

CAPÍTULO 2. OS SISTEMAS DE INFORMAÇÃO NAS ORGANIZAÇÕES

2.1 A informação nas organizações

A informação é hoje, o insumo básico para que as empresas modernas possam obter sucesso em suas atividades. Para tanto, os computadores são instrumentos perfeitos para fornecer informações, já que possuem grande capacidade de armazenamento e processamento de dados para a geração de informação.

A informação é usada dentro de um ambiente organizacional tanto para fornecer mais precisão e segurança aos processos operacionais quanto para auxílio gerencial, no sentido de fornecer informações com maior qualidade.

De acordo com Polloni (2000), o papel fundamental da informação dentro das empresas é como provedora de dados, seja na área de planejamento de crescimento das atividades comerciais ou industriais, seja na organização, controle e estabelecimento das diretrizes de atuação no mercado.

A importância estratégica da informação nas organizações é dar apoio às suas metas, como, por exemplo, aumentar os lucros, diminuir os custos, melhorar a produtividade, identificar áreas decisivas para o sucesso como criação de novos mercados, desenvolvimento ou aperfeiçoamento de produtos ou serviços, e com isso aumentar sua vantagem competitiva.

Para ter valor para os administradores e tomadores de decisão, segundo Stair (1998), a informação deve ter várias características, como ser precisa, completa, de produção econômica, flexível, confiável, relevante, de fácil compreensão, atual e verificável, pois o valor da informação está diretamente ligado ao modo como ela ajuda os tomadores de decisão a atingirem os objetivos da organização.

A informação torna-se cada vez mais a base para a competição, portanto, para se manter mais competitiva no mercado a empresa deve contar com informações

apuradas para que suas estratégias sejam realmente eficazes. A empresa que tem a informação certa, no momento certo, estrategicamente é mais competitiva do que sua concorrente.

Cada vez mais as empresas precisam estar preparadas para as mudanças que vêm ocorrendo no ambiente em que estão atuando, principalmente porque o tempo agora é de globalização, onde as tecnologias de telecomunicações estão reduzindo rapidamente o tamanho do planeta, fazendo com que as informações e os dados sejam disseminados com uma velocidade cada vez maior.

2.1.1 Dados, informação e conhecimento

Segundo Stair (1998), para ser um administrador eficiente em qualquer área de negócios, é preciso entender que a informação é um dos recursos mais importantes e valiosos de uma empresa, porém, muitas vezes este termo é confundido com o termo dados.

Dados são os fatos em sua forma primária, que quando estão organizados ou arranjados de uma forma significativa, se tornam uma informação, a qual pode ser definida então, como um conjunto de fatos organizados de tal forma que adquirem valor adicional além do valor do fato em si, ou seja, a informação é criada definindo-se e organizando as relações entre os dados, sendo que a definição de diferentes relações resulta em diferentes informações (Stair, 1998).

Segundo Vieira (2001), a diferença entre dado e informação é que os dados são os componentes básicos a partir dos quais a informação é criada e as informações são dados inseridos em um contexto, ou seja, uma situação que está sendo analisada.

Existem muitas empresas que possuem um grande número de dados em sistemas que servem apenas para registro das transações, como controle de estoque, contabilidade, controle de produção.

O grande problema de uma organização que conta primordialmente apenas com estes sistemas tradicionais para registro das transações é que ela tem muitos DADOS, mas com pouca INFORMAÇÃO, gerando, de acordo com Vieira (2001), alto volume de

dados, porém necessidade de respostas rápidas, análises complexas e questões impossíveis de se prever, e ainda, em um ambiente de negócio em mudança constante.

De acordo com Barbieri (2001), existem dados operacionais e dados informacionais, os quais possuem objetivos diferentes, onde os dados operacionais estão relacionados basicamente aos sistemas tradicionais de informações, implementados sobre bases de dados, e os dados informacionais estão relacionados basicamente aos sistemas de informações executivas, implementados sobre Data Warehouse (DW) ou Data Marts (DM). As principais diferenças entre os dados de natureza operacional e os de natureza informacional são apresentados no Quadro 1.

QUADRO 1. DIFERENÇAS ENTRE OS DADOS DE NATUREZA OPERACIONAL E OS DE NATUREZA INFORMACIONAL

Características	Dados Operacionais	Dados Informacionais
Conteúdo	Valores correntes	Valores sumariados, calculados, integrados de varias fontes
Organização dos dados	Por aplicação/sistema de informação	Por assuntos/negócios
Natureza dos dados	Dinâmica	Estática até o “ <i>refreshment</i> ” dos dados
Formato das estruturas	Relacional, próprio para computação transacional	Dimensional, simplificado, próprio para atividades analíticas
Atualização dos dados	Atualização campo a campo	Acesso, sem <i>update</i>
Uso	Altamente estruturado, processamento repetitivo	Desestruturado, com processamento analítico/heurístico
Tempo de resposta	Otimizado para 2 ou 3 seg	Análises mais complexas, com tempos de respostas maiores.

Fonte: Barbieri, 2001

Conforme Barbieiri (2001), de início a informática fez os dados, depois os transformou em informação, atualmente o objetivo é garimpar conhecimentos, a partir destas matérias-primas.

Portanto, percebe-se que é a partir da informação que vem o conhecimento, o qual permite tomar decisões adequadas, trazendo assim a tão necessária vantagem competitiva (Vieira, 2001).

2.2 Sistemas de Informação

Para Bio (1985), um sistema de informação é conceituado como sendo um conjunto de partes interdependentes no seu todo que podem ser partes de um todo maior. Baseados no acima exposto tem-se a idéia de que o 'sistema empresa' é composto por vários subsistemas, e que o 'sistema de informação' é um destes subsistemas, que por serem interagentes estão em constante troca de informação. Por exemplo, o sistema de vendas interage com o de estoque, este com o financeiro, compras, entre outros.

Ampliando o escopo do conceito de sistema de informação, Campos (1994), afirma que o sistema de informação é uma combinação estruturada de informação, recursos humanos, tecnologia de informação e práticas de trabalho, organizados de forma a permitir o melhor atendimento dos objetivos da organização.

Os objetivos determinam os critérios básicos que decidem como e quando as práticas de trabalho devem ser mudadas e adaptadas, que também podem ser influenciadas pelos recursos humanos, informações e tecnologias de informação em uso, assim como também outros fatores internos e externos ao sistema. Um exemplo é quando a empresa decide lançar um produto novo, então há uma motivação de várias práticas de trabalho, combinadas sistemicamente com os recursos humanos, informações e tecnologias de informação.

Os sistemas de informações devem tratar a informação como um elemento essencial na tomada de decisão sendo muito importante manter sua precisão e

segurança, portanto, os sistemas de informação deixam de ser apenas um instrumento de redistribuição dos dados para atuarem também como um instrumento de apoio gerencial.

Pode-se dizer ainda que os sistemas de informação devem dar qualidade às informações empresariais, bem como filtrá-las por níveis de decisão e, ainda fazer com que dos níveis menores, as informações se condensem em outras informações para os níveis acima, e assim, progressivamente até o maior escalão de decisão, que receberá as informações para decisões estratégicas resumidamente, podendo atuar da melhor maneira possível (Peres e Tait, 1996).

Segundo O'Brian (2001:9), os sistemas de informação desempenham três papéis vitais em qualquer tipo de organização, conforme Figura 1:

Suporte de seus processos e operações.

Suporte na tomada de decisões de seus funcionários e gerentes.

Suporte em suas estratégias em busca de vantagem competitiva.

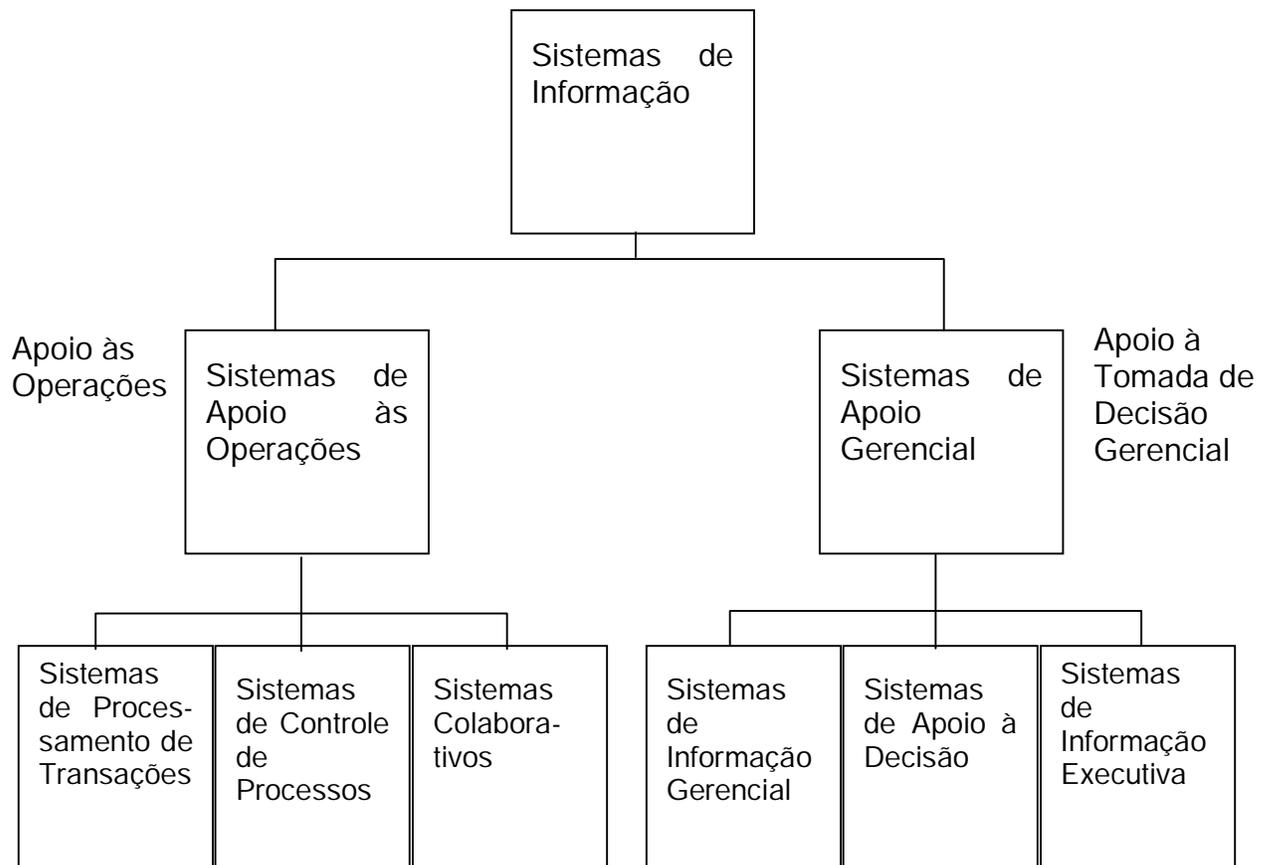


Fonte: O'Brian, 2001

FIGURA 1. PAPÉIS FUNDAMENTAIS DOS SI

Desta forma, é possível classificar os sistemas de informação sob a ótica dos níveis de informação e de decisão que obedecem a hierarquia padrão existente na maioria das empresas, e são conhecidos como estratégico, tático e operacional, onde o tipo de decisão que é tomada em cada nível requer diferente grau de agregação da informação, e os diferentes níveis de decisão requerem diferentes informações (Rezende, 2000).

