

T01: Introducción a bases de datos. Elementos.

1. Introducción.....	1
2. Definición de Base de Datos.....	2
3. Evolución histórica y tipos de bases de datos.....	3
3.1. Introducción.....	3
3.2. Evolución y tipos de Bases de Datos.....	3
Base de datos Jerárquica.....	3
Base de datos Red.....	4
Base de datos Relacional.....	5
Base de datos con estructura Multidimensional.....	6
Base de datos Orientada a Objetos.....	6
3.3. Sistemas de Ficheros Tradicionales.....	7
Tipos de Ficheros Tradicionales.....	8
Dispositivos de almacenamiento.....	10
3.4. Ejemplo de Ficheros Tradicionales.....	11
3.5. Sistema de Base de Datos Relacional.....	12
4. Conceptos básicos de una Base de Datos.....	13
4.1. Datos.....	13
4.2. Entidades.....	14
4.3. Claves primarias y claves foráneas. Relaciones.....	14
4.4. Restricción de integridad referencial.....	15
4.5. Metadatos.....	15
4.6. Otros conceptos de Bases de Datos.....	16
5. Sistemas de Gestión de Bases de Datos.....	17
5.1. Sistema Gestor de Base de Datos.....	17
5.2 Tipos de SGBD.....	17
6. Ejercicios.....	18
7. Planteamiento práctico.....	19
Problemática de una inmobiliaria.....	20
Solución propuesta según Sistema de Ficheros.....	21
Cuestiones.....	23

1. Introducción.

En el entorno del mercado actual, la competitividad y la rapidez de maniobra de una empresa son imprescindibles para su éxito. Para conseguirlo existe cada vez una mayor demanda de datos y, por tanto, más necesidad de gestionarlos. Esta demanda siempre ha estado patente en empresas y sociedades, pero cada vez aumenta más esa demanda debido al acceso multitudinario a través de Internet y al uso de dispositivos móviles que también requieren esa información.

La información es poder.

En informática se conoce como **dato** a cualquier **elemento informativo que tenga relevancia para un usuario**. Desde su nacimiento, la informática se ha encargado de proporcionar herramientas que faciliten la manipulación de los datos. Antes de la aparición de las aplicaciones informáticas, las empresas tenían como únicas herramientas de gestión de datos los ficheros con cajones, carpetas y fichas de cartón. En este proceso manual, el tiempo requerido para manipular estos datos era enorme.

Al aparecer la informática, se adaptan sus herramientas para que los elementos que el usuario utiliza manualmente en el manejo de datos se parezcan a los procesos informatizados. Por eso se sigue hablado de ficheros, formularios, carpetas, directorios,....

La clientela fundamental del profesional informático es la empresa. La empresa se puede entender como un **Sistema de Información** formado por diversos objetos: el capital, los recursos humanos, los inmuebles, los servicios que presta, etc..

Los sistemas de información actuales se basan en **Bases de Datos (BD)** y **Sistemas Gestores de Bases de Datos (SGBD)** que se han convertido en elementos imprescindibles de la vida cotidiana de la sociedad moderna.

2. Definición de Base de Datos.

Cada día, la mayoría de nosotros nos encontramos con actividades que requieren algún tipo de interacción con una Base de Datos (ingreso en un banco, reserva de una entrada para el teatro, solicitud de una suscripción a una revista, compra de productos,...). Estas interacciones son ejemplos de lo que se llama aplicaciones tradicionales de bases de datos, en las que básicamente se trabaja con información numérica o de texto.

Los avances tecnológicos han permitido que las bases de datos también puedan contener otros tipos de datos. Así podemos hablar de:

- Bases de datos multimedia. Cuyos contenidos son elementos multimedia: imágenes, sonidos o fragmentos de vídeo.
- Bases de datos espaciales o geográficas. Almacenan mapas y símbolos que representan superficies geográficas. Ejemplo: Google Earth realiza consultas a bases de datos de este tipo.
- Bases de datos documentales. Permiten indexar texto para poder realizar búsquedas de textos de gran longitud.
- Bases de datos deductivas. Almacenan hechos y reglas por las que se rigen los hechos. Esto permite mediante procedimientos de inferencia extraer nuevos hechos.

Podemos definir datos como cualquier hechos conocidos que pueden registrarse y que tienen un significado **implícito**.

Y por Base de Datos a una colección de datos perteneciente al mismo contexto que se almacenarán en ficheros. La base de datos se organiza en tablas **relacionadas** que recopilan los datos de cada objeto o suceso del que trata el Sistema de Información. Las tablas tienen un aspecto similar a una hoja de cálculo donde la información se dispone en filas y columnas.

Ejemplo:

Una agenda con los nombres y teléfonos de un conjunto de personas conocidas es una Base de Datos, puesto que es una colección de datos relacionados con un significado implícito.

La definición presentada anteriormente hace referencia a dos elementos para que un conjunto de datos constituya una Base de Datos:

1. **Relaciones** entre tablas, tema que se tratará en las secciones siguientes.
2. Significado **implícito** de los datos que se atribuye dependiendo del contexto en que se utilizan los mismos. Por ejemplo, el dato fecha en una Base de Datos de VENTAS puede referirse a la fecha de emisión de las facturas, mientras que si la Base de Datos es de MÚSICA quizás corresponda a la fecha en que se grabó un tema musical. Es decir, el significado de un dato, depende de la BD que lo contenga.

Para manipular y gestionar las bases de datos surgieron herramientas software denominadas: Sistemas Gestores de Bases de Datos (SGBD en lo sucesivo).

3. Evolución histórica y tipos de bases de datos.

3.1. Introducción.

Los predecesores de los Sistemas Gestores de Bases de Datos fueron los Sistemas Gestores de Ficheros o Sistemas de Archivos Tradicionales.

1. Ficheros Tradicionales. Consiste en almacenar los datos en ficheros individuales, exclusivos para cada aplicación particular. En este sistema los datos pueden ser redundantes (repetidos innecesariamente) y la actualización de los ficheros es más lenta que en una Base de Datos.

2. Base de Datos. Es un almacenamiento de datos formalmente definido, controlado centralmente para intentar servir a múltiples y diferentes aplicaciones. La Base de Datos es una fuente de datos que son compartidos por numerosos usuarios para diversas aplicaciones.

Así, en un Sistema de Ficheros Tradicional la información está dispersa en diferentes ficheros de datos y existe un cierto número de programas que los recuperan y agrupan. Aunque los Sistemas de Ficheros o Archivos supusieron un gran avance sobre los sistemas manuales, tienen inconvenientes bastante importantes que se solventaron, en gran medida, con la aparición de los Sistemas de Bases de Datos.

3.2. Evolución y tipos de Bases de Datos.

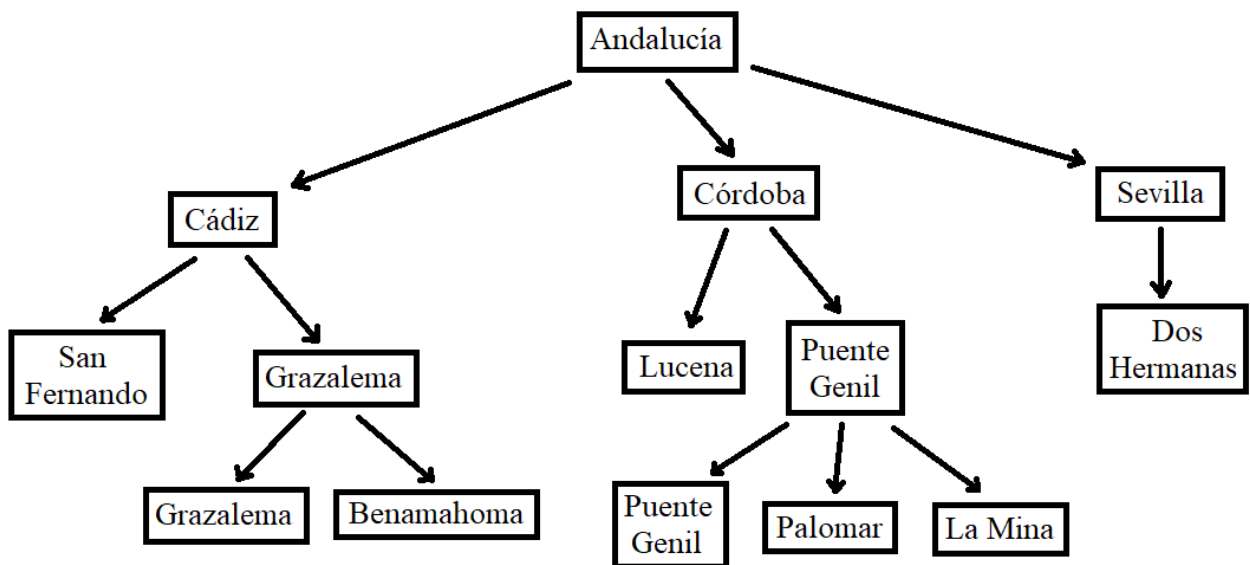
Coincidiendo con la evolución histórica de las Bases de Datos, éstas han utilizado distintos modelos:

- Jerárquicos.
- En red.
- Relacionales.
- Multi-dimensionales.
- De objetos.

