

GUÍAS CLÁSICAS DE BASES DE DATOS

1. CONCEPTOS GENERALES, ARQUITECTURA.

1.1. Preguntas de repaso.

- (1) Defina los siguientes términos: datos, base de datos, Sistema administrador de Base de Datos (DBMS), sistema de bases de datos, catálogo de base de datos, vista de usuario, Administrador de base de datos, usuario final, transacción, objeto persistente, metadato.
- (2) Discuta las principales características del enfoque de bases de datos y cómo difiere del sistema tradicional de archivos.
- (3) Discuta las funcionalidades que debiera proveer un DBMS.
- (4) ¿Cuál es la diferencia entre independencia lógica de datos e independencia física de datos?
- (5) Describa la arquitectura de un DBMS y sus principales componentes.
- (6) Mencione aplicaciones recientes de bases de datos que hacen necesaria más funcionalidades que las clásicas.
- (7) Discuta almacenamiento secundario y terciario. ¿Qué los diferencia esencialmente?
- (8) De ejemplos de datos representativos de cada orden de magnitud: KB, MB, GB, TB, PB.

2. DISEÑO

2.1. Preguntas de Repaso.

- (1) Defina modelo de datos, independencia de datos, lenguaje de Definición de datos, lenguaje de consulta.
- (2) Defina entidad, atributo, valor de atributo, instancia de relación, atributo compuesto, atributo multivaluado, atributo derivado, atributo llave, dominio.
- (3) ¿Qué es un rol de participación? ¿Cuándo es necesario usar nombre de roles en la descripción de relaciones?
- (4) ¿Cuándo es útil el concepto de entidad débil en el modelamiento de datos? Defina entidad débil, tipo de relación identificadora, llave parcial.
- (5) Discuta bajo que condiciones una relación ternaria puede ser representada por relaciones binarias.
- (6) Todo conjunto entidad débil puede ser convertido a conjunto fuerte simplemente agregando los atributos apropiados. ¿Por qué, entonces, tenemos conjuntos entidad débiles?

2.2. Ejercicios.

- (1) Construya un modelo de E/R para una compañía de seguros de auto con un conjunto de clientes, cada uno de los cuales es dueño de un número de autos. Cada auto tiene un número de acuerdo al número de accidentes asociados con él.
- (2) Construya un modelo usando diagramas de E/R que pueda responder a las siguientes consultas:
 - (a) ¿Cuántos miembros de la Facultad han sido asignados al departamento de matemáticas? ¿Cómo se llaman? ¿Quién es el encargado del depto. de música? (Nota: matemáticas, música son sólo dos ejemplos; hay más departamentos.)
 - (b) ¿Cuáles son los estudiantes que están mejor en historia? ¿Y en español?
 - (c) ¿Qué miembros de la Facultad están impartiendo cursos de sociología? ¿Qué cursos están impartiendo?
 - (d) ¿Cuántos alumnos están tomando Física 201? ¿En qué sección está Andrea Fernández?
- (3) ¿En qué circunstancias una relación es su propia inversa?

Date: Las originales son de otoño de 2002.

- (4) En este ejercicio y el siguiente consideramos dos opciones de diseño en el modelo E/R para describir nacimientos. En un nacimiento, hay un niño (los gemelos se representan con dos nacimientos), una madre, un número cualquiera de médicos. Suponga, pues, que tenemos los conjuntos entidad Guaguas, Madres, Enfermeras y Doctores. Suponga además que utilizamos una relación Nacimiento, que conecta los cuatro conjuntos entidad. Nótese que una tupla del conjunto relación Nacimiento tiene la forma (Guagua, Madre, Enfermera, Doctor). Si hay más de una enfermera, o más de un médico o de ambos que atienden el parto, entonces habrá varias tuplas con el mismo niño y madre, una para cada combinación de enfermera y médico. Hay ciertas suposiciones que podríamos incorporar al diseño. En cada una indique cómo representar en un diagrama E/R con el fin de representar la suposición.
- Para cada niño, existe una madre única.
 - Para cada combinación de niño, enfermera y médico, hay una madre única.
 - Para cada combinación de niño y madre, existe un médico único.
- (5) Otra forma de resolver el ejercicio anterior es conectar los cuatro conjuntos entidad Guaguas, Madres, Enfermeras y Doctores mediante un conjunto de entidad Nacimiento con cuatro relaciones, una entre él y cada conjunto de entidad descrito. Represente en el modelo de E/R lo siguiente:
- Cada niño es resultado de un nacimiento único y cada nacimiento es de un niño único.
 - Además de (a), cada niño tiene una madre única.
 - Además de (a) y (b), para cada nacimiento hay un médico único.
- (6) Suponga que cambiamos nuestro punto de vista y permitimos que un nacimiento pueda consistir en el nacimiento de más de un niño de una madre. ¿Cómo representaría en el enfoque del problema anterior el hecho de que todos los niños siguen teniendo una madre única?
- (7) Suponga que se desea conservar una genealogía. Tendremos una entidad Personas. La información que deseamos registrar sobre personas incluye su nombre y las siguientes relaciones: madre, padre e hijos.
- Construya un diagrama E/R.
 - Modifique el diagrama par incluir mujeres, varones, personas que no son progenitores.
- (8) Un diagrama E/R puede ser visto como un grafo. ¿Que significa en términos del dominio que se modela que:
- el grafo sea desconexo?,
 - el grafo sea acíclico?