De acordo com Stair (1998) e O'Brian (2001), os sistemas de informação, em termos conceituais podem ser classificados ora como sistemas de apoio às operações, onde se incluem os sistemas de processamento de transações, os sistemas de controle de processo e os sistemas colaborativos, ora como sistemas de informação gerencial, onde se incluem os sistemas de informação gerencial, os sistemas de apoio à decisão e os sistemas de informação executiva, conforme ilustra a Figura 2.



Fonte: O'Brian, 2001

FIGURA 2. TIPOS DE SISTEMAS DE INFORMAÇÃO.

Entre outras classificações dos sistemas de informação, mais exclusivas ou amplas a fim de suportar aplicativos operacionais ou gerenciais, O'Brian (2001) menciona os sistemas especialistas, os sistemas de administração do conhecimento, os sistemas de informação estratégica e os sistemas de informação para as operações.

Os sistemas de informação podem afetar das mais variadas formas o desempenho das modernas organizações, abrindo-lhes inclusive, novos espaços e oportunidades de atuação competitiva, motivo pelo qual, gerentes e administradores não podem deixar de compreender sua natureza e a utilizar seus recursos com eficácia (Campos, 1994).

2.2.1 Sistemas de Apoio à Decisão (SAD)

Os sistemas de apoio à decisão surgiram para suprir a necessidade dos mais altos níveis organizacionais das empresas, pois objetivam viabilizar a utilização do computador de forma interativa com os tomadores de decisão de forma a utilizar os dados e modelos nas mais diversas fases de seu processo decisório.

De acordo com Polloni (2000), a introdução de sistemas administrativos tratados via computador de grande porte (*mainframes*), no início da década de 70, resultou na expansão dos sistemas centralizados, e logo a filosofia SAD surgiu. Nessa mesma década, os sistemas centralizados devido as suas deficiências, foram substituídos pelos sistemas descentralizados, isto é, “cada um por si”, e também os SAD’s foram acoplados a essa nova abordagem dos sistemas administrativos.

Sistemas de apoio à decisão podem ser definidos conceitualmente como aplicações responsáveis pela captura e elaboração das informações dentro de uma base de dados, que sejam pertinentes no processo de tomada de decisões.

Polloni (2000, p.31), afirma que os sistemas de apoio à decisão:

“são sistemas que tratam de assuntos específicos, estatísticas, projeções, comparações de dados referentes ao desempenho da empresa, estabelecendo parâmetros para novas ações dentro do negócio da empresa. Caracterizam-se, essencialmente, pela utilização de pacotes interativos para cálculos e/ou simulações (mercado, finanças). Os resultados obtidos por estes pacotes são mostrados em vídeo, gráficos, impressos etc.”

De acordo com O'Brian (2001, p.253) os sistemas de apoio à decisão:

“são sistemas de informação computadorizados que fornecem aos gerentes apoio interativo de informações durante o processo de tomada de decisão. Os sistemas de apoio à decisão utilizam (1) modelos analíticos, (2) bancos de dados especializados, (3) os próprios *insights* e apreciações do tomador da decisão e (4) um processo de modelagem computadorizado para apoiar a tomada de decisões semi-estruturadas e não-estruturadas por parte de cada gerente.”

Segundo Rezende (2000), o SAD utiliza muito a regra ‘e se’ para a geração de dados e informações, como por exemplo:

- Determinação do local mais adequado para a alocação de uma nova loja, indústria ou ponto de venda;
- Orçamentos elaborados com várias alternativas;
- Segmentação de negócios, aliado a um possível perfil de clientes;
- Verificação de rentabilidade de vendas apontando opções de não vender determinado produto.

Portanto, os SAD's permitem a integração e coordenação de dados das mais variadas partes, fornecendo informações que permitem melhores decisões empresariais (Rezende, 2000).

As características que um sistema de apoio à tomada de decisão apresentam e que lhes proporcionam o potencial para serem eficazes ferramentas de apoio gerencial, segundo Stair (1998), são:

- Manipular grandes volumes de dados;
- Obter e processar fontes de dados diferentes;
- Proporcionar flexibilidade de relatórios e de apresentação;
- Possuir orientação tanto textual quanto gráfica;
- Executar análises e comparações complexas e sofisticadas utilizando pacotes de softwares avançados;

- Dar suporte às abordagens de otimização, satisfação e heurística;
- Executar análises de simulações e por metas.

Naturalmente nem todas estas características estão presentes em todos os SAD's, alguns possuem um escopo menor e apresentam apenas algumas dessas características. Portanto, na seleção de um SAD, devem ser observados os benefícios dessas características em termos de custo, controle e complexidade (Stair, 1998).

2.2.2 EIS (Executive Information Systems)

Os sistemas de informação executiva (EIS) destinam-se a auxiliar os executivos de alto nível administrativo na tomada de decisões estratégicas, fornecendo acesso fácil e imediato a informações sobre os fatores críticos ao sucesso de uma empresa, isto é, às atividades corretas e essenciais para o funcionamento de uma organização.

De acordo com O'Brian (2001), a maioria dos altos executivos empresariais não utilizavam diretamente os relatórios gerados por sistemas de informações gerenciais ou as capacidades de modelagem analítica dos sistemas de apoio à decisão. Então, desenvolveu-se o conceito de sistemas de informação executiva (EIS) para propiciar aos altos executivos uma maneira fácil de obter as informações críticas de que eles necessitam, quando as necessitam, elaboradas nos formatos por eles preferidos, porém, os sistemas de informação executiva combinam muitas das características dos sistemas de informação gerencial e dos sistemas de apoio à decisão, pois podem oferecer relatórios padrão (como os relatórios financeiros, por exemplo), mas também deve tornar fácil para os usuários a obtenção de seus próprios relatórios *ad hoc*.

Para Stair (1998), um EIS, é um tipo especial de SAD e, como tal, destina-se a apoiar a tomada de decisão da alta administração. Porém, os dois sistemas apresentam aspectos diferentes importantes. Um SAD fornece diversas ferramentas de modelagem e análise para capacitar usuários a analisar os problemas de forma integral, possibilitando que o usuário responda perguntas. Os EIS apresentam informações estruturadas sobre os aspectos da organização considerados importantes pelos executivos, permitindo que estes façam as perguntas certas, pois em geral, são

desenvolvidos sob medida para executivos individualmente, enquanto os SAD não são destinados a um usuário em particular. Portanto, os EIS representam o objetivo global dos sistemas de informação: fornecem as informações certas, no momento certo, para a pessoa certa e no formato certo.

Como os sistemas de informação executiva apresentam a informação de acordo com as preferências dos executivos, geralmente eles possuem recursos gráficos de alta qualidade que possam ser personalizados conforme as preferências dos executivos. Os EIS destinam-se a proporcionar informações de forma rápida para a tomada de decisões críticas, são fáceis de usar, com telas de acesso intuitivo, são desenvolvidos de modo que se enquadrem na cultura, filosofia, políticas e no modelo de gestão da empresa, e ainda, manipulam uma variedade de dados internos e externos do ambiente da empresa (Rezende, 2000).

À medida que os tomadores de decisão se aproximam dos níveis mais elevados, as situações internas e externas à empresa tornam-se mais importantes, fazendo com que os executivos estejam constantemente monitorando os processos internos da empresa, o ambiente, a concorrência, os clientes e as leis governamentais. Dessa forma, grande parte das informações de um alto executivo deriva de fontes como cartas, memorando, reuniões, telefonemas, jornais, periódicos e atividades sociais, que podem ser implementados em um EIS.

Um EIS pode ser aplicado para oferecer informações *on-line* sobre todas as atividades das funções empresariais existentes, contemplando seus respectivos acompanhamentos, desempenhos e retornos. Também pode ser aplicado para o lançamento de novos produtos no mercado, para o acompanhamento do comportamento do mercado financeiro e da movimentação dos recursos humanos envolvidos na empresa, para análise de possibilidade de fusão e apontar novas direções que a empresa pode tomar, entre outras, visando à tomada de decisões estratégicas para busca de vantagem competitiva (Rezende, 2000).

Segundo Stair (1998), um EIS pode ser usado também por pessoas dos níveis mais baixos da estrutura organizacional, porém, inicialmente era destinado aos

tomadores de decisão de alto nível como meio de acompanhar os fatores críticos de sucesso da organização, atualmente todos os níveis da organização compartilham informações e participam de redes a partir dos mesmos bancos de dados, pois os níveis de tomada de decisão gerencial ainda existem, mas as empresas estão se caracterizando por cada vez mais menos níveis hierárquicos e cada vez mais níveis decisórios possibilitando ao usuário um poder de acesso e análise de informações para a tomada de decisão rápida e de acordo com suas responsabilidades.

