

Banco de Dados: Relacional X Multidimensional*

1. Introdução

Na década de 60 a computação como um todo consistia no desenvolvimento de aplicações extremamente individuais que eram baseadas em arquivos. Todos os registros e relatórios gerados vinham de diversos arquivos, que com o tempo de uso e o aumento dos registros começavam apresentar problemas, tais como: complexidade na manutenção dos programas; complexidade no desenvolvimento de novos programas e o sincronismo dos dados a serem atualizados.

Por este motivo, eram recíprocos os esforços e investimentos em pesquisar um meio mais barato e ter uma solução mecânica eficiente. E foi então que em 1970 um pesquisador da IBM, Ted Codd, publicou um artigo sobre banco de dados relacional, onde ele imaginava o usuário sendo capaz de acessar informações através de comandos. Estes comandos que futuramente passaram a ser chamados de SQL, linguagem utilizada no primeiro DB2, tornaram-se um padrão na indústria de banco de dados.

Já na década de 70 surgiu o armazenamento em disco, ou DASD (Direct Access Storage Device: dispositivo de armazenamento de acesso direto), trazendo consigo o software conhecido hoje em dia como Sistema de Gerenciamento de Banco de Dados (SGBD). Com o DASD e o SGBD surgiu a idéia de desenvolver o “banco de dados”, também definido como a única fonte de dados para todo o processamento.

O banco de dados promoveu uma visão de organização “baseada em dados”, em que o computador poderia atuar como coordenador central das atividades de toda a empresa. Nesta visão, o banco de dados tornou-se um recurso corporativo básico, sendo que pela primeira vez as pessoas não estavam vendo os computadores apenas como misteriosos dispositivos de previsão. A partir daí os computadores passaram a ser vistos como uma verdadeira vantagem competitiva. A idéia dos sistemas de informação para os negócios começou a tomar forma, tendo as empresas alcançando mais eficiência. Com essa evolução, começou perceber-se que poderiam analisar de forma otimizada seus dados, ou seja, descobriram que poderiam incrementar e dar valor nos seus recursos de BI (Business Intelligence). Nasce um novo conceito para a tecnologia da informação (Francotti, 2006).

A partir da miniaturização dos circuitos integrados, que conseqüentemente barateou e permitiu a popularização do computador pessoal, surgiu uma nova realidade para a computação. Hoje com a internet e o uso crescente dela temos uma quantidade quase incontável de informações espalhadas em diversos servidores e em computadores pessoais, que podem estar compartilhadas e nem sabemos onde estão fisicamente. Este aumento de capacidade de processamento e de armazenagem de dados trouxe a difícil tarefa de “encontrar” conhecimento dentro deste emaranhado de dados.

Deste problema surgiram os *Date Warehouse* e a mineração de dados, com o objetivo de armazenar informações de estratégia organizacional de forma consolidada para a análise do grande volume de dados temporais para obtenção de informações, que

pudessem facilitar a tomada de decisões. Através destes dados, foi possível realizar a mineração e extrair padrões de informações que muitas vezes a olho nu não se tem o mesmo resultado.

2. Banco de Dados Relacional

O banco de dados relacional é composto de *tabelas* ou *relações*. E este segue o modelo relacional que se baseia no princípio de que todos os dados estão dispostos em tabelas e toda sua definição é teórica e fundamentada na lógica de predicados e na teoria dos conjuntos.

De uma forma simples, um Banco de Dados Relacional é um conceito abstrato que define maneiras de armazenar, manipular e recuperar dados estruturados unicamente na forma de tabelas, construindo um banco de dados (Wikipédia).

Os bancos de dados relacionais foram desenvolvidos com o intuito de prover o acesso às informações de maneira a agilizar este processo, possibilitando que os usuários utilizassem uma grande variedade de abordagens no tratamento das informações. E a linguagem padrão para o tratamento dos dados nos Bancos de Dados Relacionais é a Structured Query Language, ou simplesmente SQL, como é mais conhecida.

