

UF1469: SGBD e instalación

Elaborado por: Alberto Gómez García

Edición: 5.0

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16275-43-4 • Depósito legal: MA 1718-2014

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 **UF1469: SGDB e instalación**. Esta Unidad Formativa pertenece al Módulo Formativo **MF0224_3: Administración de sistemas gestores de bases de datos** que forma parte del Certificado de Profesionalidad **IFCT0310: Administración de bases de datos**, de la familia profesional de **Informática y comunicaciones**.

Presentación de los contenidos:

La finalidad de esta unidad formativa es conocer en profundidad los SGBD, su instalación y de qué forma nos podemos comunicar con ellos, para ello, en primer lugar analizaremos los sistemas gestores de bases de datos, posteriormente se conocerán los diccionario de datos, se analizará la estructura funcional de los SGBD, se abordará la instalación de un SGBD y por último, se describirán los mecanismos de comunicación del SGBD.

Objetivos de la unidad formativa:

- Instalar el sistema de bases de datos, determinando y aplicando la configuración del SGBD adecuada a los requisitos de rendimiento planteados.

Índice

UD1. Sistema gestores de base de datos

1.1. Introducción a la historia y evolución de los SGBD	11
1.2. Enumeración y descripción de las funciones de los SGBD	16
1.3. Clasificación de los SGBD	26
1.3.1. Modelo de datos	26
1.3.2. Número de usuarios a los que da servicio: monousuario y multiusuario	34
1.3.3. Número de sitios en los que está distribuida la BD: centralizada y distribuida	36
1.3.4. Gestión de los procesos: multiproceso y multihilo	38
1.4. Definición de la arquitectura de un SGBD atendiendo al modelo de tres capas propuesto por el comité ANSI-SPARC	40
1.4.1. Concepto de nivel interno o físico.....	42
1.4.2. Concepto de nivel externo o de visión	43
1.4.3. Concepto de nivel conceptual.....	43

UD2. Diccionario de datos

2.1. Concepto	57
2.2. Análisis de su estructura.....	59
2.3. Justificación de su importancia como elemento fundamental en la instalación y mantenimiento de la base de datos	61

UD3. Análisis de la estructura funcional del SGBD

3.1. Procesos del SGBD	71
3.2. Gestor de ficheros.....	79
3.3. Procesador y compilador de DML.....	80
3.4. Compilador del DLL.....	82
3.5. Gestión de la BD.....	84
3.6. Gestión de las conexiones de red	86

UD4. Instalación de un SGBD

4.1. Determinación de un SGBD a instalar en función de unos requerimientos planteados en un supuesto	105
4.2. Interpretación de la documentación de licencia de uso del SGBD ..	114
4.3. Identificación de las fuentes de documentación técnica. Interpretación de la documentación necesaria para la instalación..	119
4.4. Identificación y verificación de los requisitos del computador necesarios para la instalación así como los del sistema operativo..	121
4.5. Descripción de los parámetros de configuración necesarios para la puesta en marcha del SGBD tanto a nivel del propio SGBD como del entorno en el que se instala	122
4.6. Selección de componentes lógicos adicionales que puedan ser de utilidad dependiendo del supuesto de instalación	125
4.7. Determinación de la ubicación y distribución idónea del software, los datos e índices dentro del computador	127

4.8. Si el SGBD soporta varios sistemas operativos y arquitecturas de computadores, identificar las ventajas e inconvenientes de seleccionar uno u otro	128
4.9. Identificación de los posibles juegos de caracteres y elementos de internacionalización más comunes así como los posibles problemas relacionados con estos.....	134
4.10. Realización de un supuesto práctico de instalación de un SGBD (y documentación del proceso) en el que se pongan de manifiesto las relaciones entre arquitectura física del computador y las partes lógicas del SGBD	135
UD5. Descripción de los mecanismos de comunicación del SGBD	
5.1. Configuración del acceso remoto a la base de datos en al menos un SGBD del mercado	161
5.2. Descripción de la comunicación cliente-servidor con el SGBD	166
5.3. Identificación de las diferencias de medios de acceso Cliente-Servidor: Sockets, Memoria compartida, TCP/IP, etc	172
5.4. Identificación de los principales elementos que proveen de interoperabilidad al SGBD: ODBC, JDBC, etc	173
Glosario	185
Soluciones	189

