***Unidad 1: Almacenamiento de la información***

🡪 Preguntas de desarrollo

1. **INTRODUCCIÓN**

**2. Los ficheros de información.**

**2.1 ¿Qué es un fichero de datos?**

- La informática adaptó sus herramientas para que los elementos que el usuario manejaba en el ordenador se parecieran a los que utilizaba manualmente.

- El elemento que permitió llevar a cabo el almacenamiento de datos de forma permanente en dispositivos de memoria masiva fue el fichero o archivo.

**- Fichero o archivo de datos:** conjunto de información relacionada, organizada y estructurada de forma que se permita la búsqueda de datos individuales. Son identificados por un nombre y la descripción de la carpeta que lo contiene.

- Los ficheros de datos están formados por registros lógicos que contienen datos relativos a un mismo elemento. A su vez, los registros están divididos en campos que contienen cada una de las informaciones elementales que forman un registro.

- Como los ficheros suelen ser muy voluminosos, solo se pueden llevar a la memoria principal parte de ellos para poder procesarlos. La cantidad de información que es transferida por el Sistema Operativo (SO) al registro físico o bloque.

- Al número de registros lógicos que entran en un bloque se le conoce con el nombre de factor de blocaje, y a esta operación de agrupar varios registros en un bloque se le llama bloqueo de registros.

**- Una clave o identificativo:** es un campo (o conjunto de campos) que identifica cada registro del fichero. En un mismo fichero puede haber una, varias o ninguna clave. Cuando una clave se utiliza como campo de localización en el fichero se denomina llave (key).

**2.2 Tipos de ficheros.**

Es fundamental saber la función de un fichero para decidir cómo se organiza.

Según esto, se pueden clasificar de la siguiente manera:

* **Ficheros permanentes –** Información relevante para una aplicación, vida larga y no pueden generarse inmediatamente a partir de otros ficheros. Se subdividen en:

**🡪Ficheros maestros –** Datos que pueden modificarse desde la aplicación. (Son el núcleo de la aplicación).

**🡪Ficheros constantes –** Incluyen datos fijos para la aplicación. No suelen ser modificados.

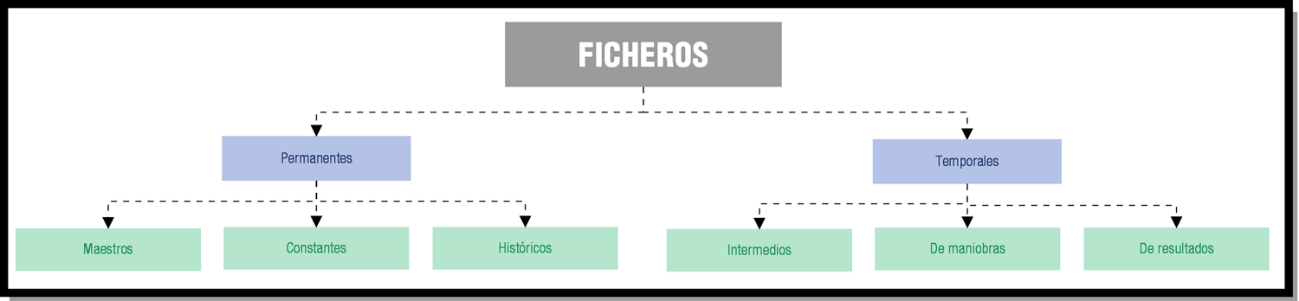
**🡪Ficheros históricos –**considerados como actuales en un período o situación anterior.

* **Ficheros temporales –** Almacenamiento de información útil para una parte de la aplicación. GENERADOS A PARTIR DE FICHEROS PERMANENTES. Tienen un período corto de existencia y se dividen en:

**🡪Ficheros intermedios –** Resultados aplicación que serán utilizados por otra.

**🡪Ficheros de maniobras –** Datos de una aplicación que no se pueden mantener en la memoria principal por espacio.

**🡪Ficheros de resultados –** Datos que van a ser transmitidos a un dispositivo de salida.



**2.3 Los soportes de información.**

- Los ficheros se almacenan en soportes de información manejados por dispositivos periféricos del ordenador, que permiten leer y grabar datos en el soporte.

