

FUNDAMENTOS DE BASES DE DATOS

Examen Febrero 2010

La duración del examen es de 3 horas y ½.

Presentar la resolución del examen:

- Con las hojas numeradas y escritas de un solo lado.
- 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.
- Comenzando cada ejercicio en una nueva hoja

Ejercicio 1 (25 puntos).

Se desea modelar la realidad correspondiente a una productora de discos y conciertos musicales. La productora cuenta con un conjunto de empresarios de los cuales se conoce su CI, nombre y teléfonos. Los empresarios pueden ser representantes de artistas o productores. De los productores interesa registrar si ha realizado trabajos en el exterior, mientras que de los representantes se registra el año en el que comenzó a trabajar como representante.

Los artistas tienen asignado un código, este código o el nombre de los artistas los identifica. Además, se les conoce el tipo (si es solista melódico, dúo folklórico, banda de rock, etc.), y el nombre de cada uno de los premios que ha ganado. Todos los artistas tienen un representante y se registra el porcentaje que obtiene el representante. Se sabe que los representantes no les cobran el mismo porcentaje a todos los artistas.

Todos los artistas graban al menos un disco. Todos los discos tienen un número que los identifica e interesa saber su año de lanzamiento y el título del disco. Un disco puede contener canciones de más de un artista e interesa registrar qué canciones graba cada artista en un disco. De cada canción se conoce su título y duración.

Los discos son producidos por un único productor. Para cada disco, cada artista realiza un ensayo con el productor de ese disco. De cada ensayo se registra la fecha y la dirección donde se realizó el mismo.

Se sabe que hay artistas que no realizan conciertos nunca. Los conciertos son identificados por la fecha que fueron realizados y se registra el país, la ciudad, el nombre del establecimiento donde se realizó el mismo y la cantidad de entradas vendidas. En un concierto pueden participar varios artistas.

Para la organización de cada concierto se registra el número de cada entrada y el número de asiento correspondiente a esa entrada. Los números de las entradas pueden repetirse en conciertos diferentes.

En un concierto puede haber diferentes sponsors para cada uno de los artistas que participa en el concierto. A su vez, un sponsor puede promover a varios artistas en conciertos.

De los sponsors se conoce la marca presentada, el nombre de la compañía que la representa (ambos datos identifican al sponsor) y el nombre de un responsable. Además, se registra la cantidad de dinero con la que el sponsor apoyó al artista en el concierto que participó.

Se pide: Modelo Entidad Relación completo.

Ejercicio 2 (25 puntos).

Considere la siguiente relación con el siguiente conjunto de dependencias funcionales:

$$R(A,B,C,D,E,G,H) F=\{BD\rightarrow A, B\rightarrow H, DH\rightarrow E, GC\rightarrow D, D\rightarrow G\}$$

- 1) Indicar cuáles de las siguientes dependencias están en F^+ y cuáles no. Justifique.
 - a. $AB\rightarrow G$
 - b. $CBG\rightarrow E$
- 2) Hallar todas las claves de R . Justificar.
- 3) Llevar a 3NF con join sin pérdida y preservación de dependencias con respecto a F , aplicando los algoritmos vistos en el curso.
- 4) Llevar el resultado en 3NF que obtuvo en la parte 3 a BCNF, aplicando los algoritmos vistos en el curso. Discuta si se pierden dependencias funcionales o no.
- 5) Dada la siguiente descomposición de R : $\rho = \{ R_1(A,B,D,G), R_2(A,C,D,H,E) \}$, decir si ρ tiene join sin pérdida con respecto a F , justificando.

Ejercicio 3 (25 puntos).

La administración de un jardín de infantes mantiene toda la información en una base de datos con el siguiente esquema:

DOCENTES (Ci, Nombre, Teléfono, añoIngreso)

Esta tabla contiene información acerca de los docentes que trabajan en el jardín de infantes. Registra la cédula de identidad del docente, su nombre, su teléfono y el año de ingreso al trabajo.

ACTIVIDADES (IdAct, Nombre, Duración)

Esta tabla contiene información acerca de las actividades que se realizan en el jardín. Registra el identificador de la actividad, el nombre (gimnasia, natación, artes plásticas, música, etc.) y la duración (en minutos) de la misma en cada clase.

ALUMNOS (Ci, Nombre, Teléfono, Nivel)

Esta tabla contiene información acerca de los alumnos del jardín de infantes. Registra la cédula de identidad del alumno, su nombre, su teléfono y el nivel (2 años, 3 años, etc.), al cual asiste el niño.

ASIGNACIÓN_ACT (IdDoc, IdAct, DiaSemana, Hora)

Esta tabla contiene la asignación de los docentes a las actividades dentro del jardín de infantes. Cada docente puede ser asignado a varias actividades y además puede ser asignado a la misma actividad en varios horarios diferentes en el mismo día o en distintos días.

ASISTENCIA_ACT (IdAlum, IdAct, IdDoc)

Esta tabla contiene información acerca de los alumnos que asisten a cada actividad con el docente correspondiente. Un alumno asiste a una actividad con un único docente.

En este esquema no existen tablas vacías y se cumplen las siguientes dependencias de inclusión.

- a. $\Pi_{IdDoc}(ASIGNACIÓN_ACT) \subseteq \Pi_{Ci}(DOCENTES)$
- b. $\Pi_{IdAct}(ASIGNACIÓN_ACT) \subseteq \Pi_{IdAct}(ACTIVIDADES)$
- c. $\Pi_{IdDoc, IdAct}(ASISTENCIA_ACT) \subseteq \Pi_{IdDoc, IdAct}(ASIGNACION_ACT)$
- d. $\Pi_{IdAlum}(ASISTENCIA_ACT) \subseteq \Pi_{Ci}(ALUMNOS)$

Resolver en álgebra relacional las siguientes consultas:

- 1) Devolver el nombre de los docentes que hayan ingresado antes del año 2000 y su asignación es solamente en el horario de la mañana (antes de las 12:00 hs.).
- 2) Devolver el nombre y teléfono de todos los alumnos del nivel 3 que asisten a actividades con duración mínima.

