

A NOVA GERAÇÃO DE SISTEMAS DE APOIO À DECISÃO

Carlos Alberto Ferreira Bispo

Professor da Academia da Força Aérea
Estrada de Aguaf S/N - Pirassununga - SP Caixa Postal 1062 - CEP 13630-000 - cafbispo@sc.usp.br

Prof. Dr. Edson Walmir Cazarini

Professor da Escola de Engenharia de São Carlos (USP)
Av. Dr. Carlos Botelho, 1465 - São Carlos - SP - CEP 13560-970 - cazarini@sc.usp.br

Abstract

This article approaches which are the new tools for Decision Support Systems that they are giving great pulses in the business. We expose basic concepts of the Data Warehouse, of the OLAP (On-Line Analytical Processing) and of the Data Mining, so that they serve, its advantages in the use as strategy of business, and the difficulties of implementation these new technologies.

Área: 10.4 - Sistemas de Apoio à Decisão

Key Words: Decision Support System, EIS (Executive Information Systems), Business Intelligence.

1. Introdução

A concorrência acirrada, a maior exigência de qualidade por parte dos clientes e as dificuldades econômicas que as empresas estão enfrentando estão forçando as empresas a se modernizarem, a serem mais criativas e mais eficientes na solução dos seus problemas. Neste ponto, as Tecnologias da Informação estão dando uma boa contribuição para os gerentes conseguirem enfrentar os seus constantes desafios. Algumas das Tecnologias da Informação se transformaram em ferramentas estratégicas para obter sucesso no gerenciamento dos negócios.

Entre as Tecnologias da Informação, uma nova categoria de Sistemas de Apoio à Decisão está surgindo nesta última década deste milênio. São os sistemas que estão sendo chamados pelos americanos de business intelligence. Os mais importantes são os que vão ser abordados neste artigo, que são o data warehouse, o OLAP e o data mining.

O data warehouse é um banco de dados, ou armazém de dados como está sendo chamado devido as suas características, que serve para dar suporte às necessidades gerenciais de informações. Faz-se necessário um banco de dados exclusivo para os gerentes pelo fato de que as necessidades de informações deles são diferentes das necessidades de informações do ambiente operacional da empresa.

O OLAP (On-Line Analytical Processing) é uma ferramenta para efetuar certos tipos de análises que só são possíveis através desta ferramenta, como por exemplo uma visão

multidimensional do negócio, onde os dados são analisados e comparados por diversos ângulos, possibilitando uma melhor compreensão do próprio negócio.

O data mining é um sistema que permite a descoberta de informações escondidas no banco de dados utilizando técnicas sofisticadas de procura, como algoritmos de Inteligência Artificial e técnicas estatísticas avançadas. Desta forma, é possível se descobrir fatos e informações que podem ser usados como estratégias nos negócios.

2. A evolução da Tecnologia da Informação

A Informática e a Tecnologia da Informação evoluíram muito nas últimas décadas. Assim como os conceitos administrativos, os modelos de negócios e o próprio mercado. Porém, somente recentemente, as duas tecnologias citadas puderam dar um melhor auxílio no mundo dos negócios.

A automação das atividades administrativas e a automatização dos processos produtivos trouxeram grandes benefícios para a área operacional, assim como trouxe bons retornos financeiros aos investimentos realizados. Porém, especificamente para a área de negócios e para a alta gerência da empresa, somente nos últimos anos é que a Tecnologia da Informação começou a trazer benefícios nesta área. Já é possível hoje a um alto gerente, diante de algum evento adverso, poder ter acesso rápido a informações precisas e confiáveis que lhe permitirão tomar as providências cabíveis em resposta àquele evento.

Felizmente, hoje com a nova geração de ferramentas de apoio à decisão, os gerentes possuem boas ferramentas para lhes proporcionar o suporte necessário nos processos decisórios para que enfrentem a concorrência acirrada e as adversidades que ocorrem frequentemente em seus negócios.

