

Segundo Parcial de Fundamentos de Bases de Datos

Noviembre 2024

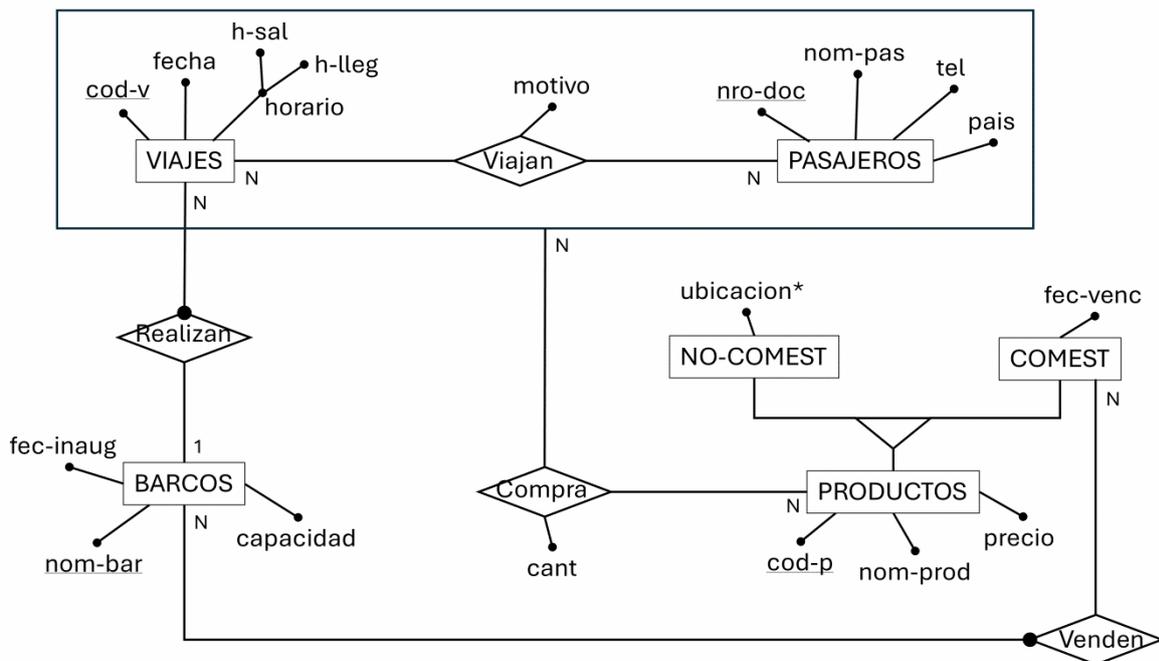
Duración: 3 horas

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 (24 puntos)

Una empresa de transporte marítimo de pasajeros realizó el siguiente modelo conceptual para construir una base de datos de su información de interés.



RNE:

- $NO-COMEST \cap COMEST = \emptyset$
- $NO-COMEST \cup COMEST = PRODUCTOS$
- $(\forall n \in NO-COMEST) (n.ubicacion \neq \emptyset)$
- Un pasajero no puede realizar más de un viaje en la misma fecha.
 $(\neg \exists p \in PASAJEROS) ((\exists v1 \in VIAJES) ((\exists v2 \in VIAJES) (\langle p,v1 \rangle \in VIAJAN \wedge \langle p,v2 \rangle \in VIAJAN \wedge v1.fecha = v2.fecha)))$

Se pide resolver, justificando las respuestas:

Parte 1.

a) Realizar el pasaje del Modelo Entidad Relación dado al Modelo Relacional. El esquema resultado debe incluir las tablas, sus claves primarias y las dependencias de inclusión.

VIAJES(cod-v, fecha, h-sal, h-lleg, nom-barc)

PASAJEROS(nro-doc, nom-pas, tel, pais)

VIAJAN(cod-v, nro-doc, motivo)

$$\prod_{cod-v} (VIAJAN) \subseteq \prod_{cod-v} (VIAJES)$$

$$\prod_{nro-doc} (VIAJAN) \subseteq \prod_{nro-doc} (PASAJEROS)$$

BARCOS(nom-barc, fec-inaug, capacidad)

$$\prod_{nro-barc} (VIAJES) \subseteq \prod_{nro-barc} (BARCOS)$$

PRODUCTOS(cod-p, nom-prod, precio)

COMEST(cod-p, f-venc)

NO-COMEST(cod-p)

NO-COMEST-UB(cod-p, ubicación)

$$\prod_{cod-p} (NO-COMEST) \subseteq \prod_{cod-p} (PRODUCTOS)$$

$$\prod_{cod-p} (COMEST) \subseteq \prod_{cod-p} (PRODUCTOS)$$

$$\prod_{cod-p} (PRODUCTOS) = \prod_{cod-p} (NO-COMEST) \cup \prod_{cod-p} (COMEST)$$

$$\prod_{cod-p} (NO-COMEST-UB) = \prod_{cod-p} (NO-COMEST)$$

$$\prod_{cod-p} (NO-COMEST) \cap \prod_{cod-p} (COMEST) = \emptyset$$

VENDEN(cod-p, nom-barc)

$$\prod_{cod-p} (VENDEN) \subseteq \prod_{cod-p} (COMEST)$$

$$\prod_{nom-barc} (VENDEN) = \prod_{nom-barc} (BARCOS)$$

COMPRA(cod-p, cod-v, nro-doc, cant)

$$\prod_{cod-p} (COMPRA) \subseteq \prod_{cod-p} (PRODUCTOS)$$

$$\prod_{cod-v, nro-doc} (COMPRA) \subseteq \prod_{cod-v, nro-doc} (VIAJAN)$$

b) Decir en qué forma normal se encuentra el esquema relacional obtenido en la parte anterior.

