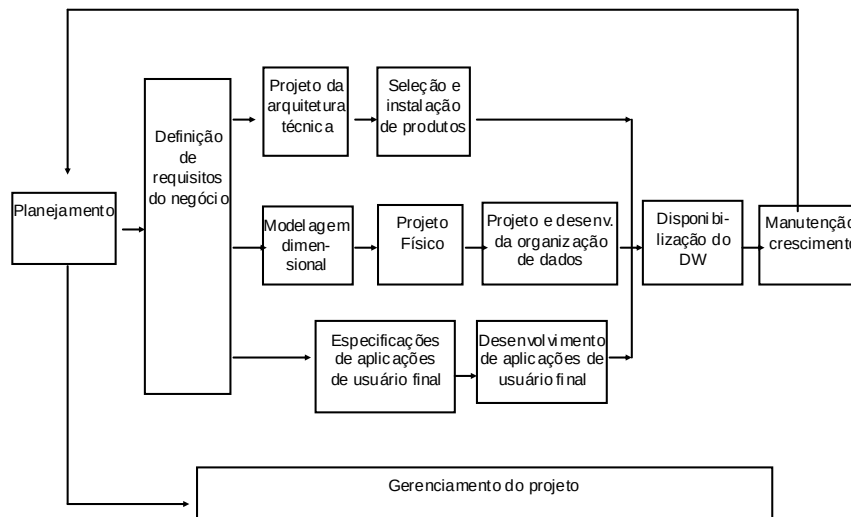


Figura 1: Desenvolvimento de um Data Warehouse de Kimball.

Fonte: Pereira (2000:p.51)

O desenvolvimento dá início com a fase de *planejamento*. Esta é uma das fases mais importantes e se deve ser o mais criterioso possível, pois todo o projeto estará envolvido.

As atividades principais desta fase são as seguintes:

- Definição do projeto, através da verificação da demanda real por informação gerencial e onde ela se localiza;
- Definição do escopo inicial do projeto, que pode ser tão amplo a ponto de incluir todo o conjunto de informações de uma empresa, ou tão restrito quanto de *Data Warehouse* de um único gerente. Quanto maior for o escopo, mais valor *Data Warehouse* tem para a empresa, sendo mais caro e trabalhosa sua criação e manutenção;
- Levantar os fatores que permitem a mensuração do sucesso, bem como os riscos do projeto;
- Elaboração de justificativas para o projeto, abordando o investimento e custos, sem esquecer o retorno financeiro e seus benefícios;

- Elaborar um Plano de Projeto contendo o nome do projeto, identificação do pessoal participante e correspondentes tarefas, atribuições e prazos a serem cumpridos etc.

A fase seguinte corresponde à *definição de requisitos do negócio*, onde é observada a importância do entendimento do negócio do usuário final, bem como suas particularidades, requisitos e exigências. Estes dados podem ser levantados através de entrevistas, análise de documentos e relatórios etc.

Algumas atividades são realizadas nesta fase, dentre as quais destacamos a identificação e preparação da equipe de entrevistadores, agendamento das entrevistas e suas realizações e análise de resultados. Nesta etapa, devem ainda ser revisados o escopo do projeto e priorização anteriormente definidos, bem como o Plano de Projeto.

A fase de *projeto de arquitetura técnica* é uma das mais importantes, pois é nesta fase que acontecem o estabelecimento de uma estrutura de arquitetura de alto nível, onde são abordados elementos fundamentais para criação de um *Data Warehouse* e a especificação da infraestrutura técnica e os componentes necessários para a criação de um *Data Warehouse* de acordo com as especificações na arquitetura técnica.

As atividades realizadas nesta fase envolvem a especificação de um grupo responsável pela criação da arquitetura *Data Warehouse*; reunião e documentação dos requisitos técnicos, revisão do ambiente, elaboração de uma arquitetura e determinação das etapas a serem cumpridas; aceitação do projeto pelo usuário final e revisão do projeto.

