**Laboratorio de Datos - Guía de ejercicios**

**Normalización**

**Preguntas de repaso.**

a. Comente las anomalías de inserción, borrado y modificación. ¿Por qué están consideradas como malas? Ilustre sus comentarios con ejemplos.

b. ¿Por qué deben evitarse en la medida de lo posible los valores NULL en una relación? Comente el problema de las tuplas espúreas y cómo pueden prevenirse.

c. ¿Qué es una dependencia funcional? ¿Cuáles son las posibles fuentes de información que definen las dependencias funcionales que se cumplen entre los atributos de un esquema de relación?

d. ¿Por qué no podemos inferir automáticamente una dependencia funcional de un estado de relación particular?

e. ¿Qué dependencias no deseables se evitan cuando una relación está en 2FN?

f. ¿Qué dependencias no deseables se evitan cuando una relación está en 3FN?

**Respuestas:**

1. El almacenaje de NATURAL JOINs en un esquema introduce *anomalías de actualización*, que refieren a problemas a la hora de la manipulación del esquema. Éstas se pueden dividir en tres:

* Anomalía de inserción: éstos son problemas que se originan al querer insertar nuevas tuplas en nuestra tabla. Pongamos un ejemplo:

EMPLEADO-DEPARTAMENTO

|  |  |  |  |
| --- | --- | --- | --- |
| CUIL | Nombre | Nro. Depto | Nombre Depto |
| 20-234567-1 | Juan | 5 | Marketing |
| 27-987653-1 | María | 7 | Diseño de Indumentaria |

Si ahora queremos insertar a un nuevo empleado, llamémosle Carla, a quién todavía no se le ha asignado un departamento específico, esto genera el ingreso de valores NULLs a nuestra tabla:

|  |  |  |  |
| --- | --- | --- | --- |
| CUIL | Nombre | Nro. Depto | Nombre Depto |
| 20-234567-1 | Juan | 5 | Marketing |
| 27-987653-1 | María | 7 | Diseño de Indumentaria |
| 27-346372-1 | Carla | NULL | NULL |

Pero eso no es el único problema que surgirá. Supongamos que Carla ha sido asignada al departamento 7, es decir, el de diseño. Se requerirá que el valor a colocar en, por ejemplo Nombre Depto, sea consistente:

|  |  |  |  |
| --- | --- | --- | --- |
| CUIL | Nombre | Nro. Depto | Nombre Depto |
| 20-234567-1 | Juan | 5 | Marketing |
| 27-987653-1 | María | 7 | Diseño de Indumentaria |
| 27-346372-1 | Carla | **7** | Diseño de Ind |

Además, contamos con un problema adicional: no puedo agregar un departamento sin antes asignarle a un empelado!. Es decir, si yo quiero cargar el departamento 3: Administración de personal, no será posible pues no puedo colocar un NULL en la columna CUIL, por lo que es necesario que se cargue a partir de uno de los empleados de ese departamento; o introducir datos falsos/imaginarios para cumplir con el propósito de añadir el departamento a la tabla.

* Anomalía de deleción: éstos son problemas que surgen ante la eliminación de tuplas de la tabla. ¿Qué ocurre si se da el caso de que Juan ya no trabaja en la empresa?

|  |  |  |  |
| --- | --- | --- | --- |
| CUIL | Nombre | Nro. Depto | Nombre Depto |
| 27-987653-1 | María | 7 | Diseño de Indumentaria |
| 27-346372-1 | Carla | **7** | Diseño de Indumentaria |

Desaparecieron los datos del departamento de Marketing! Esto puede resultar peligroso dependiendo de cómo manejemos los datos, pues podría resultarse en pérdidas de información crucial.

* Anomalías de modificación: éstos son problemas que surgen a partir de la modificación de ciertos atributos de las tuplas que se hayan en nuestra tabla. Se puede ver que, como en la anomalía anterior, sucede una pérdida de la información relacionada al departamento de Marketing si a Juan se le asigna un nuevo departamento, supongamos, por ejemplo, el de Diseño de Indumentaria; así como también se podrían introducir inconsistencias como en la inserción:

|  |  |  |  |
| --- | --- | --- | --- |
| CUIL | Nombre | Nro. Depto | Nombre Depto |
| 20-234567-1 | Juan | 7 | Diseño de Ind |
| 27-987653-1 | María | 7 | Diseño de Indumentaria |
| 27-346372-1 | Carla | **7** | Diseño de Indumentaria |

Pero no el único problema que puede surgir. ¿Qué ocurre si ahora el departamento de Diseño de Indumentaria se decide reducirse a departamento de Diseño? Hay que cambiar todas las tuplas que posea el atributo viejo!.