3. EL MODELO RELACIONAL

3.1. Preguntas de Repaso.

- Defina relación, esquema de relación, instancia de relación, atributo de relación, grado de una relación.
- ¿Por que no se permiten tuplas duplicadas en una relación?
- ¿Por que ciertas dependencias funcionales son llamadas triviales?
- Diferencias (si las hay) entre llave candidata, llave, y superllave.
- Define dependencia funcional, clausura de un conjunto de atributos bajo un conjunto de dependencias funcionales, equivalencia de dos conjuntos de dependencias funcionales.

3.2. Ejercicios.

- Convierta cada uno de los diagramas de entidad relación de la guía anterior al modelo relacional.
- Para cada una de las suposiciones sobre el problema de los nacimientos en la guía anterior, indique la llave o las llaves de la relación construída a partir de esta relación.
- Las reglas (llamados axiomas) de Armstrong son los siguientes (X, Y, Z , etc. son conjuntos de atributos).
 - Si $X \rightarrow Y$ entonces $X \rightarrow Y$.
 - Si $X \rightarrow Y$ entonces $XZ \rightarrow YZ$.

- (c) SI $X \rightarrow Y$ e $Y \rightarrow Z$, entonces $X \rightarrow Z$. Demuestre que los axiomas de Armstrong dan reglas correctas.
- (4) Indique cual de las siguientes reglas es correcta y demuéstrela, o en caso contrario muestre que no es correcta mostrando una instancia que no la satisface:
- Si $A \rightarrow B$ y $B \rightarrow C$, entonces $A \rightarrow B$.
 - Si $AB \rightarrow C$ y $BC \rightarrow D$, entonces $AB \rightarrow D$.
 - Si $AB \rightarrow C$, entonces $A \rightarrow C$ ó $B \rightarrow C$.
- (5) Deduzca, usando los Axiomas de Armstrong,
- la regla de partición,
 - la regla de combinación.
- (6) Compute la cerradura del siguiente conjunto de dependencias funcionales para el esquema de relación $R(A, B, C, D, E)$ y $f = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$. Liste las llaves candidatas de R.
- (7) Usando las dependencias funcionales del ejercicio anterior, compute $\{B\}^+$.
- (8) Considere el siguiente par de conjuntos de dependencias funcionales: $f = \{A \rightarrow C, AC \rightarrow D, E \rightarrow AD, E \rightarrow H\}$ y $g = \{A \rightarrow CD, E \rightarrow AH\}$. Determine si son equivalentes o no. Si no, determine si alguno proviene de otro o no.
- (9) Suponga que R es una relación de los atributos A_1, \dots, A_n . En función de n indique cuantas superllaves posee R si:
- La única llave es A_1 .
 - Las únicas llaves son A_1 y A_2 .
 - Las únicas llaves son $\{A_1, A_2\}$ y $\{A_3, A_4\}$.
 - Las únicas llaves son $\{A_1, A_2\}$ y $\{A_1, A_3\}$.
- (10) Sean X e Y dos conjuntos de atributos. Demuestre que si $X \subseteq Y$, entonces $X^+ \subseteq Y^+$.
- (11) El conjunto de atributos X está cerrado (respecto a un conjunto S de dependencias funcionales) si $X = X^+$. Considere el esquema $R(A, B, C, D)$ y un conjunto S de dependencias funcionales. Determine exactamente las dependencias funcionales a partir de la información dada en cada uno de los siguientes casos:
- Los conjuntos de los cuatro atributos están cerrados.
 - Los únicos conjuntos cerrados son $\{\emptyset\}$ y $\{A, B, C, D\}$.
 - Los conjuntos cerrados son $\{\emptyset\}$, $\{A, B\}$ y $\{A, B, C, D\}$.
- (12) Considere el siguiente conjunto S de dependencias funcionales:
 $RUT \rightarrow NombreDep, FechaNac, Dir, NumDep,$
 $NumDep \rightarrow NombreDep, DirDep,$
 $RUT, NumDep \rightarrow DirDep, NumDep.$
 Calcule las clausuras $\{RUT\}^+$ y $\{NumDep\}^+$ respecto de S.
- (13) ¿Es el conjunto de dependencias funcionales S del ejercicio anterior minimal? Si no, intente encontrar un conjunto minimal equivalente. Demuestre que su conjunto es equivalente con S.

