

OLAP: UMA PERSPECTIVA ESTRATÉGICA DE ANÁLISE DE DADOS

Carlos Henrique Miranda Rodrigues¹

Carlos César Oliveira Almeida¹

Eduardo Dias da Rocha¹

Eric Jader Azevedo Costa¹

Resumo:

Dado o crescente volume de dados produzidos pelas empresas, fez-se necessário a criação de ferramentas que proporcionassem uma análise concisa e consistente dos dados armazenados, de maneira a permitir que informações fossem extraídas e utilizadas de maneira estratégica na gestão empresarial. Surge então o OLAP (*On-line Analytical Processing*). Um conjunto de ferramentas que auxilia na análise quantitativa e qualitativa de grandes volumes de dados através de semi-ópticas personalizadas que fazem com que os gerentes, empresários e diretores de empresas possam desfrutar de diferentes perspectivas para tomada de decisões estratégicas.

Palavras Chaves: Análise de Dados. *Business Intelligence*. Informação. Mutidimensionalidade. OLAP.

1 Introdução

A evolução da tecnologia vem possibilitando às corporações dispor de um volume de dados cada vez maior. Paralelo a isso, os profissionais de Tecnologia da Informação (TI) desenvolveram, ao longo do tempo, ferramentas para auxiliar na extração de informações dos bancos de dados, de maneira que fosse possível uma análise personalizada, e eficiente, dos dados, a fim de melhorar o processo de tomada de decisão.

As empresas desfrutam de ferramentas que possibilitam concisa de forma a obter vantagens competitivas frente ao mercado. Esse conjunto de ferramentas pode ser conceituado por *Business Intelligence* (BI), dentre as quais se encontram: *Data Warehouses* (DW), *Data Marts* (DM) - que são técnicas de armazenamento de dados, *Data Mining* (MD), *On-Line Transaction Processing* (OLTP) - atende a manipulação de dados operacionais através de transações - e *On-line Analytical Processing* (OLAP) - conjunto de pesquisas interligadas,

¹ Acadêmicos do curso de Sistemas de Informação da Unimontes.

junto à base de dados (DB), que proporcionam otimização significativa em relação à análise dos dados.

Esse artigo tem como finalidade abordar o uso da ferramenta OLAP, na qual se baseia em uma perspectiva de análise multidimensional. Proporcionando uma consulta com maior flexibilidade e funcionalidade, permitindo ao usuário uma pesquisa ágil, não importando a quantidade de dados a serem manipulados. Para tal é necessário que os dados estejam agrupados em estruturas multidimensionais (cubos) que armazenem medidas quantitativas das informações a serem analisadas, permitindo visualização do problema em questão através de diversas perspectivas distintas, comparando as diferentes variáveis existentes no problema e relacionando-as entre si, o que é muito útil em problemas que envolvem maior número de variáveis.

2 Modelagem multidimensional

Para Rosini e Palmisano (2006. p. 47), o Processamento Analítico *On-line* trabalha com a consulta interativa de dados, de modo que o processo de análise se organize em múltiplos passos, em níveis baixos de detalhes. A multidimensionalidade se dá pelo fato de que os dados podem ser visualizados em diversas faces, causando uma ideia de cubo. Cada uma das faces apresenta uma significação, delimitando o assunto que se deseja analisar.

A forma como os dados são armazenados nas bases de dados é um assunto de grande importância para as organizações, pois influencia no tempo e no modo de inserção, remoção, alteração e busca. Para Macário e Baldo (2005. p.1) o tipo de modelagem mais utilizado atualmente é o Relacional, tratando de aspectos operacionais dos dados, ou seja, a parte transacional dos dados. É nesse ambiente de transações que os administradores do sistema, os projetistas e os usuários inserem os dados de entrada.

Laudon e Laudon (2007. p.151) afirmam que o OLAP possibilite aos usuários obterem “respostas *on-line* a questões específicas”, de forma rápida, mesmo quando os repositórios sejam considerados muito grandes ou a análise considere períodos longos para serem verificados.

