

Examen de Fundamentos de Bases de Datos

Diciembre 2022

Indicaciones Generales:

- La duración de la prueba es de **tres (3)** horas.
- En la prueba **NO** se permite consultar material alguno.
- Empezar cada ejercicio en una hoja nueva.
- Escribir con lápiz y de un solo lado de las hojas.
- Numerar todas las hojas. Incluir en cada hoja la cédula y el nombre. En la primer hoja, incluir la cantidad de hojas que se entregan.

Ejercicio 1. (20 puntos)

Se desea modelar la realidad correspondiente a las pencas del mundial propuestas por los distintos supermercados de Montevideo. En una primera etapa, se propone el modelo de la Figura 1, correspondiente a los pronósticos y resultados de la primera fase del mundial.

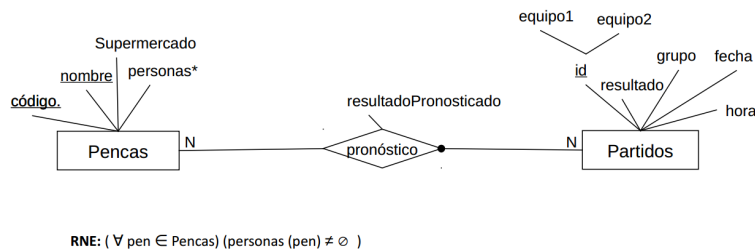


Figura 1: Primera versión del diagrama entidad relación

(a) Indique si las siguientes afirmaciones realizadas sobre el modelo de la Figura 1 son verdaderas o falsas. Justifique su respuesta.

I. Pueden existir instancias con pencas que no tengan personas asociadas.

Solución.

Falso, hay una RNE que indica que cada penca debe tener al menos una persona.

II. En toda instancia todos los partidos son pronosticados por todas las pencas.

Solución.

Falso, no hay totalidad en la relación pronóstico del lado de las pencas, por lo tanto, podría existir alguna instancia en la que haya pencas que no se relacionen con ningún partido. De todas formas esta totalidad no garantiza que todas se relacionen con todos los partidos.

III. En toda instancia las pencas pueden pronosticar más de un resultado para cada partido.

Solución.

Falso, el atributo resultado de la relación pronóstico no es multivaluado.

IV. No existe ninguna instancia en la que la penca no tenga nombre.

Solución.

Verdadero. En todas las instancias todos los atributos deben tener un valor asignado. Los únicos atributos que pueden tener asignado el conjunto vacío son los multivalorados.

- (b) El modelo de la Figura 1 no permite representar información relevante de la realidad, por lo tanto se propone realizar un nuevo modelo considerando la siguiente información:

En primer lugar se registran los datos de los supermercados, su nombre, todos sus teléfonos y la dirección que los identifica (calle, esquina y número son necesarios). A su vez, cada supermercado tiene un conjunto de pencas, a las cuales identifica internamente con un número secuencial. Las pencas pueden ser individuales o grupales. De las personas que participan en las pencas se conoce su cédula de identidad y su e-mail (cualquiera de ellos identifica a las personas). Además, se conoce su nombre, fecha de nacimiento y un teléfono. Por otro lado, se sabe que cada persona puede participar en una sola penca individual y/o en varias pencas grupales de un mismo supermercado. Una penca existe si hay personas que participan en ella. Por otro lado, se conoce información de los equipos que participan en el mundial, código de identificación, país al que representa y grupo que le tocó en el mundial. Además, se modela cada partido, los equipos que participaron, el resultado, la fecha y la hora del mismo. Recordar que los partidos solo son entre equipos del mismo grupo. Finalmente, interesa registrar el pronóstico que se hizo en cada penca a cada partido.

Se pide: Realice el Modelo Entidad-Relación completo de acuerdo a la información antes proporcionada.

Solución.