Seguindo a filosofia do sistema de apoio à decisão e dos sistemas de informação executiva, surge o conceito de *Business Intelligence* (Inteligência Aplicada aos Negócios), também com intuito de dar apoio e subsídio à tomada de decisão para a busca de vantagens competitivas, a qual será abordada no item seguinte.

2.3. BUSINESS INTELLIGENCE

2.3.1 Definição

Os sistemas de *Business Intelligence* são a evolução do Executive Information System (EIS) ou Sistema de Informação Executiva (SIE) e do Sistema de Apoio à Decisão (SAD).

Barbieri (2001), afirma que o conceito de BI, *Business Intelligence*, está na sua essência relacionado com formas alternativas de tratamento de informações para se definir estratégias de negócios da empresa, ou ainda, representa a habilidade de se estruturar, acessar e explorar informações, normalmente guardadas em DW/DM, com o objetivo de se desenvolver percepções, entendimentos, conhecimentos, os quais podem produzir um melhor processo de tomada de decisão.

Para Vieira (2001), o BI é um conjunto de conceitos e metodologias que, fazendo uso de acontecimentos (fatos) e sistemas baseados nos mesmos, apóia a tomada de decisões nos negócios, pois, a partir da informação é que vem o conhecimento, o qual permite tomar decisões adequadas, gerando então, vantagem competitiva para as empresas.

De acordo com Cerqueira (2002), o conceito de BI pode ser dado através do entendimento dos dois termos que compõem o referido conceito, onde Business (Negócio), quer dizer a intermediação de uma atividade comercial com fins lucrativos, e Intelligence (Inteligência), faculdade de aprender, compreender, capacidade de resolver situações problemáticas mediante reestruturação dos dados perceptivos.

Com a junção dos dois termos supõem-se que a inteligência do negócio está ligada à capacidade de pessoas, em posições estratégicas dentro da empresa, com poder de decisão, para alterar ou adaptar o rumo da empresa, internamente (estrutura, finanças, distribuição, etc.) ou externamente (mercado, concorrência, etc.). Portanto, o conceito de BI têm o objetivo de auxiliar estas pessoas, por meio do tratamento da base de dados existente de maneira a aprimorar o processo de tomada de decisão. (Cerqueira, 2002).

Vieira (2001), pressupõe as seguintes características dos sistemas BI:

- Analisar dados contextualizados;
- Fazer uso da experiência;
- Trabalhar com hipóteses;
- Procurar relações de causa e efeito;
- Extrair e integrar dados de múltiplas fontes;
- Transformar os registros obtidos em informação útil para o conhecimento empresarial.

A vantagem dos sistemas de BI sobre seus antepassados (EIS e SSD/SAD) é a capacidade de cruzar informações de diferentes bancos de dados, analisando-as, além de possibilitar a interpretação de forma contextualizada, pois trabalham com hipóteses e procuram relações de causa e efeito, transformando os registros dos bancos de dados em informações estratégicas para o posicionamento e direcionamento da atuação da empresa.

Já o EIS se limita a consultas pré-definidas, e o DSS/SAD dispensa a interação direta com um profissional de TI, pois gera consultas sem a necessidade de

programação prévia, porém busca respostas para a pergunta do usuário em bases de dados relacionais.

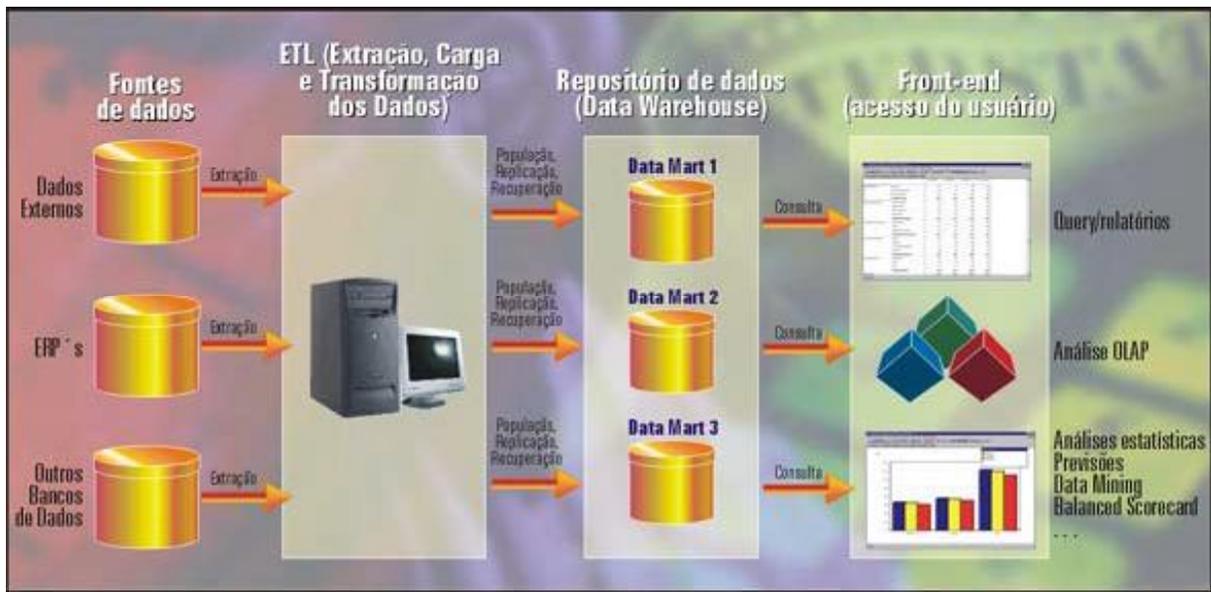
Assim, para que as ferramentas de BI funcionem de forma realmente eficiente, é necessário que a empresa tenha sólidos bancos de dados. Afinal, o sistema de BI transforma e apresenta os elementos dos bancos de dados em informações determinantes para o desenvolvimento do negócio (Monte, 2002).

Através do *Business Intelligence*, todos os dados presentes na empresa, em suas diversas bases de informação, sejam elas informatizadas (como bancos de dados e sistemas) ou não (cultura e história da empresa), são filtrados e disponibilizados de uma forma mais adequada e de fácil acesso, para a tomada de decisão, seja para gerentes, técnicos, diretores, ou quaisquer outros componentes da organização, provendo aos usuários um poder de análise e acesso a informações que possam agregar maior qualidade e valor aos processos da empresa (Bertini, 2002).

2.3.2 Arquitetura de um sistema de BI e as Tecnologias de Informação

A arquitetura de um sistema típico de *Business Intelligence* é composta pelas seguintes partes, ilustrados na Figura 3, (WG Systems, 2002a):

- Módulo de ETL (Extração, Transformação e Carga dos Dados): componente dedicado à extração, carga e transformação de dados. É a parte responsável pela coleta das informações nas mais diversas fontes, desde sistemas ERP's até arquivos TXTs ou planilhas Excel.
- Data Mart/Data Warehouse: local onde ficam concentrados todos os dados extraídos dos sistemas operacionais. A vantagem de ter um repositório de dados à parte é a possibilidade de armazenar informações históricas e agregadas, dando um suporte melhor para as análises posteriores.
- Front-end: é a parte visível ao usuário de um projeto de BI. Pode ser em forma de relatórios padronizados e *ad hoc*, portal de intranet/internet/extranet, análise OLAP e outras funções, como *data mining* e projeções de cenários futuros.



Fonte: WG Systems, 2002a

FIGURA 3. ARQUITETURA DO *BUSINESS INTELLIGENCE*

2.3.2.1 ETL (Extração, Transformação e Carga dos Dados)

O objetivo da etapa de ETL é fazer a integração de informações de fontes múltiplas e complexas, portanto, torna-se uma etapa bastante crítica, já que uma informação que for carregada erroneamente, pode trazer conseqüências imprevisíveis nas fases posteriores. Esta etapa, divide-se basicamente em três passos (WG Systems, 2002b):

Carga: inicialmente deve-se definir as origens das fontes de dados, que podem ser várias e também em diferentes formatos, como de sistemas transacionais como ERP até planilhas, arquivos textos, arquivos DBF ou do Microsoft Access, e fazer a extração deles.

- Limpeza: depois de definir as fontes, o passo seguinte é transformar e limpar esses dados. A necessidade da limpeza dos dados se dá porque os dados normalmente advêm de uma fonte muitas vezes desconhecida, concebida há muito tempo e contendo muito lixo e inconsistência, portanto, faz-se à limpeza desses dados, para haver compatibilidade entre eles.

Transformação: a origem dos dados muitas vezes pode ser de sistemas diferentes, então, é necessário padronizar os diferentes formatos, pois o

usuário não pode ver informações iguais em formatos diferentes. Portanto, o processo de ETL, transforma estes dados e deixa-os num formato uniforme normalmente sugerido pelo próprio usuário no DW, fato esse que facilitará a análise dos dados que serão recuperados pela ferramenta OLAP.

Apesar de existirem ferramentas de ETL prontas no mercado, todas com seus diferenciais e cada uma podendo ser utilizada dependendo do caso de cada empresa, às vezes é necessário criar rotinas de carga para atender determinadas situações que poderão ocorrer.

As ferramentas de ETL tem grande valia, principalmente se existirem muitos sistemas OLTP (transacionais), pois são uma poderosa fonte de geração de metadados¹, que contribuirão muito para a produtividade da equipe que está desenvolvendo o projeto de BI.

2.3.2.2 Data Warehouse

Os bancos de dados e suas teorias já existem há bastante tempo, mas inicialmente concentravam-se em um único banco de dados que consistia desde o processamento de transações ao processamento *batch* (de lote) e ao processamento analítico.

Com o passar do tempo, surge a necessidade de um novo conceito em banco de dados, que atende a necessidades operacionais separadamente de outro que atende a necessidades informacionais ou analíticas. Isto se deve ao fato de que os dados, a tecnologia de suporte, a comunidade de usuários e as características de processamento diferem em um banco de dados que atende a necessidades operacionais daquele que atende a necessidades de informações ou analíticas.

