

UF2176: Definición y manipulación de datos

Elaborado por: Antonio Cala Hurtado

Edición: 5.0

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16492-54-1

No está permitida la reproducción total o parcial de esta obra bajo cualquiera de sus formas gráficas o audiovisuales sin la autorización previa y por escrito de los titulares del depósito legal.

Impreso en España - Printed in Spain

Presentación

Identificación de la Unidad Formativa:

Bienvenido a la Unidad Formativa UF2176: Definición y manipulación de datos. Esta Unidad Formativa pertenece al Módulo Formativo MF0226_3: Programación de bases de datos relacionales que forma parte del Certificado de Profesionalidad IFCD0112: Programación de lenguajes orientados a objetos y bases de datos relacionales, de la familia de Informática y comunicaciones.

Presentación de los contenidos:

La finalidad de esta Unidad Formativa es enseñar al alumno a interpretar las estructuras de datos y el diseño de la base de datos, de forma que pueda realizar correctamente las tareas de programación encomendadas, así como manipular el contenido de bases de datos relacionales de forma interactiva y programar módulos de manipulación de la base de datos.

Para ello, se analizarán los lenguajes relacionales y el lenguaje de manipulación de la base de datos.

Objetivos de la Unidad Formativa:

Al finalizar esta Unidad Formativa aprenderás a:

- Formular consultas de manipulación y definición de datos, a partir del diseño de la BBDD y de los requisitos de usuario.

Índice

UD1. Lenguajes relacionales	9
1.1. Tipos de lenguajes relacionales	11
1.2. Operaciones en el modelo relacional	15
1.2.1. Álgebra relacional	18
1.2.1.1. Clasificación de operadores.....	22
1.2.1.2. Denominación de atributos	24
1.2.1.3. Relaciones derivadas	25
1.2.1.4. Operaciones primitivas: selección, proyec- ción, producto cartesiano, unión y diferencia .	28
1.2.1.5. Otras operaciones: intersección, join, divi- sión, etc.....	37
1.2.2. Cálculo relacional	48
1.2.2.1. Cálculo relacional orientado a dominios	50
1.2.2.2. Cálculo relacional orientado a tuplas.....	52
1.2.2.3. Transformación de consultas entre álgebra y cálculo relacional.....	55
1.3. Lenguajes comerciales: SQL (Structured Query Language), QBE (Query By Example).....	60
1.3.1. Orígenes y evolución del SQL.....	62
1.3.2. Características del SQL.....	64
1.3.3. Sistemas de Gestión de bases de datos con soporte SQL	66

UD2. El lenguaje de manipulación de la base de datos..... 75

2.1.	El lenguaje de definición de datos (DDL)	77
2.1.1.	Tipos de datos del lenguaje	79
2.1.2.	Creación, modificación y borrado de tablas	95
2.1.3.	Creación, modificación y borrado de vistas	116
2.1.4.	Creación, modificación y borrado de índices	122
2.1.5.	Especificación de restricciones de integridad	126
2.2.	El lenguaje de manipulación de datos (DML)	174
2.2.1.	Construcción de consultas de selección: Agregación, Subconsultas, Unión, Intersección, Diferencia	176
2.2.2.	Construcción de consultas de inserción.....	238
2.2.3.	Construcción de consultas de modificación	244
2.2.4.	Construcción de consultas de borrado	250
2.3.	Cláusulas del lenguaje para la agrupación y ordenación de las consultas	254
2.4.	Capacidades aritméticas, lógicas y de comparación del len- guaje.....	273
2.5.	Funciones agregadas del lenguaje.....	300
2.6.	Tratamiento de valores nulos	312
2.7.	Construcción de consultas anidadas	326
2.8.	Unión, intersección y diferencia de consultas	349
2.9.	Consultas de tablas cruzadas.....	361
2.10.	Otras cláusulas del lenguaje	363
2.11.	Extensiones del lenguaje	426
2.11.1.	Creación, manipulación y borrado de vistas	428
2.11.2.	Especificaciones de restricciones de integridad.....	444
2.11.3.	Instrucciones de autorización.....	469
2.11.4.	Control de las transacciones	489
2.12.	El lenguaje de control de datos (DCL)	514
2.12.1.	Transacciones	515
2.12.2.	Propiedades de las transacciones: atomicidad, con- sistencia, aislamiento y permanencia.....	517
2.12.2.1.	Estados de una transacción: activa, parcial- mente comprometida, fallida, abortada y comprometida	518
2.12.2.2.	Consultas y almacenamiento de estructu- ras en XML	519
2.12.2.3.	Estructura del diccionario de datos	540