Base de datos Jerárquica.

La estructura jerárquica fue usada en las primeras BD. Las relaciones entre registros forman una estructura en árbol. Actualmente las bases de datos jerárquicas más utilizadas son IMS de IBM y el Registro de Windows de Microsoft.

Veamos un ejemplo de la estructura de algunas poblaciones en Andalucía. Obsérvese que cada elemento de un determinado nivel está relacionado con uno y solo uno del nivel superior y puede estar relacionado con más de uno del nivel inferior.

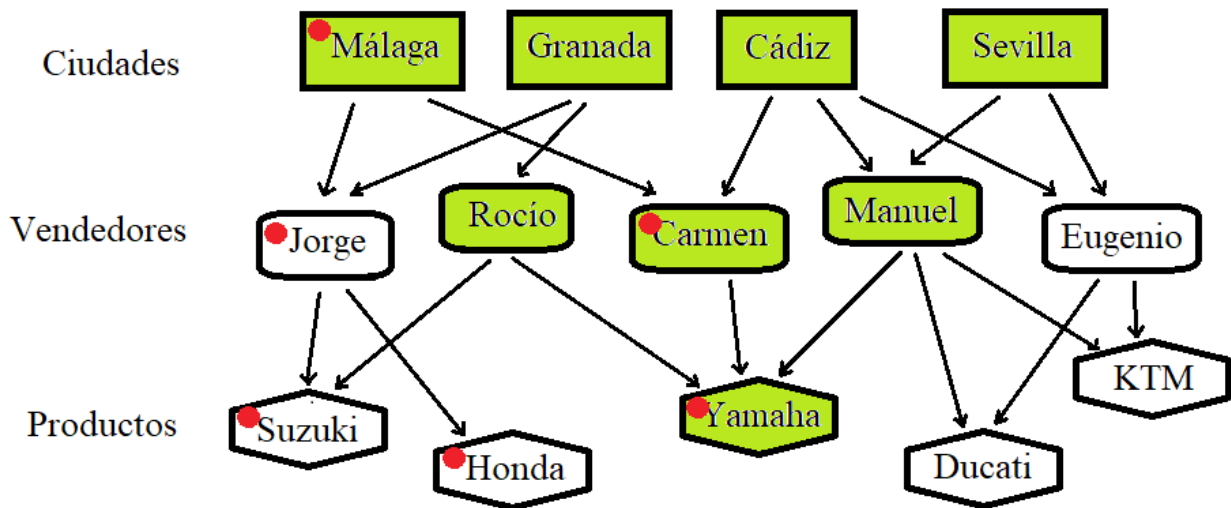


Base de datos Red.

Esta estructura contiene relaciones más complejas que las jerárquicas. Admite relaciones de los elementos de un determinado nivel con más de uno del nivel superior. Esto permite caminos alternativos para alcanzar algún elemento desde otro punto.

En el siguiente gráfico se representan la distribución de concesionarios y los vendedores en algunas provincias Andaluzas. Para comprar una moto Yamaha, puedo ir a Málaga, Granada, Cádiz o Sevilla. Incluso si estamos en Cádiz puedo elegir vendedor: Carmen o Manuel. Eso significa que podemos llegar a un nodo por distintos caminos.

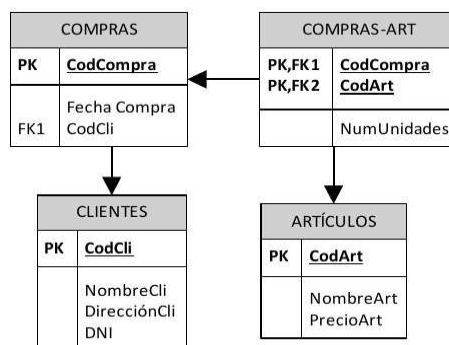
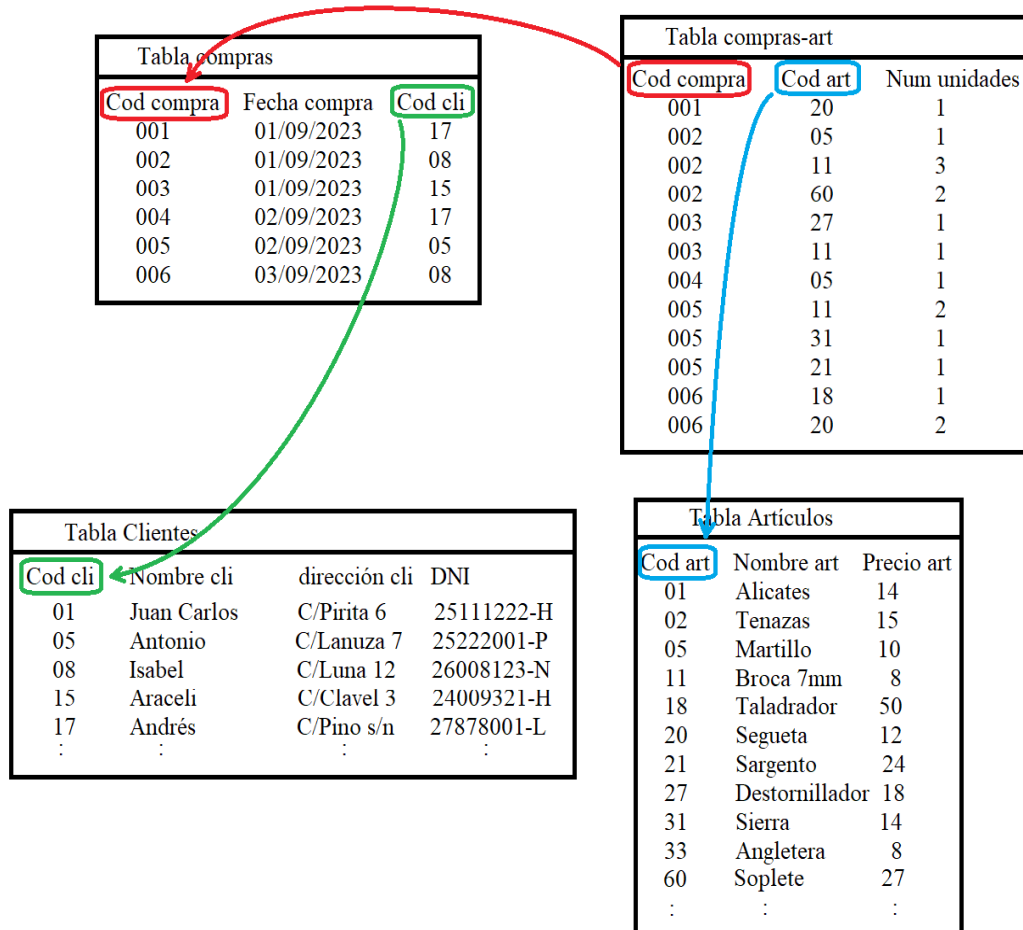
Por otra parte si estamos en Málaga y no nos queremos desplazar las opciones de compra son Suzuki, Honda y Yamaha. Las únicas marcas que disponen de vendedor en Málaga.



El inventor de este modelo fue Charles Bachman, y el estándar fue publicado en 1969 por CODASYL.

Base de datos Relacional.

La estructura relacional es la más extendida hoy en día. Almacena los datos en filas o registros (tuplas) y columnas o campos (atributos). Estas tablas pueden estar conectadas entre sí por claves comunes.

