

UNIDAD 1. FUNDAMENTOS DE BASES DE DATOS

CAPITULO 1. INTRODUCCIÓN A LAS BASES DE DATOS

La información hoy en día es el motor de toma de decisiones y el tesoro maspreciado de las organizaciones, ya que al conocer de una manera adecuada y de una forma acertada los procesos internos y externos se puede en determinado momento ayudar a tomar decisiones, esto con fin de lograr una mejor producción o posición en el mercado.

El manejo de gran cantidad de datos es la consecuencia del aumento de información que se maneja en el transcurso de nuestras vida y mas aun en el mundo empresarial, al incrementar todo este volumen de información que diariamente se puede acumular en el manejo de una empresa u organización se hace necesario organizarla para poder encontrar resultados rápidos y óptimos en el momento de utilizarla.

Debido a esta necesidad en los años setenta, para manejar toda esta información surgen las bases de datos, en la cual se integran archivos individuales para poder ser compartidos por todos los usuarios de la empresa.

Para el desarrollo de esta técnica de almacenamiento y ordenamiento de información es necesario conocer la causa que llevo el desenvolvimiento de dicho proceso, uno de ellos fue la transmisión de la información o los datos, es decir que el usuario tenga capacidad y facilidad de acceder a los datos de una forma remota.

Que el proceso de dialogo que se genere entre el usuario y la maquina sea el mas amigable y compatible en donde se pueda consultar y borrar, modificar e insertar datos en cualquier momento que se necesite en el manejo de la información.

El diseño de la base de datos es de gran importancia en el manejo de la información, ya que tiene como principal objetivo que los datos almacenados se puedan utilizar por una gran numero de aplicaciones.

Los predecesores de los sistemas de bases de datos fueron los sistemas de ficheros o archivos. No hay un momento concreto en que los sistemas de archivos hayan cesado y hayan dado comienzo los sistemas de bases de datos. De hecho, todavía existen sistemas de archivos en uso.

Los sistemas jerárquico y de red constituyen la primera generación de los SGBD. Pero estos sistemas presentan algunos inconvenientes:

- Es necesario escribir complejos programas de aplicación para responder a cualquier tipo de consulta de datos, por simple que ésta sea.
- La independencia de datos es mínima.
- No tienen un fundamento teórico.

En 1970 Codd, de los laboratorios de investigación de IBM, escribió un artículo presentando el modelo relacional. En este artículo, presentaba también los inconvenientes de los

sistemas previos, el jerárquico y el de red. Entonces, se comenzaron a desarrollar muchos sistemas relacionales, apareciendo los primeros a finales de los setenta y principios de los ochenta. Uno de los primeros es System R, de IBM, que se desarrolló para probar la funcionalidad del modelo relacional, proporcionando una implementación de sus estructuras de datos y sus operaciones. Esto condujo a dos grandes desarrollos:

- El desarrollo de un lenguaje de consultas estructurado denominado SQL, que se ha convertido en el lenguaje estándar de los sistemas relacionales.
- La producción de varios SGBD relacionales durante los años ochenta, como DB2 y SLQ/DS de IBM, y ORACLE de ORACLE Corporation.

Hoy en día, existen cientos de SGBD relacionales, tanto para microordenadores como para sistemas multiusuario, aunque muchos no son completamente fieles al modelo relacional.

Como respuesta a la creciente complejidad de las aplicaciones que requieren bases de datos, han surgido dos nuevos modelos: el modelo de datos orientado a objetos y el modelo relacional extendido. Sin embargo, a diferencia de los modelos que los preceden, la composición de estos modelos no está clara. Esta evolución representa la tercera generación de los SGBD

1.1 Propósito de los sistemas

Hoy en día, son muchas las aplicaciones que requieren acceder a datos. Bien sea un sencillo programa doméstico, bien una suite para la gestión empresarial. Estos datos se deben almacenar en algún soporte permanente, y las aplicaciones deben disponer de un medio para acceder a ellos. Normalmente, la forma en que un programa accede a un fichero es a través del Sistema operativo. Este provee de funciones como abrir archivo, leer información del archivo, guardar información, etc. No obstante, este procedimiento de acceso a ficheros es altamente ineficaz cuando se trata con un volumen elevado de información. Es aquí donde aparecen los Sistemas Gestores de Bases de Datos: proporcionan un interfaz entre aplicaciones y sistema operativo, consiguiendo, entre otras cosas, que el acceso a los datos se realice de una forma más eficiente, más fácil de implementar, y, sobre todo, más segura.

Un sistema gestor de bases de datos (SGDB) consiste en una colección de datos interrelacionados y un conjunto de programas para acceder a dichos datos. La colección de datos, normalmente denominada base de datos, contiene información relevante para una empresa. El objetivo principal de un SGDB es proporcionar una forma de almacenar y recuperar la información de una base de datos de manera que sea tanto práctica como eficiente.

Los sistemas de bases de datos se diseñan para gestionar grandes cantidades de información

Existen distintos objetivos que deben cumplir los SGBD:

- **Abstracción de la información.** Los usuarios de los SGBD ahorran a los usuarios detalles acerca del almacenamiento físico de los datos. Da lo mismo si una base de datos ocupa uno o cientos de archivos, este hecho se hace transparente al usuario. Así, se definen varios niveles de abstracción.
- **Independencia.** La independencia de los datos consiste en la capacidad de modificar el esquema (físico o lógico) de una base de datos sin tener que realizar cambios en las aplicaciones que se sirven de ella.
- **Redundancia mínima.** Un buen diseño de una base de datos logrará evitar la aparición de información repetida o redundante. De entrada, lo ideal es lograr una redundancia nula; no obstante, en algunos casos la complejidad de los cálculos hace necesaria la aparición de redundancias.
- **Consistencia.** En aquellos casos en los que no se ha logrado esta redundancia nula, será necesario vigilar que aquella información que aparece repetida se actualice de forma coherente, es decir, que todos los datos repetidos se actualicen de forma simultánea.
- **Seguridad.** La información almacenada en una base de datos puede llegar a tener un gran valor. Los SGBD deben garantizar que esta información se encuentra asegurada frente a usuarios malintencionados, que intenten leer información privilegiada; frente a ataques que deseen manipular o destruir la información; o simplemente ante las torpezas de algún usuario autorizado pero despistado. Normalmente, los SGBD disponen de un complejo sistema de permisos a usuarios y grupos de usuarios, que permiten otorgar diversas categorías de permisos.
- **Integridad.** Se trata de adoptar las medidas necesarias para garantizar la validez de los datos almacenados. Es decir, se trata de proteger los datos ante fallos de hardware, datos introducidos por usuarios descuidados, o cualquier otra circunstancia capaz de corromper la información almacenada.
- **Respaldo y recuperación.** Los SGBD deben proporcionar una forma eficiente de realizar copias de seguridad de la información almacenada en ellos, y de restaurar a partir de estas copias los datos que se hayan podido perder.
- **Control de la concurrencia.** En la mayoría de entornos (excepto quizás el doméstico), lo más habitual es que sean muchas las personas que acceden a una base de datos, bien para recuperar información, bien para almacenarla. Y es también frecuente que dichos accesos se realicen de forma simultánea. Así pues, un SGBD debe controlar este acceso concurrente a la información, que podría derivar en inconsistencias.
- **Tiempo de respuesta.** Lógicamente, es deseable minimizar el tiempo que el SGBD tarda en darnos la información solicitada y en almacenar los cambios realizados.