No entanto a competitividade empresarial fez com que os empresários e gerentes optassem por ferramentas que pudessem ajudá-los nas tomadas de decisões em relação aos dados que eles possuíam. Desse modo, surge a modelagem que chamamos de Modelagem Dimensional de Dados que tem o foco na tomada de decisões estratégicas com base em consultas que denominamos *ad-hoc*:

São consultas com acesso casual único e tratamento dos dados segundo parâmetros nunca antes utilizados, geralmente executados de forma iterativa e heurística. Isso tudo nada mais é do que próprio usuário gerar consultas de acordo com suas necessidades de cruzar as informações de uma forma não vista e com métodos que o levem a descoberta daquilo que procura (INMON, citado por MUSARDO, 2008).

Segundo Anzanello (2002. p. 3) a visualização é realizada em dados agregados, e não em dados operacionais porque a aplicação OLAP tem por finalidade apoiar os usuários finais a tomar decisões estratégicas. Desse modo temos que não se trata de uma aplicação adequada para efetuar transações operacionais, mas a consulta e análise global de todas as variáveis existentes na situação que está sendo analisada. Os usuários desse tipo de modelo são aqueles que buscam visões estratégicas para o seu negócio como, por exemplo, executivos, analistas, gerentes e administradores.

A Figura 1 apresenta um esquema de técnicas e ferramentas para modelagem de dados, desenvolvido por Leandro Salvador mestre em comunicação e semiótica, pela PUC-SP. O modelo aborda os principais pontos, positivos e negativos, das modelagens Normalizada e Dimensional. A primeira diz respeito à Modelagem Normalizada que se preocupa mais com o desempenho do processamento operacional de acordo com os dados que serão inseridos na base de dados. Já a segunda, a Modelagem Dimensional, possui o foco no desempenho da recuperação de dados, ou seja, preocupa-se com uma maneira de armazenar os dados para que futuramente eles possam estar acessíveis de forma ágil e dinâmica. Por fim a terceira vertente das Técnicas de Modelagens são algumas ferramentas que são utilizadas para que

esse armazenamento seja feito de acordo com as necessidades que forem apresentadas pelo problema.

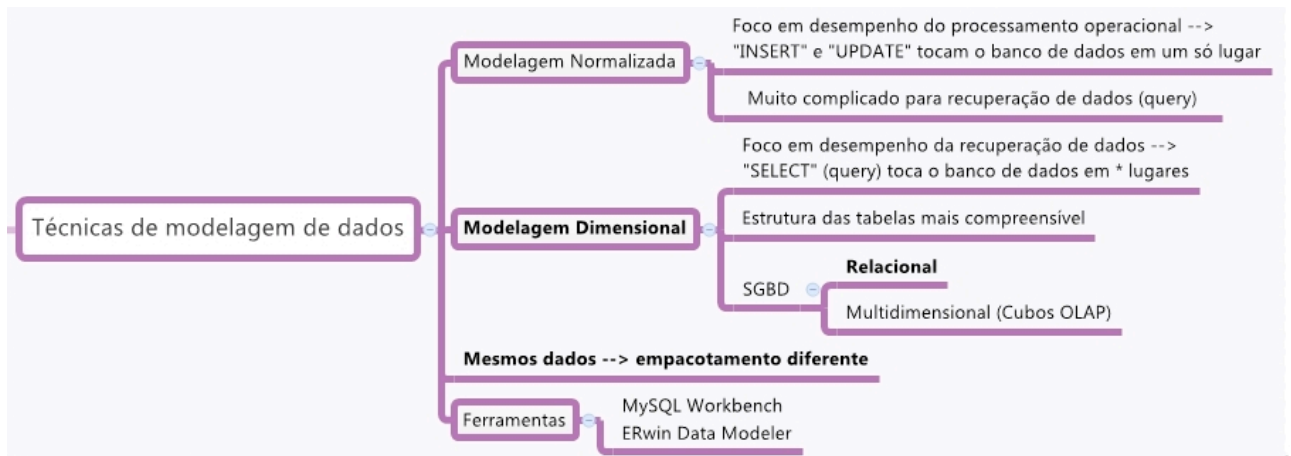


Figura 1 - Técnicas de modelagem de dados

Fonte: Modelagem Dimensional de Dados, Leandro Salvador, Consegi 2011

Segundo Anzanello (2002, p. 2) alguns conceitos característicos devem ser abordados de forma a proporcionar um melhor entendimento da definição de modelagem multidimensional. São eles:

- Medida: é um atributo numérico que representa um fato e serve de base para comparações (porcentagem dos lucros, quantidade de material no estoque);
- Fato: conjunto de medidas dotado de algum contexto (“O índice de assassinatos em Montes Claros-MG aumentou em 20% no período de 2009 a 2010.”);
- Dimensão: são as possíveis formas de visualização dos dados de forma hierarquizada (tempo, produto, fornecedor, cliente);
- Hierarquia: é uma classificação dos membros da dimensão (a dimensão tempo, por exemplo, pode se classificar em hora, dia, mês e ano).

Observa-se com esses conceitos, que modelagem multidimensional dos dados consiste em uma forma de fornecer consultas aos dados a respeito de medidas de desempenho, divididas em uma ou mais dimensões, ou seja, é possível que se faça o cruzamento de dimensões para uma análise personalizada de acordo com os interesses do usuário (Anzanello. 2002. p. 2). A

partir desse tipo de implementação é possível a análise de informações de maneira simples, por exemplo: Quais as vendas mensais de um determinado produto no ano de 1998? Ou ainda, questionamentos mais complexos como: Quais as vendas mensais dos produtos de uma dada marca nos últimos três anos, de acordo com as promoções de Natal?

A partir da definição de multidimensionalidade, bem como de seus conceitos básicos que são importantes para entendimento de como trabalha esse tipo de modelagem. Além disso, estes conceitos e definições citados anteriormente formarão a base para o entendimento de como a técnica de OLAP opera dentro das bases de dados.

3 Caracterizando a ferramenta OLAP

OLAP é uma ferramenta de *Business Intelligence* utilizada para apoiar as empresas na análise de suas informações, visando obter novos conhecimentos que são empregados na tomada de decisão (ARAÚJO, Erika M. T, et al, 2007).

O OLAP é uma solução de ambiente, integração e modelagem de dados voltada para empresas de médio e grande porte, tendo como objetivo auxiliar os analistas, gerentes e executivos na tomada de decisões estratégicas nas organizações. Utiliza processos de análise multidimensional do banco de dados de forma personalizada, ou seja, é possível obter diferentes resultados a partir de uma mesma base, mudando somente seu enfoque (dimensão). Tal característica da ferramenta que a torna um diferencial no mercado, pois através de consultas utilizando ferramentas relacionais de OLTP não seria possível fazer uma investigação tão personalizada, específica, em um espaço de tempo reduzido.

As ferramentas OLAP proporcionam condições de análise de dados *on-line* necessárias para responder as possíveis perguntas dos analistas, gerentes e executivos (MACHADO, citado por ARAÚJO, Erika M T, et al, 2007. p. 2). Para formular a topologia e o projeto de uma solução OLAP multidimensional as seguintes perguntas devem ser feitas: Quando ?, O quê ?, Onde ? e Quem ?. Essas perguntas formam a base de todos os arrays multidimensionais (Anzanello. 2002. p. 3).

Durante o projeto de OLAP a modelagem dos dados é uma preocupação importante de modo que forneça flexibilidade e ótimo desempenho para consultas complexas junto ao banco de dados. O armazenamento de dados pode ser feito tanto em um banco de dados Relacional quanto em um Multidimensional. A grande questão a ser analisada é o que o usuário deseja com a ferramenta: agilidade, detalhamento e carregamento de informações por exemplo. De acordo com essas especificações é que será definida qual a melhor arquitetura a ser utilizada.

3.1 Arquiteturas

Fernandes e Kantorski (2008 p. 2) afirmam que o tipo de armazenamento é que definirá a arquitetura a ser utilizada pelo OLAP. As arquiteturas, segundo esses autores, podem ser divididas em: ROLAP – *Relational* OLAP, MOLAP – Multidimensional OLAP e HOLAP – *Hybrid* OLAP. Por outro lado, Anzanello(2002. p. 4) cita uma outra arquitetura, o DOLAP (*Desktop* OLAP).