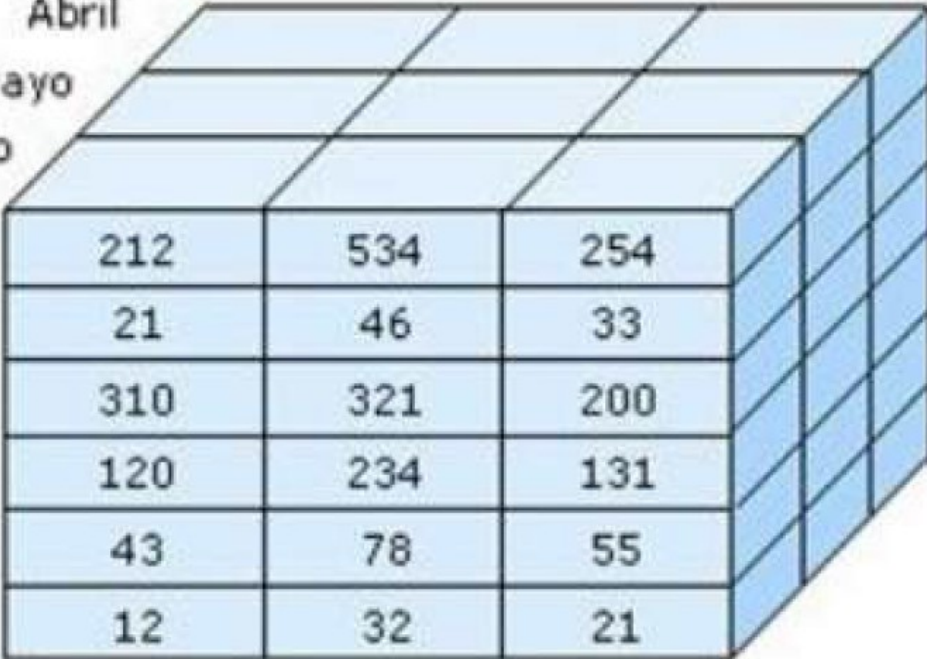


Durante el curso nos centramos en el estudio de bases de datos relacionales.

El origen de este modelo de Base de Datos fueron los estudios realizados por Edgar Frank Codd en 1970 en los que definió el modelo relacional basado en la lógica de predicados y la teoría de

conjuntos. El modelo de Codd, a pesar del tiempo, sigue siendo el más utilizado en todos los niveles en sus distintas implementaciones. Ejemplos: Oracle, SQL, DB2, PostgreSQL, MariaDB, MySQL,...

Base de datos con estructura Multidimensional.



	Abril	Mayo	Junio
Producto1	212	534	254
Producto2	21	46	33
Producto3	310	321	200
Producto4	120	234	131
Producto5	43	78	55
Producto6	12	32	21
	Argentina	Brasil	Chile

La estructura multidimensional tiene parecidos a la del modelo relacional, pero en vez de las dos dimensiones filas-columnas, tiene N dimensiones. Esta estructura se le denomina cubo.

Fue la solución dada al problema de trabajar con un enorme volumen de datos resultantes de la recopilación durante años de información. En ellas la dimensión temporal adquiere un significado relevante.

Base de datos Orientada a Objetos.

La estructura orientada a objetos está diseñada siguiendo el paradigma de los lenguajes orientados a objetos. De este modo soporta los tipos de datos gráficos, imágenes, voz y texto de manera natural. Esta estructura tiene gran difusión en aplicaciones web para aplicaciones multimedia.

Según esta filosofía en lugar de almacenar los datos en tablas y relaciones entre ellas, se almacenan objetos y el comportamiento que tiene cada uno ellos, es decir, instrucciones sobre como se han de procesar.



3.3. Sistemas de Ficheros Tradicionales.

En estos sistemas, cada programa almacena y utiliza sus propios datos. La única ventaja que conlleva esto es que los procesos son independientes, por lo que la modificación de uno no afecta al resto.

Pero tiene grandes inconvenientes:

- Datos redundantes. Ya que se repiten continuamente en cada aplicación que los utilice.
- Coste de almacenamiento elevado. Al almacenarse varias veces el mismo dato en distintas aplicaciones, se requiere más espacio en los discos.
- Tiempos de procesamiento elevados. Al no poder optimizar el espacio de almacenamiento. Si hay diferentes copias se ha de tener en cuenta cada una de ellas para mantener la consistencia.
- Probabilidad alta de inconsistencia en los datos. Ya que un proceso cambia sus datos, pero no las copias de las que disponen el resto de aplicaciones. Por lo que el mismo dato puede tener valores distintos según qué aplicación acceda a él.
- Difícil modificación en los datos. Debido a la probabilidad de inconsistencia, cada modificación se debe repetir en todas las copias del dato (algo que no siempre es posible).

En la siguiente figura se muestra un sistema de información basado en ficheros. En ella se ve que la información aparece inconexa y redundante.

JUGADORES		PORTEROS		BARCELONA		REAL_MADRID		JUG_LIGA_PASADA		JUG_LIGA_ACTUAL	
PK	<u>Codjugador</u>	PK	<u>Codjugador</u>	PK	<u>Codjugador</u>	PK	<u>Codjugador</u>	PK	<u>Codjugador</u>	PK	<u>Codjugador</u>
	Club jugador dorsal codpaís codDem		club jugador dorsal CodPaís NombDem		Club jugador dorsal codpaís codDem		Club jugador dorsal codpaís codDem		Club jugador dorsal codpaís codDem		Club jugador dorsal codpaís codDem

CLUBES		CLUB_JUG		ESTADIOS		EQUIPACIÓN		SOCIOS		ESPONSOR	
PK	<u>CodClub</u>	PK	<u>CodJugador</u>	PK	<u>CodEstadio</u>	PK	<u>CodEquip</u>	PK	<u>Numsocio</u>	PK	<u>CodEspon</u>
	Club Nombre Dirección Población Provincia CosPostal Tífn Colores Himno Fax AñoFundación presupuesto Presidente Vicepresidente CodEquip Sponsor		CodClub		CodEquipo Estadio Dirección CodPostal Población Provincia Capacidad Sentados Inauguración Dimensiones		NombreEquip Encasa		CodClub Codpost Nombre Apellidos Dirección Provincia FechaAlta CuotaActual		Esponsor

ENTRENADORES		PAÍSES		PALMARÉS		DEMARCACIÓN	
PK	<u>CodEntren</u>	PK	<u>CodPaís</u>	PK	<u>CodClub</u>	PK	<u>CodDem</u>
	CodClub Entrenador FechaNcmto Población Provincia		Nombrepais Comunitario		CodTrofeo		Demarcación

ENTREN_CLUBES		CAMPEONATO	
PK	<u>codCluB</u>	PK	<u>CodTrofeo</u>
	CodEntrenador		NombreTrofeo

El objetivo de este apartado es dar una visión de los Sistemas de Información basados en Sistemas de Ficheros Tradicionales.

Sistema de Información: toda la información que una empresa necesita para funcionar correctamente.

Esa información ha sido gestionada según la tecnología de su tiempo. Haciendo una descripción evolutiva podemos hablar de tres etapas:

- Basada en papel.
- Basada en Sistemas de Ficheros digitales. Al introducirse los ordenadores en las empresas todo el Sistema de Información se digitalizó. Fue una revolución que mejoró considerablemente el rendimiento de los sistemas. Básicamente se pasó de las estructuras de papel a digitalizarlas. Los problemas más graves que presentaban son:
 - Redundancia de datos y sus consecuencias.
 - Dependencia de datos respecto de los programas.
- Basada en Sistemas Gestores de Bases de Datos. Se desvinculan los datos de los programas, ya que son los SGBD son los que se encargan de almacenar la información y recuperarla. También se reduce la redundancia de los datos mejorando considerablemente la respuesta y eficiencia de los sistemas.

Tipos de Ficheros Tradicionales.

Según el análisis que se hace del Sistema de Información de una empresa cada Analista/Programador decide que estructuras de datos y tipos de ficheros va a utilizar que respondan

a los requerimientos del mismo. Además, cada lenguaje de programación puede ofrecer distintos tipos de ficheros o gestionarlos de distinta forma. A grandes rasgos todos se basaban en:

- La información a recopilar de cada elemento de estudio (**Entidad**) o sus relaciones (**Relación**) se guardaba en ficheros (**Tabla**) y toda la información que se recopile de una ocurrencia de esos elementos de estudio se almacena en una estructura fija (para todos es igual) llamada registro (**Tupla**).
- Los ficheros pueden tener distintas organizaciones según el uso que se les va a dar. Podemos distinguir tres clásicas: secuencial, indexado y directo.