1.2 Visión de los datos

Korth define, un sistema de bases de datos como una colección de archivos interrelacionados y un conjunto de programas que permitan a los usuarios acceder y modificar estos archivos. Uno de los propósitos principales de un sistema de bases de datos es proporcionar a los usuarios una visión abstracta de los datos.

Abstracción de datos: para que el sistema sea útil, si debe recuperar los datos eficientemente. Los tres niveles de abstracción de datos son:

- **Nivel físico:** describe cómo se almacena realmente los datos. Se describen en detalle las estructuras de datos complejas de bajo nivel.
- **Nivel lógico:** describe que datos se almacena en la base de datos y qué relaciones existen entre esos datos.
- **Nivel de vistas:** describe solo parte de la base de datos completa.
-

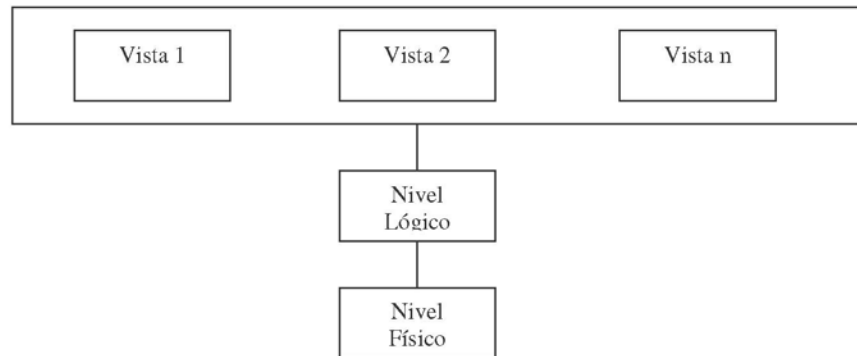
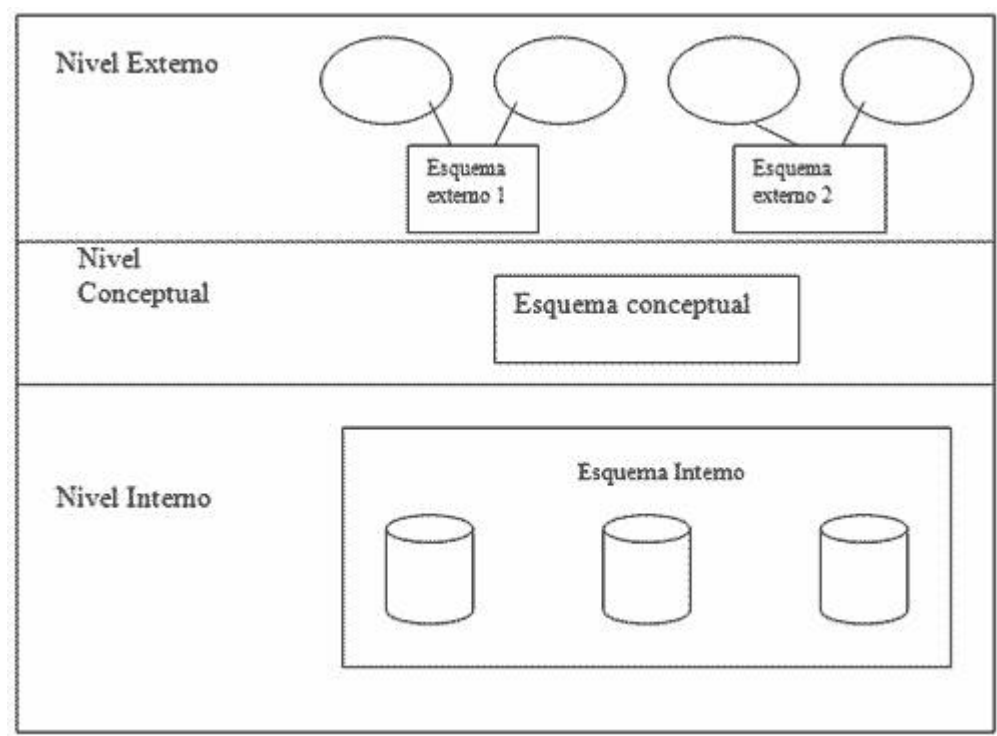


Figura 1.1
Los tres niveles de abstracción de datos



1.3 Modelos de los datos

Bajo la estructura de la base de datos se encuentra el modelo de datos: una colección de herramientas conceptuales para describir los datos, las relaciones, la semántica y las restricciones de consistencia. Los diferentes modelos de datos que sean propuesto se clasifican en tres grupos diferentes: modelos lógicos basados en los objetos, modelos lógicos basados en registros y modelos físicos.

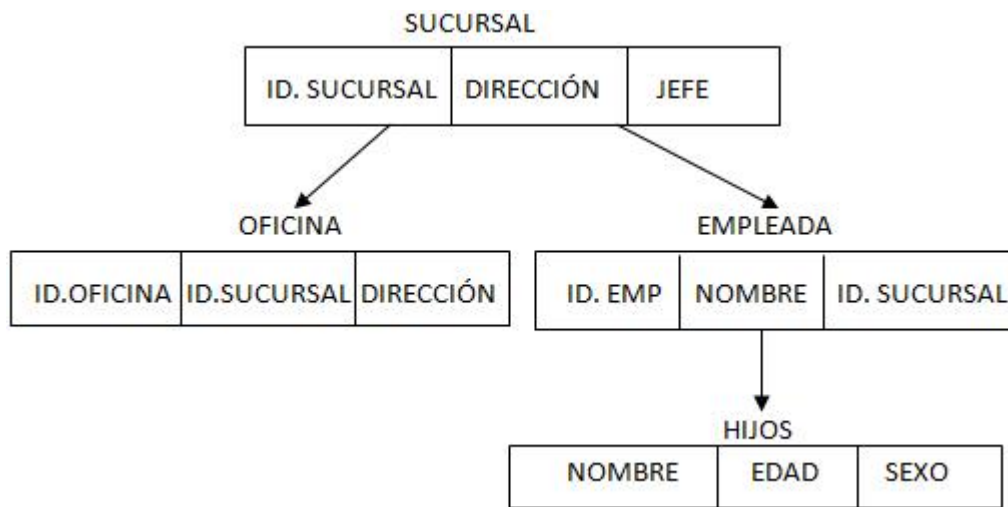
- **Modelo entidad-relación:** esta o en una percepción del mundo real que consta de una colección de éstos básicos, llamados entidades, y niveles relaciones entre esos objetos. Una entidad es una cosa hubo gesto en el mundo real que distinguible de otros objetos.
- **Modelo relacional:** Este modelo es el más utilizado actualmente ya que utiliza tablas bidimensionales para la representación lógica de los datos y sus relaciones. La tabla está compuesta por varias columnas, y cada columna tiene un nombre único.