¹ Metadados: são dados a respeito de dados. Exemplos de metadados incluem as descrições de elementos de dados, descrições de tipos de dados, atributos/propriedades, faixas/domínios, métodos e processos. O ambiente do repositório abrange todos os recursos de metadados.

Segundo Inmon (1997a), o processamento informacional ou analítico atende às necessidades dos gerentes durante o processo de tomada de decisões. Focalizado dentro do ambiente analítico, e geralmente conhecido como processamento SAD (sistema de apoio à decisão), o processamento analítico examina grandes conjuntos de dados para detectar tendências, onde os registros são constantemente acessados, agrupando seus conteúdos para análise, mas ocorrendo pouca ou nenhuma alteração nos registros individuais.

Para atender às necessidades de gerenciamento destes dados, faz-se necessário uma mudança de arquitetura, surgindo o ambiente projetado de *data warehouse* (DW). De acordo com Inmon (1997a), um *data warehouse* é um conjunto de dados baseado em assuntos, integrado, não volátil, variável em relação ao tempo e de apoio às decisões da gerência.

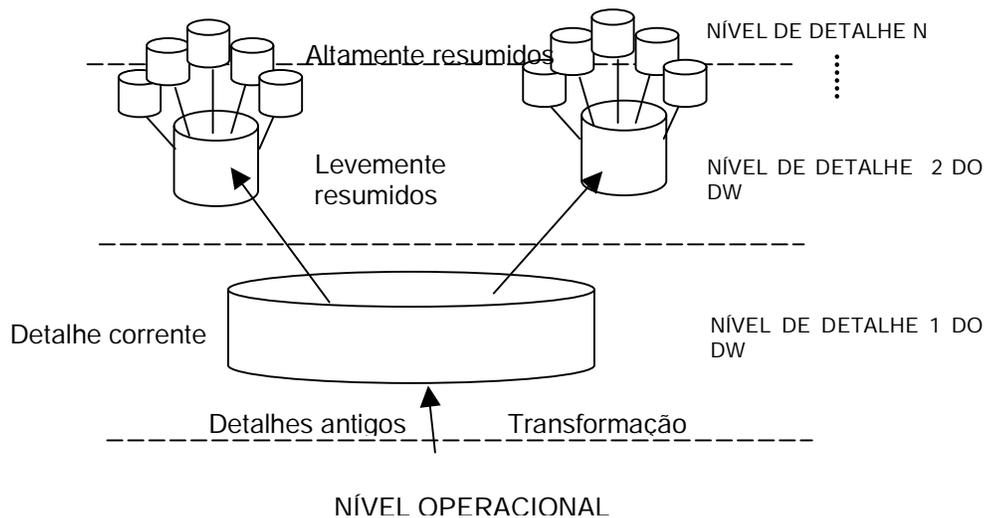
Na visão de Sell (2001), a análise desta definição de *data warehouse* é:

- Baseado em assuntos: isto é, tem o intuito de fornecer informações estratégicas sobre o negócio, e não descrever os processos do negócio, como os sistemas operacionais.
- Integrado: dados anteriormente independentes fornecendo respostas redundantes e ambíguas são integrados em fonte única, produzindo respostas únicas e certas.
- Não volátil: uma vez armazenado, o dado não sofrerá alterações, isto é, os dados são apenas acrescentados no DW e raramente atualizados.
- Variável em relação ao tempo: como os dados estão sendo sempre acrescentados no DW, há um acúmulo de dados sobre diversos períodos, possibilitando análises históricas do negócio, tendências, etc.

Neste ambiente há a percepção de duas espécies de dados: primitivos/operacionais e derivados/SAD, que diferem quanto à sua aplicação e gerenciamento. Os dados primitivos são baseados em aplicações cotidianas da empresa, enquanto os dados derivados são baseados em assuntos ou negócios de forma a atender as necessidades da gerência. A codificação deve ser feita de forma consistente e independente da aplicação de origem, de

modo que as muitas inconsistências das aplicações sejam desfeitas quando introduzidas no DW (Inmon, 1997a).

Uma estrutura simplificada do DW é mostrada na Figura 4, onde diferentes níveis de detalhe no DW são demonstrados. Os dados fluem do nível operacional para o nível de DW, ocorrendo uma grande quantidade de transformação sobre os dados. Há um nível de detalhes antigos, um nível corrente de detalhe, um nível de detalhes levemente resumidos (data mart) e um nível altamente resumidos.



Fonte: Inmon, 1997a

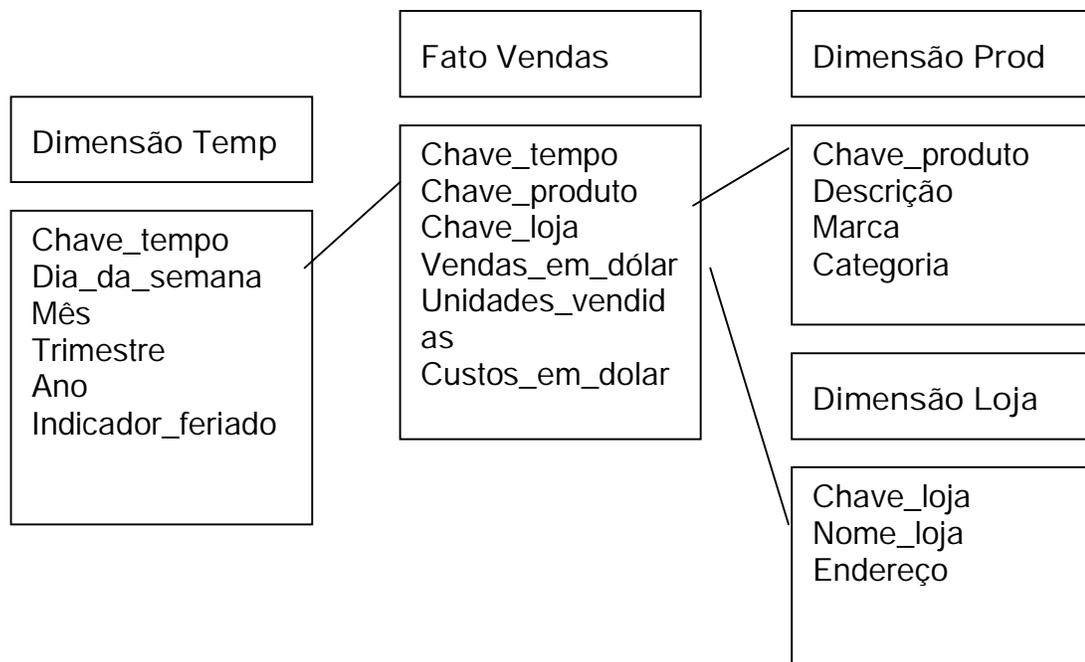
FIGURA 4. A ESTRUTURA DO DATA WAREHOUSE

Através do nível de data mart pode-se obter informações dos diversos departamentos de uma empresa. A fonte de dados deste nível é o DW. Muitas empresas começam sua implementação de DW pelos data marts, pois sua implementação é feita em menor tempo e seus custos também são menores (Kimball, 1998).

No conceito de Barbieiri (2001), o termo Data Mart (Mercado de Dados), significa depósito de dados que atende a certas áreas específicas da empresa, ou seja, ele pode responder um problema empresarial particular, voltado, assim como o DW, para o processo decisório gerencial, sendo que ambos podem ser definidos como espécie do mesmo tipo, sendo diferenciados basicamente no escopo do projeto e nos limites de sua abrangência.

De acordo com Kimball (1998a), uma forma alternativa de modelagem de dados para projetos de DW é o modelo dimensional, que dá a idéia de um cubo de dados, onde é possível ver as informações em várias dimensões. Pode-se dizer que um esquema relacional é bastante diferente, pois esquemas do modelo dimensional são em estrela (*star join schema*), devido ao fato que para descrever modelos dimensionais o diagrama é semelhante a uma estrela com uma tabela grande no centro, chamada tabela de Fato, rodeada por tabelas auxiliares, chamadas tabelas Dimensão. A Figura 5 mostra um esquema estrela de um negócio que vende produtos em algumas lojas e mede seu desempenho ao longo do tempo.

Para Kimball (1998b), o modelo dimensional é uma forma de organizar os dados que propicia flexibilidade conforme a necessidade de eventuais ajustes no modelo como adicionar novos fatos à tabela de fatos, desde que correspondam ao mesmo nível de detalhe, adicionar novas dimensões, adicionar novos atributos às dimensões existentes e redefinir o nível de detalhes dos dados.



Fonte: Kimball, 1998

FIGURA 5. UM MODELO DIMENSIONAL TÍPICO.

Em termos simples, um *Data Warehouse* é um grande banco de dados que reúne informações integradas a partir de dados operacionais de uma organização para que esta possa obter informações sobre tudo e todos.

Os dados operacionais transformam-se em uma ferramenta estratégica graças a característica principal do DW que é a integração dos dados operacionais ao seu ambiente, sendo utilizada pelos usuários de primeira linha (executivos), que tomam decisões-chave sobre negócios baseados em informações disponíveis.

As principais promessas do DW são agilidade e precisão na tomada de decisões. Com a informação *on-line* e no formato que desejam, os executivos, ganham tempo e segurança para tomar as decisões, sem precisar vasculhar em arquivos espalhados por vários servidores da organização (Militello, 1997).

2.3.2.2.1 Fases de um Projeto de Data Warehouse/Data Marts

O processo de desenvolvimento de um projeto de DW/DM possui alguns aspectos que se assemelham ao desenvolvimento tradicional de sistemas, porém com algumas diferenças que devem ser observadas com cuidado.

Segundo Inmon (1997a), o mais importante aspecto do projeto de um DW é a questão da granularidade, que diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no DW. Quanto mais detalhe, mais baixo o nível de granularidade, quanto menos detalhe, mais alto o nível de granularidade. O volume de dados contidos no DW é balanceado de acordo com o nível de detalhe de uma consulta.