3. Data Warehouse

Podemos definir o data warehouse como um grande banco de dados voltado para dar o suporte necessário nas decisões de usuários finais, geralmente gerentes e analistas de negócios. Os dados que compõem o data warehouse são derivados dos diversos bancos de dados operacionais utilizados na empresa, e estão armazenados em um local diferente dos referidos bancos de dados operacionais.

Segundo TAURION (1997) ao reunir informações dispersas nos diversos bancos de dados operacionais da empresa que podem estar em plataformas distintas, o data warehouse permite que sejam feitas consultas e análises bastante eficazes, transformando dados esparsos em informações antes inacessíveis ou subaproveitadas. Essas informações podem ser convertidas em estratégias para os negócios.

Um dos grandes benefícios proporcionados pelo data warehouse é a diminuição do tempo que os gerentes levam para obter as informações necessárias aos seus processos decisórios com a eliminação de tarefas operacionais como pesquisa e identificação dos dados necessários. O data warehouse contém somente os dados necessários aos gerentes para realizarem as referidas tarefas.

O objetivo do data warehouse deve ser o de satisfazer as necessidades de análises de informações dos seus usuários, como por exemplo, monitorar e comparar dados de

transações atuais com as passadas e prever tendências futuras, permitindo tomar as medidas cabíveis ainda a tempo de eventuais correções ou aperfeiçoamentos.

Utilizar os bancos de dados dos aplicativos operacionais para se realizar uma análise ou uma consulta é muito mais trabalhoso, e ainda pode interferir na performance do aplicativo. Outros problemas também podem ocorrer com esta prática, como enfrentar problemas oriundos da redundância de dados dispersos pelos diversos aplicativos, ou encontrar dados que não estão no formato adequado para se realizar as análises necessárias.

“A grande vantagem de um data warehouse é permitir a tomada de decisões baseadas em fatos” (TAURION, 1998). O data warehouse deve tornar-se uma ferramenta imprescindível para os gerentes conseguirem administrar seus negócios nas próximas décadas. Como o ambiente de negócios está se tornando cada vez mais dinâmico, é extremamente necessário que as regras de negócios sejam incorporados às aplicações, que as estruturas dos sistemas se ajustem aos negócios, e que o tempo de resposta dos sistemas seja cada vez menor.

4. OLAP

Segundo FIGUEIREDO (1998) os sistemas OLAP ajudam os analistas e os gerentes a sintetizarem as informações sobre a empresa através de comparações, visões personalizadas, análises estatística, previsões e simulações. São implementados para ambientes multiusuário e arquitetura cliente/servidor e oferecem respostas rápidas e consistentes às consultas iterativas executadas pelos usuários, independentemente da complexidade da consulta.

Segundo FORSMAN (1997) um sistema OLAP aumenta a produtividade dos gerentes da organização inteira, a sua flexibilidade significa que os gerentes podem se tornarem mais auto-suficientes. Permitem aos gerentes modelar problemas que seriam impossíveis usando sistemas menos flexíveis, com um tempo de resposta longo, ou incompatíveis com a estrutura analítica dos negócios.

Sistemas OLAP podem ser usados em diversas funções organizacionais. Departamentos de Finanças usam OLAP para planejar orçamentos e realizar análises financeiras. O Departamento de Vendas o utiliza nas análises e estimativas de vendas. O Departamento de Marketing o utiliza em pesquisas e análises de mercado, estimativas, análises de clientes e segmentação de mercado. A Manufatura o utiliza para o planejamento, análises da produção e análises de falhas ou defeitos. Porém, os sistemas OLAP somente serão realmente úteis, se conseguirem fornecerem as informações necessárias, de um modo absolutamente confiável, e com um tempo de resposta compatível com as necessidades de quem está necessitando dessas informações.

