

## CC42A - Examen

Profs. Claudio Gutiérrez y Gonzalo Navarro

25 de Noviembre de 2002

### P1 (1.5 ptos)

Considere el esquema  $R(A, B, C, D, E)$  con las dependencias funcionales  $A \rightarrow BC$ ,  $CD \rightarrow E$ ,  $B \rightarrow D$  y  $E \rightarrow A$ .

1. Demostrar que  $R'(A, B, C)$  y  $R''(A, D, E)$  es una descomposición sin pérdida de información. Esto significa que se pueden reconstruir las relaciones originales a partir de las descomposiciones.
2. Demostrar que la descomposición en (1) no preserva las dependencias funcionales.
3. Descomponer el esquema original (con sus dependencias funcionales) en 3FN (por supuesto no debe haber pérdida de información y debe preservar las dependencias funcionales.)

### P2 (1.5 ptos)

1. Considere la Base de Datos

Emp(nombre-emp, calle, ciudad)  
Jefe(nombre-emp, nombre-jefe)  
Trabaja(nombre-emp, empresa, sueldo)

En SQL escriba:

- (a) Consulta que responda a: Empresa que tiene el mayor número de empleados.
  - (b) Consulta que responda a: Los jefes que tienen sólo un subordinado que vive en Santiago.
  - (c) Modifique la BD incrementando en 8 % el sueldo de los jefes del BancoEstado.
2. (Algebra Relacional)
    - (a) Expresé  $R \div S$  en función de  $\pi$ ,  $\sigma$  y  $\bowtie$ . (Recuerdo:  $R(Z) \div S(X)$  para conjuntos de atributos  $X \subseteq Z$  se define como la relación  $T[Y]$ , donde  $Y = Z - X$ , que cumple: Una tupla  $t$  está en  $R \div S$  sii para cada tupla  $t_S \in S$ , hay una tupla  $t_R \in R$  con  $t_R[Y] = t$  y  $t_R[X] = t_S$ .)
    - (b) Considere las expresiones:  $\sigma_C(R) \cup \sigma_C(S)$  y  $\sigma_C(R \cup S)$ . ¿Son iguales? (Demuestre o dé un contraejemplo si son diferentes). Dé una expresión en SQL equivalente a cada una.

### P3 (1.5 ptos)

Se introduce un nuevo tipo de campo, el tipo *texto*. A éste se le aloca una cantidad fija de bytes al momento de definición del esquema, y se supone que se llena con una secuencia de palabras. Las consultas que interesa realizar sobre este tipo de campo son: (1) Dada una palabra  $w$  y un campo  $a$ , entregar las filas de la tabla donde  $w$  aparezca en el campo  $a$ . (2) Idem una frase (secuencia de palabras). (3) Join de dos campos de texto, donde se entiende que calzan cuando comparten al menos una palabra.

- (a) Diseñe un índice capaz de resolver eficientemente operaciones de selección del tipo (1). Explique la estructura y la forma de buscar. Evalúe el costo de búsqueda.
- (b) Indique cómo reflejar en este índice las inserciones y borrados de campos. Evalúe el costo.
- (c) Indique cómo realizar las consultas tipo (2) sobre su índice. Recomienda algún cambio?
- (d) Proponga estrategias para realizar el join (3) de campos de texto.

#### P4 (1.5 ptos)

Se propone el siguiente esquema de control de concurrencia. Se define un árbol donde los nodos son los elementos de la base de datos (no confundir con múltiple granularidad). Para simplificar existe sólo el lock exclusivo. Las reglas del protocolo son: (1) Una transacción puede tomar su primer lock sobre cualquier elemento. (2) Los siguientes locks sólo pueden tomarse si ya se posee un lock sobre el padre. (3) Se puede soltar un lock en cualquier momento. (4) No se puede retomar un lock sobre un elemento sobre el que ya se soltó un lock.

- (a) ¿Cómo funciona en la práctica este protocolo si el árbol consiste de una única raíz (correspondiente a un elemento falso, sólo con el propósito de hacer funcionar el protocolo) y  $n$  hijos de esa raíz?
- (b) ¿Cómo funciona en la práctica este protocolo si el árbol es en realidad una lista lineal con todos los elementos, cada uno hijo del anterior?
- (c) Muestre que este protocolo asegura serializabilidad.
- (d) Muestre que este protocolo asegura que no se producen deadlocks.
- (e) Indique ventajas y desventajas de este protocolo sobre el protocolo de dos fases.

