Exames Preliminares do Programa de Doutorado
Prova de Banco de Dados

Departamento de Ciência da Computação — IME-USP

Primeiro semestre de 1999.

Nome:

Atenção: Embora a soma dos pontos das questões seja 14, tente responder corretamente o máximo número de questões, a fim de maximizar as suas chances de uma boa avaliação. A nota é o máximo das somas das notas de questões, tomado sobre os subconjuntos de questões cujo valor soma 10. Seja claro e preciso em suas respostas. Se quiser, use figuras para ilustrar suas respostas.

Boa sorte!

Questão 1 (2,0 pontos)

Modelar a seguinte situação num diagrama de entidades e relacionamentos:

> Um banco de dados deve comportar dados sobre a mortalidade dos brasileiros. As informações sobre mortalidade devem ser dadas por sexo e pelas seguintes idades: natimorto, 0–1, 1–4, 4–12, 12–18, 18–25, 24–40, 40–60, mais que 60.
>
> Os dados de mortalidade devem ser anuais, mas infelizmente não é possível se realizar contagens populacionais nesta frequência. Desta forma, a natureza do dado pode variar de ano para ano, e pode ser: contagem, amostragem, estimativa e cálculo.
>
> Os dados de contagem são normalmente agrupados na granularidade municipal. Já as amostragens e estimativas podem ser realizadas em diversas granularidades: municipal, estadual, regional. Por fim, podemos desejar agrupar dados de menor refinamento, por exemplo, provendo dados de mortalidade infantil regional ou de todo o país a partir do dado estadual ou municipal.

Questão 2 (3,0 pontos)

Considere o seguinte esquema, inspirado pela série “Guerra nas Estrelas”:

NascidoEm(personagem, planeta)
ExisteEm(flor, planeta)
AlérgicoA(personagem, flor)

Expresse cada uma das consultas abaixo em álgebra relacional, em cálculo de tuplas e em SQL:

(a) Obter os nomes dos personagens que são alérgicos a pelo menos uma das flores existentes em seu planeta natal.

(b) Obter os nomes dos personagens que não são alérgicos a nenhuma das flores existentes em seu planeta natal.

(c) Obter os nomes dos personagens que são alérgicos a todas as flores existentes em seu planeta natal.

Questão 3 (1,0 ponto)

Considere o esquema relacional $R = (A, B, C, D)$ e sua decomposição em $R_1 = (A, B, D)$ e $R_2 = (B, C)$.

(a) Assuma que o conjunto de dependências funcionais $F = \{A \to B, D \to C\}$ está associado a R. Construa uma instância de relação $r(R)$ que ilustre o fato da decomposição de R em R_1 e R_2 não ser sem perdas de junção. Em outras palavras: exiba uma relação $r(R)$ que satisfaz F e para a qual não vale a igualdade $r = \Pi_{R_1}(r) \bowtie \Pi_{R_2}(r)$.

(b) Agora assuma que o conjunto de dependências funcionais $F = \{A \to B, D \to C, B \to C\}$ está associado a R. Construa instâncias $r_1(R_1)$ e $r_2(R_2)$ que ilustrem o fato da decomposição de R em R_1 e R_2 não preservar dependências. Em outras palavras: exiba relacões $r_1(R_1)$ e $r_2(R_2)$ tais que r_i satisfaz à restrição de F a R_i, mas $r_1 \bowtie r_2$ não satisfaz F.

Questão 4 (3,0 pontos)

Considere a relação

Ambulância(C, E, H, A, Z, R, P, T)

onde

C: Cliente, ou paciente transportado pela ambulância
E: Endereço do cliente
H: Hospital
A: N. de Ambulâcias disponíveis no hospital
Z: Zona da cidade
R: Responsável
P: Período
T: Telefone

A essa relação estão associadas as dependências funcionais

$$
\begin{aligned}
C &\rightarrow E\ H \\
C\ E &\rightarrow Z \\
E &\rightarrow H\ Z \\
H &\rightarrow A\ Z\ H \\
H\ P &\rightarrow R\ P \\
H\ R &\rightarrow R\ P
\end{aligned}
$$

bem como a dependência multivalorada H $\rightarrow\rightarrow$ T. Pede-se:

(a) Dê uma cobertura canônica para as dependências funcionais acima.

