

UF1471: Base de datos relacionales y modelado de datos

Elaborado por: María Jesús Guerrero Fernández

Edición: 5.0

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16275-47-2 • Depósito legal: MA 1722-2014

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 UF1471: Bases de datos relacionales y modelado de datos. Esta Unidad Formativa pertenece al Módulo Formativo MF0225_3: Gestión de bases de datos, que forma parte del Certificado de Profesionalidad IFCT0310: Administración de bases de datos, de la familia profesional Informática y comunicaciones.

Presentación de los contenidos

La finalidad de esta unidad formativa es enseñar al alumno a realizar y modificar el diseño físico de las bases de datos a partir del diseño lógico previo, ajustándolo a los requerimientos de explotación de la base de datos. Para ello, se estudiarán las bases de datos relacionales, se analizará el modelo relacional y los elementos que lo integran y se describirá el modelo Entidad-Relación para el modelado de datos. Además, se profundizará en el modelo orientado a objeto y en el modelo distribuido y los enfoques para realizar el diseño.

Objetivos

Al finalizar esta unidad formativa aprenderás a:

- Dominar los fundamentos conceptuales y las técnicas de las bases de datos relacionales.

Índice

UD1. Bases de datos relacionales

1.1. Concepto de base de datos relacional	11
1.2. Ejemplificación de tabla relacional.....	23
1.3. Concepto de modelo de datos. Funciones y Sublenguaje (DDL y DML)	27
1.4. Clasificación de los diferentes tipos de modelos de datos de acuerdo al nivel de abstracción.....	57
1.4.1. Modelos de Datos Conceptuales	59
1.4.2. Modelos de Datos Lógicos	64
1.4.3. Modelos de Datos Físicos.....	73

UD2. Análisis del modelo racional y de los elementos que lo integran

2.1. Concepto de Relaciones y sus propiedades	95
2.2. Concepto de Claves en el modelo relacional	102
2.3. Nociones del álgebra relacional.....	108

2.4. Nociones de Cálculo Relacional de tuplas para poder resolver ejercicios prácticos básicos	123
2.5. Nociones de Cálculo Relacional de dominios	130
2.6. Teoría de la normalización y sus objetivos	135
2.6.1. Concepto de dependencias funcionales.....	138
2.6.2. Análisis y aplicación de las distintas Formas Normales 1ª, 2ª, 3ª, 4ª y 5ª forma normal y la forma normal del Óbice-Codd.....	146
2.6.3. Ventajas e inconvenientes que justifican una desnormalización de las tablas valoración en diferentes supuestos prácticos	166
2.6.4. Desarrollo de diferentes supuestos prácticos de normalización de datos incluyendo propuestas de desnormalización.....	174

UD3. Descripción y aplicaciones del modelo de entidad-relación para el modelado de bases de datos

3.1. Proceso de realización de diagramas de entidad relación y saberlos aplicar	201
3.2. Elementos de.....	213
3.2.1. Entidad	214
3.2.2. Atributo	216
3.2.3. Relación.....	219
3.3. Diagrama entidad relación entendidos como elementos para resolver las carencias de los diagramas entidad-relación simples	223
3.4. Entidades de.....	240
3.4.1. Entidades fuertes y débiles	241
3.4.2. Cardinalidad de las relaciones	244
3.4.3. Atributos en relaciones.....	246
3.4.4. Herencia	247
3.4.5. Agregación.....	248
3.5. Desarrollo de diferentes supuestos prácticos de modelización mediante diagramas de entidad relación.....	249

UD4. Modelo orientado a objeto

4.1. Contextualización del modelo orientado a objeto dentro del modelado UML	267
4.2. Comparación del modelo de clases con el modelo entidad-relación	289
4.3. Diagrama de objetos como caso especial del diagrama de clases	306

UD5. Modelo distribuido y los enfoques para realizar el diseño

5.1. Enumeración de las ventajas e inconvenientes frente a otros modelos	321
5.2. Concepto de fragmentación y sus diferentes tipos.....	332
5.3. Vertical	335
5.4. Horizontal	339
5.5. Mixto	346
5.6. Enumeración de las reglas de corrección de la fragmentación.....	350
5.7. Enumeración de las reglas de distribución de datos.....	352
5.8. Descripción de los esquemas de asignación y replicación de los datos.....	355

Glosario	381
----------------	-----

Soluciones	385
------------------	-----

Anexo	387
-------------	-----

Área: informática y comunicaciones

UD1

Bases de datos
relacionales

- 1.1. Concepto de base de datos relacional
- 1.2. Ejemplificación de tabla relacional
- 1.3. Concepto de modelo de datos. Funciones y Sublenguaje (DDL y DML)
- 1.4. Clasificación de los diferentes tipos de modelos de datos de acuerdo al nivel de abstracción
 - 1.4.1. Modelos de Datos Conceptuales
 - 1.4.2. Modelos de Datos Lógicos
 - 1.4.3. Modelos de Datos Físicos
 - 1.4.3.1. Enumeración de las reglas de Codd para un sistema relacional

1.1. Concepto de base de datos relacional

En primer lugar daremos una serie de definiciones que nos ayudarán más a comprender el concepto de base de datos relacional.



