

## ¿Qué es un fichero? (2.1)

**Fichero o archivo:** conjunto de **información relacionada**, tratada como un todo y **organizada** de forma estructurada. Es una **secuencia** de **dígitos binarios** que **organiza información relacionada** con un mismo aspecto.

Los ficheros están formados por **registros lógicos** que contienen **datos relativos** a un **mismo elemento** u objeto (datos de usuarios).

Los datos almacenados se pueden añadir, suprimir, actualizar o consultar individualmente en cualquier momento.

**Solo se puede llevar a la memoria principal partes** de los ficheros porque son **muy pesados**.

La **cantidad** de **información** que es **transferida** entre el **soporte** donde está almacenado el fichero, y la **memoria principal**, en una sola operación de lectura/grabación, **se llama registro físico o bloque**.

En una **operación** de lectura/grabación se **transfiere** un **bloque**, con **varios registros** lógicos del fichero, a estos registros dentro de un bloque se les llama **factor de blocaje**, y a la operación de agruparlos **bloqueo de registros**.

### Tipos de ficheros (2.2)

Ficheros permanentes		
Contienen...		Ejemplo
Maestros	<b>Estado actual</b> de los datos que pueden <b>modificarse</b> desde la aplicación. Es la parte <b>central</b> de la aplicación, su <b>núcleo</b> .	Usuarios de una plataforma educativa
Constantes	<b>Datos fijos</b> para la aplicación. <b>No suelen</b> ser <b>modificados</b> y se <b>accede</b> a ellos para <b>realización</b> de <b>consultas</b> .	Códigos postales
Históricos	<b>Datos</b> que <b>fueron considerados</b> como <b>actuales</b> en un periodo o <b>situación anterior</b> . Se utilizan para la <b>reconstrucción</b> de <b>situaciones</b> .	Usuarios que han sido dados de baja en la plataforma educativa

Ficheros temporales	
Almacenan...	
Intermedios	Resultados de una aplicación que serán utilizados por otra
Maniobras	Datos de una aplicación que no pueden ser mantenidos en memoria principal por falta de espacio.
Resultados	Datos que van a ser transferidos a un dispositivo de salida.

## Los soportes de información (2.3)

Los soportes más utilizados para almacenar los ficheros son:

- Las cintas magnéticas
- Los discos (magnéticos, ópticos, o magneto-ópticos).

Tipos de soportes		
Acceso Directo a los datos	<p>Son los <b>más usados</b></p> <p>Se puede <b>acceder</b> a los datos de <b>forma directa</b></p> <p>Podemos <b>colocarnos</b> en la <b>posición</b> que nos <b>interesa</b> y <b>leer</b> a partir de ella.</p>	Discos
Acceso Secuencial	<p>Usados para <b>copias de seguridad</b></p> <p><b>No</b> se puede <b>acceder</b> a los datos de <b>forma directa</b>.</p> <p>Para leer un dato, hay que <b>leer todo</b> lo que hay <b>hasta llegar</b> a esa <b>posición</b>.</p>	Cintas magnéticas

## Métodos de acceso (2.4-2.8)

Objetivos fundamentales de estas modificaciones
<b>Proporcionar</b> un <b>acceso rápido</b> a los <b>registros</b> .
Conseguir <b>economizar</b> el <b>almacenamiento</b> .
<b>Facilitar</b> la <b>actualización</b> de los <b>registros</b> .
<b>Permitir</b> que la <b>estructura</b> refleje la <b>organización real</b> de la información.

Ficheros secuenciales	<p><b>Definición:</b> Los <b>registros</b> están <b>almacenados</b> de <b>forma contigua</b>, para acceder a un registro hay que <b>leer todos</b> los registros <b>hasta llegar</b> a él.</p> <div><div>Registro 1</div><div>Registro 2</div><div> </div><div>Registro i-1</div><div>Registro i-2</div><div> </div><div>Registro n-1</div><div>Registro n</div></div> <p><b>Características:</b></p> <p>La <b>lectura</b> siempre se realiza hacia delante.</p> <p>Son <b>ficheros monousuarios</b>, <b>no permiten</b> el <b>acceso simultáneo</b> de varios usuarios.</p> <p>Tienen una <b>estructura rígida</b> de campos. <b>Todos</b> los <b>registros</b> deben <b>aparecer en orden</b> y la <b>posición</b> de los <b>campos</b> de cada <b>registro siempre</b> ha de ser <b>la misma</b>.</p> <p>El modo de apertura del fichero, condiciona la lectura o escritura.</p> <p><b>Aprovechan al máximo</b> el <b>soporte</b> de <b>almacenamiento</b>, <b>al no dejar huecos vacíos</b>.</p> <p>Se pueden <b>grabar</b> en <b>cualquier tipo</b> de <b>soporte</b>, <b>secuenciales</b> y <b>direccionables</b>.</p> <p>Todos los <b>lenguajes</b> de programación pueden <b>trabajar</b> con este tipo de <b>ficheros</b>.</p> <p><b>No</b> se pueden <b>insertar registros entre</b> los que ya están <b>grabados</b>.</p>
--------------------------	--

**Definición:** 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 **dispositivos** de **memoria** masiva de **acceso directo**.

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

Cada uno de los **registros** se guarda en una **posición física**, que **dependerá** del **espacio disponible** en memoria masiva, de ahí que la distribución de los registros sea aleatoria dentro del soporte de almacenamiento. Para **acceder** a la **posición física** de un registro se **utiliza** una **dirección** o **índice**.

#### Características:



**Posicionamiento inmediato.**

Registros de **longitud fija**.

**Apertura** del fichero en **modo mixto**, para lectura y escritura.

Permiten **múltiples usuarios** utilizándolos.

Los registros se **borran** colocando un **cero** en la **posición** que **ocupan**.

Permiten la utilización de **algoritmos** de **compactación** de **huecos**.

Los **archivos** se crean con un **tamaño definido**, es decir, con un **máximo** de **registros** establecido durante la creación.

Esta organización sólo es posible en **soportes direccionables**.

**Se usan** cuando el acceso a los datos de un registro se hace siempre empleando la **misma clave** y la **velocidad** de **acceso** a un registro es lo que más nos importa.

Permiten la **actualización** de los **registros** en el mismo fichero, **sin necesidad** de **copiar** el **fichero**.

Permiten realizar procesos de **actualización** en **tiempo real**.

## Ficheros indexados

Utilizando **índices**, permiten el acceso a un registro del fichero de forma directa.

Existe una **zona de registros** en la que se encuentran los datos del archivo y una **zona de índices**, que contiene una tabla con las claves de los registros y las posiciones donde se encuentran estos. La tabla de índices estará ordenada por el campo clave.

La **tabla de índices** se carga en **memoria principal** para **buscar** la fila correspondiente a la **clave del registro** a encontrar, obteniéndose así la **dirección** donde se **encuentra** el **registro**.

Una vez **localizada** la dirección, sólo hay que **acceder** a la **zona de registros** en el **soporte** de almacenamiento y **posicionarnos** en la **dirección** indicada. Puesto que la tabla debe prever la inclusión de todas las direcciones posibles del archivo.

Su principal **inconveniente** resulta **determinar** su **tamaño** y **mantenerla ordenada** por los valores de la clave.

### Características:



El diseño del registro tiene que tener un campo, o combinación de campos, que permita identificar cada registro de forma única, es decir, que no pueda haber dos registros que tengan la misma información en él. A este campo se le llama campo clave y es el que va a servir de índice. Un mismo fichero puede tener más de un campo clave, pero al menos uno de ellos no admitirá valores duplicados y se le llama clave primaria. A las restantes se les llama claves alternativas.

**Permiten utilizar** el modo de **acceso secuencial** y el modo de **acceso directo** para leer la información guardada en sus registros.

### Utilizando el modo de acceso directo:

Se hace **conociendo** el **contenido** del **campo clave** del **registro** que queremos localizar. Con esa información el **sistema operativo** puede **consultar** el **índice** y **conocer** la **posición** del **registro** dentro del **fichero**.

### Utilizando el modo de acceso secuencial:

Los **registros** son **leídos ordenados** por el **contenido** del **campo clave**, **independientemente del orden** en que se **fueron grabando** (el orden lógico no es igual al orden físico), debido a que el **acceso** a los **datos** se hace **a través** del **índice**, que para **hacer más fácil** la **búsqueda** de los **registros**, permanece **siempre ordenado** por el **campo clave**.

