

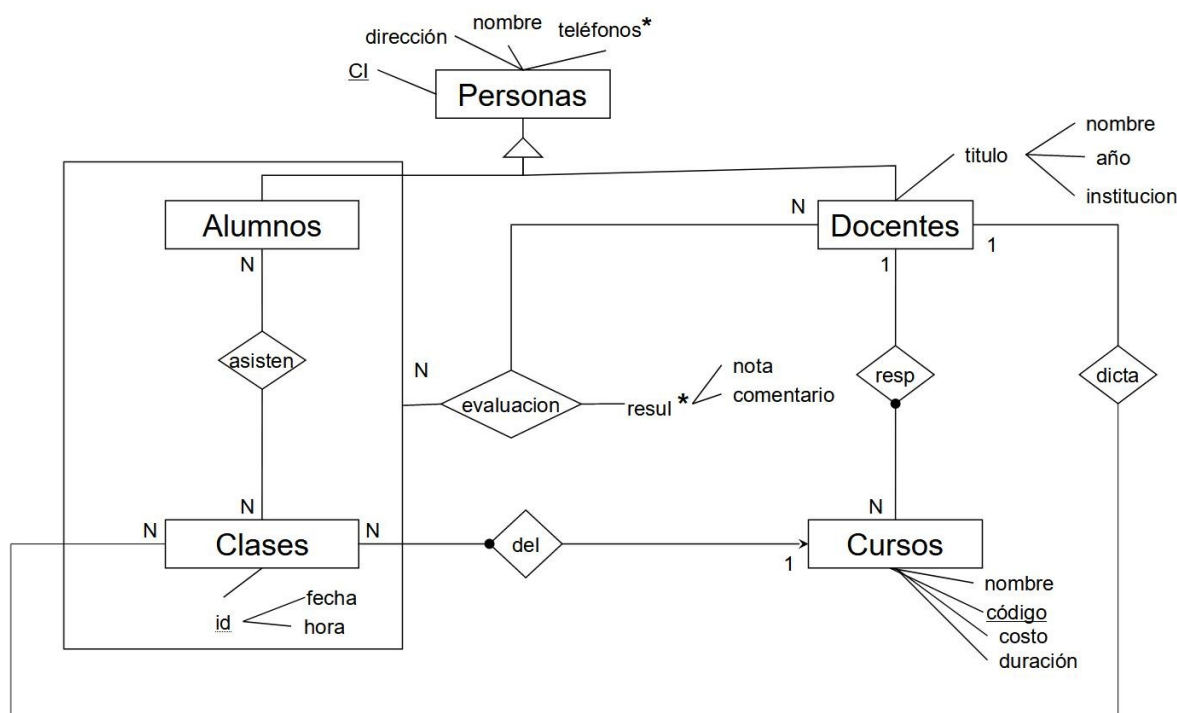
Segundo Parcial de Fundamentos de Bases de Datos

Noviembre 2021

SOLUCIÓN

Ejercicio 1 (18 puntos)

Considere el siguiente Modelo de Entidad Relación (MER) con información sobre cursos y un conjunto de Restricciones No Estructurales (RNEs) sobre él.



Restricciones no estructurales:

1) $\text{Alumnos} \cap \text{Docentes} = \emptyset$

2) Un alumno que asiste a una clase que se dicta, debe tener una evaluación donde participa el docente que la dicta

$$(\forall d \in \text{docentes})(\forall a \in \text{alumnos})(\forall c \in \text{clases}) \\ (<a,c> \in \text{asisten} \wedge <d,c> \in \text{dicta} \textcircled{R} <d,<a,c>> \in \text{evaluacion})$$

3) Un alumno no puede asistir a más de una clase en a la misma fecha, hora.

$$(\forall a \in \text{alumnos})(\forall c1,c2 \in \text{clases}) \\ (<a,c1> \in \text{asisten} \wedge <a,c2> \in \text{asisten} \textcircled{R} \text{id}(c1) \neq \text{id}(c2))$$

5) Toda persona tiene al menos un teléfono.

$$(\forall p \in \text{Personas})(\text{COUNT}(\text{telefonos}(p)) > 0)$$

6) El costo de los cursos es mayor o igual 0, mientras que la duración es mayor a 0.