A fase de *seleção e instalação dos produtos*, só deverá ser executada depois que o projeto de arquitetura técnica tiver sido criado. As atividades realizadas nesta fase envolvem a pesquisa de produtos; avaliação das funcionalidades dos produtos, sob vários pontos de vista; desenvolvimento do protótipo em um ambiente controlado, permitindo com isso uma melhor avaliação das funcionalidades; e seleção de ferramentas e produtos adequados e estabelecimento de contrato.

Na fase de *modelagem dimensional*, são utilizados para o desenvolvimento de um modelo de dados adequado, os requisitos do negócio e os dados necessários ao atendimento das exigências analíticas. Algumas das atividades desenvolvidas nesta fase são as seguintes:

- Definição do *Data Mart* que será desenvolvido e a granularidade dos dados;
- Revisão da modelagem dimensional com o usuário final e aceitação do modelo de dados;
- Atualização do Projeto Lógico da base de dados analítica;
- Revisão lógica do projeto da base de dados analítica;
- Aceitação do projeto final pelo usuário.

Nesta fase, ainda são feitas as análises de diversas fontes de dados, permitindo com isso a identificação daquelas que possuem os dados necessários ao atendimento do modelo de dados.

Na fase do *projeto físico*, são desenvolvidas as atividades de definição de nomes padronizados para os objetos da base de dados, criação da base analítica e desenvolvimento de um plano inicial de indexação, agregação e particionamento.

A fase de *projeto e desenvolvimento da área de organização de dados* faz referência as principais atividades a serem realizadas no *Data Warehouse*:

- Realização do processo de organização de dados com uma dimensão e avaliação do resultado;

- Definição e realização de modificações lógicas e atualização dos registros das dimensões, quando necessário;
- Realização do processo de organização de dados com as demais dimensões;
- Desenvolvimento e aplicação de procedimentos para assegurar a qualidade dos dados;
- Desenvolvimento e realização de atividades de arquivamento, *back-up* e recuperação de dados.

Na fase de *especificações de aplicações de usuário final*, se faz necessário à identificação das áreas consideradas prioritárias, e após esta identificação a definição de um conjunto padronizado de aplicações.

No *desenvolvimento de aplicações de usuário final*, são desenvolvidas as aplicações necessárias de acordo com os levantamentos realizados na fase anterior, inclusive:

- Desenvolvimento de relatórios;
- Desenvolvimento de procedimentos de manutenção e atualização das aplicações de usuário final.

A fase de disponibilização do *Data Warehouse* é composta das seguintes atividades;

- Plano de disponibilização do *Data Warehouse*;
- Teste completo do sistema;
- Disponibilização do *Data Warehouse* propriamente dito.

E por fim, a fase de *manutenção e crescimento do Data Warehouse*, basicamente é composta pelas atividades de suporte contínuo e treinamento dos usuários e manutenção da infra-estrutura técnica; comunicação e publicidade dos sucessos obtidos em razão do uso do *Data Warehouse*.

Em todas as fases o projeto deve estar continuamente sendo revisto e submetido ao usuário final para aceitação.

3.6 Inserindo dados em um Data Warehouse

Um dos aspectos críticos na criação e manutenção de um *Data Warehouse* consiste na carga dos dados oriundos dos sistemas transacionais e outras fontes de dados externas.

O processo de inserção de dados compreende as etapas de extração, filtragem e carga ou migração dos dados.

A extração dos dados é feita a partir das diversas fontes de dados (sistemas transacionais, planilhas eletrônicas, arquivos de texto). Os dados são extraídos de seus sistemas fonte para serem posteriormente trabalhados. Um grande desafio é determinar quais dados extrair.

Uma vez extraídos de seus sistemas de origem, os dados devem ser tratados, ou filtrados. Nesta etapa os dados são transformados para um padrão único definido para o *Data Warehouse*. Além disso, é verificada sua integridade, e estes dados poderão ser submetidos a algum tipo de agregação, normalização, cálculos etc.