1. El problema de los valores NULLs son diversos:
2. En primer lugar, al habilitarse una celda en nuestra tabla, se genera un espacio de memoria para guardar el dato a cargar. Por lo tanto, si éste es NULL, eso se traduce en un malgasto de memoria.
3. En segundo lugar, se introducen problemas de semántica (lo cual es fundamental a la hora del diseño de BD) pues, ¿qué significa NULL?, ¿Valor no aplicable a la tupla, valor desconocido, valor conocido pero no cargado?
4. En tercer lugar, pueden surgir problemas al querer realizar un JOIN, pues podría no arrojar un mismo resultado un INNER JOIN que un RIGHT OUTER JOIN.
5. En cuarto lugar, habrá que tomar decisiones en cuanto a cómo se interpretarán funciones de agregación (COUNT, SUM, etc.). No es lo mismo sacar un promedio (AVG) tomando valores NULL como 0; que directamente salteándolos.

El problema de las tuplas espúreas es que brindan información incorrecta. Es decir, son tuplas generadas a partir de NATURAL JOINs que no son ciertas/no representan información verídica.

Para evitarlas, se deben diseñar esquemas tal que puedan relacionarse por atributos que sean claves primarias/foráneas; es decir atributos que se encuentran relacionados apropiadamente mediante condiciones de igualdad.

1. Una *dependencia funcional* es una restricción entre dos conjuntos X e Y de una BD, en la que los valores que toman los atributos de Y dependen de los valores que toman los de X.

Esta funciona como una herramienta formal para el análisis de esquemas, que permite detectar y describir problemas descriptos previamente.

Ciertas propiedades que cumple toda relación R son:

1. Si X es clave candidata de R, entonces X Y para todo subconjunto Y de R.
2. Si X es clave candidata de R, entonces X R.
3. No es posible inferir DFs de un estado de relación particular pues para ello se necesita tener conocimiento de la semántica/significado de los atributos, así como las relaciones entre ellos.

Una DF puede existir si existe una instancia r(R) que la cumpla, pero para confirmar su existencia, es necesario conocer la semántica de los atributos de R; sin embargo, para descartar la existencia, bastará mostrar alguna(s) tupla(s) que violan dicha DF.

1. Para que una relación esté en 2FN, se deben evitar las *dependencias funcionales parciales*, es decir, aquellas DFs en las que son posibles eliminar algún atributo A de X y la DF continúa existiendo. Es decir, existen atributos no primos de R que no dependen de todas las claves de R.
2. Para que una relación esté en 3FN (teniendo en cuenta que está en 2FN), se deben evitar las *dependencias funcionales transitivas*, es decir, aquellas DF tal que exista un conjunto de atributos Z en R que no son clave candidata ni algún subconjunto de alguna clave de R tal que X Z y Z Y.

**Ejercicios:**

a. ¿Por qué pueden producirse tuplas espúreas en el resultado de la concatenación de las relaciones EMP\_PROY1 y EMP\_ LOCS?

EMP\_LOCS

|  |  |
| --- | --- |
| NombreE | UbicaciónProyecto |

(En conjunto forman la PK)

EMP\_PROJ1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DNI | NumProyecto | Horas | NombreProyecto | UbicaciónProyecto |

(Las tres primeras columnas forman la PK)

b. Considere las siguientes relaciones para una aplicación de base de datos para el procesamiento de pedidos de la empresa ABC, Inc.

