

# **UT1. SISTEMAS DE ALMACENAMIENTO DE LA INFORMACIÓN**

**BASE DE DATOS**

# ÍNDICE

## UT1. SISTEMAS ALMACENAMIENTO DE LA INFORMACIÓN.

1.1 INTRODUCCIÓN

1.2 SISTEMAS DE INFORMACIÓN

1.3 TIPOS DE SISTEMAS DE INFORMACIÓN

1.4 FICHEROS

1.5 SISTEMAS DE FICHEROS

1.6 SISTEMAS GESTORES DE BASE DE DATOS

1.7 MODELO DE BASE DE DATOS

1.8 FUNCIONAMIENTO DE UN SISTEMA GESTOR DE  
BASES DE DATOS

1.8.1 FUNCIONES. LENGUAJES DE LOS SGBD

1.8.2. COMPONENTES

1.8.3. ARQUITECTURA

1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE  
BASES DE DATOS

1.10 LEGISLACIÓN

1.11 BIG DATA

1.12 INTELIGENCIA DE NEGOCIOS

# ¿QUÉ ES UN DATO?

**Según la RAE:**

1. m. Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho.

Sin.:

información, noticia, nota, apunte, referencia, documento.

2. m. Documento, testimonio, fundamento.

3. m. Inform. **Información dispuesta de manera adecuada para su tratamiento por una computadora.**

# ¿QUÉ ES UN DATO?



## 1.1 INTRODUCCIÓN

En la sociedad actual existe una demanda cada vez mayor de datos, el procesamiento de la información es un sector en auge, hasta el punto de hablar de **SOCIEDAD DE LA INFORMACIÓN**. Esta demanda ha aumentado en los últimos años debido al acceso multitudinario a Internet y a la aparición de pequeños dispositivos (móviles y tabletas) cada vez más potentes. Constantemente hacemos uso de estos datos:





## 1.1 INTRODUCCIÓN

Inicialmente se necesitaba almacenar y gestionar pocos datos. Para poder almacenar estos datos, utilizaba archivadores, cajones, carpetas y fichas en las que se almacenaban los datos, el tiempo requerido para manipular estos datos era enorme. Sin embargo, el proceso de aprendizaje era relativamente sencillo ya que se usaban elementos que el usuario reconocía perfectamente.



## 1.1 INTRODUCCIÓN

Sin embargo, en la actualidad las empresas gestionan enormes sistemas de información, necesitando a veces varios ordenadores.

Estos datos que usamos a diario tienen que estar almacenados en algún sitio y debemos tener herramientas que permitan su tratamiento. Cuando la cantidad de datos es pequeña su tratamiento es sencillo, sin embargo, a medida que aumenta la cantidad de información, resulta más complejo su tratamiento y aquí es donde intervienen los sistemas informáticos y concretamente las aplicaciones software que van a ayudar a realizar esta tarea.

Antes de la aparición del ordenador, el tiempo requerido para manipular estos datos era enorme.

Por esa razón, la informática adaptó sus herramientas para que los elementos que el usuario maneja en el ordenador se parezcan a los que utilizaba manualmente. Así en informática se sigue hablando de ficheros, formularios, carpetas, directorios...

## 1.2 SISTEMAS DE INFORMACIÓN

Podemos definir un **sistema de información (SI)** como un conjunto de elementos relacionados entre sí, cuyo objetivo es el tratamiento y administración de los datos. El sistema de información en la empresa es el que se encarga de gestionar todos los datos.

Un sistema de información está formado por los siguientes elementos:

- **Recursos físicos.** Maquinaria y elementos que se utilizan para manejar la información. Carpetas, documentos, equipamiento, archivadores, ...
- **Recursos humanos.** Personal que maneja la información
- **Protocolo.** Normas que se siguen para manejar la información (formato de la información, modelo para los documentos,...)
- **Datos.** Información que se maneja.

Cuando todas o partes de las operaciones de un sistema de información se realizan mediante el uso de ordenadores al sistema se le llama sistema informático.