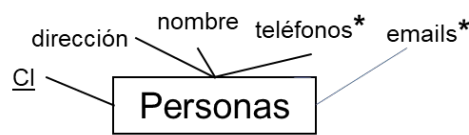
$$(\forall c \in \text{Cursos})(c.\text{costo} \geq 0 \wedge c.\text{duracion} > 0)$$

Se pide:

1. Teniendo en cuenta el MER y las restricciones no estructurales presentadas:

- a. elabore un esquema relacional que incluya los esquemas relación así como las claves (tanto las primarias como las alternativas) y las dependencias de inclusión,
- b. indicar las dependencias funcionales que se cumplen y que no han sido ya expresadas sobre cada una de las relaciones del esquema relacional obtenido en la parte (a). **NO** incluir ni triviales ni redundantes.
- c. indicar las restricciones no estructurales que no fueron posibles expresar ni en (a) y ni en (b). **NO** se pide la elaboración de un método para controlarlas.

2. Considere que ahora se quiere incluir la información de las direcciones de correo (emails) de las personas. Para ello, se modifica la entidad Personas:



y se define el siguiente esquema relación:

PERSONAS(ci, direccion, nombre, telefono, email)

- a. Indique las dependencias funcionales que se cumplen en este esquema relación dada la modificación de la entidad Personas.
- b. Se tiene información de una persona llamada Draco Malfoy, de la que se conoce que su cédula es 7.563.107-8, su dirección es Rivera 3456, sus teléfonos son 2467.879 y 4345.2222 y su emails son dracom@gmail.com y dmalfoy@hotmail.com. Dar una instancia del esquema Personas que represente a esta información.
- c. ¿Qué dificultades encuentra en este esquema? Usando conceptos de normalización vistos en el curso, ¿cómo lo mejoraría?

Solución

Parte 1

a)

PERSONAS(CI, dirección, nombre)

PERSONAS_TEL (CI, telefono)

$\Pi_{CI}(\text{PERSONAS_TEL}) \subseteq \Pi_{CI}(\text{PERSONAS})$

Por la RNE “Toda persona tiene al menos un teléfono” se agrega la siguiente dependencia de inclusión:
 $\Pi_{CI}(\text{PERSONAS}) \subseteq \Pi_{CI}(\text{PERSONAS_TEL})$

DOCENTES (CI, nombre, año, institucion)

ALUMNOS(CI)

$\Pi_{CI}(\text{DOCENTES}) \subseteq \Pi_{CI}(\text{PERSONAS})$

$\Pi_{CI}(\text{ALUMNOS}) \subseteq \Pi_{CI}(\text{PERSONAS})$

Por la RNE 1:

$\Pi_{CI}(\text{DOCENTES}) \cap \Pi_{CI}(\text{ALUMNOS}) = \emptyset$

CURSOS(codigo, nombre, costo, duración, CIResponsable)

$\Pi_{CIResponsable}(\text{CURSOS}) \subseteq \Pi_{CI}(\text{DOCENTES})$

CLASES (codigoCurso, fecha, hora,)

$\Pi_{codigoCurso}(\text{CLASES}) \subseteq \Pi_{codigo}(\text{CURSOS})$

ASISTEN (codigoCurso, fecha, hora, CIAlumno)

$\Pi_{codigoCurso, fecha, hora}(\text{ASISTEN}) \subseteq \Pi_{codigoCurso, fecha, hora}(\text{CLASES})$

$\Pi_{CIAlumno}(\text{ASISTEN}) \subseteq \Pi_{CI}(\text{ALUMNOS})$

DICTA (codigoCurso, fecha, hora, CIDocente)

$\Pi_{codigoCurso, fecha, hora}(\text{DICTA}) \subseteq \Pi_{codigoCurso, fecha, hora}(\text{CLASES})$

$\Pi_{CIDocente}(\text{DICTA}) \subseteq \Pi_{CI}(\text{DOCENTES})$

EVALUACIONES(codigoCurso, fecha, hora, CIAlumno, CIDocente)

$\Pi_{codigoCurso, fecha, hora, CIAlumno}(\text{EVALUACIONES}) \subseteq \text{ASISTEN}$