Após a filtragem os dados são carregados no banco de dados do *Data Warehouse*.

3.7 Visualizando dados de um Data Warehouse

Os dados constantes do banco de dados de um *Data Warehouse* podem ser visualizados a partir de dois tipos de ferramentas: *Softwares OLAP* ou *Data Mining*.

3.7.1 OLAP

As Ferramentas do tipo OLAP permitem que o usuário explore diversas possibilidades de pesquisa, e tem como principais características possibilitar a análise multidimensional (*"slice and dice"*) e o detalhamento ou sumarização de uma

informação (“*drill up/down*”). No entanto, este tipo de ferramenta requer que o usuário saiba aonde quer chegar, ou seja, o usuário define a questão de pesquisa. A ferramenta é usada então para confirmar ou não hipóteses já formuladas.

3.7.2 Data Mining

O *Data Mining* ou mineração de dados oferece ao analista informações que ele não solicitou. Usando das mais modernas técnicas de computação, como redes neurais, descoberta por regra, detecção de desvio, programação genética, os *Data Mining* extraem padrões e associações de dados descobrindo problemas e oportunidades nos relacionamentos entre os dados.

Um exemplo clássico de utilização do *Data Mining* foi a descoberta por uma das maiores redes de varejo dos Estados Unidos, que a venda de fraldas descartáveis estava associado a venda de cervejas, isto é, o comprador de fraldas descartáveis é um potencial consumidor de cervejas. Geralmente, os compradores eram homens e ao saírem para comprar fraldas compravam também latas de cervejas. Ao colocarem os produtos lado a lado, a venda de fraldas e cervejas aumentou consideravelmente. Um outro exemplo: a *Wal Mart*, umas das pioneiras no uso de *Data Mining* no mundo, e uma das maiores redes de varejo dos Estados Unidos, descobriu que 60% das mães que compravam a boneca *Barbie* levavam também uma barra de chocolate para a filha.

O *Data Mining* origina-se da combinação entre a Estatística e a Inteligência Artificial, chamada “*Machine Learning*”.

A Estatística usa de conceitos como distribuição normal, variância, análise de regressão, desvio simples, análise de conjuntos, análise de discriminantes e intervalos de confiança para estudar dados e descobrir relacionamentos entre eles.

A Inteligência Artificial tem base na heurística, e baseia-se na estrutura de pensamento humano.

A *Machine Learning* combina os conceitos estatísticos utilizados na análise de dados à Inteligência Artificial, fazendo com que os programas de computador “aprendam” ou seja, acrescentem novas regras de decisão, à medida que estudam os dados.

Os principais tipos de *Data Mining* são a seguir descritos brevemente, ao tempo em que são citados alguns exemplos de ferramentas referentes a cada tipo:

- **Redes Neurais** – É a tecnologia que oferece o mais profundo poder de mineração. Consiste em um número de elementos interconectados, organizados em camadas, chamados de neurônios, que estão constantemente realizando interações, com o objetivo de identificar relações. Exemplos de Ferramentas: *SPSS Neural Connection*, *IBM Neural Network Utility*, *NeuralWare*, e *Prism da Nestor Inc.*

- **Lógica baseada em casos** – Deriva regras a partir de estudos de casos. Exemplo de Ferramenta: *Remind, da Cognitive Systems.*

- **Processamento de Sinais** – Usa técnicas de processamento de sinais, tal como filtragem digital, para identificar observações com características similares. Exemplos de Ferramentas: *Data Engine, da alemã MIT GmbH.*

- **Fractais** – O uso de fractais é feito para comprimir grandes bancos de dados sem perda de informação, permitindo análise de todo o universo do banco com um excelente tempo de resposta. Exemplo de Ferramentas: *F-DBMS, da Cross/Z International.*

- **Indução de Regras** – Detecta tendências dentro de um grupo de dados formulando “regras” sobre este dado. O principal problema da Indução de Regras é que o programa não seleciona as regras formuladas em ordem de importância,

ficando esta tarefa a cargo do usuário. Exemplos de Ferramentas: *IDIS*, da *Information Discovery* e *Knowledge Seeker*, da *Agoss Software*.