## 1.2 SISTEMAS DE INFORMACIÓN

Los componentes son:

- **Hardware.** Equipamiento físico que se utiliza para gestionar los datos. cada uno de los dispositivos electrónicos que permiten el funcionamiento del sistema de información.
- **Software.** Aplicaciones informáticas que se encargan de la gestión de los datos.
- **Recursos humanos.** Personal que maneja el sistema de información.
- **Protocolo.** Normas que se siguen para manejar la información (formato de la información que se almacena)
- **Datos.** Se trata de la información relevante que almacena y gestiona el sistema de información. Ejemplos de datos son:

**Sánchez, 12764569F, Calle Mayo 5, Azul...**

## 1.3 TIPOS DE SISTEMAS DE INFORMACIÓN

Según la tecnología empleada para realizar los procedimientos:

- **Manuales:** En estos sistemas la gestión de la información se realiza mediante herramientas elementales como lápiz, papel y archivadores.
- **Automáticos o informatizados:** Realizan la gestión de la información mediante ordenadores. Las ventajas de estos sistemas son:
  - Mayor velocidad de procesamiento.
  - Recuperación más rápida de la información.
  - Mayor exactitud en las operaciones.
  - Reducción de costes (mano de obra).
  - Mejora de la comunicación interna y externa.
  - Atraen a más clientes.

## 1.3 TIPOS DE SISTEMAS DE INFORMACIÓN

Desde un punto de vista empresarial

- **Sistemas transaccionales.** Gestiona la información referente a las operaciones básicas de la empresa. Es decir, se dedican al tratamiento de las operaciones diarias. Las características son:
  - Automatizan las operaciones básicas de la empresa (pagos, cobros, entradas, almacenes...). Procura que los datos sean consistentes en todo momento. Por ejemplo, si se realiza un pago con tarjeta, se debe hacer el cargo en la cuenta del cliente y el abono en la cuenta de la empresa por la misma cantidad, si falla una de ellas, se anula también la otra.
  - Son los sistemas que se suelen instalar en primer lugar.
  - Ahorran mano de obra.
  - Se encargan de introducir los datos y suelen ser poco complejos.
  - Son utilizados por los trabajadores de nivel bajo.

## 1.3 TIPOS DE SISTEMAS DE INFORMACIÓN

- **Sistemas de soporte a decisiones (DSS).** No automatizan nada, simplemente ayudan a decidir cual es la mejor decisión. Las características son:
  - Analizan el problema, generan alternativas, simulan cada una de ellas y deciden cuál es la mejor.
  - Son utilizados por los ejecutivos.
  - Se encargan de proporcionar información y son complejos.
  - No ahorran mano de obra, simplemente ayudan a los ejecutivos a tomar decisiones.

Por ejemplo, un director de ventas puede obtener los totales de ventas mensuales por provincia, por vendedores, por productos...

## 1.4 FICHEROS

En los inicios del desarrollo de las aplicaciones, el almacenamiento de la información que gestiona cualquier aplicación informática era mediante el uso de archivos o ficheros (files en inglés). Esto provocaba que las aplicaciones estuvieran "acopladas" a los ficheros (dependencia físico-lógica), y dificultaba compartir datos entre diferentes aplicaciones, así como el acceso concurrente a los mismos.

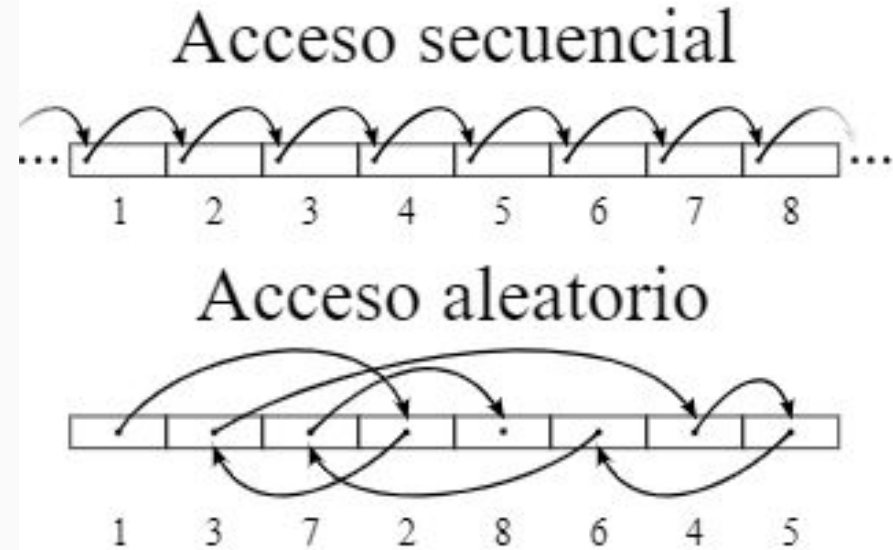
Respecto a los ficheros, conviene conocer que existen ficheros con diferente estructura y formato, como pueden ser:

- **Ficheros de texto:** suelen llamarse también ficheros **planos** o ficheros ascii. Están formados por caracteres que son directamente legibles por el ser humano, no llevan ningún tipo de formato. Dentro de estos tenemos: .c, .java, .rtf
- **Ficheros binarios:** son todos los que no son de texto. A continuación, se muestran algunos tipos de formatos de ficheros binarios: De imagen: .jpg, .gif, .tiff, .bmp, de vídeo: .mpg, .mov, .avi, ..., comprimidos o empaquetados, ejecutables o compilados, .....

## 1.4 FICHEROS

Dependiendo del acceso, tenemos diferentes tipos de ficheros:

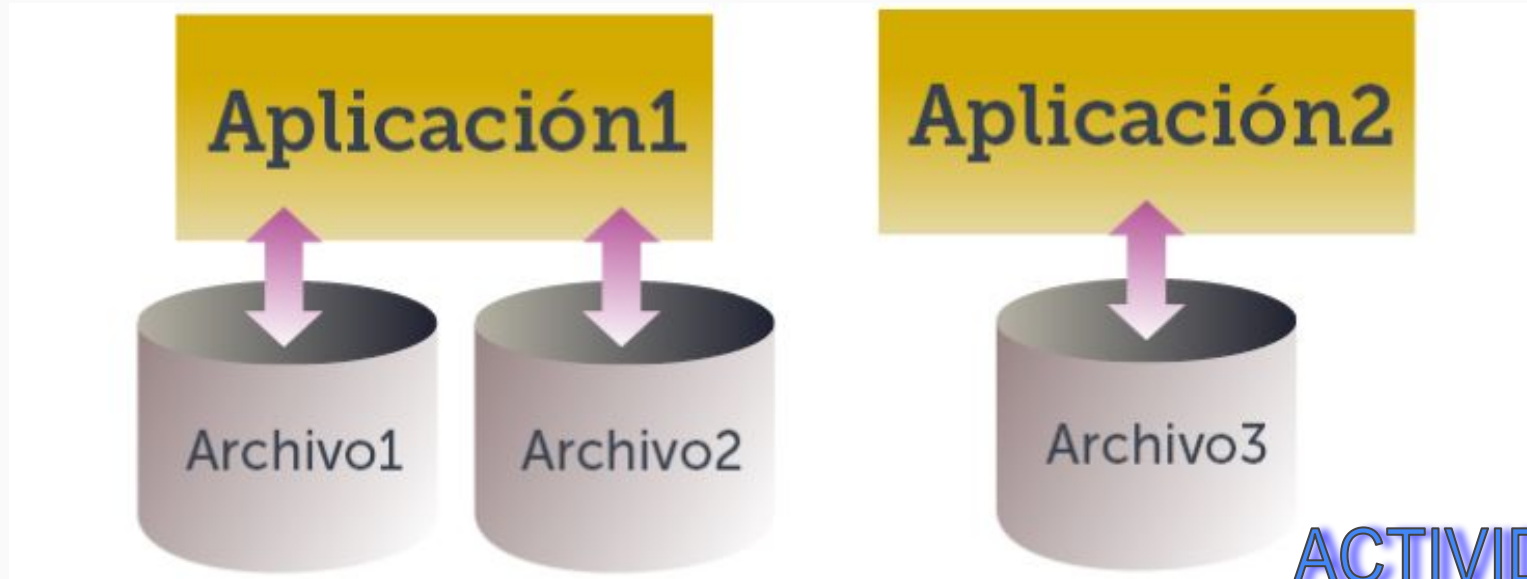
- **Secuenciales:** los registros se almacenan uno detrás de otro. Antiguamente se utilizaba cuando los ficheros se persistían mediante cintas magnéticas, y provocaba que para acceder a un registro, era necesario recorrer todos los anteriores.
- **De acceso directo o aleatorio:** permite el acceso directo al registro que nos interesa, pudiendo saltar de uno a otro sin necesidad de recorrer los anteriores.