Un **modelo de datos** consiste en un lenguaje cuya finalidad es hablar de una base de datos.

Este tipo de modelo permite describir:

- Las estructuras de datos de la base: Tipos de datos que existen y manera en la que estos están relacionados o “conectados “ entre sí.
- Las restricciones de integridad: Conjunto de condiciones que han de verificar los datos para plasmar adecuadamente la realidad pretendida.
- Operaciones de manipulación de datos: Son operaciones que insertan (agregan o suman) datos, borran o modifican estos, así como operaciones de recuperación de datos de la base.

En un modelo de datos, usualmente las oraciones constan de sujeto y predicado, donde el predicado suele ser la expresión que refleja la acción expresada en la oración.



Definición

- Un **predicado** es una expresión lingüística que puede conectarse con una o varias expresiones.
 - Una **lógica de primer orden o de predicados** es aquella en la que los predicados son considerados como funciones.
-

En matemáticas una función es una expresión en la que a cada valor de un conjunto inicial le asocia un único valor en un conjunto final. A los elementos del primer conjunto se les denomina argumentos y a los elementos del segundo imágenes.

La lógica de primer orden o de predicados estudia la inferencia en los lenguajes de primer orden, los cuales son a su vez, lenguajes formales con cuantificadores que alcanzan sólo a variables de individuo y con predicados o funciones cuyos argumentos son sólo constantes o variables de individuo.

Modelo relacional



Definición

El **modelo relacional** para la gestión de una base de datos consiste en un modelo de datos basado en la lógica de predicados y la teoría de conjuntos.

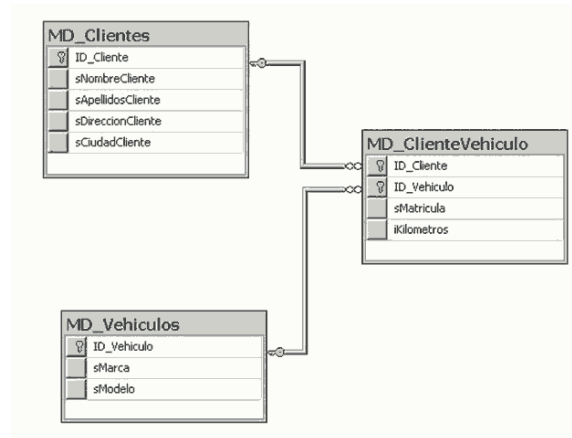


Imagen de base de datos relacional.

Se trata del modelo más utilizado en nuestros días y fundamentalmente es utilizado en el modelado de problemas reales y en la realización de una administración dinámica de los datos.

Este modelo que nace en 1970 de la mano de Edgar Frank Codd en los laboratorios estadounidenses de IBM en San José (California), se expande y consolida rápidamente como un nuevo paradigma en los modelos de bases de datos.

Este modelo gira en torno al uso de las **“relaciones”**. Estas relaciones, a su vez, podrían considerarse en forma lógica como un conjunto de datos llamados **“tuplas”**. Sin embargo en la mayoría de los casos esta teoría de bases de datos creada por Edgar Frank Codd, trata de ser vista de un modo más sencillo y se concibe pensando en cada relación como una tabla compuesta por registros (cada fila de la tabla sería un registro o tupla) y columnas (conocidas como **campos**).

En un ordenador la información puede estar almacenada de distintas formas. La forma en que esta se almacena da lugar a distintos modelos de organización de bases de datos. Así hablamos de modelos jerárquicos, de red, relacional y orientado a objetos.

En el modelo relacional para la gestión de una base de datos, los datos son almacenados en relaciones y puesto que cada relación es un conjunto de datos, como en todo conjunto, el orden en que éstos se almacenan no “im-

porta". Esto, constituye una diferencia con otro tipo de modelos como el jerárquico o de red. La falta de relevancia del almacenamiento de los datos implica que es más fácil de entender y utilizar, lo cual a su vez es una ventaja para usuarios no expertos.

La base de datos relacional está constituida por los siguientes elementos:

BASE DE DATOS
Relaciones: base o derivadas.
Restricciones.
Dominio.
Claves.
Procedimientos almacenados (*).

