

Sistemas de Bases de Dados
1º teste 2014/15
Duração: 1 hora + 30 minutos (sem consulta)

Grupo 1

Considere parte de uma base de dados registando informação sobre contribuintes identificados por um NIF, podendo ser individuais ou colectivos (empresas), empresas e seus accionistas e respectivo capital social (montante em dinheiro), e atividade económica realizada pela empresa classificada de acordo com códigos de atividade (CAEs) tendo obrigatoriamente um código primário e no máximo 3 secundários (não primários). Essa base de dados inclui as seguintes tabelas não sendo permitidos valores nulos em nenhum atributo (onde os atributos que constituem a chave primária se encontram sublinhados):

contribuintes({ <u>NIF</u> ,Rua,Local,CodPostal,...})	peessoas({ <u>NIFPess</u> ,Nome,Sexo,DataNasc,...})
empresas({ <u>NIFEmpr</u> ,Nome,Tipo,DataInício,...})	acionistas({ <u>NIFEmpr</u> , <u>NIFContribuinte</u> ,Capital})
atividades({ <u>NIFEmpr</u> , <u>CAE</u> ,Primário})	codigosAtividade({ <u>CAE</u> ,Designação,...})

Todas as chaves são formadas por atributos inteiros. O atributo Tipo em empresa tem 4 valores possíveis e o atributo Primário da tabela atividades é booleano. Para cada uma das tabelas existe um índice non-clustered de árvore B+ sobre o(s) atributo(s) da chave primária. Além disso são definidas na base de dados as seguintes *foreign keys*: de NIFPess em pessoas e NIFEmpr em empresas para contribuintes(NIF), de NIFEmpr em acionistas e atividades para empresas, NIFContribuinte para contribuintes(NIF), e CAE em atividades para codigosAtividade.

Tendo em conta o sistema de gestão de base de dados usado, tipicamente cabem num bloco 20 tuplos da tabela contribuintes, pessoas e empresas, 40 tuplos da tabela codigosAtividade e 100 tuplos das tabelas atividades e acionistas.

Sabemos ainda que num dado momento existem 6 milhões de contribuintes, dos quais 2 milhões empresas e 4 milhões pessoas, a tabela acionista tem 8 milhões de tuplos e a de atividades 5 milhões. A tabela de codigosAtividade tem 4000 tuplos.

Nota: Neste grupo, sempre que se solicitarem exemplos, estes devem ser **exclusivamente** sobre esta base de dados. Adicionalmente, **todas** as respostas deverão conter uma **breve justificação**.

1 a) Indique qual o tipo de índice que utilizaria para efetuar eficientemente consultas pelo nome da empresa, sabendo que os nomes das empresas são únicos.

1 b) Os códigos de atividades são bastante estáveis ao longo do tempo e definidos oficialmente. Discuta a opção de se utilizar um índice de hash estático para a chave primária da tabela codigosAtividade em vez do índice non-clustered em árvore B+.

1 c) Considere que está definido um índice non-clustered de árvore B+ sobre o atributo Local da tabela contribuintes. Indique se esse índice poderia ser útil para responder à seguinte consulta, sabendo que 90% dos contribuintes tem morada fiscal em Lisboa:

SELECT CodPostal FROM contribuintes WHERE Local >= 'Lisboa'

1 d) Repare que o índice para a chave primária da tabela atividades é composto por dois atributos. Indique se na sua opinião é indiferente ou não criar o índice com a seguinte ordem dos atributos (NIFEmpr,CAE) versus (CAE,NIFEmpr).

1 e) Apresente uma consulta para o qual seria aconselhável a utilização de um índice bitmap.

1 f) A seguinte interrogação é muito utilizada para obter um resumo da informação da empresa a partir do seu NIF (indicado por ?). Indique se uma organização em *multitable clustering file organization* poderia ser benéfica e em caso afirmativo indicando como organizaria a informação em cada bloco.

SELECT * FROM Empresas NATURAL JOIN atividades NATURAL JOIN codigosAtividade WHERE Primário = TRUE AND NIFEmpr = ?

Grupo 2

Nota: Dê respostas breves

2 a) Num índice em árvore B+ secundário com 4 filhos e de altura 3 (a raiz da árvore tem altura 1), quantos registos da tabela podem ser indexados?

2 b) Construa um índice esparsos sobre a seguinte relação com 2 atributos, contendo os seguintes tuplos cuja chave de pesquisa é o primeiro atributo e assumindo que cada bloco alberga no máximo 3 tuplos e 3 entradas do índice.

$$\{(7,A),(1,C),(10,B),(8,F),(5,D),(15,E),(3,G),(6,I),(4,H)\}$$

2 c) Compare em termos de desempenho a utilização de dispersão estática (*static hashing*) com dispersão extensível (*extendable hashing*), assumindo que não há *buckets* cheios.