- **Árvores de Decisão** – São quase sempre usadas em conjunto com Indução de Regras. Apresentam a solução na forma de árvore, que se assemelha a um organograma. Exemplos de Ferramentas: *Alice d'Isoft*, *Hyper Parallel/Discovery*, *Business Objects*, *BusinessMiner*, *Data Mind*, e *Angoss Knowledge Seeker*.

4 Data Warehouse e EIS (Sistemas de Informações Executivas)

No gerenciamento dos negócios, a análise dos fatos é um fator decisivo na vida dos gestores. Estes devem detectar as tendências e tomar decisões eficientes e no tempo certo. Surgiu com isso o conceito de *Business Intelligence*, sendo o *Data Warehouse* uma de suas ferramentas.

O *Data Warehouse* reduz sobremaneira a necessidade de geração de relatórios pré-definidos nos sistemas de informações. Estando o *Data Warehouse* povoado e disponível, o analista tem condições de ter uma postura pró-ativa com relação ao atendimento das necessidades da gerência, pois os executivos estão constantemente alterando seu foco a cada novo problema surgido ou uma oportunidade.

Como exemplo temos o seguinte: no dia 2, o executivo está com seu foco voltado para as atividades financeiras da empresa, devendo o analista atender este interesse; no dia 3 acontece um imprevisto na produção e a atenção da gerência é desviada, devendo o analista colher informações necessárias. No quarto dia, outro problema aparece, na expedição e o analista se vê direcionado para lá. Observa-se que não há um padrão, e o *Data Warehouse* se presta muito bem a este tipo de necessidade informativa, uma vez que o usuário pode montar suas “visões” no momento da consulta aos dados.

De acordo com Inmon (1997:p.245) o *Data Warehouse* está sempre a postos quando necessário, é integrado, possui dados detalhados e resumidos, cobre todos os tópicos que a gerência precisa e dispõe de longo horizonte de dados necessários para a análise de tendências.

Para Inmon (1997:p. 253), utilizando o *Data Warehouse*, o analista está capacitado a lidar com as necessidades da gerência em termos de:

- Informações rápidas;
- Mudança de enfoque;
- Examinar dados integrados;
- Examinar dados ao longo de um espectro de tempo.

Conclusão

O conceito de *Data Warehouse* abre uma nova perspectiva em termos de acesso à informação. A possibilidade de se reunir dados operacionais oriundos de diversos sistemas em um grande “armazém de dados”, e de se poder analisar tais dados de diferentes pontos de vista, sem a necessidade de se definir, de antemão, quais relatórios deverão ser gerados, dá grande flexibilidade à pesquisa.

Associado às ferramentas de Data Mining, os *Data Warehouse* se tornam ainda mais poderosos em termos de informação, uma vez que tais ferramentas oferecem ao gestor relações entre dados que este não havia observado, e que podem esconder problemas ainda não detectados ou oportunidades não vistas.

Tais características se coadunam com as requisitos enunciados pelos autores pesquisados, que discorreram acerca dos sistemas de informações estratégicas,

atendendo aos requisitos de flexibilidade, adaptabilidade, possibilidade simultânea de sumarização e detalhamento, dentre outros.

Esta nova forma de enxergar os dados corporativos e de tratá-los, não é, no entanto uma fórmula mágica capaz de resolver todos os problemas de informação, nem dispensa planejamento. Para desenvolver um *Data Warehouse* é preciso conhecer as regras de negócio da organização, planejar a amplitude do banco de dados, definir os tratamentos aos quais os dados serão submetidos etc. O sucesso do *Data Warehouse* depende, em grande parte, deste planejamento.

Bibliografia

AMARAL JR., Antônio, *Desmistificando Definitivamente o Data Warehouse*, Developers' Magazine, Rio de Janeiro, número 6, p. 14-17, fevereiro de 1997.