Nota (): Los procedimientos almacenados aunque no son considerados como un elemento, de la base de datos relacional, todas las aplicaciones comerciales los incluyen.*

1. **Relaciones:** Las relaciones donde son almacenados los datos se denominan "relaciones base" y su implementación se conoce como "tabla". Hay otro tipo de relaciones, las "relaciones derivadas" que no almacenan datos pero que son determinadas al aplicar operaciones relacionadas. La implementación de este tipo de relaciones se denomina "vista" o "consulta" y su utilización es interesante, ya que se utilizan para expresar información proceden de diferentes relaciones actuando como si fuesen una sola.
2. **Restricciones:** Las restricciones son limitaciones que obligan al cumplimiento de determinadas condiciones en una base de datos. Estas condiciones o limitaciones pueden venir impuestas bien por los usuarios, o bien ser inherentes a la propia base de datos.

Estas restricciones proporcionan un método para implementar reglas en una base de datos, igualmente determinan qué datos pueden ser almacenados en una tabla. Normalmente se definen expresiones que tienen como resultado un valor booleano, donde se indica si los datos satisfacen o no la restricción.

Aunque las restricciones no son formalmente un elemento del modelo relacional, suelen incluirse ya que juegan el rol de organizar mejor los datos.

3. **Dominio:** Un dominio describe un conjunto de valores posibles para un atributo dado. Puesto que un dominio restringe los valores de un atributo, puede en cierto modo, ser considerado como una restricción.

Desde un punto de vista matemático asignar un dominio a un atributo significa: "cualquier valor de este atributo ha de ser un elemento del conjunto especificado".

Se consideran dominios: los números enteros, las cadenas de texto, fecha, etc.

4. **Claves:** Una tabla puede tener uno o más campos cuyos valores identifiquen de manera univoca cada registro de la tabla; esto es, no pueden existir dos o más registros diferentes cuyos valores en dichos campos sean idénticos. A este conjunto de campos se le denomina **clave única**.

En una tabla puede haber varias claves únicas, y cada una estas recibe el nombre de clave **candidata a clave primaria**.

Se denomina **clave primaria** a una clave única seleccionada entre todas las claves candidatas para definir de manera univoca a todos los demás atributos de la tabla con objeto a determinar qué datos serán relacionados con las demás tablas. Esto se realizará mediante las **claves foráneas**.

Una **clave foránea** es una referencia de una clave en otra tabla. Establece la relación existente entre dos tablas. Las claves foráneas no tienen por qué ser claves únicas de la tabla en la que se encuentren pero sí en la tabla donde estén referenciadas.

Por ejemplo, el código de un departamento puede ser una clave foránea en una tabla de empleados. Esto es, se permite que existan varios empleados en un mismo departamento, pero habrá uno y sólo un departamento para cada clave diferente de departamento en la tabla de departamentos.

Las **claves índices** nacen a raíz de la necesidad de tener un acceso más rápido a los datos. Los índices pueden ser creados bajo cualquier combinación de campos en una tabla. Las consultas que filtran registros a través de estos campos, mediante la utilización de la clave índice, pueden encontrar los registros de una forma no secuencial.

Existen diferentes técnicas para ordenar una base de datos relacional, siendo cada una de estas óptima para una cierta distribución de datos y un determinado tamaño de la relación.

Normalmente, los índices no son considerados una parte de la base de datos, ya que estos son un detalle agregado. Sin embargo, las claves índices han sido desarrolladas por el mismo grupo de programadores que desarrollo el resto de elementos de la base de datos.

5. **Procedimientos almacenados:** Un procedimiento almacenado es un código ejecutable que se asocia y almacena junto con la base de datos. Usualmente estos procedimientos recogen y personalizan determinadas operaciones comunes, como puede ser la inserción de un registro en una tabla, recopilación de información estadística, o encapsular cálculos complejos.

Por consiguiente:

BASE DE DATOS
Constituida por varias tablas o relaciones.
No pueden existir dos tablas con un mismo nombre o registro.
Cada tabla a su vez está formado por un conjunto de registros (filas y columnas).
La relación entre una tabla padre y un hijo se efectúa mediante las claves primarias y ajenas (o foráneas).
Las claves primarias son la clave principal de un registro de una tabla, debiendo estas cumplir con la integridad de datos.
Las claves ajenas se localizan en la tabla hija, tienen el mismo valor que la clave primaria del registro padre. A través de éstas se establecen las relaciones.