	<b>Solamente se puede grabar en un soporte direccionable.</b>
--	---

Ficheros Secuenciales Indexados	<p>Al igual que en los <b>ficheros indexados</b> existe una <b>zona de índices</b> y otra <b>zona de registros de datos</b>, pero esta última se encuentra dividida en <b>segmentos</b> (bloques de registros) ordenados.</p> <p><b>Organización</b> muy <b>utilizada</b>, para <b>procesos</b> en los que <b>intervienen pocos registros</b> y en los que se <b>maneja el fichero completo</b>.</p> <p><b>Características:</b></p> <p><b>Permite el acceso secuencial</b>, interesante cuando la <b>tasa de actividad</b> es <b>alta</b> y los <b>registros</b> se <b>leen ordenados</b> por el <b>campo clave</b>.</p> <p>Permite el <b>acceso directo</b> a los registros, empleando para ello las <b>tablas de índices</b>. Primero <b>busca</b> la <b>clave</b> en el <b>área de índices</b> y luego va a <b>leer</b> al <b>área de datos</b> en la <b>dirección</b> que le <b>indica la tabla</b>.</p> <p>Se pueden <b>actualizar</b> los <b>registros</b> en el <b>mismo fichero</b>, <b>sin necesidad</b> de <b>crear un fichero nuevo</b> de copia en el proceso de actualización.</p> <p><b>Ocupa más espacio</b> en el disco <b>que los ficheros secuenciales</b>, <b>debido</b> al uso del <b>área de índices</b>.</p> <p><b>Solo</b> se puede utilizar <b>soportes direccionables</b>.</p> <p><b>Inversión económica mayor</b>, por la <b>necesidad</b> de <b>programas y hardware más sofisticado</b>.</p>
---------------------------------	--

Ficheros de Acceso  
Calculado o Hash

Permite accesos más **rápidos**, ya que, **en lugar** de **consultar** una **tabla**, se **utiliza** una transformación o **función matemática** (función de hashing) conocida, que **a partir** de la **clave** genera la **dirección** de cada **registro** del **archivo**. Si la **clave** es **alfanumérica**, deberá **previamente** ser **transformada** en un **número**.

El **mayor problema** que presenta este tipo de ficheros es que **a partir** de **diferentes claves** se **obtenga** la **misma dirección** al aplicar la función matemática o transformación.

A este **problema** se le **denomina** **colisión**, y las claves que generan la misma dirección se conocen por **sinónimos**. Para **resolver** este **problema** se aplican **diferentes métodos**, como **tener** un **bloque** de **excedentes** o zona de **sinónimos**, o **crear** un **archivo** de **sinónimos**.

#### Métodos de extracción:

**Módulo:** La dirección será igual al resto de la división entera entre la clave y el número de registros.

**Extracción:** La dirección será igual a una parte de las cifras que se extraen de la clave.

## Parámetros de utilización (2.9)

Capacidad o volumen	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	
	<b>Tasa de consulta o modificación</b>	Porcentaje de registros consultados o modificados en cada tratamiento del fichero, respecto al número total de registros contenidos en él
	<b>Frecuencia de consulta o modificación</b>	Número de veces que se accede al fichero para hacer una consulta o modificación en un periodo de tiempo fijo.
Volatilidad	Mide la cantidad de inserciones y borrados que se efectúan en un fichero	
	<b>Tasa de renovación</b>	Tanto por ciento de registros renovados en cada tratamiento del fichero, respecto al número total de registros contenidos en él
	<b>Frecuencia de renovación</b>	Número de veces que se accede al fichero para renovarlo en un periodo de tiempo fijo
Crecimiento	Variación de la capacidad del fichero y se mide con la tasa de crecimiento, que es el porcentaje de registros en que aumenta el fichero en cada tratamiento	



## Conceptos (3.1)

**Base de datos:** 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**.

Elementos de una Base de Datos		
Entidades	<b>Objeto real o abstracto</b> con <b>características diferenciadoras</b> de otros, del que se <b>almacena información</b> en la base de datos.	En una clínica veterinaria, podría ser: ejemplar, doctor, consulta
Atributos	<b>Datos</b> que se <b>almacenan de la entidad</b> . <b>Cualquier propiedad o característica</b> de una <b>entidad</b> puede ser atributo.	Raza, color, nombre, número de identificación
Registros	<b>Donde se almacena la información de cada entidad</b> . Es un <b>conjunto de atributos</b> que <b>contienen los datos</b> que <b>pertenecen</b> a una <b>misma repetición</b> de entidad.	2123056, Sultán, Podenco, Gris, 23/03/2009
Campos	<b>Donde se almacenan los atributos de cada registro</b> .	El valor Podenco

Ventajas	
Acceso múltiple	<b>Varios usuarios o aplicaciones</b> podrán <b>acceder</b> a la <b>base de datos</b> , <b>sin</b> que existan <b>problemas</b> en el acceso o los datos
Utilización múltiple	<b>Cada uno</b> de los <b>usuarios o aplicaciones</b> podrán disponer de una <b>visión particular</b> de la <b>estructura</b> de la <b>base de datos</b> , de tal manera que <b>cada uno</b> de ellos <b>accederá sólo</b> a la <b>parte</b> que <b>realmente</b> le <b>corresponde</b> .
Flexibilidad	La <b>forma</b> de <b>acceder</b> a la <b>información</b> puede ser <b>establecida</b> de <b>diferentes maneras</b> , ofreciendo <b>tiempos de respuesta</b> muy <b>reducidos</b> .
Confidencialidad y seguridad	El <b>control</b> del <b>acceso</b> a los <b>datos</b> podrá ser <b>establecido</b> para que <b>unos usuarios o aplicaciones</b> puedan <b>acceder</b> a <b>unos datos</b> y a <b>otros no</b> , <b>impidiendo</b> a los <b>usuarios no autorizados</b> la <b>utilización</b> de la <b>base de datos</b> .
Protección contra fallos	<b>En caso</b> de <b>errores</b> en la <b>información</b> , existen <b>mecanismos</b> bien <b>definidos</b> que <b>permiten</b> la <b>recuperación</b> de los <b>datos</b> de forma fiable

Independencia física	Un <b>cambio de soporte físico</b> de los datos (tipo de discos), <b>no afectaría</b> a la <b>base de datos</b> o a las aplicaciones que acceden a ellos
Independencia lógica	Los <b>cambios realizados</b> en la <b>base de datos</b> <b>no afectan</b> a las <b>aplicaciones</b> que la usan
Redundancia	Los <b>datos</b> se <b>almacenan</b> , por lo general, una <b>única vez</b> . Aunque si es necesario, <b>podríamos repetir información</b> de <b>manera controlada</b>
Interfaz de alto nivel	<b>Mediante</b> la <b>utilización</b> de <b>lenguajes</b> de <b>alto nivel</b> puede <b>utilizarse</b> la <b>base de datos</b> de <b>manera sencilla</b>
Consulta directa	<b>Existe</b> una <b>herramienta</b> para poder <b>acceder</b> a los <b>datos</b> <b>interactivamente</b>

## Usos (3.2)

Ventajas	
El administrador	<b>Persona encargada</b> de la <b>creación</b> o <b>implementación física</b> de la <b>base de datos</b> . Escoge los <b>tipos</b> de <b>ficheros</b> , los <b>índices</b> que, la <b>ubicación</b> de éstos, también, establecerá la <b>política</b> de <b>seguridad</b> y de acceso para <b>garantizar</b> el <b>menor número</b> de <b>problemas</b> .
Los diseñadores	<b>Personas encargadas</b> de <b>diseñar</b> cómo será la <b>base de datos</b> . <b>Identificarán</b> los <b>datos</b> , las <b>relaciones</b> entre ellos, sus <b>restricciones</b> ... El <b>diseñador</b> de la base de datos debe <b>implicar</b> en el <b>proceso</b> a <b>todos los usuarios</b> de esta.
Los programadores de aplicaciones	Se <b>encargan</b> de <b>implementar</b> los <b>programas</b> de <b>aplicación</b> que <b>servirán</b> a los <b>usuarios finales</b> . <b>Programas</b> que <b>realizan consultas</b> de <b>datos</b> , <b>inserción</b> , <b>actualización</b> o <b>eliminación</b> de los mismos. Para <b>desarrollar</b> estos <b>programas</b> se utilizan <b>lenguajes</b> de <b>tercera</b> o <b>cuarta generación</b> .
Los usuarios finales	Son los <b>clientes finales</b> de la <b>base de datos</b> . Al diseñar, implementar y mantener la base de datos se busca cumplir los requisitos establecidos por el cliente para la gestión de su información.