O criador do modelo relacional, Edgar Frank Codd, publicou em seu artigo algumas regras que definem o modelo, aqui cito quatro delas:

- Toda informação deve ser representada de uma única forma, como dados em uma tabela;
- Todo dado (valor atômico) pode ser acessado logicamente (e unicamente) usando o nome da tabela, o valor da chave primária da linha e o nome da coluna;
- A capacidade de manipular a relação base ou relações derivadas como um operador único não se aplica apenas a recuperação de dados, mas também a inserção, alteração e eliminação de dados;
- Independência lógica de dados.

2.1 Tabelas

Uma tabela é um conjunto não ordenado de *linhas* (tuplas). Cada linha é composta por uma série de *campos* (valor de atributo). E cada campo é identificado por um *nome de campo* (nome do atributo). E por fim o conjunto de campos homônimos de todas as linhas de uma *tabela* (relação) forma uma *coluna* (atributo).

Quando uma tabela é definida, para cada coluna deve ser especificado um conjunto de valores (alfanuméricos, numérico,...) que os campos da respectiva coluna podem assumir. Este conjunto de valores é chamado de *domínio da coluna*. Sendo que nas colunas que não são admitidos valores nulos são chamados de colunas *obrigatórias*. Do contrário são chamadas de colunas *opcionais* (Heuser, 2001).

Tabela 1. Tabela em um banco de dados relacional

Torneio	Ano	Vencedor
Indiana Invitational	1998	Al Fredrickson
Cleveland Open	1999	Bob Albertson
Des Moines Masters	1999	Al Fredrickson
Indiana Invitational	1999	Chip Masterson

2.2 Chaves

O conceito básico para identificar linhas e estabelecer relações entre linhas de tabelas de um banco de dados relacional é o de *chave*. Em um banco de dados relacional, há ao menos três tipos de chaves a considerar: a chave *primária*, a chave *alternativa*, e a chave *estrangeira* (Heuser, 2001).

2.2.1 Chave primária

Uma chave primária é uma coluna (simples) ou uma combinação de colunas (composta) cujos valores distinguem uma linha das demais dentro de uma tabela. Usa-se chave composta quando apenas um dos valores dos campos que compõem a chave não me suficiente para distinguir uma linha das demais.

Na abordagem relacional, ao definir uma chave primaria, não esta se definindo nenhum caminho de acesso. Esta se definindo apenas uma *restrição de integridade*, isto é uma regra que deve ser obedecida em todos estados válidos de um BD. No caso da chave primaria, a regra é a de unicidade de valores nas colunas que compõem a chave (Heuser, 2001).

2.2.2 Chave estrangeira

Uma chave estrangeira é uma coluna ou uma combinação de colunas, cujos valores aparecem necessariamente na chave primaria de outra tabela. A chave estrangeira é o mecanismo que permite a implementação do relacionamento em um banco de dados relacional (Heuser, 2001).

2.2.3 Restrição de integridade

É de responsabilidade dos SGBDs manter a integridade dos dados. Dizer que os dados de um banco de dados estão íntegros significa dizer que eles refletem corretamente a realidade representada pelo banco de dados e que são consistentes entre si. Essa integridade se dá com o mecanismo de *restrição de integridade* e é realizada pelo SGBD e são as seguintes categorias de restrição:

- Integridade de domínio;
- Integridade de vazio;
- Integridade de chave;
- Integridade referencial.

2.2.4 Consultas à base de dados

Como foi dito na introdução deste artigo a linguagem usada como padrão de definição e manipulação do banco de dados é a SQL. As três opções básicas de manipulação de dados com seus respectivos exemplos são:

- Seleção: `SELECT * FROM tabela WHERE...`
- Alteração: `UPDATE tabela SET atributox=2,... WHERE...`
- Exclusão: `DELETE FROM tabela WHERE...`

3. Banco de dados Multidimensional

Nos últimos anos o termo Business Intelligence (BI) tem sido largamente utilizado no mercado como sinônimo de sistemas analíticos, OLAP, cubos, entre outros. Embora essas denominações possam estar associadas entre si, são conceitualmente distintas.