El algoritmo garantiza 1NF.

Obtengo las dfs en cada esquema:

$$F_{VIAJES} = \{ cod-v \rightarrow fecha, h-sal, h-lleg, nom-barc \}$$

$$F_{PASAJEROS} = \{ nro-doc \rightarrow nom-pas, tel, pais \}$$

$$F_{VIAJAN} = \{ cod-v, nro-doc \rightarrow motivo \}$$

$$F_{\text{BARCOS}} = \{ \text{nom-barc} \rightarrow \text{capacidad}, \text{fecha-inaug} \}$$

$$F_{\text{PRODUCTOS}} = \{ \text{cod-p} \rightarrow \text{nom-produ}, \text{precio} \}$$

$$F_{\text{COMEST}} = \{ \text{cod-p} \rightarrow \text{fec-venc} \}$$

$$F_{\text{NO-COMEST}} = \emptyset$$

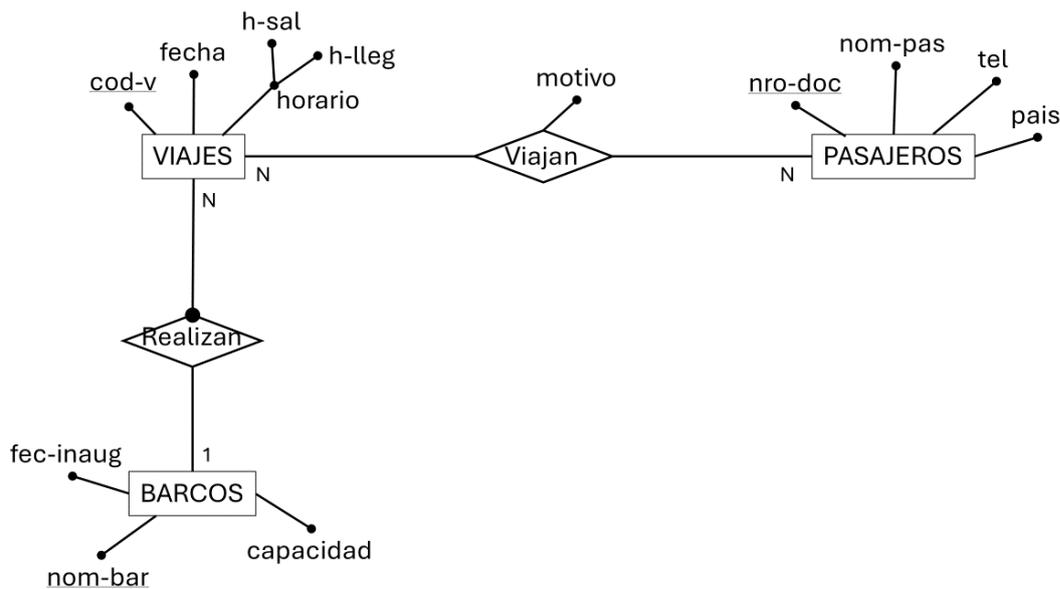
$$F_{\text{NO-COMEST-UB}} = \emptyset$$

$$F_{\text{COMPRA}} = \{ \text{cod-p}, \text{cod-v}, \text{nro-doc} \rightarrow \text{cant} \}$$

Todos los esquemas están en BCNF porque las dfs que se cumplen en cada esquema cumplen que el lado izquierdo es superclave. **Por lo tanto el esquema relacional completo está en BCNF (y 4NF si consideramos mvds).**

Parte 2.

Ahora considere solamente la siguiente parte de la realidad presentada:



RNE:

- Un pasajero no puede realizar más de un viaje en la misma fecha.

$$(\neg \exists p \in \text{PASAJEROS}) ((\exists v1 \in \text{VIAJES}) ((\exists v2 \in \text{VIAJES}) (\langle p, v1 \rangle \in \text{VIAJAN} \wedge \langle p, v2 \rangle \in \text{VIAJAN} \wedge v1.\text{fecha} = v2.\text{fecha})))$$

a) Indicar los nombres de las tablas de la solución dada en la **Parte 1 - a)**, que corresponden a esta porción de la realidad.

VIAJES(cod-v, fecha, h-sal, h-lleg, nom-barc)

PASAJEROS(nro-doc, nom-pas, tel, pais)

VIAJAN(cod-v, nro-doc, motivo)

BARCOS(nom-barc, fec-inaug, capacidad)

b) Considerar la Relación Universal U , que incluye todos los atributos de esta porción de la realidad. Construir el conjunto F de dependencias funcionales que se cumplen en U , considerando **toda** la información proporcionada por este nuevo MER.