Resolver en cálculo relacional las siguientes consultas:

- 3) Devolver el nombre y teléfono de los alumnos que asistieron a todas las actividades que tienen una duración mayor de 30 minutos.
- 4) Obtener los nombres de los docentes que están asignados solamente a una actividad.

Resolver en SQL, sin utilizar vistas ni sub-consultas en el FROM, las siguientes consultas:

- 5) Obtener el nombre y el año de ingreso de los docentes que realizan más de 4 actividades distintas a las cuales asisten niños del nivel 5.
- 6) Dar los nombres de los docentes que dictan alguna de las actividades de menor duración los martes o los jueves.

Ejercicio 4 (25 puntos).

Considere el siguiente esquema:

ACTIVIDADES (IdAct, Nombre, Duracion)
ASIGNACION_ACT (IdDoc, IdAct, DiaSemana, Hora)

A continuación se presenta una consulta sobre dicho esquema:

```
SELECT S.IdDoc
FROM ACTIVIDADES A, ASIGNACION_ACT S
WHERE S.IdAct = A.IdAct
AND A.Nombre = 'GIMNASIA'
```

Con el siguiente catálogo:

Tabla	Columna	Valores distintos
ACTIVIDADES	IdAct	50
ACTIVIDADES	Nombre	50
ASIGNACION_ACT	IdAct	50
ASIGNACION_ACT	IdDoc	100

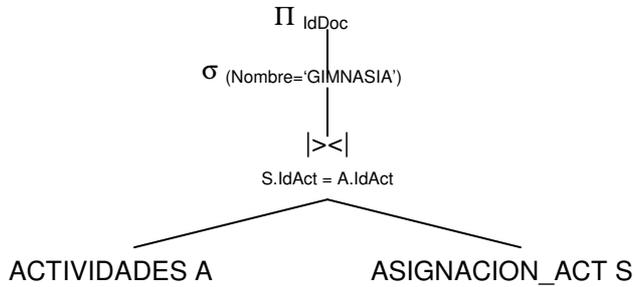
Tabla	Cant. tuplas	Factor de bloqueo
ACTIVIDADES	50	4
ASIGNACION_ACT	400	10
ACTIVIDADES >< ASIGNACION_ACT		2

Indice	Tabla/Atributo	Tipo	Cant. Niveles
Act_Nombre	ACTIVIDADES/Nombre	Secundario	2

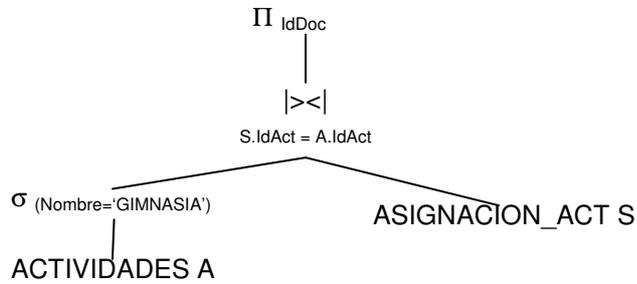
Se pide:

a) Dados los siguientes planes lógicos, indique cuál debería tener mejor performance y justifique su respuesta.

1)



2)



b) Construya el mejor plan físico que pueda para cada uno de los planes anteriores. Calcule los costos. Justifique sus respuestas.

NOTA: Asuma que tiene disponibles 3 buffers para el intercambio de datos entre disco y memoria.

Implementaciones de los Operadores.

Oper.	Algoritmo	Costo	Condición	Organización
$\sigma_c(R)$	Búsqueda Lineal	b_R peor caso, $b_R/2$ promedio	Cualquier Caso	Cualquiera
	Búsqueda Binaria	$\log_2 b_R + \lceil s/bf_R \rceil - 1$	Cualquier caso	Registros ordenados
	Índice Primario	$x + 1$	Por igualdad a un valor	Registros Ordenados
	Hash	1 o 2 según el tipo	Por igualdad a un valor	Cualquiera
	Índices Primario	$x + (b/2)$ (promedio)	Por relación de orden.	Índice ordenado
	Índice Cluster	$x + \lceil s/bf_R \rceil$	Cualquier Caso	Registros Ordenados

CSI-INCO

Fundamentos de Bases de Datos

45

Implementaciones de los Operadores.

Oper.	Algoritmo	Costo	Cond.	Organización
$\sigma_c(R)$	Índice secundario B+	$x + s$ peor caso	Cualquier Caso	Cualquiera
	Grabación Intermedia	s/bf_R	Cualquier caso	Cualquiera
$R \bowtie X \mid_c S$	Loop Anidado (registros)	$b_R + (n_R * b_s)$	Cualquier caso	Cualquiera
	Loop Anidado (bloque)	$b_R + \lceil b_R / (M - 2) \rceil * b_s$	Cualquier caso	Cualquiera
	Sort Merge	$b_R + b_s + \text{costo ords.}$	Cualquier caso	índice en disco
	Index join	$b_R + n_R * Z$	Cualquier caso	índice en disco

CSI-INCO

Fundamentos de Bases de Datos

46

donde Z depende del tipo de índice:

secundario: $Z = x + sS$, cluster: $Z = x + \lceil sS/bfS \rceil$, primario: $Z = x + 1$, hash = h