Para definir o nível de granularidade adequado é importante colocar o usuário final frente aos dados, para então construir o DW, pois só depois de o usuário ter visto efetivamente os dados é que pode ser dada uma resposta definitiva quanto ao nível de granularidade adequado. Pode haver ainda, o nível dual de granularidade, que se enquadra nos requisitos da maioria das empresas, com os dois tipos de dados, levemente resumidos (compactos de fácil acesso) e dados históricos detalhados (em que um nível maior de detalhe deve ser investigado) (Inmon, 1997a).

Outra importante questão do projeto dos dados contidos no DW (depois da granularidade) é o particionamento, que ocorre quando dados de uma mesma estrutura são divididos em mais de uma unidade física de dados e toda unidade de dados pertence a uma e somente uma partição. Há vários critérios por meio dos quais é possível dividir dados, como por ex: por data, por área de negócio, por clientes, por todos estes critérios. No ambiente de DW é praticamente obrigatório que um dos critérios de particionamento seja a data. Se tanto a granularidade, quanto o particionamento forem apropriadamente executados, então quase todos os outros aspectos de projeto e implementação do DW brotarão naturalmente (Inmon, 1997a).

Barbieri (2001), ressalta que o sucesso de um projeto de DW está ligado com a escolha correta dos usuários que definirão as necessidades de informação, que deverão ter poder de decisão influência nas estruturas das áreas envolvidas.

Os principais passos para o projeto de DW/DM são ilustrados na Figura 6 sendo que os cinco primeiros segundo Barbieri (2001), são considerados críticos e os quatro últimos passíveis de muitos problemas, caso não sejam muito bem planejados e realizados.

1. Planejamento
2. Levantamento de Necessidades
3. Modelagem Dimensional
4. Projeto Físico dos BDs
5. Projeto de ETC
6. Desenvolvimento de Aplicações
7. Validação e Teste
8. Treinamento
9. Implantação

Fonte: Adaptado Barbieri, 2001

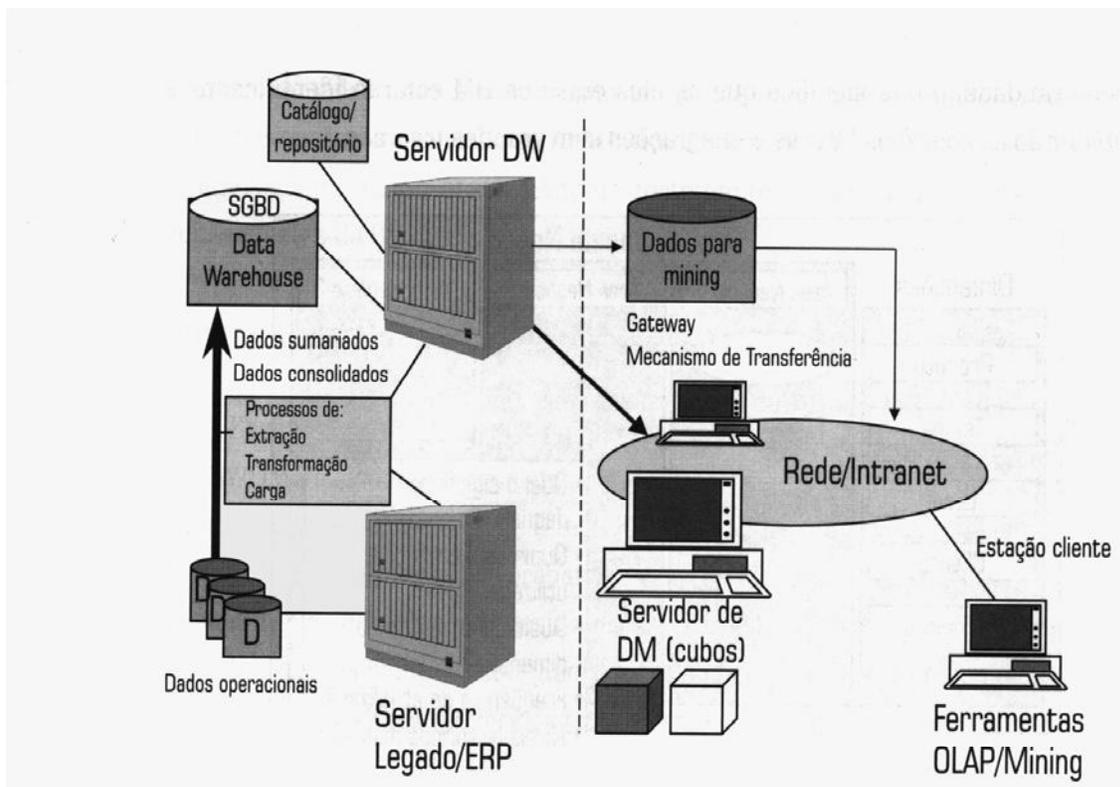
FIGURA 6. FASES DE UM PROJETO DE DW/DM

1. Planejamento: esta primeira etapa do projeto de DW/DM divide-se em quatro passos: foco no negócio, definição da abordagem, planejamento para integração, e definição da arquitetura tecnológica, sendo que os componentes básicos da arquitetura tecnológica que servirá de base para o projeto devem ser definidos antes do início do projeto de DW/DM, pois fatores relacionados à performance e

disponibilidade podem definir níveis de serviços e graus de compromissos variados durante o projeto. Os componentes tecnológicos básicos, conforme Figura 7, são:

- Sistema Gerenciador de Banco de Dados: é o ambiente onde o DW reside, devendo ser máquina mais robusta, e dependendo do projeto, um ambiente de alta disponibilidade, performance e segurança.

Ferramentas de Desenvolvimento de Sistemas OLAP e Mining: são os produtos que desenvolvem e executam aplicações OLAP e de Data Mining.



Fonte: Barbieri, 2001

FIGURA 7. ARQUITETURA TECNOLÓGICA PARA A IMPLANTAÇÃO DE UM AMBIENTE DE BI.

- Ferramentas para Processos de Extração, Transformação e Carga: estas ferramentas podem ser desenvolvidas internamente, ou adquiridas no mercado, com o objetivo de realizar os processos de transformação dos dados, sendo que a escolha dessas ferramentas, aliada à integridade dos

dados fonte se tornam fatores críticos de sucesso no projeto de DW/DM, proporcionando maior qualidade aos dados.

Catálogo para controle de Metadados: os metadados representam uma importante camada de documentação dos DW/DM, pois permite que a empresa conheça os cubos de dados disponíveis, suas dimensões e métricas, e informações sobre os dados que lhe deram origem.

- Mecanismos para transferência de dados entre ambientes heterogêneos: ferramenta de natureza variada, com o objetivo de facilitar o processo de transferência entre ambientes heterogêneos. Genericamente são denominados *middlewares*.

Servidor de Data Mart/Cubos: é o ambiente onde reside o gerenciador de Data Mart, onde também residem os DM, ou parte deles, na forma de extrações ou cubos, como são denominados, voltados para aplicações específicas.

- Extrato de dados para Data Mining: representa o armazenamento de informações, definido de forma especial, objetivando atender os requisitos de mineração de dados, aplicados a sistemas de inferência de informações.

2. Levantamento de Necessidades: nessa etapa deverão ser identificados dois modelos, o modelo dimensional e o modelo de fonte dos dados. O modelo dimensional representa os blocos conceituais de dados necessários ao alcance dos objetivos do sistema de suporte a decisão. O modelo de fonte de dados é relacionado com as fontes das informações, onde deverão ser registrados os blocos conceituais de dados existentes, com suas respectivas descrições e formas atuais de armazenamento e de uso nos sistemas.

3. Modelagem Dimensional: é a técnica de projeto que conduz os dados a uma fase, onde a informação reside na intersecção de várias dimensões. Num projeto de DW/DM o dado fundamentalmente importante é aquele consolidado e/ou sumariado

nas dimensões específicas, além daqueles com maior nível de granularidade ou detalhe.

4. Projeto Físico dos Bancos de Dados: nessa fase serão desenhadas as estruturas lógicas do modelo Dimensional, com as definições de tabelas Fatos e tabelas Dimensão, relacionamentos, indexação, atributos de tabelas e implantação de regras. É também nessa fase que a equipe projetista deverá optar pelo uso do SGBD Relacional ou o uso de um SGBD Dimensional como depósito das informações do armazém de dados.

5. Projeto de ETC – Extração, Transformação e Carga: nessa etapa deverão ser definidos os processos requeridos de transformação do modelo Fonte para o modelo Dimensional. Os conceitos de extração dos dados e de seu tratamento já foram definidos no capítulo 2.3.2.1

6. Desenvolvimento de Aplicações: nessa etapa será projetado o sistema aplicativo, objeto do trabalho, sendo que o sistema deverá priorizar a interface *Web*, facilitando o acesso aos dados, via *browser*.

7. Validação e Teste: nessa fase o sistema é testado considerando-se, o máximo possível, as simulações de volume e de processamento. Inicialmente o sistema deverá ser utilizado para um grupo restrito de usuários e, após análise do *feedbacks*, o sistema poderá ser entregue ao ambiente produtivo.

8. Treinamento: o treinamento deverá ser realizado para o grupo de usuários voltados para atividades de negócios, além de gerentes das áreas envolvidas.

9. Implantação: essa etapa deverá ser seguida de um acompanhamento rigoroso de uso das aplicações disponibilizadas, sendo que a equipe do projeto deverá incentivar os usuários a apresentarem críticas e sugestões de melhorias para as próximas versões do sistema.

Além destas etapas, é necessária a construção do diretório de metadados do projeto, que deverá conter os metadados fundamentais para descrever os dados do modelo Fonte, suas transformações, os dados do modelo Dimensional e suas formas

de acesso e de disponibilização, e também um *front-end* amigável, que permita ao usuário obter as informações necessárias ao uso do sistema.

2.3.2.3 ODS (*Operational Data Store*)

De acordo com Barbieri (2001), enquanto DW e DM estão relacionados com estruturas dimensionais de dados, remodeladas com o objetivo de prover análises diferenciais, o ODS (*Operational Data Store*/Armazenamento de Dados Operacionais), conceitualmente refere-se ao armazenamento e tratamento de dados operacionais, de forma também consolidada, porém sem as características dimensionais.

