

UF1472: Lenguajes de definición y
modificación de datos SQL

Elaborado por: Ester Chicano Tejada

Edición: 5.0

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16275-49-6 • Depósito legal: MA 1747-2014

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 **UF1472: Lenguajes de definición y modificación de datos SQL**. 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 utilizar los lenguajes de definición y modificación de datos SQL. Para ello, se analizarán los objetos y estructuras de almacenamiento de la información para diferentes SGBD (Sistemas de Gestión de Bases de Datos), así como los lenguajes de definición, manipulación y control y los conceptos de transaccionalidad y concurrencia.

Objetivos

Al finalizar esta unidad formativa aprenderás a:

- Realizar el diseño físico de las Bases de Datos según las especificaciones del diseño.

Área: informática y comunicaciones

Índice

UD1. Análisis de los objetos y estructuras de almacenamiento de la información para diferentes SGBD

1.1. Relación de estos elementos con tablas, vistas e índices	11
1.1.1. Ventajas y desventajas de un SGBD	12
1.1.2. Visión de los datos	14
1.1.3. Relación del almacenamiento de datos con la estructura de un SGBD	17
1.1.4. Medios físicos de almacenamiento	19
1.1.5. Métodos de acceso a la base de datos (BD).....	25
1.1.6. Métodos de organización primaria de los ficheros.....	30
1.1.7. Métodos de organización secundaria (o secuencial) de ficheros	31
1.2. Consecuencias prácticas de seleccionar los diferentes objetos de almacenamiento	47
1.2.1. Consecuencias prácticas de seleccionar índices densos o índices dispersos	51

1.2.2. Consecuencias prácticas de seleccionar índices multinivel o índices de árbol B+ o B	54
1.2.3. Consecuencias prácticas de seleccionar técnicas de asociación estáticas o dinámicas (hashing)	59
1.2.4. Elección entre indexación y asociación para la organización de los archivos de registros	61
1.3. Diferentes métodos de fragmentación de la información en especial para bases de datos distribuidas	62
1.3.1. Tipos de fragmentación	65
1.3.2. Grados de fragmentación	75
1.3.3. Información necesaria para la fragmentación	78
1.3.4. La fragmentación en las bases de datos distribuidas	79

UD2. Lenguaje de definición, manipulación y control

2.1. Conceptos básicos, nociones y estándares	95
2.1.1. Introducción a SQL	98
2.2. Lenguaje de definición de datos (DDL SQL) y aplicación en SGBD actuales	103
2.3. Discriminación de los elementos existentes en el estándar SQL-92 de otros elementos existentes en bases de datos comerciales	105
2.4. Sentencias de creación: CREATE	107
2.4.1. Bases de datos	110
2.4.2. Tablas	111
2.4.3. Vistas	124
2.4.4. Disparadores o triggers	129
2.4.5. Procedimientos	132
2.4.6. Sentencias de modificación: ALTER	138
2.4.6.1. Bases de datos	139
2.4.6.2. Tablas	143
2.4.6.3. Vistas	154

2.4.6.4. Disparadores o triggers	156
2.4.6.5. Procedimientos	159
2.4.7. Sentencias de borrado: DROP, TRUNCATE.....	161
2.4.7.1. Bases de datos.....	162
2.4.7.2. Tablas.....	163
2.4.7.3. Vistas	174
2.4.7.4. Disparadores o triggers	175
2.4.7.5. Procedimientos	177
2.4.8. Lenguaje de manipulación de datos (DML SQL)	178
2.4.9. Consultas de datos: SELECT	183
2.4.10. Inserción de datos: INSERT	231
2.4.11. Modificación de datos: UPDATE	242
2.4.12. Eliminación de datos: DELETE	246
2.4.13. Otros elementos de manipulación de datos	251
2.4.13.1. DO	252
2.4.13.2. REPLACE	254
2.4.13.3. Otros elementos	256
2.4.14. Agregación de conjuntos de datos para consulta:	
JOIN, UNION	265
2.4.15. Subconsultas.....	282
2.4.16. Manipulación del diccionario de datos.....	284
2.5. Nociones sobre el almacenamiento de objetos en las	
bases de datos relacionales	288
2.6. Nociones sobre almacenamiento y recuperación de XML en	
las bases de datos relacionales	292
2.6.1. Introducción del estándar SQL 2006.....	294