$\Pi_{CIDocente}(\text{EVALUACIONES}) \subseteq \Pi_{CI}(\text{DOCENTES})$

RES_EVALUACIONES(codigoCurso, fecha, hora, CIAlumno, CIDocente, nota, comentario)

$\Pi_{codigoCurso, fecha, hora, CIAlumno, CIDocente}(\text{RES_EVALUACIONES}) \subseteq \Pi_{codigoCurso, fecha, hora, CIAlumno, CIDocente}(\text{EVALUACIONES})$

Por la RNE2 sabemos que todos los alumnos que asisten a una clase deben tener al menos una evaluación de un docente que la dicta. Si bien esto **no** puede representarse completamente en el esquema, surgen las siguientes dependencias de inclusión:

- $\Pi_{codigoCurso, fecha, hora, CIDocente}(\text{EVALUACIONES}) \subseteq \text{DICTA}$, ya que el docente que dicta la clase es evaluador de alumnos de su clase)
- $\text{ASISTEN} \subseteq \Pi_{codigoCurso, fecha, hora, CIAlumno}(\text{EVALUACIONES})$, ya que quienes asisten tienen al menos una evaluación.

NOTA: una solución sin la tabla Evaluaciones, sólo con Res_Evaluaciones, también es una solución posible si se asume que el atributo multivaluado result no puede ser vacío.

b) Por la RNE 3 se cumple:

ASISTEN (codigoCurso, fecha, hora, CIAlumno)

CIAlumno, fecha, hora -> codigoCurso

c) Los números de las RNEs que faltan controlar en el esquema son la 2 y la 6

Parte 2

a) En este nuevo diagrama aparece la DF

ci → direccion, nombre

b)

Ci	Direccion	Nombre	Telefono	Email
7.563.107-8	Rivera 3456	Draco Malfoy	2467.879	dracrom@gmail.com
7.563.107-8	Rivera 3456	Draco Malfoy	4345.2222	dmalfoy@hotmail.com

Si sólo se tienen estas tuplas, se estaría dejando a la (falsa) interpretación de que en presencia de determinado telefono se tendría determinado email (o viceversa). Para reflejar que no hay relación entre ellos como se representa en el MER entonces también son necesarias las tuplas:

7.563.107-8	Rivera 3456	Draco Malfoy	2467.879	dmalfoy@hotmail.com
7.563.107-8	Rivera 3456	Draco Malfoy	4345.2222	dracrom@gmail.com

d) En el esquema PERSONAS(ci, direccion, nombre, telefono, email) las instancias presentan valores redundantes en el caso general donde una persona tiene varios telefonos y emails, tanto en los atributos telefono y email como en dirección y nombre. Esto dificulta el mantenimiento de la consistencia entre las tuplas al aplicar las operaciones de inserción, borrado y actualización.

Además, la clave del esquema son todos los atributos y la DF de la parte (a) hace que direccion y nombre dependan parcialmente de la clave, por lo tanto el esquema propuesto no está en 2NF, está en 1NF.

Se puede aplicar el algoritmo visto en el curso para llevar este esquema a BCNF con join sin pérdida.

1. PERSONAS_UBICACION(ci, direccion, nombre)

en ella se proyecta: ci → direccion, nombre y por lo tanto la única clave es CI

2. PERSONAS_CONTACTO(ci, telefono, email)

no se proyecta ninguna DF, y por lo tanto la única clave la forman todos los atributos.

Con PERSONAS_UBICACION se mejora la redundancia en los atributos direccion y nombre. En cambio PERSONAS_CONTACTO sigue teniendo redundancia.

En PERSONAS_CONTACTO se cumple la siguiente dependencia multivaluada:

ci ->> telefono | email

Dado que la clave es (ci, telefono, email), dicha relación no está en 4NF. Se puede aquí aplicar el algoritmo visto en el curso para llevar este esquema a 4NF con join sin pérdida. Aplicando un sólo paso se logra una descomposición en 4NF.

