

FUNDAMENTOS DE BASES DE DATOS

Solución Examen Julio 2006

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.

Parte 1. Modelo Entidad-Relación (25 puntos)

Ejercicio 1. (25 pts)

Una droguería desea mantener una base de datos con la información relativa a su negocio. Básicamente la empresa se dedica a la intermediación en las compras de fármacos entre farmacias y los laboratorios. Recibe las solicitudes de productos farmacéuticos realizados por las farmacias afiliadas, los agrupa convenientemente en pedidos para los laboratorios y realiza una gran compra a cada uno de los laboratorios correspondientes, obteniendo de esta forma un descuento por cantidad y/o pago en efectivo. Posteriormente, una vez recibidos los productos de los laboratorios, se encarga de la distribución y cobro a los solicitantes.

De las farmacias afiliadas se conoce su nombre, RUC, dirección de la empresa, dirección de entrega y calificación. La calificación es un número del 1 al 10 que indica la calidad del cliente.

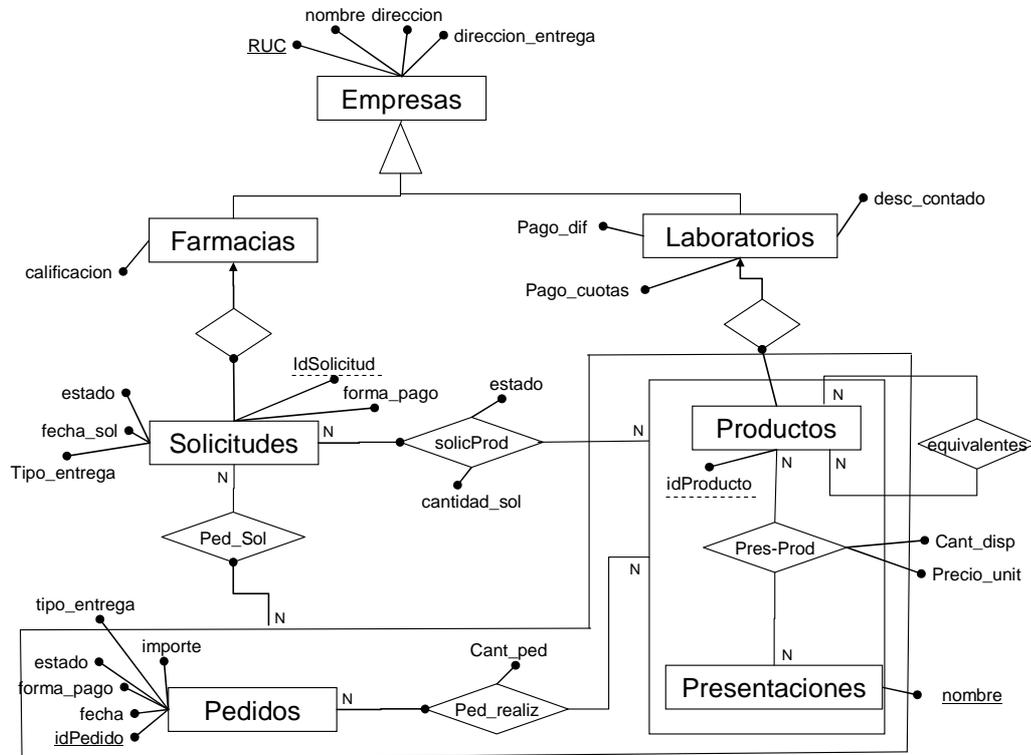
De los laboratorios se conoce el nombre, RUC, dirección de la empresa, dirección de entrega, así como también el descuento obtenido por pago contado y si acepta pago diferido y/o en cuotas. De los productos de los laboratorios, se mantienen los siguientes datos: un identificador según el laboratorio, la presentación de cada producto (comprimidos 20mg, Comprimidos 40 mg, solución, etc.), la cantidad disponible y el precio unitario (estos dos últimos para cada presentación). También se conocen las equivalencias entre los diferentes productos.