Para o autor, ODS pode ser entendido como um cadastro consolidador de informações, no qual são mantidas ainda as características de granularidade e de estruturação não-dimensional, originada dos sistemas legados e ERP (*Enterprise Resourcing Planning*), sendo que seu conceito nasceu como uma solução intermediária entre os muitos arquivos e dados espalhados pela empresa, que precisavam de uma certa uniformização e a proposta final de DW e DM. Portanto, pode se dizer que o ODS é a metade do caminho entre o legado e o DW, oferecendo também importantes informações ao processo decisório, devido a sua característica de consolidação e integração de várias fontes de dados.

Segundo Inmon (1997b), o ODS é a extensão do DW ao mundo dos ambientes operacionais, sendo que não se aplica ao processamento informacional fora do ambiente das operações do dia-a-dia. Existem muitas diferenças importantes entre o ODS e o DW, que estão especificadas no Quadro 2.

Ainda, conforme Inmon (1997b, p. 44), a definição do ODS, poder ser “uma coleção de dados atuais ou quase atuais para o suporte das decisões operacionais detalhadas do dia-a-dia; orientado a assunto; integrado; volátil (i.e. passível de atualização)”.

Portanto, pode se concluir que o ODS se difere basicamente do DW quanto à volatilidade, onde dados no DW não são voláteis e no ODS são; quanto ao conteúdo, onde no ODS os dados são atuais ou quase atuais e no DW os dados são históricos; e

especificamente o ODS refere-se ao domínio operacional, já o DW refere-se ao domínio estratégico da organização.

QUADRO 2. DIFERENÇAS ENTRE O ODS E O DW

ODS	DW
Valores atuais ou quase atuais dos dados	Valores históricos e quase atuais dos dados
Dados detalhados, quase que, exclusivamente	Conjunto rico de dados detalhados e resumidos
Capaz de ser atualizado	Sem atualizações
Valores correntes homogêneos dos dados (dados atuais)	Dados heterogêneos (altamente resumidos, levemente resumidos, detalhe atual, detalhe antigo)
Tecnologia requer um ambiente orientado à atualização de registros, totalmente funcional	Tecnologia simples de carga-e-acesso
Atende a um público de escritório ou tomadores de decisão corporativos no dia-a-dia	Atende a um público de analistas de SAD ou comunidade orientada ao gerenciamento
Usado para decisões “instantâneas”	Análises de longo prazo ou determinação de tendências

Fonte: Adaptado Inmon, 1997b.

2.3.2.4 Data Mining

Harrison (1998), afirma que o Data Warehouse fornece memória à empresa, porém a memória tem pouco uso sem a inteligência, a qual permite a análise da memória observando modelos, estabelecendo mecanismos e tendo novas idéias para fazer previsões. A inteligência da empresa é proporcionada pelo *Data Mining*, através dele é possível explorar gigantescas montanhas de dados gerados pelas suas interações e conhecê-los melhor.

Data Mining (mineração de dados) pode ser conceituado, então, como um conjunto de técnicas utilizadas de maneira automática para explorar exhaustivamente e tornar claras relações complexas existentes em bancos de dados de grandes dimensões. Estas técnicas são baseadas em descobertas por meio de procura de padrões de dados, através de uma série de algoritmos para achar relações fundamentais entre os dados, que se caracterizam por serem automáticos, ou seja, são de baixa interação com o ser humano, restringindo sua participação aos processos de ajuste necessários aos mecanismos de busca (Strech, 1998).

Os dados analisados em Bancos de Dados podem prover conhecimento adicional sobre um negócio, objetivando benefícios para qualquer empreendimento. Ou seja, a tecnologia *Data Mining* visa explorar grandes BD para obter de forma automática, valiosas informações que poderão causar diferenciais efetivos no negócio. Para se identificar/garimpar estas informações valiosas em grandes BDs, ou ainda no DW/DM, são utilizadas técnicas de Inteligência Artificial, redes neurais, sistemas baseados em regras, raciocínios baseados em casos e algoritmos específicos, como: árvores de decisão, vizinho mais próximo.

As técnicas de *Data Mining* podem ser aplicadas em diversas áreas, como financeiras, científicas, governamentais e comerciais. Utilizando por exemplo, técnicas de *Data Mining* em aplicações comerciais, é possível identificar as preferências do consumidor e seus padrões de compra. De posse de tais informações, pode-se direcionar campanhas de vendas, programar estoques, avaliar o comportamento do mercado e detectar novas tendências mercadológicas ou necessidades de consumo, dentre outras.

Normalmente, no processo de *Data Mining*, a busca é iterativa, de modo que à medida que os analistas examinam as saídas, um novo conjunto de perguntas é reformulado para refinar a busca ou elaborar algum aspecto das descobertas. Terminado o processo de busca iterativo, o processo de *Data Mining* gera relatórios de descobertas, de onde é tarefa de humanos interpretar os resultados do processo de mineração e tomar iniciativas com base nessas descobertas (Hedberg,1995).

As principais técnicas de descoberta do conhecimento que podem ser empregadas ao processo de *Data Mining* são na definição de Strech (1998):

- Clusterização: os algoritmos de clusterização transformam registros com grande número de atributos automaticamente em conjuntos relativamente menores, chamados de segmentos, por meio da identificação das características que distinguem o conjunto de dados e pelo posterior particionamento dos mesmos. São utilizados em casos como de modelos de segmentação de população, como segmentação demográfica de mercados de consumidores, implicando em possível comparação dos hábitos de consumo de múltiplos segmentos da população visando determinar campanhas de vendas.
- Classificação: as técnicas de classificação criam automaticamente um modelo a partir de um conjunto inicial de registros, que é chamado de conjunto de treinamento e serve de exemplo, e que devem pertencer a um pequeno grupo de classes predefinidas pelo analista. O modelo é composto de padrões, essencialmente generalizações em relação aos registros, os quais são usados para diferenciar as classes, que uma vez obtido, é usado para classificar automaticamente os demais registros. São utilizados em aplicações típicas de cartão de crédito, onde dada uma base de dados dos usuários de cartões cujo histórico de crédito é conhecido atribui-se a estes um grau de risco de acordo com seu histórico (alto, médio ou baixo), então por meio das técnicas de classificação têm-se a caracterização do grau de risco de um usuário visando dados como faixa de renda familiar, idade, área de moradia, por exemplo: usuários de 25 a 30 anos com renda mensal superior a R\$ 1000,00, moradores da região X apresentam baixo risco.
- Associação: o objetivo da técnica é encontrar tendências, a partir de grande número de transações, que possam entender e explorar padrões de compra. Através de técnicas de associações são realizadas operações sobre um conjunto de registros retornando afinidades entre a coleção de itens expressas por um conjunto de regras da forma “um determinado percentual x dos registros que contém os itens A,B e C também contém os itens D, E e F”, sendo que A, B e C podem corresponder a qualquer número de itens (o mesmo

para D,E e F) e onde o fator de confiança é o percentual x. Por exemplo: 80% das vendas de cervejas também correspondem a vendas de batatas fritas.

Análise por meio de Seqüências: uma operação que envolva seqüências pode detectar padrões de compra de produtos com o decorrer do tempo, sendo que em operações de associação, a coleção de itens corresponde a uma operação de compra para um ponto específico no tempo, enquanto que em operações com seqüências, estas correspondem a um intervalo de tempo sendo analisado, onde é relevante não só a coleção de itens, mas também a ordem e o intervalo de tempo entre as compras. Por exemplo, observando procedimentos médicos aplicados com freqüência ao longo do tempo podem ser detectados casos de erro médico, ou ainda, em aplicações de vendas no varejo, poderia ser utilizada uma seqüência de compras que pudesse prever a posterior compra de determinados itens assim como campanhas de mala direta.

- Visualização: esta técnica não é propriamente de *Data Mining*, mas é uma técnica utilizada para analisar e observar os dados de uma determinada base de dados, pois oferece meios de obter sumários visuais dos dados. Pode ainda ser usada como um mecanismo de compreensão da informação extraída por meio das técnicas de *Data Mining*, pois se observadas graficamente características difíceis de detectar pela simples observação de linhas e colunas podem se tornar óbvias.

O *Data Mining* é um dos passos do processo de KDD (*Knowledge Discovery In Database*) - Descoberta de Conhecimento em Bancos de Dados, pois para extrair o que é julgado conhecimento de acordo com especificações, o processo de KDD utiliza algoritmos de *Data Mining* (Fayyad, 1996).