UD1

Sistema de gestores de
base de datos

- 1.1. Introducción a la historia y evolución de los SGBD
- 1.2. Enumeración y descripción de las funciones de los SGBD
- 1.3. Clasificación de los SGBD atendiendo a
 - 1.3.1. Modelo de datos
 - 1.3.2. Número de usuarios a los que da servicio: monousuario y multiusuario
 - 1.3.3. Número de sitios en los que esta distribuida la BD: centralizada y distribuida
 - 1.3.4. Gestión de los procesos: multiproceso y multihilo
- 1.4. Definición de la arquitectura de un SGBD atendiendo al modelo de tres capas propuesto por el comité ANSI-SPARC
 - 1.4.1. Concepto de Nivel interno o físico
 - 1.4.2. Concepto del Nivel externo o de visión
 - 1.4.3. Concepto del Nivel conceptual

1.1. Introducción a la historia y evolución de los SGBD

Un Sistema de Gestión de Bases de Datos (SGBD) consiste en un conjunto de aplicaciones informáticas para el manejo de bases de datos. Estas aplicaciones, nos proveen de herramientas para la gestión de la información almacenada en dichas bases de datos. Son un tipo de programa muy específico que sirve de puente entre el usuario y las bases de datos.

Un SGBD no solo posee los programas o aplicaciones para la gestión de la información, sino que engloba también a los datos almacenados, que normalmente estarán relacionados entre sí.

Estos sistemas hicieron su aparición en la década de los setenta, pero hasta ese momento, la información era tratada mediante ficheros planos, cuya gestión dependía de los sistemas operativos en donde se almacenaban.

Los programadores creaban aplicaciones específicas que obtenían y procesaban los datos almacenados en esos archivos. Normalmente la estructura de los archivos era específica para ser tratados con sus programas correspondientes.

Si se modificaba la estructura de los archivos, conllevaba la modificación de los programas para que estos pudieran volver a acceder a los archivos, con el consiguiente “trastorno” de reprogramación.

Por ejemplo, si se contaba con un archivo para almacenar los clientes de una empresa, en el cual se guardaban seis campos por cada registro: id, nombre, apellidos, dirección, teléfono, fax. Si posteriormente se tenía que añadir algún campo más, como por ejemplo el email, habría que modificar el programa que leía los datos, ya que este había sido programado inicialmente para extraer solo los seis primeros campos.

Los programas de aplicación además controlaban el acceso a la información, así como la definición de los datos almacenados y sus restricciones.

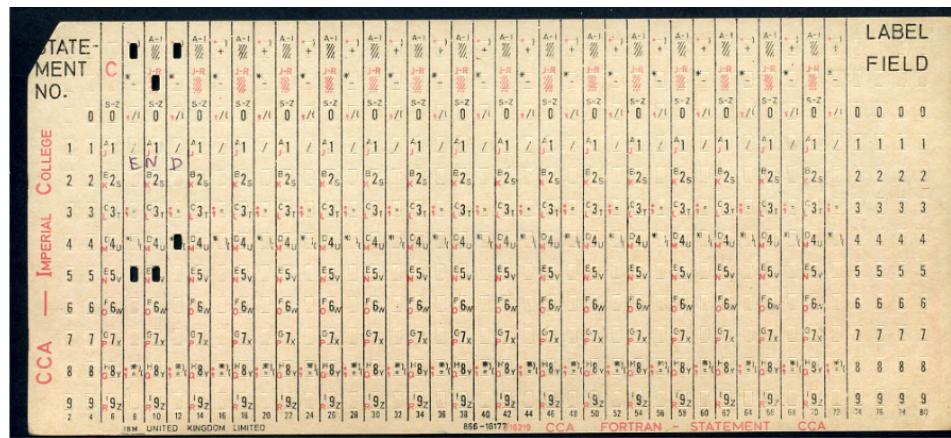
Estos sistemas provocaban con asiduidad numerosos contratiempos cuando se manipulaban grandes cantidades de información, no resultando nada prácticos.