UD3. Transaccionalidad y concurrencia

3.1. Conceptos fundamentales.....	305
3.2. Identificación de los problemas de la concurrencia	309
3.3. Actualizaciones perdidas	312
3.4. Lecturas no repetibles	314

3.5. Lecturas ficticias.....	315
3.6. Nociones sobre el control de la concurrencia	317
3.7. Optimista	318
3.8. Pesimista.....	319
3.9. Conocimiento de las propiedades fundamentales de las transacciones	321
3.10.ACID.....	325
3.11.Atomicidad	326
3.12.Consistencia.....	328
3.13.Aislamiento (Isolation).....	330
3.14.Durabilidad.....	331
3.15.Análisis de los niveles de aislamiento	333
3.16.Lectura confirmada.....	336
3.17.Lectura repetible	337
3.18.Lectura serializable	339
3.18.1.Desarrollo de un supuesto práctico en el que se ponga de manifiesto la relación y las implicaciones entre el modelo lógico de acceso y definición de datos y el modelo físico de almacenamiento de los datos..	341
Glosario	351
Soluciones.....	355

UD1

Análisis de los objetos
estructuras de almacenamiento
de la información para
diferentes SGBD

- 1.1. Relación de estos elementos con tablas, vistas e índices
 - 1.1.1. Ventajas y desventajas de un SGBD
 - 1.1.2. Visión de los datos
 - 1.1.3. Relación del almacenamiento de los datos con la estructura de un SGBD
 - 1.1.4. Medios físicos de almacenamiento
 - 1.1.5. Métodos de acceso a la base de datos (BD)
 - 1.1.6. Métodos de organización primaria de los ficheros
 - 1.1.7. Métodos de organización secundaria (o secuencial) de ficheros
- 1.2. Consecuencias prácticas de seleccionar los diferentes objetos de almacenamiento
 - 1.2.1. Consecuencias prácticas de seleccionar índices densos o índices dispersos
 - 1.2.2. Consecuencias prácticas de seleccionar índices multinivel o índices de árbol B+ o B
 - 1.2.3. Consecuencias prácticas de seleccionar técnicas de asociación estáticas o dinámicas (hashing)
 - 1.2.4. Elección entre indexación y asociación para la organización de los archivos de registros
- 1.3. Diferentes métodos de fragmentación de la información en especial para bases de datos distribuidas
 - 1.3.1. Tipos de fragmentación
 - 1.3.2. Grados de fragmentación
 - 1.3.3. Información necesaria para la fragmentación
 - 1.3.4. La fragmentación en las bases de datos distribuidas

1.1. Relación de estos elementos con tablas, vistas e índices

Una base de datos es un conjunto de datos que tienen cierta relación entre sí; datos que se organizan, estructuran y almacenan de un modo sistemático para poder ser utilizados en momentos posteriores.

Las bases de datos están gestionadas por sistemas gestores de bases de datos o SGBD, conjunto de aplicaciones software cuya función principal es el almacenamiento, organización y gestión de los datos contenidos en una base de datos.

Características

Las características principales de los sistemas gestores de bases de datos son:

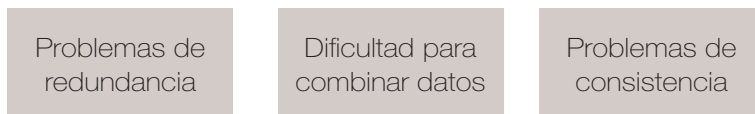
Control de redundancia	Los SGBD permiten al administrador controlar la redundancia de la base de datos.
Control de accesos	Ofrecen la opción de controlar los accesos de los usuarios a la base de datos.
Relaciones complejas	Puede representar relaciones entre datos de carácter complejo.
Restricciones de integridad	El administrador puede establecer restricciones de integridad de la información.

