

FUNDAMENTOS DE BASES DE DATOS

SOLUCIÓN

Examen Febrero 2009

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 registrar en una base de datos la información de películas de cine.

De cada película se conoce su título y además un identificador único de película. Otros datos de la película a registrar son su sitio web y la duración en minutos de la misma.

Existen actores y directores (un actor puede ser también director) los cuales se identifican por un documento, el cual consiste de tipo y número. De ellos se desea registrar además su nombre y sexo.

En una película trabajan actores (al menos uno), y ésta es dirigida por un solo director. Interesa registrar cuantos minutos duró la participación del actor en cada película.

De cada actor o director se desea registrar de que origen es, o sea en que país nació. A su vez de cada película se desea registrar en que países se filmó. Para cada país en que se filmó, interesa registrar los nombres de las locaciones en las cuales se filmó en ese país. No existen dos países con el mismo nombre.

Pueden existir varias versiones de una película. Cada una de éstas es la misma película, pero doblada en un idioma particular. Interesa saber si una versión es en el idioma original o no. Cada versión es exhibida en uno o varios países.

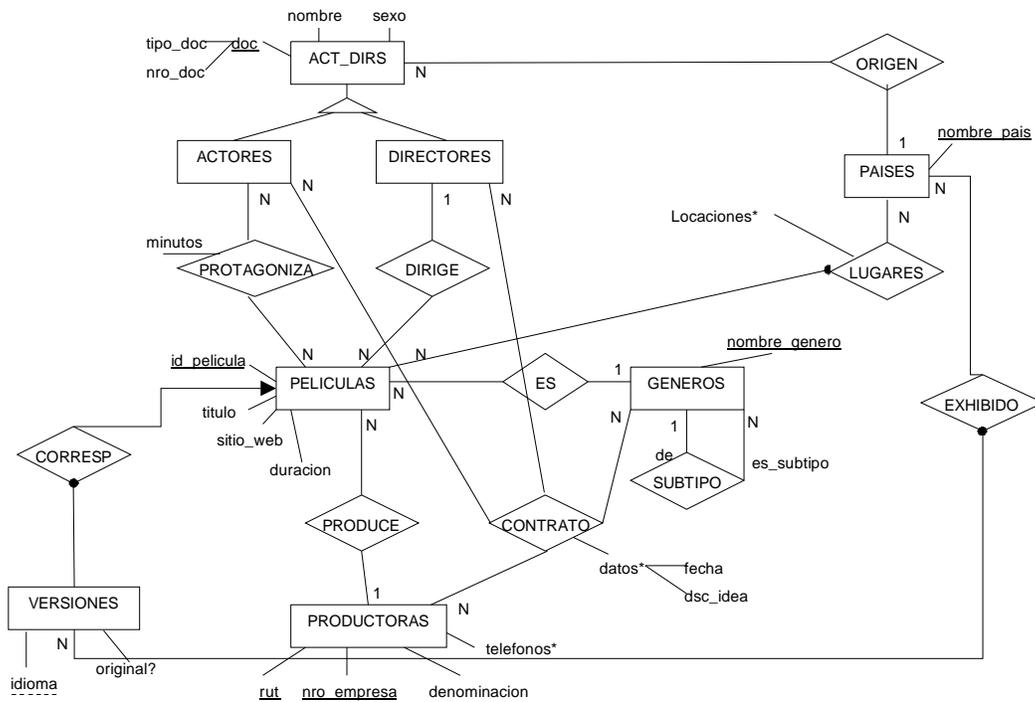
Las películas se clasifican en géneros. Cada película está asociada a un solo género. Existen géneros mas generales y otros más específicos. Un género más específico es subtipo de otro más general. De los géneros solo se registra su nombre, no existiendo dos géneros con el mismo nombre.

Se registran las productoras, las cuales se identifican por su número de empresa o su rut indistintamente. Se conoce su denominación y los teléfonos de contacto. Las productoras producen películas, una película es producida por una única productora.

