

Classificação e Pesquisa de Dados

Aulas 21 e 22

Operações sobre Arquivos; Arquivos Seqüenciais e
Arquivos Seqüenciais Indexados

UFRGS

INFO1124

1

O que é um índice?

- Estrutura de acesso formada por uma coleção de pares do tipo <chave, endereço>, cada um deles associando o valor de uma chave de acesso a um endereço do arquivo na memória secundária (disco)
- Principais vantagens
 - Maior rapidez de busca
 - Permite múltiplos caminhos de acesso
 - Oferece maior facilidade de acesso a registros de tamanho variável

2

Tipos de Índices

- Índice primário
 - Índice associado a chave de ordenação
- Índices secundários
 - Não estão associados a chave de ordenação
- Índice *clustering* (agrupamento)
 - Campo de classificação diferente da chave primária

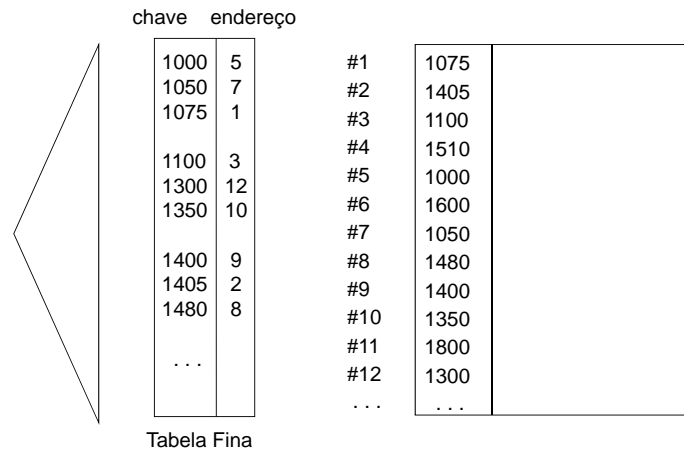
3

Arquivo Indexado

- Caracterização
 - Nesta organização, os registros são acessados, sempre, através de um ou mais índices, não havendo qualquer compromisso com a ordem física de instalação dos registros no arquivo
 - Consequência direta → maior eficiência na operação de inserção de um registro

4

Arquivo Indexado



5

Índices

- Podem existir tantas quantas forem as chaves de acesso aos registros
- Um índice contém uma entrada para cada registro

ID	#
00101	1
00102	2
00104	3
:	:
2193	n

#	ID	Nome	Salário
1	00101	Leonardo	2.500,00
2	00102	Aline	3.500,00
3	00104	João	1.800,00
:	:	:	:
n	2193	Maria	4.500,01

Nome	#
Aline	2
João	3
Leonardo	1
:	:
Maria	n

Salário	#
1.800,00	3
2.500,00	1
3.500,00	2
:	:
4.500,01	n

6

Índices

- As entradas do índice são ordenadas pelo valor da chave de acesso
- Cada entrada em um índice é constituída por um par :
 - <Chave_registro, endereço_registro>
- A seqüencialidade física no índice visa tornar mais eficiente o processo de busca e permitir o acesso serial ao arquivo

7

Índices na prática

- Arquivo de dados organizado pelo índice primário (chave-primária de uma tabela em um BD)
- Arquivo de índice primário aponta para elemento inicial de um bloco (seletivo)

Nome	Bloco
Adão	1
Alberto	2
:	:
Viviane	n

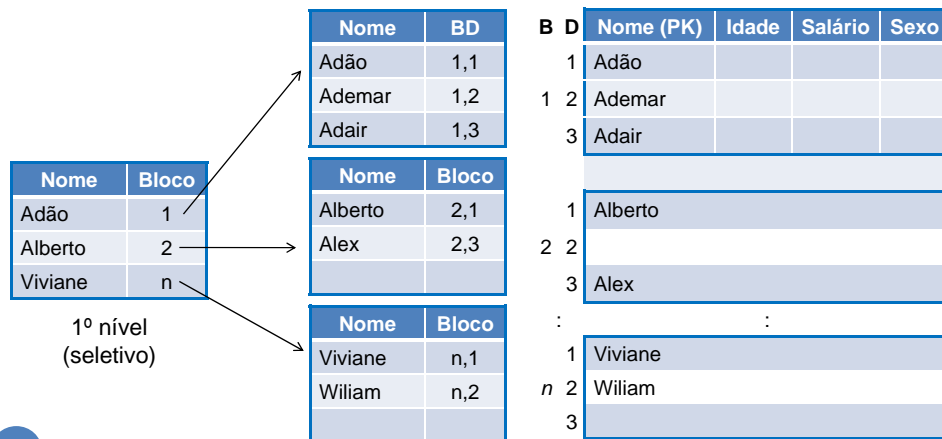
Índice seletivo

B D	Nome (PK)	Idade	Salário	Sexo
1	Adão			
2	Ademar			
3	Adair			
1	Alberto			
2				
3	Alex			
:	:	:	:	:
1	Viviane			
2	William			
3				

8

Índices na prática

- Arquivo de índice pode ter múltiplos níveis



9

Tipos de Índices

- Exaustivo (**Denso**)
 - Contém uma entrada para cada registro do arquivo
- Seletivo (**Esparsos**)
 - Possui entradas apenas para um subconjunto dos registros
 - Ex.: Subconjunto que define todos os alunos cujo curso = "ciência da computação" e ano_ingresso = "2010"

10

Desvantagens

- Necessidade de atualização de todos os índices quando
 - Um registro é inserido no arquivo
 - Alteração de um registro envolve atributos associados a índices
- Quanto maior é o número de índices, mais onerosa é a operação de inserção

11

Acesso a um registro

- Acesso serial
 - Através de um dos índices, de acordo com a sequência desejada
 - A maioria dos acessos requer apenas uma leitura do disco
- Acesso aleatório
 - Efetivação de uma busca sobre o índice
 - De um modo geral requer acessos a um único cilindro
 - Após a determinação do endereço, a leitura do registro de dados pode requerer o acesso a um cilindro distinto

12

Inserção de um registro

- O registro é armazenado em qualquer endereço vago do arquivo
- Seus pares <chave_registro, endereço_registro> relativos aos índices existentes são inseridos nos índices correspondentes
- Ocorrência de overflow
 - Alternativa de solução: árvores B

13

Exclusão de um registro

- Área de dados ocupada é liberada
- São removidas as entradas do índices correspondentes por
 - Retirada física do par
 - Utilização de uma marca
- Devem ser determinados os índices seletivos afetados

14

Alteração de um registro

- Identificação do registro
 - Via argumento de pesquisa
 - Endereço é determinado por uma busca sobre o índice
 - O endereço do registro é conhecido
- Tipo da alteração
 - Normal
 - Lê → altera → grava na mesma posição
 - Com aumento do tamanho do registro
 - Lê → exclui → insere registro atualizado

15

Procedimento de reorganização

- Não é muito necessária
 - Registros excluídos são ocupados pelas novas inserções
 - Alteração/exclusão refazem os índices

16