Strech (1997), informa que extrair informação de um DW/DM para aplicá-la ao processo de tomada de decisão envolve diferentes fases, que são apresentadas a seguir:

Seleção de dados: nem todos os dados encontrados em um DW/DM são necessários para realização de buscas de ferramentas de *Data Mining*. Portanto, o primeiro passo é selecionar os tipos de dados que serão usados, então agrupá-los em múltiplas tabelas sendo talvez necessária a realização

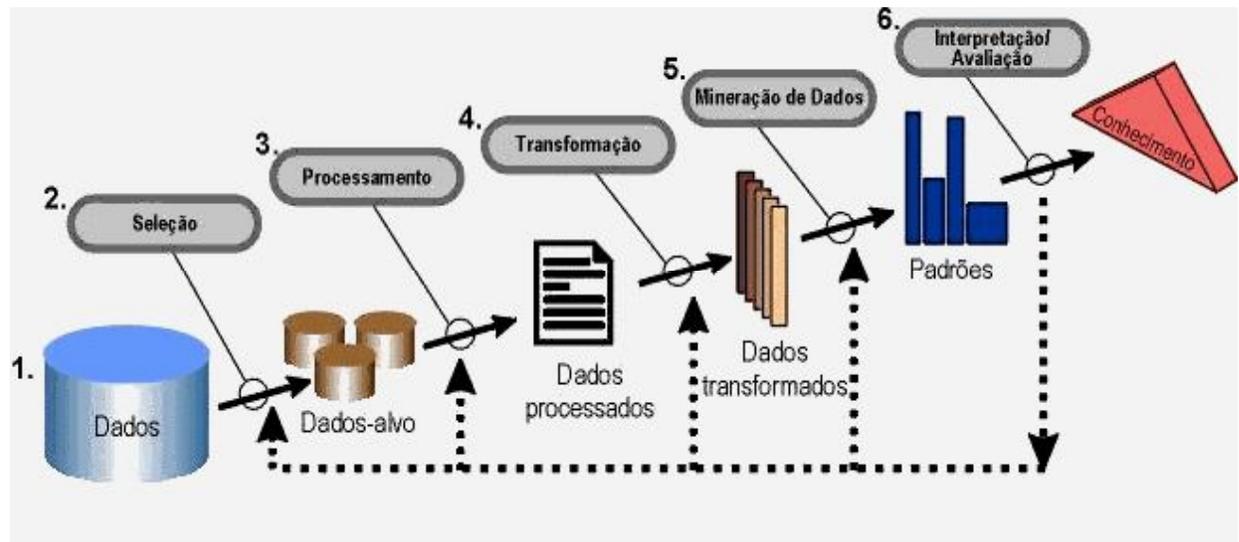
de *joins*. Mesmo após a seleção da tabela desejada pode ser desnecessário aplicar as ferramentas de *Data Mining* sobre toda a tabela, pois dependendo da técnica a ser empregada basta que se use uma amostragem.

Transformação de dados: uma transformação dos dados é geralmente necessária após a seleção das tabelas e a identificação dos dados que devem ser manipulados, que pode variar de acordo com o tipo de técnica a ser utilizada. Pode haver necessidade de conversão de um tipo de dado para outro, como por exemplo, conversão de valores literais para valores numéricos.

- *Data Mining*: após a transformação dos dados, estes são manipulados por uma ou mais técnicas de *Data Mining* para a obtenção da informação desejada, podendo ser necessário acessar dados adicionais do DW/DM e/ou realizar transformações adicionais sobre os dados relacionados.

Interpretação dos resultados: extraída a informação, então esta é analisada para ser aplicada ao processo de tomada de decisão.

O diagrama da Figura 8 resume os estágios/processos identificados na descoberta do conhecimento em Bancos de Dados (DW/DM) por Usama Fayyad (1996), um dos principais conhecedores desta área.



Fonte: Usama Fayyad, 1996, p. 10

FIGURA 8. ETAPAS DO PROCESSO DE KDD.

2.3.2.5 OLAP (On-Line Analytical Processing)

No início dos anos 90 começaram a surgir os primeiros protótipos de SGBDs (Sistema de Gerenciamento de Bancos de Dados) Multidimensionais. Após alguns anos de aprimoramento da tecnologia, os SGBDs Multidimensionais foram submetidos à análise e então foram batizados com o nome OLAP (*On-Line Analytical Processing*, derivado do termo OLTP – *On-Line Transactional Processing* – que foi atribuído aos bancos de dados relacionais no início da década de 70, quando foi definido os padrões para modelo relacional).

A tecnologia OLAP foi definida em 1993, em decorrência da forte necessidade de análises dos dados de forma fácil e flexível, mas ao mesmo tempo, analisando múltiplas visões do negócio em diferentes níveis de detalhes. Os SGBDs Multidimensionais foram a resposta para atender a essas necessidades analíticas (Harrison, 1998).

OLAP representa um conjunto de tecnologias projetadas para suportar análises e consultas *ad hoc*, objetivando ajudar analistas e gerentes a sintetizarem informações sobre a empresa através de comparações, visões personalizadas, análise histórica e projeção dos dados em vários cenários de ‘e se...’ (Informativo Técnico no. 52, 2002).

Através de suas funções mais complexas, que incluem um conjunto robusto de capacidades computacionais e de navegação dos dados, o OLAP permite que os usuários executem funções de análise de dados passando livremente pelas várias dimensões do DW/DM (Harrison, 1998).

Dessa forma, os usuários podem estudar os dados de maneira multidimensional, de modo que os mesmos podem ‘perfurar’ os dados até os seus detalhes (função comumente chamada de *drill dow*), ou ainda ver porções sumarizadas desses dados (função *slice-and-slice*), do ponto de vista que desejarem, enquanto ‘perseguem’ as respostas que procuram. Assim essa função permite que o usuário veja os dados de várias e diferentes perspectivas, e a numerosos níveis de detalhe ou agregação (DWBrasil, 2002).

Para Cielo (2000), a característica *slice and slice* das ferramentas OLAP possibilita analisar as informações de diferentes prismas limitados somente pela nossa imaginação. Já com a característica *drill-down/up* é possível fazer uma exploração em diferentes níveis de detalhe das informações, subindo ou descendo dentro do detalhamento do dado, como, por exemplo, analisar uma informação tanto diariamente como anualmente, partindo da mesma base de dados, pois o conceito de *drill down* está diretamente relacionado com o fato de sairmos de um nível mais alto da hierarquia e buscarmos informações mais detalhada, ou seja, em níveis menores, já o conceito de *drill-up* é o inverso.

Segundo Ikematu (1998), a principal diferença entre OLAP e *Data Mining* é como eles operam o dado. A tecnologia OLAP mostra os dados para análise em várias dimensões e níveis de agregação, já o processo de *Data Mining* apresenta proporções e padrões de comportamento dos dados.

Barbieri (2001), esclarece que enquanto as técnicas OLAP objetivam trabalhar os dados existentes, buscando consolidações em vários níveis, trabalhando fatos em dimensões variadas, a técnica de *Data Mining* busca algo a mais que a interpretação dos dados existentes. Através de inferências, tenta descobrir possíveis fatos e correlações não explicitadas em um DW/DM. Por exemplo, numa empresa de crédito, as técnicas de OLAP produziriam gráficos mostrando os percentuais comparativos de compras com cartões de crédito roubados e válidos, já as ferramentas de *Data Mining* indicariam os padrões associados a certo comportamento fraudulento com cartões de crédito.

A diferença básica entre OLAP e *Data Mining* está na maneira como a exploração dos dados é realizada. Na análise OLAP a exploração é feita através da verificação, isto é, o analista conhece a questão, elabora uma hipótese e utiliza a ferramenta para refutá-la ou confirmá-la. Com o *Data Mining*, a questão é total ou parcialmente desconhecida e a ferramenta é utilizada para a busca de conhecimento (DWBrasil, 2002).

O Quadro 3 mostra em resumo, algumas das principais diferenças entre as tecnologias OLAP e *Data Mining*.

QUADRO 3. AS PRINCIPAIS DIFERENÇAS ENTRE OLAP E *DATA MINING*.

OLAP	Data Mining
Objetiva trabalhar os dados existentes	Busca algo a mais que interpretação dos dados
Mostra os dados para análise em várias dimensões e níveis de agregação.	Apresenta proporções e padrões de comportamento dos dados
Resultado de uma sessão são dados calculados e a questão é conhecida, a exploração é feita através da verificação, elaborando-se uma hipótese e utilizando a ferramenta para refutá-la ou confirmá-la.	Resultado de uma sessão é uma regra e a questão é total ou parcialmente desconhecida, e a ferramenta é utilizada para busca de conhecimento.
O resultado é a descoberta da razão de um problema conhecido. Para isto é necessário conhecer a base de dados.	Tenta-se descobrir uma nova oportunidade. Para isto é necessário conhecimento da base de dados e, em muitos casos, conhecimentos matemáticos e estatísticos.

Fonte: autora.

A principal característica do OLAP é permitir uma visão conceitual multidimensional dos dados de uma empresa, isto é, o tomador de decisão pode a partir de qualquer dimensão do DW/DM iniciar uma análise, navegando então para outras dimensões para analisar posteriormente as informações.

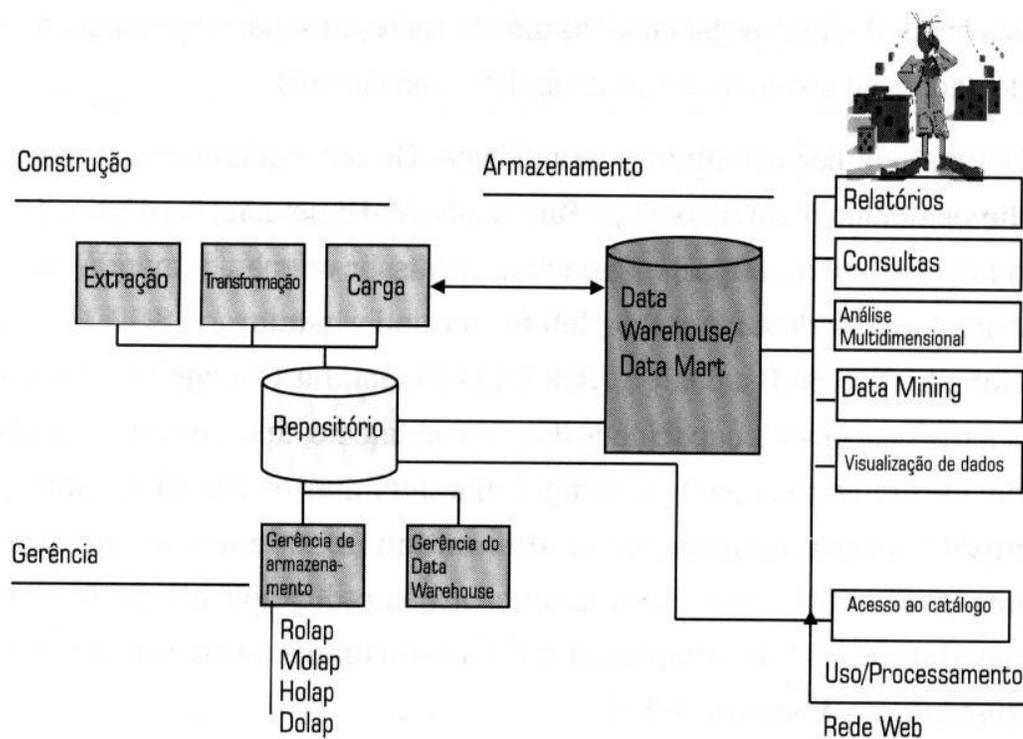
2.4 Visão Geral do Mercado e das Ferramentas de BI

Barbieri (2001), afirma que de maneira geral, as ferramentas para um ambiente BI podem ser classificadas como de Construção, Gerência, Uso e Armazenamento, de acordo com a Figura 9 e descrito a seguir:

Ferramentas de Construção: auxiliam no processo de Extração de dados das diversas fontes, seu tratamento de preparação, transformação e sua carga nas estruturas finais do DW/DM. Realizam processos de classificação de fontes diferentes, facilitando a busca em ambientes heterogêneos, a sua transformação através de procedimentos escritos, ou tabelas de pesquisas e permitindo a sumarização para os níveis de granularidade desejados.