Almacenamiento de datos	Almacena tanto datos simples como complejos.
Datos de gran tamaño	Es capaz de gestionar bases de datos de gran tamaño (incluso base de datos con varios Terabytes).
Utilización conjunta	Permiten compartir bases de datos entre usuarios y aplicaciones.
Copias de seguridad	Dispone de sistemas de copias de seguridad y reparaciones, incluso sin necesidad de suspender la actividad del SGBD.
Interfaces de usuario	Pueden utilizarse con distintas interfaces de usuario como interfaces gráficas, lenguajes de consulta,...
Eficiencia	Destacan por su eficiencia en la gestión de bases de datos.

1.1.1. Ventajas y desventajas de un SGBD

Los sistemas gestores de bases de datos surgieron por la necesidad de tener herramientas que fuesen capaces de almacenar y ordenar la información para poder utilizarla y localizarla con mayor facilidad en momentos posteriores.

Anteriormente se utilizaban sistemas de archivos que ocasionaban numerosos problemas para gestionar datos:



Los SGBD solucionan la gran mayoría de problemas de los sistemas de archivos, pero aun así no son perfectos y ocasionan ciertas desventajas.

En la tabla siguiente se mencionan las ventajas y desventajas de un sistema gestor de bases de datos:

Ventajas y desventajas de un SGBD	
Ventajas	Desventajas
Garantizan una mayor consistencia de los datos aunque haya varios usuarios gestionando la base de datos simultáneamente o surjan errores en el disco, entre otros.	Suelen ser necesarios varios profesionales que administren la base de datos, lo que supone un incremento en costes de personal.
La modificación del código de la aplicación que gestiona la base de datos es mínima cuando que quiere modificar la organización de los datos.	Cuando hay muy pocos datos en la base de datos y su gestión es mínima, es mejor opción la utilización de plantillas de cálculo en vez de SGBD.
Permite una gestión centralizada de la seguridad de los datos.	Ocupan bastante tamaño por su alta complejidad y su elevado número de funcionalidades.
Mayor facilidad y menor tiempo de desarrollo de aplicaciones para los administradores.	Los requisitos de hardware para que un SGBD funcione correctamente suelen ser bastante elevados, lo que implica un coste mayor en recursos hardware.
Permite a los administradores de la base de datos definir estándares vía normalización.	Lentitud de procesamiento en ciertas aplicaciones.
Hay una mayor facilidad para añadir o eliminar datos.	Mayor vulnerabilidad del sistema al estar todos los datos centralizados en un solo gestor.
Inmediatez en la disponibilidad de las actualizaciones del SGBD para todos sus usuarios.	Mayor dificultad de recuperación de datos ante la ocurrencia de incidencias.
Mayor eficiencia, tanto en las operaciones a realizar como en el almacenamiento de los datos.	



Antes de elegir la implantación de un sistema gestor de base de datos será necesario efectuar un análisis de la información que se pretende gestionar. Si son pocos datos y la gestión es mínima, no compensa la implantación de un SGBD por los costes que conlleva, tanto en coste del software, como de los profesionales que son necesarios para su correcta utilización.

1.1.2. Visión de los datos

El objetivo principal de un sistema gestor de base de datos es facilitar una imagen global y abstracta de los datos que contiene, ocultando los detalles sobre su almacenaje y gestión.