Fue por este motivo por el que se planteó el desarrollar sistemas que pudiesen separar los datos y su gestión, de las aplicaciones que accedían a ellos, pudiendo modificar la estructura de los archivos de información, por ejemplo añadiendo o eliminando campos, sin tener que tocar los programas de aplicación.

História de los SGBD

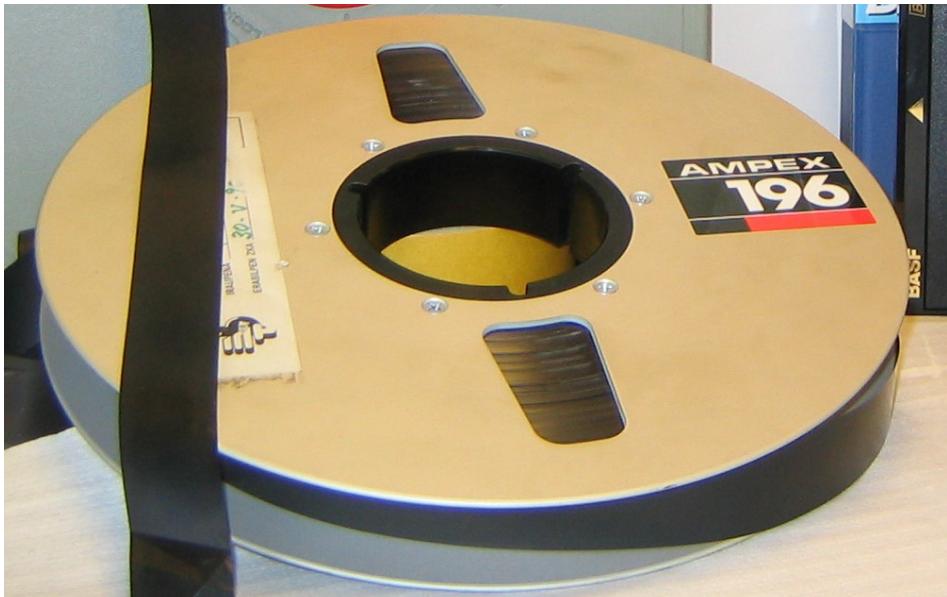
Ya en la antigüedad se almacenaba gran cantidad de información en bibliotecas por medio de libros y diversos documentos impresos, pero la historia de las bases de datos automatizadas la podemos empezar con **Herman Hollerit**, él fue el inventor en 1884 de las tarjetas perforadas para su "Máquina Automática Perforadora de Tarjetas".

Gracias a las tarjetas perforadas, se pudo almacenar el censo de Estados Unidos de 1890 en el tiempo record de dos años y medio.



A partir de estas primeras bases de datos rudimentarias que eran almacenadas en tarjetas perforadas de acceso secuencial, tuvimos que esperar hasta la década de los cincuenta para tener otra gran revolución gracias a las cintas magnéticas.

Estas cintas eran mucho más rápidas que las tarjetas perforadas y eran capaces de almacenar muchísima más información. Su principal inconveniente era su acceso secuencial.