Algunas de sus principales características son:

- Puede ser entendido y usado por cualquier usuario.
- Permite ampliar el esquema conceptual sin modificar las aplicaciones de gestión.
- Los usuarios no necesitan saber donde se encuentran los datos físicamente.
- El elemento principal de este modelo es la relación que se representa mediante una tabla.
- **Otros modelos de datos:** el modelo de datos orientado objetos es otro modelo de datos que se puede observar como una extensión del modelo E-R con las nociones de encapsulación, métodos e identidad de objeto. El modelo de datos relacional orientado objetos combinan las características del modelo de datos orientado objetos y el modelo de datos relacional. El modelo de datos de red y el modelo de datos jerárquico, precedieron al modelo de datos relacional.

Como el modelo Jerárquico: Este modelo utiliza árboles para la representación lógica de los datos. Este árbol está compuesto de unos elementos llamados nodos. El nivel más alto del árbol se denomina raíz. Cada nodo representa un registro con sus correspondientes campos.

La representación gráfica de este modelo se realiza mediante la creación de un árbol invertido, los diferentes niveles quedan unidos mediante relaciones.



En este modelo solo se pueden representar relaciones 1:M, por lo que presenta varios inconvenientes:

- No se admiten relaciones N:M
- Un segmento hijo no puede tener más de un padre.
- No se permiten más de una relación entre dos segmentos.
- Para acceder a cualquier segmento es necesario comenzar por el segmento raíz
- El árbol se debe de recorrer en el orden designado.

Modelo de datos en red

En este modelo las entidades se representan como nodos y sus relaciones son las líneas que los unen. En esta estructura cualquier componente puede relacionarse con cualquier otro.

A diferencia del modelo jerárquico, en este modelo, un hijo puede tener varios padres. Los conceptos básicos en el modelo en red son:

- El tipo de registro, que representa un nodo.
- Elemento, que es un campo de datos.
- Agregado de datos, que define un conjunto de datos con nombre.

Este modelo de datos permite representar relaciones N:M

1.4 Usuarios y administradores de la Base de Datos

Las personas que trabajan con una base de datos se pueden catalogar como usuarios de bases de datos o como administradores de bases de datos.

1.4.1 Usuarios de bases de datos e interfaces de usuarios

Hay cuatro tipos diferentes de usuarios de un sistema de bases de datos, diferenciados por la forma en que ellos esperan interactuar con el sistema.

Usuarios normales
programadores de aplicaciones
usuarios sofisticados
usuarios especializados

1.4.2 Administrador de la base de datos

Una de las principales razones de usar SGBDs es tener un control centralizado tanto de los datos como de los programas que acceden a esos datos la persona que tiene este control central sobre el sistema se llama administrador de la base de datos (ABD). Las funciones del ABD incluyen las siguientes:

- Definición del esquema.
- Definición de la estructura y del método de acceso.
- Modificación del esquema y de la organización física.
- Concesión de autorización para el acceso a los datos.
- Mantenimiento rutinario.

1.5 Gestión de transacciones y de Almacenamiento

Una transacción es una colección de operaciones que se lleva a cabo como una única función lógica. Es responsabilidad del programador definir adecuadamente las diferentes transacciones, de tal manera que cada una preserve la consistencia de la base de datos. Asegurar la consistencia de las propiedades es responsabilidad del propio sistema de bases de datos, específicamente del componente de gestión de transacciones. El sistema de bases de datos debe realizar la recuperación de fallos, es decir, detectar los fallos de sistema y restaurar la base de datos al estado que existía antes de que ocurriera el fallo. Finalmente, cuando varias transacciones actualizan la base de datos concurrentemente, la consistencia de los datos puede no ser preservada, incluso aunque cada transacción individualmente sea correcta. Es responsabilidad del gestor de control de concurrencia controlar la interacción entre las transacciones concurrentes para asegurar la consistencia de la base de datos.

1.6 Estructura de un sistema de bases de datos

Un sistema de bases de datos se divide en módulos que se encargan de cada una de las responsabilidades del sistema completo. Los componentes funcionales de un sistema de bases de datos se pueden dividir a grandes rasgos en los componentes, gestor de almacenamiento y procesador de consultas.

1.6.1 Gestor de almacenamiento

Es un módulo de programas que proporciona la interfaces entre los datos de bajo nivel en la base de datos y los programas de aplicación y consultas emitidas al sistema el gestor de almacenamiento es responsable de la interacción con el gestor de archivos. Los

componentes del gestor de almacenamiento incluyen:

- Gestor de autorización e integridad, que comprueba que se satisfagan las restricciones de integridad y la autorización de los usuarios para acceder a los datos.
- Gestor de transacciones, que asegura que la base de datos quiere un estado consistente a pesar de los fallos del sistema, y que las ejecuciones de transacciones concurrentes ocurren sin conflictos.
- Gestor de archivos, que gestiona la reserva de espacio de almacenamiento de disco y las estructuras de datos usadas para representar la información almacenada en disco.
- Gestor de memoria intermedia, que es responsable de traer los datos del disco de almacenamiento a memoria principal y describir que datos tratar en memoria caché.

El gestor de almacenamiento implementa varias estructuras de datos como parte de la implementación física el sistema:

Archivos de datos.
Diccionario de datos.
Índices.

1.6.2 Procesador de consultas

Los componentes del procesador de consultas incluyen:

- Intérprete del LDD, que interpreta las instrucciones del LDD y registra las definiciones en el diccionario de datos. **Para definir los esquemas de la base de datos.**
- Compilador del LMD, que traduce las instrucciones del LMD en un lenguaje consultas a un plan de evaluación que consiste en instrucciones de bajo nivel que entiende el motor de evaluación de consultas. **Para manipular los datos de la base de datos.**
- Motor de evaluación de consultas, que ejecutan las instrucciones de bajo nivel generadas por el compilador del LMD.
- Lenguaje de control de datos (DCL). Para la administración de usuarios y seguridad en la base de datos.