## 1.5 SISTEMAS DE FICHEROS

En estos sistemas de ficheros se crean varias aplicaciones para gestionar diferentes aspectos del sistema. Cada aplicación realiza una determinada operación. Los datos de dichas aplicaciones se almacenan en ficheros. Cada programa almacena y utiliza sus propios datos y por tanto crea y usa sus propios ficheros.



## 1.5 SISTEMAS DE FICHEROS

Los sistemas de ficheros surgieron al tratar de informatizar el manejo de los archivadores manuales con el objetivo de proporcionar un acceso más eficiente a los datos. En la empresa, en lugar de establecer un sistema centralizado en donde almacenar todos los datos de la organización, se escogió un modelo descentralizado en el que cada sección o departamento almacena y gestiona sus propios datos.

La ventaja de este sistema (la única ventaja), es que los procesos son independientes por lo que la modificación de uno no afecta al resto. Pero tiene grandes inconvenientes:

- **Separación y aislamiento de los datos.** Los datos están distribuidos en distintos ficheros que en muchos casos no mantienen ningún tipo de relación. (No hay integridad referencial)
- **Datos redundantes.** Los mismos datos aparecen en diferentes ficheros, con el consiguiente desperdicio de espacio y coste de almacenamiento. (se requiere más espacio en disco).

## 1.5 SISTEMAS DE FICHEROS

- **Probabilidad alta de inconsistencia en los datos.** El mismo dato puede tener valores distintos en diferentes lugares. Pues un programa puede cambiar el valor en un lugar y otro no hacerlo.
- **Formatos de ficheros incompatibles.** Ya que la estructura de los ficheros se define en los programas de aplicación, es completamente dependiente del lenguaje de programación. La incompatibilidad entre ficheros generados por distintos lenguajes hace que los ficheros sean difíciles de procesar de modo conjunto.
- **Falta de integridad.** Es difícil mantener ciertas condiciones. Ejemplo: un inmueble puede estar asociado a un propietario inexistente.
- **Dificultad para administrar seguridad.** No se puede controlar que ciertos usuarios no accedan a ciertos datos.

## 1.6 SISTEMAS GESTORES DE BASE DE DATOS

Para solucionar los inconvenientes que presentaban los sistemas de ficheros, inconvenientes que se hicieron más patentes con la aparición de las redes de ordenadores, y la explosión de internet, surgen las bases de datos como un conjunto, colección o depósito de datos interrelacionados, con características de integridad y redundancia mínima, accesibles por varios usuarios y aplicaciones.. A su vez alrededor de las bases de datos nacen los **sistemas gestores de base de datos (SGBD)** (DBMS, DataBase Management System) conjunto de utilidades que permiten la implantación, acceso y mantenimiento de una base de datos.

Las aplicaciones que creen los programadores, no acceden directamente a los datos, de modo que la **base de datos** es común para todas las aplicaciones. Cuando una aplicación modifica un dato, la modificación será visible inmediatamente para el resto de aplicaciones; ya que todas utilizarán la misma base de datos.

## 1.6 SISTEMAS GESTORES DE BASE DE DATOS



## 1.6 SISTEMAS GESTORES DE BASE DE DATOS

Tienen numerosas ventajas:

- **Independencia de los datos y los programas:** los SGBD separan las descripciones de los datos de las aplicaciones. Esto permite modificar los datos sin modificar el código de las aplicaciones.
- **La redundancia es mínima:** en estos sistemas los mismos datos no están duplicados en diferentes ficheros, con lo cual el espacio de almacenamiento es mínimo y la duplicidad también.
- **Mejora la accesibilidad a los datos:** se proporcionan lenguajes de consultas o generadores de informes que permiten al usuario hacer cualquier tipo de consulta sobre los datos, sin que sea necesario que un programador escriba una aplicación que realice tal tarea. (Ejemplo: Microsoft Access, MySql).
- **No hay inconsistencias en los datos:** No hay inconsistencia, pues gracias a las relaciones si un dato se modifica en un sitio automáticamente se modifica en todos los sitios donde aparece.