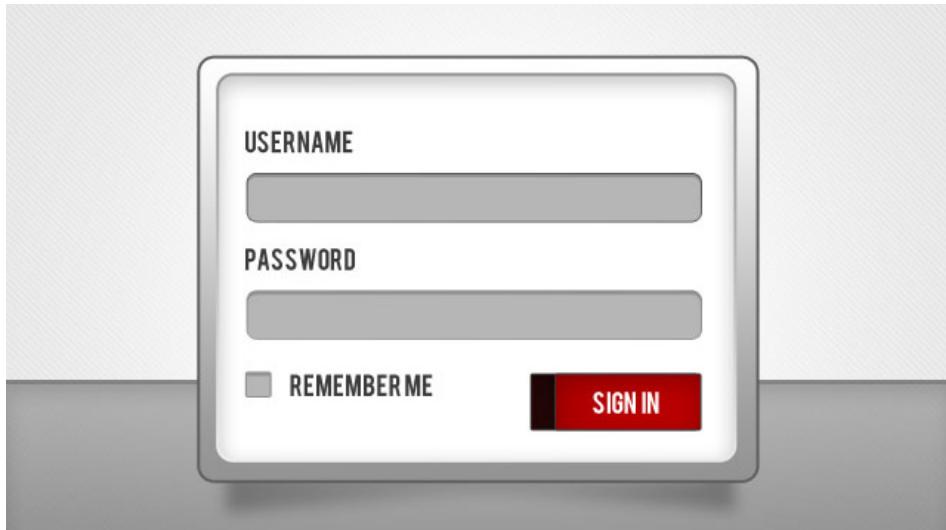


Primera generación de los SGBD

Ya en la década de los **sesenta**, se pudo vivir el abaratamiento de los ordenadores, dando la posibilidad a muchas más empresas a disponer de estos aparatos para almacenar su información. Antes de esto, solo algunas empresas privilegiadas podían contar con un sistema informático debido a su alto precio. También en estas fechas hizo su aparición el disco duro, dispositivo que permitió no solo acceder más rápido a la información sino que además hacerlo de una forma aleatoria y no secuencial.

Ya en estos años surgieron las primeras bases de **datos jerárquicas y en red**, de esta forma ya se podían almacenar los datos en forma de listas y árboles. Esta sería la **primera generación** de los SGBD.

Uno de los productos más relevantes de esta época fue desarrollado por **IBM** en lo que ahora se conoce como **IMS** (Information Management System), también de la década de los sesenta fue el desarrollo de **IDS** (Integrated Data Store), de **General Electric**.



Este proyecto fue dirigido por uno de los pioneros en los sistemas de bases de datos, **Charles Bachmann**. IDS fue el primer sistema de bases de datos de red, que mejoraba las prestaciones de los sistemas jerárquicos que había hasta la fecha.

Segunda generación de los SGBD

En la década de los **setenta**, se propone el estándar **CODASYL** (Conference on Data Systems Languages), formado por representantes del gobierno de EEUU y representantes del mundo empresarial, formaron un grupo denominado **DBTG** (Data Base Task Group), cuyo objetivo era definir unas especificaciones estándar que permitieran la creación de bases de datos y el manejo de los datos.

También en esta época, **Edgar Frank Codd**, de los laboratorios de investigación de IBM, presentó el modelo **relacional**, que venía a solventar los problemas que padecían los sistemas jerárquicos y de red.

Esto daría paso a la **segunda generación** de los SGBD. Uno de los primeros sistemas en adoptar este nuevo modelo fue el **System R** de **IBM**, sistema que dio pie al desarrollo de un lenguaje de consultas estructurado denominado **SQL**, que a la postre, se ha convertido en el lenguaje estándar de los sistemas relacionales.

Ya en la década de los **ochenta**, surgieron numerosos productos **SGBD relacionales** como: **DB2** y **SLQ/DS de IBM**, y **ORACLE** de **ORACLE Corporation**.

También empezaron los desarrollos de los sistemas relacionales de código abierto actuales.

Tercera generación de los SGBD

Ya en la década de los **noventa**, el desarrollo de las herramientas **Excel** y **Access de Microsoft**, supusieron el inicio de los sistemas **Orientados a Objetos**. También en esta época, la facilidad que dieron las redes para la intercomunicación entre ordenadores, permitieron que un programa pudiese trabajar con diferentes bases de datos como si se tratase de una sola. Es lo que se conoce como base de datos **distribuida**.