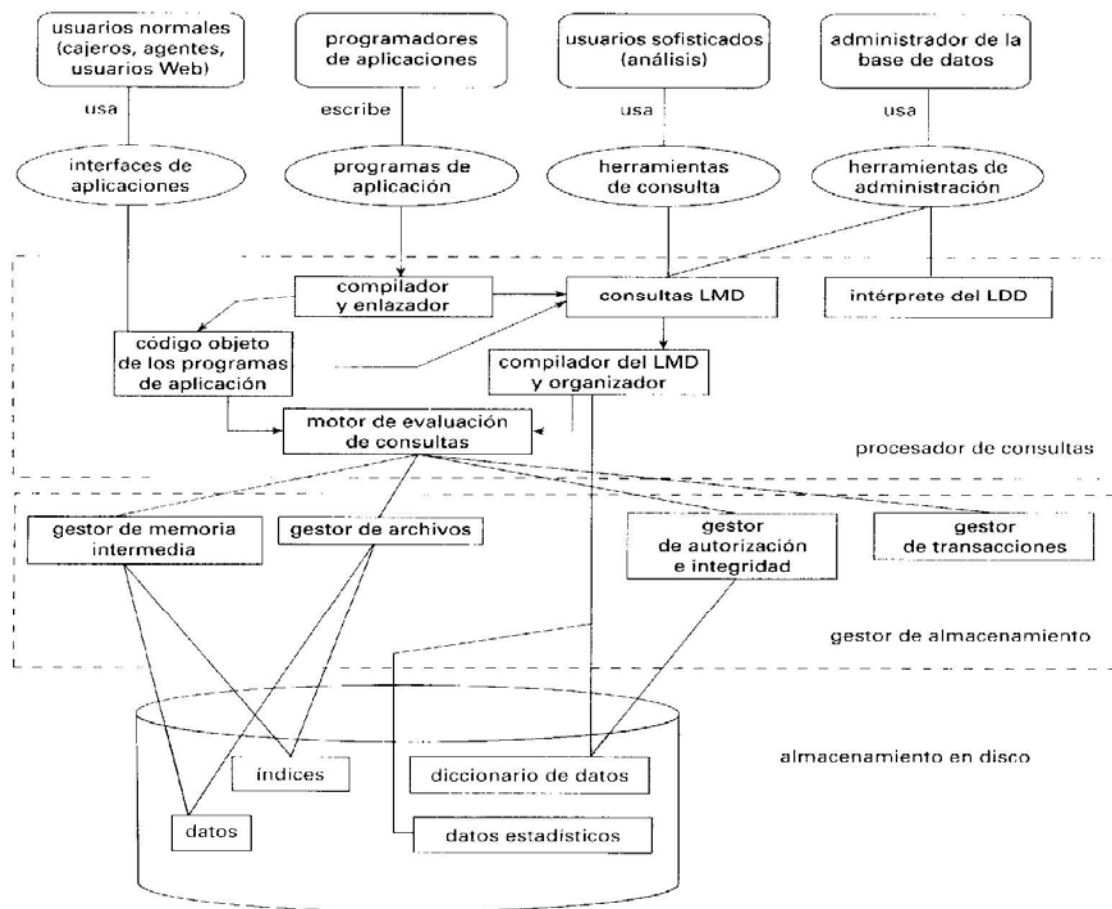


Figura 1.2 Estructura de un sistema Completo

1.7 Jerarquía de los datos campo, registro, archivos y Bases de datos

Los datos se organizan en una jerarquía que inician con una unidad mínima de información que es utilizada por los computadores denominado bit, y se desplaza a través de una escala hasta una base de datos. Un bit es un dígito binario que representa un circuito que puede estar encendido o apagado. Los bits se pueden organizar en unidades denominadas bytes. Un byte esta compuesto por ocho bits, cada byte representa un carácter, que es la unidad de información básica, un carácter puede consistir de letras mayúsculas (A, B, C,...Z), letras minúsculas (a,b,c,.....z), dígitos numéricos (0,1,2,....9), o símbolos especiales(![+]^*[-]/...).

Los caracteres se reúnen para formar un campo. Un campo por lo general es un nombre, un número o una combinación de caracteres que describen un aspecto de un objeto (por ejemplo, un empleado, una ubicación, etc.), o una actividad (Una venta). Un grupo de campos relacionados representan un registro. Al combinar descripciones de varios aspectos de un objeto o actividad, se obtiene una descripción mas completa de estos. Por ejemplo, el registro de un empleado es una compilación de campos referentes a ese individuo. El nombre del empleado, su dirección, su número telefónico, sueldo, etc., representarían cada

uno un campo. Un conjunto de registros relacionados representan un archivo, por ejemplo el archivo de empleados es la recopilación de todos los registros de los empleados de la compañía. De igual forma, un archivo de inventarios es la suma de todos los registros de inventarios de una compañía u organización en particular. Para el manejo de los archivos las aplicaciones se refieren a ellos como tablas.

En el nivel más alto de la jerarquía se encuentran las bases de datos, una recopilación de archivos integrados y relacionados. Bits, caracteres, campos, registros, archivos y bases de datos forman la jerarquía de datos.