Las solicitudes tienen un identificador que depende de la farmacia, es decir, que identifican a la solicitud dentro de una misma farmacia, y además incluyen: el tipo de entrega (semanal, mensual, inmediata), la fecha de solicitud, la forma de pago (contado, pago diferido, cuotas), y el estado de la misma. Cada ítem de la solicitud se refiere a un producto en cierta presentación, y además se incluye la cantidad solicitada. Para cada ítem de la solicitud se tiene un estado del ítem (pendiente, pedido, no disponible, etc.).

Los pedidos de presentaciones de productos que hace la droguería a los laboratorios, tienen un identificador único, se conoce el importe total del pedido y el tipo de entrega (mensual, semanal, inmediata), la fecha de realización, la forma de pago, el estado del mismo y la cantidad pedida por cada ítem. Cada ítem del pedido está asociado a una o más solicitudes de las farmacias.

SE PIDE: Esquema Entidad Relación completo, incluyendo restricciones de integridad.

Solución:



RNE:

Un producto no es equivalente a si mismo ni directa ni indirectamente

La cantidad pedida por presentación de producto debe ser menor o igual que la suma de las cantidades solicitadas de esa presentación de ese producto.

Si $\exists s \in \text{Solicitudes}, u \in \text{Ped_realiz} / \langle s, u \rangle \in \text{Ped_Sol} \wedge u = \langle p, \langle \text{pres}, \text{prod} \rangle \rangle \in \text{Ped_realiz} \Rightarrow \exists \langle s, \langle \text{pres}, \text{prod} \rangle \rangle \in \text{solicProd}$

$\text{Farmacias} \cap \text{Laboratorios} = \emptyset$

Parte 2. Consultas (25 puntos)

Ejercicio 2 (25 pts)

A continuación se presenta parte del esquema relacional correspondiente a una escuela de música. En esta escuela los profesores dictan cursos individuales, de 2 o de 3 alumnos. Los cursos son de canto o de un instrumento determinado y todos se dictan en clases de 1 hora semanal.

PROFESORES (CIProf, NomProf, DirProf, TelProf, Fechalng)

Esta tabla guarda información acerca de los profesores de la escuela.

ALUMNOS (CIAlum, NomAlum, DirAlum, TelAlum, AñoIng)

Esta tabla guarda información acerca de los alumnos de la escuela

CURSOS (IdCurso, instrumento, DuracionCurso)

Esta tabla guarda información de los cursos dictados. Los cursos de canto utilizan el instrumento "voz". La duracion del curso indica el total de horas a dictar.

CURSO_PROF (IdCurso, CIProf, NivelProf)

Esta tabla vincula a los profesores con los cursos. Para cada docente, en cada curso que dicta, se indica el nivel de ese docente en esa disciplina.

HORARIOS (IdCurso, CIAlum, CIProf, Dia, Hora, Salon)

Esta tabla guarda información acerca de los horarios asignados a los alumnos de cada curso, y que docente del curso le corresponde. Recordar que cada alumno tiene un única clase por semana por curso.

- No existen tablas vacias
- $\Pi_{IdCurso}(CURSO_PROF) \subseteq \Pi_{IdCurso}(CURSOS)$
- $\Pi_{CIProf}(CURSO_PROF) \subseteq \Pi_{CIProf}(PROFESORES)$
- $\Pi_{CIAlum}(HORARIOS) \subseteq \Pi_{CIAlum}(ALUMNOS)$
- $\Pi_{IdCurso}(HORARIOS) \subseteq \Pi_{IdCurso}(CURSOS)$
- $\Pi_{IdCurso, CIProf}(HORARIOS) \subseteq \Pi_{IdCurso, CIProf}(CURSO_PROF)$

Resolver en álgebra relacional las siguientes consultas:

- 1) Devolver la cédula de aquellos profesores que participan en cursos con mas de un alumno inscripto.

$$A = \Pi_{IdCurso, CIAlum}(HORARIOS)$$
$$\text{Resultado} = \Pi_{CIProf}(CURSO_PROF * \Pi_{\$1}(A |><| \$1=\$3 \wedge \$2<>\$4 A))$$