Fichero secuencial.

Este tipo de fichero se caracteriza por tener la siguiente organización: cada registro se graba en el espacio dedicado al fichero según van surgiendo. El fichero empieza en el primer registro que se escribió y termina en el último registro guardado. Los registros están distribuidos en el orden en el que aparecieron.

Características:

- Reducción mínima de espacio. Todo el espacio ocupado corresponde a registros.
- Solo permite acceso secuencial. Para acceder a un determinado registro es necesario empezar por el principio del fichero y leer uno a uno hasta llegar al que se está buscando.
- Difícil de mantener: añadir, modificar o eliminar registros. Pues requiere en la mayoría de los casos reescribir el fichero.

Se utilizaba en ficheros muy estables (pocas actualizaciones) y ficheros históricos.

Pueden utilizar tanto cinta magnética como discos.

Fichero indexado.

Se caracteriza por tener dos áreas diferenciadas: una de índice y otra de datos.

Se ha de elegir alguna información de los registros como clave, que identifique de forma única al registro. Por ejemplo en alumnos del centro: código matrícula o DNI; en un fichero de artículos de un almacén: código de artículo;..

Área de índice: contendrá solo los códigos de los registros ordenados y un enlace al lugar exacto del disco duro donde está almacenado el registro completo. Para acceder a un registro concreto:

- Se solicita la clave.
- Se busca en la zona de índice.
- Se va directamente al registro siguiendo el enlace de la zona de índice.

Área de datos: Contiene los datos según van apareciendo. La mayoría de las implementaciones se mantiene ordenado por el campo clave.

Características:

- Permite acceso secuencial y aleatorio (llamado así pues no sabemos donde va a estar el registro al buscarlo, pues estará donde se escribió cuando llegó al sistema).
- Utiliza más espacio de almacenamiento que el secuencial: área de índice + área de datos.

Solo se puede utilizar en discos magnéticos.

Fichero directo.

También se ha de elegir un campo clave para poder acceder a los registros. A partir de la clave se utiliza un algoritmo que nos indica la posición relativa de ese registro dentro del fichero. Con esa posición el sistema va directamente a esa posición y espera encontrar ahí el registro.

Se produce un fenómeno de colisión: el campo clave de más de un registro puede ocurrir que el algoritmo nos lleve a la misma posición relativa. Para resolver esta situación se define una zona de colisión, donde se almacenarán secuencialmente los registros que den colisión.

Ej: la letra de los DNI. Hay un algoritmo que a partir del número de DNI nos calcula la letra que le corresponde. Esto se utiliza como medida de seguridad para evitar errores al teclear un DNI. Ocurre que hay más de un número con la misma letra.

Para que pueda funcionar esta organización el sistema ha de reservar espacio para todos los registros que se esperan recibir y de esa forma al aplicar el algoritmo se puede calcular donde estará el registro.

Esta organización se caracteriza por:

- Dos áreas definidas: de datos y de colisiones.
- Se ha de estimar el número de registros que se esperan y reservar su espacio en el disco duro a priori. Si tenemos más registros que los reservados la zona de colisión se disparará. Por otra parte, si dimensionamos por arriba, se quedarán huecos en la zona de datos sin cubrir. .

Ejemplo: Se estima que se matriculen 900 alumnos/as.

- Si finalmente aparecen 800. Queda sin utilizar espacio para 100 (11% de disco previsto perdido).
- Si finalmente aparecen 1000. Habrá al menos 100 colisiones, lo que ralentizará el sistema.
- Solo permite acceso directo.
- A más colisiones, mayor cantidad de huecos quedarán en la zona de datos y mayor será la zona de colisión. Esto hará que el acceso a los datos se relente.

Solo se puede utilizar en discos magnéticos.

Dispositivos de almacenamiento.

Clásicamente se han utilizado cintas y discos magnéticos.

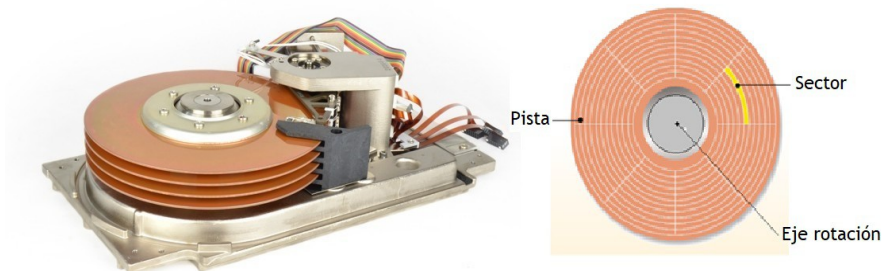
Cinta magnética.

Similar a una cinta de casete o vídeo. Se escribe una marca de inicio de fichero, su contenido y una marca de final. Esto permite poder escribir más de un fichero en la misma cinta.

Solo se puede tener un acceso secuencial.

Disco magnético.

La organización de un disco magnético es en: pista/cilindro, superficie y sectores.



El brazo de lectura/escritura se mueve acercándose o alejándose al centro de los discos y el disco está girando constantemente. La combinación de ambos movimientos hace que las cabezas de lectura/escritura se puedan posicionar sobre cualquier sector. Esto es un acceso aleatorio. Indicando: un cilindro, superficie y sector de forma aleatoria, rápidamente el sistema se posiciona y lee o escribe.

El concepto de cilindro indica todas las pistas que están a la misma distancia del centro. Por tanto un sector concreto se define por:

- Número de cilindro.
- Número de superficie. Un cilindro incluye todas las pistas que están a la misma distancia del centro. Cada superficie se identifica por un número según el disco en el que está.
- Y número de sector.

3.4. Ejemplo de Ficheros Tradicionales.

Supongamos un Sistema de Información que trabaja el proceso de facturación de una empresa. Para ello disponemos de dos ficheros: CLIENTES y FACTURAS.

El primer fichero tiene los datos básicos de los clientes, mientras que en el segundo se almacenan las ventas realizadas. Al emitir cada factura se recogen nuevamente los datos del cliente: num, nombre, domicilio,...

Fichero1:CLIENTES.

Num	Nombre	Dirección	Teléfono	FechaNacimiento	E-mail
1225	Juan García	Guaná 102	985674863	13/05/80	jgarcia@gmail.com
1226	Fernando Martínez	Rincón 10	985445522	20/07/93	fmar@hotmail.com
...

Fichero 2: FACTURAS.

NumFac	NumCli	Nombre	Dirección	Teléfono	Artículo	Total
101	1225	Joaquín García	Guaná 102	985674863	Pintura 25 litros	56
102	1226	Fernando Martínez	Rincón 10	985445522	Escalera 10 pasos	27
...

Desventajas:

- Se presentan redundancias de datos (datos repetidos innecesariamente: nombre, dirección y teléfono). Se duplican esfuerzos.

- Se pueden producir contradicciones entre los datos, si por ejemplo se introduzcan nombres diferentes para un mismo cliente (Juan por Joaquín).

3.5. Sistema de Base de Datos Relacional.

En este tipo de sistemas los datos se centralizan en una Base de Datos común a todas las aplicaciones.