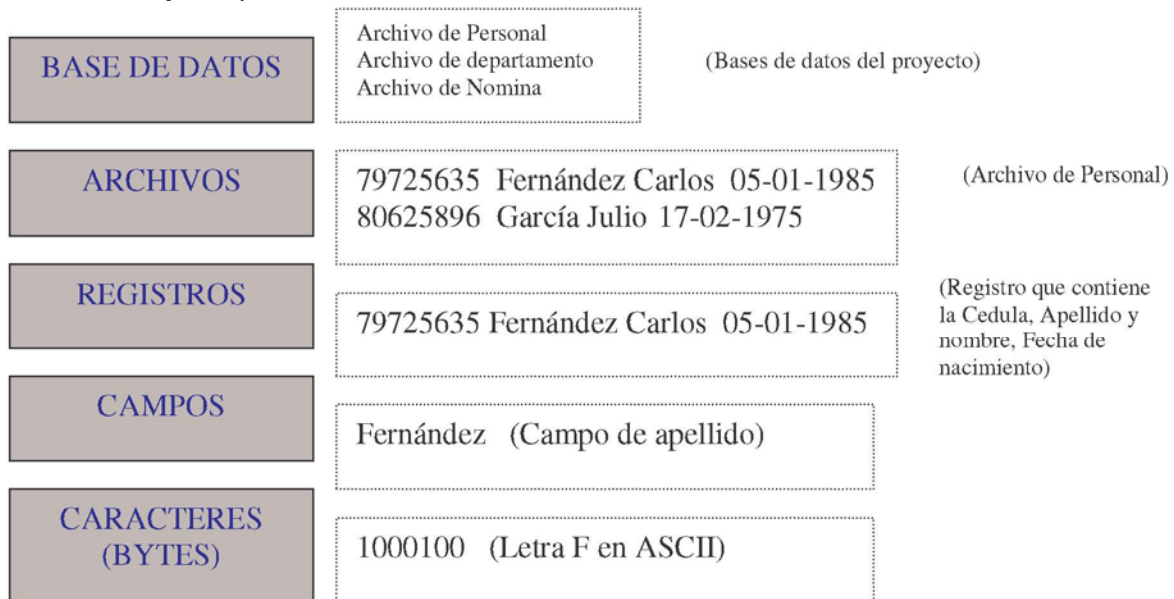


Figura 1.3.
La Jerarquía de los datos

1.8 Bases de datos distribuidas y centralizadas

1.8.1 Bases de datos centralizadas

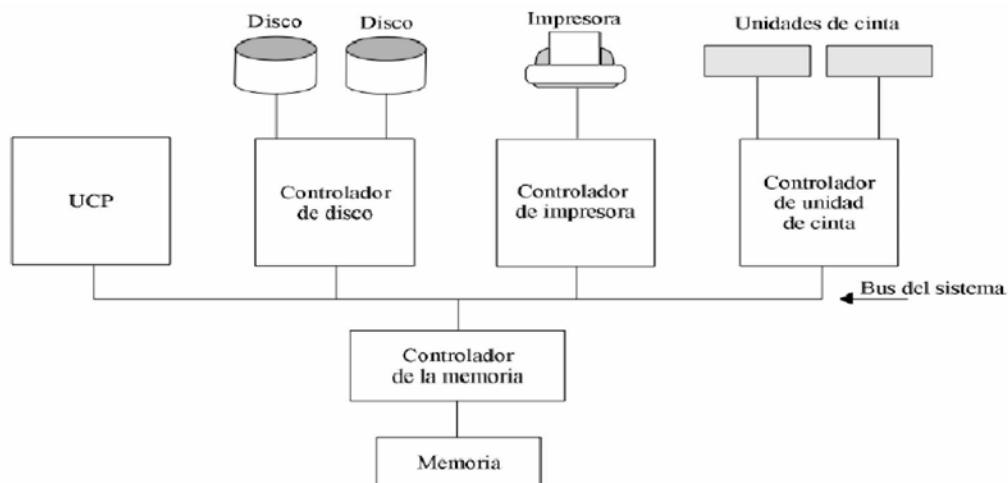


Figura 1.4 Sistema de información Centralizado

Los sistemas de bases de datos centralizados son aquellos que se ejecutan en un único sistema informático sin interaccionar con ninguna otra computadora. Tales sistemas comprenden el rango desde los sistemas de bases de datos monousuario ejecutándose en computadoras personales hasta los sistemas de bases de datos de alto rendimiento ejecutándose en grandes sistemas.

Una computadora moderna de propósito general consiste en una o unas pocas CPU's y un número determinado de controladores para los dispositivos que se encuentren conectados a través de un bus común, el cual proporciona acceso a la memoria compartida. Las CPU's poseen memorias caché locales donde se almacenan copias de ciertas partes de la memoria para acelerar el acceso a los datos. Cada controlador de dispositivo se encarga de un tipo específico de dispositivos (por ejemplo, una unidad de disco, una tarjeta de sonido o un monitor). La CPU y los controladores de dispositivo pueden ejecutarse concurrentemente, compitiendo así por el acceso a la memoria. La memoria caché reduce el acceso a la memoria, ya que la CPU necesita acceder a la memoria compartida un número de veces menor. Se distinguen dos formas de utilizar las computadoras: como sistemas monousuario o como sistemas multiusuario. En la primera categoría están las computadoras personales y las estaciones de trabajo. Un sistema monousuario típico es una unidad de sobremesa utilizada por una única persona que dispone de una sola CPU, de uno o dos discos fijos y que trabaja con un sistema operativo que sólo permite un único usuario. Por el contrario, un sistema multiusuario típico tiene más discos y más memoria, puede disponer de varias CPU y trabaja con un sistema operativo multiusuario. Se encarga de dar servicio a un gran número de usuarios que están conectados al sistema a través de terminales. Estos sistemas se denominan con frecuencia sistemas servidores.

Normalmente, los sistemas de bases de datos diseñados para funcionar sobre sistemas monousuario, como las computadoras personales, no suelen proporcionar muchas de las facilidades que ofrecen los sistemas multiusuario. En particular, no tienen control de concurrencia, que no es necesario cuando solamente un usuario puede generar modificaciones. Las facilidades de recuperación en estos sistemas, o no existen o son primitivas; por ejemplo, realizar una copia de seguridad de la base de datos antes de cualquier modificación. La mayoría de estos sistemas no admiten SQL y proporcionan un lenguaje de consulta muy simple, que en algunos casos es una variante de QBE (Query By Example).

Aunque hoy en día las computadoras de propósito general tienen varios procesadores, utilizan paralelismo de grano grueso, disponiendo de unos pocos procesadores (normalmente dos o cuatro) que comparten la misma memoria principal. Las bases de datos que se ejecutan en tales máquinas habitualmente no intentan dividir una consulta simple entre los distintos procesadores, sino que ejecutan cada consulta en un único procesador, posibilitando la concurrencia de varias consultas. Así, estos sistemas soportan una mayor productividad, es decir, permiten ejecutar un mayor número de transacciones por segundo, a pesar de que cada transacción individualmente no se ejecuta más rápido.

Las bases de datos diseñadas para las máquinas monoprocesador ya disponen de multitarea, permitiendo que varios procesos se ejecuten a la vez en el mismo procesador, usando tiempo compartido, mientras que de cara al usuario parece que los procesos se están ejecutando en paralelo.

1.8.2 Bases de datos Distribuidas

Los sistemas de bases de datos distribuidas son un caso particular de los sistemas de cómputo distribuido en los cuales un conjunto de elementos de procesamiento autónomos (no necesariamente homogéneos) se interconectan por una red de comunicaciones y cooperan entre ellos para realizar sus tareas asignadas. Entre los términos más comunes que se utilizan para referirse al cómputo distribuido podemos encontrar: funciones distribuidas, procesamiento distribuido de datos, multiprocesadores, multicomputadoras, procesamiento satelital, procesamiento tipo "backend", computadoras dedicadas y de propósito específico, sistemas de tiempo compartido, sistemas funcionalmente modulares.

En el cómputo distribuido existen muchos componentes para realizar una tarea. Los elementos que se pueden distribuir son:

Control. Las actividades relacionadas con el manejo o administración del sistema.

Datos. La información que maneja el sistema.

Funciones. Las actividades que cada elemento del sistema realiza.

Procesamiento lógico. Las tareas específicas involucradas en una actividad de procesamiento de información.