A rigor, Business Intelligence pode ser obtido por qualquer artefato, seja tecnológico ou não, que permita a extração de conhecimento a partir de análises do negócio. Por razões óbvias, a efetividade destas análises será maior se os dados estiverem disponíveis de modo consistente e, preferencialmente, consolidado. Este é um dos objetivos dos Data Warehouses (Nardi, 2007).

Segundo o autor Harry S. Singh, Data Warehouse é o processo de integração dos dados corporativos de uma empresa em um único repositório. É um ambiente de *suporte à decisão* que alavanca dados armazenados em diferentes fontes e os organiza e entrega aos tomadores de decisões. Resumindo, é uma tecnologia de gestão e análise de dados. Já W. H. Inmon (1993) define da seguinte maneira: “Um Data Warehouse é um conjunto de dados orientado por assunto, integrado, variável com o tempo e não-volátil, que fornece suporte ao processo de tomada de decisão”.

3.1 Processos para a construção de um Data Warehouse

Para os processos de construção de um Data Warehouse é altamente importante a compreensão do negócio que envolve a empresa ou instituição que se esta desenvolvendo o trabalho.

O processo inicial é a extração de dados das bases de dados transacionais, dados de sistemas ERP, dados locais, externos ou web. Esta extração sem chama ETL (*Extract Transform Load*, Extração Transformação Carga) e é aqui que são analisados cuidadosamente os dados e assim traduzidas às necessidades de negócio da empresa. Após a extração os dados estes devem ser transformados para que seja possível a carga dos dados em um Data Warehouse ou Data Mart dependendo do método de construção adotado.

Como métodos de construção, existem formalmente dois:

- Top-down, no qual é realizada a modelagem integral do DW, seguida pelas extrações de dados. A principal vantagem é a criação de um modelo único. O revés fica por conta do maior tempo de projeto;

- Bottom-up, onde o foco é em uma área por vez, com o crescimento gradual do DW. A vantagem é a obtenção de resultados a intervalos mais curtos, garantindo muitas vezes sustentação ao projeto. A desvantagem é a maior dificuldade de se consolidar informações entre as diversas áreas.

Como vemos na imagem abaixo há duas formas de construção, uma onde o DW gera os DM (Data Mart) e outra em que os DM geram o DW. O DM nada mais é que um subconjunto de dados de um DW, onde tipicamente desempenham o papel de um DW departamental, regional ou funcional. Alguns autores e especialistas dizem que o DW é uma evolução do DM que começou localizado e cresceu para atender um escopo maior.

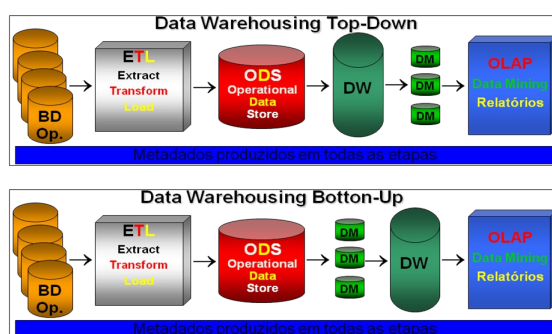


Figura 1. Exemplos das duas formas de construção de um Data Warehouse

3.2 Modelagem Multidimensional

A modelagem multidimensional é a técnica de projeto mais frequentemente utilizada para a construção de um Data Warehouse. O objetivo é buscar um padrão de apresentação de dados que seja facilmente visualizado pelo usuário final e que possua um bom desempenho para consultas.

O modelo dimensional é formado por uma tabela central (tabela de fatos) e várias outras a ela interligadas (tabelas de dimensão), sempre por meio de chaves especiais, que associam o fato a uma dimensão do cubo.

Alguns conceitos importantes dentro da modelagem multidimensional estão listados abaixo:

- **Dimensões:** estabelecem a organização dos dados, determinando possíveis consultas/cruzamentos. Por exemplo: região, tempo, canal de venda,... Cada dimensão pode ainda ter seus elementos, chamados membros, organizados em diferentes níveis hierárquicos. A dimensão tempo, por exemplo, pode possuir duas hierarquias: calendário gregoriano (com os níveis ano, mês e dia) e calendário fiscal (com os níveis ano, semana e dia);
- **Medidas:** são os valores a serem analisados, como médias, totais e quantidades;
- **Fatos:** são os dados a serem agrupados, contendo os valores de cada medida para cada combinação das dimensões existentes. O tamanho da tabela que contém os fatos merece atenção especial do analista;
- **Agregações:** totalizações calculadas nos diversos níveis hierárquicos.

