

**APLICATIVO COMUM PARA COMPARAÇÃO DE DESEMPENHO DE TÉCNICAS DE CLUSTERING EM SGBDOO.** Willian dos Santos Lima, Carlos Roberto Valêncio – Bacharelado em Ciência da Computação – Departamento de Ciências de Computação e Estatística – Instituto de Biociências, Letras e Ciências Exatas – Campus São José do Rio Preto.

Os Sistemas Gerenciadores de Banco de Dados Orientado a Objeto (SGBDOO) devem ser capazes de suportar dados variados e não convencionais, cuja complexidade é significativa quando comparados aos dados comuns, tipicamente os de tipo texto. Podemos citar como não convencionais dados como imagens, vídeos, arquivos de áudio, e provenientes de aplicações CAD/CAM, CASE, entre outros. Desta forma, é necessário que se tenha uma estratégia capaz de manipular os registros lógicos de uma base de dados de maneira eficaz, a fim de que se consiga reduzir o número de acessos a disco e, conseqüentemente, o tempo necessário para armazenamento e recuperação desses dados.

O Núcleo Gerenciador de Dados Multimídia (NuGeM), proposto por Valêncio (2000), tem como objetivo a manipulação de dados multimídia (POLIZELLI JUNIOR, 2005), feita através da montagem de objetos a partir desses dados. Cada objeto no NuGeM pode ser composto por outras peças lógicas, como tuplas, listas e BLOB (*Binary Large Objects*), sendo este último o responsável pelo armazenamento de parte dos dados multimídia (GARCIA-MOLINA; ULLMAN; WIDOM, 2001). O NuGeM possui uma camada em sua estrutura denominada de Subsistema Gerenciador de Registros Lógicos (SGLog), a qual é responsável pela alocação dos registros lógicos de diversos tipos (colônia, objeto, tupla, lista, BLOB e índice), e, por este motivo, é junto a esta camada que as estratégias de *clustering* são implementadas.

Entre as várias formas de otimização de desempenho para gerenciadores de banco de dados, as políticas de agrupamento de registros lógicos, conhecidas como *clustering*, visam a fornecer este suporte para os gerenciadores de dados não convencionais. O principal objetivo de uma política de *clustering* é proporcionar alocação eficiente dos registros no banco de dados de modo que sua recuperação, baseada essencialmente na obtenção de unidades físicas (páginas do disco), seja feita de forma eficaz.

Tendo em vista que a recuperação de dados do disco é realizada através da obtenção das páginas do disco, e também que os vários registros lógicos podem ser alocados em uma página, então quando um registro lógico é requisitado, todos os outros alocados na mesma unidade física são recuperados. Desta forma, se os registros de objetos semanticamente relacionados forem armazenados na mesma unidade física ou o mais próximo possível, tem-se uma probabilidade maior de que, quando deseja-se recuperar registros relacionados entre si, seja necessário apenas um acesso a disco.

O trabalho teve como meta mostrar um aplicativo criado para comparação de desempenho entre diferentes políticas de *clustering* (DARMONT; GRUENWALD, 1996) para NuGeM, de forma que se possa dimensionar a de melhor comportamento segundo alguns critérios, como tempo, número de acessos a disco e número de acessos à *cache*, partindo de parâmetros como número de registros manipulados, tamanho desses registros e seus respectivos tipos. O aplicativo, desenvolvido em C++ dentro de um ambiente de desenvolvimento integrado com suporte para a criação de aplicativos gráficos para o sistema operacional Windows®, atua sobre o SGLog e contempla políticas de *clustering* definidas dentro do mesmo.

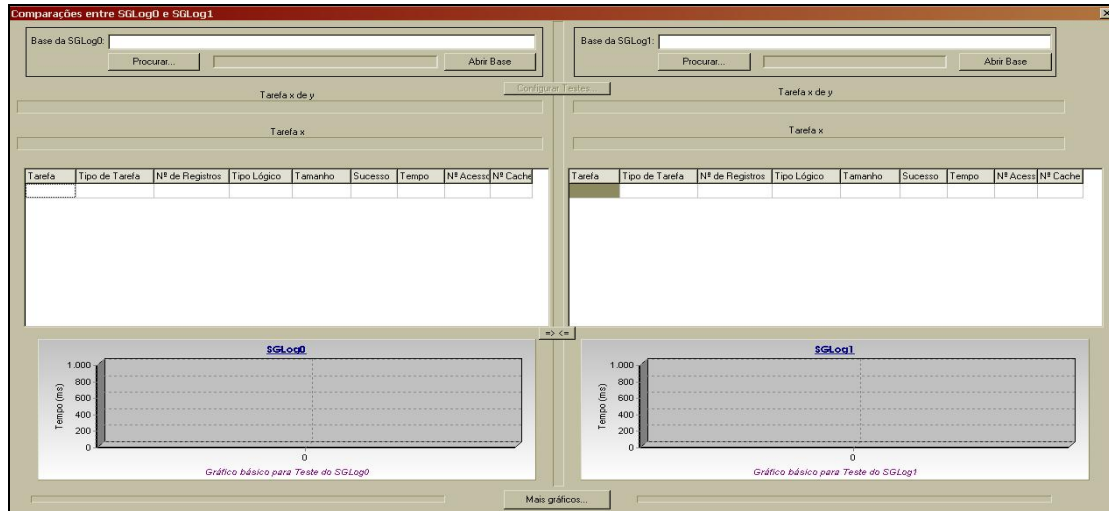
O aplicativo, apresentado na Figura 1, funciona através dos seguintes passos:

1. Abertura de duas bases de dados, cada uma contendo uma política de *clustering* distinta. É importante que essas duas bases sejam idênticas, para que não haja deturpação nos resultados da análise;
2. Com as bases abertas, deve-se formular os conjuntos de testes a partir da interface mostra na Figura 2, que são as operações, e suas respectivas ordem de acontecimento, que deverão ser executadas;
3. Por fim, executam-se os testes e os resultados aparecem nas tabelas correspondentes às estratégias analisadas.

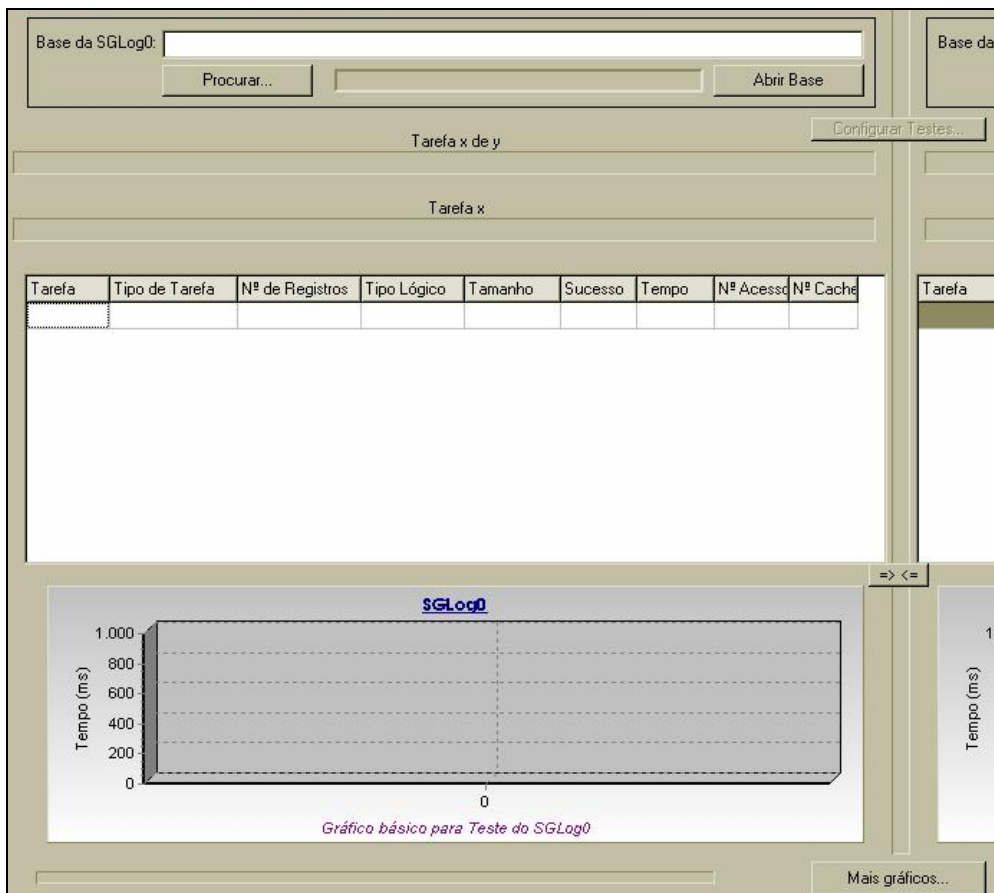
Com os dados resultantes do teste obtidos, o aplicativo proporciona ao usuário algumas facilidades para a visualização e comparação dos resultados, como gráficos que relacionam:

- Cada tarefa executada com o tempo consumido para tal;
- Quantidade de registros lógicos manipulados e tempo para execução;
- Quantidade de registros lógicos manipulados e número de acessos a disco requisitados;
- Cada tipo de registro lógico com o tempo gasto para a realização das tarefas.

Entende-se por tarefa uma seqüência de registros lógicos de um mesmo tipo, com mesmo tamanho, a serem submetidos à execução de um mesmo método do SGLog. Cada linha da tabela mostrada na Figura 1(b) é uma tarefa diferente.



**Figura 1(a):** Interface para comparação entre técnicas.



**Figura 1(b):** Ampliação de uma parte da interface de (a). O outro lado da interface possui a mesma tabela e os mesmos campos que o mostrado.