Esta visión abstracta que ofrecen los SGBD, puede observarse a tres niveles distintos:

- **Nivel interno o físico:** es el nivel más cercano al almacenamiento físico; muestra una imagen sobre cómo se almacenan los datos dentro de la base de datos. Más concretamente, se muestran detalles de almacenamiento como los archivos que contienen cada información, la organización de los archivos, las formas de acceso a los registros, los tipos de registro, los campos que componen cada registro, su longitud,...
- **Nivel lógico:** muestra toda la estructura de la base de datos y ofrece una visión sobre qué datos están almacenados y qué relaciones hay entre ellos. En este nivel se describen las entidades, atributos, relaciones, restricciones,... sin mostrar los detalles de la estructura física de almacenamiento de datos.
- **Nivel externo o de vistas:** al contrario que el nivel interno, este nivel es el más cercano al usuario. En éste se muestran varios esquemas o vistas de usuarios y resulta de gran utilidad para definir qué información de la base de datos mostrar a cada usuario o grupo de usuarios, pudiendo crear vistas o esquemas diferenciados para cada uno de ellos.

En el siguiente esquema se puede observar la estructura de un SGBD:



Se añade también un nuevo nivel, el nivel conceptual situado entre el usuario y el nivel de vistas. Este nivel ofrece una visión organizada de los datos (una visión no informática), con independencia del SGBD que se utilice.

Los tres niveles de visión (abstracción) descritos anteriormente forman la estructura de un SGBD. El nivel físico es el único que contiene los datos almacenados: el SGBD es el encargado de transformar la petición del usuario a través del nivel de vistas a un esquema conceptual para ser, posteriormente, una petición al esquema interno que se procesará en dicho nivel sobre la base de datos almacenada.

De este modo, el SGBD transforma las peticiones de los usuarios a través de cada nivel y, más concretamente, a través de una serie de pasos:

1. El usuario crea una consulta solicitando una serie de datos de la base de datos.
2. El SGBD analiza, verifica y acepta el esquema externo de la petición del usuario.

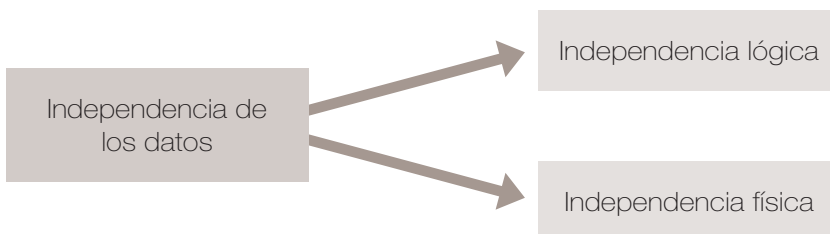
3. A continuación transforma la petición al esquema conceptual.
4. Verifica y acepta el esquema conceptual en el nivel lógico.
5. Transforma la petición al esquema interno.
6. Selecciona las tablas de la petición del usuario y realiza la consulta.
7. Transforma los resultados de la consulta del nivel físico al nivel lógico.
8. Transforma los resultados del nivel lógico al nivel de vistas.
9. El SGBD pone los resultados a disposición del usuario para que pueda visualizarlos.



Aunque pueden coexistir varios esquemas externos (de vistas) para una misma base de datos, sólo puede haber un esquema físico y un esquema lógico.

Independencia de los datos

El SGBD debe garantizar la independencia de los datos en todos sus niveles:



- **Independencia lógica:** el SGBD debe poder modificar el esquema conceptual sin que afecte a los esquemas externos. Por ejemplo, si se elimina un atributo de una base de datos, los esquemas externos que hacen referencia a ellos deben permanecer inalterados.
- **Independencia física:** además, debe ser capaz de modificar el esquema interno sin que se altere ni el esquema conceptual ni los externos mediante la separación entre las aplicaciones y las estructuras de almacenamiento físicas.

1.1.3. Relación del almacenamiento de datos con la estructura de un SGBD

Un sistema gestor de base de datos tiene la capacidad de almacenar elevados volúmenes de datos además de realizar una gestión eficiente de los datos y de su almacenamiento.