## Ubicación de la información (3.3)

<b>Discos SATA</b>	<p><b>Interfaz de transferencia de datos</b> entre la <b>placa base</b> y algunos <b>dispositivos de almacenamiento</b>:</p> <p>Disco duro, lectores y grabadores de CD/DVD/BD, Unidades de Estado Sólido...</p> <p>SATA proporciona:</p> <ul style="list-style-type: none"><li>• <b>Mayores velocidades</b></li><li>• <b>Mejor aprovechamiento</b> cuando hay <b>varias unidades</b></li><li>• <b>Mayor longitud</b> del <b>cable de transmisión de datos</b></li><li>• <b>Capacidad</b> para <b>conectar unidades</b> al <b>instante</b>, es decir, insertar el dispositivo <b>sin</b> tener que <b>apagar</b> el <b>ordenador</b></li></ul> <p>SATA 150 MB/s   SATA II 300 MB/s   SATA III 600 MB/s.</p>
<b>Discos SCSI</b>	<p><b>Interfaces</b> para <b>discos duros</b> de <b>gran capacidad</b> de <b>almacenamiento</b> y <b>velocidad</b> de <b>rotación</b>.</p> <p>Su <b>tiempo medio de acceso</b> puede llegar a <b>7 milisegundos</b> y su velocidad de <b>transmisión secuencial</b> de información es de:</p> <ul style="list-style-type: none"><li>• SCSI Estándar (Standard SCSI) - 5 MB/s en los discos</li><li>• SCSI Rápido (Fast SCSI) - 10 MBps en los discos</li><li>• SCSI Ancho-Rápido (Fast-Wide SCSI) - 20 MBps en los discos</li></ul> <p>Un controlador SCSI puede manejar hasta 7 discos duros SCSI.</p>
<b>RAID</b>	<p>Acrónimo de <b>Redundant Array of Independent Disks</b> o matriz de discos independientes,</p> <p>Es un <b>contenedor</b> de <b>almacenamiento redundante</b>.</p> <p>Se basa en el <b>montaje en conjunto</b> de <b>dos o más discos</b> duros, <b>formando</b> un <b>bloque</b> de trabajo, para <b>obtener</b> desde una <b>ampliación</b> de <b>capacidad</b> a <b>mejoras</b> en <b>velocidad</b> y <b>seguridad</b> de <b>almacenamiento</b>.</p> <p><b>Según</b> las <b>características</b> que queramos primar, se establecen <b>distintos sistemas</b> de RAID.</p>

<b>Sistemas NAS</b>	<p>Acrónimo de <b>Network Attached Storage</b> ó sistema de almacenamiento masivo en red.</p> <p>Estos sistemas de almacenamiento <b>permiten compartir</b> la <b>capacidad de almacenamiento de</b> un <b>computador</b> (Servidor) con <b>ordenadores personales</b> o <b>servidores</b> clientes a <b>través de</b> una <b>red</b>, <b>haciendo uso</b> de un <b>sistema operativo</b> optimizado para <b>dar acceso</b> a los <b>datos a través</b> de <b>protocolos de comunicación específicos</b>.</p> <p>Suelen ser <b>dispositivos</b> para <b>almacenamiento masivo</b> de datos con <b>capacidades muy altas</b>, de varios Terabytes, generalmente superiores a los discos duros externos y además se <b>diferencian</b> de estos al <b>conectar</b> por <b>red</b>.</p>
<b>Sistemas SAN</b>	<p>Acrónimo de <b>Storage Area Network</b> o red de área de almacenamiento. Se trata de una <b>red concebida</b> para <b>conectar servidores</b>, matrices (arrays) de discos y librerías de soporte.</p> <p>La arquitectura de este tipo de sistemas <b>permite</b> que los <b>recursos de almacenamiento</b> estén <b>disponibles</b> para <b>varios servidores en una red de área local o amplia</b>.</p> <p>Debido a que la <b>información almacenada no reside</b> directamente en <b>ninguno</b> de los <b>servidores</b> de la <b>red</b>, se <b>optimiza</b> el poder de <b>procesamiento para aplicaciones comerciales</b> y la <b>capacidad de almacenamiento</b> se puede <b>proporcionar</b> en el <b>servidor</b> donde <b>más</b> se <b>necesite</b>.</p>

## Modelos (4-4.3)

Modelos	
Jerárquico	<p>Se utiliza una <b>jerarquía</b> en la que la <b>relación</b> entre las entidades de este modelo siempre es del <b>tipo padre/hijo</b>. De tal manera que existen <b>nodos</b> que <b>contienen atributos</b> o <b>campos</b> y que se <b>relacionarán</b> con sus <b>nodos hijos</b>.</p> <p><b>Cada nodo</b> puede <b>tener más</b> de un <b>hijo</b>, pero <b>únicamente</b> tendrá un <b>padre</b>.</p> <p>Los <b>datos</b> de este <b>modelo</b> se <b>almacenan</b> en <b>estructuras lógicas</b> llamadas <b>segmentos</b>. Los segmentos se <b>relacionan</b> entre sí <b>utilizando arcos</b>. La <b>forma visual</b> de este <b>modelo</b> es de <b>árbol invertido</b>, en la <b>parte superior</b> están los <b>padres</b> y en la <b>inferior</b> los <b>hijos</b>.</p> <p><b>Hoy</b> en día, debido a sus limitaciones, <b>está en desuso</b>.</p>
Red	<p><b>Organiza</b> la <b>información</b> en <b>registros</b> (también llamados nodos) y enlaces. En los <b>registros</b> se <b>almacenan</b> los <b>datos</b>, mientras que los <b>enlaces</b> permiten <b>relacionar</b> estos <b>datos</b>. Las <b>bases de datos</b> en <b>red</b> son <b>parecidas</b> a las <b>jerárquicas</b> sólo que en ellas <b>puede haber más</b> de un <b>padre</b>.</p> <p>Se pueden <b>representar perfectamente</b> cualquier tipo de <b>relación</b> entre los <b>datos</b>, pero hace <b>muy complicado</b> su <b>manejo</b>.</p> <p>Al <b>no</b> tener que <b>duplicar</b> la <b>información</b> se <b>ahorra espacio</b> de almacenamiento.</p> <p>El <b>sistema</b> de <b>gestión</b> de <b>información</b> <b>basado</b> en el modelo en red <b>más popular</b> es el <b>sistema IDMS</b>.</p>
Relacional	<p><b>Hoy en día</b> las bases de datos <b>relacionales</b> son las más <b>utilizadas</b>.</p> <p>En el <b>modelo relacional</b> la base de datos es un <b>conjunto</b> de <b>tablas</b>.</p> <p>Esta percepción es <b>sólo</b> a <b>nivel lógico</b>, ya que a <b>nivel físico</b> puede estar <b>implementada</b> mediante <b>distintas estructuras</b> de almacenamiento.</p> <p>El modelo relacional utiliza <b>tablas bidimensionales</b> (relaciones) para la <b>representación lógica</b> de los <b>datos</b> y las <b>relaciones</b> entre <b>ellos</b>. Cada <b>relación</b> (tabla) <b>posee</b> un <b>nombre</b> que es <b>único</b> y contiene un <b>conjunto</b> de <b>columnas</b>.</p> <p><b>Registro, entidad o tupla:</b> cada fila de la tabla y <b>campo o atributo</b> a cada <b>columna</b> de la <b>tabla</b>.</p>

	<p><b>Dominio:</b> Conjuntos de valores que puede tomar un <b>determinado atributo</b>.</p> <p><b>Clave:</b> atributo o conjunto de atributos que identifique de forma única a una tupla.</p> <p><b>Requisitos</b></p> <p>Todos los <b>registros</b> son del <b>mismo tipo</b>.</p> <p>La <b>tabla sólo</b> puede tener <b>un tipo</b> de registro.</p> <p><b>No existen campos o atributos repetidos</b>.</p> <p><b>No existen registros duplicados</b>.</p> <p><b>No existe orden</b> en el <b>almacenamiento</b> de los <b>registros</b>.</p> <p><b>Cada registro</b> o tupla es <b>identificada</b> por una <b>clave</b> que puede estar formada por uno o <b>varios campos o atributos</b>.</p>
Orientado a Objetos	<p><b>Base de datos</b> en términos de <b>objetos</b>, sus <b>propiedades</b> y sus <b>operaciones</b>.</p> <p>Los objetos con la misma estructura y comportamiento pertenecen a una <b>clase</b>, y las clases se organizan en jerarquías.</p> <p>Las <b>operaciones</b> de cada <b>clase</b> se <b>especifican</b> en <b>términos</b> de <b>procedimientos predefinidos</b> denominados <b>métodos</b>.</p> <p>Algunos sistemas basados en el modelo relacional, han evolucionado incorporando conceptos orientados a objetos llamados <b>objeto-relacionales</b>.</p> <p>El <b>objetivo</b> de este modelo es <b>cubrir</b> las <b>limitaciones</b> del <b>modelo relacional</b> ya que se <b>incorporan mejoras</b> como la <b>herencia</b> entre <b>tablas</b>, los <b>tipos definidos</b> por el <b>usuario</b>, <b>soporte multimedia...</b></p> <p><b>Conceptos más importantes:</b></p> <p><b>Encapsulación:</b> Propiedad que <b>permite ocultar</b> la <b>información</b> al <b>resto</b> de los <b>objetos</b>, <b>impidiendo</b> así <b>accesos incorrectos</b> o <b>conflictos</b>.</p> <p><b>Herencia:</b> Propiedad a través de la cual los <b>objetos heredan comportamiento dentro</b> de una <b>jerarquía</b> de <b>clases</b>.</p> <p><b>Polimorfismo:</b> Propiedad de una <b>operación mediante</b> la cual puede ser <b>aplicada</b> a <b>distintos</b> tipos de <b>objetos</b>.</p> <p>Desde la <b>aparición</b> de la <b>programación</b> orientada a objetos (<b>POO</b> u <b>OOP</b>) se empezó a pensar en bases de datos adaptadas a estos lenguajes.</p>