2.1 PERSONAS_TEL(ci, telefono)

2.2 PERSONAS_EMAIL(ci, email)

Ejercicio 2 (18 puntos)

Considere el esquema de relación $R(A,B,C,D,E,G,H)$, el conjunto de dependencias funcionales $F = \{C \rightarrow D, A \rightarrow BH, CA \rightarrow E, H \rightarrow G, HE \rightarrow CA\}$, y la siguiente descomposición:

$$\rho = \{R1(A,B,C,E), R2(A,D,E,G,H)\}$$

1) Para cada una de las afirmaciones, indique si es verdadera o falsa, **justificando**.

- a) C no pertenece a ninguna clave de R según F.
- b) (AEH) es clave de R según F.
- c) R está en 2NF respecto a F.
- d) La descomposición ρ tiene JSP.
- e) $\Pi_{R2}(F) = \{H \rightarrow G\}$

2)

- a) Encuentre todas las claves de R según F.
- b) Obtenga una descomposición en 3NF de R con JSP y preservación de dependencias, aplicando el algoritmo visto en el curso y justificando cada paso.

Solución

Parte 1

- a) **Falso.** AC es clave de R porque determina a todos los atributos, y además A y C no son superclaves, como se muestra a continuación:

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

$$C^+ = \{C, D\}$$

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

- b) **Falso.** Como AE es clave, entonces AEH es superclave, y no es clave. Para ver que AE es clave, vemos que A y E no son claves y que AE determina a todos los atributos:

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

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

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

- c) **Falso.** Vimos anteriormente que AE es clave de R según F, y B es un atributo no primo ya que no pertenece a ninguna clave (dado que aparece solo a la derecha de las dependencias funcionales). Entonces la dependencia funcional $A \rightarrow B$ viola 2NF, ya que un atributo no primo depende parcialmente de una clave.

- d) **Verdadero.** La descomposición tiene JSP ya que se cumple la DF $(R1 \cap R2) \rightarrow (R1 - R2)$, que en este caso sería $AE \rightarrow BC$.

- e) **Falso.** Faltan dependencias en la proyección, por ejemplo $A \rightarrow H$.

Parte 2

- a) Atributos que aparecen solo a la derecha: $D = \{B, D, G\}$
Atributos que nunca aparecen a la derecha: $ND = \{\}$

Los atributos que están en D no pertenecen a ninguna clave, y los que están en ND a todas.

Como ND es vacío, probamos con los atributos que aparecen a la izquierda y a la derecha:

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

$$C^+ = \{C, D\}$$

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

$$H^+ = \{H, G\}$$

Ninguno es superclave así que probamos combinaciones:

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

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

$$(A,H)^+ = \{A, B, H, G\}$$

$$(C,E)^+ = \{C, E, D\}$$

$$(C,H)^+ = \{C, H, D, G\}$$

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

AC, AE y EH son todas las claves. No hay claves de 3 atributos porque todas las posibles contienen a una de las tres claves que ya encontramos.

b) Paso 1 – Encontrar un cubrimiento minimal G.

Primero separo los atributos de la derecha de las Dfs:

$C \rightarrow D$
 $A \rightarrow B$
 $A \rightarrow H$
 $CA \rightarrow E$
 $H \rightarrow G$
 $HE \rightarrow A$
 $HE \rightarrow C$

Luego elimino los atributos redundantes a la izquierda de las Dfs:

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

Como E no pertenece a C^+ ni a A^+ , la df $CA \rightarrow E$ queda igual

Como A no pertenece a H^+ ni a E^+ , la df $HE \rightarrow A$ queda igual

Como C no pertenece a H^+ ni a E^+ , la df $HE \rightarrow C$ queda igual

El último paso es eliminar las DFs redundantes. Observando las DF se puede ver que en todas ellas el atributo que aparece a la derecha no aparece a la derecha de ninguna otra DF. Debido a esto, ninguno de los atributos de la derecha de una DFs se podría derivar a partir de las otras DFs, por lo que el conjunto ya es minimal.

$G = \{C \rightarrow D, A \rightarrow B, A \rightarrow H, CA \rightarrow E, H \rightarrow G, HE \rightarrow A, HE \rightarrow C\}$

Paso 2 – Generar un esquema para cada parte izquierda de las DFs.