$U = (\text{cod-v}, \text{fecha}, \text{h-sal}, \text{h-lleg}, \text{nom-barc}, \text{nro-doc}, \text{nom-pas}, \text{tel}, \text{pais}, \text{capacidad}, \text{fec-inaug}, \text{motivo})$

$F = \{$
 $\text{cod-v} \rightarrow \text{fecha}, \text{h-sal}, \text{h-lleg}, \text{nom-barc}$
 $\text{nro-doc} \rightarrow \text{nom-pas}, \text{tel}, \text{pais}$
 $\text{nom-barc} \rightarrow \text{capacidad}, \text{fecha-inaug}$
 $\text{cod-v}, \text{nro-doc} \rightarrow \text{motivo}$
 $\text{fecha}, \text{nro-doc} \rightarrow \text{cod-v}$
 $\}$

c) Aplicar a U el algoritmo de 3NF con preservación de dependencias funcionales y JSP, dado en el curso.

$F = \{$
 $\text{cod-v} \rightarrow \text{fecha}, \text{h-sal}, \text{h-lleg}, \text{nom-barc}$
 $\text{nro-doc} \rightarrow \text{nom-pas}, \text{tel}, \text{pais}$
 $\text{nom-barc} \rightarrow \text{capacidad}, \text{fecha-inaug}$
 $\text{cod-v}, \text{nro-doc} \rightarrow \text{motivo}$
 $\text{fecha}, \text{nro-doc} \rightarrow \text{cod-v}$
 $\}$

Claves:

SOLO_IZQ = {nro-doc}

IZQ_DER = {nom-barc, fecha, cod-v}

SOLO_DER = {h-sal, h-lleg, motivo, capacidad, fecha-inaug, nom-pas, pais, tel}

$(\text{nro-doc})^+ = \{\text{nro-doc}, \text{nom-pas}, \text{tel}, \text{pais}\}$

$(\text{nro-doc}, \text{fecha})^+ = \{\text{nro-doc}, \text{nom-pas}, \text{tel}, \text{pais}, \text{fecha}, \text{cod-v}, \text{h-sal}, \text{h-lleg}, \text{nom-barc}, \text{capacidad}, \text{fecha-inaug}, \text{motivo}\} = U$

$(\text{nro-doc}, \text{fecha})$ es clave.

$(\text{nro-doc}, \text{cod-v})^+ = \{\text{nro-doc}, \text{nom-pas}, \text{tel}, \text{pais}, \text{fecha}, \text{cod-v}, \text{h-sal}, \text{h-lleg}, \text{nom-barc}, \text{capacidad}, \text{fecha-inaug}, \text{motivo}\} = U$

$(\text{nro-doc}, \text{cod-v})$ es clave

$(\text{nro-doc}, \text{nom-barc})^+ = \{\text{nro-doc}, \text{nom-pas}, \text{tel}, \text{pais}, \text{capacidad}, \text{fecha-inaug}\}$

Las únicas claves son: $(\text{nro-doc}, \text{fecha})$ y $(\text{nro-doc}, \text{cod-v})$

1 - Cubrimiento minimal:

Paso 1: Todas las dfs deben tener un único atributo del lado derecho.

$F1 = \{$
 $\text{cod-v} \rightarrow \text{fecha},$
 $\text{cod-v} \rightarrow \text{h-sal}$
 $\text{cod-v} \rightarrow \text{h-lleg}$
 $\text{cod-v} \rightarrow \text{nom-barc}$
 $\}$

nro-doc \rightarrow nom-pas
 nro-doc \rightarrow tel
 nro-doc \rightarrow pais
 nom-barc \rightarrow capacidad
 nom-barc \rightarrow fecha-inaug
 cod-v, nro-doc \rightarrow motivo
 fecha, nro-doc \rightarrow cod-v
 }

Paso 2: Se eliminan atributos redundantes

Para analizar si hay atributos redundantes analizo las dfs cod-v, nro-doc \rightarrow motivo y fecha, nro-doc \rightarrow cod-v.

cod-v, nro-doc \rightarrow motivo:

(cod-v)⁺ = {cod-v, fecha, h-sal, h-lleg, nom-barc}
 (nro-doc)⁺ = {nro-doc, nom-pas, tel, pais}

motivo no pertenece a ninguna de las dos clausuras, por lo tanto en cod-v, nro-doc \rightarrow motivo no hay atributos redundantes.

fecha, nro-doc \rightarrow cod-v:

(fecha)⁺ = {fecha}

(nro-doc)⁺ = {nro-doc, nom-pas, tel, pais}

cod-v no pertenece a ninguna de las dos clausuras, por lo tanto en fecha, nro-doc \rightarrow cod-v no hay atributos redundantes.

Paso 3: Se eliminan dfs redundantes

No hay dfs redundantes porque no hay atributos que aparezcan más de una vez del lado derecho de las dfs.

Por lo tanto, F1 es un cubrimiento minimal de F.