Habitualmente una productora, un actor y un director registran un contrato para trabajar sobre un género en particular. Los datos que se registran de este contrato son la fecha en que se comenzó a trabajar y la descripción de la idea, que más tarde se podrá materializar en una película o no. Una productora, un actor y un director pueden registrar varios contratos con el paso del tiempo sobre un mismo género. Nunca se producen películas que no sean de un género sobre el que la productora hizo un contrato.

Se pide: Modelo Entidad Relación completo.

Solución:



Restricciones no estructurales:

- No existe un genero que sea subtipo de si mismo ni directa ni indirectamente.
- Si un genero es subtipo de otro, este no puede ser subtipo del anterior.
- ACTORES U DIRECTORES = ACT_DIRS
- Si una película es producida por una productora, su género debe existir en algún contrato de la productora.

Ejercicio 2 (10 puntos).

Considere el siguiente conjunto de dependencias y el siguiente esquema de relación $R(A,B,C,D,E)$:

$$F = \{ A \twoheadrightarrow BE, C \twoheadrightarrow EB \}$$

Se pide:

- a) Deduzca (si es posible) alguna dependencia multivaluada no trivial que tenga sólo el atributo A del lado izquierdo. Justifique su respuesta.

● **Solución.**

Dado que las dependencias funcionales son también multivaluadas, entonces $A \twoheadrightarrow B$ y $A \twoheadrightarrow E$ son dependencias multivaluadas que se cumplen en ese contexto.

- b) Deduzca (si es posible) alguna dependencia funcional no trivial que tenga sólo el atributo C del lado izquierdo. Justifique su respuesta.

● **Solución.**

Dada la dependencia multivaluada $C \twoheadrightarrow EB$, la dependencia funcional $A \twoheadrightarrow BE$ y considerando la regla de coalescencia, se cumple la dependencia funcional $C \twoheadrightarrow EB$.

Ejercicio 3 (15 puntos).

PARTE 1

Dado el esquema relacional $R(A,B,C,D,E,G,H)$ y los siguientes conjuntos de dependencias funcionales:

$$F = \{ AB \rightarrow E, AB \rightarrow G, C \rightarrow D, C \rightarrow H, HC \rightarrow A, D \rightarrow A, G \rightarrow E \}$$

$$J = \{ CH \rightarrow A, A \rightarrow E, A \rightarrow D, D \rightarrow G, B \rightarrow C, C \rightarrow B, C \rightarrow G, G \rightarrow H, H \rightarrow C \}$$

- a) Determinar si F y J son equivalentes. Justificar la respuesta.

Solución.

F y J son equivalentes sii $F^+ = G^+$, o sea $F^+ \subseteq J^+$ y $J^+ \subseteq F^+$.

Observando las dependencias funcionales vemos que C no aparece en el lado derecho de ninguna dependencia funcional de F pero si aparece a la derecha en J, en la dependencia funcional $B \rightarrow C$.

Sabemos que para que F cubra a J ($J^+ \subseteq F^+$), se debe cumplir:

$$\forall X \rightarrow Y \in J \Rightarrow X \rightarrow Y \in F^+$$

o lo que es lo mismo,

$$\forall X \rightarrow Y \in J \Rightarrow Y \subseteq (X)^+_F$$

Tomando $B \rightarrow C \in J$ vemos que $C \notin (B)^+_F$ pues $(B)^+_F = \{B\}$. Luego, F y J no son equivalentes.

- b) Determinar si F y J son minimales. Justificar la respuesta.

Solución.

Para que un conjunto de dependencias sea minimal debe cumplir tres condiciones:

1. los lados derechos de las dependencias tienen un único atributo
2. los lados izquierdos de las dependencias no tienen atributos redundantes
3. no hay dependencias redundantes en el conjunto.

