

Bases de Dados (CC2005), 2019/20, DCC/FCUP

Exame de época de recurso — 27/07/2020 — Duração: 2:00 ★

Número: Nome:

GRUPO A (32.5 %) — Responda na folha do enunciado.

A cotação individual deste grupo pode substituir a nota do teste SQL (caso seja melhor).

Em qualquer caso é obrigatória a sua resolução (contará para a nota global do exame).

1. Complete o código para criação de uma tabela `DOCUMENT` em uma BD de documentos “online” com os seguintes atributos: `DocId`, de tipo inteiro para o identificador do documento, a chave primária da tabela; `URL`, uma string de tamanho 128, sendo o URL único para cada documento (chave secundária); `Name`, uma string de tamanho 32 para o nome; `Type`, uma enumeração dos valores `' FOLDER '` ou `' FILE '`, onde `' FOLDER '` indica que o documento é uma pasta que pode conter outros documentos e `' FILE '` indica que o documento é um ficheiro; `Owner`, um atributo inteiro que é chave externa para `USER.UserId`, onde `USER` é outra tabela para os utilizadores do sistema, indicando o dono do documento; `Parent`, um atributo opcional que é chave externa para `DOCUMENT.DocId` indicando a pasta de que faz parte o documento. Todos os atributos excepto `Parent` são não-opcionais (não-nulos).

```
CREATE TABLE DOCUMENT (  
  DocId _____,  
  URL _____,  
  Name _____,  
  Type _____,  
  Owner _____,  
  Parent _____  
);
```

2. Tenha em conta as tabelas `DOCUMENT` e `USER` referidas na questão 1, e assuma adicionalmente que a tabela `USER` tem um atributo `Login` para o “login” de um utilizador. Escreva uma consulta SQL para obter o nome e URL de documentos do tipo ficheiro (`' FILE '`) que pertençam ao utilizador com login `' nick '`. Não precisa de especificar critério de ordenação.

3. Expresse a consulta anterior através de uma sequência equivalente de operações de álgebra relacional (AR). Notação para alguns operadores AR – σ : seleção; π : projeção; \mathcal{F} : agregação; \bowtie : junção; $*$: junção natural.

4. Escreva uma consulta SQL para obter o nome de documentos do tipo pasta do utilizador 'nick' que contenham 10 ou menos documentos (possivelmente 0), e o correspondente nº de documentos contidos em cada pasta. Os resultados deverão ser ordenados pelo de forma decrescente nº de documentos numa pasta e depois de forma crescente pelo nome da pasta (para pastas com igual nº de documentos).

5. Assinale com **F**, **P**, **T**, ou **V** cada uma das características consoante estas se relacionem respectivamente com (**F**) “stored functions”, (**P**) “stored procedures”, (**T**) “triggers” ou (**V**) vistas/“views”. Indique apenas **uma** das opções para cada uma das características abaixo.

- A invocação é feita através de uma expressão num comando SQL.
- Pode ser objecto de uma consulta.
- Pode ter as variáveis OLD e/ou NEW implicitamente definidas.
- Os parâmetros podem ser de entrada (IN), saída (OUT), ou entrada-saída (INOUT).
- Executa automaticamente em reacção a uma operação de manipulação de dados.
- A definição tem associada uma consulta de dados.
- A invocação é feita através da instrução CALL.
- Tem um tipo de retorno definido.

6. [escolha múltipla] Considere o fragmento de código de uma aplicação Python que acede a uma base de dados com a tabela DOCUMENT:

```
results = db.execute(  
    'SELECT * FROM DOCUMENT WHERE Name = %s AND UserId = %s',  
    name, userId  
) .fetchall()
```

Suponha que name e userId são valores arbitrário obtido de uma fonte externa à aplicação (ex. através de uma ligação à rede).

- O código é passível de injeção SQL.
- O código não é passível de injeção SQL porque usa uma instrução parameterizada.
- O código não é passível de injeção SQL porque a instrução em causa é do tipo SELECT.
- Uma instrução parameterizada não deixa que código SQL arbitrário seja injectado dinamicamente a partir de valores “maliciosos” passados à aplicação.
- Uma instrução parameterizada altera o esquema da BD dinamicamente por forma a inibir a injeção de SQL.
- O uso de instruções parameterizadas acarreta um custo acrescido para a execução de instruções ao nível do SGBD.