2 - Para cada df de la forma X \rightarrow A1, X \rightarrow A2, ..., X \rightarrow An se crea un esquema de la forma Ri(X, A1, A2, ..., An)

R1 (cod-v, fecha, h-sal, h-lleg, nom-barc)

R2 (nro-doc, nom-pas, tel, pais)

R3 (nom-barc, capacidad, fecha-inaug)

R4 (cod-v, nro-doc, motivo)

R5 (fecha, nro-doc, cod-v)

R5 ya incluye las 2 claves, no hay que agregar otro esquema.

d) Decir en qué forma normal se encuentra la descomposición obtenida en la parte anterior. *Nota:* Para esto deberá proyectar cuidadosamente todas las dfs que se cumplen en cada sub-esquema.

Está en 3NF por aplicación del algoritmo.

No está en BCNF porque en **R5** se cumplen las siguientes dfs:

$\prod_{R5}(F) = \{ \text{fecha, nro-doc} \rightarrow \text{cod-v}; \text{cod-v} \rightarrow \text{fecha} \}$

entonces: cod-v no es superclave de R5, ya que $(\text{cod-v})^+ = \{\text{cod-v}, \text{fecha}\}$

e) Enumerar las diferencias entre el resultado obtenido en la **Parte 2 - a)** y el obtenido en la **Parte 2 - d)**, explicando por qué estas suceden.

Diferencia 1 – En la Parte 2 – d el esquema resultado contiene la relación R5, que no se encuentra en la Parte 2 – a). Esto sucede porque en la parte 2 – d) influye una df que se deriva de una RNE del MER, que no se tiene en cuenta en la Parte 2 – a).

Diferencia 2 – En la Parte 2 – d) se preserva la dependencia funcional, mientras que en la Parte 2 – a), se pierde. Esto sucede porque en la Parte 2 – d) se aplica el algoritmo de 3NF que asegura preservación de dependencias.

Diferencia 3 – En la Parte 2 – d) el esquema relacional obtenido está en 3NF, mientras que en la Parte 2 – a) el esquema se encuentra en BCNF (y en 4NF, si consideramos dep. multivaluadas no triviales, que no hay).

Ejercicio 2 (18 puntos)

Una compañía de automóviles posee una base de datos para la gestión del mantenimiento que realizan, con el siguiente esquema:

AUTOS (matricula, marca, modelo, año)

Representa los automóviles registrados en la empresa.

MECANICOS (ci, nombre, teléfono, direccion)

Representa los mecánicos que trabajan en el mantenimiento de los autos.

MANTENIMIENTOS (matricula, ci, fecha, km)

Contiene el registro de mantenimientos realizados a los autos. De cada uno se registra la matrícula del auto, la cédula del mecánico responsable, la fecha del último mantenimiento realizado por ese mecánico a ese auto, y la cantidad de kilómetros recorridos por el auto hasta ese momento.

$$\begin{aligned} \Pi_{matricula}(MANTENIMIENTOS) &\subseteq \Pi_{matricula}(AUTOS) \\ \Pi_{ci}(MANTENIMIENTOS) &\subseteq \Pi_{ci}(MECANICOS) \end{aligned}$$

De cada tabla se conoce la siguiente información:

Relación R	n _R	bf _R	Atributos	Índices
AUTOS A	2000	10	V(marca, A) = 4	Índice primario IndMat, por atributo matricula. Índice secundario IndMarca, por atributo marca.
MECANICOS M	80	20		Índice primario IndCi, por ci
MANTENIMIENTOS T	10000	10	El 40% de los mantenimientos se realizan a marca VW	Índice secundario IndMantCi por ci.
M $\mid \gg \mid_{M.ci = T.ci}$ T		8		

Se sabe que todos los atributos tienen distribución uniforme. Recuerde que $V(A,R)$ es la cantidad de valores distintos que tiene un atributo A en una tabla R, n_R es la cantidad de tuplas de R y bf_R es la cantidad de tuplas de R que entran en un bloque.

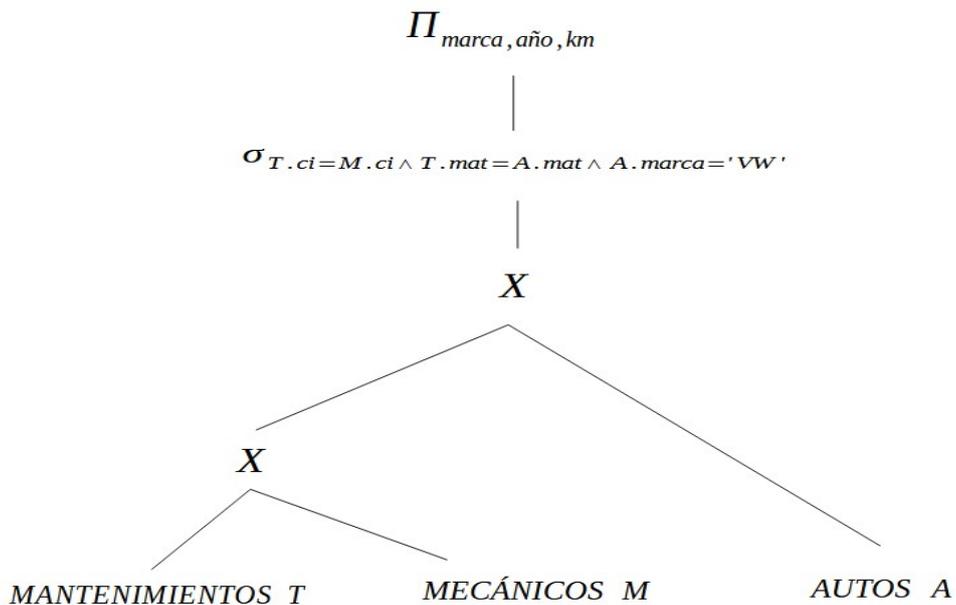
Considere la siguiente consulta sobre el esquema dado:

```

SELECT marca, año, Km
FROM MANTENIMIENTOS T, MECANICOS M, AUTOS A
WHERE T.ci = M.ci AND T.matricula = A.matricula
        AND A.marca = 'VW';
    
```