Observando las dependencias funcionales en F vemos que se cumplen $C \rightarrow H$ y $HC \rightarrow A$. Luego, H es redundante en $HC \rightarrow A$ pues $A \in (C)^+_F = \{C, D, H, A\}$. Por tanto, F no es minimal.

Siguiendo el mismo razonamiento, vemos que J tampoco es minimal. Considerando las dependencias funcionales $H \rightarrow C$ y $CH \rightarrow A$, vemos que C es redundante en $CH \rightarrow A$ pues $A \in (H)^+_J = \{H, C, B, G, A, E, D\}$. Luego, J no es minimal.

- c) Indicar cuál es la máxima forma normal en que se encuentra R según F. Justificar su respuesta.

Solución.

Cálculo de claves de R según F.

B y C no pertenecen a los lados derechos de las dependencias funcionales por lo tanto pertenecen a todas las claves.

$$(BC)^+_F = \{B, C, D, H, A, E, G\}$$

Por lo tanto BC es la única clave de R según F.

Sea $C \rightarrow D$, C es parte de una clave y D no es primo (no pertenece a una clave) por lo tanto $BC \rightarrow D$ es una dependencia parcial de una clave a un atributo no primo y viola 2NF.

Por lo tanto **la máxima forma normal en que se encuentra R según F es 1NF.**

- d) Indicar cuál es la máxima forma normal en que se encuentra R según J. Justificar su respuesta.

Solución.

Cálculo de claves de R según J.

No hay atributos que no pertenezcan a los lados derechos de las dependencias funcionales por lo tanto pertenezcan a todas las claves.

Se consideran los atributos de a uno y se calcula su clausura con respecto a J:

$$A^+ = \{A, E, D, G, H, C, B\}$$

$$B^+ = \{B, C, G, H, A, E, D\}$$

$$C^+ = \{C, B, G, H, A, E, D\}$$

$$D^+ = \{D, G, H, C, A, E, D\}$$

$$E^+ = \{E\}$$

$$G^+ = \{G, H, C, B, A, E, D\}$$

$$H^+ = \{H, C, B, G, A, E, D\}$$

Por lo tanto todos los atributos menos E son clave de R según J y son las únicas claves ya que no es posible armar un subconjuntos de atributos con dos o más atributos que no contenga una clave.

En los lados izquierdos de las dependencias de J no figura el atributo E por lo tanto los lados izquierdos son todos claves o superclaves \Rightarrow R se encuentra en BCNF y como las únicas dependencias multivaluadas son las funcionales \Rightarrow R se encuentra en 4NF.

Por lo tanto **la máxima forma normal en que se encuentra R según J es 4NF.**

PARTE 2

Dado un esquema relacional R y dos conjuntos de dependencias funcionales F y J sobre los atributos de R, determinar para cada una de las siguientes afirmaciones, si es verdadera o falsa. Justifique su respuesta.

- a) Si F y J no son equivalentes entonces la máxima forma normal en que se encuentra R según F es distinta a la máxima forma normal en que se encuentra R según J.

Solución.

FALSO

Sea $R(A,B,C)$, $F = \emptyset$ y $J = \{A \rightarrow BC\}$

F y J no son equivalentes ($A \rightarrow BC \notin F^+$)

La máxima forma normal en que se encuentra R según F es 4NF

La máxima forma normal en que se encuentra R según J es 4NF ya que A es clave y la única dependencia multivaluada es la funcional.

Por lo tanto la máxima forma normal coincide para los dos conjuntos de dependencias.

- b) Si $F^+ \subseteq J^+$ y $X \subseteq R$ es superclave de R según F entonces X es superclave de R según J .

Solución.

VERDADERO

X es superclave de R según $F \Leftrightarrow X \rightarrow R \in F^+$. (i)

De $F^+ \subseteq J^+$ y (i) $\Rightarrow X \rightarrow R \in J^+ \Leftrightarrow X$ es superclave de R según J .

Ejercicio 4 (30 puntos).