- El tamaño de los ficheros está limitado por el tamaño de los dispositivos que los albergan y el sistema de archivos implantado.

- Se distinguen dos tipos de soportes para el almacenamiento de datos:

* **Soportes de acceso secuencial o no direccionables (ej.: cintas magnéticas):** se suelen usar en copias de seguridad y si se desea leer un dato que está en la mitad de la cinta, se tendrá que leer todo lo que hay hasta llegar a esa posición.
* **Soportes de acceso directo o direccionables (ej.: discos):** son los más empleados y el acceso a los datos puede hacerse de forma directa, pudiendo colocarnos en la posición que nos interese y leer a partir de ella (no es necesario recorrer o leer otros bloques)

**2.4 Parámetros de utilización.**

En función del uso que se le vaya a dar al fichero, serán adecuados unos tipos u otros de organización. Mediante la utilización de parámetros de referencia, podremos determinar el uso de un fichero. Estos parámetros son:

* **Capacidad o volumen:** es el espacio, en caracteres, que ocupa el fichero. La capacidad podrá calcularse multiplicando el número previsto de registros por la longitud media de cada registro.
* **Actividad:** permite conocer la cantidad de consultas y modificaciones que se realizan en el fichero. Para poder especificar la actividad se deben tener en cuenta:

• Tasa de consulta o modificación

• Frecuencia de consulta o modificación

* **Volatilidad:** mide la cantidad de inserciones y borrados que se efectúan en un fichero. Para determinar la volatilidad es necesario conocer:

**• Tasa de renovación:** es el porcentaje de registros renovados en cada tratamiento del fichero, respecto al número total de registros contenidos en él. **• Frecuencia de renovación:** es el número de veces que se accede al fichero para renovarlo en un periodo de tiempo fijo.

* **Crecimiento:** es la variación de la capacidad del fichero y se mide con la tasa de crecimiento.

**2.5 Organización de ficheros**

La información contenida en los tipos de ficheros ha cambiado mucho con objetivo de:

* Proporcionar un acceso rápido a los registros.
* Conseguir economizar el almacenamiento.
* Facilitar la actualización de los registros.
* Permitir que la estructura refleje la organización real de la información.

Hay distintas formas de estructurar u organizar los ficheros sobre un soporte de información. Las características de utilización del fichero dependen de la organización que se adopte. Se debe optar por una u otra atendiendo a la forma en que se va a usar el fichero. Tipos:

**2.5.1 Organización secuencial.**

Sus registros están almacenados de forma contigua.

Los registros almacenados se identifican por medio de una información ubicada en uno de sus campos, a este campo se le denomina clave o llave.

Otras características relevantes de los ficheros secuenciales son:

* La lectura siempre se realiza hacia delante.
* No permiten el acceso simultáneo de varios usuarios (monousuario).
* Tienen una estructura rígida de campos.
* El modo de apertura del fichero, condiciona la lectura o escritura.
* Aprovechan al máximo el soporte de almacenamiento.
* Se pueden grabar en secuenciales y direccionables.
* Todos los lenguajes.
* No se pueden insertar registros entre los que ya están grabados.

Si se necesita actualizar un archivo con organización secuencial se debe crear de nuevo el archivo (esto se realiza mediante un programa escrito para tal fin).

**2.5.2 Organización secuencial encadenada.**

Aquí los registros están conectados mediante un puntero (enlaces). Los ficheros con esta organización sólo pueden ser gestionados en soportes direccionables.Cada vez que se lee un registro se lee la posición del siguiente.

Para insertar un registro es necesario localizar la posición en que se desea insertar. El registro se crea en una zona libre y se modifican los punteros (enlaces) para que el registro quede insertado entre los dos registros deseados. Para borrar un registro se asigna al puntero del registro anterior la dirección del registro siguiente al que se desea borrar. El SO puede o no liberar el espacio ocupado por el registro borrado.