	<p>Este <b>modelo</b> es considerado como el <b>fundamento</b> de las <b>bases de datos de tercera generación</b>, siendo consideradas las bases de datos en red como la primera y las bases de datos relacionales como la segunda generación.</p> <p>Aunque no han reemplazado a las bases de datos relacionales, si son el <b>tipo</b> de base de datos que más está <b>creciendo</b> en los últimos años.</p>
--	--

Otros Modelos	
<b>Objeto-Relacional</b>	<p>Son un <b>híbrido</b> entre las bases del <b>modelo relacional</b> y el <b>orientado a objetos</b>.</p> <p>El mayor <b>inconveniente</b> son los <b>costes</b> de la <b>conversión</b> de las BD <b>relacionales</b> a BD <b>orientadas a objetos</b>.</p> <p>Se busca <b>obtener</b> lo <b>mejor</b> del <b>modelo relacional</b>, <b>incorporando</b> las <b>mejoras ofrecidas</b> por la <b>orientación a objetos</b>.</p> <p>Se siguen <b>almacenando tuplas</b>, aunque la <b>estructura</b> de las <b>tuplas</b> <b>no</b> está <b>restringida</b>, sino que las <b>relaciones</b> pueden ser <b>definidas</b> en <b>función</b> de <b>otras</b>, que es lo que denominamos <b>herencia directa</b>.</p> <p>El <b>estándar</b> en el que se <b>basa</b> este <b>modelo</b> es <b>SQL99</b>. Este estándar <b>ofrece</b> la <b>posibilidad</b> de <b>añadir</b> a las <b>bases de datos relacionales</b> <b>procedimientos almacenados</b> de <b>usuario</b>, <b>triggers</b>, <b>tipos definidos</b> por el <b>usuario</b>, <b>consultas recursivas</b>, <b>bases de datos OLAP</b>, <b>tipos LOB</b>, ...</p> <p>Otra <b>característica</b> a <b>destacar</b> es la <b>capacidad</b> para <b>incorporar funciones</b> que tengan un <b>código</b> en algún <b>lenguaje</b> de <b>programación</b> como, por ejemplo: <b>SQL</b>, <b>Java</b>, <b>C...</b></p> <p>La gran <b>mayoría</b> de las <b>bases de datos relacionales clásicas</b> de <b>gran tamaño</b>, como <b>Oracle</b>, <b>SQL Server</b>, son <b>objeto-relacionales</b>.</p>
<b>Bases de datos deductivas</b>	<p><b>Almacenan</b> la <b>información</b> y <b>permiten</b> realizar <b>deducciones</b> a través de <b>inferencias</b>. Es decir, se <b>derivan</b> <b>nuevas informaciones</b> a <b>partir</b> de <b>las</b> que se han <b>introducido explícitamente</b> en la <b>base de datos</b> por parte del <b>usuario</b>.</p> <p><b>También</b> son <b>llamadas</b> <b>BD lógicas</b>, al <b>basarse</b> en <b>lógica matemática</b>.</p> <p>Surgieron para <b>contrarrestar</b> las <b>limitaciones</b> del <b>modelo relacional</b> para la <b>respuesta</b> a <b>consultas recursivas</b> y la <b>deducción</b> de <b>relaciones indirectas</b> entre los <b>datos almacenados</b>.</p>

<b>Bases de datos multidimensionales</b>	<p>Ideadas para <b>desarrollar aplicaciones</b> muy <b>concretas</b>.</p> <p><b>Almacena</b> sus <b>datos</b> con <b>varias dimensiones</b>, en vez de <b>un valor</b>, encontramos <b>varios dependiendo</b> de los <b>ejes definidos</b> o <b>una BD</b> de <b>estructura basada en dimensiones</b> orientada a <b>consultas complejas</b> y <b>alto rendimiento</b>.</p> <p>La <b>información</b> se <b>representa</b> como <b>matrices multidimensionales</b>, <b>cuadros de múltiples entradas</b> o <b>funciones</b> de <b>varias variables</b> sobre <b>conjuntos finitos</b>.</p> <p><b>Cada una</b> de <b>estas matrices</b> se denomina <b>cubo</b>. Eso <b>facilita</b> el <b>manejo</b> de <b>grandes cantidades</b> de <b>datos</b>, dándole a esto una <b>amplia aplicación</b> dentro de <b>varias áreas</b> y <b>diferentes campos</b> del <b>conocimiento humano</b>.</p>
<b>Bases de datos transaccionales</b>	<p><b>Caracterizadas</b> por su <b>velocidad</b> para <b>gestionar</b> el <b>intercambio</b> de <b>información</b>, <b>utilizadas</b> en <b>sistemas bancarios</b>, <b>análisis</b> de <b>calidad</b> y <b>datos</b> de <b>producción</b> industrial.</p> <p>Son <b>bases</b> de <b>datos</b> muy <b>fiables</b>, ya que <b>las operaciones</b> de <b>inserción</b>, <b>actualización</b> o <b>borrado</b> se <b>realizan completamente</b> o se <b>descartan</b>.</p> <p>Entre la más destacadas se encuentra <b>Oracle</b>.</p>
<b>Modelo de bases de datos orientadas a documentos</b>	<p>El <b>principal objeto</b> de <b>gestión</b> es el <b>documento</b> que <b>contiene datos semiestructurados</b> que podrán estar <b>almacenados</b> en <b>algún formato</b>, por ejemplo, XML.</p> <p><b>MongoDB</b> y <b>CouchDB</b></p>



## Tipos de bases de datos (5)

Según su contenido		
<b>Información actual</b>	Contienen <b>información</b> muy <b>concreta</b> y <b>actualizada</b> , normalmente, de <b>tipo numérico</b> : <b>estadísticas</b> , series <b>históricas</b> , <b>resultados</b> de <b>encuestas</b> , <b>convocatorias</b> de <b>becas</b> ...	
<b>Directorios</b>	<b>Recogen</b> datos sobre <b>personas</b> o <b>instituciones especializadas</b> en una <b>actividad</b> o <b>materia concreta</b> .  Hay directorios de <b>profesionales</b> , de <b>bibliotecas</b> , de <b>revistas científicas</b> ...	
<b>Documentales</b>	<b>Cada registro</b> se <b>corresponde</b> con un <b>documento</b> , sea éste <b>de cualquier tipo</b> :  Una <b>publicación impresa</b> , un <b>documento audiovisual</b> , gráfico.  <b>Dependiendo</b> de si <b>incluyen o no</b> el <b>contenido completo</b> de los <b>documentos</b> que <b>describen</b> , podremos tener:	
	<b>Texto completo</b>	<b>Constituidas</b> por los propios <b>documentos</b> en <b>formato electrónico</b> , por un <b>volcado completo</b> de su texto.
	<b>Archivos electrónicos de imágenes</b>	<b>Constituidos</b> por <b>referencias</b> que <b>permiten</b> un <b>enlace directo</b> con la <b>imagen</b> del <b>documento</b> original.
	<b>Referenciales</b>	Sus <b>registros no contienen</b> el <b>texto original</b> sino <b>tan sólo</b> la <b>información fundamental</b> para <b>describir</b> y <b>permitir</b> la <b>localización</b> de <b>documentos</b> obtener <b>referencias</b> sobre <b>documentos</b> que <b>habrá</b> que <b>localizar posteriormente</b> en <b>otro servicio</b> .

Según su uso	
<b>Individual</b>	<p><b>Utilizada</b> básicamente por <b>una persona</b>.</p> <p>El <b>sistema administrador</b> de la <b>base de datos</b> y los <b>datos</b> son <b>controlados</b> por el <b>mismo usuario</b>.</p> <p><b>Puede</b> estar <b>almacenada</b> en la <b>unidad de disco duro</b> del usuario <b>o</b> en el <b>servidor de archivos</b> de una <b>red de área local</b>.</p> <p>Por ejemplo, un <b>gerente de ventas</b> podría contar con una <b>base de datos</b> para el <b>control</b> de sus <b>vendedores</b> y su <b>desempeño</b>.</p>
<b>Compartida</b>	<p>Son <b>bases de datos</b> con <b>múltiples usuarios</b> y que muy probablemente pertenezcan a la <b>misma organización</b>.</p> <p>Se <b>encuentra almacenada</b> en una <b>computadora potente</b> y bajo el <b>cuidado</b> de un <b>profesional</b> en el <b>área</b>, el <b>administrador</b> de la <b>base de datos</b>.</p> <p>Los <b>usuarios</b> tienen <b>acceso</b> a la <b>base de datos</b> mediante una <b>red de área local</b> o una <b>red de área extensa</b>.</p>
<b>Acceso Público</b>	<p>Son <b>bases de datos accesibles</b> por <b>cualquier persona</b>.</p> <p>Puede <b>no</b> ser <b>necesario pagar</b> un <b>canon</b> para hacer <b>uso</b> de los <b>datos</b> contenidos en ellas.</p>
<b>Propietarias o Bancos de Datos</b>	<p><b>Bases de datos de gran tamaño, desarrolladas</b> por una <b>organización</b> y que <b>contienen temas especializados</b> o de <b>carácter particular</b>.</p> <p>El <b>público general</b> puede tener <b>acceso</b> a estas <b>bases</b> a veces de forma <b>gratuita</b> y otras <b>mediante el pago</b> de una cuota.</p> <p>Pueden <b>ofrecer información</b> que va desde <b>negocios</b>, economía, <b>inversión</b>, técnica y <b>científica</b>....</p> <p>Permiten <b>encontrar</b> en <b>minutos</b> lo que tardaría <b>horas ojeando revistas</b>.</p>