**Configurar Testes em Massa**

Testes de Escrita:

☐ Todos N° Registros: Tipo Lógico: Tamanho:

☐ 1: 1. Objeto

☐ 2: 1. Objeto

☐ 3: 1. Objeto

☐ 4: 1. Objeto

☐ 5: 1. Objeto

OK Limpar

Testes de Leitura:

☐ Todos N° Registros: Tipo Lógico: Tamanho:

☐ 1: 1. Objeto

☐ 2: 1. Objeto

☐ 3: 1. Objeto

☐ 4: 1. Objeto

☐ 5: 1. Objeto

OK Limpar

Testes de Remoção:

☐ Todos N° Registros: Tipo Lógico: Tamanho:

☐ 1: 1. Objeto

☐ 2: 1. Objeto

☐ 3: 1. Objeto

☐ 4: 1. Objeto

☐ 5: 1. Objeto

OK Limpar

Tarefas:

Para Cima

Para Baixo

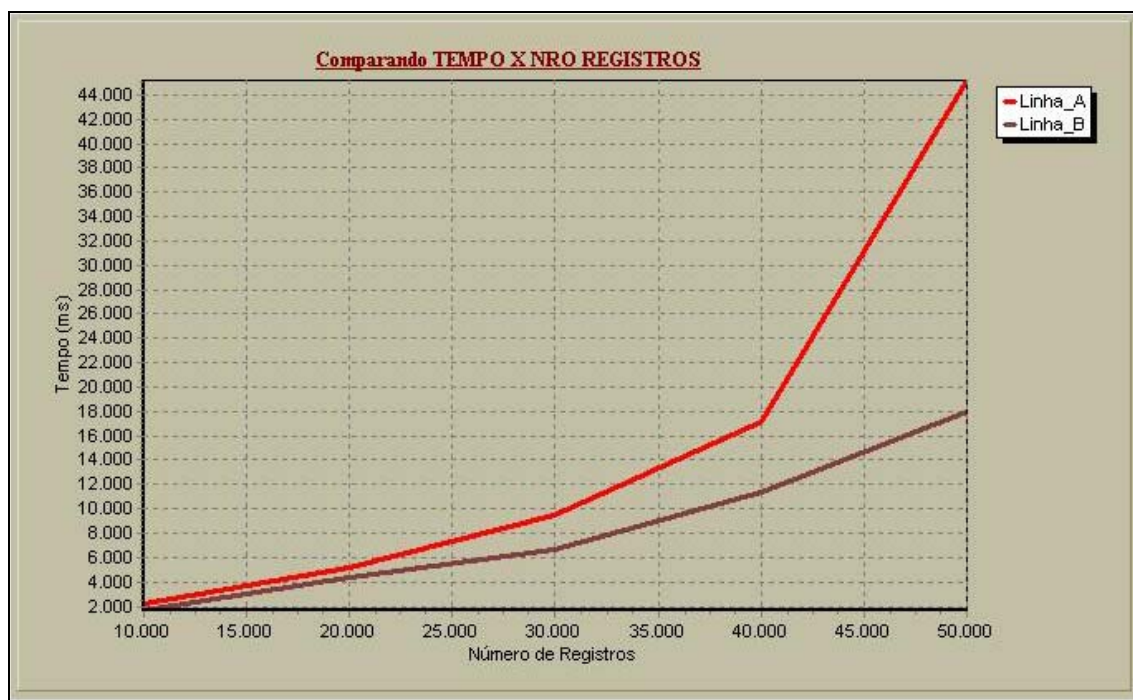
Limpar

Colônia: 0 Transação: 0

Executar Cancelar

**Figura 2:** Interface para configuração dos testes.

A Figura 3 mostra um exemplo de gráfico gerado a partir da execução de um conjunto de testes.



**Figura 3:** Exemplo de gráfico gerado pelo aplicativo.

Através do gráfico da Figura 3, podemos perceber que uma das estratégias, representada pela linha vermelha mais clara, tem aumento significativamente maior que a outra quando o número de registros lógicos manipulados aumenta.

Portanto, o aplicativo apresentado fornece a possibilidade de comparação entre políticas de *clustering* implementadas dentro do NuGeM, possibilitando ao usuário métricas para auxiliá-

lo nas decisões de implementação e, também, para testar a eficiência de uma possível nova estratégia de *clustering* a ser contemplada no NuGeM.

#### **Referências:**

DARMONT, J.; GRUENWALD, L. A Comparison Study of Object-Oriented Database Clustering Techniques. **Information Sciences** **94** New York, NY, v. 94, n. 1-4, p. 55-86, 1996.

GARCIA-MOLINA, H; ULLMAN, J. D.; WIDOM, J. **Implementação de Sistemas de Banco de Dados**. Editora Campus, 2001.

POLIZELLI JUNIOR, V. **Armazenamento e Recuperação de Imagens Baseados em Conteúdo sobre o NuGeM**. Trabalho de conclusão de curso (Bacharelado em Ciência da Computação) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2005.

VALENCIO, C. R. **Núcleo Gerenciador de Objetos – Compatibilizando eficiência e flexibilidade**. Tese (Doutorado em Física) - Instituto de Física, Universidade de São Paulo, São Carlos, Set. 2000.

**Bolsa:** Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

**ÁREA:** *Ciências Exatas*

**SUB-ÁREA:** 1.03 *Ciência da Computação*