

UF2215: Herramientas de los sistemas
gestores de bases de datos. Pasarelas
y medios de conexión

Elaborado por: José Luis Ávila Jiménez

Edición: 5.1

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16424-48-1

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 **UF2215: Herramientas de los sistemas gestores de bases de datos. Pasarelas y medios de conexión**. Esta Unidad Formativa pertenece al Módulo Formativo **MF0966_3: Consulta y manipulación de información** contenida en gestores de datos que forma parte del Certificado de Profesionalidad **IFCD0211: Sistemas de gestión de información**, de la familia de **Informática y Comunicaciones**.

Presentación de los contenidos

La finalidad de esta unidad formativa es enseñar al alumno a configurar las pasarelas y medios de conexión entre las herramientas cliente y los sistemas gestores de datos, para establecer la comunicación entre ambos según necesidades de la organización.

Para ello, se analizarán los sistemas de gestión de BD, los sistemas de análisis de la información y las herramientas CASE. También se estudiarán los entornos 4GL, el desarrollo de aplicaciones, y se profundizará en las tecnologías de conectividad a BD.

Objetivos del módulo formativo

- Aplicar procedimientos de extracción y consulta de información en el sistema gestor de datos, utilizando herramientas específicas.
- Aplicar los procedimientos de configuración de pasarelas y medios de comunicación para extraer y consultar información de sistemas gestores de datos, según necesidades funcionales.

Índice

UD1. Sistemas de gestión de bases de datos.....	9
1.1. Funcionalidades y objetivo.....	11
1.2. Arquitectura funcional.....	22
1.3. El componente de procesamiento de peticiones	34
1.3.2. Concurrencia	41
1.3.3. Transaccionalidad	43
1.3.4. Procesamiento de consultas.....	47
1.3.4.1. Optimización semántica.....	53
1.3.4.2. Optimización sintáctica	55
1.3.4.3. Optimización física.....	60
1.4. El componente de gestión de los datos	67
1.4.1. Restauración.....	70
1.4.2. Reconstrucción.....	72
1.5. Gestión del rendimiento	74
1.5.1. Plan de consultas.....	78
1.5.2. Monitores	80
1.5.3. Benchmarks	83
1.6. Herramientas	87
1.6.1. Administración	97
1.6.2. Importación/exportación masiva de datos	107
1.6.3. Acceso y consulta de datos	110

UD2. Sistemas de análisis de la información	127
2.1. Minería de datos	129
2.2. OLAP	140
2.3. Dashboard/Scorecard.....	166
2.4. Herramientas	174
2.4.1. Administración	185
2.4.2. Migración.....	190
2.4.3. ETL	200
2.4.4. Información estadística	211
UD3. Herramientas case	221
3.1. Concepto	223
3.2. Historia y evolución	228
3.3. Clasificación	236
3.4. Componentes y funcionalidades de una herramienta CASE.	261
UD4. Entornos 4GL.....	289
4.1. Tipología.....	291
4.2. Componentes y funcionalidades.....	297
4.3. Generadores de informes.....	300
4.4. Generadores de formularios	316
4.5. Generadores de gráficos.....	325
4.6. Generadores de aplicaciones	327
4.7. Nuevas tendencias	339
UD5. Desarrollo de aplicaciones	349
5.1. Ordenes embebidas	351
5.2. Uso de bibliotecas de funciones.....	380
5.3. Diseño de nuevos lenguajes.....	401

UD6. Tecnologías de conectividad a BD	433
6.1. OLEDB.....	435
6.2. ODBC	452
6.3. JDBC	462
6.4. SQLJ	476
6.5. SQL/CLI	485
Glosario	511
Soluciones	513
Anexo	515

UF2215: Herramientas de los sistemas gestores de bases
de datos. Pasarelas y medios de conexión

UD1

Sistemas de
gestión de BD

- 1.1. Funcionalidades y objetivo
- 1.2. Arquitectura funcional
- 1.3. El componente de procesamiento de peticiones
 - 1.3.1. Seguridad
 - 1.3.2. Concurrencia
 - 1.3.3. Transaccionalidad
 - 1.3.4. Procesamiento de consultas
 - 1.3.4.1. Optimización semántica
 - 1.3.4.2. Optimización sintáctica
 - 1.3.4.3. Optimización física
- 1.4. El componente de gestión de los datos
 - 1.4.1. Restauración
 - 1.4.2. Reconstrucción
- 1.5. Gestión del rendimiento
 - 1.5.1. Plan de consultas
 - 1.5.2. Monitores
 - 1.5.3. Benchmarks
- 1.6. Herramientas
 - 1.6.1. Administración
 - 1.6.2. Importación/exportación masiva de datos
 - 1.6.3. Acceso y consulta de datos

1.1. Funcionalidades y objetivo



El concepto de bases de datos surgió en los años 60. Anteriormente solamente se utilizaban ficheros planos para el almacenamiento de la información procesada por un sistema informático.