R1 (C, D)
R2 (A, B, H)
R3 (C, A, E)
R4 (H, G)
R5 (H, E, A, C)

Paso 3 – Si ningún esquema tiene una clave de R, crear uno con ella.

El esquema R3 contiene la clave CA, así que no es necesario agregar nada.

Paso 4 – Si hay algún esquema redundante, eliminarlo.

$R3 \subseteq R5$

La descomposición resultante resultante es entonces (R1, R2, R4, R5)

Ejercicio 3 (12 puntos)

Considere el esquema relacional de un subconjunto de la base de datos del sitio IMDB, que almacena, entre otras cosas, información sobre películas, y que fue utilizada en la tarea de laboratorio del curso.

PELICULAS(pele_id, nombre, fecha, duracion, presupuesto, ganancia, web, votos_prom)

Contiene información sobre las películas: su identificador, su nombre (en el idioma original), su fecha de estreno, su duración en minutos, el presupuesto que se tenía para realizar la película, los ingresos que generó, un sitio web (que puede no existir) y el promedio de votos que la película ha recibido en el sitio IMDB hasta el momento.

SAGAS(saga_id, nombre, fecha)

Contiene información sobre las sagas: su identificador, su nombre y la fecha de estreno de la última película perteneciente a la saga.

PELI_EN_SAGA(saga_id, pele_id)

Almacena las películas que pertenecen a cada saga.

$$\begin{aligned}\Pi_{pele_id}(PELI_EN_SAGA) &\subseteq \Pi_{pele_id}(PELICULAS) \\ \Pi_{saga_id}(PELI_EN_SAGA) &\subseteq \Pi_{saga_id}(SAGAS)\end{aligned}$$

Se sabe que todos los atributos tienen distribución uniforme (recuerde que $V(A, T)$ es la cantidad de valores distintos que tiene un atributo A en una tabla T). Además, de cada tabla se conoce la siguiente información:

Relación R	n_R	Atributos	Índices
PELICULAS P	5000	$V(\text{fecha}, P) = 1000$	Índice primario sobre pele_id
SAGAS S	80		Índice primario sobre saga_id
PELI_EN_SAGA PS	1000		Índice Primario sobre (saga_id, pele_id)

Considere la siguiente consulta sobre el esquema dado:

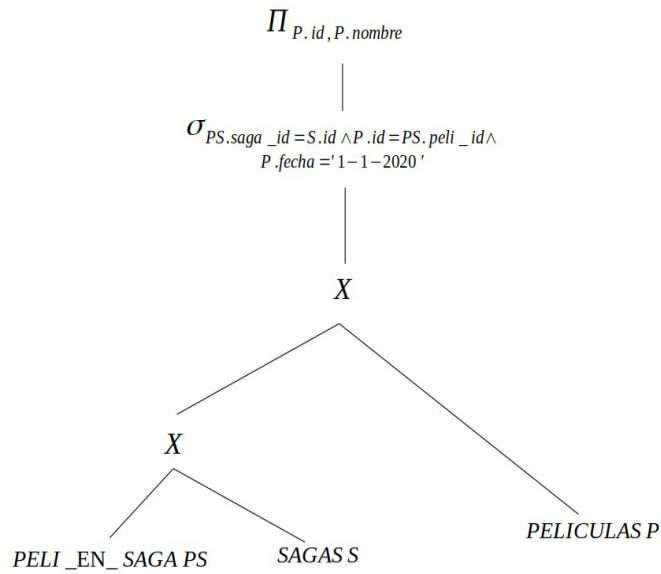
```
SELECT p.id, p.name
FROM PELI_EN_SAGA ps, SAGAS s, PELICULAS p
WHERE ps.saga_id = s.id AND p.id=ps.pele_id AND p.fecha= '1-1-2020';
```

Se pide:

1. Dar el árbol canónico para la consulta.
2. Aplicar las heurísticas para optimización llegando al plan lógico optimizado. Explique cada uno de los pasos ejecutados.
3. Dar un plan físico para el plan lógico obtenido en la **parte 2**, utilizando índices en los casos en que sea posible.
4. Considere ahora que se agrega un índice secundario en el atributo fecha de la tabla PELÍCULAS, ¿haría algún cambio al plan físico propuesto en la **parte 3**?
5. Se decide modificar el índice primario de la tabla PELI_EN_SAGA, de forma que el nuevo índice primario sea (pele_id, saga_id), ¿haría algún cambio al plan físico propuesto en la **parte 3**?
6. Finalmente, se desea agregar a la consulta una selección sobre la tabla SAGAS que no afecta el orden de las hojas del árbol obtenido en la **parte 2**. Además, esa selección se resuelve en el plan físico mediante la operación de búsqueda lineal. Dar el nuevo plan físico adecuado para esta consulta.

