

CK0117 - Sistemas de Bancos de Dados - 2015-1

Javam C. Machado

TRABALHO I - ÍNDICE HASH

1 Considerações

- Um banco de dados será mapeado para um arquivo do sistema operacional, onde cada registro do arquivo é uma página do banco de dados com 128 bytes. O conteúdo de cada página tem partes de tamanho fixo e um vetor de endereçamento de cada uma dessas partes. Um caractere especial deve ser usado para indicar uma parte de conteúdo vazio.
- Um índice hash inicialmente com 4 buckets primários onde cada bucket tem duas páginas do banco de dados – no início o bucket é formado portanto de 2 registros do arquivo e o índice tem 8 registros uma vez que tem inicialmente 4 buckets.
- O índice vai indexar valores de chave numéricos, como por exemplo idade. A estrutura das entradas de dados é formada pelo valor da chave de busca e o RID da tupla indexada. Como o trabalho não vai implementar as páginas da tabela indexada, o RID pode ser um número aleatório, mas único, para o índice. As entradas de dados são mapeadas portanto para cada parte de página do arquivo.
- O índice deverá estar armazenado em disco e a unidade de I/O será uma página de índice, portanto um registro de arquivo considerando o mapeamento adotado.
- Uma página de bucket deve ser lida para a memória quando necessário, portanto não é permitido ler todo o arquivo de índice para a memória para realizar alterações. Quando modificada, a página deverá ser armazenada em disco após alteração.

2 Tipos de Trabalho

Cada equipe de, *no máximo*, dois alunos fará um dos seguintes trabalhos.

Índice Hash Extensível Implementar um índice de Hash Extensível considerando as especificações acima. Implemente os algoritmos de busca, inserção e remoção de entradas de dados considerando a profundidade global e local, e a atualização desses valores caso a divisão de algum bucket implique na duplicação do diretório de ponteiros.

Índice Hash Linear Implementar um índice de Hash Linear considerando as especificações acima. Para isto você utilizará uma família de funções ao invés de diretórios. Dessa forma, você precisará manter na estrutura a quantidade corrente de buckets e o nível (número da rodada de divisões). Da mesma forma que na implementação do Hash Extensível, é necessário que as operações básicas de inserção, remoção e recuperação de entradas de dados sejam implementadas, bem como as operações específicas inerentes à estrutura, como a divisão de páginas.

Data da entrega: segunda-feira - 13 de abril de 2015 durante a aula.