Los ficheros planos son ficheros generalmente no relacionados entre si, y por tanto para reflejar la interrelación entre la información se ha de utilizar enlaces lógicos o físicos. Pero este sistema conlleva numerosas desventajas; entre ellas podemos citar la dificultad para mantener la integridad de los datos y la gran cantidad de redundancia (datos repetidos) necesaria.

Además los ficheros dependen fuertemente de la aplicación con la que eran tratados, es decir, cada aplicación utiliza sus propios ficheros y esto hace complicado compartir información entre ellas.

Por todos estos motivos, en una etapa temprana del desarrollo de la informática, comenzaron a desarrollarse los sistemas de bases de datos, incorporando, conforme la tecnología ha ido avanzando, las características que en la actualidad poseen.

Una base de datos debe cumplir una serie de características. Para que un sistema pueda ser considerado una base de datos ha de poseer una serie de características para ser considerado como tal:



Independencia de los datos con respecto a los procedimientos: como ya se ha mencionado, uno de los factores que influyó en la génesis de las bases de datos fue la necesidad de mantener una independencia entre los datos y los programas que los utilizan.

Generalmente en las organizaciones los programas de aplicación suelen cambiar cada poco tiempo, sin embargo los datos, y sobre todo la estructura subyacente a esos datos suele permanecer con pocas variaciones durante bastante tiempo. Por ello en épocas tempranas de la informática se vio la necesidad de independizar los datos con respecto a los procedimientos que los manejan.

La independencia de los datos con respecto a los procedimientos también supone que la visión conceptual de dichos datos no ha de ser necesariamente la misma que la visión física, es decir, los datos relativos a un mismo problema han de ser representados conceptualmente de la misma forma independientemente del dispositivo físico que se utilice.



Es por ello que la independencia de los datos con respecto a los procedimientos que la manejan ha de ser satisfecha a dos niveles de abstracción: inde-

pendencia lógica de los datos por la que la modificación de la representación lógica del dominio del problema no debe de afectar a los programas de aplicación que maneja los datos e independencia física por la que la distribución en las unidades de almacenamiento y la estructura física de la información debe de ser independiente de los cambios en la estructura lógica de la información.

Otra característica importante de las bases de datos es la versatilidad para la representación de la información: si bien la estructura de la información es única, diferentes procedimientos pueden hacer un uso distinto de ella.

Por esta razón la organización de la información en las bases de datos debe permitir que diferentes procedimientos puedan construir diferentes registros a partir de la información existente en la base de datos. Estos registros lógicos estarán formados por ítems de datos pertenecientes al dominio del problema que se quiere representar y son derivados de los ítems de datos existentes en dicho dominio. Por supuesto para garantizar la independencia física antes mencionada, estos registros han de ser independientes de los registros físicos que utilice la base de datos para almacenar la información.



Se denomina **redundancia** a la existencia de varias copias de un mismo ítem de datos, las cuales en un momento determinado pueden tener distintos valores.

Además de por el espacio desaprovechado la redundancia es nefasta debido a las posibilidades de inconsistencia en la información almacenada; de hecho uno de los principales motivos por los que surgió la tecnología de las bases de datos fue para evitar la alta redundancia que implica un sistema de archivos planos.

Sin embargo las bases de datos no elimina, ni sería deseable que eliminaran, cualquier redundancia existente en la información perteneciente al dominio de un problema determinado. Esto es debido a que es necesario representar

todas las relaciones existentes entre las entidades que pertenecen a dicho dominio.



Si han de existir diversas copias de un mismo ítem de datos es necesario establecer procedimientos que garanticen la consistencia de la Información cuando dichas copias vayan a ser utilizadas por diversos procedimientos, posiblemente al mismo tiempo.

Por otro lado también es necesario establecer procedimientos para permitir que un mismo ítem de datos pueda ser accedido por varios procedimientos al mismo tiempo.



El concepto de **integridad** hace referencia a la veracidad de los datos almacenados con respecto a la información que realmente caracteriza el dominio del problema en un momento dado.

Ya que los datos van a ser accedidos por diversos procedimientos de diversa índole, es necesario garantizar que dichos datos no sean destruidos ni modificados de forma anómala. Esto también incluye el hecho de que el sistema que maneje la información ha de garantizar la consistencia de esta frente a todo tipo de errores tanto hardware como software que se puedan producir.