Ferramentas de Gerência: auxiliam no processo de armazenamento e de utilização dos DW/DM e do Repositório, onde residem as informações de metadados, responsáveis pela definição das estruturas e dos processos de transformação desejados.

- Ferramentas de Uso: são através destas ferramentas que os usuários manipulam os dados no DW/DM e obtêm as informações requeridas. Oferecem um arsenal de operadores, como *drill-down*, *drill-up*, *drill-cross*, *drill-through*, etc. e trabalham em interfaces Web.



Fonte: Barbieri, 2001

FIGURA 9. VISÃO GERAL DAS FERRAMENTAS DE UM AMBIENTE DE BI.

Dentre os produtos que, poderão desempenhar um papel importante nos próximos anos, na aplicação de conceitos de BI nos seus respectivos domínios, segundo Barbieri (2001) destaca-se:

O BW (Business Information Warehouse) da SAP, com o intuito de se acoplar ao sucesso do ERP-SAP R/3, presente hoje em oito dentre dez grandes

empresas de qualquer segmento de negócio, mas concentrado no ambiente transacional, enquanto o BW é dedicado exclusivamente ao tratamento de informações gerenciais.

O Oracle Express, produto oferecido pela líder em bancos de dados e também uma das líderes do segmento de Data Warehouse e Data Marts. A Oracle, foi desenvolvido originalmente no MIT (Massachusetts Institute of Technology) nos anos 70, para modelagem de dados de mercado, com sua primeira versão para o ambiente *mainframe*. Ao longo desse período, vem sofrendo alterações, mas mantém a sua essência, centrada no perfil de servidor para análises multidimensionais.

- O Analysis Services da Microsoft, um produto da Microsoft, com a proposta de um OLAP interno no SQL-Server 7.0/2000, popularizando os servidores OLAP, disponibilizando-os, como um agregado do SQL-Server 7.0/2000, para os usuários fiéis da Microsoft.

As empresas de tecnologia estão cada vez mais se preparando para o mercado de BI, que tende a crescer, visto os benefícios que tais sistemas tendem a oferecer quando bem implantados dentro das empresas, as quais diante da concorrência cada vez mais forte, buscam soluções através de tecnologias que proporcionem um maior conhecimento do negócio através da interpretação de grandes quantidades de informação.

2.4.1 Casos: *Business Intelligence*

Caso1: A SomLivre.com, loja virtual da organização SIGLA (Sistema Globo de Gravações Audiovisuais) oferece cerca de 20 mil títulos de CDs de mais de 250 gravadoras, 50 mil livros, DVD's e games entre outros produtos, além de resenhas e faixas de músicas para que o usuário possa ouvir diretamente do site. O site é visitado diariamente por cerca de 30 mil pessoas. Através da ferramenta de BI é possível conhecer melhor quem compra os produtos da SomLivre.com, e desta forma, customizar as ofertas de acordo com o seu perfil, o que é feito graças aos recursos de Data Warehouse e *Business Intelligence*, que além de relatórios

concisos, possibilitam ao site, durante a própria navegação, gerar novas sugestões, oferta de serviços e *cross-selling*. A partir do comportamento de navegação e compra dos usuários, é possível que promoções-relâmpagos possam ser colocadas no ar instantaneamente, proporcionando ao sistema maior funcionalidade e inteligência para análise dos hábitos e preferências dos clientes (Microsoft, 2002).

Caso 2: A Bradesco Seguros decidiu investir em tecnologia de informação a fim de identificar fraudes e abusos de Planos de Saúde, pois segundo alguns estudos e estatísticas, mais de 10% dos custos das seguradoras estão relacionados a fraudes a abusos, desta forma, objetivando aumentar a produtividade, a empresa decidiu intensificar as ações contra esta situação. A solução deveria ser capaz de analisar as atividades fraudulentas, reduzir custos, ganhar lucratividade, aumentar *market share* e ainda maximizar a efetividade e eficácia dos recursos humanos, melhor equipando-os para que conseguissem acompanhar e identificar referenciados que estivessem submentendo solicitações fraudulentas e/ou abusivas, voltadas a seguros de saúde. Foi utilizado uma solução de *Business Intelligence* baseado em análises de comportamento e que emprega técnicas estatísticas padrão, ferramentas de visualização de dados e algoritmos avançados de data mining, proporcionando os seguintes benefícios: identificação proativa de suspeitos (eliminando a necessidade de investigações para descobrir atividades fraudulentas efetuando-se buscas em bancos de dados enormes), ajuste para detecção de padrões de fraude (como as tendências médicas mudam, o sistema se ajusta automaticamente a elas), sistema não-paramétrico (não necessidade de selecionar padrões arbitrários e/ou especificar relacionamentos complexos entre variáveis e dados, mas simplesmente identificar as categorias de risco), necessidade de baixo volume de dados (o sistema não é baseado em técnicas de modelagem matemáticas ou estatísticas complexas, portanto, não requer grandes amostras de dados para detectar potenciais fraudes). Como o sistema possui dados históricos sobre os sinistros é possível criar diferentes *reports* e gráficos apontando situações irregulares, ajudando os auditores a identificar a necessidade de análise de fraude, diretamente de suas estações de trabalho, sem que haja necessidade de utilização de qualquer outra ferramenta (IBM, 2002).

Os casos descritos a seguir foram elencados por Balieiro (2000):

Caso 3: O Banco HSBC, também investiu cerca de 3 milhões de dólares na instalação de um projeto de BI, afim de direcionar melhor as ofertas dos produtos do banco, permitindo que os gerentes de relacionamento identificassem em sua base de clientes os que têm maior propensão para adquirir determinados produtos, assim, ao invés de se oferecer um seguro de vida para todos os correntistas, ocasionado perda de tempo e dinheiro, a proposta pode ser feita apenas aos clientes que têm maior propensão a adquiri-lo, aumentando a eficácia das ofertas do banco.

Caso 4: O Unibanco investiu 7 milhões de dólares em um projeto de BI, para que quando um cliente liga para o *call center* ou acessa um caixa eletrônico, seja automaticamente reconhecido pelo sistema, que avalia se o usuário tem ou não as características para obter um crédito pessoal. Então, se a resposta for positiva, o cliente recebe uma proposta imediata de crédito, fazendo com que o índice de aceitação dos créditos chegue a 50% (índice considerado alto entre as instituições). Dessa forma é possível individualizar a oferta para o cliente do Unibanco.

Caso 5: Porém, não é apenas no relacionamento com o cliente que o BI pode ser utilizado. Na GVT, o projeto já está sendo instalado e será utilizado para fazer uma avaliação dos serviços da operadora, que junto com as informações da empresa está incluindo os indicadores da Anatel, que determinam os serviços que devem ser realizados pelas empresas de telefonia privatizadas. Desta forma, a GVT pretende comparar os seus números com os exigidos pelo órgão, monitorando corretamente o que fazer para atingir as metas, evitando multas. As informações que serão incluídas em um projeto de BI são definidas pelas próprias empresas, sendo que o preço das soluções pode variar de cinquenta mil a milhões de reais, dependendo do tamanho do sistema, já o tempo da implementação normalmente começa, mas nunca termina, pois é sempre preciso incluir um novo dado ou fazer adaptações no sistema.

Para Giurliani (2002), apesar do crescimento do BI no Brasil, ainda existem algumas dificuldades, como a resistência por parte dos profissionais de informática, que apresentam falta de confiança no produto nacional preferindo os programas

estrangeiros, mesmo que não sejam os mais adequados. Porém, mesmo os fornecedores estrangeiros sentem certas dificuldades, pois os profissionais de informática se mostram mais preocupados com detalhes técnicos e de performance, do que com os benefícios da ferramenta, o que para alguns fornecedores trata-se de um problema de miopia e para outros, esses profissionais temem perder o controle sobre os dados e tornarem-se obsoletos.

Outros problemas são as experiências malsucedidas que as empresas tiveram com projetos enormes, mal dimensionados, que custaram milhões de dólares e mais de um ano para serem concluídos, e ainda não trouxeram os resultados esperados, tornando as empresas mais cuidadosas para investir em soluções deste tipo.

No entanto, de maneira geral, o mercado brasileiro já está amadurecendo e nem está tão atrasado em relação aos outros países, inclusive os EUA. Percebe-se também que, a grande maioria dos investimentos em BI são feitos por empresas de grande porte, o que é reforçado pelas reportagens na mídia que focalizam apenas os grandes projetos, dando a impressão de que as médias empresas não precisam desse tipo de solução (Giurliani, 2002).

2.5 Considerações do capítulo

Neste capítulo, apresentou-se uma revisão teórica sobre os sistemas de informações nas organizações, seus conceitos, e sua evolução até se chegar ao conceito de *Business Intelligence*, que são a evolução do Executive Information System (EIS) ou Sistema de Informação Executiva (SIE) e do Sistema de Apoio à Decisão (SAD).

Pode-se dizer que o *Business Intelligence* é uma combinação de todas as ferramentas que servem de suporte à decisão, ou seja, reúne um conjunto de ferramentas e tecnologias para gerenciar informações, bem como, os processos para coleta, armazenamento e distribuição de dados gerenciais.