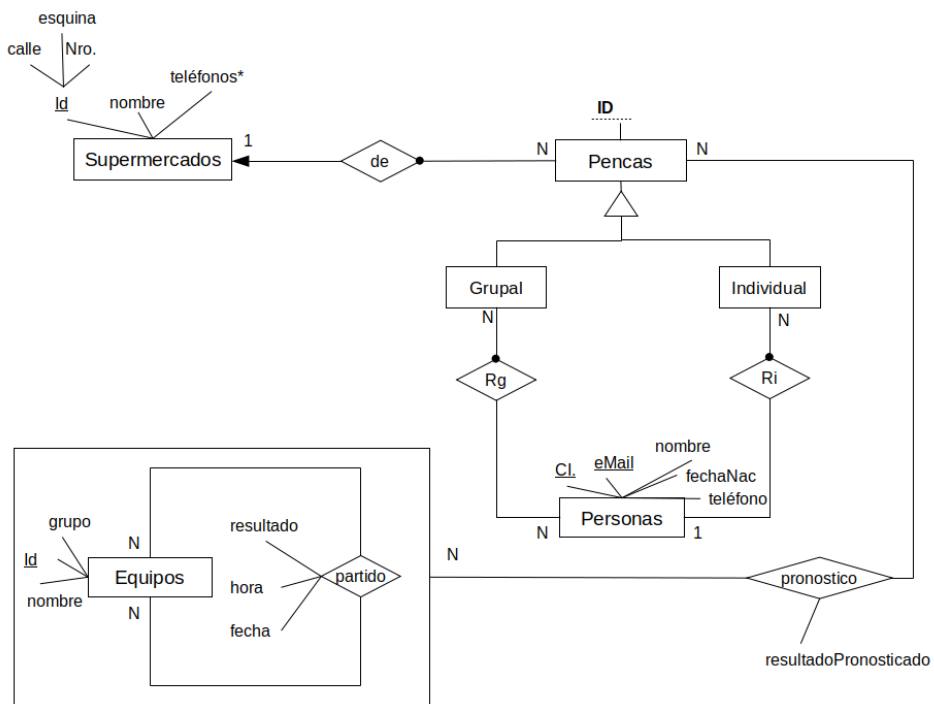


Figura 2: Diagrama entidad relación sol de la parte b

RNEs

- 1—Partido es simétrica e irreflexiva.
- 2— $Grupal \cup Individual = Pencas$
- 3— $Grupal \cap Individual = \emptyset$
- 4—Los partidos son entre equipos diferentes, pero del mismo grupo.
 $(\forall e_1 \in Equipos)(\forall e_2 \in Equipos)(\langle e_1, e_2 \rangle \in partido \rightarrow (e_1 \neq e_2 \wedge e_1.grupo = e_2.grupo))$
- 5—Cada persona sólo puede participar en una penca individual por supermercado
 $(\forall p_1 \in Personas)(\forall pi \in Individual)(\forall s \in Supermercados)$
 $(\langle p_1, s \rangle \in de \wedge \langle p_1, pi \rangle \in Ri \rightarrow \neg(\exists pi_1 \in Individual.$
 $\langle p_1, pi_1 \rangle \in Ri \wedge \langle pi_1, s \rangle \in de \wedge p_i \neq pi_1))$
- 6—Un equipo no puede participar en más de un partido en cierta fecha/hora
 $(\forall e_1 \in Equipos)(\forall e_2 \in Equipos)(\langle e_1, e_2 \rangle \in partido \rightarrow$
 $\neg \exists e_3 \in Equipos(\langle e_1, e_3 \rangle \in partido \vee \langle e_3, e_1 \rangle \in partido \wedge$
 $fecha(e_1, e_3) \neq fecha(e_1, e_2) \vee hora(e_1, e_3) \neq hora(e_1, e_2) \wedge e_2 \neq e_3))$

Ejercicio 2. (30 puntos)

Considere una aerolínea que realiza vuelos directos. Cada vuelo tiene una identificación (idV), un origen (origen) y un destino (destino). Además, se conoce una fecha-hora de salida (fhSal), su piloto (idPil) y su copiloto (idCopil). En varias fechas y horas pueden salir vuelos con la misma identificación que necesariamente tienen el mismo origen y destino. Ej: El vuelo Montevideo-Madrid sale los lunes a las 18:30 y los jueves a las 9:10.

Cada vuelo es realizado por un avión (idAv) y lleva pasajeros y encomiendas de terceros. De los pasajeros se conoce su documento de identidad (docPas) y de las encomiendas un identificador de paquete (idP).