## 1.6 SISTEMAS GESTORES DE BASE DE DATOS

- **Mejora en la integridad de los datos:** la integridad se expresa mediante restricciones o reglas que no se pueden violar. Estas restricciones se pueden aplicar tanto a los datos, como a sus relaciones, y es el SGBD quien se debe encargar de mantenerlas. Ejemplo: el alquiler de un piso no puede ser un valor negativo.
- **No hay aislamiento de información:** los datos están relacionados unos con otros, de manera que pertenecen a empresa y los mismos datos son compartidos por los diferentes usuarios de los departamentos. Además, las nuevas aplicaciones que se vayan creando pueden utilizar los datos de la base de datos existente.
- **Mayor seguridad en los datos.** La seguridad de la base de datos es la protección de la base de datos frente a usuarios no autorizados. Los SGBD permiten mantener la seguridad mediante el establecimiento de usuarios y claves para identificar al personal autorizado a utilizar la base de datos.
- **Aumento de la concurrencia.** En algunos sistemas de ficheros, si hay varios usuarios que pueden acceder simultáneamente a un mismo fichero, es posible que

## 1.6 SISTEMAS GESTORES DE BASE DE DATOS

el acceso interfiera entre ellos de modo que se pierda información o, incluso, que se pierda la integridad. La mayoría de los SGBD gestionan el acceso concurrente a la base de datos y garantizan que no ocurran problemas de este tipo.

### **Desventajas:**

- **Un fallo en la BD afecta a todo el entorno.**
- **Instalación costosa.** El control y administración de bases de datos requiere de un software y hardware poderoso.
- **Requiere personal cualificado.** Debido a la dificultad de manejo de este tipo de sistemas.
- **Implantación larga y difícil.** En relación a los puntos anteriores. La adaptación del personal y del equipamiento es mucho más complicada y lleva bastante tiempo.

Como se puede observar las ventajas son muy superiores a las desventajas.

Una base de datos constará de los siguientes elementos:

## 1.6 SISTEMAS GESTORES DE BASE DE DATOS

- **Entidades:** objeto real o abstracto con características diferenciadoras de otros, del que se almacena información en la base de datos. En una base de datos de una clínica veterinaria, posibles entidades podrían ser: animal, personal, consulta, etc.
- **Atributos:** son los datos que se almacenan de la entidad. Cualquier propiedad o característica de una entidad puede ser atributo. Continuando con nuestro ejemplo, podrían ser atributos: raza, color, nombre, número de identificación, etc.
- **Registros:** donde se almacena la información de cada entidad. Es un conjunto de atributos que contienen los datos que pertenecen a una misma repetición de entidad. En nuestro ejemplo, un registro podría ser:

**2123056, Sultán, Podenco, Gris, 23/03/2009.**

- **Campos:** donde se almacenan los atributos de cada registro. Teniendo en cuenta el ejemplo anterior, un campo podría ser el valor Podenco.

La base de datos también almacena una descripción de los datos. Esta descripción es lo que se denomina **metadatos**, se almacena en el **diccionario de datos o catálogo** y es lo que permite que exista independencia de datos lógica-física.

**ACTIVIDAD 2**

## 1.7 MODELO DE BASE DE DATOS

Un modelo simboliza una parte del mundo real, de forma que se simplifique su manipulación. Por ejemplo, un modelo de una ciudad sería un mapa a escala 1:100.

Un **modelo de datos** es un conjunto de herramientas conceptuales para describir los datos, sus relaciones, su semántica y sus restricciones.

