

Segundo Parcial de Fundamentos de Bases de Datos

Noviembre 2018

Duración: 3 horas y media

Presentar la resolución del parcial:

- Con las hojas numeradas y escritas de un solo lado. Comience cada ejercicio en una hoja nueva.
- Con la cantidad de hojas entregadas en la primer hoja.
- Con cédula de identidad y nombre en cada hoja.
- **Escrita a lápiz y en forma prolija.**

Ejercicio 1 (15 puntos)

Dado el esquema relación $R(A,B,C,D,E,G,H)$ y el conjunto de dependencias F sobre R .

$F = \{A \rightarrow DG, D \rightarrow BA, G \rightarrow C, CD \rightarrow EH\}$

Para cada una de las siguientes afirmaciones indicar si son verdaderas o falsas. Justificar todas las respuestas:

- (DG) es clave de R según F .
- A y D son las únicas claves de R según F .
- $\Pi_{R_1}(F) = \{G \rightarrow C\}$ siendo $R_1(C,D,G,H)$
- Sea $F_1 = \{A \rightarrow D, A \rightarrow G, D \rightarrow B, D \rightarrow A, G \rightarrow C, CD \rightarrow E, CD \rightarrow H\}$. F_1 es un cubrimiento minimal de F .
- R está en 3NF respecto a F .
- R no está en BCNF respecto a F .
- La siguiente descomposición tiene JSP: $R_1(A,H), R_2(B,C,D,E,G,H)$

Ejercicio 2 (15 puntos)

Dada la siguiente realidad:

En una agencia de viajes se quiere construir una base de datos con información acerca de las ventas de paquetes turísticos a clientes por parte de los vendedores de la agencia. Para esto se consideran los datos que se detallan a continuación.

Sobre los clientes se conoce su cédula de identidad, su nombre, fecha de nacimiento, dirección y teléfono. Cada paquete tiene un identificador, un destino principal, la aerolínea correspondiente al pasaje aéreo, la fecha de partida y un conjunto de nombres de hoteles (que son ofrecidos al cliente para que elija). Por otro lado, cada paquete incluye varios paseos. De cada paseo se conoce su nombre.

De los vendedores se conoce su cédula de identidad, su nombre, su teléfono y su fecha de ingreso a la agencia.

Cada vendedor puede realizar varias ventas de paquetes a varios clientes. En esta realidad, un mismo cliente puede realizar varias compras en la misma fecha, pero en una misma fecha es atendido por un sólo vendedor. Además un mismo cliente puede comprar un mismo paquete en varias fechas diferentes.

Utilizando los siguientes nombres para los atributos:

ci-cli	destino	ci-ven
nom-cli	aero	nom-ven
f-nac	f-part	tel-ven
dir-cli	nom-hotel	f-ing
tel-cli	nom-paseo	f-venta
paquete		

Se pide:

- Deducir todas las dependencias funcionales que se cumplen.
- Para la relación universal R (tabla que contiene todos los atributos), y el conjunto de dependencias funcionales hallado en la parte a), hallar todas las claves. Justificar.
- Dar una descomposición de R en 3NF sin pérdida de dependencias y con JSP. Aplicar el algoritmo dado en el curso (mostrando los pasos).
- Decir si la descomposición anterior está en BCNF, justificando. Si no está en BCNF, llevarla a dicha forma normal, aplicando el algoritmo dado en el curso (mostrando los pasos).
- Deducir todas las dependencias multivaluadas (no triviales) que se cumplen en cada tabla de la última descomposición obtenida. Para cada multivaluada que proponga, mostrar qué parte de la letra lo induce a pensar que ésta se cumple.
- Llevar la última descomposición obtenida a 4NF, aplicando el algoritmo dado en el curso (mostrando los pasos).