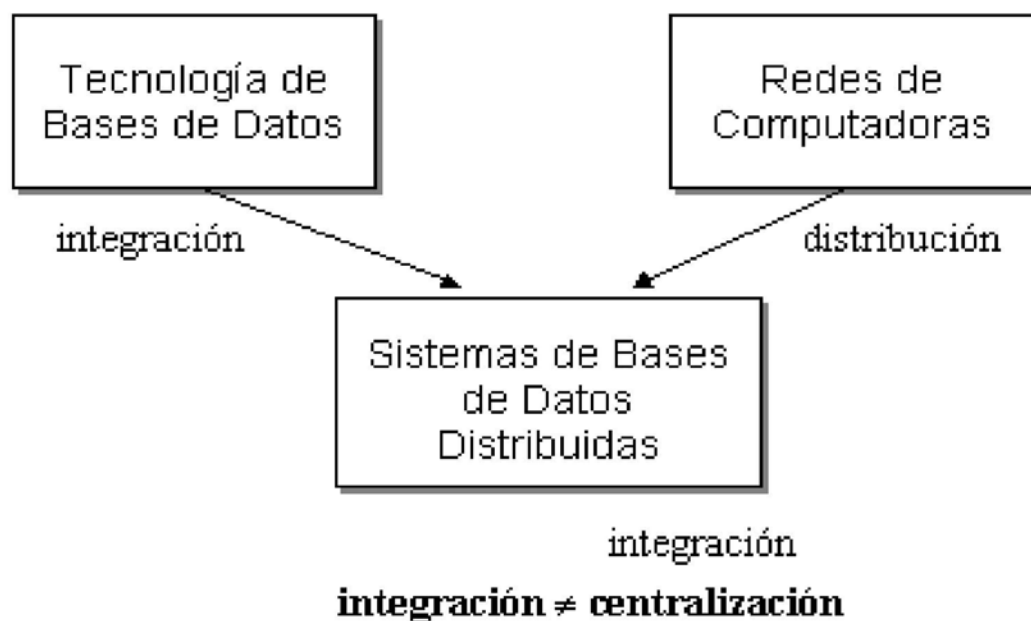


Figura 1.5 Motivación de los SBDD

Una base de datos distribuida (BDD) es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas entre diferentes sitios interconectados por una red de comunicaciones.

Un sistema de bases de datos distribuida (SBDD) es un sistema en el cual múltiples sitios de bases de datos están ligados por un sistema de comunicaciones, de tal forma que, un usuario en cualquier sitio puede acceder los datos en cualquier parte de la red exactamente como si los datos estuvieran almacenados en su sitio propio.

Un sistema de manejo de bases de datos distribuidas (SMBDD) es aquel que se encarga del manejo de la BDD y proporciona un mecanismo de acceso que hace que la distribución sea transparente a los usuarios. El término transparente significa que la aplicación trabajaría, desde un punto de vista lógico, como si un solo SMBD ejecutado en una sola máquina, administrara esos datos.

Un sistema de base de datos distribuida (SBDD) es entonces el resultado de la integración de una base de datos distribuida con un sistema para su manejo.

Dada la definición anterior, es claro que algunos sistemas no se pueden considerar como SBDD. Por ejemplo, un sistema de tiempo compartido no incluye necesariamente un sistema de manejo de bases de datos y, en caso de que lo haga, éste es controlado y administrado por una sola computadora.

Un sistema de multiprocesamiento puede administrar una base de datos pero lo hace usualmente a través de un solo sistema de manejo de base de datos; los procesadores se utilizan para distribuir la carga de trabajo del sistema completo o incluso del propio SMBD pero actuando sobre una sola base de datos. Finalmente, una base de datos la cual reside en un solo sitio de una red de computadoras y que es accesada por todos los nodos de la red no es una base de datos distribuida. Este caso se trata de una base de datos cuyo control y administración esta centralizada en un solo nodo pero se permite el acceso a ella a través de la red de computadoras.

El medio ambiente típico de un SMBDD consiste de un conjunto de sitios o nodos los cuales tiene un sistema de procesamiento de datos completo que incluye una base de datos local, un sistema de manejo de bases de datos y facilidades de comunicaciones. Si los diferentes sitios pueden estar geográficamente dispersos, entonces, ellos están interconectados por una red de tipo WAN. Por otro lado, si los sitios están localizados en diferentes edificios o departamentos de una misma organización pero geográficamente en la misma ubicación, entonces, están conectados por una red local (LAN).

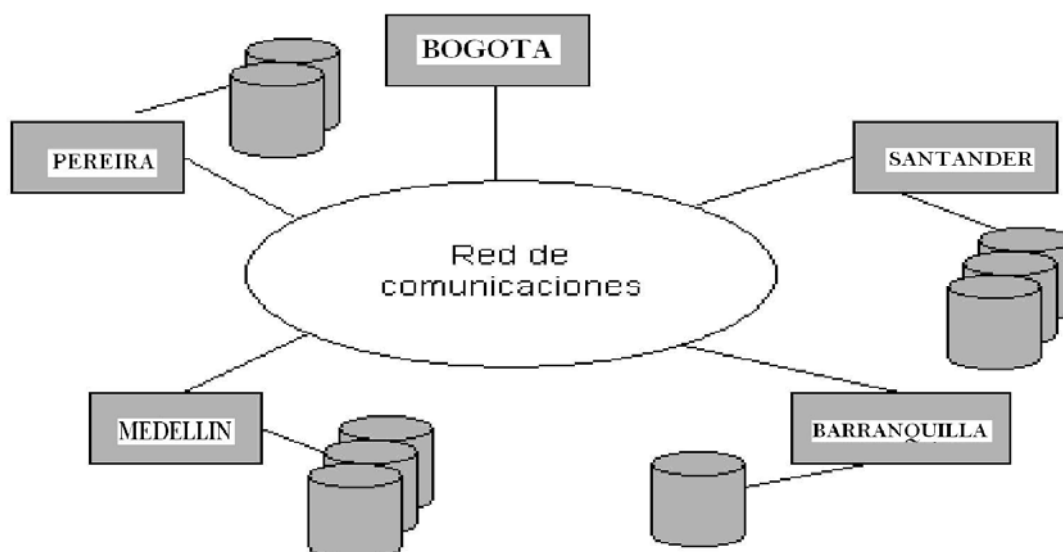


Figura 1.6 Medio ambiente de SBDD.

1.8.2.1 Aplicaciones

Los ambientes en los que se encuentra con mayor frecuencia el uso de las bases de datos distribuidas son:

- Cualquier organización que tiene una estructura descentralizada.
- Casos típicos de lo anterior son: organismos gubernamentales y/o de servicio público.
- La industria de la manufactura, particularmente, aquella con plantas múltiples. Por ejemplo, la industria automotriz.
- Aplicaciones de control y comando militar.
- Líneas de transportación aérea.
- Cadenas hoteleras.
- Servicios bancarios y financieros.

1.8.2.2 Ventajas

Los SMBDD tienen múltiples ventajas. En primer lugar los datos son localizados en lugar más cercano, por tanto, el acceso es más rápido, el procesamiento es rápido debido a que varios nodos intervienen en el procesamiento de una carga de trabajo, nuevos nodos se pueden agregar fácil y rápidamente. La comunicación entre nodos se mejora, los costos de operación se reducen, son amigables al usuario, la probabilidad de que una falla en un solo nodo afecte al sistema es baja y existe una autonomía e independencia entre los nodos.

Las razones por las que compañías y negocios migran hacia bases de datos distribuidas incluyen razones organizacionales y económicas, para obtener una interconexión confiable y flexible con las bases de datos existentes, y por un crecimiento futuro. El enfoque distribuido de las bases de datos se adapta más naturalmente a la estructura de las organizaciones. Además, la necesidad de desarrollar una aplicación global (que incluya a toda la organización), se resuelve fácilmente con bases de datos distribuidas. Si una organización crece por medio de la creación de unidades o departamentos nuevos, entonces, el enfoque de bases de datos distribuidas permite un crecimiento suave.