Al modificar, si no implica un aumento de la longitud del registro, éste puede reescribirse en el mismo espacio. En el caso de que el registro aumente de longitud se debe insertar el registro y posteriormente borrar la versión anterior a la modificación. La principal ventaja es su flexibilidad, y su principal inconveniente es su limitación a consulta secuencial.

**2.5.3 Organización secuencial indexada.**

Están formados por tres estructuras o zonas diferentes.

* **Zona de registros –** Registros grabados en un soporte de almacenamiento directo en secuencia ascendente, de acuerdo con los valores de la clave y en bloques de longitud fija.
* **Zona de índices –** Tabla que asocia claves con direcciones de los registros en el área de datos. Entrada formada con valor más alto de la clave de cada conjunto de registros y un puntero con la dirección del primer registro del grupo.
* **Zona de desbordamiento –** Sitio donde se graban los registros que no caben en la zona de registros.

La zona de registros está dividida en tramos lógicos, cada uno está formado por una serie de registros consecutivos. Por cada tramo de la zona de registros hay un registro en la zona de índices. Para permitir actualizaciones es necesario incluir en la estructura una zona de desbordamientos. Esta organización es muy usada para los procesos en los que intervienen pocos registros como en aquellos en los que se maneja el fichero completo.

**Las principales características son:**

* **Permiten acceso secuencial**
* **Permite el acceso directo a los registros**
* **Se pueden actualizar los registros en el mismo fichero**
* **Ocupa más espacio en el disco que los ficheros secuenciales**
* **Únicamente se pueden utilizar con soportes direccionables.**
* **Inversión económica más alta**.

**2.5.4 Organización directa o aleatoria.**

Se puede acceder a un registro indicando la posición relativa del mismo dentro del archivo o a través de una clave que forma parte del registro como un campo más. Estos archivos deben almacenarse en discos.

* **Campo clave:** permite identificar y localizar un registro de manera ágil y organizada.

La distribución de los registros es aleatoria dentro del soporte de almacenamiento. Para acceder a la posición física de un registro se utiliza una dirección o índice, no siendo necesario recorrer todo el fichero para encontrar un determinado registro.

A través de una transformación específica aplicada a la clave, se obtendrá la dirección física en la que se encuentra el registro. El problema fundamental de esta organización es la elección de la transformación que se ha de usar: direcciones que no corresponden a ninguna clave y direcciones que corresponden a más de una clave.

- Hay dos formas de resolver el problema de los sinónimos, siempre a costa de complicar la estructura del fichero:

1. Cuando se asocia a una clave una dirección ya ocupada por un registro distinto, se busca en el archivo, por algún procedimiento, una posición libre donde escribir el registro.

2. Se reserva una zona de desbordamiento donde se escribirán los registros que no se puedan escribir en la posición que les corresponde según la transformación. Según la forma de realizar esta transformación, existen diferentes modos de acceso:

**2.5.4.1** **Direccionamiento directo.** Se utiliza como dirección la propia clave.

**2.5.4.2 Direccionamiento asociado.** Se puede utilizar para cualquier tipo de clave. Si se utiliza este método debe construirse una tabla en la que figurarán todas las claves y la dirección donde se encuentra el registro correspondiente.

**2.5.4.3 Direccionamiento calculado** (hashing). Permite accesos más rápidos, se utiliza una transformación (función de hashing) conocida, que a partir de la clave genera la dirección de cada registro del archivo.

**- Problema:** que presenta este tipo de ficheros es que a partir de diferentes claves se obtenga la misma dirección al aplicar la función matemática. Para resolver este problema se aplican diferentes métodos crear una zona de sinónimos, un archivo de sinónimos, etc.

2.6 Inconvenientes de un sistema de gestión de ficheros

* **Redundancia e inconsistencia de los datos:** pueden tener distintos formatos y los datos pueden estar duplicados.
* **Dependencia de los datos física-lógica:** la estructura física de los datos se encuentra codificada en los programas.
* **Dificultad para tener acceso a los datos:** cada vez que se necesite una consulta no prevista es necesario codificar el programa correspondiente.
* **Separación y aislamiento de los datos:** es difícil escribir nuevos programas que aseguren la manipulación correcta de los datos.
* **Dificultad para el acceso concurrente:** es complicado que los usuarios actualicen los datos simultáneamente.
* **Dependencia de la estructura del archivo con el lenguaje de programación:** al haber distintos lenguajes de programación dificulta la modificación de los mismos.
* **Problemas en la seguridad de los datos:** difícil implantar restricciones de seguridad.
* **Problemas de integridad de datos:** los valores almacenados en los archivos deben cumplir con restricciones de consistencia.