Simplicidad: una base de datos debe de estar basada en una representación lógica simple pero potente que permita tanto la representación de la información y sus relaciones de cualquier problema que se pueda plantear como la

fácil verificación del problema que representa, así como las modificaciones necesarias de los requerimientos del mismo de manera que la inclusión de nuevos ítems de datos o nuevas relaciones entre ellos no cause una complejidad excesiva.



Importante

Seguridad y privacidad: la seguridad de una base de datos hace referencia a la capacidad de ésta para proteger los datos contra pérdidas accidentales, causadas por fallos en la infraestructura informática (tanto hardware como software). La privacidad se refiere a la reserva de la información a personas no autorizadas.



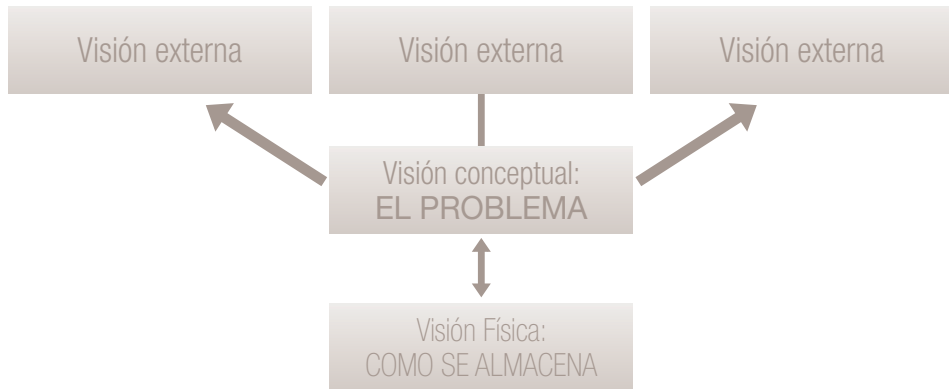
Definición

La denominada “**arquitectura de referencia**” hace referencia a los diferentes niveles de visión de los datos que pueden aparecer en una base de datos.

Una base de datos, dependiendo del tipo de usuario que acceda a los datos debe presentarle una visión de éstos que sea capaz de manejar, interpretar y reconocer. Además la organización física de los datos debe ser lo más independiente posible de los procedimientos que manejen dicha información, así como de los posibles cambios que puedan ir apareciendo en el dominio del problema.

Es por esta razón que en una base de datos deben de poder distinguirse tres tipos de visiones de los datos:

- **Visión externa:** es la visión de los datos que tiene los usuarios de la base de datos. Habrá tantas visiones externas como tipos de usuarios diferentes utilicen una base de datos.
- **Visión conceptual:** es una representación del dominio del problema tal y como este se presenta en el mundo real. Es una representación abstracta e independiente de cómo va a ser tratada dicha información. La visión conceptual no debe experimentar cambios a menos que cambie la naturaleza del problema a representar por la base de datos.
- **Visión física:** es una representación de cómo la información va ser guardada en los dispositivos de almacenamiento.



Estas tres formas de descripción son en realidad tres niveles de abstracción diferentes a los que se describe el mismo problema. La descripción a tres niveles garantiza la independencia de los datos de tal manera que se cumplan las siguientes características:

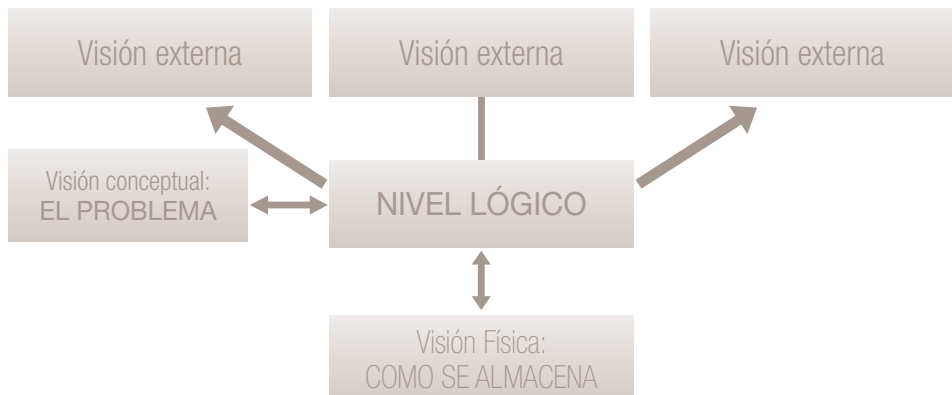
Pueda ser cambiada la organización física de los datos sin necesidad de que sea cambiada la descripción conceptual, ni los programas de usuario que utilizan las visiones externas.