No MOLAP os dados são armazenados de forma multidimensional e sua implementação varia de acordo com a ferramenta de OLAP que está sendo empregada, mas frequentemente implementa-se essa arquitetura em uma base de dados relacional, porém não na terceira forma normal. Já no ROLAP os dados são armazenados no modelo relacional como também suas consultas são processadas pelo gerenciador da base de dados relacional. O DOLAP por sua vez é uma variação que existe para fornecer portabilidade dos dados, tendo como característica marcante a redução no tráfego na rede. O HOLAP por sua vez se baseia em uma combinação das principais características das arquiteturas ROLAP e MOLAP: a alta escalabilidade do ROLAP e a alta performance do MOLAP (Anzanello. 2002. p. 4).

As mais utilizadas são a Rolap e Molap. A arquitetura Rolap é a mais indicada para grandes repositórios, como os Datawarehouses, onde os dados estão armazenados de maneira relacional em colunas e tabelas, pelo grande volume de dados e pelo maior número de funções e diversas regras a serem aplicadas. Já o Molap é mais indicado para Data Mining onde os dados são mais específicos e será direcionado na análise com dimensionalidade limitada e pouco detalhamento das informações. Essas duas arquiteturas possuem vantagens e desvantagens, por exemplo: em relação a desempenho de consulta, o Molap é mais rápido,

pois são gerados previamente todas as combinações e resumos possíveis no Modelo Multidimensional, no entanto é uma arquitetura que possui um alto custo de investimento e uma baixa escalabilidade, enquanto no Rolap a maior parte do processamento é feito em tempo de execução, ao passo que possui um baixo desempenho; em relação a carregamento dos dados o Rolap é mais rápido, devido a sua estrutura de linhas e colunas ser menos complexa que a multidimensional do Molap (ARAÚJO et al, 2007. p. 4).

Ainda segundo Araújo *et al* o Holap, que é a arquitetura mais recente, caracteriza-se por ser a junção das vantagens do Rolap e do Molap, ou seja, a escalabilidade do Rolap (capacidade de trabalhar com grandes volumes de dados) e o alto desempenho do Molap. A desvantagem é que essa é a arquitetura de maior custo de aquisição.

Outra arquitetura é a DOLAP que é uma arquitetura desktop do OLAP, ou seja, é uma ferramenta para usuários que possuam uma cópia da base multidimensional ou de um subconjunto dela ou ainda, que queiram acessar um repositório de dados central localmente. O usuário ao acessar este repositório, dispara uma instrução SQL e acessa os cubos já existentes no banco de dados multidimensional residente no servidor OLAP e obtém de volta um para ser analisado em sua estação de trabalho (CAVALCANTI, OLIVEIRA e MONTEIRO, 1998 apud ARAÚJO et al, 2007. p. 4).