Sus **ventajas** son las siguientes:

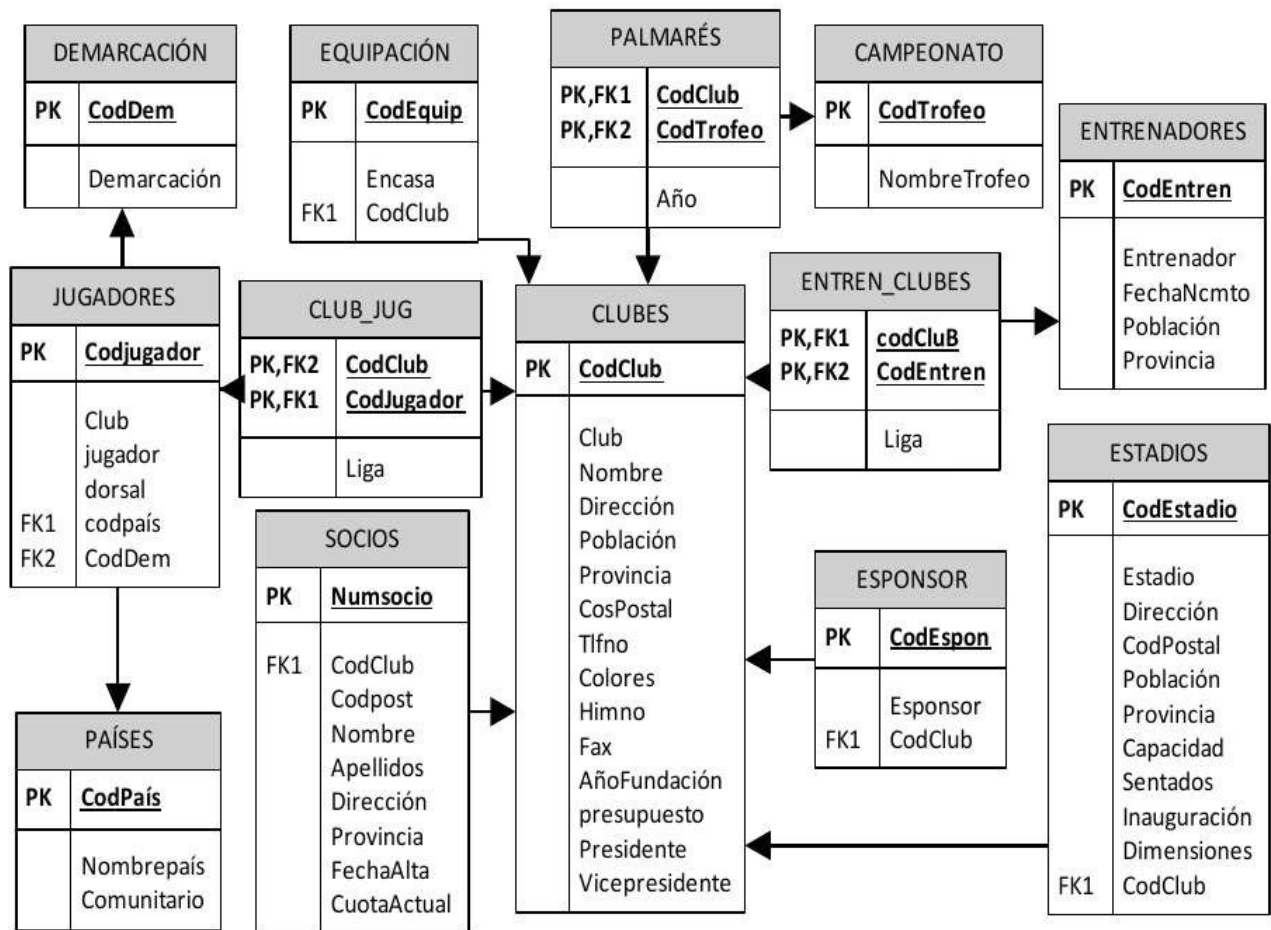
- Menor **redundancia**. No hace falta tanta repetición de datos. Requiere un buen diseño de la estructura de los datos.
- Menor espacio de almacenamiento. Debido a la reducción de redundancia de datos.
- Acceso a los datos más eficiente. La organización de los datos optimiza el rendimiento.
- Documentación de datos. Aparece el concepto de metadatos, que permiten describir la información que se almacena en la Base de Datos.
- **Independencia** de los datos respecto a los programas y procesos. Esto permite modificar los datos sin modificar el código de las aplicaciones.
- Integridad de los datos. Mayor dificultad de perder los datos o que aparezcan incoherencias en ellos.
- Mayor **seguridad** en los datos. Permite definir el perfil de cada usuario, según a la información a la que pueda acceder.

Como contrapartida encontramos los siguientes **inconvenientes**:

- **Instalación** costosa. El control y administración de bases de datos requiere de un software y hardware potente que requiere una instalación particular.
- **Requiere personal cualificado**. Debido a la dificultad de manejo de este tipo de sistemas. Figura de administrador de la Base de Datos.
- Implantación larga y difícil. Debido a los puntos anteriores. La adaptación del personal es mucho más complicada y requiere su tiempo.

Los sistemas actuales están muy optimizados, por lo que estas dificultades se están reduciendo.

En la siguiente figura se muestra un Sistema de Información basado en Bases de Datos. La información está relacionada y no es redundante.



4. Conceptos básicos de una Base de Datos.

4.1. Datos.

Datos son hechos conocidos que deben registrarse y que tienen un significado implícito.

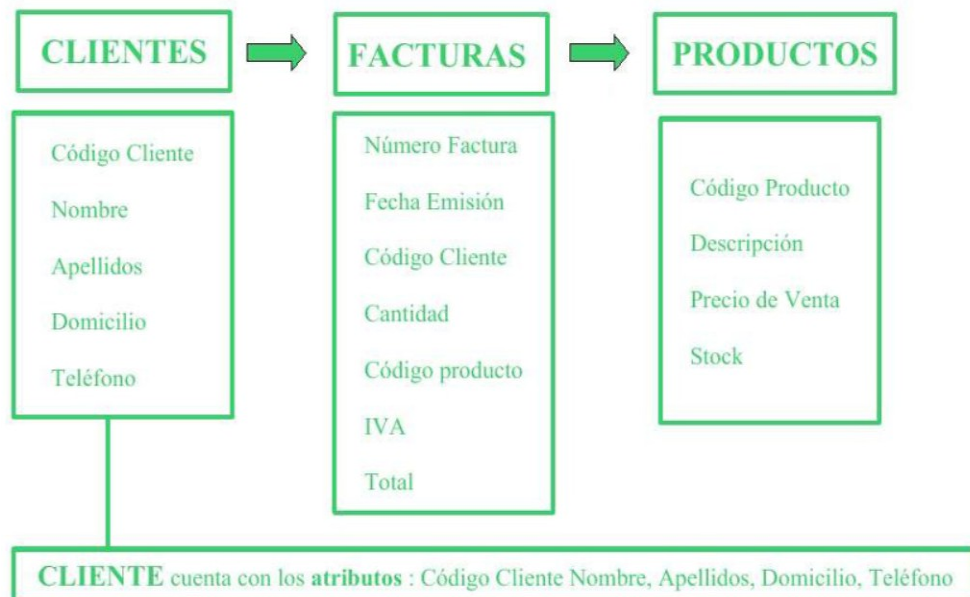
Ramez Elmasri y Shamkant B. Navathe



Ejemplo: Pueden constituir datos los nombres, números telefónicos y direcciones de personas que conocemos,...

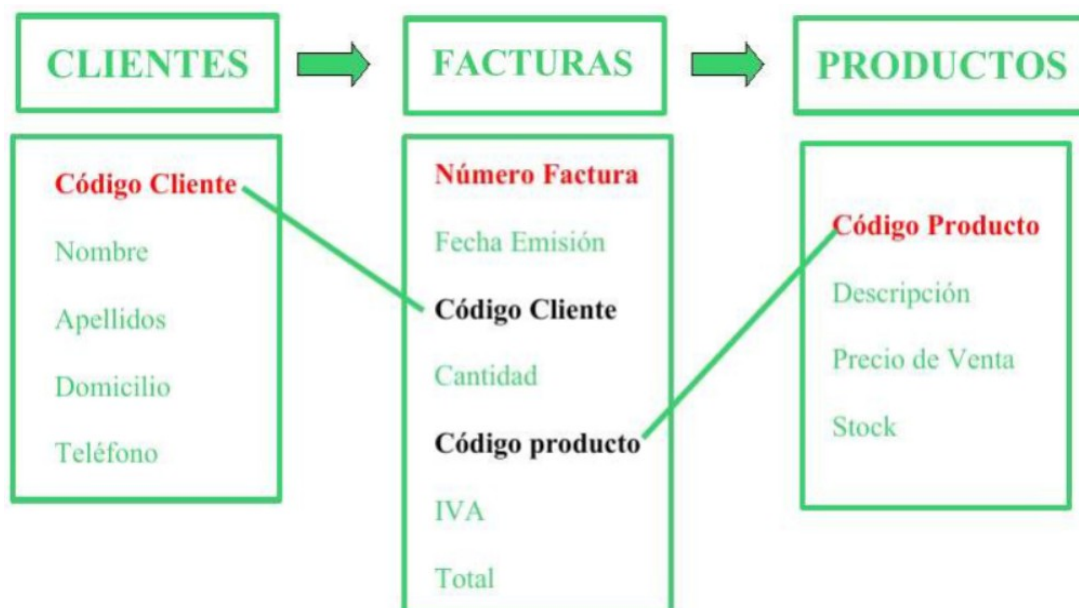
4.2. Entidades.