- 2) Devolver las cédulas de aquellos alumnos que han tomado al menos un curso de cada instrumento.

$$\text{Resultado} = \Pi_{CIAlum, instrumento}(CURSOS * HORARIOS) \% \Pi_{instrumento}(CURSOS)$$

Resolver en cálculo relacional las siguientes consultas:

- 3) Devolver el nombre de aquellos profesores que sólo dan clases los jueves

$$\{ p.NomProf / PROFESORES(p) \wedge$$
$$(\exists h)(HORARIOS(h) \wedge h.CIProf = p.CIProf \wedge h.Dia = 'jueves') \}$$

$$\neg((\exists h)(\text{HORARIOS}(h) \wedge h.\text{CIProf} = p.\text{CIProf} \wedge h.\text{Dia} \neq \text{'jueves'})$$

$$)$$

$$\}$$

- 4) Devolver el nombre de los alumnos de la escuela que nunca repitieron profesor, es decir en cada curso tuvieron un profesor diferente.

$$\{ a.\text{NomAlum} / \text{ALUMNOS}(a) \wedge$$

$$(\forall h)(\text{HORARIOS}(h) \wedge h.\text{CIAlum} = a.\text{CIAlum}) \rightarrow \neg((\exists h1)(\text{HORARIOS}(h1) \wedge$$

$$h1.\text{CIAlum} = h.\text{CIAlum} \wedge h1.\text{IdCurso} \neq h.\text{IdCurso} \wedge h1.\text{CIProf} = h.\text{CIProf}) \wedge$$

$$(\exists h2)(\text{HORARIOS}(h2) \wedge h2.\text{CIAlum} = a.\text{CIAlum})$$

$$)$$

$$\}$$

Resolver en SQL las siguientes consultas:

- 5) Devolver el identificador de curso y la cantidad de docentes asignados para cada uno de los cursos de mayor duración

```
SELECT c.IDCurso, count(*)
FROM CURSOS c, CURSO_PROF cp
WHERE c.IDCurso=cp.IDCurso AND c.duracion IN
                                (SELECT max(c1.duracion)
                                 FROM CURSOS c1)
GROUP BY IDCurso
```

- 6) Devolver id de curso, día, hora, salón de los cursos con más de 2 profesores de nivel 'alto'.

```
SELECT IdCurso, Dia, Hora, Salon
FROM HORARIOS
WHERE IdCurso IN (
    SELECT IdCurso
    FROM CURSO_PROF
    WHERE NivelProf = 'alto'
    GROUP BY IdCurso
    HAVING count(*) > 2)
```

Parte 3. Diseño Relacional (25 puntos)

Ejercicio 3. (9 pts.)

- a) Detectar los atributos y dependencias funcionales en la siguiente realidad:

“Una universidad quiere representar datos de estudiantes. Se sabe que cada estudiante está identificado por su cédula y se le conoce el nombre, la dirección y cualquier cantidad de teléfonos e E-mails.”

Sol:

Los atributos son CI, Nom, Dir, Tel e Email para Cédula del estudiante, Nombre del

estudiante, Dirección del estudiante, teléfono del estudiante e Email del estudiante respectivamente.

Dado que “Se sabe que cada estudiante está identificado por su cédula y se le conoce el nombre, la dirección y cualquier cantidad de teléfonos e E-mails”, las dependencias funcionales son:

$CI \rightarrow \text{Nom,dir}$

NO SE CUMPLEN, $CI \rightarrow \text{Tel}$ ni $CI \rightarrow \text{Email}$

b) Considere la siguiente consulta: $\Pi_{\text{Email}}(\sigma_{\text{tel}='555-55-55'}(R_k))$ donde R_k tiene sólo la cédula, los teléfonos y los emails de cada estudiante.

Que dependencia se debería cumplir sobre esa tabla para que esta consulta devolviera todos los emails asociados al teléfono 5555-55-55, sabiendo que la única relación que tienen los email con los teléfonos es que son usados por el mismo estudiante?