Pueda ser cambiada la representación conceptual sin necesidad de modificar la organización física de los datos ni las visiones externas (siempre que no se eliminen descripciones de datos que utilicen).

Pueda ser cambiadas las visiones externas conforme aparezcan nuevos requisitos para los programas de usuario.



El nivel conceptual describe el sistema tal y como es, de manera independiente a como esté va ser tratado por los procedimientos automatizados, y por esto es necesario añadir otro nivel de abstracción, que llamaremos nivel lógico o canónico, que será derivado del conceptual, aplicándole una serie de reglas y restricciones que tendrán en cuenta cómo la información representada puede ser tratada por los procedimientos que van a manejar y definir la información.



Una vez introducidos todos los conceptos anteriores, tanto las características deseables de una base de datos como los niveles de descripción de los datos, podemos dar una definición formal de base de datos de la siguiente manera:



Se denomina **base de datos** a una colección de archivos que almacena tanto una representación abstracta del dominio de un problema como los datos correspondientes a la información acerca del mismo. Tanto la representación como los datos estarán sujetos a una serie de restricciones, las cuales también forman parte del dominio del problema y que se almacenarán también en dichos ficheros.



Icono utilizado para representar las bases de datos a modo de bloques



Es importante diferenciar entre lo que es una base de datos y los que es un Sistema Gestor de Bases de datos (SGBD), dos términos que suelen confundirse a menudo cuando se utilizan estas tecnologías.

Anteriormente se ha definido lo que es una base de datos, definiéndola y describiendo una serie de características que debe de cumplir y restricciones que deben de darse para que una base de datos pueda ser considerada como tal.



Para que la información pueda ser almacenada de la forma descrita anteriormente, es necesario disponer de una serie de procedimientos, es decir, un sistema software, que sea capaz de llevar a cabo esta labor. A este sistema se le denomina [Sistema Gestor de Bases de datos](#).

Un sistema gestor de bases de datos debe de permitir llevar a cabo las siguientes funciones:

- Definición de los datos: ha de permitir definir tanto la estructura como el contenido de los datos a almacenar a los distintos niveles de abstracción (físico, lógico y externo).

Manipulación de los datos de la base de datos: realización de consultas sobre los datos almacenados, con criterios arbitrarios y tanto consultas previstas como no previstas por el desarrollador; además de las operaciones de inserción, borrado y modificación de dichos datos.

Mantenimiento de la integridad de la base de datos: evitando cualquier tipo de pérdida accidental de los datos en sí, las relaciones existentes entre ellos, así como control de las restricciones que se impongan que deben de satisfacerse se cumplan en cualquier operación de inserción, borrado y modificación.

Control de la privacidad y seguridad de los datos: evitando cualquier pérdida accidental de éstos por cualquier causa de fallo proveniente del entorno, del hardware o del software. Además debe de controlar que solamente los usuarios autorizados puedan acceder a aquellos datos que les esté permitido manipular.

En definitiva, el SGBD debe de poner los medios necesarios para el establecimiento de las características anteriormente mencionadas que deben de exigirse a una base de datos.

Hay muchos tipos de sistemas de bases de datos. Se diferencian fundamentalmente entre sí en el modelo de datos lógico que utilizan. Se pueden distinguir la siguiente tipología de sistemas gestores de bases de datos:

- **SGBD Relacionales:** la información se almacena a nivel lógico en forma de tablas, denominadas relaciones, almacenándose la estructura de la información utilizando diferentes tipos de definiciones de tablas.
- **SGBD Jerárquicos:** la información es almacenada siguiendo una estructura arborescente. Los nodos de los árboles representan registros de información y la jerarquía se utiliza para guardar las relaciones entre los distintos tipos de datos.
- **SGBD en red:** similares a los jerárquicos, con la diferencia de que no necesariamente ha de utilizarse una estructura en forma de árbol para representar las relaciones entre los datos.
- **SGBD Orientados a Objeto:** la información se almacena en forma de objetos que pertenecen a varias clases distintas. Las clases se organizan de forma jerárquica, utilizando la herencia y encapsulan tanto datos como los procedimientos que permiten el acceso a los mismos

Además de la clasificación anterior en base al modelo lógico utilizado, podemos distinguir entre SGBD Centralizados, en los cuales la información se almacena en una sola máquina, y SGBD Distribuidos, en los cuales la base de datos no reside en un solo ordenador, sino en varios host pertenecientes a una red y no necesariamente con la misma ubicación geográfica.