Una entidad es todo aquello de lo que interesa guardar datos, por ejemplo: clientes, facturas, productos,.. en el Sistema de Información de una tienda.



4.3. Claves primarias y claves foráneas. Relaciones.

Cada entidad tiene una clave primaria o campo clave o llave que identifica unívocamente al conjunto de datos. Cuando en una entidad figura la clave primaria de otra entidad, ésta se denomina clave foránea o clave ajena. Las entidades se relacionan entre sí a través de las claves foráneas.



CLAVES PRIMARIAS:

- Código Cliente es la clave primaria de CLIENTES. A cada cliente se le asocia un código y a cada código le corresponde un cliente.
- Número Factura es clave primaria de FACTURAS.
- Código Producto es clave primaria de PRODUCTOS.

CLAVES FORÁNEAS:

- En FACTURAS, son claves foráneas Código Cliente y Código Producto. CLIENTES se relaciona con FACTURAS a través del Código Cliente que figura en ambas tablas y con PRODUCTOS mediante el Código Producto.

4.4. Restricción de integridad referencial.

- Código Cliente en Facturas debe cumplir que exista en la tabla Clientes y que sea clave primaria.
- Código Producto en Facturas debe cumplir que exista en la tabla Productos y que sea clave primaria.

Retomando la Definición de Base de Datos, que dice que “... es un conjunto de datos relacionados entre sí y que tienen un significado implícito”.

Se observa en la imagen que los datos de las tablas se relacionan a través de las claves, haciendo coincidir el contenido de la clave foránea de una tabla con el contenido de la clave principal en la otra tabla.

Por otra parte, el significado implícito de cada dato se atribuye en el contexto. Así, por ejemplo, el significado del dato "Nombre" se refiere al nombre del CLIENTE, el de "Fecha emisión" a la de la FACTURAS y el de "Descripción" a la del PRODUCTO.

Por tanto, podemos decir que dadas dos tablas que están relacionadas, se cumple la restricción de integridad referencial cuando todos los valores que toman el atributo que es clave foránea en una de ellas está representado como clave primaria en la otra tabla.

4.5. Metadatos.

Metadatos son datos acerca de los datos presentes en la Base de Datos.

Por ejemplo:

- Qué tipo de datos se van a almacenar (si son texto, números, fechas,...).
- Qué nombre se le da a cada dato (nombre, apellidos, fecha, precio, edad,...).
- Cómo están agrupados los datos.
- Cómo se relacionan.
- Qué valores puede tomar o alguna restricción de valores.

Ejemplo de Metadatos:

El tipo de datos puede ser Numérico, alfabético, fecha, lógico (Si /NO).		
Dato	Tipo	Longitud
Num	Numérico	4
Nombre	Alfabético	20
.....

La longitud indica la cantidad máxima de caracteres que puede asumir el dato.

Ejemplo de Restricción de Dominio:
Num >0 y <9999

4.6. Otros conceptos de Bases de Datos.

Además de los conceptos básicos anteriores, a modo de aclarar algunos de los componentes que se pueden encontrar en una Base de Datos, y que se verán en las próximas unidades, se definen los siguientes conceptos:

- **Tabla (fichero):** Es un conjunto de filas y columnas bajo un mismo nombre que representa el conjunto de valores almacenados para una serie de instancias de una entidad. Por ejemplo, la información de todos los clientes (cada cliente es una instancia de la entidad Clientes) de una BD se almacenarán en una tabla llamada CLIENTES.
- **Atributo (campo):** Cada una de las columnas de una tabla. Por ejemplo, el atributo Nombre representa los nombres de cada uno de los clientes que contiene una tabla CLIENTES.
- **Tupla (registro):** Corresponde a cada una de las filas de la tabla. Por ejemplo en la tabla CLIENTES, una tupla está formado por todos los datos que se recojan de un determinado cliente: nombre, apellidos, dirección,...

CLIENTES				
Código cliente	Nombre	Apellidos	Domicilio	teléfono
12	Juan	García López	C/Pino 27	952 121212
13	Inés	Martínez Durán	C/Camino del pinar 7	636 077000
14	Luis	Moreno Sánchez	C/Pirita 31	956 090990

Diagram annotations: A red box highlights the 'Nombre' column, labeled 'Atributo (campo)'. A green box highlights the value 'Inés', labeled 'Dato'. A blue box highlights the entire row for client 13, labeled 'Tupla (registro)'.

- **Tipo de Dato:** El tipo de dato indica la naturaleza del campo: qué contenidos puede almacenar y qué operaciones se pueden hacer con ellos. Así, se pueden tener datos numéricos, que son aquellos con los que se pueden realizar cálculos aritméticos (sumas, restas, multiplicaciones...); los datos alfa-numéricos, que son los que contienen caracteres alfabéticos y números (concatenar dos cadenas o partir una cadena); datos tipo fecha, contendrán día, mes y año (sumar un entero y calcular una nueva fecha, restar dos fechas,...);...
- **Consulta:** Es una instrucción para hacer peticiones a una BD.
- **Índice:** Es una estructura que almacena los campos clave de una tabla, organizándolos para hacer más fácil encontrar y ordenar los registros.
- **Vista:** Se obtienen al guardar una consulta de una o varias tablas. De esta forma se obtiene una tabla virtual, es decir, no está almacenada en los dispositivos de almacenamiento del ordenador, aunque sí se almacena su definición.
- **Informe:** Es un listado ordenado de los campos y registros seleccionados en un formato fácil de leer. Por ejemplo, un informe de las facturas impagadas del mes de enero ordenadas por nombre de cliente.
- **Guiones o scripts:** Son un conjunto de instrucciones, que ejecutadas de forma ordenada, realizan operaciones avanzadas o mantenimiento de los datos almacenados en la BD.
- **Procedimientos:** Son un tipo especial de script que están almacenados en la BD y forman parte de su esquema.

5. Sistemas de Gestión de Bases de Datos.

5.1. Sistema Gestor de Base de Datos.

Un Sistema Gestor de Bases de Datos (SGBD) es una aplicación que permite a los usuarios definir, crear y mantener una Base de Datos, y proporciona acceso controlado a la misma.

En general, un SGBD proporciona los siguientes servicios:

- Permite la **definición** de la Base de Datos mediante el lenguaje de definición de datos (DDL - Data Description Language). Este lenguaje permite especificar la estructura y el tipo de los datos, así como las restricciones sobre los datos. Todo esto se almacenará en la Base de Datos.
- Permite la **inserción, actualización, eliminación y consulta** de datos mediante el lenguaje de manejo o manipulación de datos (DML - Data Manipulation Language).
- Proporciona un **acceso controlado** a la Base de Datos mediante:
 - Un sistema de seguridad, de modo que los usuarios no autorizados no puedan acceder a la Base de Datos, se consigue mediante el lenguaje de control de datos (DCL - Data Control Language).
 - Un sistema de integridad que mantiene la integridad y la consistencia de los datos.
 - Un sistema de control de concurrencia y gestión de transacciones. Cuando una base de datos tiene accesos compartidos por varios usuarios aparece el concepto de concurrencia. Estas situaciones pueden dar lugar a efectos no deseados como inconsistencias. Generalmente ocurren al intercalarse las acciones de los procesos que realizan los diferentes usuarios. Esos problemas se pueden evitar considerando cada proceso o acción que realizan los usuarios como transacciones.
 - Un sistema de control de recuperación que restablece la Base de Datos después de que se produzca un fallo hardware o software.
- Un **diccionario de datos** o catálogo contiene la descripción de los datos de la Base de Datos y es accesible por los usuarios.

La principal herramienta de un SGBD es la interfaz de programación con el usuario. Este interfaz consiste en un lenguaje que permite al usuario interactuar con el servidor. Este lenguaje comúnmente se denomina SQL (Structure Query Language) y está estandarizado por la ISO 1. Es decir, todas las BD que soporten SQL deben tener la misma sintaxis a la hora de aplicar el lenguaje.

