

DI- FCT NOVA

Sistemas de Bases de Dados**1º teste (modelo)****Duração: 1 hora + 30 minutos (sem consulta)****Grupo 1**

Considere uma base de dados para registar consultas feitas num hospital e receitas de fármacos feitas em cada consulta. Essa base de dados inclui as seguintes tabelas (onde em cada uma delas apenas se apresentam parte dos atributos, e os atributos que constituem a chave primária estão sublinhados):

fármacos({ <u>NumF</u> , NomeF, ...})	consultas({ <u>NumCons</u> , codMed, NomePaciente, Data, ...})
receitas(<u>NumCons</u> , <u>NumF</u> , quantidade, ...)	médicos({ <u>CodMed</u> , Nome, Sx, Especialidade, ...})

Para cada uma das tabelas há um índice clustered de B+-tree sobre o(s) atributo(s) da chave primária.

Tendo em conta o sistema de gestão de bases de dados usado, tipicamente cabem num bloco 20 tuplos da tabela de fármacos ou da tabela médicos, ou 40 tuplos da tabela consultas ou da tabela receitas. Assuma que o sistema que suporta esta base de dados permite ter em memória apenas 100 blocos.

Sabemos ainda que num dado momento a tabela fármacos tem 500.000 tuplos, a de médicos 50 tuplos, a de consultas 10.000 tuplos e a de receitas 50.000 tuplos.

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

1 a) Apresente um exemplo de pergunta em que seria vantajoso existir uma organização de ficheiros com “multi-table clustering”. Apresente uma outra pergunta para a qual a organização que apresentou seria muito desvantajosa.

1 b) Assuma que o índice da tabela de receitas foi criado com o comando:

```
create index p_pk on receitas(NumCons, NumF)
```

Qual das seguintes perguntas lhe parece ter uma execução mais eficiente?

select * from receitas	select * from receitas
where NumCons = 1111 and NumF < 1111	where NumCons < 1111 and NumF = 1111

1 c) Dê um exemplo de uma pergunta envolvendo pelo menos dois atributos da tabela consultas para a qual a criação de índices multi-atributo não traria nenhuma vantagem, comparativamente ao uso apenas dos índices criados para as chaves.

1 d) Suponha agora que o sistema lhe permitia definir índices de bitmap. Apresente um exemplo duma pergunta em que a existência dum índice de bitmap tornaria muito mais eficiente a execução.

1 e) Considere que está definido na tabela receitas um índice não clustered de árvore B+ sobre o atributo quantidade. Considere ainda as perguntas resultantes de substituir <VALOR> por 1 ou por 100 em: **select * from receitas where quantidade = <VALOR>**

Para cada uma destas duas perguntas, seria mais ou menos eficiente usar o índice?

Grupo 2

Nota: Dê respostas **breves**.

2 a) Porque é que, em geral, em sistemas de bases de dados quando se apaga um tuplo a tabela não é rearranjada por forma a ocupar o espaço deixado livre por esse tuplo?

2 b) Apresente, graficamente, uma árvore B+, em que a dimensão máxima de cada nó é 4, contendo as seguintes chaves de pesquisa (e assumindo a ordem usual em inteiros).

(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

2 c) Porque é que, em geral, em sistemas de bases de dados não faz sentido usar a técnica de “static hashing” para organização de dados?