Considere el esquema Universal y los comentarios de la Figura 3:

$$R(idAv, idPil, idCopil, fhSal, idV, origen, destino, docPas, idP)$$

- a. Un avión sólo puede realizar un vuelo en cada fecha-hora de salida.
- b. Una salida de un vuelo sólo puede ser piloteada por un único piloto y un único copiloto.
- c. Hay un único avión que realiza un vuelo determinado en una fecha-hora de salida.
- d. Ninguna persona puede estar en dos aviones al mismo tiempo.
- e. Ningún paquete puede estar en dos aviones al mismo tiempo.
- f. Un vuelo tiene un único origen y un único destino, independientemente de su fecha-hora de salida.

Figura 3: Comentarios sobre los vuelos.

(a) Considere el conjunto F con las siguientes dependencias funcionales:

- 1. $idAv, fhSal \rightarrow idV$
- 2. $idV, fhSal \rightarrow idAv, idPil, idCopil$
- 3. $idPil, fhSal \rightarrow idAv$
- 4. $idCopil, fhSal \rightarrow idAv$
- 5. $docPas, fhSal \rightarrow idAv$
- 6. $idP, fhSal \rightarrow idAv$
- 7. $idV \rightarrow origen, destino$

Se pide: Para cada dependencia de F indique el identificador del comentario que la justifica.

Solución.

1.	a
2.	b,c
3.	d
4.	d
5.	d
6.	e
7.	f

- (b) Identifique al menos una dependencia multivaluada embebida en R . Justifique su respuesta.

Solución.

En una salida de un vuelo, los pasajeros son independientes de los paquetes enviados por terceros. Esto hace que se cumpla la siguiente multivaluada embebida en R_2 de la parte c):

$$idV, fhSal \twoheadrightarrow docPas | idP$$

- (c) Considere la siguiente descomposición de R :

$$\rho = \{ \begin{array}{l} R_1(idAv, idPil, idCopil, fhSal, idV), \\ R_2(idV, fhSal, docPas, idP) \\ R_3(idV, origen, destino) \end{array} \}$$

- I. Determine las proyecciones de F sobre cada uno de los esquemas de ρ . Si corresponde, incluya la dependencia multivaluada embebida como una dependencia multivaluada normal.

Solución.

$$\begin{aligned} \pi_{R_1}(F) &= \{ \\ &\quad idAv, fhSal \rightarrow idV \\ &\quad idV, fhSal \rightarrow idAv, idPil, idCopil \\ &\quad idPil, fhSal \rightarrow idAv \\ &\quad idCopil, fhSal \rightarrow idAv \\ &\quad \} \\ \pi_{R_2}(F) &= \{ \\ &\quad docPas, fhSal \rightarrow idV \\ &\quad idP, fhSal \rightarrow idV \\ &\quad idV, fhSal \twoheadrightarrow docPass \\ &\quad \} \\ \pi_{R_3}(F) &= \{ \\ &\quad idV \twoheadrightarrow origen, destino \\ &\quad \} \end{aligned}$$

- II. Determine todas las claves de R_1 .

Solución.

Como primer paso, se clasifican los atributos según su ubicación en las dependencias funcion-

les:

Nunca a la Derecha – $ND = \{fhSal\}$

Sólo a la Derecha – $SD = \{\}$

A la izquierda y a la Derecha – $ID = \{idAv, idV, idPil, idCopil\}$

Ahora comenzamos a hacer las clausuras de las combinaciones de atributos de $ND \cup ID$ comenzando por el conjunto ND .

$(fhSal)^+ = \{fhSal\}$	
$(fhSal, idAv)^+ = \{fhSal, idAv, idPil, idCopil, idV\}$	Es superclave
$(fhSal, idV)^+ = \{fhSal, idV, idAv, idPil, idCopil\}$	Es superclave
$(fhSal, idPil)^+ = \{fhSal, idPil, idAv, idCopil, idV\}$	Es superclave
$(fhSal, idCopil)^+ = \{fhSal, idCopil, idAv, idPil, idV\}$	Es superclave

Dado que todas las superclaves encontradas no incluyen propiamente a ninguna clave, son todas claves. No hay más claves porque cualquier otra combinación incluiría alguna de las claves.

- III. Indique en que forma normal están cada uno de los esquemas de relación de ρ . Indique también en qué forma normal está ρ . Justifique su respuesta.

