Introducción

La información es un recurso cada vez más crítico e importante para las empresas y organizaciones. Por ello es crucial disponer de herramientas que permitan y faciliten su administración.

El administrador de bases de datos debe asegurar y controlar la información de manera que siempre esté accesible al usuario.

Para ello los administradores tienen como tareas principales el diseño, instalación y configuración de sistemas gestores de bases de datos, automatización de tareas mediante la programación de rutinas y scripts, la realización copias de seguridad, la gestión usuarios y permisos sobre los datos y la monitorización y optimización del rendimiento de los SGBD

De todo ello nos ocupamos en este libro, que está concebido para introducir al lector en todos los aspectos importantes de la administración.

Por motivos de sencillez la mayoría de ejercicios y ejemplos del libro se basan en el uso de MySQL sobre máquinas *Windows*, sin embargo hay numerosas referencias e incluso ejercicios en entornos *Linux* de forma que el alumno pueda tener una visión más completa y heterogénea.

Administrar sistemas no es una tarea sencilla, ni mucho menos sistemática, este y otros libros pueden ayudar pero solo la experiencia y el trabajo diario con sistemas reales en producción permitirán la formación completa de un administrador.

Además, es imposible detallar en un solo libro todos los contenidos y todas las posibilidades que pueden darse en el trabajo administrativo. Es por ello crucial para un administrador saber hacer uso de las herramientas de ayuda que proporcionan tanto las herramientas de su gestor de bases de datos cómo de la documentación oficial, en este caso de MySQL, disponible en la web.

A partir del segundo capítulo se ha incluirlo un caso base hipotético para trabajar sobre él lo explicado en cada capítulo. Éste se basa en un caso o ejemplo concreto. En nuestro caso hemos elegido una aplicación de plena actualidad como son las dedicadas a ofrecer música y vídeo bajo demanda.

Conviene señalar que la mayoría de ejemplos del libro se realizan usando comandos en lugar de interfaces gráficas, ya que estos proporcionan mayor flexibilidad y potencia al administrador. Además permiten conocer mejor las interioridades de nuestros sistemas.

Este libro surge con el propósito de acercar al lector a los aspectos más importantes que encierran los sistemas gestores de bases de datos, ante la creciente demanda de personal cualificado para su administración. Con tal propósito, puede servir de apoyo también para estudiantes del Ciclo Formativo Administración de Sistemas Informáticos en Red de Grado Superior y de Ingenierías Técnicas.

Para todo aquel que use este libro en el entorno de la enseñanza (Ciclos Formativos o Universidad), se ofrecen varias posibilidades: utilizar los conocimientos aquí expuestos para inculcar aspectos genéricos de los sistemas gestores de bases de datos o simplemente centrarse en preparar a fondo alguno de ellos. La extensión de los contenidos aquí incluidos hace imposible su desarrollo completo en la mayoría de los casos.

Ra-Ma pone a disposición de los profesores una guía didáctica para el desarrollo del tema que incluye las soluciones a los ejercicios expuestos en el texto. Puede solicitarlo a *editorial@ra-ma.com*, acreditándose como docente y siempre que el libro sea utilizado como texto base para impartir las clases.

1 Revisión de conceptos de bases de datos

OBJETIVOS DEL CAPÍTULO

- ✓ Instalar el servidor MySQL.
- ✓ Conocer las opciones de configuración de MySQL.
- ✓ Optimizar el funcionamiento de MySQL.
- ✓ Monitorizar MySQL.
- ✓ Aprender a gestionar ficheros de registro.
- ✓ Conocer la estructura del diccionario de datos de MySQL.

1.1 INTRODUCCIÓN. DEFINICIÓN DE BASES DE DATOS Y SGBD

Conviene, antes de comenzar con el tema principal del libro, recordar los conceptos más relevantes relacionados con sistemas gestores de bases de datos, así como las herramientas relacionadas.

En primer lugar y, aunque probablemente ya se ha visto en otros módulos, es importante diferenciar entre el concepto de base de datos y el de sistema gestor, ya que es habitual confundirlos y, sin embargo, son cosas muy distintas.