Según la variabilidad de la información	
<b>Estáticas</b>	<p>Son <b>BD</b> de <b>sólo lectura</b>.</p> <p>Se <b>utilizan</b> para el <b>almacenamiento</b> de <b>datos históricos</b> que pueden ser <b>analizados</b> y <b>utilizados</b> para el <b>estudio</b> del <b>comportamiento</b> de un <b>conjunto de datos</b> a través del <b>tiempo</b>.</p> <p>Permiten <b>realizar proyecciones</b> y <b>toma de decisiones</b>.</p>
<b>Dinámicas</b>	<p>Son <b>BD</b> donde la <b>información almacenada</b> se <b>modifica</b> con el <b>tiempo</b>, <b>permitiendo operaciones</b> como <b>actualización</b> y <b>adición de datos</b>, además de las <b>operaciones fundamentales</b> de <b>consulta</b>,</p>

Según la localización de la información		
<b>Distribuidas</b>	<p>Se <b>sitúan</b> en un <b>lugar</b> o <b>lugares diferentes</b> a donde se <b>encuentran</b> los <b>usuarios</b>.</p> <p>Es la <b>unión</b> de las <b>bases de datos</b> mediante <b>redes</b>.</p> <p>Los <b>usuarios</b> se <b>vinculan</b> a los <b>servidores</b> de <b>BD</b> <b>distantes</b> mediante una <b>amplia variedad</b> de <b>redes</b> de <b>comunicación</b>.</p>	
<b>Centralizadas</b>	<p><b>Ubicadas</b> en un <b>único lugar</b>, un <b>único computador</b>.</p> <p>Pueden ser <b>bases de datos monousuario</b> que se <b>ejecutan</b> en <b>ordenadores personales</b> o <b>sistemas</b> de <b>BD</b> de <b>alto rendimiento</b> que se <b>ejecutan</b> en <b>grandes sistemas</b>.</p> <p><b>Facilita</b> las <b>labores</b> de <b>mantenimiento</b>, sin embargo, <b>hace</b> que la <b>información</b> contenida en <b>dicha base</b>, sea más <b>vulnerable</b> a <b>posibles fallos</b> y <b>limita</b> su <b>acceso</b>.</p> <p><b>Ofrece dentro</b> de la <b>arquitectura Cliente/Servidor</b> dos configuraciones:</p>	
	<b>Anfitrión</b>	<p>La <b>máquina cliente</b> y la <b>máquina servidor</b> son la <b>misma</b>.</p> <p>Los <b>usuarios</b> se <b>conectarán directamente</b> a la <b>máquina</b> donde se <b>encuentra</b> la <b>BD</b>.</p>
	<b>Cliente/Servidor</b>	<p>La <b>BD reside</b> en una <b>máquina servidor</b> y los <b>usuarios acceden</b> a la <b>BD</b> desde su <b>máquina cliente</b> a través de una <b>red</b>.</p>

## Bases de datos según el organismo productor

- ✓ **Bases de datos de organismos públicos y de la Administración:** Las bibliotecas y centros de documentación de los ministerios, instituciones públicas, universidades y organismos públicos de investigación elaboran gran cantidad de recursos de información. Estos sistemas pueden ser:
  - ◆ **Bases de datos de acceso público**, sean gratuitas o no.
  - ◆ **Bases de datos de uso interno**, con información de acceso restringido.
- ✓ **Bases de datos de instituciones sin ánimo de lucro:** Fundaciones, asociaciones, Sindicatos y organizaciones no gubernamentales elaboran frecuentemente sus propios sistemas de información especializados.
- ✓ **Bases de datos de entidades privadas o comerciales:** Los centros de documentación, bibliotecas y archivos de las empresas pueden elaborar distintos tipos de sistemas de información:
  - ◆ Bases de datos de uso interno para facilitar la circulación de información dentro de la empresa
  - ◆ Bases de datos de uso interno que ocasionalmente ofrecen servicio hacia el exterior (usuarios particulares u otras instituciones).
  - ◆ Bases de datos comerciales, diseñadas específicamente para ser utilizadas por usuarios externos.
- ✓ **Bases de datos realizadas por cooperación en red:** Se trata de sistemas de información cuya elaboración es compartida por diversas instituciones. bases de datos internacionales se elaboran a través de este sistema de trabajo, con diversos centros nacionales responsables de la información perteneciente a cada país.

## Bases de datos según el modo de acceso

- ✓ **Bases de datos de acceso local:** Para consultarlas es necesario acudir al organismo productor, a su biblioteca o centro de documentación. Pueden ser consultables en monopuesto o en varios puntos de una red local.
- ✓ **Bases de datos en CD-ROM:** Pueden adquirirse por compra o suscripción bien directamente por un particular o por una biblioteca o centro de documentación que permita su consulta a sus usuarios. En algunas instituciones se instalan diferentes CD-ROM en una red local para permitir su consulta desde cualquier ordenador conectado a la misma.
- ✓ **Bases de datos en línea:** Pueden consultarse desde cualquier ordenador conectado a Internet. La consulta puede ser libre (gratuita) o exigir la solicitud previa de una clave personal de entrada (denominada comúnmente con el término inglés password). Para obtener un password puede exigirse la firma de un contrato. Hay diferentes tipos de acceso en línea:
  - ◆ **Acceso vía telnet o mediante línea de Internet:** el usuario realiza una conexión estable al host (gran ordenador) en donde se halla la base de datos, a través de Internet. La interfaz de usuario instalada en dicho ordenador remoto determinará si la interrogación debe realizarse por menús o por comandos o expresiones de un lenguaje determinado. Cuando un usuario entra en una base de datos vía telnet establece una sesión de trabajo interactiva con el programa que gestiona la base de datos, que le permite aplicar todas las posibilidades de interrogación que tenga el sistema: selección, combinación y visualización o impresión de resultados. En cualquier momento podrá visualizar todas las búsquedas realizadas hasta ese instante y establecer combinaciones entre ellas.
  - ◆ **Acceso vía web:** conexión a través de un formulario existente en una página web de Internet, diseñado para lanzar preguntas a una base de datos.

Una misma base de datos puede tener acceso local y además una edición en CD-ROM y un sistema de acceso en línea. Sin embargo, puede haber diferencias en el contenido presente en cada uno de estos formatos o en el grado de actualización de la información. Por ejemplo, el productor de una base de datos puede ofrecer la conexión en línea a la base de datos completa con actualización diaria y, en cambio, editar un CD-ROM que tan sólo contenga los últimos cinco años de información y se actualice semestralmente.

## Bases de datos según cobertura temática

- ✔ **Bases de datos científico-tecnológicas:** contienen información destinada a los investigadores de cualquier ámbito científico o técnico. A su vez, este grupo puede dividirse en:
  - ◆ **Bases de datos multidisciplinarias:** abarcan varias disciplinas científicas o técnicas.
  - ◆ **Bases de datos especializadas:** recopilan y analizan documentos pertinentes para una disciplina o subdisciplina concreta: investigación biomédica, farmacéutica, química, agroalimentaria, social, humanística, etc.
- ✔ **Bases de datos económico-empresariales:** contienen información de interés para empresas, entidades financieras,....
- ✔ **Bases de datos de medios de comunicación:** contienen información de interés para los profesionales de medios de comunicación de masas: prensa, radio, televisión,...
- ✔ **Bases de datos del ámbito político-administrativo y jurídico:** contienen información de interés para los organismos de la administración y los profesionales del Derecho: legislación, jurisprudencia,...
- ✔ **Bases de datos del ámbito sanitario:** además de las propias del primer grupo especializadas en ciencias de la salud, existen otros sistemas con información de interés sanitario: historiales médicos, archivos hospitalarios,...
- ✔ **Bases de datos para el gran público:** contienen información destinada a cubrir necesidades de información general, de interés para un gran número de usuarios.

# Sistemas gestores de base de datos (6)

**Sistema Gestor de Base de Datos:** Conjunto coordinado de **programas, procedimientos, lenguajes**, 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 **base de datos**, **manteniendo** su **integridad, confidencialidad y seguridad**.

El SGBD **permite** a los **usuarios** la **creación** y el **mantenimiento** de una **base de datos**, facilitando la **definición, construcción y manipulación** de la **información** contenida en ésta. **Definir** una **base de datos** consistirá en **especificar** los **tipos de datos**, las **estructuras** y las **restricciones** que han de **cumplir los datos** a la **hora de almacenarse** en esta.

Por otro lado, la **construcción** de la **base** será el **proceso** de **almacenamiento** de datos **concretos** en algún **medio** o **soporte** de almacenamiento que esté **supervisado** por el SGBD.