Todos estos inconvenientes fomentan al desarrollo de las BD y los SGBD.

**3. Bases de datos.**

Una BD permitirá reunir toda la información relacionada en un único sistema de almacenamiento, pudiendo cualquier aplicación utilizarla de manera independiente y ofreciendo una mejora en el tratamiento de la información, así como una evolución para el desarrollo de aplicaciones.

Las BD son instrumentos de gran utilidad para gestionar grandes ficheros y facilitar la consulta de información. En muchas, además, puede definirse un esquema de permisos que establece qué personas o programas pueden acceder a los datos, y a cuáles, con el objetivo de presentar el contenido de forma adecuada y clara.

Los distintos sistemas de BD se diferencian conceptualmente entre sí y tienen, por lo tanto, sus propias ventajas y desventajas. Pero, antes que nada, es conveniente diferenciar entre la BD en sí y el sistema que la gestiona. Como BD se designa al conjunto de los datos que se ha de almacenar, mientras que el SGBD es responsable de su administración, determinando así su estructura, el orden, los permisos de acceso, las dependencias, etc

La gestión de las BD ha experimentado gran cantidad de cambios, partiendo de aplicaciones especializadas hasta llegar a convertirse en el núcleo de los entornos informáticos modernos.

**3.2 Conceptos.**

La BD no sólo contiene los datos de la organización, también almacena una descripción de dichos datos. Esta descripción es lo que se denomina metadatos, se almacena en el diccionario de datos o catálogo y es lo que permite que exista independencia de datos lógica-física.

Base de datos (BD): es una colección de datos relacionados lógicamente entre sí, con una definición y descripción comunes y que están estructurados de una determinada manera. Es un conjunto estructurado de datos que representa entidades y sus interrelaciones, almacenados con la mínima redundancia y posibilitando el acceso a ellos eficientemente por parte de varias aplicaciones y usuarios.

Una BD constará de los siguientes elementos:

**✓ Entidades**: objeto real o abstracto con características diferenciadoras de otros, del que se almacena información en la BD.

**✓ Atributos:** son los datos que se almacenan de la entidad.

**✓ Registros:** donde se almacena la información de cada entidad.

**✓ Campos:** donde se almacenan los atributos de cada registro.

**Las ventajas fundamentales que ofrece el uso de BD se resumen a continuación:**

Las ventajas que ofrece el uso de BD son:

* **Acceso múltiple:** Diversos usuarios podrán acceder a la BD.
* **Utilización múltiple:** Usuarios o aplicaciones disponen de una visión particular de la estructura la BD.
* **Flexibilidad:** Forma de acceder a la información establecida de diferentes maneras y con tiempos de respuesta muy reducidos.
* **Confidencialidad y seguridad:** Control de acceso de datos establecido para que se pueda acceder a una información determinada.
* **Protección contra fallos:** Mecanismos que permiten la recuperación de los datos de forma fiable.
* **Independencia lógica:** Cambios en la BD no afectan a las aplicaciones que las usan.
* **Redundancia:** Los datos se almacenan por lo general una única vez
* **Interfaz de alto nivel:** Mediante utilización de lenguajes de alto nivel de forma sencilla y cómoda.
* **Consulta directa:** Herramientas para poder acceder a los datos interactivamente.

**3.3 Usos.**

**El administrador** 🡪 Es la persona encargada de la creación o implementación física de la BD. Es quien escoge los tipos de ficheros, los índices que hay que crear, la ubicación de éstos, etc. El administrador establecerá la política de seguridad y de acceso para garantizar el menor número de problemas.