La oficina de trabajo de una facultad mantiene la información de su sistema en una base de datos con el siguiente esquema:

LLAMADOS (nro, descrip, fchAp, fchLim, nombreInst, figura)

Contiene la información de los llamados. Su número, descripción, fecha de aparición, la fecha límite de presentación al mismo, nombre de la institución que realiza el llamado y si la misma desea figurar en el aviso o no (figura = Si o No).

Se sabe que no existen dos llamados distintos para una misma institución con la misma fecha de aparición.

ESTUDIANTES (ci, nombreEst, fchNac, dir, tel, mail, carrera, cv)

Contiene la información de los estudiantes. Su cédula de identidad, nombre, fecha de nacimiento, dirección, teléfono, dirección de correo, carrera a la cual está inscripto y su curriculum vitae.

INSCRIPTOS (ci, nro, fecha)

Contiene la información de los estudiantes que se inscriben a los llamados. Se registra la cédula de identidad del estudiante, el número de llamado y la fecha de inscripción.

LLAMADOS_DESIERTOS (nro, motivo)

Contiene la información de los llamados que quedaron desiertos, o sea para los cuales ningún estudiante se inscribió en él o la empresa decidió no contratar a nadie. Los datos que se registran son el número de llamado y el motivo por el cual quedó desierto.

CONTRATOS (ci, nro)

Contiene la información de los contratos realizados a partir de los llamados. Los datos que se registran son número del llamado y cédula de los estudiantes contratados.

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

- $\Pi_{nro}(\text{LLAMADOS_DESIERTOS}) \subseteq \Pi_{nro}(\text{LLAMADOS})$
- $\Pi_{nro}(\text{CONTRATOS}) \subseteq \Pi_{nro}(\text{LLAMADOS})$
- $\Pi_{ci,nro}(\text{CONTRATOS}) \subseteq \Pi_{ci,nro}(\text{INSCRIPTOS})$
- $\Pi_{ci}(\text{INSCRIPTO}) \subseteq \Pi_{ci}(\text{ESTUDIANTES})$
- $\Pi_{nro}(\text{INSCRIPTO}) \subseteq \Pi_{nro}(\text{LLAMADOS})$

Resolver las siguientes consultas en Álgebra Relacional:

1. Cédula de identidad, nombre y correo electrónico de los estudiantes que se inscribieron a todos los llamados de la institución "Facultad de Ingeniería" pero que no están contratados por ningún llamado.

$$\begin{aligned}
 A &= \Pi_{\text{nro}} (\sigma_{\text{nombreInst}='Facultad de Ingeniería'} \text{ LLAMADOS}) \\
 B &= \Pi_{\text{ci, nro}} (\text{INSCRIPTO}) \% A \\
 \text{SOL} &= \Pi_{\text{ci, nombreEst, mail}} ((B - \Pi_{\text{ci}} (\text{CONTRATOS})) * \text{ESTUDIANTES})
 \end{aligned}$$

2. Número y descripción de los llamados con fecha límite de presentación posterior al 13/02/09 para los cuales sólo se inscribieron estudiantes de la carrera "Computación".

$$\begin{aligned}
 A &= \Pi_{\text{nro}} (\sigma_{\text{fchLim} > 13/02/09} \text{ LLAMADOS}) \\
 B &= \Pi_{\text{ci}} (\sigma_{\text{carrera} = \text{Computación}} \text{ ESTUDIANTES}) \\
 C &= \Pi_{\text{nro}} (\text{INSCRIPTO} * A * B) \\
 D &= \Pi_{\text{nro}} (\text{INSCRIPTO}) * (\sigma_{\text{carrera} \neq \text{Computación}} \text{ ESTUDIANTES}) \\
 \text{SOL} &= \Pi_{\text{nro, descrip}} ((C - D) * \text{LLAMADOS})
 \end{aligned}$$

Resolver las siguientes consultas en Cálculo Relacional:

3. Nombre de las instituciones que han realizado llamados con fecha de aparición anterior al 01/01/09 y ninguno de éstos figura como desierto.