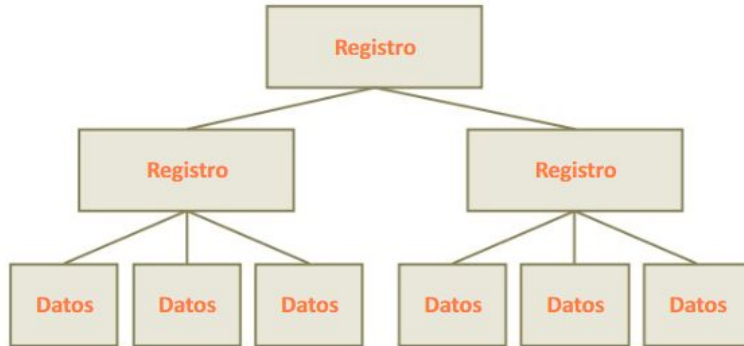
A lo largo de la historia los **SGBD** han ido evolucionando ligados a los **modelos de datos** que pueden implementar, desde los años sesenta con bases de datos estructurados de forma **jerárquica y en red**, hasta los actuales que permiten desarrollar aplicaciones gráficas multiusuario y multiplataforma, en potentes sistemas en red, podemos distinguir una evolución en la que se distinguen tres etapas o generaciones:

- **Primera generación de los SGBD:** Los sistemas jerárquico y de red constituyen la primera generación de los SGBD. El **modelo jerárquico** (finales de los 60) estructura los datos de manera escalonada, existiendo una relación del tipo padre-hijo entre sus registros. La representación de la información es similar a una estructura en árbol. El **modelo en red** (principios de los 70), elimina la restricción del modelo jerárquico, en el

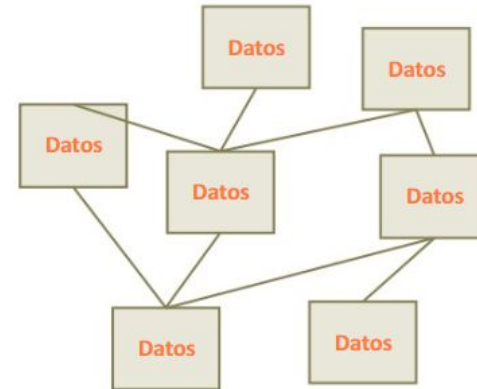
## 1.7 MODELO DE BASE DE DATOS

sentido de que un registro puede tener más de un padre, por lo que resulta más flexible, dando lugar a una estructura de grafo. Pero estos sistemas presentan algunos inconvenientes:

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



a) Modelo jerárquico



b) Modelo en red

## 1.7 MODELO DE BASE DE DATOS

- **Segunda generación de los SGBD:** En 1970 Codd, de los laboratorios de investigación de IBM, escribió un artículo presentando el modelo relacional. En este artículo, presentaba también los inconvenientes de los sistemas previos, el jerárquico y el de red. Entonces, se comenzaron a desarrollar muchos sistemas relacionales, apareciendo los primeros a finales de los setenta y principios de los ochenta. Esto condujo a dos grandes desarrollos:
  - El desarrollo de un lenguaje de consultas estructurado denominado **SQL**, que se ha convertido en el lenguaje estándar de los sistemas relacionales.
  - La producción de varios SGBD relacionales durante los años ochenta, como DB2 y SLQ/DS de IBM, y ORACLE de ORACLE Corporation.

ALUMNO				
P- dni	Nombre	prApellido	sgApellido	Bilingue
44102321	Rosa	Blanco	Montero	S
29600501	Dolores	García	Ramos	N
48300100	Pedro	Marín	Espinosa	N
84501495	Ángel	Luque	Nieto	S
48103100	Josefa	Muñoz	Marín	S
90100200	Pilar	Cea	Ruíz	N
28900194	David	Chaparro	Gómez	N

c) Modelo relacional



## 1.7 MODELO DE BASE DE DATOS

- **Tercera generación de los SGBD:** En los años 80 y 90, los **modelos orientados a objetos** intentaron integrar conceptos de la programación orientada a objetos en las bases de datos. En paralelo, el advenimiento de la web y el crecimiento masivo de los datos trajo consigo una nueva categoría:

**Bases de datos NoSQL**(década del 2000 hasta hoy en día) Estas se caracterizan por una mayor flexibilidad en el esquema, soporte para datos no estructurados y escalabilidad horizontal, lo que las hace ideales para big data y aplicaciones distribuidas. Como complemento a los modelos relacionales, ofrecen alternativas donde prima el rendimiento y escalabilidad de los datos. Destacan los modelos:

- **Clave-Valor:** se almacenan datos a los cuales sólo se puede acceder mediante una clave.
- **Documental:** utiliza estructuras de documentos JSON que contiene la información de varias entidades en un mismo documento
- **Grafos:** Uso de nodos y vértices para navegar por los datos.