Segundo FORSMAN (1997), RADEN (1995) e FIGUEIREDO (1998) as aplicações OLAP necessitam ter as seguintes características:

- ⇒ Permitir a visão multidimensional dos dados;
- ⇒ Possuir inteligência de tempo;
- ⇒ Fazer a atualização do banco de dados;
- ⇒ Realizar cálculos complexos ;
- ⇒ Criar agregações e consolidações;
- ⇒ Fazer previsão, análise de tendências, otimização e análises estatísticas;
- ⇒ Construir cenários a partir de suposições e fórmulas aplicadas pelos analistas aos dados históricos disponíveis;

⇒ Fazer cálculos e manipular dados através de diferentes dimensões;

CODD *et al.* (1998) criaram, em 1993, doze regras que servem para avaliar produtos OLAP. Na verdade, essas regras são propriedades que os sistemas OLAP devem apresentar. Essas regras logo se espalharam pelo mercado e hoje servem como referência para desenvolvedores de sistemas e usuários. Infelizmente, não nos é possível, neste espaço, descrever e discutir as seguintes doze regras: 1. Visão Conceitual Multidimensional, 2. Transparência, 3. Acessibilidade, 4. Desempenho, 5. Arquitetura Cliente/Servidor, 6. Dimensionalidade genérica, 7. Manipulação dinâmica da matriz esparsa, 8. Suporte multiusuário, 9. Irrestritas operações dimensionais cruzadas, 10. Manipulação intuitiva de dados, 11. Relatórios flexíveis, 12. Dimensões e níveis de agregação ilimitados

Existem dois tipos de tecnologias OLAP, o ROLAP e o MOLAP. O ROLAP - Relational OLAP - trabalha com banco de dados relacionais e com a linguagem SQL para realizar as suas consultas e análises. A grande vantagem do ROLAP é de trabalhar com bancos de dados que possuem um modelo (relacional) que é conhecida por todos. Outra vantagem é o limite de armazenamento de dados é a própria capacidade dos dispositivos de armazenamento, ou seja, praticamente ilimitada. A desvantagem do ROLAP é quando se deseja realizar análises multidimensionais, ou seja, análises onde se verifica os diversos ângulos e as diversas dimensões a que os negócios estão sujeitos. As tabelas dos banco de dados relacionais possuem apenas duas dimensões, e quando queremos ter uma visão que utilize mais de duas dimensões, é necessário se realizar diversos e complexos processamentos de dados para se obter o resultado desejado, o que torna a consulta ou análise mais demorada, e ainda exige-se equipamentos mais robustos.

O MOLAP - Multidimensional OLAP - trabalha com banco de dados multidimensionais, o que facilita as pesquisas e as análises multidimensionais, e permite uma modelagem dos dados muito parecida com os modelos dos negócios. Desta forma, podemos fazer, por exemplo, uma pesquisa ou uma análise do tipo: verificar as vendas realizadas pela empresa por produto, por mês e por loja, ou do tipo saber a rentabilidade de um banco por funcionário, por mês e por aplicação. A grande vantagem de se trabalhar com a tecnologia MOLAP é rapidez nestes tipos de consultas ou análises. A desvantagem é que esta tecnologia não é aberta, ou seja, cada fornecedor desta tecnologia possui a sua própria estrutura, e é um pouco mais difícil de se fazer a sua integração com outras tecnologias. Outra desvantagem é que devido a estrutura multidimensional do armazenamento dos dados é necessário um espaço muito maior para o armazenamento de dados e, normalmente, ocorrem os dados esparsos, ou seja, células do banco de dados que não possuem dados devido ao fato de não existirem transações que preencham aqueles dados. No exemplo acima da rentabilidade por funcionário por mês e por aplicação, se algum funcionário não trabalhar com algum tipo de aplicação não haverá dados para serem armazenados referentes esta aplicação, porém o seu espaço no banco de dados está assegurado mas não será preenchido. A maior parte dos bancos de dados multidimensionais armazenam os índices na memória para melhorar o desempenho e, atualmente, isto limita o tamanho destes bancos de dados a no máximo alguns gigabytes.