$$\{ \text{l.nombreInst} / \text{LLAMADOS}(l) \wedge \text{l.fchAp} < 1/1/09 \wedge \\
 \neg ((\exists a)(\text{LLAMADOS}(a) \wedge \text{a.fchAp} < 1/1/09 \wedge \text{a.nombreInst} = \text{l.nombreInst} \wedge \\
 (\exists d)(\text{LLAMADOS_DESIERTOS}(d) \wedge \text{d.nro} = \text{a.nro})) \\
) \\
 \}$$

4. Nombre de los estudiantes que: o se inscribieron a todos los llamados de la institución "Ute"; o se inscribieron a todos los llamados que quedaron desierto pero no se inscribieron a ningún llamado donde la institución no figure en el aviso del llamado.

$$\{ \text{e.nombreEst} / \text{ESTUDIANTES}(e) \wedge \\
 ((\forall d)(\text{LLAMADOS_DESIERTOS}(d) \rightarrow \\
 (\exists i)(\text{INSCRIPTO}(i) \wedge \text{i.ci} = \text{e.ci} \wedge \text{i.nro} = \text{d.nro})) \\
) \wedge \\
 \neg ((\exists i)(\text{INSCRIPTO}(i) \wedge \text{i.ci} = \text{e.ci} \wedge \\
 (\exists l)(\text{LLAMADO}(l) \wedge \text{l.figura} = \text{NO} \wedge \text{i.nro} = \text{l.nro})) \\
) \\
) \\
 \vee \\
 ((\forall l)(\text{LLAMADO}(l) \wedge \text{l.nombreInst} = \text{Ute} \rightarrow \\
 (\exists i)(\text{INSCRIPTO}(i) \wedge \text{i.ci} = \text{e.ci} \wedge \text{i.nro} = \text{l.nro})) \\
) \\
) \\
) \\
 \}$$

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

5. Nombre de los estudiantes que se inscribieron a todos los llamados de la institución de nombre 'antel'.

```
SELECT e.nombreEst
FROM ESTUDIANTES e
WHERE NOT EXISTS
  (SELECT *
   FROM LLAMADOS I
    WHERE I.nombreInst = 'Antel'
     AND NOT EXISTS
      (SELECT *
       FROM INSCRIPTO i
        WHERE i.ci = e.ci
         AND i.nro = I.nro
      )
  )
```

6. Número de los llamados que tuvieron la misma cantidad de inscriptos que el último llamado publicado (fecha de aparición) por la institución de nombre 'Ute'.

```
SELECT I.nro
FROM INSCRIPTO I
GROUP BY I.nro
HAVING count(*) =
  (SELECT count(*)
   FROM INSCRIPTO I2
    WHERE I2.nro =
      (SELECT L.nro
       FROM LLAMADOS L
        WHERE L.nombreInst = UTE
         AND L.fchAp =
           (SELECT max(LL.fchAp)
            FROM LLAMADOS LL
             WHERE LL.nombreInst = Ute
           )
      )
  )
```

Ejercicio 5 (20 puntos).

El siguiente es parte del esquema de un sistema de planificación de la producción (M-R-P).

Maquinas (tipoMaquina, capacidad, cantidad)

Registra los tipos de máquina existentes en la empresa, y para cada tipo, la capacidad que tienen las máquinas, y la cantidad de máquinas disponibles.

RecursosHumanos (ci, edad, profesion, nombre)

Registra los recursos humanos que trabajan en la empresa, cédula de identidad, edad, profesión y nombre de cada individuo.

Produccion (producto, profesion, tipoMaquina, horas)

Registra para cada producto cuantas horas son necesarias de una persona de determinada profesión en determinado tipo de máquina para producirlo.