Solución.

- R_1 : Dado que todas las dependencias proyectadas tienen alguna superclave del lado izquierdo, está en **4NF**.
- R_2 : Aquí hay que tener particular cuidado porque hay una dependencia multivaluada. Si se observan las funcionales, se ve que $docPas$, idP , y $fhSal$ no aparecen del lado derecho de ninguna de las **dependencias funcionales**. Esto hace que deban estar en todas las claves. Si calculamos su clausura, se obtiene lo siguiente:

$$(docPas, idP, fhSal)^+ = \{docPas, idP, fhSal, idV\}$$

Es por este resultado que $docPas$, idP , y $fhSal$ conforman la única clave posible y que idV es un atributo no primo.

Por este motivo, la dependencia $idP, fhSal \rightarrow idV$ tiene del lado izquierdo una parte de una clave y del lado derecho un atributo no primo por lo que induce que $docPas, idP, fhSal \rightarrow idV$ sea una dependencia parcial que viola 2NF. Por esto, esta tabla está en **1NF**.

- R_3 : En esta tabla hay sólo una dependencia y que es funcional y contiene a todos los atributos. Por esto, la parte izquierda es la clave de la tabla y por lo tanto, está en **4NF**
 - ρ : Por todo lo anterior, la descomposición se encuentra en la forma normal más baja encontrada entre los esquemas de relación que lo componen. Por esto, la descomposición está en **1NF**.
- IV. Aplique un paso del algoritmo para llevar a 4NF sobre R_2 . Indique si se pierde alguna dependencia funcional en ese paso. Justifique su respuesta.

Solución.

La dependencia $docPas, fhSal \rightarrow idV$ viola BCNF y por lo tanto 4NF. Aplicando el algoritmo

por esa dependencia se obtiene la siguiente descomposición:

$$R_{21}(docPas, fhSal, idV)$$

$$\pi_{R_{21}}(F) = \{docPas, fhSal \rightarrow idV\}$$

$$R_{22}(fhSal, docPas, idP)$$

$$\pi_{R_{22}}(F) = \{\}$$

La dependencia funcional que parece perderse es $idP, fhSal \rightarrow idV$. Para verificar si se pierde o no, se debe calcular la clausura de su lado izquierdo contra la unión de las proyecciones anteriores.

$$(idP, fhSal)^+_{\pi_{R_{21}}(F) \cup \pi_{R_{22}}(F)} = \{idP, fhSal\}$$

Este resultado evidencia que **se pierde la dependencia**.

Ejercicio 3. (25 puntos)

Considere un sistema que trabaja con medios de pagos en Uruguay. El siguiente esquema relacional sólo refiere a cuentas bancarias y personas.

CUENTAS_BANCARIAS(id_cuenta, tipo_cuenta, nombre_banco, moneda): Contiene información sobre las cuentas bancarias: el identificador, el tipo de cuenta, el nombre del banco y la moneda de la cuenta. Existen sólo 2 tipos de cuenta bancaria: cuenta corriente o caja de ahorro., que son representados por los valores CC y CA, respectivamente. En cuanto a la moneda, sólo se trabajan con 3: pesos uruguayos, dólares y euros, que son representados con los valores 'UYU, DOLAR y EURO

PERSONAS(id_pers, nombre, pais, email, celular) : Contiene información sobre las personas de todo el sistema: el identificador, el nombre, el nombre del país (que representa a su nacionalidad), el email y un celular.

TIENE(id_pers, id_cuenta, relación) : Contiene información sobre las cuentas bancarias que tienen las personas. El atributo relación representa, por ejemplo, si la persona es titular, titular representativo, ordenatario, entre otras posibilidades.

En este esquema se cumplen las siguientes dependencias de inclusión:

$$\pi_{id_cuenta}(TIENE) \subseteq \pi_{id_cuenta}(CUENTAS_BANCARIAS)$$

$$\pi_{id_pers}(TIENE) \subseteq \pi_{id_pers}(PERSONAS)$$

Considere que:

- Los atributos tipo-cuenta y moneda de la tabla CUENTAS_BANCARIAS tienen distribución uniforme
- El 80% de las personas de la tabla PERSONAS son uruguayas.