PEDIDO (NúmeroPedido, FechaPedido, NúmeroCliente, CosteTotal) LÍNEA\_PEDIDO(NúmeroPedido, CódigoObjeto, CantidadSolicitada, PrecioTotal, PorcentajeDto)

Asumimos que cada línea de pedido tiene un descuento diferente. El PrecioTotal se refiere a una línea, FechaPedido es la fecha en la que el pedido se realizó y el CosteTotal es el coste total del mismo. Si aplicamos una concatenación natural en la relaciones LÍNEA\_PEDIDO y PEDIDO, ¿qué aspecto tendría el esquema de relación resultante? ¿Cuál sería su clave? Muestre las DF resultantes de esta relación. ¿Está en 2FN? ¿Está en 3FN? Razone sus respuestas.

c. Considere la siguiente relación:

VENTA\_COCHE(NúmeroCoche,FechaVenta, NúmeroVendedor, PorcentajeComisión, Descuento)

Asumimos que un coche lo pueden vender varios vendedores, por lo que {NúmeroCoche, NúmeroVendedor} es la clave principal. Las siguientes son otras dependencias adicionales:

FechaVenta → Descuento

NúmeroVendedor → PorcentajeComisión

Basándonos en la clave principal anterior, ¿está la relación en 1FN, 2FN o 3FN?

Razone su respuesta. ¿Cómo podría normalizarla completamente?

d. Considere la siguiente relación para registrar los libros publicados: LIBRO(TítuloLibro, Autor, TipoLibro, ListaPrecio, AfiliaciónAutor, Editorial)

AfiliaciónAutor hace referencia a la afiliación del autor. Suponga que existen las siguientes dependencias:

TítuloLibro → Editorial, TipoLibro

TipoLibro → ListaPrecio

Autor → AfiliaciónAutor

¿En qué forma normal está la relación? Razone su respuesta.

Aplique una normalización hasta que la relación no pueda descomponerse más. Explique las razones que se esconden tras cada descomposición.

e. Este ejercicio le pide que convierta sentencias de empresa en dependencias. Considere la relación DISCO (NúmeroSerie, Fabricante, Modelo, Versión, Capacidad, Distribuidor). Cada tupla de esta relación contiene información acerca de un disco con un NúmeroSerie único, fabricado por un fabricante, que tiene un modelo particular, distribuido con un número de versión concreto, que tiene una determinada capacidad y que es vendido por un distribuidor concreto. Por ejemplo, la tupla Disco (‘1978619’, ‘WesternDigital’, ‘A2235X’, ‘765234’, 500, ‘CompUSA’) especifica que WesternDigital fabricó un disco con el número de serie 1978619, el modelo A2235X y la versión 765234; su capacidad es de 500GB y lo vende CompUSA. Escriba cada una de las siguientes dependencias como una DF:

DF1. El fabricante y el número de serie identifican el disco de forma única.

DF2. Un modelo está registrado por un fabricante y, por consiguiente, no puede utilizarlo ningún otro.

DF3. Todos los discos de una versión particular son del mismo modelo.

DF4. Todos los discos de un determinado modelo de un fabricante particular tienen exactamente la misma capacidad.

f. Considere la siguiente relación:

VENTA\_COCHE (IdCoche, Extra, PrecioExtra, FechaVenta, PrecioDescontado)

Esta relación hace referencia a los extras instalados en un coche (por ejemplo, el control de velocidad) y que se vendieron a un distribuidor, y la lista y los precios descontados de los extras. Si se cuenta con las siguientes dependencias funcionales:

IdCoche → FechaVenta

Extra → PrecioExtra

IdCoche, Extra → PrecioDescontado

Argumente usando la definición generalizada de la 3FN que esta relación no está en 3FN. A continuación, y desde sus conocimientos de la 2FN, indique por qué tampoco está en 2FN.

**Respuestas:**

1. Al no ser UbicaciónProyecto parte de la clave primaria del esquema EMP\_PROJ1, si éste esquema presenta una misma ubicación más de una vez, lo que generará el NATURAL JOIN es la unión de tuplas que compartan esa ubicación, lo que conlleva que quede un DNI asociado a varios empleados (cuando el DNI debería ser algo único).
2. PEDIDO\_LINEA\_PEDIDO

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Número  Pedido | Código  Objeto | Fecha  Pedido | Número  Cliente | Coste  Total | Cantidad  Solicitada | Precio Total | PorcentajeDto |

La clave será NúmeroPedido, CódigoObjeto.