## 1.7 MODELO DE BASE DE DATOS

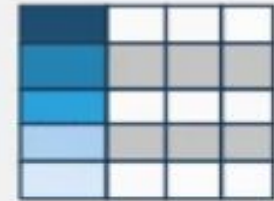
- **Basado en columnas:** gira las tablas para almacenar los datos por columnas en vez de hacerlo por filas.

### NoSQL

#### Key-Value



#### Column-Family



#### Graph



#### Document



## 1.7 MODELO DE BASE DE DATOS

**Tecnologías emergentes y nuevas arquitecturas diseñadas para las necesidades actuales y futuras del manejo de datos.**

- **Bases de datos distribuidas globalmente y multi-modelo:** La tendencia hacia bases de datos distribuidas geográficamente que soportan múltiples modelos (relacional, documental, grafos, etc.) de forma nativa. Un ejemplo es Google Spanner o Amazon Aurora. Estas permiten transacciones distribuidas a nivel global con fuerte consistencia, lo que antes era muy difícil de lograr.
- **Bases de datos impulsadas por la inteligencia artificial y el aprendizaje automático:** El uso de AI/ML para la optimización automática de consultas, administración de índices, y otras tareas. Un ejemplo son los sistemas auto-gestionados como Microsoft SQL Server con Adaptive Query Processing o Oracle Autonomous Database, que integran funciones inteligentes para autoconfigurarse, optimizarse y repararse.

## 1.7 MODELO DE BASE DE DATOS

- **Bases de datos cuánticas:** A medida que la computación cuántica avanza, algunos investigadores ya están explorando el diseño de bases de datos cuánticas, que aprovecharían la capacidad de procesamiento paralelo masivo de la computación cuántica. Estas BD aún están en etapas teóricas y experimentales.
- **Bases de datos inmersivas y basadas en blockchain:** Con el auge de blockchain y la descentralización, bases de datos distribuidas seguras y sin intermediarios están ganando tracción, como las bases de datos blockchain (por ejemplo, BigchainDB). Además, la creciente importancia del metaverso y la realidad aumentada/virtual podría requerir nuevos tipos de bases de datos optimizadas para entornos inmersivos y tridimensionales.
- **Bases de datos de memoria persistente:** Con el avance en tecnologías de almacenamiento como la memoria persistente (persistent memory), se están desarrollando bases de datos que pueden trabajar con acceso más directo y rápido a los datos sin depender completamente del almacenamiento en disco tradicional.

# 1.8 FUNCIONAMIENTO DE UN SISTEMA GESTOR DE BASE DE DATOS.

Un sistema gestor de base de datos (**SGBD**), también conocido como DataBase Management System (**DBMS**), es la herramienta de software que permite la creación, manipulación y administración de bases de datos. Así pues, tanto **Oracle**, **MySQL**, **MariaDB**, **OpenOffice/LibreOffice Base**, **Access**, **MongoDB** o **PostgreSQL** son sistemas gestores de bases de datos, y no bases de datos en sí.

Formalmente, un sistema gestor de bases de datos es una herramienta de propósito general que permiten crear bases de datos de cualquier tamaño y complejidad, con propósitos específicos distintos. Ofrece la interfaz entre el usuario y las bases de datos, mediante herramientas para la manipulación y administración de las mismas.

## 1.8 FUNCIONAMIENTO DE UN SISTEMA GESTOR DE BASES DE DATOS



**ACTIVIDAD 3**



## 1.8 FUNCIONAMIENTO DE UN SISTEMA GESTOR DE BASES DE DATOS

Entre las herramientas que proporciona están:

- **Herramientas para la creación y especificación de los datos.** Así como la estructura de la base de datos.
- **Herramientas para la manipulación de los datos de las bases de datos,** para añadir, modificar, suprimir o consultar datos.
- **Herramientas para la creación de copias de seguridad**
- **Herramientas de recuperación en caso de desastre**
- **Herramientas para la gestión de la comunicación de la base de datos**
- **Herramientas para la creación de aplicaciones** que utilicen esquemas externos de los datos.
- **Herramientas para la exportación e importación de datos**
- **Herramientas para el control de seguridad** de manera que los usuarios no autorizados no puedan acceder a la base de datos

## 1.8.1. FUNCIONES. LENGUAJES DE LOS SGBD

Los SGBD tienen que realizar tres tipos de funciones para ser considerados válidos. A continuación se describen estas tres funciones:

- **Función de descripción o definición.** Permite la definición de la base de datos mediante el lenguaje de definición de datos o **DDL**.