Además, de cada tabla se conoce la siguiente información:

R	n _R	Indices (Nombre: Definición)
CUENTAS_BANCARIAS	600	IDXCB_idcuenta: Índice primario sobre id-cuenta IDXCB_tipocuenta: Índice secundario sobre tipo-cuenta IDXCB_moneda: Índice secundario sobre moneda
PERSONAS	800	IDXP_idpers : Índice primario sobre id-pers IDXP_pais : Índice secundario sobre pais
TIENE	900	IDXT_idpers_idcuenta : Índice primario sobre (id-pers, id-cuenta) IDXT_idpers : Índice secundario sobre id-pers IDXT_idcuenta : Índice secundario sobre id-cuenta

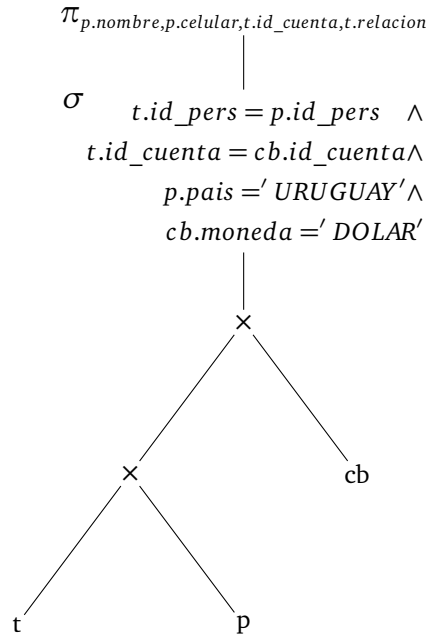
Considere la siguiente consulta sobre el esquema dado:

```

SELECT p.nombre, p.celular, t.id_cuenta, t.relacion
FROM TIENE t, PERSONAS p, CUENTAS_BANCARIAS cb
WHERE t.id_pers = p.id_pers AND t.id_cuenta = cb.id_cuenta
AND p.pais = 'URUGUAY' AND cb.moneda = 'DOLAR';
    
```

(a) Dar el árbol canónico para la consulta.

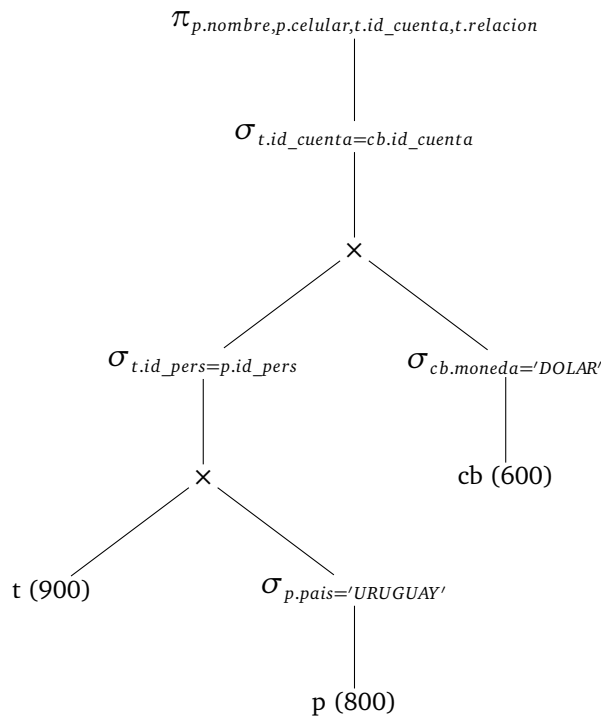
Solución.



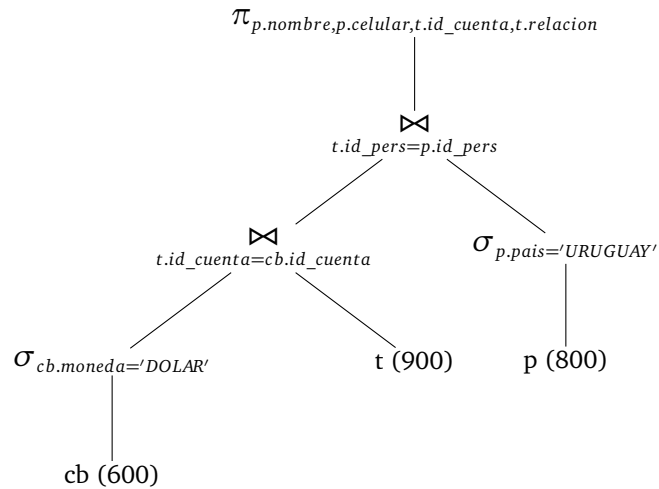
(b) A partir del árbol canónico de la parte a, obtenga el plan lógico optimizado aplicando todas las heurísticas vistas en el curso. Explique cada uno de los pasos ejecutados.