5.2 Tipos de SGBD.

Los SGBD se pueden clasificar según el tipo de Base de Datos que gestionen (jerárquicas, relacionales, orientadas a objetos,...). Actualmente la mayoría de los SGBD integran múltiples filosofías, los clasificaremos según su capacidad y potencia del propio gestor, resultando los siguientes SGBD:

- SGBD ofimáticos: manipulan Bases de Datos pequeñas orientadas a almacenar datos domésticos o de pequeñas empresas. Ejemplos típicos son **ACCESS** de Microsoft y **Base** de LibreOffice.
- SGBD corporativos: tienen la capacidad de gestionar BD enormes, de medianas o grandes empresas con una carga de datos y transacciones que requieren un servidor de gran

capacidad. Ejemplos de BD corporativas son ORACLE, DB2 o SQL Server. Estas soluciones son de pago (independientemente que se ofrezcan versiones limitadas sin pago), para evitar esos gastos han aparecido otros gestores intermedios entre los ofimáticos y los corporativos.

- SGBD intermedios. En este segmento se encuentra MySQL. Es un SGBD gratuito, fácil de instalar, que permite trabajar con volúmenes considerables de datos y está integrado en paquetes XAMPP (Windows/Linux - Apache - MySQL - PHP/Perl).

Otros tipos de Bases de Datos son las conocidas como no relacionales o no SQL. Se caracterizan por :

- No requieren estructuras fijas como son las tablas.
- Trabajando en entornos de concurrencia, no garantizan completamente las características propias de las transacciones: ACID:
 - A: Atomicidad. Considera todas las acciones de un proceso como una única acción.
 - C: Consistencia. Después de una transacción la Base de Datos queda coherente.
 - I: Aislamiento. Cada proceso que realiza un usuario se considera independiente uno de otro.
 - D: Durabilidad. Una vez dada por buena una transacción quedan definitivamente hechos los cambios que ha realizado.

Ejemplos de Bases de Datos no relacionales:

- Mongo-DB. Muy popular orientada a ficheros. Utilizada por Google, Facebook, eBay,...
- Redis. Se basa en almacenamiento clave – valor.
- Cassandra. Se basa en almacenamiento clave – valor. Distribuida y masivamente escalable.

6. Ejercicios.

1 Para cada una de las siguientes cuestiones elige razonadamente las respuestas correctas.

I ¿Cuáles de los siguientes puntos representan inconvenientes de los Sistemas de Ficheros?

- a Redundancia e Inconsistencia.
- b Sistema de Gestión de Datos independiente de la máquina y del SO.
- c Control de concurrencia.
- d Dificil modificación de los datos.

II Los sistemas orientados a BD presentan las siguientes ventajas...

- a Integridad de los datos.
- b Redundancia.
- c Cada aplicación maneja sus propios datos.
- d Independencia entre los datos y las aplicaciones que los usan.

III Los datos son...

- a todo aquello de lo cual interesa guardar información.
- b hechos conocidos que pueden registrarse y que tienen un significado implícito.
- c información acerca de los metadatos.
- d las claves primarias y foráneas de cada entidad.

IV Un SGBD ...

- a esta formado por datos acerca de los datos presentes en la Base de Datos.

- b es una aplicación que permite a los usuarios definir, crear y mantener una Base de Datos, y proporciona acceso controlado a la misma.
 - c permite a los usuarios tener acceso a la BD completa impidiendo restricciones.
 - d permite la inserción, actualización, eliminación y consulta de datos mediante el lenguaje de manejo o manipulación de datos.
- 2 Explica brevemente los antecedentes de las BD actuales.
 - 3 Enumera y explica brevemente los inconvenientes que presentan los antiguos sistemas de Archivos o de Ficheros.
 - 4 Tipos de organizaciones de los ficheros y principales características.
 - 5 ¿Qué significa que los datos de una BD tienen un significado implícito? Por tres ejemplos, diferentes de los expuesto en el tema, que ilustren tu explicación.
 - 6 Soportes de almacenamiento. Explica la organización de un disco duro. Elementos que aparecen y descríbelos brevemente.
 - 7 Definir que es un SGBD.
 - 8 Indica las principales ventajas de las BD frente a los antiguos sistemas de ficheros.
 - 9 Nombra los distintos tipos de Bases de Datos que existen según el modelo que siguen.
 - 10 Describe el significado de las siguientes siglas: DDL, DML y DCL. Explica la utilidad de cada una.
 - 11 ¿Qué es un script o guión? Define los siguientes conceptos:
 - a Dato.
 - b Tipo de Dato.
 - c Campo.
 - d Registro.
 - e Tabla.
 - f Relación.
 - g Consulta.
 - h Procedimiento.
 - 12 ¿Qué es el diccionario de datos?
 - 13 ¿Qué quiere decir que una Base de Datos permita la concurrencia?

7. Planteamiento práctico.

Se pretende comparar un Sistema de Ficheros con un Sistema de Bases de Datos.

Partiendo de un problema real, se muestra la solución utilizando un Sistema de Ficheros. Se analizará dicha solución y se detectarán en ella ejemplos de cada uno de los inconvenientes que presenta. Es decir, buscaremos ejemplos en la solución propuesta que ilustren los siguientes inconvenientes:

- Coste de almacenamiento elevado.
- Datos redundantes.
- Probabilidad alta de inconsistencia en los datos.
- Difícil modificación en los datos.
- Tiempos de procesamiento elevados.

Problemática de una inmobiliaria.

Una empresa se encarga de dar publicidad a los inmuebles que ofrece en alquiler, tanto en prensa local como nacional, entrevista a los posibles inquilinos, organiza las visitas a los inmuebles y negocia los contratos de alquiler. Una vez firmado el alquiler, la empresa asume la responsabilidad del inmueble, realizando inspecciones periódicas para comprobar su correcto mantenimiento. A continuación se describen los datos que se manejan en las oficinas de la empresa para llevar a cabo el trabajo diario.

OFICINAS

La empresa tiene varias oficinas en todo el país. Cada oficina tiene un código de identificación que es único, tiene una dirección (calle, número y ciudad), un número de teléfono y un número de fax. Cada oficina tiene su propia plantilla.

PLANTILLA

Cada oficina tiene un director que se encarga de supervisar todas sus gestiones. La empresa sigue muy de cerca el trabajo de los directores y tiene registrada la fecha en que cada director empezó en el cargo en su oficina. Cada director tiene un pago anual por gastos de vehículo y una bonificación mensual que depende de los contratos de alquiler que haya realizado su oficina. En cada oficina hay varios supervisores. Cada uno es responsable del trabajo diario de un grupo de entre cinco y diez empleados que realizan las gestiones de los alquileres. El trabajo administrativo de cada grupo lo lleva un administrativo. Cada miembro de la plantilla tiene un código único que lo identifica en la empresa. De cada uno de ellos se quiere conocer el nombre, la dirección, el número de teléfono, la fecha de nacimiento, el número del DNI, su puesto en la empresa, el salario anual y la fecha en que entró en la empresa. De los administrativos se desea conocer también la velocidad con que escriben a máquina (en pulsaciones por minuto). Además, de cada empleado se debe guardar información sobre uno de sus parientes más próximos: nombre, relación con el empleado, dirección y número de teléfono.

INMUEBLES

Cada oficina de la empresa tiene una serie de inmuebles para alquilar. Estos inmuebles se identifican por un código que es único dentro de la empresa. Los datos que se guardan de cada inmueble son los siguientes: dirección completa (calle, número y ciudad), tipo de inmueble, número de habitaciones y precio del alquiler en euros (este precio es mensual). El precio del alquiler se revisa de forma anual. Cada inmueble se asigna a un empleado que es el responsable de su gestión. Cada miembro de la plantilla puede tener asignados hasta veinte inmuebles para alquilar.

PROPIETARIOS

Los propietarios de los inmuebles pueden ser particulares o empresas. A cada propietario se le asigna un código que es único en la empresa. De los particulares se guarda el nombre, la dirección y el número de teléfono. De las empresas se guarda el nombre comercial, tipo de empresa, la dirección, el número de teléfono y el nombre de la persona de contacto.