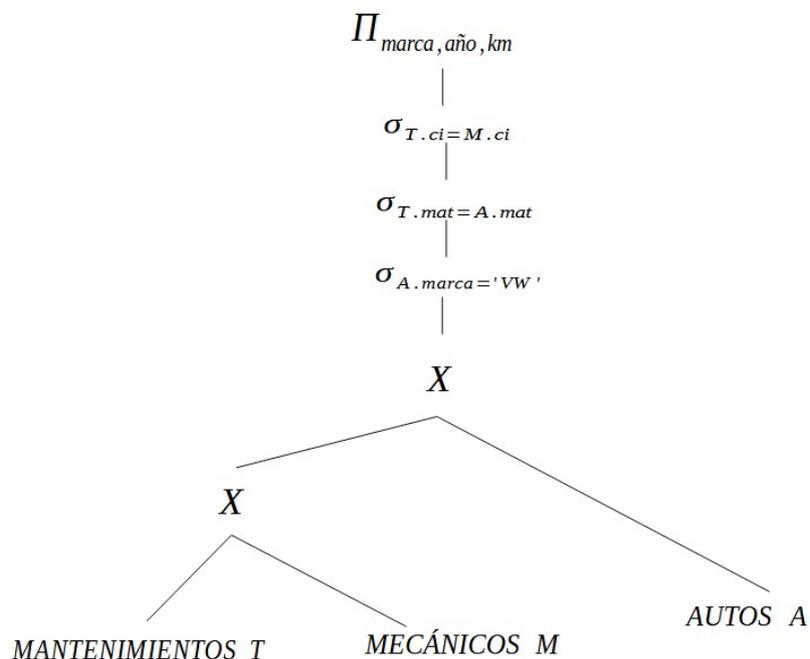
Se pide:

1. Dar el árbol canónico para la consulta.

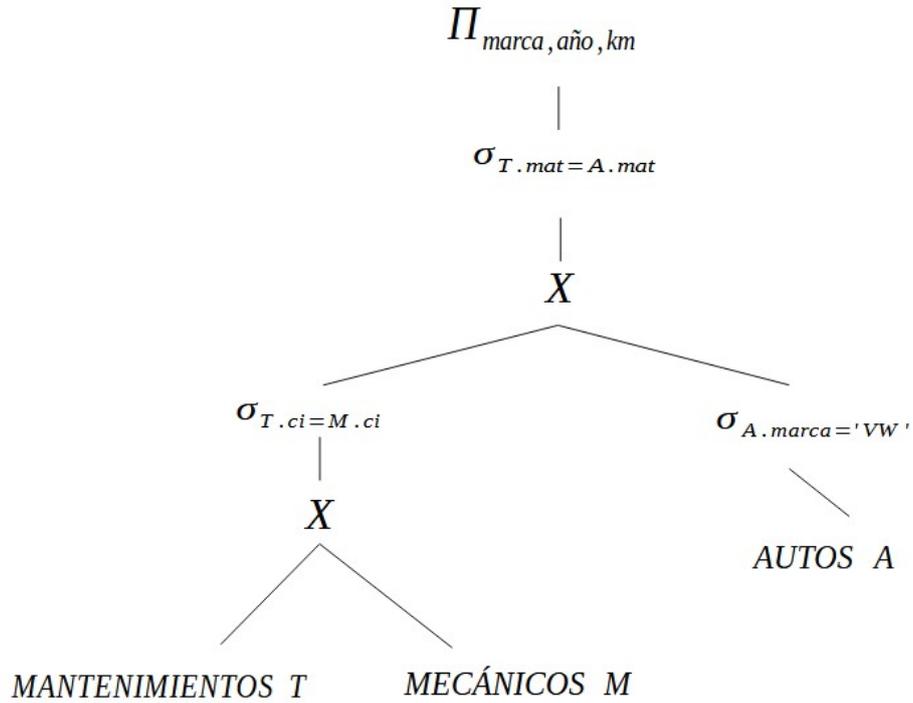


2. Aplicar las heurísticas para optimización llegando al plan lógico optimizado, mostrando cada uno de los pasos ejecutados.