Solución.

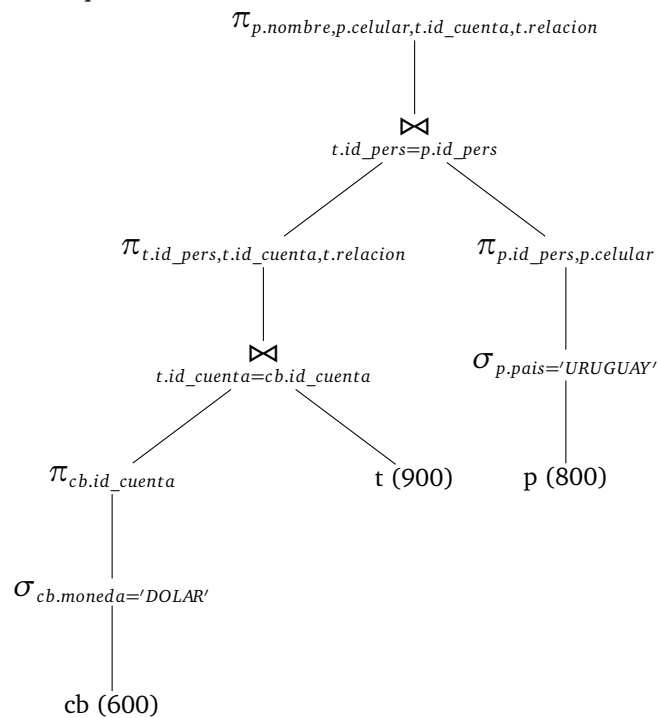
En la primer etapa, se aplican las heurísticas 1 y 2 (cascada de selecciones y bajada de esas selecciones) y se anotan los tamaños de las tablas para simplificar la siguiente etapa. El árbol resultante es el siguiente:



En la segunda etapa, se mueven las tablas con menos tuplas a la izquierda sin generar productos cartesianos. Esto puede cambiar el orden de las selecciones que están en los nodos sobre los productos cartesianos. Aquí también se aplican la transformación de las secuencias de selecciones y productos cartesianos por joins. El árbol resultante es el siguiente:



La última heurística por aplicar es el movimiento de las proyecciones lo más abajo posible. Hay que tener cuidado al aplicarla que no se eliminen índices. El árbol resultante sería el siguiente:



- (c) Dar un plan físico para el plan lógico obtenido en la parte b, utilizando índices en los casos en que sea posible, indicando explícitamente el nombre del índice elegido.

Solución.

Para dar el plan físico, se asocia un algoritmo posible a cada selección y a cada join. Se solicita además, utilizar los índices siempre que sea posible. Las operaciones son las del árbol resultante de la parte anterior.

$\sigma_{p.pais='URUGUAY'}$	Sobre la tabla <i>PERSONAS</i> existe un índice sobre el país. Esto hace viable utilizar una selección por el índice IDXP_pais .
$\sigma_{cb.moneda='DOLAR'}$	Dado que existe un índice sobre la moneda en <i>CUENTAS_BANCARIAS</i> , es posible realizar una selección por el índice IDXCB_moneda .
\bowtie $t.id_cuenta=cb.id_cuenta$	Como existe el índice secundario IDXT_idcuenta , se utiliza un index join para esta operación.
\bowtie $t.id_pers=p.id_pers$	Los índices sobre la tabla <i>PERSONAS</i> fueron descartados cuando se realiza la selección por país. Es por esto que se elige utilizar un join anidado por bloque .

Ejercicio 4. (25 puntos)

Considere el siguiente esquema relacional correspondiente a la base de datos de una tienda de ropa, que incluye información de las prendas, las personas que trabajan en la tienda y las ventas.