**Los diseñadores** 🡪Son las personas encargadas de diseñar cómo será la BD. Llevarán a cabo la identificación de los datos, las relaciones entre ellos, sus restricciones, etc. Para obtener un buen resultado, el diseñador de la BD debe implicar en el proceso a todos los usuarios de la BD, tan pronto como sea posible.

**Los programadores de aplicaciones** 🡪Hacen el programa de aplicación para los usuarios finales. Estos programas de aplicación ofrecerán la posibilidad de realizar consultas de datos, inserción, actualización o eliminación de los mismos. Para desarrollar estos programas se utilizan lenguajes de programación.

**Los usuarios finales** 🡪 Son los clientes finales de la BD. Al diseñar, implementar y mantener la BD se busca cumplir los requisitos establecidos por el cliente para la gestión de su información.

**3.4 Ubicación de la información.**

todas las BD normalmente se almacenan y localizan en discos duros y otros dispositivos de almacenamiento, a los que se accede a través de un ordenador. Una gran BD puede necesitar servidores en lugares diferentes, y viceversa, pequeñas BD pueden existir como archivos en el disco duro de un único equipo.

**3.4.1 Discos.**

En la actualidad existen dos tipos de discos:

✓ Discos duros: los discos HDD (Hard Drive Disk) son dispositivos mecánicos que utilizan el magnetismo para grabar los datos.

✓ Discos de estado sólido: las unidades de estado sólido SSD (Solid State Drive) almacenan la información en chips con memorias flash interconectadas.

El sistema de disco duro NAS (Network Attached Storage) permite colocar tantas unidades como se necesiten. Los sistemas NAS son en sí mismos pequeños ordenadores conectados a la red que permiten una gestión de datos empresarial muy eficiente y garantizan la seguridad en caso de un ciberataque.

**3.4.2 Cintas magnéticas.**

Este tipo de dispositivo de almacenamiento graba los datos en pistas sobre una banda plástica que cuenta con material magnético. Existen distintos tipos de cintas de almacenamiento según su composición química o formatos de grabación que utilicen. La gran durabilidad y su reducido coste las han convertido en un medio de almacenamiento que es aún utilizado en la actualidad para copias de seguridad.

**3.4.4 Almacenamiento en red.**

✓ SAN: su uso principal es en servidores de aplicaciones. Es una solución de almacenamiento en la red donde múltiples dispositivos actúan como bloques de disco.

✓ NAS: como se ha dicho anteriormente, es un sistema de almacenamiento de datos que utiliza un único dispositivo accesible desde la red para guardar y compartir información.

**3.4.5 Almacenamiento en la nube.**

Se trata de contratar un espacio externo (en los servidores de un proveedor de almacenamiento en la nube) donde es posible almacenar toda la información y acceder a ella independientemente del SO, ubicación, hora o dispositivo que se utilice.

**4. Evolución de los modelos de BD.**

**-Modelar:** se hace que lo abstracto se la realidad.

**-Modelo:** instrumento que se aplica a una parcela del mundo real para obtener una estructura de datos a la que denominamos esquema. El modelo será un conjunto de conceptos, reglas y convenciones que nos permiten describir los datos del mundo real.

**-Modelo de datos:** conjunto de conceptos o herramientas conceptuales que sirven para describir la estructura de una BD.

**-Esquema de la BD:** descripción de una BD mediante un modelo de datos.

**4.1 Modelo jerárquico.**

Los datos de este modelo se almacenan en estructuras lógicas llamadas segmentos. Los segmentos se relacionan entre sí utilizando arcos. Ya no se usa.

**4.2 Modelo en red.**

El modelo en red organiza la información en registros (también llamados nodos) y enlaces. En los registros se almacenan los datos, mientras que los enlaces permiten relacionar estos datos. Ya no se usa.

**4.3 Modelo relacional.**

Es posterior a los dos anteriores y fue desarrollado por Codd en 1970 y hoy en día son las BD más utilizadas. Es perceptible por el usuario como un conjunto de tablas.

El modelo relacional utiliza tablas bidimensionales (relacionales) para la representación lógica de los datos y las relaciones entre ellos.