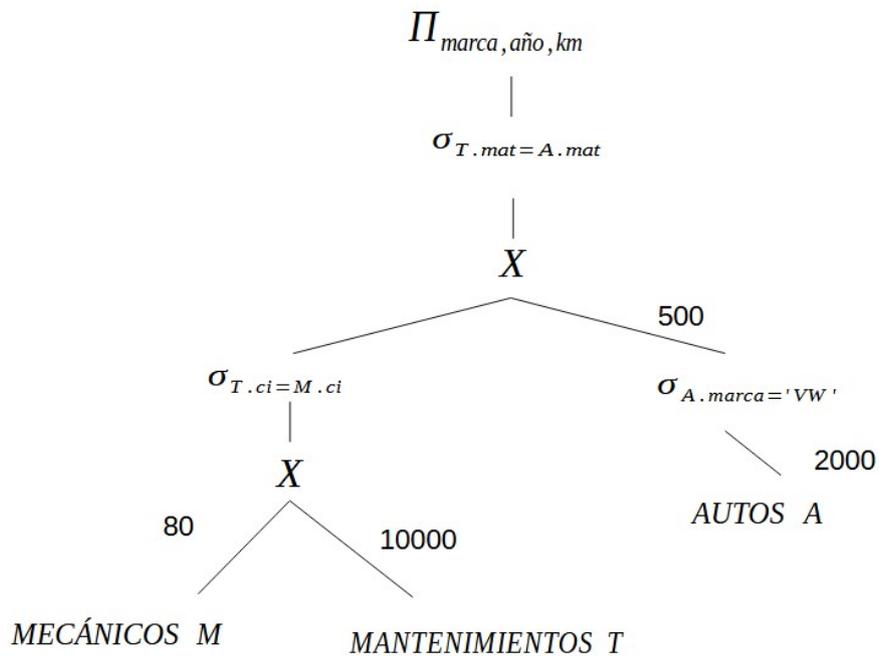
Paso 1: separar las selecciones en cascada de selecciones



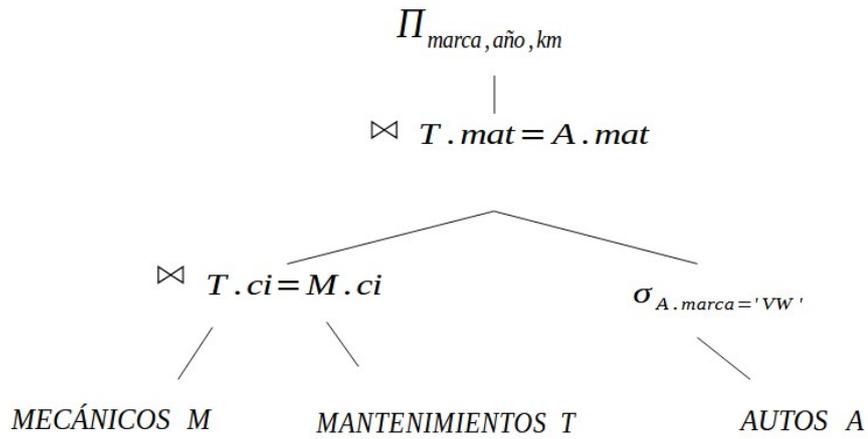
Paso 2: Llevar las selecciones lo más abajo posible



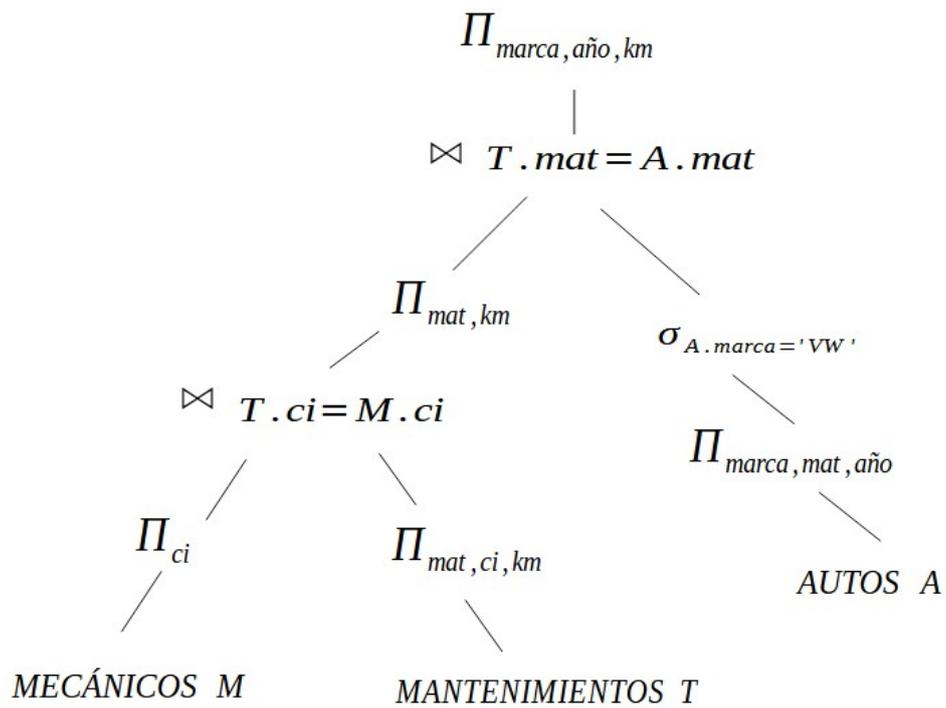
Paso 3: Mover a la izquierda las hojas con menos tuplas, sin generar productos cartesianos innecesarios