(b) Usando as dependências funcionais e multivalorada acima, decomponha Ambulância na quarta forma normal.

(c) Mostre quais dependências funcionais da cobertura canônica em (a) estão associadas a cada uma das relações obtidas em (b).

(d) A decomposição em (b) é sem perdas de junção? Ela preserva as dependências?

(e) É verdade que em Ambulância vale a seguinte dependência multivalorada: CEHA $\rightarrow\rightarrow$ ZRPT? Prove ou refute.

Questão 5

(2,0 pontos)

Considere as relações $R_1 = (A, B, C)$ e $R_2 = (A, D, E)$. Para cada uma das situações abaixo esboçe um algoritmo que compute a junção natural $R_1 \bowtie R_2$ e estime o número de acessos a bloco numa execução desse algoritmo. Seus algoritmos devem explorar tanto quanto possível as estruturas físicas existentes. Descreva-os num nível de abstração bem alto, em português ou pseudo-código de alto nível. Suas estimativas devem ser feitas em termos de parâmetros como o número de blocos de dados de R_i, o número de tuplas de R_i, etc.

(a) Um índice de hash secundário é mantido sobre o atributo A de R_1 e nenhum índice é mantido sobre o atributo A de R_2. Assuma que cada "hash bucket" tem comprimento de 1 bloco.

(b) Duas árvores-B+ são mantidas como índices secundários, uma sobre o atributo A de R_1, a outra sobre o atributo A de R_2.

(c) Estruturas primárias são mantidas sobre ambas as relações, de modo que cada uma delas é fisicamente armazenada em ordem crescente dos valores do atributo A.

Questão 6

(1,0 ponto)

Um sistema S consiste de 3 máquinas M_i ($i = 1, \ldots, 3$), cada uma com 4 processadores $P_{i,j}$ ($j = 1, \ldots, 4$) e com 2 discos $D_{i,k}$ ($k = 1, \ldots, 2$). O sistema operacional permite U usuários ao mesmo tempo, com um máximo de C processos concorrentes. O sistema S é usado para sediar um banco de dados (com propriedades ACID) que aceita até T transações concorrentes (T pode ou não ser maior que C), não tem replicação, e faz controle de concorrência no nível de registro.

(a) Suponha que o controle de concorrência é por travamento bifásico (*two-phase locking*). Quantas transações podem ser simultaneamente executadas no sistema?

(b) Responda novamente à pergunta acima, agora supondo que o controle de concorrência é por ordenação de marcas de tempo (*timestamp ordering*).

(c) Quantas transações que accessam o mesmo registro r no disco $D_{1,1}$ podem ser simultaneamente executadas no sistema caso o controle de concorrência seja por travamento bifásico?

(d) E caso o controle de concorrência seja por ordenação de marcas de tempo?

Questão 7 (2,0 pontos)

Considere as três transações

T_1: read(X); write(X); read(Y); write(Y)

T_2: read(Z); read(Y); write(Y); read(X); write(Y)

T_3: read(Y); read(Z); write(Y); write(Z)

A figura abaixo mostra 2 escalonamentos (A e B) para essas transações.

T_1	T_2	T_3
	read(Z)	
	read(Y)	
	read(Y)	
		read(Y)
		read(Z)
read(X)		
read(X)		
		write(Y)
		write(Z)
	read(X)	
read(Y)		
write(Y)	write(X)	

Escalonamento A

T_1	T_2	T_3
		read(Y)
		read(Z)
read(X)		
write(X)		
	read(Z)	
read(Y)		
write(Y)	read(Y)	
	write(Y)	
	read(X)	
	write(X)	

Escalonamento B

(a) Determine se os escalonamentos A e B são aceitáveis pelo protocolo do travamento bifásico (*two-phase locking*).

(b) Determine se eles são aceitáveis pelo esquema de ordenação de marcas de tempo (*timestamp ordering*).

(c) Idem, pelo esquema multiversão com marcas de tempo (*multi-version timestamping*).

(d) Determine se esses escalonamentos são serializáveis.

Indique, nos casos (a), (b) e (c), que operações são efetuadas pelo sistema.