INQUILINOS (CLIENTES)

Cuando un cliente contacta con la empresa por primera vez, se toman sus datos: nombre, dirección, número de teléfono, tipo de inmueble que prefiere e importe máximo que está dispuesto a pagar al mes por el alquiler. Ya que es un posible inquilino, se le asigna un código que es único en toda la empresa. De la entrevista inicial que se realiza con cada cliente se guarda la fecha, el empleado que la realizó y unos comentarios generales sobre el posible inquilino.

VISITAS A LOS INMUEBLES

En la mayoría de los casos, los posibles inquilinos desean ver varios inmuebles antes de alquilar uno. De cada visita que se realiza se guarda la fecha y los comentarios realizados por el cliente respecto al inmueble.

ANUNCIOS

Cuando algún inmueble es difícil de alquilar, la empresa lo anuncia en la prensa local y nacional. De cada anuncio se guarda la fecha de publicación y el coste económico del anuncio. De los periódicos se guarda el nombre, la dirección, el número de teléfono, el número de fax y el nombre de la persona de contacto.

CONTRATOS DE ALQUILER

La empresa se encarga de redactar los términos de cada contrato de alquiler. Cada contrato tiene un número, un importe mensual, un método de pago, el importe del depósito, si se ha realizado el depósito, las fechas de inicio y finalización del contrato, la duración del contrato en meses y el miembro de la plantilla que lo formalizó. La duración mínima de un contrato es de tres meses y la duración máxima es de un año. Cada cliente puede tener alquilados uno o varios inmuebles al mismo tiempo.

INSPECCIONES

Como parte del servicio que presta la empresa, ésta se encarga de realizar inspecciones periódicas a los inmuebles para asegurarse de que se mantienen en buen estado. Cada inmueble se inspecciona al menos una vez cada seis meses. Se inspeccionan tanto los inmuebles alquilados, como los que están disponibles para alquilar. De cada inspección se anota la fecha y los comentarios sobre su estado que quiera incluir el empleado que la ha llevado a cabo.

ACTIVIDADES DE CADA OFICINA

En cada oficina se llevan a cabo las siguientes actividades para garantizar que cada empleado tenga acceso a la información necesaria para desempeñar su tarea de modo efectivo y eficiente. Cada actividad está relacionada con una función específica de la empresa. Cada una de estas funciones corresponde a uno o varios puestos de los que ocupan los empleados, por lo que éstos se indican entre paréntesis.

Solución propuesta según Sistema de Ficheros.

DEPARTAMENTO DE VENTAS

En esta inmobiliaria, el departamento de ventas se encarga de alquilar inmuebles. Por ejemplo, cuando *un propietario pasa por el departamento de ventas para ofrecer en alquiler su piso*, se rellena un formulario en donde se recogen los datos del piso, como la dirección y el número de habitaciones, y los datos del propietario. El departamento de ventas también se encarga de atender a los clientes que desean alquilar un inmueble. Cuando *un cliente (posible inquilino) pasa por este departamento se rellena un formulario con sus datos y sus preferencias*: si quiere un piso o una casa, el importe mensual que está dispuesto a pagar por el alquiler, etc. Para gestionar toda esta información, el departamento de ventas posee un Sistema de Información. El sistema del departamento de ventas tiene **tres ficheros**: fichero de **inmuebles**, fichero de **propietarios** y fichero de **inquilinos**.

INMUEBLE

Inum	Calle	Área	Población	Tipo	Hab.	Alquiler	Pnum
IA14	En medio, 128	Centro	Castellón	Casa	6	600	P46
IL94	Riu Ebre, 24	Ronda Sur	Castellón	Piso	4	350	P87
IG4	Sorell, 5	Grac	Castellón	Piso	3	300	P40
IG36	Alicante, 1		Segorbe	Piso	3	325	P93
IG21	San Francisco, 10		Vinaroz	Casa	5	550	P87
IG16	Capuchinos, 19	Rafalafena	Castellón	Piso	4	400	P93

PROPIETARIO

Pnum	Nombre	Apellido	Dirección	Teléfono
P46	Amparo	Felipe	Asensio 24, Castellón	964 230 680
P87	Manuel	Alejandro	Av. Libertad 15, Vinaroz	964 450 760
P40	Alberto	Estrada	Av. del Puerto 52, Castellón	964 200 740
P93	Yolanda	Robles	Purísima 4, Segorbe	964 710 430

INQUILINO

Qnum	Nombre	Apellido	Dirección	Teléfono	Tipo	Alquiler
Q76	Juan	Felip	Barceló 47, Castellón	964 282 540	Piso	375
Q56	Ana	Grangel	San Rafael 45, Almazora	964 551 110	Piso	300
Q74	Elena	Abaso	Navarra 76, Castellón	964 205 560	Casa	700
Q62	Alicia	Mori	Alloza 45, Castellón	964 229 580	Piso	550

DEPARTAMENTO DE CONTRATOS

El departamento de contratos se ocupa de gestionar los contratos de alquiler de los inmuebles. Cuando un cliente desea formalizar un contrato, un empleado de la empresa rellena un formulario con los datos del inquilino y los datos del inmueble. Este formulario se pasa al departamento de contratos, que asigna un número al contrato y completa la información sobre el pago y el período del contrato. Para gestionar esta información, el departamento de contratos posee un Sistema de Información con **tres ficheros**: el fichero de los **contratos**, el fichero de los **inmuebles alquilados** y el fichero de los **inquilinos** que tienen en vigor un contrato de alquiler.

Importante: Al tratarse de una solución basada en Sistemas de Ficheros, cada departamento accede a sus propios ficheros mediante una serie de programas o aplicación escritos especialmente para ellos. Estos programas son totalmente independientes entre un departamento y otro, y se utilizan para introducir datos, mantener los ficheros y generar los informes que cada departamento necesita. Es importante destacar que la estructura física de los ficheros de datos y de sus registros está definida dentro de los programas de aplicación.

CONTRATO

Cnum	Inum	Qnum	Importe	Pago	Depósito	Pagado?	Inicio	Fin	Meses
10024	IA14	Q62	600	Visa	1200	S	1/6/99	31/5/00	12
10075	IL94	Q76	350	Efectivo	700	N	1/1/00	30/6/00	6
10012	IG21	Q74	550	Cheque	1100	S	1/7/99	30/6/00	12

INMUEBLE

Inum	Calle	Área	Población	Alquiler
IA14	Enmedio, 128	Centro	Castellón	600
IL94	Riu Ebre, 24	Ronda Sur	Castellón	350
IG21	San Francisco, 10		Vinaroz	550

INQUILINO

Qnum	Nombre	Apellido	Dirección	Población	Teléfono
Q76	Juan	Felip	Barceló, 47	Castellón	964 282 540
Q74	Elena	Abaso	Navarra, 76	Castellón	964 205 560
Q62	Alicia	Mori	Alloza, 45	Castellón	964 229 580

La situación es muy similar en el resto de departamentos.

DEPARTAMENTO DE NÓMINAS

En el departamento de nóminas tiene **un fichero** con los datos de los salarios de los empleados. Los registros de este fichero tienen los siguientes campos: número de empleado, nombre, apellido, dirección, fecha de nacimiento, salario, DNI y número de la oficina en la que trabaja.

DEPARTAMENTO DE PERSONAL

El departamento de personal tiene **un fichero** con los datos de los empleados. Sus registros tienen los siguientes campos: número de empleado, nombre, apellidos, dirección, teléfono, puesto, fecha de nacimiento, salario, DNI y número de la oficina en la que trabaja.

Cuestiones.

1. Completa una tabla buscando ejemplos en el enunciado de cada uno de los inconvenientes estudiados en los sistemas de ficheros. Deberás completar la respuesta con una breve explicación que ilustre el porqué dicho ejemplo presenta el inconveniente en cuestión.

INCONVENIENTE	EJEMPLO	EXPLICACIÓN
Coste de almacenamiento elevado		
Datos redundantes		
Probabilidad alta de inconsistencia de datos		
Difícil modificación de los datos		
Tiempo de procesamiento elevado		

Nota: La tabla se mostrará en formato apaisado.

2. Supongamos ahora que quisiéramos resolver el problema real usando el enfoque actual de BD. Detecta en el enunciado ejemplos de los siguientes elementos fundamentales de una BD:

- Datos
- Entidades

- Claves principales
- Claves foráneas
- Integridad referencial
- Metadatos