5. Data Mining

O termo “data mining” ou “mineração de dados” ou ainda “garimpagem de dados” é o processo de extrair informações potencialmente úteis a partir de dados “crus” que estão armazenados em um data warehouse ou nos bancos de dados dos sistemas transacionais.

“Data Mining é o processo de descoberta de novas correlações, padrões e tendências entre as informações de uma empresa, através da análise de grandes quantidades de dados armazenados em bancos de dados usando técnicas de reconhecimento de padrões, estatísticas e matemáticas” (NIMER & SPANDRI, 1998). Através do data mining é possível, por exemplo, se descobrir o perfil de diversas categorias de clientes, o perfil de suas compras e até alguns dados sobre seus hábitos.

Temos um padrão quando duas colunas de uma tabela compartilham os mesmos dados, por exemplo, quando diversos clientes adquiriram o mesmo produto ou o mesmo conjunto de produtos. O data mining pode fazer esta descoberta automaticamente, sem fornecermos nenhuma informação ao sistema. O data mining estuda o nível de agrupamento de cada coluna e elimina as colunas que não formam agrupamentos ou formam agrupamentos muito pequenos. É possível, também, se fazer as configurações necessárias para ajustarmos o nível de relacionamento entre as colunas. Por exemplo, podemos configurar o sistema para descobrir padrões que ocorram em um nível superior a 5%, neste caso os padrões que representem um valor inferior a 5% de relacionamento entre as colunas serão desprezados. Um bom sistema de data mining deve permitir, também, que se elimine do processo de descoberta os relacionamentos óbvios, como por exemplo, relacionar as cidades onde residem os clientes com os seus respectivos estados e CEPs.

Segundo MENCONI (1998) o data mining trabalha com um conjunto de técnicas estatísticas avançadas e princípios de Inteligência Artificial para identificar os padrões de compras e de comportamento dos consumidores. Com base em dados a respeito dos hábitos de consumo dos clientes, seus hobbies e dados sobre suas transações comerciais e financeiras, é possível traçar associações que revelem grandes filões de mercado.

“Nenhuma técnica de análise pode substituir a experiência e o conhecimento nos negócios e seus mercados. Pelo contrário, a mineração de dados torna o conhecimento e a experiência mais importante que nunca” (SMALL, 1997). Hoje, com esta nova ferramenta, a possibilidade de se aprofundar no conhecimento sobre os negócios é muito maior. Porém, é necessário já se conhecer muito bem os negócios e utilizar esta ferramenta para se aprofundar mais ainda neste conhecimento, passando a conhecer novos fatores que auxiliarão os gerentes nos desafios que enfrentam no gerenciamento dos seus negócios.

O processo de descobrimento realizado pelo data mining pode ser utilizado a partir de sistemas transacionais. Porém, é muito mais eficiente utilizá-lo a partir de um data warehouse onde os dados já estão sem erros, sem duplicidade, são consistentes e habilitam descobertas mais abrangentes e precisas.

O data mining oferece algumas funções muito sofisticadas, porém a tecnologia utilizada está embutida no software e os usuários ao utilizá-lo não precisam ser experts em técnicas estatísticas avançadas ou em Inteligência Artificial.

Quando formos adquirir um sistema de data mining, a primeira coisa a se avaliar é saber que tipos de modelos ele constrói e se esses modelos atendem às necessidades dos nossos negócios. Combinando os conceitos apresentados por SIMOUDIS (1995) e por EDELSTEIN (1996 e 1997) levantamos os principais modelos que os sistemas de data mining constroem para realizarem suas descobertas:

1. **Modelo de descoberta de associação.** Um modelo de associação poderia descobrir que quando qualquer cliente que compra um item **X**, em 65% das vezes ele compra também o item **Y**, a menos que haja uma promoção na qual o item **Y** é comprado em 85% das vezes.
2. **Modelo de descoberta de sucessão.** Por exemplo, se uma casa é comprada, então em 60% das vezes um novo fogão será comprado dentro de um mês e em 45% das vezes um novo refrigerador será comprado dentro de dois meses. Pode-se, então providenciar para que não faltem esses produtos, e um contato devidamente planejado com o cliente poderia ser feito.
3. **Modelo de agrupamento.** Segmentam um conjunto de dados em grupos diferentes cujos os itens são semelhantes.
4. **Modelo de classificação.** Reconhece padrões que descreve o agrupamento ao qual um determinado item pertence.
5. **Modelo de regressão.** Usam uma série de valores existentes para prever que outros valores.
6. **Visualização.** Usado para compreender (através de uma melhor visualização) a informação adquirida por outros modelos de mineração de dados.

Segundo SIMOUDIS (1995) transformar dados em informação requer um processo de quatro passos:

1. **Seleção.** Seleção dos tipos de dados que serão utilizados.
2. **Transformação.** Os dados normalmente sofrem transformações convertendo um tipo de dados em outro (valores nominais em numérico) ou transformando seus atributos.
3. **Mineração.** O dados são minerados usando as técnicas necessárias.
4. **Interpretação dos resultados.** Se os resultados apresentados não são satisfatórios, pode ser necessário repetir o processo da mineração de dados, ou fazer repetidas iterações em alguns passos. A informação resultante deve ser útil, confiável e compreensível.

Segundo EDELSTEIN (1996) são essas as principais técnicas que estão sendo utilizados pelos sistemas de data mining:

1. **Redes neurais.** São coleções de nós conectados que possuem uma camada de entrada e outra de saída, e que realizam diversos processamentos em seus nós. A rede neural é capaz de aprender sozinha; através de um treinamento apropriado. Em cada passo do treinamento o conjunto de dados de saída é comparado com um resultado já conhecido, se for diferente, uma correção é calculada e processada automaticamente nos nós na rede. Estes passos são repetidos até que uma condição de parada seja atingida, ou quando um número de correções estipulados é alcançado.

2. **Árvores de decisão.** Dividem os dados em subgrupos baseado nos valores das variáveis. O resultado é uma hierarquia de declarações tipo “Se ... então ...” que são utilizadas principalmente para classificar dados.
3. **Regra de indução.** Criam conjuntos não hierárquicos de condições que podem se sobrepor. Por exemplo, um software de um provedor de soluções faz as regras de indução gerando árvores de decisão parciais e as combina com técnicas estatísticas para escolher qual é o melhor resultado.

Segundo EDELSTEIN (1996) alguns produtos combinam as características de redes neurais e de árvores de decisão em uma tentativa de se construir um modelo mais preciso e mais rápido.

6. A independência das novas Tecnologias da Informação

Estas três tecnologias podem ser implementadas separadamente e independentemente. Em algumas empresas, a implantação do data warehouse já é o suficiente para dar o suporte necessário no gerenciamento dos negócios. A simples separação do banco de dados com fins corporativos dos bancos de dados dos aplicativos operacionais já possibilita um grande auxílio aos gerentes.

Existem empresas que implantaram a tecnologia OLAP sem a implantação de um data warehouse, o é recomendado mas não é obrigatório. Estas empresas possuem um servidor exclusivo para o OLAP e sempre que necessário se realizar uma consulta ou uma análise, os dados são buscados nos bancos de dados dos aplicativos operacionais, são feitos os ajustes necessários e realizada a consulta ou a análise necessária.

Também existem empresas que estão utilizando a tecnologia de mineração de dados independentemente do data warehouse. As dificuldades aumentam, os resultados podem se tornarem menos expressivos, mas não é obrigatório a utilização do data mining em conjunto com o data warehouse.