En este esquema se cumplen las siguientes restricciones de inclusión:

- $\Pi_{\text{profesion}}(\text{Produccion}) \subseteq \Pi_{\text{profesion}}(\text{RecursosHumanos})$
- $\Pi_{\text{tipoMaquina}}(\text{Produccion}) \subseteq \Pi_{\text{tipoMaquina}}(\text{Maquinas})$

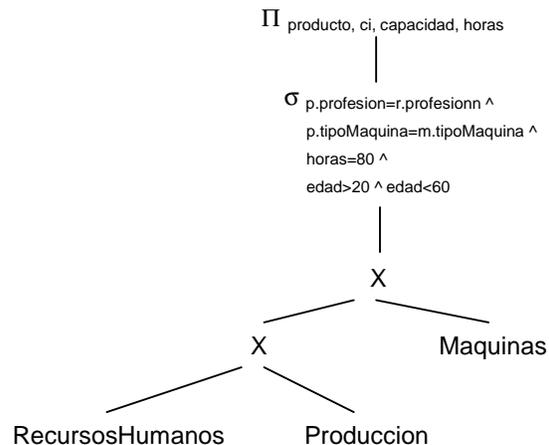
Sea la siguiente consulta SQL:

```
SELECT producto, ci, capacidad, horas
FROM RecursosHumanos r, Produccion p, Maquinas m
WHERE p.profesion = r.profesion
AND p.tipoMaquina = m.tipoMaquina
AND p.horas = 80
AND r.edad > 20 AND r.edad < 60
```

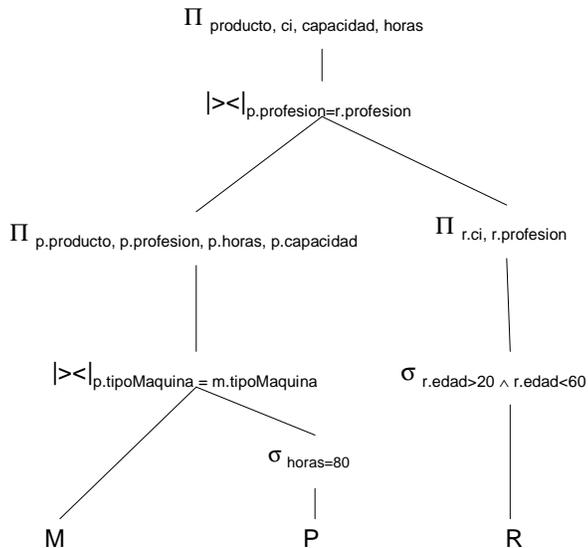
Se pide:

- a) Dibuje el árbol canónico para la consulta SQL anterior.

Solución.



b) Luego de aplicar heurísticas de optimización vistas en el curso, se obtuvo el siguiente árbol:



Enumere los pasos aplicados para obtenerlo.

Solución.

Las heurísticas aplicadas son las siguientes:

- Descomponer selección en selecciones en cascada.
- Desplazar selecciones hacia abajo en el árbol lo más posible.
- Posicionar las hojas de las tablas con menos tuplas (o que tengan una selección que produzca menos tuplas) más a la izquierda en el árbol, cuidando que no se genere un producto cartesiano.
- Combinar los productos cartesianos con las selecciones correspondientes, generando joins.
- No se aplicó la heurística de bajar lo más posible las proyecciones, dado que ser perderían índices.

c) Dado el árbol optimizado de la parte b), qué índices sugeriría y por qué? Asuma que existe un índice primario por la clave primaria de cada tabla.

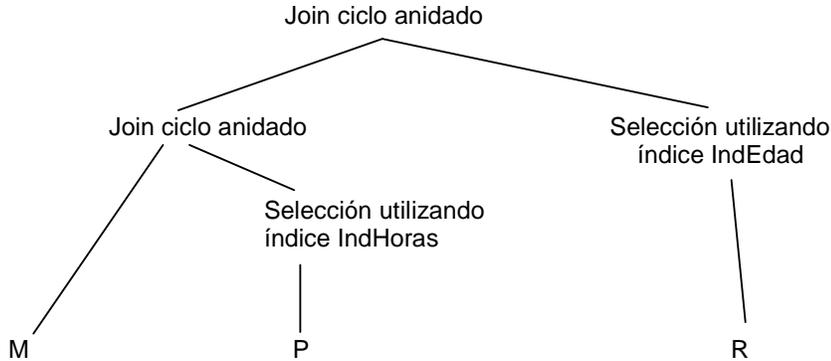
Solución.

