

GBC053 - Gerenciamento de Bancos de Dados

Aula 7 Indexação

Humberto Razente
humberto.razente@ufu.br

O que é um índice?

- Conceito básico: chaves e referências
- Permite impor uma ordem em um arquivo sem ter que rearranjar o arquivo
 - e mais, considere que se esteja procurando um livro na biblioteca:
 - deseja-se procurar por autor, título ou assunto
 - se houvesse 3 cópias de cada livro, poderia-se ter 1 prateleira ordenada por autor, outra prateleira por título, outra por assunto → 3 arquivos de dados ordenados com atributos diferentes?
 - com índices, é possível realizar buscas com atributos diferentes em um mesmo arquivo

Operações requeridas para manter um arquivo de índice

- Criar arquivos de dados e índices
- Carregar o arquivo de índice na memória
- Reescrever o arquivo de índice após usá-lo
- Adicionar registros no arquivo de dados
- Atualizar registros no arquivo de dados
- Remover registros no arquivo de dados
- Atualizar índice para refletir as mudanças nos registros no arquivo de dados

Operações requeridas para manter um arquivo de índice

- Operações devem envolver o uso de buffers
- Deve-se haver um mecanismo para saber se o índice foi alterado
 - para determinar se o índice em disco está desatualizado, e nesse caso, deverá ser reescrito

Índices muito grandes para serem mantidos em memória

- Até agora, assumiu-se que os índices eram pequenos o suficiente para caberem em memória principal
- O acesso a um índice diretamente no disco tem as seguintes desvantagens:
 - uma busca binária envolve vários **seeks** no disco
 - um rearranjo no índice por adição ou remoção requer um **shift** nos registros do arquivo
- Deve-se considerar:
 - organização **hash**
 - índice com estrutura em árvore ou multinível

Índices muito grandes para serem mantidos em memória

- Mas antes de descartar o uso de índices simples, deve-se considerar:
 - um índice simples torna possível usar uma busca binária → o índice pode ser uma estrutura com registros de tamanho fixo, mesmo se indexação for de registros de tamanho variável
 - se as entradas do índice forem bem menores que os registros de dados, ordenar e manter o índice pode ser menos caro que ordenar e manter o arquivo de dados

Índexação para prover acessos por múltiplas chaves

- Em geral não se procura na biblioteca pelo livro com número do catálogo QA331T5
 - identificador único é necessário → chave primária
 - procura dá-se por consultas tipo
 - "procuro um livro sobre estruturas de dados"
 - "procuro um livro do Donald Knuth"
 - "procuro um livro de estruturas de dados do Donald Knuth"
 - chaves secundárias

Índexação para prover acessos por múltiplas chaves

Composer index

<i>Secondary key</i>	<i>Primary key</i>
BEETHOVEN	ANG3795
BEETHOVEN	DG139201
BEETHOVEN	DG18807
BEETHOVEN	RCA2626
COREA	WAR23699
DVORAK	COL31809
PROKOFIEV	LON2312
RIMSKY-KORSAKOV	MER75016
SPRINGSTEEN	COL38358
SWEET HONEY IN THE R	FF245

- Manutenção de índices secundários deve ser feita
 - inserções de registros no arquivo de dados
 - alterações
 - remoções

Recuperação baseada em combinações de chaves secundárias

- A combinação de duas ou mais chaves secundárias pode ser usada para buscar um subconjunto de registros no arquivo de dados
 - Uso de operadores booleanos AND e OR
 - Intersecções
 - Uniões
 - Possibilita múltiplas perspectivas dos registros

Recuperação baseada em combinações de chaves secundárias

- Exemplo
 - "encontrar todos os livros de estruturas de dados do Donald Knuth"
 - assunto = "estruturas de dados"
 - autor = "Donald Knuth"
- Envolve o "casamento" de valores em dois índices

Leitura complementar

- Capítulo "**Indexing**" do livro
 - Folk et al. "File Structures: An Object-Oriented Approach with C++", Editora Pearson, 3ª edição, 1998