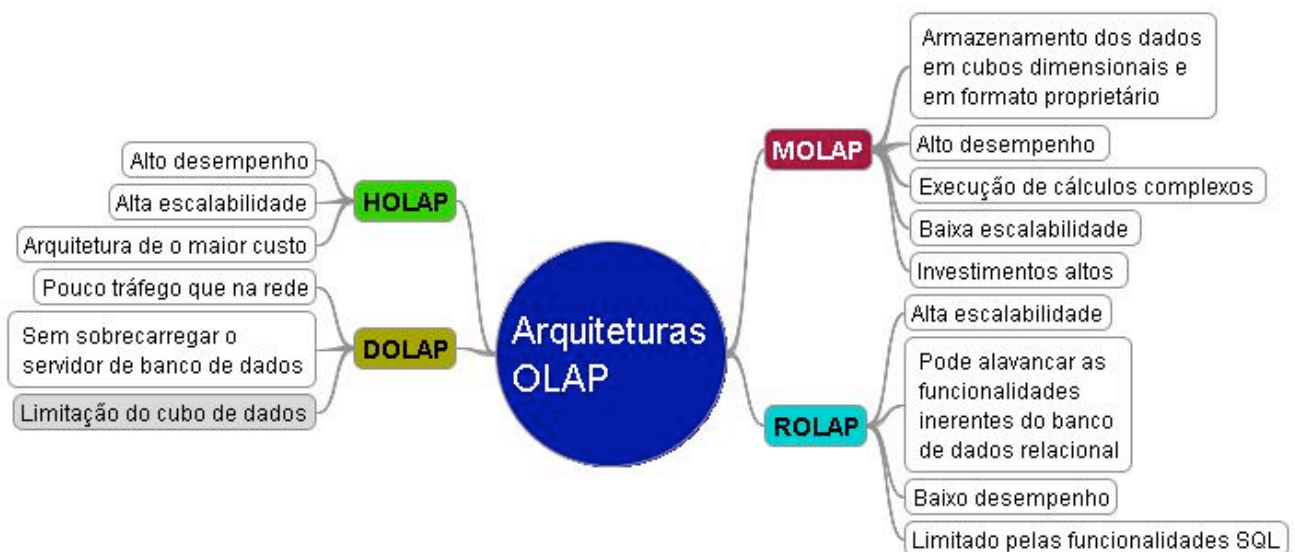


Figura 2 - Características das arquiteturas

Fonte: http://rogerioaraujo.files.wordpress.com/2009/05/mm01_arquiteturasolap.jpg, 2009.

A figura 2 resume esquematicamente os tipos de arquiteturas analisadas acima, bem como as suas respectivas vantagens e desvantagens.

3.2 Tipos de operação

Para que as ferramentas OLAP possam desempenhar as suas funções de maneira satisfatória, elas usufruem de características específicas. Segundo Anzanello(2002 p. 5):

Consultas ad-hoc: geradas pelos usuários finais de acordo com as suas necessidades de cruzar informações de uma forma não vista e que o levem a descoberta do que procuram.

Slice and Dice: possibilita a alteração da perspectiva de visão. Serve para modificar a posição de uma informação, trocar linhas por colunas de maneira facilitar a compreensão dos usuários e girar o cubo sempre que houver necessidade.

Drill down/up: consiste em realizar exploração em diferentes níveis de detalhes da informação. Com drill down dividi-se um item de resumo em seus componentes detalhados, como por exemplo, ano, semestre trimestre, mensal e diário.

Além dessas características os dados devem ser apresentados de maneira que possibilitem uma visualização de diversas formas, como por exemplo, gráficos. Outra forma muito utilizada para a visualização desses dados é o famoso Cubo OLAP.

Na figura 3 é apresentado um exemplo de cubo OLAP, no qual é possível observar a maneira como as dimensões (*Customer*, que são os clientes de uma determinada organização, *Product*, que são os produtos que esta organização oferece e *Time* que será um período de tempo, podendo ser meses e anos) estão agrupadas e como o cubo completo pode ser dividido para que seja possível uma visualização específica de um determinado fato, Quais Clientes compraram uma Camera Digital em Março de 2005, por exemplo.

Sendo assim, a forma como os dados são armazenados nas bases de dados é um dos elementos fundamentais para que a técnica de OLAP possa ser aplicada com o máximo de eficiência,

possibilitando, de acordo com Anzanello (2002. p. 5), vários modelos de visualização em uma variedade de formatos, e não apenas em simples tabelas, sendo muitas vezes apresentadas em gráficos.

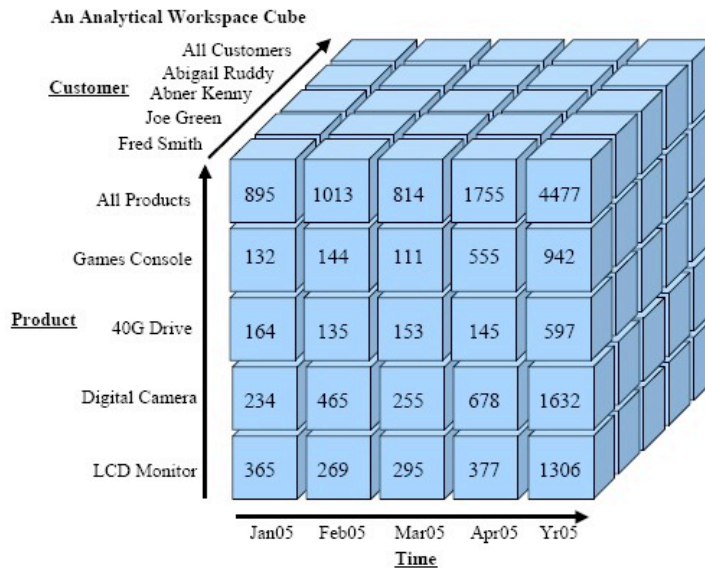


Figura 3 - Exemplo de cubo OLAP

Fonte: http://gerardnico.com/wiki/_media/database/oracle/oracle_olap_aw_cube.jpg, 2008.