Como los únicos índices que pueden ser utilizados en ese árbol son índices útiles para la selección, se sugieren:

- IndHoras – índice secundario en tabla *Produccion*, por atributo *horas*.
- IndEdad – índice secundario en tabla *RecursosHumanos*, por atributo *edad*.

d) Dar un plan físico para el árbol de la parte b) que considere los índices propuestos en la parte c).

Solución.



e) Calcular los costos de las operaciones de selección y de la operación de join entre *Maquinas* y *Producción*, considerando las grabaciones de los resultados de cada una.

Nota: - Para los índices asuma cantidad de niveles = 3.
- Suponga que se cuenta con 3 buffers en memoria.

Datos sobre las tablas:

	Cantidad tuplas	Factor de bloqueo	Observaciones
Maquinas	40	10	
RecursosHumanos	220	8	El 85% de los empleados tienen entre 20 y 60 años.
Produccion	2100	8	Existen 30 valores diferentes en el atributo <i>horas</i> , los cuales se distribuyen uniformemente.
Maquinas >< Produccion		5	

Solución.

Notación: - $|T_i|$: Tamaño de T_i
- $C(T_i)$: Costo para obtener T_i (incluyendo grabación)

$$|\sigma_{\text{horas}>80}(P)| = 2100 / 30 = 70$$

$$|\sigma_{r.\text{edad}>20 \wedge r.\text{edad}<60}(R)| = 220 * 0.85 = 187$$

$$|M |><| \sigma_{\text{horas}>80}(P)| = |\sigma_{\text{horas}>80}(P)| = 70$$

$$C(\sigma_{\text{horas}>80}(P)) = C(\text{selección}) + C(\text{grabación}) = x + s + b_S = 3 + 70 + \lceil 70/8 \rceil = 82$$

$$b_S = \lceil |\sigma_{\text{horas}>80}(P)| / fb \rceil$$

$$C(\sigma_{r.\text{edad}>20 \wedge r.\text{edad}<60}(R)) = C(\text{selección}) + C(\text{grabación}) = x + s + b_S = 3 + 187 + \lceil 187/8 \rceil = 214$$

$$b_S = \lceil |\sigma_{r.\text{edad}>20 \wedge r.\text{edad}<60}(R)| / fb \rceil$$

$$C(M |><| \sigma_{\text{horas}>80}(P)) = C(\text{join}) + C(\text{grabación}) =$$

$$= b_R + \lceil b_R / (\text{buffers} - 2) \rceil * b_S + \lceil |M |><| \sigma_{\text{horas}>80}(P)| / fb \rceil = 40/10 + \lceil 4/1 \rceil * \lceil 70/8 \rceil + \lceil 70/5 \rceil = 54$$

$$b_R = \lceil |M| / fb \rceil$$

$$b_S = \lceil |\sigma_{\text{horas}>80}(P)| / fb \rceil$$

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
	Indice Primario	$x + 1$	Por igualdad a un valor	Registros Ordenados
	Hash	1 o 2 según el tipo	Por igualdad a un valor	Cualquiera
	Indices Primario	$x + (b/2)$ (promedio)	Por relación de orden.	Indice ordenado
	Indice 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
$\sigma_c(R)$	Indice secundario B+	$x + s$ peor caso	Cualquier Caso	Cualquiera
	Grabacion Intermedia	s/bf_R	Cualquier caso	Cualquiera
$R \bowtie X \bowtie S$	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	indice en disco
	Index join	$b_R + n_R * Z$	Cualquier caso	indice 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