SOL:

Si se considera la cédula del estudiante, los teléfonos y los emails, una instancia posible sería

| CI | Tel | Email |
|-------|------------|-------|
| C_1 | 5555-55-55 | E_1 |
| C_1 | 5555-57-59 | E_2 |

Dado que la única relación que tienen los teléfonos y los emails es que son usados por el mismo estudiante, la respuesta correcta a la consulta debería ser E_1 y E_2 . Para que eso se cumpla, y garantizar el resultado correcto para esa consulta, se debe cumplir entonces que

$CI \twoheadrightarrow \text{Tel} \mid \text{Email}$

c) En qué forma normal está la relación formada por todos los atributos? Justificar.

Sol:

La única clave de la tabla es $CI, \text{Tel}, \text{Email}$, por lo que la dependencia $CI \rightarrow \text{Nom,Dir}$ es parcial, lo que viola 2NF y hace que la relación esté en 1NF.

d) Llevar a 4NF con JSP aplicando el algoritmo.

Sol:

Partimos R en $R_1(Ci, \text{Nom}, \text{Dir})$ y $R_2(Ci, \text{Tel}, \text{Email})$. Al proyectar las dependencias se puede ver que en R_1 se cumplen $\{ CI \rightarrow \text{Nom}, \text{Dir} \}$ y en R_2 $\{ CI \twoheadrightarrow \text{Tel}, CI \twoheadrightarrow \text{Email} \}$.

Por lo que R_1 está en 4NF pero R_2 no lo está, porque en R_2 , CI no es superclave.

Por esto, partimos R_2 en $R_{21}(Ci, \text{Tel})$ y $R_{22}(CI, \text{Email})$. C

El resultado final en 4nf es:

$R_1(Ci, \text{Nom}, \text{Dir}), R_{21}(Ci, \text{Tel})$ y $R_{22}(Ci, \text{Email})$.

Ejercicio 4. (16 pts).

Considere el siguiente esquema, las siguientes dependencias funcionales y la siguiente descomposición D de R:

- $R(A,B,C,D,E,G,H)$
- $F=\{CD \rightarrow E, G \rightarrow H, D \rightarrow BG, CH \rightarrow D\}$
- $D=\{R_1(B,C,D,E,G), R_2(A,C, D,H)\}$

Considere el siguiente esquema, las siguientes dependencias funcionales y la siguiente descomposición D de R:

- $R(A,B,C,D,E,G,H)$
- $F=\{CD \rightarrow E, G \rightarrow H, D \rightarrow BG, CH \rightarrow D\}$
- $D=\{R_1(B,C,D,E,G), R_2(A,C, D,H)\}$

a) Projete F sobre D.

SOL:

Observar que $G \rightarrow H$ no se proyecta sobre R_1 por lo que debemos buscar los encadenamientos que utilizan esta dependencia y que podríamos estar perdiendo. Para lo que buscamos otras dependencias que tengan H e intentamos deducir dependencias combinándolas con las que se pierden. De esta forma, al aplicar pseudotransitividad a $CH \rightarrow D$ con $G \rightarrow H$ surge $GC \rightarrow D$. Por este motivo, la proyección sobre R_1 es:

$$\Pi_{R_1}(F)=\{CD \rightarrow E, D \rightarrow BG, GC \rightarrow D\}$$

Razonando de forma análoga con R_2 obtenemos:

$$\Pi_{R_2}(F)=\{CH \rightarrow D, D \rightarrow H\}$$

b) Indicar si se pierden dependencias y cuáles. Justificar

Sol:

Consideremos la unión de las proyecciones de las dependencias:

$$F'=\Pi_{R_1}(F) \cup \Pi_{R_2}(F)=\{CD \rightarrow E, D \rightarrow BG, GC \rightarrow D, CH \rightarrow D, D \rightarrow H\}$$

Es claro que las dependencias que se proyectaron directamente no se perdieron. Por lo que debemos considerar sólo $G \rightarrow H$.