ARIMA, Carlos Hideo. Sistemas de informações gerenciais. SCHMIDT, Paulo (org.) *Controladoria: agregando valor para a empresa*. Porto Alegre: Bookman. 2002. P.79-90.

BERSON, Alex e SMITH, Stephen J., *Data Warehousing, Data Mining & OLAP*. Nova York, Ed. McGrawhill, 1997.

BEUREN, Ilse Maria; MARTINS, Luciano Waltrick. *Sistema de informações executivas: suas características e reflexões sobre sua aplicação no processo de gestão*. Revista Contabilidade & Finanças FIECAFI – FEA – USP. São Paulo, v.15, n. 26, p. 6-24, maio/agosto 2001.

CAMPOS, Maria Luiza e ROCHA Filho, Arnaldo V. *Data Warehouse* / Rio de Janeiro, setembro 1999. Disponível em: <http://genesis.nce.ufrj.br/dataware/tutorial/> Acesso em: 20. Agosto de 2001.

CIELO, Ivã Rafael. *Metadados*. Disponível em: http://www.datawarehouse.inf.br/metadados/body_metadados.html Acesso em 20. Agosto de 2001.

COUTINHO, Fernando Vieira et alli. *Data Mining* Disponível em: <http://www.dwbrasil.com.br/html/dmining.html> Acesso em 20. Agosto de 2001.

COUTINHO, Fernando Vieira et alli. *Metadados*. Disponível em: <http://www.dwbrasil.com.br/html/metadados.html> Acesso em 20. Agosto de 2001.

COUTINHO, Fernando Vieira et alli. *Características do Data Warehouse*. Disponível em: http://www.dwbrasil.com.br/html/artdw_carac.html Acesso em 20. Agosto de 2001.

COUTINHO, Fernando Vieira et alli. *Área de Stage*. Disponível em: <http://www.dwbrasil.com.br/html/stage.html> Acesso em 20. Agosto de 2001.

FAGUNDES, Eduardo. *Que Informações devem ser armazenadas no data warehouse?* Disponível em: <http://www.efagundes.com/disc1/0000002a.htm> Acesso em 20.Agosto de 2001.

GUROVITZ, Hélio *O que cerveja tem a ver com fraldas?* Disponível em: <http://www2.uol.com.br/info/exame/armaz.html> Acesso em 25 Agosto de 2001

INMON, W. H. *Como construir o Data Warehouse* Trad. Ana Maria Netto Guz. Rio de Janeiro, Ed. Campus, 1997.

KIMBALL, Ralph. *Aprofundando-se no Data Mining. DBMS*. Rio de Janeiro, número 7, p. 13-15, novembro de 1997.

_____, *Joint Effort* 24.05.2001 Disponível em : http://www.intelligententerprise.com/010524/webhouse1_1.shtml Acesso em 26 de agosto de 2001

MACHADO, Carlos. *A abc da mineração dos dados*. Info Exame. São Paulo, número 154, janeiro de 1999.

PELEIAS, Ivam Ricardo. *Controladoria: gestão eficaz utilizando padrões*. São Paulo: Saraiva, 2002.

PEREIRA, Walter Adel Leite. *Uma Metodologia de Inserção de Tecnologia de Data Warehouse em Organizações*. Dissertação de Mestrado PUC-RS-Faculdade de Informática Programa de Pós Graduação em Ciência da Computação. Porto Alegre, Outubro/2000.

SANTOS, Kellyne Marques. *Um Mecanismo para Redirecionamento de Consultas e Agregados em Ambientes de Data Warehouse*. Dissertação de Mestrado PUC-RS-Faculdade de Informática Programa de Pós Graduação em Ciência da Computação. Porto Alegre, 2001.

VILAROUCA Jr, Josafá *Acerte nos dados 07/Fev/2000* Disponível em: www.informationweek.com.br/noticias/artigo.asp?id=1403 Acesso em 25 Agosto de 2001