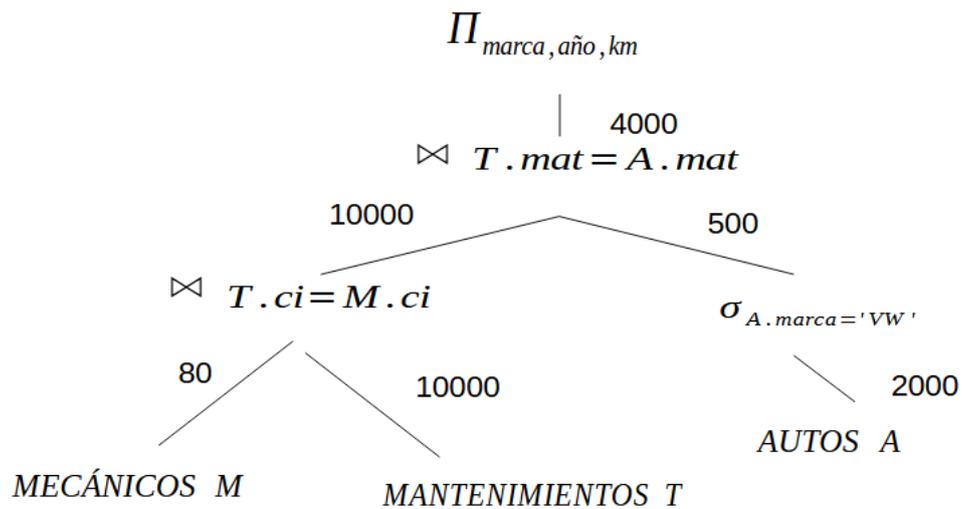
Paso 4: Cambiar productos cartesianos y selecciones por joins



Paso 5: Agregar todas las proyecciones necesarias para manejar la menor cantidad de datos posibles



3. Calcular todos los tamaños intermedios y el tamaño final de la consulta.

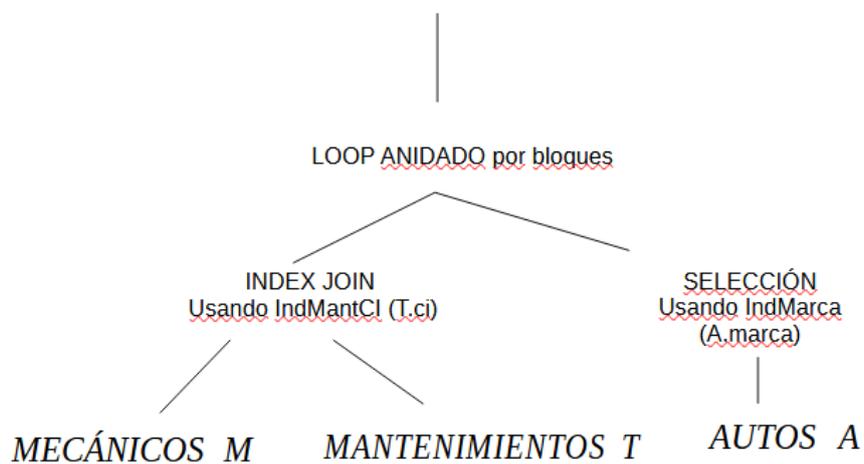


4. Calcular los costos de grabación de todos los resultados intermedios de la consulta.

Costo de grabación de la selección sobre AUTOS: $500/10 = 50$

Costo de grabación del join entre MECÁNICOS y MANTENIMIENTOS: $10000/8 = 1250$

5. Dar un plan físico para el plan lógico obtenido en la **parte 2**, utilizando índices en los casos en que sea posible.



Ejercicio 3 (18 puntos)

Sea una base de datos con 3 items: x, y, z . En ella se cumple la restricción de integridad que la suma de x e y es igual a 100 ($x + y = 100$)

Considere las siguientes transacciones, donde cada una cumple (como se puede verificar) que a partir de un estado consistente termina con un estado consistente con respecto a la restricción.

T1	T2	T3
$\text{tmpX} := r1(x);$ $\text{tmpX} := \text{tmpX} - 15;$ $w1(x, \text{tmpX});$ $\text{tmpY} := r1(y);$ $\text{tmpY} := \text{tmpY} + 15;$ $w1(y, \text{tmpY});$ $c1$	$\text{tmpX} := r2(x);$ $\text{tmpY} := r2(y);$ $\text{tmpZ} := \text{tmpX} + \text{tmpY};$ $w2(z);$ $c2$	$\text{tmpY} := r3(y);$ $\text{tmpY} := \text{tmpY} + 27;$ $w3(y, \text{tmpY});$ $\text{tmpX} := r3(x);$ $\text{tmpX} := \text{tmpX} - 27;$ $w3(x, \text{tmpX});$ $c3$

Se pide:

1. Escribir las transacciones T1, T2 y T3 incluyendo **sólo** las operaciones que se encarga el manejador de transacciones.

