

Exercícios de estruturas de indexação

- 1) Se índices aceleram a busca, quais seriam as razões para não criar índices em todos os atributos de uma tabela?
- 2) Consider the following relations:
Emp(eid: integer, ename: varchar, sal: integer, age: integer, did: integer)
Dept(did: integer, budget: integer, floor: integer, mgr eid: integer)

Salaries range from \$10,000 to \$100,000, ages vary from 20 to 80, each department has about five employees on average, there are 10 floors, and budgets vary from \$10,000 to \$1 million. You can assume uniform distributions of values.

For each of the following queries, which of the listed index choices would you choose to speed up the query? If your database system does not consider index-only plans (i.e., data records are always retrieved even if enough information is available in the index entry), how would your answer change? Explain briefly.

1. Query: Print ename, age, and sal for all employees.

- (a) Clustered hash index on #ename, age, sal fields of Emp.
- (b) Unclustered hash index on #ename, age, sal fields of Emp.
- (c) Clustered B+ tree index on #ename, age, sal fields of Emp.
- (d) Unclustered hash index on #eid, did fields of Emp.
- (e) No index.

2. Query: Find the dids of departments that are on the 10th floor and have a budget of less than \$15,000.

- (a) Clustered hash index on the floor field of Dept.
- (b) Unclustered hash index on the floor field of Dept.
- (c) Clustered B+ tree index on #floor, budget fields of Dept.
- (d) Clustered B+ tree index on the budget field of Dept.
- (e) No index

3. Considere um disco com tamanho de bloco $B = 512$ bytes. Um ponteiro de bloco é do tamanho $P = 6$ bytes, e um ponteiro de registro é do tamanho $Pr = 7$ bytes. Um arquivo b possui $r = 30.000$ registros de EMPREGADO de tamanho fixo. Cada registro possui os seguintes campos: NOME (30 bytes), SSN (9 bytes), CODIGODEPARTAMENTO (9 bytes), ENDE-RECO (40 bytes), TELEFONE (9 bytes), DATANASC (8 bytes), SEXO (1 bytes), CODIGOCARGO (4 bytes), SALARIO (4 bytes, número real). Um byte adicional é usado como marcador de exclusão.

- (a) Calcule o tamanho do registro R em bytes.
- (b) Quantos registros “cabem” em um bloco?
- (c) Calcule o número de blocos do arquivo b .

4) Considere um disco com tamanho de setor igual a 512 bytes, 2000 trilhas por superfície, 50 setores por trilha, cinco pratos de lado duplo e tempo de busca médio de 10mseg. Suponha que o tamanho do bloco é de 1024 bytes. Suponha que um arquivo contendo 100.000 registros de 100 bytes cada deva ser armazenado neste disco.

- 4.1) Quantos registros cabem em um bloco?
- 4.2) Quantos blocos são necessários para armazenar a tabela inteira? Se o arquivo estiver organizado sequencialmente em disco, quantas superfícies serão necessárias?
- 4.3) Quantos registros de 100 bytes cada podem ser armazenados neste disco?
- 4.4) Qual o custo da busca sem índice nesta tabela?
- 4.5) Qual o custo da busca em um índice primário com 15 bytes?

4.6) Qual o custo da busca de um índice secundário denso com 15 bytes?

5) Considere a seguinte relação Aluno:

<u>co</u> <u>dA</u> <u>lu</u> <u>no</u>	nome	cidade	codCurso	renda
1	Ana	POA	1	1300.50
2	Paulo	Gravataí	1	2641.56
3	José	POA	2	1300.50
4	Maria	Canoas	4	2300.54
5	Marcos	Canoas	6	2300.57
6	Mariana	POA	4	3300.88
7	Juliana	Canoas	7	2300.98
8	Eduardo	POA	9	4305.38
9	Ronaldo	Viamão	11	3137.38
10	Carlos	Viamão	4	2360.28
11	Pedro	Gravataí	7	2370.88
12	Sérgio	POA	7	1360.98
...				
1000	Vando	POA	5	1500.52

Estruturas básicas de indexação. Considere as situações:

1. Supondo que o arquivo físico de dados está ordenado pelo campo chave codAluno.
 - i. Dos tipos de índices vistos até o momento, qual tipo poderia ser criado para este campo (primário, secundário, clustering)? Esparso ou denso? Justifique.
 - ii. Qual seria o arquivo de índices correspondente a este campo, considerando um índice primário esparsos com 10 entradas?
2. Supondo que um bloco armazene 10 tuplas do arquivo de dados da relação Alunos.
 - i. Quais acessos/operações de acessos seriam feitos para selecionar a tupla do aluno de código 214, supondo um índice esparsos primário na coluna codAluno, com 10 entradas no arquivo de índices, supondo que o arquivo de índices aponta para o primeiro registro do bloco? Uma vez localizado o bloco, como poderia ser feita a busca do registro dentro do bloco?
3. Supondo que o arquivo físico de dados está ordenado pelo campo cidade.
 - i. Dos tipos de índices vistos até o momento, qual tipo poderia ser criado para este campo (primário, secundário, clustering)? Esparsos ou denso? Justifique.
 - ii. Qual seria o arquivo de índices correspondente a este campo, considerando um índice esparsos com 10 entradas?
 - iii. Qual seria o arquivo de índices correspondente a este campo, considerando um índice denso? Quantas entradas existiriam no arquivo de índices?
 - iv. Supondo que um bloco armazene 5 tuplas do arquivo de dados da relação Alunos. Considere que existe um índice clustering esparsos na coluna cidade.
 1. Quais acessos/operações de acesso seriam feitos para selecionar as tuplas dos alunos que moram em POA, supondo que o arquivo de índices aponta para o primeiro registro do bloco que contém a informação pesquisada? (considere que o grupo de registros que compartilham diferentes valores de clustering podem estar no mesmo bloco)
 2. Quais acessos/operações de acesso seriam feitos para selecionar as tuplas dos alunos que moram em POA, supondo que o arquivo de

índices aponta para o primeiro registro do bloco que contém a informação pesquisada? (considere que o grupo de registros que compartilham o mesmo valor de clustering podem estar em blocos separados)