Los datos se pueden colocar físicamente en el lugar donde se accesan más frecuentemente, haciendo que los usuarios tengan control local de los datos con los que interactúan. Esto resulta en una autonomía local de datos permitiendo a los usuarios aplicar políticas locales respecto del tipo de accesos a sus datos.

Mediante la replicación de información, las bases de datos distribuidas pueden presentar cierto grado de tolerancia a fallas haciendo que el funcionamiento del sistema no dependa de un solo lugar como en el caso de las bases de datos centralizadas.

1.8.2.3 Desventajas

La principal desventaja se refiere al control y manejo de los datos. Dado que éstos residen en muchos nodos diferentes y se pueden consultar por nodos diversos de la red, la

probabilidad de violaciones de seguridad es creciente si no se toman las precauciones debidas.

La habilidad para asegurar la integridad de la información en presencia de fallas no predecibles tanto de componentes de hardware como de software es compleja. La integridad se refiere a la consistencia, validez y exactitud de la información.

Dado que los datos pueden estar replicados, el control de concurrencia y los mecanismos de recuperación son mucho más complejos que en un sistema centralizado.

1.9 Componentes de sistemas de bases de datos

Los SGBD son paquetes de software muy complejos y sofisticados que deben proporcionar los servicios para el buen funcionamiento de la base de datos. No se puede generalizar sobre los elementos que componen un SGBD ya que varían mucho unos de otros. 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.

Un SGBD tiene varios módulos, cada uno de los cuales realiza una función específica. El sistema operativo proporciona servicios básicos al SGBD, que es construido sobre él.

- El procesador de consultas es el componente principal de un SGBD. Transforma las consultas en un conjunto de instrucciones de bajo nivel que se dirigen al gestor de la base de datos.
- El gestor de la base de datos es el interface con los programas de aplicación y las consultas de los usuarios. El gestor de la base de datos acepta consultas y examina los esquemas externo y conceptual para determinar qué registros se requieren para satisfacer la petición. Entonces el gestor de la base de datos realiza una llamada al gestor de ficheros para ejecutar la petición.
- El gestor de ficheros maneja los ficheros en disco en donde se almacena la base de datos. Este gestor establece y mantiene la lista de estructuras e índices definidos en el esquema interno. Si se utilizan ficheros dispersos, llama a la función de dispersión para generar la dirección de los registros. Pero el gestor de ficheros no realiza directamente la entrada y salida de datos. Lo que hace es pasar la petición a los métodos de acceso del sistema operativo que se encargan de leer o escribir los datos en el buffer del sistema.
- El preprocesador del LMD convierte las sentencias del LMD embebidas en los programas de aplicación, en llamadas a funciones estándar escritas en el lenguaje anfitrión. El preprocesador del LMD debe trabajar con el procesador de consultas para generar el código apropiado.
- El compilador del LDD convierte las sentencias del LDD en un conjunto de tablas que contienen metadatos. Estas tablas se almacenan en el diccionario de datos.
- El gestor del diccionario controla los accesos al diccionario de datos y se encarga de mantenerlo. La mayoría de los componentes del SGBD acceden al diccionario de datos.

Los principales componentes del gestor de la base de datos son los siguientes:

- Control de autorización. Este módulo comprueba que el usuario tiene los permisos necesarios para llevar a cabo la operación que solicita.
- Procesador de comandos. Una vez que el sistema ha comprobado los permisos del usuario, se pasa el control al procesador de comandos.
- Control de la integridad. Cuando una operación cambia los datos de la base de datos, este módulo debe comprobar que la operación a realizar satisface todas las restricciones de integridad necesarias.
- Optimizador de consultas. Este módulo determina la estrategia óptima para la ejecución de las consultas.
- Gestor de transacciones. Este módulo realiza el procesamiento de las transacciones.
- Planificador (scheduler). Este módulo es el responsable de asegurar que las operaciones que se realizan concurrentemente sobre la base de datos tienen lugar sin conflictos.
- Gestor de recuperación. Este módulo garantiza que la base de datos permanece en un estado consistente en caso de que se produzca algún fallo.
- Gestor de buffers. Este módulo es el responsable de transferir los datos entre memoria principal y los dispositivos de almacenamiento secundario. A este módulo también se le denomina gestor de datos.

1.10 Funciones de sistemas manejadores de bases de datos

Codd, el creador del modelo relacional, ha establecido una lista con los ocho servicios que debe ofrecer todo SGBD.

1. Un SGBD debe proporcionar a los usuarios la capacidad de almacenar datos en la base de datos, acceder a ellos y actualizarlos. Esta es la función fundamental de un SGBD y por supuesto, el SGBD debe ocultar al usuario la estructura física interna (la organización de los ficheros y las estructuras de almacenamiento).
2. Un SGBD debe proporcionar un catálogo en el que se almacenen las descripciones de los datos y que 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 almacena:
 - Nombre, tipo y tamaño de los datos.
 - Nombre de las relaciones entre los datos.
 - Restricciones de integridad sobre los datos.
 - Nombre de los usuarios autorizados a acceder a la base de datos.
 - Esquema externo, conceptual e interno, y correspondencia entre los esquemas.
 - Estadísticas de utilización, tales como la frecuencia de las transacciones y el número de accesos realizados a los objetos de la base de datos.

Algunos de los beneficios que reporta el diccionario de datos son los siguientes:

- La información sobre los datos se puede almacenar de un modo centralizado. Esto ayuda a mantener el control sobre los datos, como un recurso que son.
- El significado de los datos se puede definir, lo que ayudará a los usuarios a entender el propósito de los mismos.

- La comunicación se simplifica ya que se almacena el significado exacto. El diccionario de datos también puede identificar al usuario o usuarios que poseen los datos o que los acceden.
 - Las redundancias y las inconsistencias se pueden identificar más fácilmente ya que los datos están centralizados.
 - Se puede tener un historial de los cambios realizados sobre la base de datos.
 - El impacto que puede producir un cambio se puede determinar antes de que sea implementado, ya que el diccionario de datos mantiene información sobre cada tipo de dato, todas sus relaciones y todos sus usuarios.
 - Se puede hacer respetar la seguridad.
 - Se puede garantizar la integridad.
 - Se puede proporcionar información para auditorías.
3. Un SGBD debe proporcionar 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. Una transacción en el sistema informático de la empresa inmobiliaria sería dar de alta a un empleado o eliminar un inmueble. Una transacción un poco más complicada sería eliminar un empleado y reasignar sus inmuebles a otro empleado. En este caso hay que realizar varios cambios sobre la base de datos. Si la transacción falla durante su realización, por ejemplo porque falla el hardware, la base de datos quedará en un estado inconsistente. Algunos de los cambios se habrán hecho y otros no, por lo tanto, los cambios realizados deberán ser deshechos para devolver la base de datos a un estado consistente.
 4. Un SGBD debe proporcionar un mecanismo que asegure que la base de datos se actualice correctamente cuando varios usuarios la están actualizando concurrentemente. Uno de los principales objetivos de los SGBD es el permitir que varios usuarios tengan acceso concurrente a los datos que comparten. El acceso concurrente es relativamente fácil de gestionar si todos los usuarios se dedican a leer datos, ya que no pueden interferir unos con otros. Sin embargo, cuando dos o más usuarios están accediendo a la base de datos y al menos uno de ellos está actualizando datos, pueden interferir de modo que se produzcan inconsistencias en la base de datos. El SGBD se debe encargar de que estas interferencias no se produzcan en el acceso simultáneo.
 5. Un SGBD debe proporcionar un mecanismo capaz de recuperar la base de datos en caso de que ocurra algún suceso que la dañe. Como se ha comentado antes, cuando el sistema falla en medio de una transacción, la base de datos se debe devolver a un estado consistente. Este fallo puede ser a causa de un fallo en algún dispositivo hardware o un error del software, que hagan que el SGBD aborte, o puede ser a causa de que el usuario detecte un error durante la transacción y la aborte antes de que finalice. En todos estos casos, el SGBD debe proporcionar un mecanismo capaz de recuperar la base de datos llevándola a un estado consistente.
 6. Un SGBD debe proporcionar un mecanismo que garantice que sólo los usuarios autorizados pueden acceder a la base de datos. La protección debe ser contra accesos no autorizados, tanto intencionados como accidentales.
 7. Un SGBD debe 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.