Respuesta:

T1: $r1(x), w1(x), r1(y), w1(y), c1$
T2: $r2(x), r2(y), w2(z), c2$
T3: $r3(y), w3(y), r3(x), w3(x), c3$

2. Dar un ejemplo de ejecución (historia) concurrente (entrelazada) completa usando las transacciones dadas en la parte (1) para cada una de las siguientes situaciones:

(a) contenga solamente a T1 y T3 y que resulte en un estado inconsistente con respecto a la restricción,

Respuesta: Una solución posible (Ha)

$\{x = x_0, y = y_0, x_0 + y_0 = 100\}$	
T1	T3
$r1(x) // x_0$	
$w1(x) // x_0 - 15$	
	$r3(y) // y_0$
$r1(y) // y_0$	
	$w3(y) // y_0 + 27$
$w1(y) // y_0 + 15$	
	$r3(x) // x_0 - 15$
	$w3(x) // x_0 - 15 - 27$
$c1$	
	$c3$
$\{x = x_0 - 15 - 27, y = y_0 + 15, x + y = x_0 + y_0 - 27\}$	

(b) contenga solamente a T1 y T2 y que el resultado de T2 no sea el esperado ($z \neq 100$)

Respuesta: Una solución posible (Hb)

{x = x0, y = y0, x0 + y0 = 100}	
T1	T2
r1(x)	
w1(x) // x0 -15	
	r2(x) // x0 -15
	r2(y) // y0
r1(y)	
w1(y) // y0 + 15	
	w2(z) // x0 + y0 - 15
c1	
	c2
{x = x0, y = y0, x + y = 100, z = 85}	

3. Para cada una de las historias dadas en 2(a) y 2(b), indicar si son serializables. Justificar la respuesta. En los casos en que sea serializable indicar el orden serial equivalente (por conflicto) de las transacciones.

Respuesta:

Se contruyen los grafos de serializabilidad de cada historia. La historia es serializable (por conflicto) sii el grafo es aciclico.

- En Ha: hay un arco de T1 a T3 ($w1(x) < r3(x)$) y hay un arco de T3 a T1 ($r3(y) < w1(y)$). Hay ciclo. Por lo tanto, no es serializable.
- En Hb: hay un arco de T1 a T2 ($w1(x) < r2(x)$) y hay un arco de T2 a T1 ($r2(y) < w1(y)$). Hay ciclo. Por lo tanto, no es serializable.

4. Usando las transacciones dadas en la parte (1):

(a) escribir T1 y T2 agregándoles locks y unlocks respetando las reglas de uso de los locks para locks de tipo read/write,

Respuesta:

T1: r1(x), r1(x), w1(x),w1(x), ul1(x), r1(y), r1(y), w1(y), w1(y), ul1(y), c1
T2: r2(x), r2(x), ul2(x), r2(y), r2(y), ul2(y), w2(z), w2(z), ul2(z), c2

(b) dar una historia completa con las T1 y T2 dadas en 4(a) que no sea serializable.

r1(x), r1(x), w1(x),w1(x), ul1(x), r2(x), r2(x), ul2(x), r2(y), r2(y), ul2(y), w2(z), w2(z), ul2(z), c2, r1(y), r1(y), w1(y), w1(y), ul1(y), c1

Para comprobar que esta historia no sea serializable, se contruye el grafo de serializabilidad. Hay ciclo ya que hay un arco de T1 a T2 ($w1(x) < r2(x)$) y hay un arco de T2 a T1 ($r2(y) < w1(y)$)

5. Usando las transacciones dadas en la parte (1):

(a) ¿qué protocolo de locking dado en el curso utilizaría para garantizar que toda ejecución de T1 y T3

resulte en una historia serializable y estricta?

Respuesta: 2PL estricto

(b) escribir T1 y T3 aplicando el protocolo indicado en 5(a),

Respuesta:

T1: rl1(x), r1(x), wl1(x), w1(x), rl1(y), r1(y), wl1(y), w1(y), c1, ul1(x), ul1(y)
T3: rl3(y), r3(y), wl3(y), w3(y), rl3(x), r3(x), wl3(x), w3(x), c3, ul3(y), ul3(x)

(c) dar una historia de las T1 y T3 dadas en 5(b), donde se genere una situación de deadlock.

Respuesta: Una solución posible

T1	T3
rl1(x)	
r1(x)	
	rl3(y)
	r3(y)
wl1(x)	
w1(x)	
	wl3(y)
<i>Al intentar rl1(y), T1 queda bloqueada a la espera de la liberación del lock de escritura sobre 'y' que tiene T3. T1 no puede continuar. En cambio, T3 si puede hacerlo.</i>	
	w3(y)
	<i>Al intentar rl3(x), T3 queda bloqueada a la espera de la liberación del lock de escritura sobre 'x' que tiene T1. T3 no puede continuar y T1 tampoco. Se está en una situación de deadlock</i>