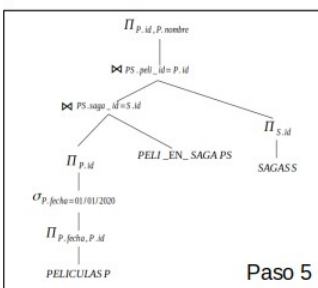
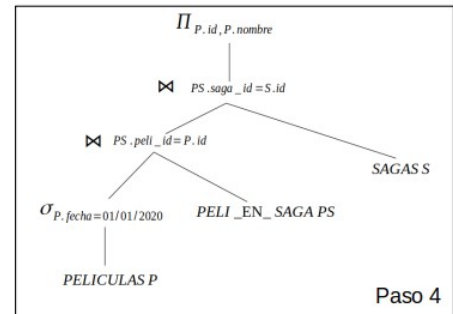
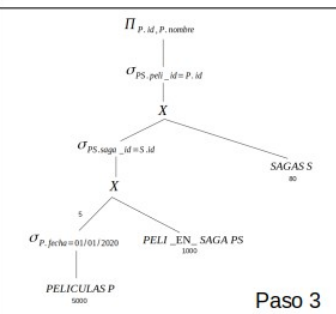
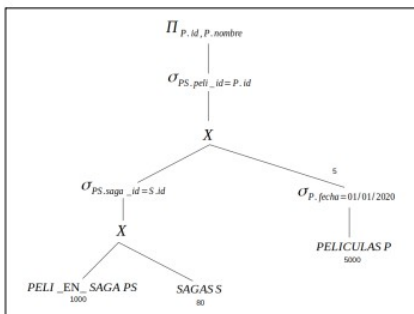
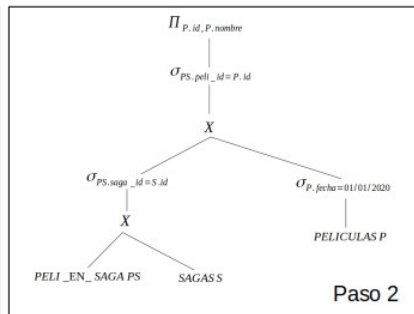
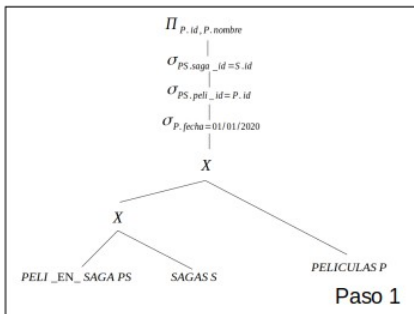
Solución

1)

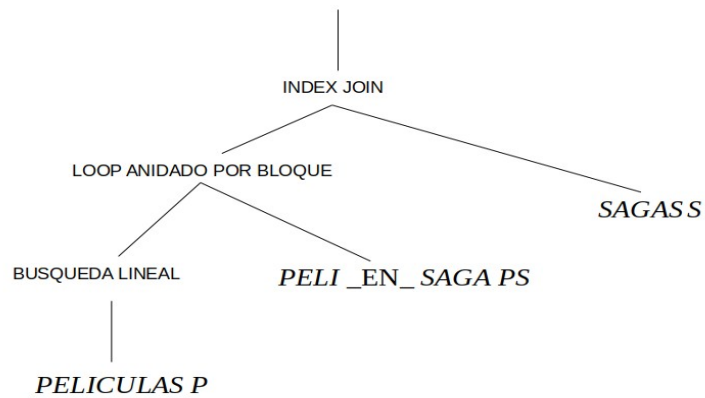


2) Heurísticas:

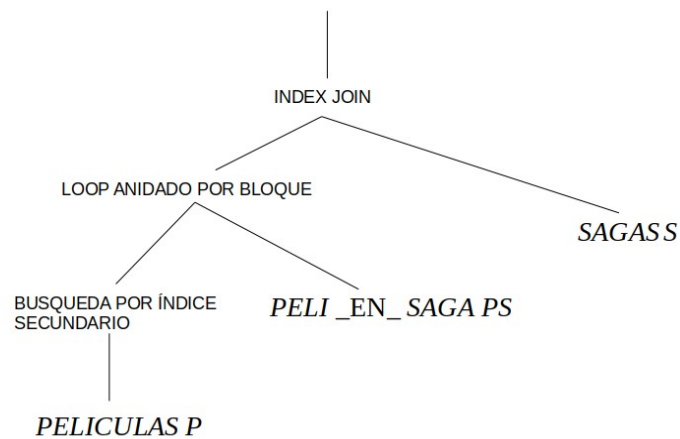
1. Cambiar las selecciones conjuntivas por una “cascada” de selecciones simples.
2. Mover las selecciones lo más abajo que se pueda en el árbol.
3. Poner a la izquierda de los productos las hojas que generen menos tuplas, asegurando que el orden de las hojas no cause operaciones de producto cartesiano (que no pueden convertirse en join).
4. Cambiar secuencias de selecciones y productos por join's.
5. Mover las proyecciones lo más abajo posible en el árbol, agregando las proyecciones que sean necesarias.



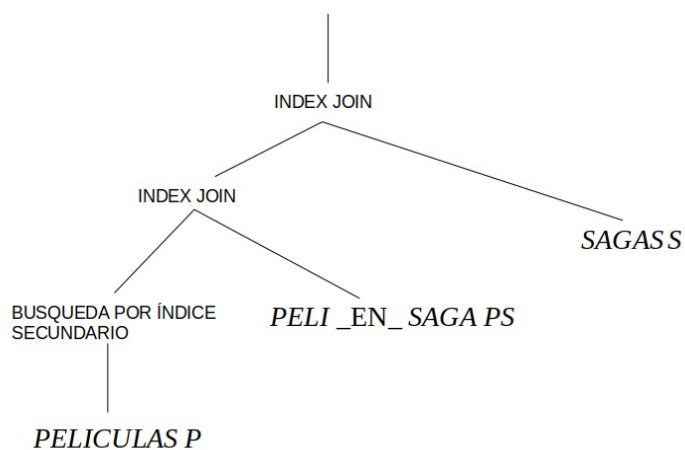
3)



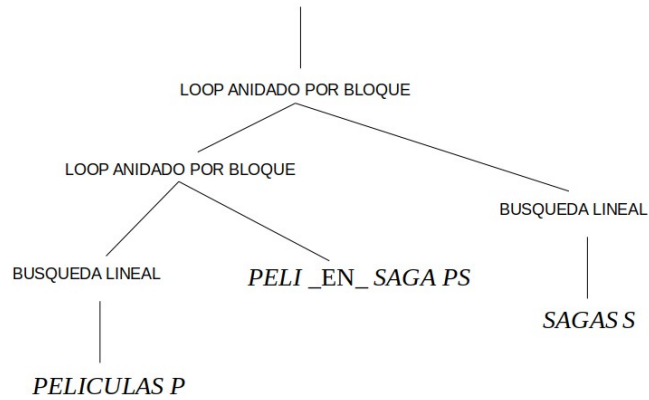
4) Se puede aplicar búsqueda por índice secundario. No hay restricción para aplicar búsqueda por este tipo de índice.