4 OLAP e data Warehouse

Segundo Anzanello(2002) DataWarehouse é utilizada para armazenar informações e OLAP para recuperá-las, ambos são específicos para exercer suas funções de maneira eficiente. Desse modo, para explorar o Data Warehouse completamente, é necessário o OLAP para extrair e alavancar todas as informações nele contidas. Diferentemente dos bancos de dados transacionais, onde as informações históricas não ficam armazenadas por muito tempo, pois eles são utilizados somente para registrar e executar operações que já estão pré-definidas, ou seja, não há a necessidade de uma grande capacidade de armazenamento dessas informações.

Um Data Warehouse armazena dados analíticos, destinados às necessidades de gerência no processo de tomada de decisões. Isto pode envolver consultas complexas que necessitam acessar um grande número de registros, por isso é importante a existência de muitos índices criados para acessar as informações de maneira mais rápida possível. Um Data warehouse

armazena informações históricas de muitos anos e deve ter uma grande capacidade de processamento e armazenamento de dados que se encontram de duas maneiras, detalhados e resumidos (Nunes, 2007. p. 6).

Segundo Nunes (2007) com base nesses conceitos podemos concluir que o Data Warehouse é um conjunto de técnicas e bancos de dados integrados, projetados para suportar as funções dos Sistemas de Apoio à Decisão, onde cada unidade de dados está relacionada a um determinado assunto ou fato.

Como o DataWarehouse é uma base de dados utilizada para o armazenamento de dados analíticos é necessário que haja uma maneira de analisá-los que seja diferente das tecnologias de OLTP (que são utilizadas para operações convencionais, transacionais). A tecnologia OLAP fornece ferramentas específicas para auxiliar na extração de informações estratégicas a partir desses dados armazenados.

5 OLAP e data Mining

Segundo Anzanello(2002. p. 5) o OLAP e *Data Mining* são partes integrantes de todo e qualquer processo de decisão. Ainda, nos dias de hoje, a maioria dos sistemas OLAP tem o foco no provimento de acesso aos dados multidimensionais, enquanto os sistemas de *Data Mining* lidam com a análise para influência dos dados de uma única dimensão.

Desse modo, pode-se observar que a grande diferença entre essas duas tecnologias é a maneira como os dados são explorados. Na análise com a utilização de ferramentas OLAP a exploração é feita por meio de verificações, ou seja, o usuário conhece a informação e utiliza das ferramentas para comprová-la, acessá-la. Já no *Data Mining*, esse mesmo usuário não conhece a informação, ou conhece parcialmente, e utiliza das ferramentas na busca de padrões e na descoberta de informações (Nunes, 2007. p. 15).

6 Tendências

O JOLAP é um esforço da *Java Community Process* (JCP) de projetar uma API Java para servidores e aplicações OLAP, aderentes ao ambiente *Java 2Platform Enterprise Edition*

(J2EE). Ela está sendo especificada para suportar a criação e manutenção de dados e metadados OLAP, independente de fornecedor. JOLAP é baseada em uma forte generalização, orientada a objeto e nos conceitos de OLAP. Este modelo suporta conceitos referentes a três áreas que são chave para as aplicações OLAP, metadados, dados e pesquisas (ANZANELLO, 2002. p. 5).

Outra arquitetura é a WOLAP que se baseia na utilização de uma ferramenta OLAP a partir de um navegador Web. Essa arquitetura tem duas tecnologias que estão em constante evolução, a primeira é a Web e a segunda são as ferramentas OLAP. A diferença desta ferramenta para as outras é que ela utiliza a Internet, facilitando assim, a distribuição da ferramenta, o acesso remoto dos dados a serem analisados e a utilização da aplicação independente de plataforma (ANDREATO, 1999, apud ARAÚJO et al, 2007, p. 6).