7. Conclusão

Com essa nossa explanação quisemos abordar os conceitos básicos das três ferramentas que constituem a mais recente geração de Sistemas de Apoio à Decisão. Todas as três ferramentas são novas e ainda existe muito a se desenvolver e a se aprender sobre elas. Porém, as empresas que já implementaram algumas ou as três ferramentas estão satisfeitas com os resultados alcançados. A nossa intenção com este artigo é divulgar um pouco destas novas tecnologia, tendo em vista que existe pouca publicação a respeito delas no nosso idioma. Em outros artigos apresentados neste mesmo congresso, fazemos uma explanação um pouco mais aprofundada sobre cada uma das três novas tecnologias. Continuaremos pesquisando a respeito destas tecnologias e em futuros artigos apresentaremos mais detalhes das mesmas. Quaisquer comentários, críticas, elogios ou sugestões a respeito deste artigo serão muito bem aceitos nos endereços eletrônicos dos autores.

Bibliografia

- CODD, E. F.; CODD, S. B.; SALLEY, C. T. (1998). *Providing OLAP (On-Line Analytical Processing) to Users-Analysts: An IT Mandate*. Arbor Software. http://www.arborsoft.com/essbase/wht_ppr/coddc0.html (28 Jan. 1998).
- EDELSTEIN, H. (1996). *Technology How To: Mining Data Warehouses*. Information Week; 8 Jan. <http://techweb.cmp.com/iw/561/61oldat.htm> (30 Dez. 1997).
- EDELSTEIN, H. (1997). *Mining For Gold*. Information Week; 21 Abr. <http://techweb.cmp.com/iw/627/27olmin.htm> (30 Dez 1997).
- FIGUEIREDO, A. M. C. M. (1998) Molap x Rolap: Embate de Tecnologias para Data Warehouse, *Developers' Magazine*, ano 2, n. 18, p. 24-25, fev.
- FORSMAN, S. (1997). *OLAP Council White Paper*. OLAP Council. <http://www.olapcouncil.org/research/whtpapco.html> (01 Fev. 1998).
- INMON, W. H. (1997). *Como Construir o Data Warehouse*. Rio de Janeiro, Campos.
- KIMBALL, R. (1998). *Data Warehouse Toolkit*. São Paulo, Makron Books.
- LAMBERT, B. (1996). Data Warehousing Fundamentals: What You Need to Know to Succeed. *Data Management Review*; March 1996. <http://www.data-warehouse.com/resource/articles/lamber8.htm> (29 Dez. 1997)
- MENCONI, D. (1998). A Mineração de Informações. *Info Exame*, ano 13, n. 144, março, p. 92-93.
- NIMER, F.; SPANDRI, L. C. (1998). Obtendo Vantagem Competitiva com o Uso de Data Mining. *Developers Magazine*, ano 2, n. 18, p. 30-31, fev.
- PARSAYE, K. (1996). *Surveying Decision Support: New Realms of Analysis*. Database Programming & Design; April. <http://www.dbpd.com/parsaye.htm> (27 Dez 1997).
- RADEN, N. (1995). Data, Data Everywhere. *Information Week*, 30 Outubro. http://members.aol.com/nraden/iw_mct01.htm (26 Dez. 1997).
- SIMOUDIS, E. (1995). *Data Mining: A Techonology Comes of Age*. IBM Software Quaterly, Vol. 24. <http://www.software/ibm.com/sq/issues/vol24/datatech.htm> (29 Dez 1997).
- SMALL, R. D. (1997). *Debunking Data Mining Miths*. Information Week; 20 Jan. <http://techweb.cmp.com/iw/614/14oldat.htm> (02 Jan. 1998).
- TAURION, C. (1997). Data Warehouse: Estado de Arte e Estado de Prática. *Developers' Magazine*, ano 1, n. 6, p. 10-11, fev.
- TAURION, C. (1998). Data Warehouse: Vale a Pena gastar Milhões Investindo em um? *Developers' Magazine*, ano 2, n. 18, p. 10-11, fev.