Bases de Dados (CC2005), 2019/20, DCC/FCUP

Exame de época de recurso — 27/07/2020 — Duração: 2:00 ★

Número: Nome:

GRUPO B (10 %) — Responda no enunciado.

Considere uma tabela `ACCOUNT` para uma conta bancária com atributos inteiros `Num` para o número da conta e `Value` para o saldo da conta, e que está definido um “stored procedure”

```
transfer(IN a INT, IN b INT, IN v INT)
```

de tal forma que `transfer(a, b, v)` executa internamente a seguinte sequência de passos:

— Inicia uma transação com `START TRANSACTION`.

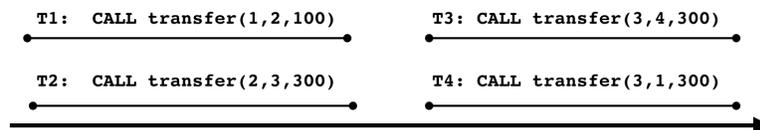
— Se `a` e `b` forem números de contas válidas e o saldo da conta `a` for igual ou superior a `v` (o código é omitido para o teste destas condições) então são executadas as seguintes instruções:

```
UPDATE ACCOUNT SET Value = Value - v WHERE Num = a;  
UPDATE ACCOUNT SET Value = Value + v WHERE Num = b;  
COMMIT;
```

isto é, o saldo de `a` é decrementado (debitado) no montante de `v`, o de `b` é incrementado (creditado) no montante de `v`, e a transação toma efeito permanente com `COMMIT`.

— Caso contrário, a transação iniciada é abortada com a execução de uma instrução `ROLLBACK`.

1. [escolha múltipla] Considere que as contas 1, 2, 3, e 4 estão definidas na BD com saldos iniciais de, respectivamente, 100, 200, 300, e 400. Na execução de 4 transações da forma ilustrada abaixo, temos que T1 e T2 executam em simultâneo, seguidas de T3 em simultâneo com T4.



Quais das seguintes opções são possíveis tendo em conta execuções serializáveis das transações, em termos de transações que completam com sucesso (i.e., terminam com `COMMIT`) e correspondentes estado da BD no para os saldos finais das contas 1 a 4?

	Transações completadas	Conta 1	Conta 2	Conta 3	Conta 4
<input type="checkbox"/>	Todas	300	0	0	700
<input type="checkbox"/>	T1 e T2	0	0	600	400
<input type="checkbox"/>	Todas menos T3	300	0	300	400
<input type="checkbox"/>	T1 e T3	0	300	0	700
<input type="checkbox"/>	Todas menos T4	0	0	300	700
<input type="checkbox"/>	T1 e T4	300	300	0	400

2. [escolha múltipla] Durante a execução de uma transação, uma anomalia de leitura irrepitível:

- É caracterizada pela leitura de dois valores diferentes para o mesmo registo durante a execução, em virtude do valor escrito por outra transação que terá completado entre as duas leituras.
- É caracterizada pela leitura de um valor escrito por outra transação ainda não completada e que acaba por abortar.
- É caracterizada pela leitura de um valor de uma tabela cujo acesso se encontra bloqueado por um “lock”.
- Uma execução serializável da transação e outras simultaneamente activas pode ser comprometida.
- Uma execução serializável da transação e outras simultaneamente activas nunca é comprometida.

GRUPO C (12.5 %) — Responda no enunciado.

x	2]2, 4]]4, 8]]8, 16]]16, 32]]32, 64]]64, 128]]128, 256]]256, 512]]512, 1024]]1024, 2048]]2048, 4096]
$\lceil \log_2(x) \rceil$	1	2	3	4	5	6	7	8	9	10	11	12

Neste grupo justifique as suas respostas apresentando todos os cálculos relevantes que efectuou.

Considere uma tabela T com $N = 100,000$ registos de tamanho fixo igual a $R = 100$ bytes, armazenada num ficheiro ordenado pela chave primária K de T , onde o tamanho tomado por um valor de K é de 56 bytes. Assuma ainda que cada bloco de disco tem um tamanho de $B = 4096$ (2^{12}) bytes, e que uma referência a um bloco de disco tem um tamanho de 8 bytes.