Definición 1:

Una base de datos es un conjunto de datos relacionados y organizados con cierta estructura. Según dicha organización distinguimos entre diferentes modelos de bases de datos como el relacional, jerárquico o en red.

El modelo de bases de datos más extendido es el **relacional** y es el que trabajaremos en este libro.

Para su manipulación y gestión surgieron los sistemas gestores de bases de datos **(SGBD** en lo sucesivo).

Definición 2:

El sistema de gestión de la base de datos (SGBD) es una aplicación que permite a los usuarios definir, crear y mantener bases de datos, proporcionando acceso controlado a las mismas. Es una herramienta que sirve de interfaz entre el usuario y las bases de datos.

Es decir, por un lado tenemos los datos organizados según ciertos criterios y, por otro, un software que nos permite o facilita su gestión con distintas herramientas y funcionalidades que describimos a continuación.

1.2 ARQUITECTURA DE SISTEMAS DE BASES DE DATOS

Hay tres características importantes inherentes a los sistemas de bases de datos: la separación entre los programas de aplicación y los datos, el manejo de múltiples vistas por parte de los usuarios y el uso de un catálogo para almacenar el esquema de la base de datos. En 1975, el comité ANSI-SPARC (*American National Standard Institute - Standards Planning and Requirements Committee*) propuso una arquitectura de tres niveles para los sistemas de bases de datos, que resulta muy útil a la hora de conseguir estas tres características.

El objetivo de la arquitectura de tres niveles es el de separar los programas de aplicación de la base de datos física. En esta arquitectura, el esquema de una base de datos se define en tres niveles de abstracción distintos como se aprecia en la siguiente imagen:

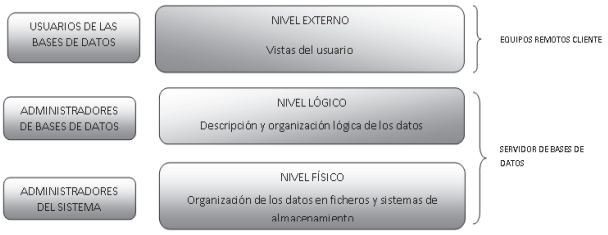


Figura 1.1. *Arquitectura de Sistemas de Bases de Datos*

En el nivel interno se describe la estructura física de la base de datos mediante un esquema interno. Este esquema se especifica mediante un modelo físico y describe todos los detalles para el almacenamiento de la base de datos, así como los métodos de acceso.

En el nivel conceptual se describe la estructura de toda la base de datos para una comunidad de usuarios (todos los de una empresa u organización), mediante un esquema conceptual. Este esquema oculta los detalles de las estructuras de almacenamiento y se concentra en describir entidades, atributos, relaciones, operaciones de los usuarios y restricciones. En este nivel se puede utilizar un modelo conceptual o un modelo lógico para especificar el esquema.

En el nivel externo se describen varios esquemas externos o vistas de usuario. Cada esquema externo describe la parte de la base de datos que interesa a un grupo de usuarios determinado y oculta a ese grupo el resto de la base de datos.

La arquitectura de tres niveles es útil para explicar el concepto de **independencia de datos,** que podemos definir como *la capacidad para modificar el esquema en un nivel del sistema sin tener que modificar el esquema del nivel inmediato superior*. Se pueden definir dos tipos de independencia de datos:

La **independencia lógica** es la capacidad de modificar el esquema conceptual sin tener que alterar los esquemas externos ni los programas de aplicación. Se puede modificar el esquema conceptual para ampliar la base de datos o para reducirla. Si, por ejemplo, se reduce la base de datos eliminando una entidad, los esquemas externos que no se refieran a ella no deberán verse afectados.