Actualmente han surgido diferentes modelos para mejorar y potenciar el modelo relacional.

Los modelos actuales de datos orientados a objetos y el modelo **relacional extendido** representan la **tercera generación** de los SGBD. Ambos modelos persiguen adaptarse a tres tecnologías: la multimedia, la de orientación a objetos (OO) e Internet y la web.

Durante los últimos años se ha comenzado a extender un tipo de aplicación de las BD denominado **Data Warehouse**, o almacén de datos. Un Data Warehouse es una colección de datos orientada a un determinado ámbito, que ayuda a la toma de decisiones en la entidad en la que se utiliza; por consiguiente, el sistema no solo posee herramientas para el tratamiento y gestión típica de la información, sino que también posee herramientas para el análisis (inteligencia empresarial) y herramientas para gestionar y recuperar los **metadatos**.



Inteligencia Empresarial o Business Intelligence, es el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización. Es decir, ayuda a la toma de decisiones a partir del análisis de la información.



Los **metadatos** son los datos de otros datos. Los metadatos engloban la información referente a los datos almacenados, como su tipo, cantidad o funcionalidad.

1.2. Enumeración y descripción de las funciones de los SGBD

Antes de enumerar y comentar las funciones que deben tener todos los SGBD, habría que enumerar los objetivos que deben alcanzar, que en muchas ocasiones son confundidos con sus funciones.

Los objetivos principales que deben perseguir todos los sistemas gestores de bases de datos son:

1. Abstracción de la información

Los usuarios que hacen uso de los sistemas de gestión de las bases de datos, desconocen los detalles sobre el almacenamiento físico de los datos. A los usuarios les da lo mismo el número de archivos que ocupa una base de datos o la cantidad de memoria que está ocupando, esta información se hace transparente al usuario.

2. Independencia

La independencia de los datos consiste en la capacidad que tienen los sistemas gestores, que si se modifica el esquema físico o lógico de una base de datos, no es necesario realizar cambios en las aplicaciones que se sirven de ella.

3. Redundancia mínima

El diseño ideal de una base de datos, es aquel que provoca una redundancia nula de los datos, es decir, que no existirán ningunos datos duplicados. En la práctica, en algunos casos la complejidad de los cálculos hace necesaria la aparición de redundancias.

4. Consistencia

Es la característica que asegura que solo se empieza aquello que se puede acabar. Solo se ejecutarán aquellas operaciones que no van a romper las reglas y directrices de integridad de la base de datos.

5. Seguridad

Dado que la información almacenada en una base de datos puede ser de gran valor. Los sistemas gestores de bases de datos tienen que contar con los medios necesarios para asegurar dicha información. En general, todos cuentan con sistemas para el control de usuarios y privilegios de acceso, con el fin de que ningún usuario sin permiso, acceda a información que no deba.

6. Manejo de transacciones

Una transacción es un programa de aplicación, generalmente de duración breve, que accede y actualiza una parte también generalmente pequeña de la base de datos. El manejo de transacciones consiste en controlar múltiples transacciones ejecutándose en paralelo sobre una misma base de datos. El sistema gestor de bases de datos deberá evitar que las transacciones interfieran unas con otras al ejecutarse en paralelo, y garantizar que la base de datos no sea dañada de forma irreparable si las transacciones provocasen errores.

7. Tiempo de respuesta

Un objetivo que debe ser perseguido siempre es el de la velocidad, que aplicado a los sistemas gestores de bases de datos, consiste en tardar el menos tiempo posible en proporcionar los datos que el usuario demanda. Esto ha de poder llevarse a cabo con múltiples peticiones concurrentes.

Una vez aclarados los objetivos a perseguir por todo sistema gestor de bases de datos, vamos a enumerar las diferentes funciones o servicios que han de poseer:

1. Creación y definición de la base de datos

La función principal de todo sistema gestor de bases de datos es la de poder definir y crear las bases de datos en donde se almacenará la información.