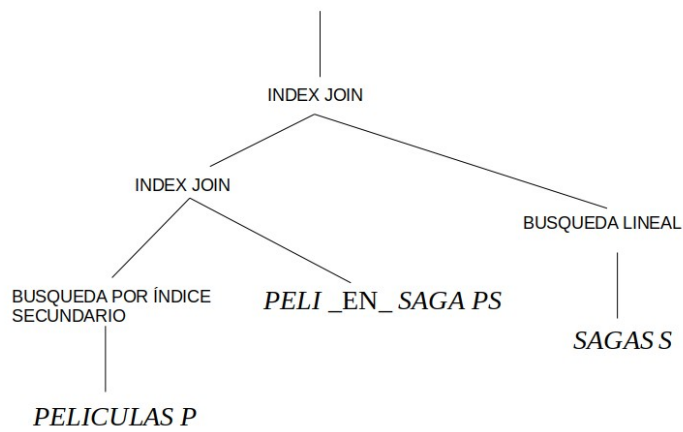
5) Sí, ahora se puede usar un index join entre PELI_EN_SAGA y PELÍCULAS, ya que se toma el primer atributo del índice.



6) No se puede aplicar index join porque se pierde el índice al realizar la selección, el plan físico queda de la siguiente forma si se parte del plan lógico obtenido en la parte 2.



Si se continúa trabajando sobre el plan físico obtenido en las partes anteriores, el nuevo plan es el siguiente:



Ejercicio 4 (12 puntos)

Considere las siguientes transacciones:

$T_1: r_1(x), w_1(y), w_1(x), r_1(y) c_1$

$T_2: r_2(y), w_2(x), w_2(y), c_2$

a) Justificando su respuesta, escriba una historia no serializable con estas transacciones.

b) En esas transacciones se agregaron los bloqueos de read-write, sin seguir ninguno de los protocolos 2PL visto en el curso:

$T_{11}: rl_1(x), r_1(x), u_1(x), wl_1(y), w_1(y), u_1(y), wl_1(x), w_1(x), u_1(x), rl_1(y), r_1(y), u_1(y), c_1$

$T_{21}: rl_2(y), r_2(y), u_2(y), wl_2(x), w_2(x), u_2(x), wl_2(y), w_2(y), u_2(y), c_2$

Si es posible, escriba una historia no serializable con estas transacciones. Justifique su respuesta.

c) Escriba las transacciones T_{11}' y T_{21}' de forma de que sean iguales con respecto a los read y write, pero teniendo los bloqueos y desbloqueos de manera que sigan 2PL estricto. Justifique su respuesta.

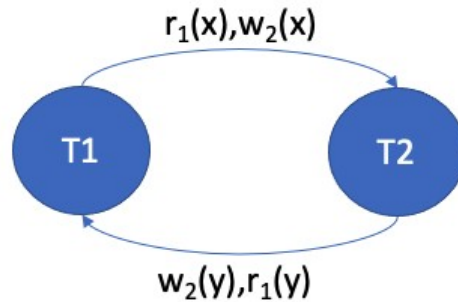
d) Si es posible, escriba con las transacciones de la parte c, una historia no serializable. Justifique su respuesta.

Solución

a) La siguiente historia cumple con lo pedido en la letra:

$r_2(y), r_1(x), w_2(x), w_2(y), w_1(y), w_1(x), r_1(y), c_2, c_1$

Siendo el siguiente su grafo de serialidad:



Al tener un ciclo, es posible afirmar que la historia es no serializable.

b) Las operaciones de bloqueo y desbloqueo no cambiaron el orden de los read y write, por esto, la historia resultante puede ser idéntica a la planteada en la parte anterior, que como vimos, es no serializable.

c) En 2PL estricto debemos de garantizar:

1. Se respeta el concepto de 2 fases: crecimiento y contracción
2. Se liberan los bloqueos de escritura luego de la finalización de la transacción
3. Promoción de bloqueos: los bloqueos aumentan en calidad, no hay rl sobre un elemento luego de que se realizó un wl
4. Los bloqueos de escritura se liberan luego de finalizada la transacción

Las siguientes transacciones cumplen con los requisitos:

$T_{11}': r_1(x), r_1(x), w_1(y), w_1(y), w_1(x), w_1(x), r_1(y), c_1, u_1(y), u_1(x)$

$T_{21}': r_2(y), r_2(y), w_2(x), w_2(x), w_2(y), w_2(y), c_2, u_2(x), u_2(y)$

d) La transacción de la parte C sigue 2PL estricto, por lo que, no es posible generar historias no serializables, el protocolo lo asegura.