4. NORMALIZACIÓN

- (1) Considere el esquema de relación $R(A, B, C, D, E, F, G, H, I, J)$ y el conjunto de dependencias funcionales $AB \rightarrow C, A \rightarrow DE, B \rightarrow F, F \rightarrow GH, D \rightarrow IJ$. ¿Cual es la llave de R? Descomponga R en 2FN, y luego en 3FN.
- (2) Suponga que en el esquema $R(A, B, C, D, E)$ valen las dependencias funcionales $A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A$.
- Demuestre que si descomponemos R en $R'(A, B, C)$ y $R''(A, D, E)$, es posible recuperar toda la información de R.
 - Demuestre que si descomponemos R en $R'(A, B, C)$ y $R''(C, D, E)$ no es posible. (Ayuda: de un ejemplo I de una instancia de R tal que $\pi_{A,B,C}(I) \bowtie \pi_{C,D,E}(I) \neq I$)

- (3) Suponga que descomponemos la relación $R(A, B, C, D)$ en la relación $S(A, B, C)$ y otras relaciones. Indique las dependencias funcionales que se realizan en S si las dependencias son (en cada caso basta ofrecer un conjunto mínimo de dependencias):
- $AB \rightarrow DE, C \rightarrow E, D \rightarrow C$ y $E \rightarrow A$.
 - $A \rightarrow D, BD \rightarrow E, AC \rightarrow E$ y $DE \rightarrow B$.
 - $AB \rightarrow D, AC \rightarrow E, BC \rightarrow D$ y $E \rightarrow B$.
 - $A \rightarrow B, B \rightarrow C, C \rightarrow D$ y $E \rightarrow A$.
- (4) De un ejemplo de un esquema de relación R' y un conjunto F' de dependencias funcionales tales que existen al menos dos descomposiciones (que no pierden información) de R' en FNBC.
- (5) ¿Porque pudiéramos elegir un diseño que no esté en FNBC?
- (6) De un ejemplo de un esquema de relación R y un conjunto de dependencias tales que R este en BCNF pero no en 4NF.
- (7) Suponga que tenemos una relación $R(A, B, C)$ con una dependencia multivaluada $A \twoheadrightarrow B$. Si sabemos que las tuplas $(a, b1, c1)$, $(a, b2, c2)$ y $(a, b3, c3)$ están en la instancia actual de R , ¿Que otras tuplas necesariamente estarán en R ?
- (8) Demuestre que son ciertas o dé contraejemplos para las siguientes reglas de dependencias multivaluadas (X, Y, Z son conjuntos de atributos):
- Si $X \twoheadrightarrow Y$ y $Y \twoheadrightarrow Z$ entonces $X \twoheadrightarrow Z$.
 - Si $X \twoheadrightarrow YZ$ entonces $X \twoheadrightarrow Z$.
 - Si $X \twoheadrightarrow Y$ entonces $X \rightarrow Y$.
 - Si $Y \subseteq X$ entonces $X \twoheadrightarrow Y$ en cualquier relación.
 - Si $X \twoheadrightarrow Y$ y $X \twoheadrightarrow Z$ entonces $X \twoheadrightarrow (Y - Z)$.