Se llamará registro, entidad o tupla a cada fila de las tablas y campo o atributo a cada columna de la tabla.

A los conjuntos de valores que puede tomar un determinado atributo, se le denomina dominio.

Una clave será un artículo o conjunto de atributos que identifique de forma única una tupla.

Las tablas deben cumplir los siguientes requisitos:

* Todos los registros son del mismo tipo.
* La tabla sólo puede tener un tipo de registro.
* No existen campos o atributos repetidos.
* No existe orden en el almacenamiento de los registros.

El lenguaje habitual para construir las consultas a BD relacionales es SQL, un estándar implementado por los principales motores o SGBD relacionales.

**4.4 Modelo orientado a objetos.**

El modelo orientado a objetos define una BD en términos de objetos, sus propiedades y sus operaciones. Los objetos con la misma estructura y comportamiento pertenecen a una clase, y las clases se organizan en jerarquías.

**4.5 Modelo orientado a documentos.**

En este modelo, los documentos son la unidad básica para el almacenamiento de datos. Estas unidades son las que estructuran los datos y no deben confundirse con los documentos de los programas de procesamiento de texto. Aquí, los datos se guardan en los llamados pares clave-valor, comprendiendo así, una clave y un valor.

Estos sistemas son especialmente interesantes para las aplicaciones web.

**4.6 Otros modelos.**

**4.6.1 Modelo objeto-relacional.**

Las BD pertenecientes a este modelo, son un híbrido entre las BD del modelo relacional y el orientado a objetos.

**4.6.2 Modelo de BD deductivas o lógicas.**En este modelo las BD almacenan la información y permiten realizar deducciones a través de inferencias.

**4.6.3 BD multidimensionales.**

Son BD ideadas para desarrollar aplicaciones muy concretas, básicamente almacena sus datos con varias dimensiones.

**4.6.4 BD transaccionales.**

Son BD caracterizadas por su velocidad para gestionar el intercambio de información, se utilizan sobre todo en sistemas bancarios, análisis de calidad y datos de producción industrial, y son muy fiables.

**4.7 Tabla comparativa de los principales modelos.**

(preguntar si hay que saberse la tabla)

**5. Sistemas Gestores de Bases de Datos (SGBD).**

Para poder tratar la información contenida en las BD se utilizan los Sistemas Gestores de Bases de Datos o SGBD, que ofrecen un conjunto de programas que permiten acceder y gestionar dichos datos. El objetivo fundamental de los SGBD es proporcionar eficiencia y seguridad a la hora de recuperar o insertar información en las BD.

* Sistema Gestor de Base de Datos: conjunto coordinado de programas, procedimientos, lenguajes, etc., que suministra, tanto a los usuarios no informáticos, como a los analistas programadores, o al administrador, los medios necesarios para describir y manipular los datos contenidos en la BD, manteniendo su integridad, confidencialidad y seguridad.

El SGBD permite a los usuarios la creación y el mantenimiento de una BD, facilitando la definición, construcción y manipulación de la información contenida en éstas.

Las ventajas del uso de SGBD son:

* Proporcionan al usuario una visión más sencilla de cómo se almacenan y mantienen los datos.
* Ofrecen independencia física.
* Disminuyen la redundancia y la inconsistencia de datos.
* Aseguran la integridad de los datos.
* Facilitan el acceso a los datos, aportando rapidez y evitando la pérdida de datos.
* Aumentan la seguridad y privacidad de los datos.
* Mejoran la eficiencia.
* Permiten compartir datos y accesos concurrentes.
* Facilitan el intercambio de datos entre distintos sistemas.
* Incorporan mecanismos de copias de seguridad y recuperación para restablecer la información en caso de fallos en el sistema

**5.1 Funciones.**

Desarrolla tres funciones fundamentales:

* **Función de descripción o definición:** Permite al diseñador de la BD crear estructuras apropiadas para integrar adecuadamente los datos. Permite definir las tres estructuras de la BD:
  + Estructura interna, Estructura conceptual y Estructura externa.
* Se realiza mediante el lenguaje de descripción de datos o DDL. Se especificarán las características de los datos a cada uno de los tres niveles:
  + **A nivel interno** – Indicar espacio reservado para BD, longitud de datos y representación.
  + **A nivel conceptual** – Definición de entidades e identificación, atributos, interrelación y restricciones de integridad.
  + **A nivel externo –** Se deben definir las vistas de los distintos usuarios.
* **Función de manipulación:** Permite a los usuarios de una BD buscar, añadir, suprimir o modificar los datos de la misma, siempre de acuerdo con las normas de seguridad dictadas por medio del administrador. Se lleva a cabo mediante un sistema de manipulación de datos (DML) que proporciona los instrumentos necesarios. Por manipulación de datos entendemos:
  + **Recuperación** de información en la BD (Consultas)
  + **Inserción** de información nueva en la BD.
  + **Borrado** de información de una BD.
  + La **modificación** de información almacenada en la BD.
* **Función de control:** Permite al administrador de la BD establecer mecanismos de protección de las diferentes visiones de los datos asociadas a cada usuario, proporcionando elementos de creación y modificación de dichos usuarios.

**5.2 Componentes.**

* **Lenguajes de la BD**: Cualquier sistema de gestión de bases de datos ofrece la posibilidad de utilizar lenguajes e interfaces adecuadas para sus distintos tipos de usuarios.
* **El diccionario de datos:** Descripción de los datos almacenados. Es el lugar donde se deposita la información sobre la totalidad de los datos que forman la BD. Contiene las características lógicas de las estructuras que almacenan los datos.
* **El gestor de BD:** Es la parte del software encargada de garantizar el correcto, seguro, íntegro y eficiente acceso al almacenamiento de los datos. Es el componente encargado de proporcionar una interfaz entre los datos almacenados y los programas de aplicación que los manejan.
* **Usuarios de la BD:** En los SGBD existen diferentes perfiles de usuario, cada uno de ellos con una serie de permisos sobre los objetos de la BD. Generalmente existirán:
  + Administrador de la BD.
  + Usuarios de la BD
* **Herramientas de la BD:** Son un conjunto de aplicaciones que permiten a los administradores la gestión de la BD, de los usuarios, permisos, generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.

**5.3 Arquitectura.**

Proporciona tres niveles de abstracción:

* **Nivel interno o físico:** En este nivel se describe la estructura física de la BD a través de un esquema interno encargado de detallar si el sistema de almacenamiento de la BD y sus métodos de acceso.
* **Nivel lógico o conceptual:** Se describe la estructura completa de la BD a través de un esquema conceptual que detalla las entidades, atributos, relaciones, operaciones de los usuarios y restricciones.
* **Nivel externo o de visión del usuario:** Se describen las diferentes vistas que los usuarios percibirán de la BD.

**5.4 Tipos**

Según el modelo lógico en que se basan: (apartado 4)

Según el número de usuarios a los que da servicio el sistema:

* **Monousuario:** sólo atienden a un usuario a la vez, y su principal uso se da en los ordenadores personales.
* **Multiusuario:** entre los que se encuentran la mayor parte de los SGBD, atienden a varios usuarios al mismo tiempo.
* Según el número de sitios en los que está distribuida la BD:
  + **Centralizados:** sus datos se almacenan en un solo ordenador. Los SGBD centralizados pueden atender a varios usuarios, pero el SGBD y la BD en sí residen por completo en una sola máquina.
  + **Distribuidos** (homogéneos o heterogéneos): la BD real y el propio software del SGBD pueden estar distribuidos en varios sitios conectados por una red.
* El cuarto criterio toma como referencia el coste
* **Según el propósito:** 
  + **Propósito general:** pueden ser utilizados para el tratamiento de cualquier tipo de BD y aplicación.
  + **Propósito específico:** cuando el rendimiento es fundamental, se puede diseñar y construir un software de propósito especial para una aplicación específica, y este sistema no sirve para otras aplicaciones.