1. Indique o número de blocos em disco necessários para o armazenamento de T , e o número máximo de acessos a blocos de disco para localizar um registo com base num valor para K .

2. Se criarmos um índice primário sobre o ficheiro para a tabela T , quantos blocos em disco ocupará o índice? Usando o índice, qual será o número máximo de acessos a blocos de disco para localizar um registo com base num valor para K ?

3. Caracterize um índice de nível 2 sobre o índice anterior. Indique novamente quantos blocos em disco serão necessários e o número máximo de acessos a blocos de disco para localizar um registo com base num valor para K usando o índice de nível 2.

GRUPO D (32.5 %) — Responda em folha de exame.

Considere os seguintes requisitos para uma base de dados de um serviço de video-conferência (ex. análogo ao Zoom):

— Um utilizador é caracterizado por um identificador de login único, um nome, um endereço de email, um nº de telefone opcional, e dados de cartão de crédito compostos por número, data validade e dígitos de verificação CVV.

— Cada utilizador subscreve um plano de utilização e em que cada plano tem um nome único (ex. “Free”, “Pro”, ...) e um preço mensal associado. Um plano pode ser subscrito por múltiplos utilizadores mas também ainda não ter subscritores. Cada subscrição tem uma data de activação associada.

— Cada utilizador pode configurar várias salas de video-conferência de sua pertença, em que cada sala tem um URL (“link”) único associado.

— Para cada sala de video-conferência, podem ser activadas várias sessões ao longo do tempo. Cada sessão é definida por um identificador único, data e hora de início, e duração opcional. Uma duração indefinida significa que a sessão está activa e não terminou ainda.

— Em cada sessão participam um ou mais utilizadores. Uma sessão pode ser gravada, existindo nesse caso um único vídeo correspondente à gravação da sessão. Cada vídeo tem associado um URL único e um tamanho. Esse vídeo pode ser partilhado depois da sessão por utilizadores. Ao longo do tempo, cada utilizador pode participar em sessões de várias salas de video-conferência ou aceder a vários vídeos que tenham sido partilhados consigo.

— Podem haver utilizadores sem qualquer sala criada, participação em sessões, ou acesso a vídeos de gravação de sessões.

1. Apresente um modelo ER na forma de um diagrama para a base de dados.
2. Apresente um esquema relacional correspondente à tradução do modelo ER.

GRUPO E (12.5 %) — Responda em folha de exame.

Considere uma BD simplificada para uma biblioteca em que:

— Os livros são identificados unicamente por um ISBN (ISBN) e têm associado um título (Título).

— Um livro pode ter várias cópias, sendo cada cópia (Cópia) numerada para cada livro (1,2, ...).

— Um utente da biblioteca é identificado de forma única pelo número de utente (UNum) e tem associado um nome (UNome).

— Um empréstimo por parte de um utente diz respeito a uma ou mais cópias de livros distintos, tem um identificador único (EId) e associado uma data (EData).

Assuma o seguinte esquema 1NF para a BD e exemplos de entradas correspondentes na BD:

BIBLIOTECA (ISBN, Cópia, Título, UNum, UNome, EId, EData)

ISBN	Cópia	Título	UNum	UNome	EId	EData
9789722526289	1	Sonetos	1	José Manuel	1	2020-07-20
9789722709620	1	Os Lusíadas	1	José Manuel	1	2020-07-20
9789722526289	2	Sonetos	2	Maria José	2	2020-07-20
9789720049759	1	A Mensagem	2	Maria José	2	2020-07-20
9789720049759	2	A Mensagem	3	Xavier Gusmão	3	2020-08-01
9789722526289	2	Sonetos	3	Xavier Gusmão	4	2020-08-15

1. Identifique as dependências funcionais entre atributos em BIBLIOTECA e o conjunto de atributos que formam a chave primária.
2. Explique porque é que o esquema não é 2NF e normalize-o para 2NF.
3. Explique se o esquema que obteve na questão anterior é ou não também 3NF. Normalize-o para 3NF se não for o caso.