En la estructura de una base de datos se organiza fundamentalmente dos marcadas secciones el esquema y las instancias. Esto es:

$$\text{ESQUEMA} + \text{INSTANCIA (DATOS)} = \text{BASE DE DATOS}$$



Definición

Un **esquema** determina la identidad de la relación y el tipo de información que podría ser almacenada en ella; en otras palabras un esquema contiene los metadatos de la relación.

Los elementos que integran un esquema son:

ESQUEMA
Nombre de la tabla.
Nombre de la columna.
El tipo de dato de cada columna.
Tabla a la que pertenece cada columna.



Definición

Una **instancia** es el resultado de aplicar un esquema a un conjunto finito de datos.

En palabras menos técnicas, una instancia puede ser definida como el contenido de una tabla en un momento dado, aunque también se puede utilizar este término cuando trabajamos o mostramos únicamente un subconjunto de información contenida en una relación o tabla, como por ejemplo:

- Ciertos caracteres y números (una sola columna de una sola fila).
- Algunas o todas las filas con todas o algunas columnas.
- Cada fila es una tupla. El número de filas se llama cardinalidad.
- El número de columnas se llama aridad o grado.

La base de datos relacional, es considerada como una colección de relaciones. De una forma sencilla, una relación es una tabla que a su vez está constituida por un conjunto de fila (tuplas o registros) , donde cada fila es a su vez un conjunto de campos (atributos) y cada campo representa un valor que interpretado describe el mundo real.



Una **base de datos relacional** es un conjunto formado por una o más tablas organizadas en registros (líneas) y campos (columnas) vinculadas entre sí por un campo en común, los cuales en ambos casos (líneas y columnas) poseen las mismas características.

Por ejemplo: nombre de campo, tipo y longitud; a este campo se le denomina ID, identificador o clave.

A esta forma de construir bases de datos se le denomina **modelo relacional**.

Estrictamente hablando el término se refiere a una colección específica de datos, pero a menudo es usado de forma errónea como sinónimo del software usado para gestionar esta colección de datos. Este software se conoce como SGBD (sistema gestor de base de datos) relacional o RDBMS (del inglés relationaldatabasemanagementsystem).

Las bases de datos relacionales pasan por un proceso conocido como normalización.



La **normalización** de una base de datos, se define como aquel proceso, cuyo resultado permite que la base de datos sea utilizada de manera óptima.

La manipulación de la información contenida en estas tablas precisa de la utilización un lenguaje relacional. Hoy día existen dos lenguajes formales, que permiten operar con dicha información: el álgebra relacional que permite describir cómo efectuar una consulta y el cálculo relacional que indica sólo cual sería el resultado de la consulta.

El lenguaje más utilizado en la construcción de las consultas a una base de datos relaciones es SQL (StructuredQueryLanguage). SQL es un lenguaje estándar implementado por los principales sistemas de gestión de datos relacionales integrados.



Definición

El **SGBD** (Sistema de Gestión de Bases de Datos) es un software dedicado exclusivamente a tratar con bases de datos relacionales.

Algunos de los gestores o manejadores actuales más populares son:

MySQL	Oracle	Microsoft SQL Server
DB2	Interbase	Sybase

Entre las **ventajas del modelo relacional**:

1. Se garantizan herramientas que **evitan la duplicidad de registros**, mediante campos claves también conocidas como llaves.
2. Se **garantiza la integridad referencial**. De este modo al eliminar un registro se eliminan todos los registros relacionados dependientes de este.
3. **Favorece la normalización** ya que hace a la base de datos más comprensible y aplicable.

Por tanto con la normalización se **reducen los costos de almacenamiento y acceso** además de **minimizar** la posible **inconsistencia** de los datos.

Son desventajas de una base de datos relacional:

1. Las bases de datos relacionales **presentan deficiencias con datos gráficos**, multimedia, CAD y sistemas de información geográfica.
2. **No permiten una cómoda utilización los bloques** de datos como tipos de datos.

Las bases de datos orientadas a objetos (BDOO) tienen como objetivo principal satisfacer las deficiencias detectadas en las bases de datos relacionales, pero en ningún caso buscaban ni pretendían sustituirlas.

Desventajas BD Relacional	Ventajas BD Relacional
No permiten una cómoda utilización de los bloques.	Favorecen la normalización.
Presentan deficiencias en el tratamiento de los datos gráficos.	Garantiza la integridad referencial.
	Evita la duplicidad de registros.

Objetivos de las bases de datos

Los objetivos principales de un sistema de base de datos es disminuir los siguientes aspectos:

1. Redundancia.
2. Integridad.
3. Dificultad para el acceso a los datos.
4. Aislamiento de los datos.
5. Anomalías de acceso concurrente.
6. Seguridad e integridad de los datos.