**6. SGBD comerciales y libres.**

En la siguiente tabla se exponen los SGBD comerciales más utilizados y sus características más relevantes:

|  |  |
| --- | --- |
| SGBD | DESCRIPCIÓN |
| Oracle | Ofrece una versión gratuita |
| Enterprise | No ofrece una versión gratuita |
| DB2 |  |
| Informix |  |
| Microsoft SQL SERVER | No ofrece una versión gratuita |
| SyBase / SAP ASE | No ofrece una versión gratuita |

La alternativa a los SGBD comerciales la encontramos en los SGBD de código abierto o libres, también llamados Open Source. Son sistemas distribuidos y desarrollados libremente.

|  |  |
| --- | --- |
| SGBD | DESCRIPCIÓN |
| MySQL Community | No gratuita / libre |
| MariaDB | No gratuita / libre |
| PostgreSQL | Libre |
| Firebird |  |
| Apache Derby |  |
| SQLite |  |

**7. BD centralizadas**

Sistema de BD centralizado: es aquella estructura en la que el SGBD está implantado en una sola plataforma u ordenador desde donde se gestiona directamente, de modo centralizado, la totalidad de los recursos. Es la arquitectura de los centros de proceso de datos tradicionales. Se basa en tecnologías sencillas, muy experimentadas y de gran robustez.Las principales características de las BD centralizados son:

* Todos los componentes del sistema residen en un solo ordenador o sitio.
* No posee múltiples elementos de procesamiento ni intercomunicación como las BD distribuidas.
* Los componentes de las BD centralizadas son: los datos, el software de gestión de BD y los dispositivos de almacenamiento secundario asociados.
* Son sistemas en los que su seguridad puede verse comprometida más fácilmente.

**8. BD distribuidas**

**- BD distribuida (BDD):** se encuentran distribuidas entre diferentes nodos interconectados por una red de comunicaciones.

**- Sistema de BD distribuida (SBDD):** múltiples sitios de BD están ligados por un sistema de comunicaciones, 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.

**- SGBD distribuido (SGBDD):** 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.

Un SGBDD desarrollará su trabajo a través de un conjunto de sitios o nodos, que poseen un sistema de procesamiento de datos completo con una BD local, un SGBD e interconectados entre sí.

**8.1. Fragmentación**

El problema de la fragmentación se refiere al particionamiento de la información para distribuir cada parte a los diferentes sitios de la red. Pero hay que tener en cuenta el grado de fragmentación.

El grado de fragmentación deberá estar equilibrado y dependerá de las particularidades de las aplicaciones que utilicen dicho BD.

Cuando se lleva a cabo una fragmentación, existen tres reglas fundamentales a cumplir:

* **Completitud**: si una relación R se descompone en fragmentos R1, R2, ..., Rn, cada elemento de datos que pueda encontrarse en R deberá poder encontrarse en uno o varios fragmentos.
* **Reconstrucción**: si una relación R se descompone en una serie de fragmentos R1, R2, ..., Rn, la reconstrucción de la relación a partir de sus fragmentos asegura que se preservan las restricciones definidas sobre los datos.
* **Disyunción**: si una relación R se descompone verticalmente, sus atributos primarios clave normalmente se repiten en todos sus fragmentos.

**Existen tres tipos de fragmentación:**

* **Fragmentación horizontal**: Se hace dividiendo la relación en subrelaciones. Los fragmentos se eligen mediante una operación de selección. Existen de dos tipos:
  + **Primaria:** particionado tuplas, con subconjuntos que tengan propiedades comunes.
  + **Derivada:** Consiste en dividir una relación partiendo de las condiciones definidas sobre alguna otra.
* **Fragmentación vertical**: Se basa en atributos de la relación para efectuar la división. Una relación R produce fragmentos R1, R2, R3, cada uno de los cuales contiene un subconjunto de los atributos de R, así como la clave primaria de R.
* **Fragmentación híbrida o mixta**: Podemos combinar ambas (horizontal y vertical), tipos:
  + **Mixta HV:** Fragmentación horizontal y luego una vertical, sobre cada uno de los fragmentos horizontales-
  + **Mixta VH**: Si tras una fragmentación vertical se lleva a cabo otra horizontal.
  + Celdas: Se aplican fragmentaciones simultáneamente sobre la relación.