5. ALGEBRA RELACIONAL

- (1) Para el esquema:

```

Product(maker, model, type)
PC(model, speed, ram, hd, cd, price)
Laptop(model, speed, ram, hd, screen, price)
Printer(model, color, type, price)

```

escriba en cada caso una expresión relacional que responda a la siguiente consultas:

- Encuentre tamaños de disco duro que ocurren en dos o más computadoras personales.
 - Encuentre el o los fabricantes de la computadora (personal o portátil) de la máxima velocidad disponible actualmente.
 - Encuentre los fabricantes de computadoras personales al menos con tres velocidades distintas.
 - Encuentre los fabricantes que venden exactamente tres modelos diferentes de computadora personal.
- (2) Dibuje los árboles de la consulta para cada una de sus expresiones del ejercicio anterior.
- (3) Suponga que R y S tienen respectivamente m y n tuplas. De los números mínimo y máximo de tuplas que pueden tener los resultados de las siguientes expresiones:
- $R \cup S$.
 - $R \bowtie S$.
 - $\sigma_C(R) \times S$, para alguna condición C .
 - $\pi_L(R) \setminus S$ para alguna lista de atributos L .
- (4) Suponga que las relaciones R, S, T, U tienen respectivamente 100, 1:000 y 10:000 y 50 tuplas cada una. Indique cómo agrupar la siguiente operación para hacerla más eficiente: $R \bowtie S \bowtie T \bowtie U$.
- (5) Defina en términos de los operadores de diferencia, producto, proyección y selección los operadores join natural e intersección.

6. SQL

- (1) Sean R, S relaciones. Para cada una de las expresiones más abajo, de una expresión en SQL equivalente:
 - (a) $\pi_{A,F}(\sigma_{C=D}(R \times S))$
 - (b) $(R - S) \cup (S - R)$
 - (c) $(R \cap S) \neq \emptyset$
- (2) SQL permite especificar llaves. ¿Por qué cree Ud. que SQL no permite la especificación de dependencias funcionales arbitrarias?
- (3) Usando la BD **Product**, **PC**, **Laptop**, **Printer** de un ejercicio anterior, escriba las siguientes consultas en SQL:
 - (a) Las parejas de modelos de computadoras personales (PC) que tienen la misma velocidad y la misma RAM. Deben listarse una sola vez; por ejemplo, liste (i,j), pero no (j,i).
 - (b) Los fabricantes de al menos dos computadoras diferentes (PC o Laptop) con una velocidad mínima de 133, junto con la velocidad mínima y la velocidad máxima de las computadoras que fabrican.
- (4) Muestre cómo expresar en SQL una consulta del álgebra relacional como $\pi_L(\sigma_C(R_1 \bowtie \dots \bowtie R_n))$. (\bowtie indica reunión (join) natural.)
- (5) Escriba la consulta

```
SELECT title FROM movie AS old
WHERE year < ANY (SELECT year
                  FROM Movie
                  WHERE title = Old.title);
```

sin subconsultas.
- (6) Usando la BD **Product**, **PC**, **Laptop**, **Printer** de un ejercicio anterior, escriba las siguientes consultas en SQL:
 - (a) Calcule el precio promedio para cada velocidad de las computadoras personales (PC) de más de 150.
 - (b) Calcule el tamaño promedio del disco duro de un PC para cada fabricante que produzca impresoras.
 - (c) El promedio de los precios de los PC que ofrece cada fabricante.
- (7) Para el ejercicio de **Product**, **PC**, **Laptop**, **Printer**, suponiendo que modelo es llave de **Product**, modelo es llave de **PC**, modelo y pantalla es llave de **Laptop** y modelo y color es llave de **Printer**, modelo en **PC**, **Laptop** y **Printer** referencian a modelo de **Product**.
 - (a) Cree las tablas y sus restricciones correspondientes.
 - (b) Guarde en la BD el hecho de que el modelo 1100 de PC es producido por el fabricante C, tiene una velocidad de 240, RAM 32, disco duro 2.5, un CD de 12x y cuesta 2.499 dólares.
 - (c) Elimine los PC con menos de 2 giga de disco duro.
 - (d) Elimine los laptops producidos por un fabricante que no fabrique impresoras.
 - (e) Para cada laptop producida por el fabricante E, agregue una pulgada al tamaño de la pantalla y reste 100 dólares al precio.
- (8) Proponga todas las llaves adecuadas al ejemplo de **Product**, **PC**, **Laptop**, **Printer**.
- (9) Explique la diferencia entre **DROP R;** y **DELETE FROM R;**.