La relación no está en 2FN pues hay atributo no primo que depende parcialmente de la clave. Por ejemplo, la FechaPedido depende solamente del NúmeroPedido, es decir, que retiro el CódigoObjeto y aún así se mantiene la DF.

Como no está en 2FN, no estará en 3FN.

1. VENTA\_COCHE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NúmeroCoche | NúmeroVendedor | FechaVenta | Porcentaje  Comisión | Descuento |

Está en 1FN pues no hay relaciones dentro de relaciones, ni relaciones como valores de atributos dentro de tuplas; y se cumple que los dominios de los atributos solo incluyen valores atómicos, y cada tupla solo puede tomar un valor del dominio.

Sin embargo no está en 2FN pues se presenta una dependencia funcional parcial de NúmeroVendedor con PorcentajeComisión; luego, no estará en 3FN (además de que se presenta una dependencia transitiva de FechaVenta (que depende funcionalmente de la clave) con Descuento.

Para conseguir que esté en 3FN (y, por lo tanto, en 2FN), propongo el siguiente diseño:

VENTA\_COCHE

|  |  |  |
| --- | --- | --- |
| NúmeroCoche | NúmeroVendedor | FechaVenta |

PORCENTAJE\_VENDEDOR

|  |  |
| --- | --- |
| NúmeroVendedor | PorcentajeComisión |

DESCUENTOS\_DEL\_DIA

|  |  |
| --- | --- |
| FechaVenta | Descuento |

De esta forma consigo, además de lo descripto para 1FN, que no haya dependencias parciales (logrando 2FN) ni transitivas; consiguiendo así la 3FN.

1. LIBRO

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TítuloLibro | Autor | TipoLibro | ListaPrecio | AfiliaciónAutor | Editorial |

(Asumo que {TítuloLibro, Autor} es la clave; y que el TipoLibro es único)

La relación está en 1FN pues no hay relaciones dentro de relaciones y los dominios son atómicos.

Sin embargo, sí existen dependencias funcionales parciales y transitivas. Es por eso que haré descomposiciones para ir alcanzando las siguientes FN.

LIBRO

|  |  |  |  |
| --- | --- | --- | --- |
| TítuloLibro | TipoLibro | ListaPrecio | Editorial |

AUTOR\_DEL\_LIBRO

|  |  |
| --- | --- |
| TítuloLibro | Autor |

AFILIACIÓN\_AUTOR

|  |  |
| --- | --- |
| Autor | Afiliación |

De esta manera, logro alcanzar 2FN pues las descomposiciones consiguieron que todas las dependencias funcionales sean completas. Sin embargo, no está en 3FN pues en LIBRO se presenta una dependencia funcional transitiva de TipoLibro con ListaPrecio. Es por eso que haré una descomposición en esa tabla para evitarla:

LIBRO

|  |  |  |
| --- | --- | --- |
| TítuloLibro | TipoLibro | Editorial |

LISTA\_DE\_PRECIOS

|  |  |
| --- | --- |
| TipoLibro | Precio |

AUTOR\_DEL\_LIBRO

|  |  |
| --- | --- |
| TítuloLibro | Autor |

AFILIACIÓN\_AUTOR

|  |  |
| --- | --- |
| Autor | Afiliación |

1. DISCO\_DF1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NúmeroSerie | Fabricante | Versión | Modelo | Capacidad | Distribuidor |

DISCO\_DF2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NúmeroSerie | Fabricante | Versión | Modelo | Capacidad | Distribuidor |

DISCO\_DF3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NúmeroSerie | Fabricante | Versión | Modelo | Capacidad | Distribuidor |

DISCO\_DF4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NúmeroSerie | Fabricante | Versión | Modelo | Capacidad | Distribuidor |

1. (Respuesta en base a suponer {IdCoche, Extra} como clave primaria):

IdCoche es la clave

En este caso, la relación presenta una dependencia funcional; pues:

{IdCoche, Extra} → PrecioExtra pues la clave determina cada atributo de la relación.

Extra → PrecioExtra por enunciado

Por lo tanto, la relación no está en 2FN, pues es posible eliminar el atributo IdCoche en la DF y aun así ésta se mantiene, es decir, la dependencia es parcial.

Esto implica que la relación no estará en 3FN pues es condición necesaria que esté en 2FN.