La **independencia física** es la capacidad de modificar el esquema interno sin tener que alterar el esquema conceptual (o los externos). Por ejemplo, puede ser necesario reorganizar ciertos ficheros físicos con el fin de mejorar el rendimiento de las operaciones de consulta o de actualización de datos. Dado que la independencia física se refiere solo a la separación entre las aplicaciones y las estructuras físicas de almacenamiento, es más fácil de conseguir que la independencia lógica.

1.3 FUNCIONES DEL SISTEMA GESTOR DE BASE DE DATOS (SGBD)

La función principal de un **SGBD** es permitir a los usuarios realizar las cuatro operaciones fundamentales posibles, tanto sobre las estructuras de datos como sobre los datos que albergan, es decir, operaciones de inserción o creación, consulta, actualización y borrado, de una manera eficiente y coherente.

Para tal fin, la mayoría de SGBD incorporan las siguientes características y funciones:

Un catálogo

Donde se almacenen las descripciones de los datos y sea accesible por los usuarios. Este **catálogo** es lo que se denomina *diccionario de datos* y contiene información que describe los datos de la base de datos (metadatos). Normalmente, un diccionario de datos describe entre otras cosas:

- Nombre, tipo y tamaño de los datos.
- Relaciones entre los datos.
- Restricciones de integridad sobre los datos.
- Usuarios autorizados a acceder a los objetos de base de datos.
- Estadísticas de utilización, tales como la frecuencia de las transacciones y el número de accesos realizados a los objetos de la base de datos.

Garantizar la integridad

Disponer de un mecanismo que garantice que todas las actualizaciones correspondientes a una determinada transacción se realicen, o que no se realice ninguna. Una transacción es un conjunto de acciones que cambian el contenido de la base de datos.

Permitir actualizaciones concurrentes

Asegurar que la base de datos se actualice correctamente cuando varios usuarios la están actualizando concurrentemente.

Recuperación de datos

Permitir recuperar las bases de datos en caso de que ocurra algún suceso que la dañe.

Integración

Ser capaz de integrarse con algún software de comunicación. Muchos usuarios acceden a la base de datos desde terminales. En ocasiones estos terminales se encuentran conectados directamente a la máquina sobre la que funciona el SGBD. En otras ocasiones los terminales están en lugares remotos, por lo que la comunicación con la máquina que alberga al SGBD se debe hacer a través de una red. En cualquiera de los dos casos, el SGBD recibe peticiones en forma de mensajes y responde de modo similar. Todas estas transmisiones de mensajes las maneja el gestor de comunicaciones de datos. Aunque este gestor no forma parte del SGBD, es necesario que el SGBD se pueda integrar con él para que el sistema sea comercialmente viable.

Cumplir restricciones

Proporcionar los medios necesarios para garantizar que tanto los datos de la base de datos, como los cambios que se realizan sobre estos datos, sigan ciertas reglas o condiciones que mantengan su integridad.

Herramientas de administración

Proporcionar herramientas que permitan administrar la base de datos de modo efectivo, lo que implica un diseño óptimo de las mismas, garantizar la disponibilidad e integridad de los datos, controlar el acceso al servidor y a los datos, monitorizar el funcionamiento del servidor y optimizar su funcionamiento.

1.4 COMPONENTES

Son los elementos que deben proporcionar los servicios comentados en la sección anterior. No se puede generalizar ya que varían mucho según la tecnología. Sin embargo, es muy útil conocer sus componentes y cómo se relacionan cuando se trata de comprender lo que es un sistema de bases de datos.

El **SGBD** es la aplicación que interacciona con los usuarios de los programas de aplicación y la base de datos. En general, un SGBD suele incluir los siguientes componentes:

- **Lenguaje de definición de datos** (**DDL**: *Data Definition Language*)

  Sencillo lenguaje artificial para definir y describir los objetos de la base de datos, su estructura, relaciones y restricciones.

- **Lenguaje de control de datos** (**DCL**: *Data Control Language*)

  Encargado del control y seguridad de los datos (privilegios y modos de acceso, etc.). Este lenguaje permite especificar la estructura y el tipo de los datos, así como las restricciones sobre los datos. Todo esto se almacenará en la base de datos.