8. Un SGBD debe 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. La integridad de la base de datos requiere la validez y consistencia de los datos almacenados. Se puede considerar como otro modo de proteger la base de datos, pero además de tener que ver con la seguridad, tiene otras implicaciones. La integridad se ocupa de la calidad de los datos. Normalmente se expresa mediante restricciones, que son una serie de reglas que la base de datos no puede violar. Por ejemplo, se puede establecer la restricción de que cada empleado no puede tener asignados más de diez inmuebles. En este caso sería deseable que el SGBD controlara que no se sobrepase este límite cada vez que se asigne un inmueble a un empleado.

Además, de estos ocho servicios, es razonable esperar que los SGBD proporcionen un par de servicios más:

1. Un SGBD debe permitir que se mantenga la independencia entre los programas y la estructura de la base de datos. La independencia de datos se alcanza mediante las vistas o subesquemas. La independencia de datos física es más fácil de alcanzar, de hecho hay varios tipos de cambios que se pueden realizar sobre la estructura física de la base de datos sin afectar a las vistas. Sin embargo, lograr una completa independencia de datos lógica es más difícil. Añadir una nueva entidad, un atributo o una relación puede ser sencillo, pero no es tan sencillo eliminarlos.
2. Un SGBD debe proporcionar una serie de herramientas que permitan administrar la base de datos de modo efectivo. Algunas herramientas trabajan a nivel externo, por lo que habrán sido producidas por el administrador de la base de datos. Las herramientas que trabajan a nivel interno deben ser proporcionadas por el distribuidor del SGBD. Algunas de ellas son:
 - Herramientas para importar y exportar datos.
 - Herramientas para monitorizar el uso y el funcionamiento de la base de datos.
 - Programas de análisis estadístico para examinar las prestaciones o las estadísticas de utilización.
 - Herramientas para reorganización de índices.
 - Herramientas para aprovechar el espacio dejado en el almacenamiento físico por los registros borrados y que consoliden el espacio liberado para reutilizarlo cuando sea necesario.

1.11 Arquitecturas 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:

1. 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.
2. 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.
3. 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 determinados y oculta a ese grupo el resto de la base de datos. En este nivel se puede utilizar un modelo conceptual o un modelo lógico para especificar los esquemas.

La mayoría de los SGBD no distinguen del todo los tres niveles. Algunos incluyen detalles del nivel físico en el esquema conceptual. En casi todos los SGBD que se manejan vistas de usuario, los esquemas externos se especifican con el mismo modelo de datos que describe la información a nivel conceptual, aunque en algunos se pueden utilizar diferentes modelos de datos en los niveles conceptual y externo.

Hay que destacar que los tres esquemas no son más que descripciones de los mismos datos pero con distintos niveles de abstracción. Los únicos datos que existen realmente están a nivel físico, almacenados en un dispositivo como puede ser un disco. En un SGBD basado en la arquitectura de tres niveles, cada grupo de usuarios hace referencia exclusivamente a su propio esquema externo. Por lo tanto, el SGBD debe transformar cualquier petición expresada en términos de un esquema externo a una petición expresada en términos del esquema conceptual, y luego, a una petición en el esquema interno, que se procesará sobre la base de datos almacenada. Si la petición es de una obtención (consulta) de datos, será preciso modificar el formato de la información extraída de la base de datos almacenada, para que coincida con la vista externa del usuario. El proceso de transformar peticiones y resultados de un nivel a otro se denomina correspondencia o transformación. Estas correspondencias pueden requerir bastante tiempo, por lo que algunos SGBD no cuentan con vistas externas.

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 sólo a la separación entre las aplicaciones y las estructuras físicas de almacenamiento, es más fácil de conseguir que la independencia lógica.

En los SGBD que tienen la arquitectura de varios niveles es necesario ampliar el catálogo o diccionario, de modo que incluya información sobre cómo establecer la correspondencia entre las peticiones de los usuarios y los datos, entre los diversos niveles. El SGBD utiliza una serie de procedimientos adicionales para realizar estas correspondencias haciendo referencia a la información de correspondencia que se encuentra en el catálogo. La independencia de datos se consigue porque al modificarse el esquema en algún nivel, el esquema del nivel inmediato superior permanece sin cambios, sólo se modifica la correspondencia entre los dos niveles. No es preciso modificar los programas de aplicación que hacen referencia al esquema del nivel superior.

Por lo tanto, la arquitectura de tres niveles puede facilitar la obtención de la verdadera independencia de datos, tanto física como lógica. Sin embargo, los dos niveles de correspondencia implican un gasto extra durante la ejecución de una consulta o de un programa, lo cual reduce la eficiencia del SGBD. Es por esto que muy pocos SGBD han implementado esta arquitectura completa.

1.12 ACTIVIDADES COMPLEMENTARIAS

- 1 ¿Cuáles son las cuatro diferencias principales entre un sistema de procesamiento de archivos y un SGDB?
- 2 Investigue y determine cuáles son los inconvenientes de un sistema gestor de base de datos.
- 3 Elabore un cuadro y explique la diferencia entre independencia de datos física y lógica.
- 4 Explique las cinco responsabilidades del sistema gestor de la base de datos, ¿Que ocurriría si no se realizara alguna de estas funciones?
- 5 ¿Cuáles son las cinco funciones principales del administrador de la base de datos?
- 6 Investigue y amplíe la información relacionada con la arquitectura de dos y tres capas.
- 7 Explique mediante un cuadro comparativo las ventajas y desventajas entre una base de datos centralizada y una base de datos distribuida.

CONSULTAS WEB

- La página inicial del grupo especial de interés de la ACM en gestión de datos proporciona una gran cantidad de información sobre la investigación en bases de datos. <http://www.acm.org/>