Ejercicio 3 (15 puntos)

Considere las siguientes transacciones:

T1: r1(X) w1(X) r1(Y) w1(Y) r1(Z) w1(Z) c1

T2: w2(Z) r2(Y) w2(Y) c2

1. Considere la siguiente historia H1: w2(Z) r2(Y) r1(X) w1(X) r1(Y) w1(Y) w2(Y) c2 r1(Z) w1(Z) c1 **Justificando cada una de sus respuestas**, indique si la historia H1 cumple con alguna de las siguientes propiedades:

- (a) serializable
- (b) recuperable
- (c) evita abortos en cascada
- (d) estricta

2. Dar un historia entrelazada H2 de T1 y T2, cuyas transacciones sigan 2PL básico, pero no siga 2PL estricto, usando lock's del tipo read/write.

3. ¿La historia H2 es serializable? **Justifique su respuesta.**

Ejercicio 4 (15 puntos)

Considere una porción de la base de datos de música construída en el laboratorio del curso, con el siguiente esquema relacional:

Temas (tema, título)

Contiene un conjunto de identificadores de temas musicales y sus títulos.

Lugares (idLugar, descLugar, pais)

Contiene información sobre los lugares. De cada lugar se conoce su identificador, su descripción y el país al que pertenece.

Compositores (comp, nombreComp, idLugar, nacionalidad)

Contiene información de los compositores de temas. De cada compositor se conoce su identificador, su nombre, el identificador de su lugar de nacimiento y su nacionalidad.

TemasComp (tema, comp)

Contiene los temas con los compositores que los crearon.

Además se cuenta con la siguiente información:

	Cant. Tuplas	Factor de bloqueo	Atributos	Índices
Temas	300000	150		Índice Primario sobre tema
Compositores	40000	80		Índice Primario sobre comp
TemasComp	630000	150		Índice Primario sobre <tema, comp>
Lugares	5000	100	Existen 100 países diferentes en la tabla, y se asume distribución uniforme.	Índice Primario sobre idLugar (cant niveles = 1) Índice Secundario sobre pais

Considere la siguiente consulta SQL sobre esta base de datos:

```
SELECT T.titulo
FROM Temas T, TemasComp TC, Compositores C, Lugares L
WHERE T.tema = TC.tema AND TC.comp = C.comp AND
      C.idLugar = L.idLugar AND L.pais = 'Francia'
```

Se pide:

1. Partiendo del árbol canónico de la consulta, aplicar las heurísticas para obtener el plan lógico optimizado. Mostrar los pasos seguidos.
2. Dar un plan físico correspondiente al plan lógico anterior, donde se utilicen índices cuando sea posible.
3. Si existiera un índice secundario (árbol B+) en Compositores por el atributo idLugar, ¿se mejoraría el costo de la consulta? Justificar. (Sugerencia: Calcular solamente los costos de la parte del plan que se ve afectada)

Implementaciones de los Operadores.

	Algoritmo	Costo	Condición	Organización
$\sigma_c(R)$	Búsqueda Lineal	b_R (peor caso) $b_R/2$ (prom)	Todas	--
	Búsqueda Binaria	$\log_2 b_R + \lceil s/bf_R \rceil - 1$	Todas	Ordenado
	Índice Primario	$x + 1$	Igualdad	Ordenado
	Hash	1 o 2	Igualdad	--
	Índice Primario	$x + (b/2)$ (prom)	de orden	Ordenado
	Índice Cluster	$x + \lceil s/bf_R \rceil$	Todas	Ordenado
	Índice secundario B+	$x + s$ (peor caso)	Todas	--
	Grabación Intermedia	s/bf_R	Todas	--

	Algoritmo	Costo	Condición	Organización
$R \gg _c S$	Loop Anidado (registros)	$b_R + (n_R * b_s)$	Todas	--
	Loop Anidado (bloque)	$b_R + \lceil b_R / (M-2) \rceil * b_s$	Todas	--
	Sort Merge	$b_R + b_s +$ costo ordenación	Todas	Índice
	Index join	$b_R + n_R * Z$	Todas	Índice

Z depende del tipo de índice:

secundario: $Z = x + sS$

cluster: $Z = x + \lceil sS/bf_S \rceil$

primario: $Z = x + 1$