- **Lenguaje de manipulación de datos** (**DML**: *Data Manipulation Language*)

  Para la inserción, actualización, eliminación y consulta de datos. Para tal fin el lenguaje por excelencia es el conocido SQL (*Structured Query Language*). Incluye instrucciones para los tres tipos de lenguajes comentados y por su sencillez y potencia se ha convertido en el lenguaje estándar de los **SGBD relacionales**.

- **Diccionario de datos**

  Esquemas que describen el contenido del **SGBD** incluyendo los distintos objetos con sus propiedades.

- **Objetos**: **Tablas base y vistas (tablas derivadas)**
  - Consultas.
  - Dominios y tipos definidos de datos.
  - Restricciones de tabla y dominio y aserciones.
  - Funciones y procedimientos almacenados.
  - Disparadores o *triggers*.

- **Distintas herramientas para**:
  - Seguridad: de modo que los usuarios no autorizados no puedan acceder a la base de datos.
  - Integridad: que mantiene la integridad y la consistencia de los datos.
  - El control de concurrencia: que permite el acceso compartido a la base de datos.
  - El control de recuperación: que restablece la base de datos después de que se produzca un fallo del hardware o del software.
  - Gestión del diccionario de datos (o catálogo): accesible por el usuario que contiene la descripción de los datos de la base de datos.
  - Programación de aplicaciones.
  - Importación/exportación de datos (migraciones).
  - Distribución de datos.
  - Replicación (arquitectura maestro-esclavo).
  - Sincronización (de equipos replicados).

- **Optimizador de consultas**

  Para determinar la estrategia óptima para la ejecución de las consultas.

- **Gestión de transacciones**

  Este módulo realiza el procesamiento de las transacciones.

- **Planificador** (*scheduler*)

  Para programar y automatizar la realización de ciertas operaciones y procesos.

- **Copias de seguridad**

  Para garantizar que la base de datos se puede devolver a un estado consistente en caso de que se produzca algún fallo o error grave.

1.5 USUARIOS DE LOS SGBD

Generalmente distinguimos cuatro grupos de usuarios de sistemas gestores de bases de datos: los usuarios administradores, los diseñadores de la base de datos, los programadores y los usuarios de aplicaciones que interactúan con las bases de datos.

Administrador de la base de datos

Se encarga del diseño físico de la base de datos y de su implementación, realiza el control de la seguridad y de la concurrencia, mantiene el sistema para que siempre se encuentre operativo y se encarga de que los usuarios y las aplicaciones obtengan buenas prestaciones. El administrador debe conocer muy bien el SGBD que se esté utilizando, así como el equipo informático sobre el que esté funcionando.

Diseñadores de la base de datos

Realizan el diseño lógico de la base de datos, debiendo identificar los datos, las relaciones entre datos y las restricciones sobre los datos y sus relaciones. El diseñador de la base de datos debe tener un profundo conocimiento de los datos de la empresa y también debe conocer sus reglas de negocio.

Programadores de aplicaciones

Se encargan de implementar los programas de aplicación que servirán a los usuarios finales. Estos programas de aplicación son los que permiten consultar datos, insertarlos, actualizarlos y eliminarlos. Estos programas se escriben mediante lenguajes de tercera generación o de cuarta generación.

Usuarios finales

Clientes de la base de datos que hacen uso de ella sin conocer en absoluto su funcionamiento y organización. Son personas con pocos o nulos conocimientos de informática.

1.6 TIPOS DE SGBD

Existen numerosos SGBD en el mercado que podemos clasificar según distintos criterios:

- **Modelo lógico en el que se basan**
  - Modelo Jerárquico.
  - Modelo de Red.
  - Modelo Relacional.
  - Modelo Objeto-Relacional
  - Modelo Orientado a Objetos.
- **Número de usuarios**
  - Monousuario.
  - Multiusuario.
- **Número de sitios**
  - Centralizados.
  - Distribuidos: homogéneos y heterogéneos.
- **Ámbito de aplicación**
  - Propósito General.
  - Propósito Específico.
- **Estructura de datos**
  - Tipo tabla.
  - Tipo clave valor (NoSQL).
- **Arquitectura hardware**
  - Paralelas: se ejecutan sobre multiples procesadores y discos.
  - Utilizan un solo procesador.