La organización física del SGBD está formada por varios elementos:

- **Sistemas software**

Para una gestión y almacenamiento eficaces, es necesario un despliegue de aplicaciones de software que ejecuten las acciones definidas por los usuarios.

- **Sistemas operativos**

Según el sistema operativo que se utilice, el almacenamiento de los datos se realizará de modos diferentes. Hay que tener en cuenta que la tipología de archivos que soporta Linux no es la misma que la que soporta Windows u otros sistemas operativos disponibles.

- **Sistemas de gestión de ficheros**

Los sistemas de gestión de ficheros se encargan de distribuir los datos almacenados para maximizar la eficiencia del sistema y una mayor velocidad de procesamiento del sistema.

- **Controladores de dispositivos**

Los controladores de dispositivos son un conjunto de programas que permiten la interacción del sistema operativo y de los sistemas software con los distintos dispositivos hardware (internos o externos) utilizados en el equipo.



Los controladores de dispositivos (en inglés *drivers*) se pueden definir como un manual de instrucciones que indica al sistema operativo cómo debe comunicarse con un dispositivo hardware concreto. Si no hay un controlador del dispositivo instalado en el sistema operativo, éste no lo reconocerá y no podrá utilizarse.

– Hardware

Se trata de los dispositivos de almacenamiento físico de la información, su estructura y capacidad serán decisivas para la correcta gestión de la información de la base de datos.

Relación elementos de almacenamiento/niveles del sgbd:

El nivel interno de un SGBD correctamente estructurado (con las estructuras de almacenamiento adecuadas) puede ser de gran utilidad para los administradores de la base de datos, ya que facilita mecanismos que le permiten optimizar el almacenamiento de los datos y el acceso a los mismos.

Por otra parte, el nivel externo es un nivel de abstracción sobre el hardware implantado en el sistema operativo. El nivel externo permite el acceso a los dispositivos de almacenamiento a través de peticiones al nivel físico.

1.1.4. Medios físicos de almacenamiento

Los medios físicos de almacenamiento de información fundamentales se representan en la tabla siguiente:

Dispositivos físicos	Características principales
Memoria caché	Es el modo de almacenamiento más rápido y costoso
	Está gestionada por el hardware del sistema.
	Es una forma de almacenamiento volátil
Memoria principal	Destaca por la rapidez en su acceso.
	Suele tener un tamaño demasiado reducido para almacenar una base de datos al completo.
	También es volátil e insegura.
Memoria flash	Su velocidad de lectura es similar a la de la memoria principal pero su velocidad de escritura es bastante lenta.
	Su coste es similar al de la memoria principal.
	Al tener volatilidad mínima aumenta su seguridad
	Es útil sobretodo en volúmenes reducidos de datos.

Discos magnéticos	Suelen almacenar bases de datos enteras y son el medio de almacenamiento de datos más utilizado.
	Permiten el acceso directo a la lectura de los datos.
	Su capacidad de almacenamiento es elevada.
	Es seguro ya que generalmente soportan bien una caída del sistema o algún fallo eléctrico.
Dispositivos de almacenamiento óptico (CDROM, DVD,...)	Lentitud en la lectura y escritura de datos.
	Su utilización se recomienda para bases de datos no volátiles.
Cintas magnéticas	Su coste es bastante reducido.
	Son de alta capacidad.
	Se utilizan para realizar copias de seguridad debido a su no volatilidad



Una **memoria volátil** es aquella que pierde la información que contiene en cuanto se produce alguna interrupción o fallo eléctrico. Por el contrario, la memoria no volátil mantiene los datos ante fallos eléctricos ofreciendo más garantías de seguridad.

De los medios de almacenamiento descritos, son volátiles la memoria principal y la memoria caché. Sin embargo, no son volátiles la memoria flash, los discos magnéticos, los dispositivos de almacenamiento óptico y las cintas magnéticas.