Ahora veamos cuáles de éstas dependencias se siguen cumpliendo. Para eso hacemos las clausuras de los atributos de los lados izquierdos de las dependencias de F y buscamos los lados derechos de las mismas en esa clausura:

$G_{F'}^+ = \{G\}$ por lo que la dependencia se perdió.

c) D, tiene JSP? Justificar

SOL:

Basta ver si se cumple que $R_1 \cap R_2 \rightarrow R_1 - R_2$ o $R_1 \cap R_2 \rightarrow R_2 - R_1$. Por lo que hacemos la clausura de CD con respecto a F':

$CD_{F'}^+ = \{C,D,E,H,B,G\}$ Por lo que se cumple que $CD \rightarrow BEG$ lo que es $R_1 - R_2$.

Esto significa que tenemos JSP.

d) Indique en que forma normal está D y en qué forma normal están cada uno de sus esquemas de relación. Justificar.

Sol:

Si observamos $\Pi_{R_1}(F)$ vemos que B no determina a nadie por lo que claramente es un atributo no primo. Por otro lado, en la parte anterior se vió que CD es una clave de R_1 por lo que $D \rightarrow B$ induce una dependencia parcial de clave sobre un atributo no primo por lo que viola 2NF. Conclusión: R_1 está en 1NF.

Con respecto a R_2 vemos que A tiene que estar en todas las claves, por lo que las únicas claves son ACH y ACD. Con esto vemos que todos los atributos son primos, pero las dependencias que se proyectaron no tienen una superclave del lado izquierdo. Por este motivo, esas dependencias violan BCNF pero no violan 3NF. Conclusión: R_2 está en 3NF.

Por lo anterior, D está en 1NF.

- e) Llevar R (el esquema original) a 3NF con JSP.

SOL:

Lo primero entonces es calcular el cubrimiento minimal de $F = \{CD \rightarrow E, G \rightarrow H, D \rightarrow BG, CH \rightarrow D\}$, por lo que:

- 1) Descomposición de dependencias: $F_1 = \{CD \rightarrow E, G \rightarrow H, D \rightarrow B, D \rightarrow G, CH \rightarrow D\}$
- 2) Eliminación de atributos Redundantes:

Para eso vemos si se pueden eliminar atributos del lado izquierdo de las dependencias que tienen más de un atributo del lado izquierdo en F_1

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

Dado que C^+ no contiene a E, entonces no se puede eliminar D en $CD \rightarrow E$.

Por el mismo motivo con respecto a D^+ no se puede eliminar C en $CD \rightarrow E$.

Dado que C^+ no contiene a D, entonces no se puede eliminar H en $CH \rightarrow D$.

Por el mismo motivo con respecto a H^+ no se puede eliminar C en $CH \rightarrow D$.

Por todo lo anterior, entonces no hay atributos redundantes del lado izquierdo.

- 3) Eliminación de dependencias Redundantes:

No hay dependencias redundantes porque no hay lados derechos repetidos.

El cubrimiento minimal es:

$$\{CD \rightarrow E, G \rightarrow H, D \rightarrow B, D \rightarrow G, CH \rightarrow D\}$$

Ahora construimos una tabla para cada lado izquierdo en donde ponemos los atributos de todas las dependencias en que aparece ese lado izquierdo, obteniendo:

$$R_1(C, D, E)$$

$$R_2(G, H)$$

$$R_3(D, B, G)$$

$$R_4(C, H, D)$$

Como no hay ninguna clave de R incluida en ninguna tabla, entonces agregamos una tabla que tenga una de estas claves. Pero primero debemos determinar al menos una de esas claves.

Observando el cubrimiento minimal, vemos que los atributos que nunca a parece a la derecha son A y C. El único que sólo aparece del lado derecho es B.

Por esto verificamos si AC es clave: $AC^+ = \{A, C\}$ por lo que no es clave.

Luego verificamos con ACD, $ACD^+ = \{A, C, D, B, G, H, E\}$ Que sí es clave. Por lo tanto agregamos $R_5(A, C, D)$.