2.12.3. Control de las transacciones	541
2.12.4. Privilegios: autorizaciones y desautorizaciones	545
2.13. Procesamiento y optimización de consultas	550
2.13.1. Procesamiento de una consulta.....	551
2.13.2. Tipos de optimización: basada en reglas, basada en costes, otros.....	553
2.13.3. Herramientas de la BBDD para la optimización de consultas.....	556
Glosario	565
Soluciones	567
Anexo	569

Área: informática y comunicaciones

UD1

Lenguajes relacionales

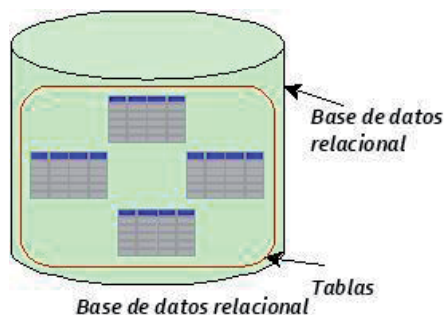
- 1.1. Tipos de lenguajes relacionales
- 1.2. Operaciones en el modelo relacional
 - 1.2.1. Álgebra relacional
 - 1.2.2. Cálculo relacional
- 1.3. Lenguajes comerciales: SQL (Structured Query Language), QBE (Query By Example)
 - 1.3.1. Orígenes y evolución del SQL
 - 1.3.2. Características del SQL
 - 1.3.3. Sistemas de Gestión de bases de datos con soporte SQL

1.1. Tipos de lenguajes relacionales

Una base de datos es un conjunto de datos que están organizados de tal forma que pueden ser consultados, actualizados y gestionados de una forma sencilla por parte del usuario. La organización de una base de datos relacional se basa en lo que se denomina modelo de datos relacional, consistiendo este básicamente en que los datos son almacenados de forma lógica en estructuras de tablas, y la relación entre estos datos están asimismo contenidas en estas tablas y representadas por estos datos.



Una bases de datos relacional es un conjunto de tablas (relaciones), cada tabla está formada por un conjunto de columnas (atributos), cada columna (atributo) puede tomar una serie de valores (dominio), cada fila (tupla) en la tabla contiene los datos que se deseen, los datos de una fila están relacionados.



Base de datos relacional. Imagen propia



Dado los conjuntos C_1, C_2, \dots, C_n , se dice que R es una relación sobre estos conjuntos si R es un conjunto de n -tuplas en donde cada una de estas tuplas el primer elemento pertenece a C_1 , el segundo pertenece a C_2 y así sucesivamente hasta C_n . El conjunto C_i representa el i -ésimo dominio de R , esto es, el conjunto de valores que puede tomar una tupla en el elemento i -ésimo. Cada conjunto C_i de una relación R tiene asociado un nombre de atributo A_i que identifica la columna i -ésima de forma única. Se habla del dominio del atributo A_i . Al número de atributos se le llama cardinalidad u orden de la relación.

Atributo A_1	Atributo A_2	Atributo A_i	
C_1	C_2	C_n	Dominios
e_{11}	e_{12}	e_{1n}	Primera tupla
e_{21}	e_{22}	e_{2n}	
.....	
e_{i1}	e_{i2}	e_{in}	Tupla i -ésima
.....	

Esquema de una relación R

Esquema de una relación R

La expresión $R(A)$ se puede leer como la relación R sobre el conjunto de atributos A , donde $A = \{A_1, \dots, A_n\}$.

Ejemplo

La empresa **El cisne blanco** tiene centros comerciales distribuidos por diferentes ciudades que ofertan todo tipo de productos de vestir, zapatería, complementos, y perfumería. Esta base de datos tiene el nombre de **El cisne**. Se requiere representar los datos relativos al esquema de la relación centro comercial. Esta base de datos tiene el nombre de **el_cisne**.

Para más información, consulta *El cisne blanco* en el anexo al final del libro.

Centro_comercial			
ID_Centro_comercial	Ciudad	Dirección	Nombre

En la relación Centro_comercial se tienen los atributos ID_Centro_comercial que es un número que identifica al centro comercial de forma única, Ciudad donde se localiza el centro comercial, la Dirección del centro comercial y el Nombre del centro comercial.

Añadiendo algunas tupla:

Centro_comercial			
ID_Centro_comercial	Ciudad	Dirección	Nombre
1	Madrid	C/ Goya, 10	Cisne Goya
2	Barcelona	C/ Diagonal, 5	Cisne Diagonal
3	Sierpes	C/ Sierpes, 8	Cisne Sierpes

En el modelo relacional se tiene una vista u organización lógica de los datos que es independiente de como realmente se almacenen estos físicamente. Además, este modelo supone la base para la derivabilidad, redundancia y consistencia de las relaciones:

- Derivabilidad: se dice que una relación es derivable o se puede obtener a partir de la manipulación de otras relaciones.
- Redundancia: se dice que una relación de la base de datos es redundante cuando puede ser obtenida a partir de otras relaciones de la base de datos. En general y en lo posible se intenta evitar la redundancia en el diseño de la base de datos ya que puede llevar a inconsistencias.
- Consistencia: se habla de que el conjunto de relaciones en una base de datos es consistente o inconsistente si cumple o no una serie de restricciones.