PRENDAS(id-prenda, nombre, descripcion, precio)

Esta tabla contiene la información sobre las prendas que se venden en la tienda. Cada registro corresponde a un tipo de prenda, por ej., remera con bolsillo.

STOCK(id-prenda, talle, color, cantidad)

Esta tabla contiene la cantidad de prendas que hay en stock, de cada talle y color. Si la prenda no está disponible en cierto talle y color tiene valor 0 en el atributo cantidad.

VENDEDORES(id-vendedor, nombre, telefono, fecha-ingreso)

Esta tabla contiene los datos de los vendedores de la tienda.

VENTAS(id-prenda, talle, color, id-vendedor, fecha-venta, cantidad)

En esta tabla se registra cada venta. Cada registro tiene los datos de la venta de una o más prendas de un talle y color particulares junto con la cantidad vendida. Por ej., un registro podría corresponder a la venta de 2 remeras con bolsillo de color azul y talle S. Además se guarda el vendedor que hizo la venta y la fecha.

En este esquema se cumplen las siguientes dependencias de inclusión:

$$\Pi_{id-prenda}(STOCK) = \Pi_{id-prenda}(PRENDAS)$$

$$\Pi_{id-vendedor}(VENTAS) \subseteq \Pi_{id-vendedor}(VENDEDORES)$$

$$\Pi_{jid-prenda,talle,color}(VENTAS) \subseteq \Pi_{id-prenda,talle,color}(STOCK)$$

(a) Resolver en Álgebra Relacional las siguientes consultas

1. Devolver el nombre de los vendedores que solo realizaron ventas en Diciembre 2022.

Solución.

Empleados que vendieron en diciembre 2022

$$EDic = \Pi_{id-vendedor}(\sigma_{fecha-venta \geq '01/12/2022' \wedge fecha-venta \leq '31/12/2022'}(VENTAS))$$

Empleados que vendieron en otros momentos

$$EOtras = \Pi_{id-vendedor}(\sigma_{mes(fecha-venta) <> 'Diciembre' \vee \text{anio}(fecha-venta) <> 2022}(VENTAS))$$

Empleados que solo vendieron en diciembre 2022

$$ESOL = EDic - EOtras$$

Solucion

$$S = \Pi_{nombre}(ESOL * VENDEDORES)$$

II. Devolver el nombre y descripción de las prendas que fueron vendidas por todos los vendedores.

Solución.

Empleados y prendas que vendieron

$$VE = \Pi_{id-prenda, id-vendedor}(VENTAS)$$

Empleados

$$E = \Pi_{id-vendedor}(VENDEDORES)$$

Solucion

$$S = \Pi_{nombre, descripcion}(PRENDAS * (VE \% E))$$

(b) Resolver en Cálculo Relacional la siguiente consulta

- I. Devolver el nombre y descripción de las prendas que fueron vendidas exclusivamente por el vendedor con $id-vendedor=23$. Se considera que una prenda es vendida por un vendedor, si existe alguna venta para el vendedor y el identificador de prenda independientemente del talle y color de la misma.

Solución.

$$\{p.nombre, p.descripcion / PRENDAS(p) \wedge (\exists v)(VENTAS(v) \wedge v.id-prenda = p.id-prenda \wedge v.id-vendedor = 23 \wedge \neg(\exists v2)(VENTAS(v2) \wedge v2.id-prenda = v.id-prenda \wedge v2.id-vendedor <> 23))\}$$

(c) Resolver en SQL las siguientes consultas

- I. Devolver el identificador y nombre de las prendas que están disponibles en al menos tres colores diferentes (sin considerar el talle)

Solución.

```
SELECT id-prenda, nombre
FROM stock s NATURAL JOIN prendas p
WHERE s.cantidad > 0
GROUP BY s.id-prenda
HAVING COUNT(DISTINCT s.color) >=3
```

II. Devolver el monto total de las ventas realizadas el 12/12/2022

Solución.

```
SELECT SUM(cantidad * precio) as monto
FROM ventas v NATURAL JOIN prendas p
WHERE v.fecha-venta = '12/12/2022'
GROUP BY v.fecha-venta
```