Finalmente, la **manipulación** de la **base de datos** incluirá la **posibilidad** de **realización** de **consultas** para **recuperar información** específica, la **actualización** de los datos y la **generación** de **informes a partir** de su **contenido**.

## Ventajas

**Proporcionan** al **usuario** una **visión abstracta** de los **datos**, **ocultando** parte de la **complejidad** relacionada con **cómo se almacenan y mantienen** los datos.

Ofrecen **Independencia física**: la **información** que ve el **usuario**, y la **manipulación** de los **datos** almacenados en la Base de Datos, es **independiente** de **cómo** estén **almacenados físicamente**.

**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**.

## Funciones (6.1)

Descripción o Definición	
<p><b>Permite</b> al <b>diseñador</b> de la base de datos <b>crear</b> las <b>estructuras apropiadas</b> para <b>integrar adecuadamente</b> los <b>datos</b>, permite <b>definir</b> las <b>tres estructuras</b> de la base de datos.</p> <p>Esta <b>función</b> se <b>realiza mediante</b> el <b>lenguaje de descripción de datos</b> o <b>DDL</b>.</p> <p>Mediante ese lenguaje: se <b>definen</b>:</p> <ul style="list-style-type: none"><li>• Las <b>estructuras</b> de <b>datos</b>,</li><li>• Las <b>relaciones</b> entre los <b>datos</b></li><li>• Las <b>reglas</b> (restricciones) que han de cumplir los datos.</li></ul> <p>Se <b>especificarán</b> las <b>características</b> de los <b>datos</b> a <b>cada uno</b> de los <b>tres niveles</b> y el <b>SGBD</b> se <b>ocupará</b> de la <b>transformación</b> de las <b>estructuras externas</b> orientadas a los usuarios a las <b>estructuras conceptuales</b> y de la <b>relación</b> de ésta y la <b>estructura física</b>.</p>	
<b>Interno</b>	<b>Indicar</b> el <b>espacio</b> de <b>disco</b> reservado para la base de datos, la <b>longitud</b> de los <b>campos</b> , su <b>modo</b> de <b>representación</b> (lenguaje para la definición de la estructura externa).
<b>Conceptual</b>	<p><b>Proporcionan herramientas</b> para la <b>definición</b> de las <b>entidades</b> y su <b>identificación</b>, <b>atributos</b> de las mismas, <b>interrelaciones</b> entre ellas, <b>restricciones</b> de <b>integridad</b>, etc.</p> <p>Es decir, el <b>esquema</b> de la <b>base</b> de <b>datos</b> (lenguaje para la definición de estructura lógico global).</p>
<b>Externo</b>	<b>Definir</b> las <b>vistas</b> de los <b>distintos usuarios</b> a <b>través</b> del <b>lenguaje</b> para la <b>definición</b> de <b>estructuras externas</b> .

## Manipulación

Permite a los **usuarios** de la base **buscar, añadir, suprimir o modificar** los **datos** de la misma, siempre de **acuerdo** con las **especificaciones** y las **normas** de **seguridad dictadas** por el **administrador**. Se llevará **a cabo** por medio de un **lenguaje de manipulación de datos** (DML) que **facilita** los **instrumentos necesarios** para la **realización** de estas **tareas**.

También se encarga de **definir** la **vista externa** de **todos** los **usuarios** de la **base de datos** o **vistas parciales** que **cada usuario** tiene **de los datos definidos** con el **DDL**.

**Por manipulación de datos entenderemos:**

La **recuperación** de **información almacenada** en la **base de datos**, llamado, **consultas**.

La **inserción** de **información nueva**.

El **borrado** de **información**.

La **modificación** de **información almacenada**.

## Control

Permite al **administrador** de la base de datos **establecer mecanismos** de **protección** de las **diferentes visiones** de los **datos asociadas** a **cada usuario**, **proporcionando elementos** de **creación** y **modificación** de dichos **usuarios**.

Adicionalmente, **incorpora sistemas** para la **creación** de **copias** de **seguridad**, **carga** de **ficheros**, **auditoría**, **protección** de **ataques**, **configuración** del **sistema**, etc.

El **lenguaje** que **implementa** esta **función** es el **lenguaje de control de datos** o **DCL**



## Componentes (6.2)

Lenguajes de la base de datos	<p>A través de los lenguajes se pueden <b>especificar</b> los <b>datos</b> que <b>componen</b> la <b>BD</b>, su <b>estructura</b>, <b>relaciones</b>, <b>reglas de integridad</b>, <b>control de acceso</b>, <b>características físicas</b> y <b>vistas externas</b> de los usuarios.</p> <p>Los lenguajes del SGBD son:</p> <ul style="list-style-type: none"><li>• Lenguaje de Definición de los Datos (<b>DDL</b>)</li><li>• Lenguaje de Manejo de Datos (<b>DML</b>)</li><li>• Lenguaje de Control de Datos (<b>DCL</b>).</li></ul>
El diccionario de datos	<p><b>Descripción</b> de los <b>datos almacenados</b>.</p> <p>Es el <b>lugar</b> donde se <b>deposita</b> la <b>información</b> sobre la <b>totalidad</b> de los <b>datos</b> que <b>forman</b> la <b>base de datos</b>.</p> <p>Contiene las <b>características lógicas</b> de las <b>estructuras</b> que <b>almacenan</b> los <b>datos</b>, su <b>nombre</b>, <b>descripción</b>, <b>contenido</b> y <b>organización</b>.</p> <p>El diccionario de datos aportará información sobre:</p> <ul style="list-style-type: none"><li>• <b>Estructura lógica y física</b> de la BD.</li><li>• <b>Definición</b> de <b>tablas</b>, <b>vistas</b>, <b>índices</b>, <b>disparadores</b>, <b>procedimientos</b>, <b>funciones</b>.</li><li>• <b>Cantidad</b> de <b>espacio asignado</b> y <b>utilizado</b> por los <b>elementos</b> de la BD.</li><li>• <b>Descripción</b> de las <b>restricciones de integridad</b>.</li><li>• <b>Información</b> sobre los <b>permisos asociados</b> a cada usuario.</li><li>• <b>Auditoría</b> de <b>acceso</b> a los <b>datos</b>, <b>utilización</b>.</li></ul>
El gestor de la base de datos	<p>Es la <b>parte</b> de <b>software encargada</b> de <b>garantizar</b> el <b>correcto</b>, <b>seguro</b>, <b>íntegro</b> y <b>eficiente acceso</b> y <b>almacenamiento</b> de los <b>datos</b>.</p> <p>Es el <b>encargado</b> de:</p> <ul style="list-style-type: none"><li>• <b>Proporcionar</b> una <b>interfaz</b> entre los <b>datos almacenados</b> y los <b>programas de aplicación</b> que los manejan.</li></ul>

	<ul style="list-style-type: none"> <li>• <b>Garantizar la privacidad, seguridad e integridad de los datos, controlando los accesos concurrentes e interactuando con el sistema operativo</b></li> </ul> <p>Es un <b>intermediario</b> entre el <b>usuario</b> y los <b>datos</b>.</p>
<b>Usuarios de la base de datos</b>	<p>En los SGBD existen <b>diferentes perfiles de usuario, cada uno, con unos permisos</b> sobre los <b>objetos</b> de la BD:</p> <p><b>El administrador de la base de datos</b> (DBA): <b>persona o conjunto de ellas encargadas de la función de administración</b> de la base de datos.</p> <p>Tiene el <b>control centralizado</b> de la BD y es el <b>responsable</b> de su <b>buen funcionamiento</b>. Es el <b>encargado de autorizar el acceso</b> a la <b>base de datos</b>, de <b>coordinar y vigilar su utilización</b> y de <b>adquirir los recursos software y hardware</b> que sean <b>necesarios</b>.</p> <p><b>Los usuarios de la base de datos:</b> <b>usuarios</b> de la BD con <b>diferentes necesidades</b> sobre los datos, <b>diferentes accesos y privilegios</b>.</p> <p>Podemos establecer la siguiente clasificación:</p> <ul style="list-style-type: none"> <li>• Diseñadores.</li> <li>• Operadores y personal de mantenimiento.</li> <li>• Analistas y programadores de aplicaciones.</li> <li>• Usuarios finales: ocasionales, simples, avanzados y autónomos.</li> </ul>
<b>Herramientas de la base de datos</b>	<p><b>Conjunto de aplicaciones que permiten a los administradores la gestión de la base de datos, de los usuarios y permisos, generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones...</b></p>

## Componentes (6.3)

Para una **base de datos**, sólo existirá un **único esquema interno**, un **único esquema conceptual** y podrían **existir varios esquemas externos definidos** para uno o varios usuarios.