Sin embargo, son los SGBD relacionales los que se han impuesto hasta llegar a dominar casi totalmente el mercado actual. Ello se ha debido principalmente a su flexibilidad y sencillez de manejo. Igualmente conviene destacar la amplia implantación del lenguaje SQL, que se ha convertido en un estándar para el manejo de datos en el modelo relacional, lo que ha supuesto una ventaja adicional para su desarrollo.

Sin embargo, debido al aumento de recursos y prestaciones de los equipos informáticos, así como del auge de otros tipos de aplicaciones orientadas al trabajo con grandes volúmenes de datos poco estructurados más que a la gestión de información, se está extendiendo el uso de sistemas basados en NoSQL (Not Only SQL), cuyas características principales son la velocidad en almacenamiento y recuperación de datos y la distribución de los mismos en distintos nodos o equipos.

Estos sistemas se utilizan sobre todo para aplicaciónes de carácter estadístico y de tiempo real.

1.7 SISTEMAS GESTORES DE BASE DE DATOS COMERCIALES Y LIBRES

Con el advenimiento de Internet, el software libre se ha consolidado como alternativa, técnicamente viable y económicamente sostenible al software comercial, contrariamente a lo que a menudo se piensa, convirtiéndose el software libre como otra alternativa para ofrecer los mismos servicios a un coste cada vez más reducido.

Estas alternativas se encuentran tanto para herramientas de ofimática como LibreOffice o Microsoft Office. También disponemos de herramientas mucho más avanzadas a un nivel de propósito general como MySQL, SQL Server y si hablamos de software con más potencia y funcionalidad vale la pena señalar a Postgresql u Oracle, entre otros.

No conviene decantarse por uno en concreto, debemos usar en primer lugar lo que mejor nos funciona dadas nuestras restricciones particulares. Así, si solamente quiero una aplicación de agenda puede ser perfectamente válido un Access o incluso un LibreOffice Base (y en según qué casos incluso un *Bloc de notas*).

Si, por el contrario, mi base de datos requiere cierta cantidad de accesos de usuarios diversos, control de integridad y otras funcionalidades, debería plantearme algo como MySQL.

Finalmente, si hablamos de una gran corporación que requiere herramientas avanzadas para grandes bases de datos quizás debamos plantearnos sistemas más potentes como Oracle o Postgresql.

En cuanto a si debe ser libre o no la decisión dependerá de si disponemos de personal cualificado, en cuyo caso un sistema libre es más barato y potente. En caso contrario el sistema de pago es la elección más adecuada. No obstante todas las tecnologías disponen de servicios de soporte de gran calidad.

Por otro lado, los sistemas libres cada vez proveen una mejor y más eficiente documentación tanto a nivel oficial como a través de múltiples foros y blogs, lo que hace que cada vez lo use más gente y mejore continuamente. Sin embargo, siempre existe el riesgo de que sea comercializado y deje de estar disponible de forma abierta.

En resumen, para la toma de esta decisión se tendrán en cuenta factores como:

- Documentación.
- Seguridad, control de acceso a los recursos.
- Volúmenes de información que soportará y número de accesos esperable.
- Complejidad en la migración de los datos.
- Soporte ofrecido.

Como siempre, al final la solución estará determinada por las características de la organización en cuanto a requerimientos de su sistema de información y al personal de que dispone, recursos económicos, etc.

En todo caso, una solución general de compromiso podría ser aquella que involucre software libre y un contrato de soporte.

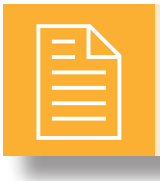
RESUMEN DEL CAPÍTULO

En este capítulo introductorio hemos repasado someramente las principales características de los sistemas gestores de bases de datos actuales.

