

UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA

2004.2

FRAGMENTAÇÃO VERTICAL DE DATA WAREHOUSE EM
TERMOS DE MEDIDAS NUMÉRICAS: UM ALGORITMO BÁSICO.

PROPOSTA DE TRABALHO DE
GRADUAÇÃO EM
BANCO DE DADOS

Aluno – Marcelo Victor Calado de Sousa Costa, mvsc@cin.ufpe.br.

Orientador – Fernando da Fonseca de Souza, fdfd@cin.ufpe.br.

08 de Novembro de 2004

Índice

1. CONTEXTO.....	3
2. PROJETO	4
3. CRONOGRAMA	5
4. REFERÊNCIAS	6
5. DATA E ASSINATURAS	7

1. Contexto

O data warehouse nasceu a partir do reconhecimento da importância do valor da informação nas organizações. Segundo Inmon [5], a definição clássica de data warehouse é: “uma coleção de dados orientada por assuntos, integrada, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão”.

Assim, o data warehouse forma uma base de dados que permite efetuar um tratamento adequado da informação, o qual pode habilitar a descoberta e exploração de tendências empresariais importantes.

Entretanto, dados extremamente centralizados podem resultar em perda de disponibilidade e queda de desempenho das consultas. Daí surge a necessidade de fragmentação do data warehouse.

De acordo com Özsu e Valduriez [7], o objetivo da fragmentação vertical é particionar uma relação em um conjunto de relações menores, de tal modo que muitos dos aplicativos do usuário possam atuar apenas sobre um fragmento. É com esta motivação que surge a necessidade de desenvolvimento de algoritmos para realizar uma fragmentação vertical em um ambiente de data warehouse.

Neste contexto, um esquema de fragmentação vertical deve minimizar o tempo de processamento dos aplicativos do usuário que atuam sobre esses fragmentos [7]. Baseado nisto é que podemos identificar a necessidade de algoritmos para fragmentação vertical.

Grandes empresas necessitam trabalhar com medidas numéricas de forma veloz para que possam estar sempre à frente de seus concorrentes. A fragmentação vertical dos dados numéricos faz elevar a velocidade com que esses dados são manipulados, visto que eles podem ser situados em sítios de alta performance.

É com esta motivação que surge a necessidade para o desenvolvimento de um algoritmo voltado a resolver o problema da fragmentação vertical de um data warehouse em termos de medidas numéricas.

Como arquitetura, tem-se como base o sistema WebD²W (Web Distributed Data Warehousing), desenvolvido pela necessidade de distribuição dos dados do data warehouse e a utilização da tecnologia Web como infra-estrutura para garantir o acesso ao ambiente de data warehousing distribuído [1]. Uma abordagem de fragmentação vertical representa uma extensão do WebD²W, pois este fornece apenas soluções de fragmentação horizontal.

2. Projeto

Este trabalho de graduação propõe a elaboração de um algoritmo em pseudocódigo para fragmentação vertical de um data warehouse em termos de medidas numéricas.

De acordo com Ciferri [1], uma medida numérica é definida como uma função de suas dimensões correspondentes. Desta forma, as medidas numéricas utilizadas como base no processo de fragmentação e suas respectivas restrições são necessárias como entrada para o algoritmo, assim como a agregação do grafo de derivação sobre o qual essas medidas numéricas são contextualizadas.

O algoritmo desenvolvido será baseado nos conceitos de grafos de derivação, de propagação das dimensões sendo fragmentadas aos vértices do grafo e de fragmentação ou reconstrução de agregações. Sendo assim uma maneira adequada para o armazenamento dos dados do data warehouse em diferentes níveis de agregação, conforme cita Ciferri [1]. O algoritmo proposto aqui constitui uma das fundamentações para o sistema WebD²W.

3. Cronograma

O cronograma abaixo visa demonstrar a evolução das atividades elaboradas ao longo do período de novembro a fevereiro, dividido em semanas, para o desenvolvimento do trabalho de graduação:

Atividade	Mês														
	Novembro			Dezembro				Janeiro				Fevereiro		Março	
1. Estudo detalhado da arquitetura do sistema WebD ² W.	■	■	■												
2. Estudo aprofundado de grafos de derivação.			■	■											
3. Estudo de tópicos avançados relativos à fragmentação vertical.			■	■											
4. Definição do Algoritmo para fragmentação vertical em termos de medidas numéricas.				■	■										
5. Detalhamento do Algoritmo em pseudocódigo.					■	■	■	■	■						
6. Finalização do algoritmo e dos estudos.											■				
7. Revisão geral do trabalho de graduação.												■			
8. Elaboração da apresentação final.													■		
9. Elaboração do documento final.				■	■	■	■	■	■	■	■				
10. Defesa do Trabalho de Graduação.														■	

4. Referências

- [1] CIFERRI, C.D.A., *Distribuição dos Dados em Ambientes de Data Warehousing: O Sistema WebD²W e Algoritmos Voltados à Fragmentação Horizontal dos Dados*, Universidade Federal de Pernambuco, 2002.
- [2] COME, C., *Contribuição ao Estudo da Implementação de Data Warehousing: Um Caso no Setor de Telecomunicações*, Universidade de São Paulo, 2001.
- [3] GREENFIELD, L., *The Data Warehousing Information Center*, consultada em 04/11/2004, disponível em <<http://www.dwinfocenter.org>>.
- [4] HAMMER, M., NIAMIR, B., *A Heuristic Approach to Attribute Partitioning*, ACM, 1979.
- [5] INMON, W. H., *Data Mart ≠ Data Warehouse*, DM Review, consultada em 04/11/2004, disponível em <<http://www.dmreview.com>>.
- [6] NAVATHE, S., CERI S., WIEDERHOLD G., DOU, J., *Vertical Partitioning of Algorithms for Database Design*, ACM Trans, 1984.
- [7] ÖZSU, M.T., VALDURIEZ P., *Principles of Distributed Database Systems*, pages 108-165, Upper Saddle River, New Jersey, USA. Prentice-Hall, 1999.

5. Data e Assinaturas

Recife; 08 de Novembro de 2004.

Fernando da Fonseca de Souza
Orientador

Marcelo Victor Calado de Sousa Costa
Aluno