**Importante**

El modelo relacional proporciona la base para gestionar una base de datos mediante relaciones (en forma de tablas) que el usuario puede manipular de forma sencilla, independientemente del medio en que los datos sean almacenado físicamente. Se dice que la representación lógica de los datos y su almacenamiento físico son independientes.

El término relación es un concepto matemático que está basado en la teoría de conjuntos. Los operadores usados en el modelo relacional son los operadores que se usan para manipular conjuntos, como pueden ser por ejemplo el operador de unión o el de intersección, y que tienen como operandos y resultados conjuntos de relaciones.

**Sabías que**

El modelo de datos relacional fue ideado por el investigado de IBM Edgar Frank Codd en su artículo “ RelationalModel of Data forLargeShared Data Banks” (Un modelo de datos relacional para grandes bancos de datos compartidos. Nació en Inglaterra en 1923 y murió en EEUU en 2003. En el año 1981 ganó el premio ACM TuringAward por su contribución a la teoría y práctica en la gestión de los sistemas de bases de datos. Como dato curioso durante la segunda guerra mundial interrumpió sus estudios de matemáticas y química para convertirse en piloto de la fuerza aérea británica.

Con la finalidad de llevar a cabo la interacción entre usuario y una base de datos, y reflejar la dinámica en los cambios de estados de los datos, es necesario el uso de lo que se denomina un lenguaje relacional. El lenguaje relacional posibilita la realización de las operaciones de consultas, actualizaciones o gestión de los datos. Existen actualmente dos tipo de lenguajes formales relacionales, el álgebra relacional y el cálculo relacional.

- Álgebra relacional: en este tipo de lenguaje se indica cómo se va a obtener el resultado. Esto se lleva a cabo mediante diversos operadores descritos inicialmente por Codd en su modelo relacional. Los lenguajes comerciales basados en el álgebra relacional se llaman lenguajes procedimentales porque describen que pasos hay que seguir para obtener el resultado deseado.
- Cálculo relacional: en este tipo de lenguaje se indica que resultado se quiere obtener. Los lenguajes basados en el cálculo relacional no indican un procedimiento de cómo obtener el resultado sino que resultado se obtiene, por este motivo a este tipo de lenguajes se les denomina lenguajes declarativos.

Tipos de lenguajes relacionales	<p>Álgebra relacional</p> <p>Cálculo relacional</p>
---------------------------------	---

Basándose en estos dos lenguajes formales existen diversos lenguajes relacionales comerciales como puede ser SQL (StructuredQueryLanguage) o QBE (QueryByExample).

1.2. Operaciones en el modelo relacional

Mediante las operaciones del modelo relacional se puede gestionar el funcionamiento de una base de datos. Estas operaciones manipulan las relaciones de una base de datos, de forma que se pueden actualizar o se puede consultar la base de datos tomando una o varias de estas relaciones, y aplicándoles una o varias operaciones, obteniendo como resultado una nueva relación.

Ejemplo

La empresa **Espacio3i** tiene una red de franquicias inmobiliarias que alquila o vende inmuebles por toda España. En la base de datos central de la red tienen que ser registrados datos relativos a los inmuebles, empleados, inmobiliarias y las ventas realizadas. Esta base de datos tiene el nombre de **espacio3i**.

Esta base de datos tiene las siguientes relaciones:

- **Empleado:** relación donde se recogen los datos de los empleados de Espacio3i.
- **Inmueble:** relación con datos de inmuebles y características.
- **Inmobiliaria:** relación con datos de cada franquicia.
- **Ciudad:** relación donde se registran las ciudades donde trabaja la red de franquicias.
- **Zona:** relación donde se guardan datos de las zonas de cada ciudad.
- **Venta:** relación donde se registran datos de la venta de inmuebles.
- **Cliente:** relación donde se registran datos de los clientes.

Las operaciones básica del modelo relacional son las de: inserción, modificación, borrado y consulta:

- **Inserción:** usada para añadir una ó más tuplas a una relación.

Ejemplo

En la tabla de Empleado de la base de datos espacio3i se inserta dos nuevas tuplas.

Empleado				
ID_Empleado	Nombre	Apellido1	Apellido2	ID_Inmobiliaria
1	Pedro	Pérez	García	1
2	Antonio	Cortés	Galán	1

- **Modificación:** mediante esta operación se pueden cambiar los valores de una o más tuplasasi como uno o más atributos de una relación.