# CC42A - Examen

Profes. Claudio Gutiérrez y Gonzalo Navarro

2 de Julio de 2003

Tiempo: 2.5 horas

Una hoja de apuntes manuscrita

Responder en hojas separadas

## P1 (1.5 ptos)

Notown Records ha decidido almacenar información sobre músicos que tocan en sus álbumes (así como otra información de la compañía) en una base de datos. La compañía, inteligentemente, decidió contratarlo a Ud. como diseñador.

La siguiente información describe la situación que la base de datos de Notown debe modelar:

1. Cada músico que graba para Notown tiene un RUT, un nombre, una dirección y un número de teléfono. Algunos músicos que ganan poco usualmente comparten una misma dirección, y ninguna dirección tiene más de un teléfono.
2. Cada instrumento que se usa en canciones grabadas en Notown tiene un nombre (ej. guitarra, sintetizador, flauta, etc.) y una llave musical (ej. Do menor, Si bemol, etc.)
3. Cada álbum que se graba en el sello Notown tiene un título, un derecho de autor, y un identificador de álbum.
4. Cada canción grabada en Notown tiene un título, un autor y una duración.
5. Cada músico puede tocar varios instrumentos, y un instrumento dado puede ser tocado por varios músicos.
6. Cada álbum tiene un número de canciones, pero ninguna canción debe aparecer en más de un álbum.
7. Cada canción es tocada por varios músicos, y un músico puede tocar más de una canción.
8. Cada álbum tiene exactamente un músico que actúa como productor. Por supuesto, un músico puede producir varios álbumes.

Diseñe un esquema conceptual para Notown y dibuje un esquema ER. Indique todas las llaves y las restricciones de cardinalidad, y cualquier suposición que Ud. haga. En caso que existan restricciones que no pueda capturar en el diagrama ER, explique por qué no puede hacerlo.

P2 (1.5 ptos)  $AB \rightarrow A, B, C, D, E, G$      $BC \rightarrow B, C, A, D, E, G$   
 $AD \rightarrow A, D, E, G$      $E \rightarrow E, G$   
 $B \rightarrow B, D$

Considere el esquema  $R(A, B, C, D, E, G, H)$  y el conjunto de dependencias funcionales  $\{AB \rightarrow C, AD \rightarrow E, B \rightarrow D, BC \rightarrow A, E \rightarrow G\}$ .

Para cada uno de los siguientes conjuntos de atributos, haga lo siguiente:

1. Compute el conjunto de dependencias que vale sobre el conjunto y escriba un conjunto cobertor minimal.

2. Mencione la forma normal más fuerte que no es violada por el esquema que contiene estos atributos.

- (a) ABCD, (b) ABCEG, (c) BCEH.

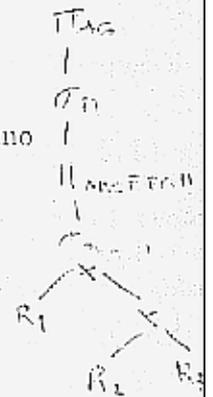
### P3 (1.5 ptos)

Dados los esquemas  $R_1(A, B, C)$ ,  $R_2(D, E, F)$  y  $R_3(G, H, I)$ , y la vista  $V$  definida como

$$V = \pi_{ABCEFGH} \sigma_{A=D \wedge (B \neq I \vee C \neq F \vee H)} (R_1 \times R_2 \times R_3),$$

considere la consulta

$$\pi_{AG} \sigma_{A < C \vee G \wedge E = H} V$$



(a) Dibuje el árbol sintáctico de la consulta (expandiendo  $V$ ).

(b) Optimice la consulta.

(c) Considere los árboles alternativos que se obtienen utilizando las reglas conmutativa y asociativa del producto cartesiano. Optimícelos, y discuta cuál elegiría para ejecutar la consulta y por qué.

### P4 (1.5 ptos)

Cuando un sistema que permite concurrencia deba recuperarse de una caída, se debe examinar el log para rehacer las transacciones que terminaron y tal vez deshacer las que no terminaron. Como la ejecución es concurrente, en el log se mezclan las acciones de distintas transacciones simultáneas. Supongamos que se usa el protocolo de dos fases.

- (i) Si las operaciones se rehacen en el orden en que aparecen en el log, ¿se preservan los resultados originales? Justifique.
- (ii) Si se rehacen las operaciones serializando las transacciones en el orden en que hicieron COMMIT, ¿se preservan los resultados originales? Justifique.
- (iii) Si se usa log con updates inmediatos, ¿se deben deshacer las transacciones no terminadas mientras se rehacen las de transacciones terminadas? Justifique.
- (iv) Si se usa log con updates diferidos, ¿se pueden simplemente ignorar las operaciones realizadas por transacciones no terminadas? ¿Qué ocurriría si otra transacción que depende de uno de esos valores sí terminó? Justifique.