3.3 Ferramentas e acesso aos dados

Como ferramenta de consulta a banco de dados multidimensionais temos um conjunto de aplicações que se denominam ferramentas OLAP (*On-Line Analytical Processing*, Processamento Analítico On-line) as quais tem capacidade para manipular e analisar um grande volume de dados sob múltiplas perspectivas.

Típicas questões que uma ferramenta OLAP consegue responder dentro de um banco de dados multidimensional:

- Quais os produtos mais bem vendidos no mês passado?
- Quais os 10 piores vendedores dos departamentos da filial X?
- Qual a média salarial dos funcionários de informática na região sul nos últimos cinco anos?

E dentro das ferramentas OLAP temos 3 distinções ainda que são: Relacional OLAP (ROLAP); Multidimensional OLAP (MOLAP); Híbrido OLAP (HOLAP).

3.4 Conclusão

Uma das principais diferenças entre estas duas tecnologias são seus objetivos e a interação entre elas. O banco de dados relacional vem com o intuito básico de relacionar as operações diárias de negócios de uma instituição ou empresa com foco no operacional e com atualizações contínuas.

Já o banco de dados multidimensional é gerado e alimentado pelas bases de dados relacionais através de processos de coleta, transformação e carga de dados e seu objetivo se dá no âmbito gerencial, e sua atualização é temporal (em lote). E juntamente com ferramentas apropriadas essas bases de dados geram gráficos, planilhas que ajudam na tomada de decisões estratégicas para a empresa.

O uso da base de dados multidimensional com ferramentas OLAP vem crescendo muito porque permitem armazenar e analisar grandes volumes de dados e auxiliam a compreensão do comportamento destes dados sobre o negócio e Fornecem apoio ao suporte estratégico de decisão.

Referências Bibliográficas

- Heuser, Carlos A. (2001) “Projeto de Banco de Dados”, Séries Livros Didáticos, 4º edição, Editora Sagra Luzzatto. Instituto de Informática da UFRGS.
- Singh, Harry S. (2001) “Data Warehouse: Conceitos, Tecnologias, Implementações e Gerenciamento”, Editora MAKRON Books.
- Nardi, Alexandre R. (2007) “Fundamentos e Modelagem de Bancos de Dados Multidimensionais”, MSDN - Microsoft Developer Network, <http://msdn.microsoft.com/pt-br/library/cc518031.aspx#XSLTsection125121120120>, acessado em 18/03/2009
- Francotti, Doubleday K. (2006). “Como será o futuro sem eles? Dados, Informação e Conhecimento a Alma do ‘Data Warehouse’”, [http://www.artigos.com/artigos/exatas/tecnologia/como-sera-o-futuro-sem-eles?-](http://www.artigos.com/artigos/exatas/tecnologia/como-sera-o-futuro-sem-eles?)

dados,-informacao-e-conhecimento-a-alma-do-data-warehouse-661/artigo/, acessado em 15/03/2009.

Inmon, W. H., e C. Kelley. In: *Developing the Data Warehouse*. QED Publishing Group, Boston, 1993

Andrade, Fábio B. e Santiago, Luciano D. G., “Introdução aos Conceitos de Modelagem Multidimensional Aplicados a Data Warehouses”, CienteFico, Salvador, julho-dezembro 2003, <http://www.frb.br/ciente/Textos%20CienteFico%202003.2/INFO/Banco%20de%20Dados/INTRODU%C7%C3O%20AOS%20CONCEITOS%20DE%20MODELAGEM%20MULTIDIMENSIONAL%20APLICADOS%20A%20DATA%20WAREHOUSES.pdf>, acessado em 21/03/2009

Rafael B. Rohden

rafaelrohden@gmail.com

Departamento de Tecnologia

Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUI)

CEP 98.700-000 – Ijuí – RS – Brazil