La descomposición final es:

$$R_1(C, D, E)$$

$$R_2(G, H)$$

$$R_3(D, B, G)$$

$$R_4(C, H, D)$$

$$R_5(A, C, D)$$

La que tiene JSP y no pierde dependencias porque lo garantiza el algoritmo

aplicado.

Parte 4 Optimización y Concurrencia (25 puntos)

Ejercicio 5. (15 pts).

Dado el siguiente esquema:

ACTIVIDADES (NroAct, Fecha, Puntaje, TipoAct, IdEstudiante)
ESTUDIANTES (IdEstudiante, AñoIng, Direccion, Telefono, FNac)
 $\Pi_{IdEstudiante} (ACTIVIDADES) \subseteq \Pi_{IdEstudiante} (ESTUDIANTES)$

y la siguiente consulta:

```
SELECT *  
FROM ACTIVIDADES A, ESTUDIANTES E  
WHERE A.IdEstudiante = E.IdEstudiante AND A.Puntaje = 0
```

- Dar 2 planes lógicos posibles para la consulta.
- Dar un plan físico asociado a cada plan lógico dado.
- Calcular los costos de los planes físicos dados y compararlos. No olvidar considerar los costos de grabar para los resultados intermedios. Considere que se tienen 3 buffers disponibles.

Datos:

| Tabla | Columna | Valores distintos |
|-------------|--------------|----------------------|
| Estudiantes | IdEstudiante | 50 |
| Actividades | IdEstudiante | 50 |
| Actividades | Puntaje | 10 (distr. uniforme) |

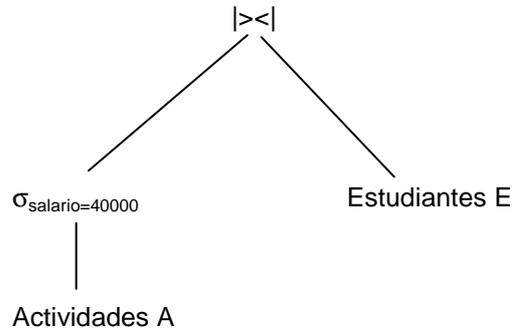
| Tabla | Cant. tuplas | Factor de bloqueo |
|-----------------------------|--------------|-------------------|
| Estudiantes | 50 | 10 |
| Actividades | 10000 | 5 |
| Estudiantes >< Actividades | | 3 |

| Indice | Tabla/Atributo | Tipo | Cant. niveles |
|------------------|--------------------------|---------------|---------------|
| Est_IdEstudiante | Estudiantes/IdEstudiante | Primario | 1 |
| Activ_Puntaje | Actividades/Puntaje | Secundario B+ | 1 |

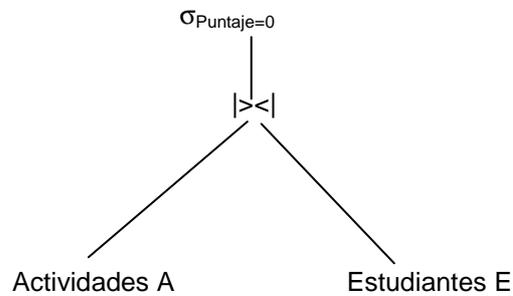
Solución:

a)

Plan lógico 1:

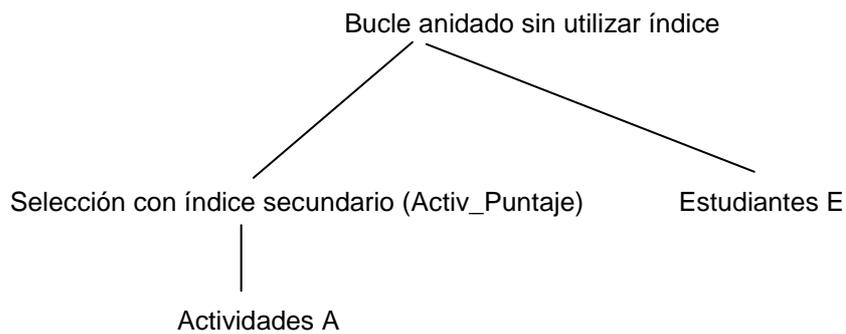


Plan lógico 2:

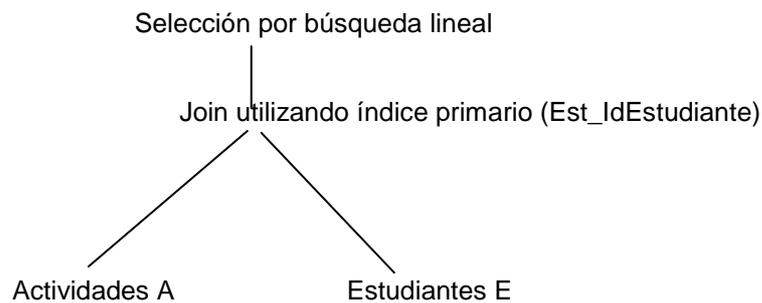


b)

Plan físico 1:



Plan físico 2:



c)

Costo Plan 1:

$$C(\text{seleccion}) = x + |\text{selección}| = 1 + 1000 = 1001$$

$$C(\text{grabar}) = b_{\text{seleccion}} = |\text{seleccion}| / \text{factor de bloqueo} = 1000 / 5 = 200$$

$$C(\text{join}) = b_{\text{seleccion}} + b_{\text{seleccion}} * b_E = 200 + 200 * 5 = 1200$$

Aclaración:

$$|\text{seleccion}| = |A| / V(\text{puntaje}, A) = 10000 / 10 = 1000$$

Ctotal = 2401

Costo Plan 2:

$$C(\text{join}) = b_R + n_R * (x + 1) = 2000 + 10000 * 2 = 22000$$

$$C(\text{grabar}) = b_{\text{join}} = |\text{join}| / \text{factor de bloqueo} = 10000 / 3 = 3334$$

$$C(\text{selección}) = b_{\text{join}} = 3334$$

Aclaración:

$|\text{join}| = |A| = 10000$, porque IdEstudiante es clave en Estudiantes y clave foránea en Actividades

Ctotal = 28668

Vemos que el Plan 1 es mucho menos costoso que el Plan 2.

Ejercicio 6. (10 pts).

Dadas 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

a) Marcar en el cuadro con "S" o "N" si cada propiedad es cumplida o no por cada historia dada.

| | Ser | Rec | EAC | Estr |
|---|-----|-----|-----|------|
| r1(X) w2(Z) w1(X) r1(Y) w1(Y) r1(Z) w1(Z) r2(Y) w2(Y) c1 c2 | N | N | N | N |
| r1(X) w1(X) r1(Y) w2(Z) r2(Y) w1(Y) r1(Z) w1(Z) w2(Y) c2 c1 | N | S | N | N |
| w2(Z) r2(Y) r1(X) w1(X) r1(Y) w1(Y) w2(Y) c2 r1(Z) w1(Z) c1 | N | S | S | N |

Notación:

Ser – Serializable
 Rec – Recuperable
 EAC – Evita Abortos en Cascada
 Estr – Estricta

b) Dar un historia entrelazada de T1 y T2 con bloqueos y desbloqueos, tal que las transacciones sigan 2PL. Utilizar bloqueos de escritura y de lectura. Decir si la historia es serializable, justificando.

H: r1(X) r1(X) w1(X) w1(X) w2(Z) w2(Z) r2(Y) r2(Y) w2(Y) w2(Y) u2(Z) u2(Y) r1(Y) r1(Y) w1(Y) w1(Y) r1(Z) r1(Z) w1(Z) w1(Z) u1(Y) u1(Z) c1 c2

La historia es serializable, ya que toda historia formada por transacciones que siguen 2PL es serializable.

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 |

| 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 |
| $R X _c S$ | 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 |

donde Z depende del tipo de índice:

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