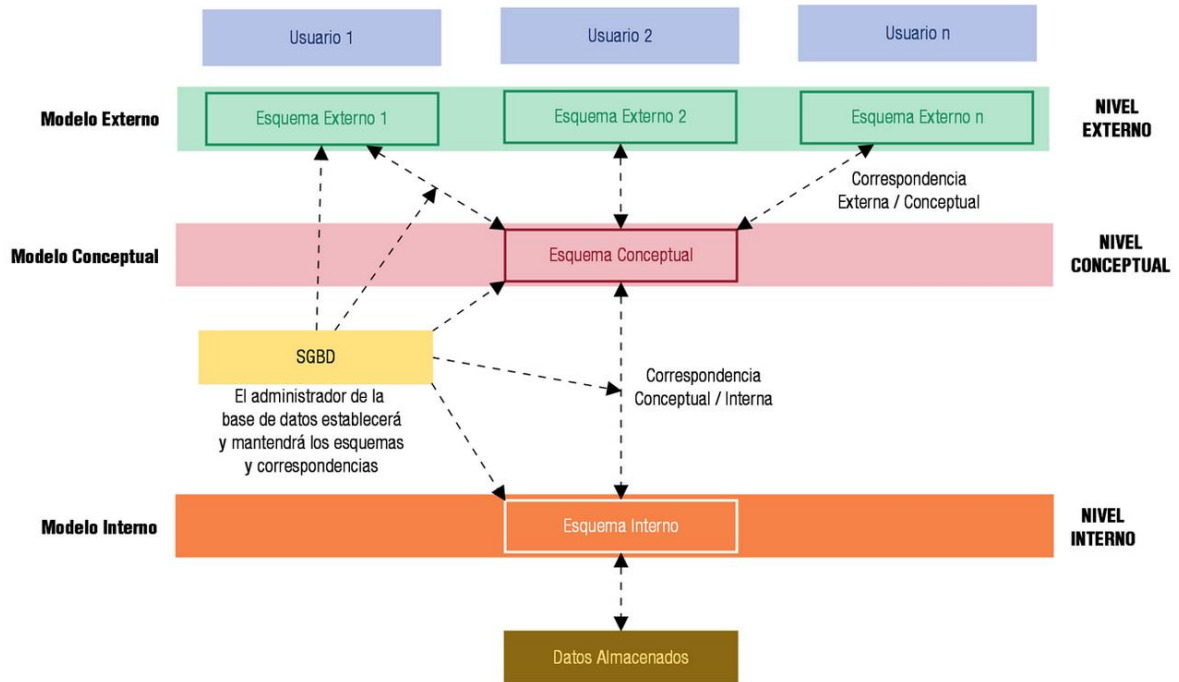
La arquitectura propuesta proporciona **tres niveles de abstracción**:

<b>Nivel interno o físico</b>	<p>Se <b>describe</b> la <b>estructura física</b> de la <b>BD</b> a través de un <b>esquema interno</b> encargado de <b>detallar</b> el <b>sistema de almacenamiento</b> de la <b>BD</b> y sus <b>métodos de acceso</b>.</p> <p>Es el <b>nivel más cercano</b> al <b>almacenamiento físico</b>.</p> <p>A <b>través</b> del <b>esquema físico</b> se <b>indican</b>, los <b>archivos</b> que <b>contienen</b> la <b>información</b>, su <b>organización</b>, los <b>métodos de acceso</b> a los <b>registros</b>, los <b>tipos de registros</b>, la <b>longitud</b>, los <b>campos</b> que los <b>componen</b>, las <b>unidades de almacenamiento...</b></p>
<b>Nivel lógico o conceptual</b>	<p>Se <b>describe</b> la <b>estructura completa</b> de la <b>BD</b> a través de un <b>esquema</b> que <b>detalla</b> las <b>entidades, atributos, relaciones, operaciones</b> de los <b>usuarios</b> y <b>restricciones</b>.</p> <p>Los <b>detalles relacionados</b> con las <b>estructuras de almacenamiento</b> se <b>ocultan</b>, permitiendo realizar una <b>abstracción</b> a más <b>alto nivel</b>.</p>
<b>Nivel externo o de visión del usuario</b>	<p>Se describen las <b>diferentes vistas</b> que los <b>usuarios percibirán</b> de <b>BD</b>.</p> <p><b>Cada tipo de usuario o grupo</b> de ellos <b>verá sólo</b> la <b>parte</b> de la <b>BD</b> que le <b>interesa, ocultando</b> el resto.</p>

Gracias a esta arquitectura se consigue la **independencia de datos** a dos niveles:

<b>Independencia lógica</b>	Podemos <b>modificar</b> el <b>esquema conceptual</b> sin <b>alterar</b> los <b>esquemas externos</b> ni los <b>programas de aplicación</b> .
<b>Independencia física</b>	<p>Podemos <b>modificar</b> el <b>esquema interno</b> sin <b>necesidad</b> de <b>modificar</b> el <b>conceptual</b> o el <b>externo</b>.</p> <p>Se puede <b>cambiar</b> el <b>sistema de almacenamiento</b>, <b>reorganizar</b> los <b>ficheros</b>, <b>añadir nuevos...</b>, sin que <b>esto afecte</b> al <b>resto de esquemas</b>.</p>

# ARQUITECTURA SGBD



## Tipos (6.4)

Según el modelo lógico en que se basan

**Actualmente**, el **modelo lógico** que **más** se **utiliza** es el relacional.

Los modelos en **red** y **jerárquico** han quedado **obsoletos**.

Otro de los modelos que **más extensión** está teniendo es el **modelo orientado a objetos**.

Por tanto, en esta primera clasificación tendremos:

- Modelo Jerárquico.
- Modelo de Red.
- Modelo Relacional.
- Modelo Orientado a Objetos.

Según el número de usuarios

<b>Monousuario</b>	<b>Sólo</b> atienden a <b>un usuario a la vez</b> , y su <b>principal uso</b> se da en los <b>ordenadores personales</b> .
<b>Multiusuario</b>	Entre los que se encuentran la mayor parte de los SGBD, <b>atienden a varios usuarios al mismo tiempo</b> .

Según el número de sitios en los que está distribuida la base de datos

<b>Centralizados</b>	<p>Sus <b>datos</b> se <b>almacenan</b> en <b>un solo</b> ordenador.</p> <p>Los SGBD centralizados <b>atienden a varios usuarios, pero</b> el <b>SGBD</b> y la <b>BD residen por completo</b> en una <b>sola máquina</b>.</p>
<b>Distribuidos</b>	<p>La <b>BD real</b> y el <b>propio software</b> del <b>SGBD</b> están <b>distribuidos</b> en <b>varios sitios conectados</b> por una <b>red</b>.</p> <p>Los <b>sistemas homogéneos</b> utilizan el <b>mismo SGBD</b> en <b>múltiples sitios</b>.</p> <p>Una tendencia reciente consiste en crear software para tener acceso a varias bases de datos autónomas preexistentes almacenadas en sistemas distribuidos heterogéneos.</p> <p>Esto da lugar a los <b>SGBD federados</b> o <b>sistemas multibase</b> de datos en los que los <b>SGBD</b> participantes <b>tienen cierto grado de autonomía local</b>.</p>

### Según el coste

La mayor parte de los paquetes cuestan entre 10.000 y 100.000 euros. Los sistemas monousuario más económicos para microcomputadores cuestan entre 0 y 3.000 euros. En el otro extremo, los paquetes más completos cuestan más de 100.000 euros.

### Según el propósito o finalidad

#### General

Utilizados para el **tratamiento** de **cualquier** tipo **BD** y **aplicación**.

#### 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**.

Muchos **sistemas** de **reservas** de **líneas aéreas** son de **propósito especial** y **pertenecen** a la **categoría** de **sistemas de procesamiento de transacciones en línea**, que deben **atender** un **gran número** de **transacciones concurrentes sin imponer excesivos retrasos**.

### Sistemas Gestores de Bases de Datos Comerciales.

SGBD	Descripción	URL
ORACLE	Reconocido como uno de los mejores a nivel mundial. Es multiplataforma, potente a nivel transaccional, confiable y seguro. Es Cliente/Servidor. Basado en el modelo de datos Relacional. De gran potencia, aunque con un precio elevado hace que sólo se vea en empresas muy grandes y multinacionales. Ofrece versiones gratuitas de las cuales la última es Oracle Database 18c Express Edition (XE)	<a href="#">Oracle</a>
MYSQL	Sistema muy extendido que se ofrece bajo dos tipos de licencia, comercial o libre. Para aquellas empresas que deseen incorporarlo en productos privativos, deben comprar una licencia específica. Es Relacional, Multihilo, Multiusuario y Multiplataforma. Su gran velocidad lo hace ideal para consulta de bases de datos y plataformas web.	<a href="#">MySQL</a>
DB2	Multiplataforma, el motor de base de datos relacional integra XML de manera nativa, lo que IBM ha llamado pureXML, que permite almacenar documentos completos para realizar operaciones y búsquedas de manera jerárquica dentro de éste, e integrarlo con búsquedas relacionales.	<a href="#">DB2</a>
INFORMIX	Otra opción de IBM para el mundo empresarial que necesita un DBMS sencillo y confiable. Es un gestor de base de datos relacional basado en SQL. Multiplataforma. Consume menos recursos que Oracle, con utilidades muy avanzadas respecto a conectividad y funciones relacionadas con tecnologías de Internet/Intranet, XML, etc.	<a href="#">Informix</a>
Microsoft SQL SERVER	Sistema Gestor de Base de Datos producido por Microsoft. Es relacional, sólo funciona bajo Microsoft Windows, utiliza arquitectura Cliente/Servidor. Constituye la alternativa a otros potentes SGBD como son Oracle, PostgreSQL o MySQL.	<a href="#">Microsoft SQL Server 2017</a>
SYBASE	Un DBMS con bastantes años en el mercado, tiene 3 versiones para ajustarse a las necesidades reales de cada empresa. Es un sistema relacional, altamente escalable, de alto rendimiento, con soporte a grandes volúmenes de datos, transacciones y usuarios, y de bajo costo.	<a href="#">Sybase</a>