De los mismos, hemos visto sus componentes, funcionalidades principales y usuarios que trabajan con los denominados SGBD. Las posibilidades han crecido enormemente de un tiempo a esta parte tanto en lo que respecta a sistemas de pago, como de código libre, y hoy en día disponemos de un repertorio bastante amplio y potente de herramientas que hacen más fácil nuestra tarea, tanto como administración de bases de datos como de diseño y uso de las mismas.

EJERCICIOS PROPUESTOS

- **1.** Comenta qué se entiende por software libre considerando aspectos como:
  - Gratuidad.
  - Código fuente.
  - Uso comercial.
- **2.** Lista al menos 3 ventajas e inconvenientes de los productos de pago respecto a los libres.
- **3.** ¿Qué tiene que ver la administración SGBD con el diseño de bases de datos?
- **4.** Cita al menos 3 ventajas de usar bases de datos frente a los tradicionales sistemas de ficheros.
- **5.** Enumera al menos tres objetos típicos de una base de datos indicando su función.
- **6.** ¿Qué es una base de datos distribuida?
- **7.** Indica resumidamente las fases involucradas en el desarrollo de una base de datos desde su concepción hasta su puesta en marcha.
- **8.** ¿Para qué sirve un disparador en un SGBD?
- **9.** Explica con tus palabras qué es el diccionario de datos en un SGBD.
- **10.** Eres administrador de la base de datos. Indica un problema y su posible solución que te pueda surgir considerando dos casos: una base de datos con miles de usuarios y centrada en consultas, como un buscador, y otra de venta *on line* con miles de usuarios y operaciones por segundo.



TEST DE CONOCIMIENTOS

1 ¿Qué es una base de datos?
- **a)** Un programa para organizar datos.
- **b)** Un software que facilita la gestión de datos.
- **c)** Un conjunto de datos organizados.
- **d)** Todo lo anterior.

2 ¿Cuál es el significado de GPL en el contexto informático?
- **a)** *General Public Library*.
- **b)** *Great Politic Licence*.
- **c)** *General Public Licence*.

3 ¿Cuáles de los siguientes objetos no son equivalentes a una tabla?
- **a)** Vista.
- **b)** Consulta.
- **c)** *Trigger*.
- **d)** Procedimiento.

4 ¿Qué se quiere decir cuando se habla de nivel conceptual?
- **a)** Lo que percibe el usuario.
- **b)** La imagen de la base de datos vista por el ordenador.
- **c)** El código para crear la base de datos.
- **d)** Una imagen de la base de datos independiente de la implementación física.

5 Las bases de datos son:
- **a)** Relacionales.
- **b)** Relacionales o jerárquicas.
- **c)** Primero eran en red y ahora son relacionales.
- **d)** La mayoría son relacionales.

6 Un modelo es:
- **a)** Una forma de representar información.
- **b)** Un programa para dibujar cajas y flechas.
- **c)** Una forma de representar un sistema.
- **d)** Una representación de un conjunto de datos.

7 Un sistema de información:
- **a)** Describe los datos de un sistema.
- **b)** Permite controlar la información de una empresa.
- **c)** Es el conjunto de elementos para gestionar la información de un sistema.
- **d)** Todo lo anterior.

8 ¿Qué es cierto respecto a los SGBD y bases de datos?
- **a)** No hay diferencia.
- **b)** Uno hace referencia a un software y una base es conceptual.
- **c)** Las bases de datos se crean necesariamente con un SGBD.
- **d)** Un SGBD es una herramienta CASE.

9 Los sistemas libres:
- **a)** Son más potentes y mejores que los comerciales.
- **b)** Son más baratos.
- **c)** Son más difíciles.
- **d)** Ninguno de los anteriores necesariamente.

10 La independencia física:
- **a)** Hace que podamos acceder a los datos desde cualquier equipo.
- **b)** Permite modificar los modelos independientemente de su almacenamiento.
- **c)** Evita problemas de redundancia.
- **d)** Hace que podamos usar las bases de datos independientemente del sistema operativo.