Ejemplo

En la tabla de **Empleado** de la base de datos espacio3i se modifica la última tupla introducida ya que el identificador de la inmobiliaria es erróneo en vez del número 1 es el número 2.

Empleado				
ID_Empleado	Nombre	Apellido1	Apellido2	ID_Inmobiliaria
1	Pedro	Pérez	García	1
2	Antonio	Cortés	Galán	2

- **Borrado:** con esta operación se pueden eliminar una o más tuplas de una relación.

Ejemplo

En la relación **Inmueble** se conservan los inmuebles vendidos o alquilados con una antigüedad máxima de 2 años, de forma que pasado este plazo hay un borrado de las tuplas cuyos inmuebles tengan más de este tiempo. Naturalmente, como fue una venta realizada se borrará asimismo de la relación **Venta**.

- **Consulta:** se puede extraer un conjunto de tuplas a partir de las relaciones y las tuplas ya existente en una base de datos.

Ejemplo

Se quiere consultar los empleados que trabajan en la franquicia de Almería. Esta franquicia tiene como identificador el número 12. Para realizar esta consulta se tendría que contar en la operación de consulta con las relaciones de Empleado, Ciudad e Inmobiliaria.

Empleado				
ID_Empleado	Nombre	Apellido1	Apellido2	ID_Inmobiliaria
31	Gerardo	Sanz	González	12
35	Rogelio	Fernández	Medina	12

1.2.1. Álgebra relacional

Un álgebra es un sistema matemático que está formado por:

- Operandos: son valores que pueden venir dados por constantes o variables.
- Operadores: son operaciones que tienen como argumentos los operandos y como resultado nuevos valores.

Álgebra	Formado por	Operandos Operadores
---------	-------------	-------------------------

Ejemplo

En el álgebra elemental se tienen como operadores las operaciones aritméticas elementales de suma, resta, multiplicación y división. Como operandos se tienen número o variables.

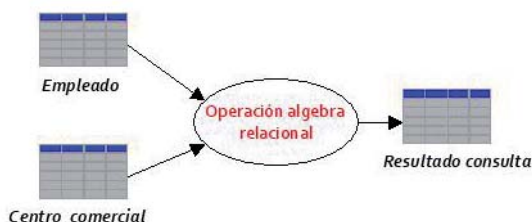
La operación $a+5$ está formado por dos operando (a y 5) y el operador $+$.



Un álgebra relacional tiene como operandos valores que pueden ser relaciones o variables que representan relaciones. Los operadores del álgebra relacional tienen como argumentos relaciones y producen relaciones. Codd añadió el álgebra relacional a su modelo relacional para la manipulación de relaciones.

Ejemplo

En una operación de consulta de una base de datos se tienen una o más relaciones como operandos y como resultado se tiene una nueva relación. En la base de datos el_cisne para obtener los empleados que trabajan en los centros comerciales ubicados en Málaga se requieren las relaciones de Empleado y Centro_comercial.



Operación con relaciones. Imagen propia.

Operación que tiene dos relaciones como operandos, y como resultado se obtiene una nueva relación.

El álgebra relacional se basa en la teoría de conjuntos. Es un lenguaje matemático formal y sobre este se basan la mayoría de los lenguajes de gestión de bases de datos relacionales.



Conceptos básicos de la teoría de conjuntos

Se dice que una serie de elementos pertenecen a un mismo grupo o conjunto, cumpliendo quizás una serie de requisitos comunes. Un conjunto es la reunión de estos elementos.

Si tenemos los elementos a, b, c, y se quiere representar que el conjunto A está formado por estos elementos se haría de la siguiente forma:

$$A=\{a,b,c\}$$

Un operador básico en la teoría de conjunto es el de pertenencia o no a un determinado conjunto.



Se dice que el elemento a pertenece al conjunto A :

$$a \in A$$

Se dice que el elemento z no pertenece al conjunto A :

$$z \notin A$$

Mediante los operadores de intersección y unión se pueden realizar ambas operaciones usando dos conjuntos como operandos.

Si se tienen los conjuntos $A=\{a,b,c\}$ y $B=\{c,e,f\}$, se pueden realizar las operaciones de unión e intersección:

- **Unión:** la operación de unión sobre dos conjuntos, denotado como $A \cup B$, da como resultado otro conjunto formado por todos los elementos de los dos conjuntos.

$$A \cup B = \{a,b,c,e,f\}$$

- **Intersección:** la operación de intersección sobre dos conjuntos, denotado como $A \cap B$, da como resultado otro conjunto formado por los elementos comunes presentes en ambos conjuntos.

$$A \cap B = \{c\}$$

Igualmente son importantes las operaciones de diferencia y producto cartesiano.