### Sistemas Gestores de Bases de Datos Libres.

SGBD	Descripción	URL
MySQL	Es un sistema de gestión de base de datos relacional, multihilo y multiusuario con más de seis millones de instalaciones. Distribuido bajo dos tipos de licencias, comercial y libre. Multiplataforma, posee varios motores de almacenamiento, accesible a través de múltiples lenguajes de programación y muy ligado a aplicaciones web.	<a href="#">MySQL</a>
PostgreSQL	Sistema Relacional Orientado a Objetos. Considerado como la base de datos de código abierto más avanzada del mundo. Desarrollado por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre y/o apoyados por organizaciones comerciales. Es multiplataforma y accesible desde múltiples lenguajes de programación.	<a href="#">PostgreSQL</a>
Firebird	Sistema Gestor de Base de Datos relacional, multiplataforma, con bajo consumo de recursos, excelente gestión de la concurrencia, alto rendimiento y potente soporte para diferentes lenguajes.	<a href="#">Firebird</a>
Apache Derby	Sistema Gestor escrito en Java, de reducido tamaño, con soporte multilenguaje, multiplataforma, altamente portable, puede funcionar embebido o en modo cliente/servidor.	<a href="#">Apache Derby</a>
SQLite	Sistema relacional, basado en una biblioteca escrita en C que interactúa directamente con los programas, reduce los tiempos de acceso siendo más rápido que MySQL o PostgreSQL, es multiplataforma y con soporte para varios lenguajes de programación.	<a href="#">SQLite</a>

# Bases de datos centralizadas (9)

**Sistema de base de datos 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.

Características
Se almacena completamente en una ubicación central, es decir, todos los componentes del sistema residen en un solo computador o sitio.
No posee múltiples elementos de procesamiento ni mecanismos de intercomunicación como las bases de datos distribuidas.
Los componentes de las bases de datos centralizadas son: los datos, el software de gestión de bases de datos y los dispositivos de almacenamiento secundario asociados.
Son sistemas en los que su seguridad puede verse comprometida más fácilmente.

<b>Ventajas</b>	<p>Se evita la redundancia debido a la posibilidad de inconsistencias y al desperdicio de espacio.</p> <p>Se evita la inconsistencia. Ya que, si un hecho específico se representa por una sola entrada, la no-concordancia de datos no puede ocurrir.</p> <p>La seguridad se centraliza.</p> <p>Puede conservarse la integridad.</p> <p>El procesamiento de los datos ofrece un mejor rendimiento.</p> <p>Mantenimiento más barato. Mejor uso de los recursos y menores recursos humanos.</p>
<b>Inconvenientes</b>	<p>Un mainframe en comparación de un sistema distribuido no tiene mayor poder de cómputo.</p> <p>Cuando un sistema de bases de datos centralizado falla, se pierde toda disponibilidad de procesamiento y sobre todo de información confiada al sistema.</p> <p>En caso de un desastre o catástrofe, la recuperación es difícil de sincronizar.</p> <p>Las cargas de trabajo no se pueden difundir entre varias computadoras, ya que los trabajos siempre se ejecutarán en la misma máquina.</p>



	<p>Los departamentos de sistemas retienen el control de toda la organización.</p> <p>Los sistemas centralizados requieren un mantenimiento central de datos.</p>
--	--

## Bases de datos distribuidas (10)

<b>Base de datos distribuida (BDD)</b>	Conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas entre diferentes nodos interconectados por una red de comunicaciones.
<b>Sistema de bases de datos distribuida (SBDD)</b>	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.
<b>Sistema gestor de bases de datos distribuida (SGBDD)</b>	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 SGBD ejecutado en una sola máquina administrara esos datos.

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 base de datos local, un sistema de gestor de bases de datos e interconectados entre sí. Si estos nodos están dispersos geográficamente se interconectarán a través de una red de área amplia o WAN, pero si se encuentran en edificios relativamente cercanos, pueden estar interconectados por una red de área local o LAN. Este tipo de sistemas es utilizado en: organizaciones con estructura descentralizada, industrias de manufactura con múltiples sedes (automoción), aplicaciones militares, líneas aéreas, cadenas hoteleras, servicios bancarios, etc.

<b>Ventajas</b>	<p>El acceso y procesamiento de los datos es más rápido ya que varios nodos comparten carga de trabajo.</p> <p>Desde una ubicación puede accederse a información alojada en diferentes lugares.</p>
-----------------	---

	<p>Los costes son inferiores a los de las bases centralizadas.</p> <p>Existe cierta tolerancia a fallos. Mediante la <a href="#">replicación</a>, si un nodo deja de funcionar el sistema completo no deja de funcionar.</p> <p>El enfoque distribuido de las bases de datos se adapta más naturalmente a la estructura de las organizaciones. Permiten la incorporación de nodos de forma flexible y fácil.</p> <p>Aunque los nodos están interconectados, tienen independencia local.</p>
<b>Inconvenientes</b>	<p>La probabilidad de violaciones de seguridad es creciente si no se toman las precauciones debidas.</p> <p>Existe una complejidad añadida que es necesaria para garantizar la coordinación apropiada entre los nodos.</p> <p>La inversión inicial es menor, pero el mantenimiento y control puede resultar costoso.</p> <p>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.</p> <p>El intercambio de mensajes y el cómputo adicional necesario para conseguir la coordinación entre los distintos nodos constituyen una forma de sobrecarga que no surge en los sistemas centralizados.</p> <p>Dada la complejidad del procesamiento entre nodos es difícil asegurar la corrección de los algoritmos, el funcionamiento correcto durante un fallo o la recuperación.</p>

# Fragmentación (10.1)

Sabemos que en los sistemas de bases de datos distribuidas la información se encuentra repartida en varios lugares. La forma de extraer los datos consultados puede realizarse mediante la fragmentación de distintas tablas pertenecientes a distintas bases de datos que se encuentran en diferentes servidores. El problema de 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** que se aplicará, ya que éste es un factor determinante a la hora de la ejecución de consultas. Si no existe fragmentación, se tomarán las relaciones o tablas como la unidad de fragmentación. Pero también puede fragmentarse a nivel de tupla (fila o registro) o a nivel de atributo (columna o campo) de una tabla. No será adecuado un grado de fragmentación nulo, ni tampoco un grado de fragmentación demasiado alto. El grado de fragmentación deberá estar equilibrado y dependerá de las particularidades de las aplicaciones que utilicen dicha base de datos. Concretando, el objetivo de la fragmentación es encontrar un nivel de particionamiento adecuado en el rango que va desde tuplas o atributos hasta relaciones completas.

Reglas fundamentales a cumplir	
<b>Compleitud</b>	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 Ri
<b>Reconstrucción</b>	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
<b>Disyunción</b>	Si una relación R se descompone verticalmente, sus atributos primarios clave normalmente se repiten en todos sus fragmentos

Tipos de fragmentación	
<b>Horizontal</b>	<p>La fragmentación horizontal se realiza sobre las tuplas de la relación, dividiendo la relación en subrelaciones que contienen un subconjunto de las tuplas que alberga la primera.</p> <p>Existen dos variantes de la fragmentación horizontal: la primaria y la derivada.</p>
<b>Vertical</b>	<p>La fragmentación vertical, en cambio, se basa en los atributos de la relación para efectuar la división.</p>

	<p>Una relación R produce fragmentos R1, R2, ..., Rr, cada uno de los cuales contiene un subconjunto de los atributos de R así como la llave primaria de R.</p> <p>El objetivo de la fragmentación vertical es particionar una relación en un conjunto de relaciones más pequeñas de manera que varias de las aplicaciones de usuario se ejecutarán sobre un fragmento.</p> <p>En este contexto, una fragmentación óptima es aquella que produce un esquema de fragmentación que minimiza el tiempo de ejecución de las consultas de usuario.</p> <p>La fragmentación vertical es más complicada que la horizontal, ya que existe un gran número de alternativas para realizarla.</p>
<b>Híbrida o mixta</b>	<p>Podemos combinar ambas, utilizando por ello la denominada fragmentación mixta.</p> <p>Si tras una fragmentación vertical se lleva a cabo otra horizontal, se habla de la fragmentación mixta (HV).</p> <p>Para el caso contrario, estaremos ante una fragmentación (VH). Para representar los dos tipos de fragmentación, se utilizan los árboles.</p>