7 Conclusão

Como a quantidade de informações aumenta a cada dia, há a necessidade de tecnologias tanto de armazenamento quanto de análise cada vez mais sofisticadas e eficientes. O OLAP tem despontado como um conjunto de ferramentas ideal para a análise de dados, pois proporciona ao usuário visualizar os resultados através de diferentes perspectivas graças à sua característica de análise multidimensional, de modo que o usuário pode navegar entre os diferentes níveis de informação (tempo: ano, mês, dia, por exemplo) e estabelecer comparações entre suas diferentes dimensões através do microcubo que é gerado.

Pelo fato de permitir ao usuário a personalização da busca, o que deixa os resultados mais consistentes, o OLAP é uma ferramenta muito útil na tomada de decisões para as empresas, uma vez que, a partir de um determinado conhecimento a respeito das informações que estão armazenadas, é possível fazer um cruzamento de dados em busca de melhores combinações. Esse conhecimento adquirido é que servirá como vantagem competitiva da empresa frente o mercado.

Por fim, observa-se que as técnicas de armazenamento (*Data Marts* e *Data Warehouse*), extração (*Data Mining*) e análise (OLAP) de dados estão intrinsecamente ligadas, ou seja,

para que os dados sejam utilizados com o potencial máximo é necessário: que estes estejam armazenados de maneira adequada em uma base de dados para facilitar na busca e pesquisa; a utilização de ferramentas eficientes para a extração das informações dessa base de dados e o uso de técnicas que possibilitem a análise dessas informações para gerar o conhecimento que o usuário espera de forma ágil e eficiente.

REFERÊNCIAS

ANZANELLO, C. A. **OLAP Conceitos e Utilização**. Instituto de Informática - Universidade federal do Rio Grande do Sul (UFRGS). Porto Alegre - RS. 2002.

ARAÚJO, E. M. T. ; BATISTA, M. de L. S. ; MAGALHÃES, T. M. de. **OLAP: Características, Arquitetura e Ferramentas**. Instituto Vianna Júnior. Faculdades Vianna Júnior. Juiz de Fora - MG. 2007

FERNANDES, M. de S.; KANTORSKI, G. Z. . Neuro BI: **Uma Ferramenta Web open source para apoio à Tomada de Decisão**. Disponível em: <<http://www.cin.ufpe.br/~ejvm/OLAP/3484.pdf>>; Acesso em 13 de Maio de 2011.

FILHO, Ly F. **Business Intelligence**. Disponível em: <<http://analgesi.co.cc/html/t32973.html>> Acesso em 17 de maio de 2011.

HELBER, P, A. **Ferramentas Olap**. Disponível em:<<http://www.inf.ufsc.br/~helber/FerramentasOlap.ppt>> Acesso em 12 de maio 2011

LAUDON, K. C. e LAUDON, J. P. **Sistemas de Informações Gerenciais**. 7. ed. São Paulo: Pearson Prentice Hall, 2007.

NUNES, Orlando Augusto. **Business Intelligence**. Disponível em <<http://www.webartigos.com/articles/2437/1/Business-Intelligence/pagina1.html>> Acesso em 15 de Maio de 2011.

ROCHA, K. A. F.; JARDIM, M. U. B., et al. **Business Intelligence**. Faculdades NDA FACNET. Brasília - DF. 1 semestre de 2004.

ROSINI, Alessandro Marco. **Administração de Sistemas de Informação e a Gestão do Conhecimento**. São Paulo: Pioneira Thomson Learning, 2003.

SALVADOR, L. **Modelagem Dimensional de Dados**. Consegi, 12/05/2011. Palestra.

VIEIRA, A. P. . OLAP - Motivação e conceitos; Disponível em: <http://www.ppgia.pucpr.br/~scalabrin/Especializacao/BI/aula01/olap_Aroldo.ppt> Acesso em 10 de Maio de 2011.

Abstract:

Because the increasing volume of data produced by companies, it was necessary to create tools that could give them a concise and consistent basis in order to allow that critical information could be extracted and used strategically in business management. Then comes the OLAP(Online Analytical Processing). A set of tools that helps in quantitative and qualitative analysis of large volumes of data through semi-custom optics that make managers, businessmen and company directors may enjoy different perspectives to make strategic decisions.

Keywords: Analysis. Business Intelligence. Information. Multidimensionality. OLAP.