Los SGBD han de proporcionar las herramientas necesarias para la especificación de la estructura, el tipo de los datos, las restricciones y las relaciones que tendrán entre ellos, todo esto mediante lenguajes de definición de datos.

Un ejemplo de la creación de una tabla con SQL seria:

```
CREATE TABLE Alumnos  
  (Nombre char(50),  
   Apellidos char(50),  
   Direccionchar(50),  
   Ciudad char(50),  
   Paischar(25),  
   Fecha_nacimiento datetime);
```

Otras posibles sentencias serían las de DROP o ALTER.

Igualmente el SGBD debe ocultar al usuario la estructura física interna, es decir, la organización de los ficheros y las estructuras de almacenamiento.

Toda esta información será almacenada dentro del diccionario de datos, el sistema gestor de bases de datos también proporcionará herramientas para la gestión del diccionario de datos.

Un diccionario de datos es un catálogo en el que se almacenen las descripciones de los datos y este debe ser accesible por los usuarios.

2. Manipulación de los datos

Otra de las funciones principales es la de posibilitar la manipulación de la información que almacena.

Esta manipulación incluirá la posibilidad de realizar consultas, inserciones y actualizaciones de los datos almacenados, utilizando para ello diferentes lenguajes de manipulación de datos.

Los usuarios podrán, dependiendo de los permisos que posean, realizar diferentes operaciones con los registros.

Un ejemplo de insertar un registro en la tabla anterior mediante SQL sería:

```
INSERT INTO Alumnos (Nombre, Apellidos, Ciudad)  
VALUES ('Pedro', 'Pérez Ruiz', 'Madrid');
```

Para modificar el registro anterior:

```
UPDATE Alumnos  
SET Ciudad = 'Barcelona'  
WHERE Nombre = 'Pedro';
```

Por último, para eliminarlo:

```
DELETE FROM Alumnos  
WHERE Nombre = 'Pedro';
```

(Podrás ampliar información sobre los comandos SQL en la Guía básica de comandos SQL en <http://sistemas.uniandes.edu.co>)

3. Control de transacciones

Una de las áreas principales de aplicación de los sistemas gestores de bases de datos es el denominado procesamiento de transacciones.

Como ya comentamos anteriormente, una transacción es un programa de aplicación, generalmente de duración breve, que accede y actualiza una parte también generalmente pequeña de la base de datos. El manejo de transacciones consiste en controlar múltiples transacciones ejecutándose en paralelo sobre una misma base de datos.

El sistema gestor de bases de datos deberá evitar que las transacciones interfieran unas con otras al ejecutarse en paralelo, y garantizar que la base de datos no sea dañada de forma irreparable si las transacciones provocasen errores.

Para asegurar que un SGBD cumple con un control de transacciones completo, existe el estándar ACID, acrónimo de Atomicity, Consistency, Isolation and Durability (Atomicidad, Consistencia, Aislamiento y Durabilidad en español).

Un SGBD que soporte ACID garantiza lo siguiente:

a) *Atomicidad*

La atomicidad es la propiedad que asegura que la operación se ha realizado o no, no quedando a medias por motivo de un fallo en el sistema.

b) *Consistencia*

La consistencia es la propiedad que asegura que solo se empieza aquello que se puede acabar. Solo se ejecutarán aquellas operaciones que no van a romper las reglas y directrices de integridad de la base de datos.

c) *Aislamiento*

El aislamiento es la propiedad que asegura que una operación no puede afectar a otras. Si se realizan dos transacciones sobre los mismos datos, estas sean independientes y no generen errores.

d) *Durabilidad*

La durabilidad es la propiedad que asegura que una vez realizada la operación, está persistirá y no se podrá deshacer aunque falle el sistema.

Actualmente no todos los SGBD ofrecen un soporte ACID completo, limitándose muchos de ellos a solo soportarlo en parte.

4. Control de los usuarios

Todo SGBD debe implementar los mecanismos de seguridad necesarios para el control de acceso de los usuarios a la información almacenada.