El DDL permite:

- Definir los datos junto con sus tipos.
- Definir las relaciones entre los datos
- Definir las reglas que han de cumplir los datos (restricciones).
- Definir las vistas de los distintos usuarios.

Las instrucciones SQL estándar que se diseñaron para realizar esta función son CREATE, ALTER y DROP.

- **Función de manipulación:** Permite cambiar y consultar los datos de la base de datos. Se realiza mediante un lenguaje de modificación de datos o **DML**. Mediante este lenguaje se puede:

## 1.8.1. FUNCIONES. LENGUAJES DE LOS SGBD

- Añadir datos
- Eliminar datos
- Modificar datos

En SQL se realiza mediante las instrucciones **INSERT, DELETE y UPDATE**.

- Consultar datos

En SQL se realiza mediante la instrucción **SELECT**.

Actualmente se suele diferenciar la función de consulta de datos, diferenciándose del resto de operaciones de manipulación de datos. Se habla de que la función de consulta se realiza con un lenguaje de consulta de datos o **DQL** (Data Query Language).

- **Función de control:** 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.

## 1.8.1. FUNCIONES. LENGUAJES DE LOS SGBD

- Un sistema de integridad que mantiene la integridad y la consistencia de los datos.
- Un sistema de control de concurrencia que permite el acceso compartido a la base de datos.
- Un sistema de control de recuperación que restablece la base de datos después de que se produzca un fallo del hardware o del software.
- Un diccionario de datos o catálogo accesible por el usuario que contiene la descripción de los datos de la base de datos.

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

Mediante este lenguaje se puede realizar las siguientes tareas:

- Creación y modificación de esos usuarios, permisos (CREATE USER, GRANT, REVOKE)
  - Control de transacciones. (ROLLBACK y COMMIT)
- Entre otras.

## 1.8.2. COMPONENTES

Una vez descritas las funciones que un SGBD debe llevar a cabo, imaginarás que un SGBD es un paquete de software complejo que ha de proporcionar servicios relacionados con el almacenamiento y la explotación de los datos de forma eficiente. Para ello, cuenta con una serie de componentes que se detallan a continuación:

- **Lenguajes de la base de datos.** Cualquier sistema gestor de base de datos ofrece la posibilidad de utilizar lenguajes e interfaces adecuadas para sus diferentes tipos de usuarios. A través de los lenguajes se pueden especificar los datos que componen la BD, su estructura, relaciones, reglas de integridad, control de acceso, características físicas y vistas externas de los usuarios. Los lenguajes del SGBD son: Lenguaje de Definición de los Datos (DDL), Lenguaje de Manejo de Datos (DML) y Lenguaje de Control de Datos (DCL).
- **El diccionario de datos.** Descripción de los datos almacenados. Se trata de información útil para los programadores de aplicaciones. Es el lugar donde se

## 1.8.2. COMPONENTES

deposita la información sobre la totalidad de los datos que forman la base de datos. Contiene las características lógicas de las estructuras que almacenan los datos, su nombre, descripción, contenido y organización. En una base de datos relacional, el diccionario de datos aportará información sobre:

- Estructura lógica y física de la BD.
  - Definición de tablas, vistas, índices, disparadores, procedimientos, funciones, etc.
  - Cantidad de espacio asignado y utilizado por los elementos de la BD.
  - Descripción de las restricciones de integridad.
  - Información sobre los permisos asociados a cada perfil de usuario.
  - Auditoría de acceso a los datos, utilización, etc.
- **El gestor de la base de datos.** Es la parte de software encargada de garantizar el correcto, seguro, íntegro y eficiente acceso y almacenamiento de los datos. Su **núcleo** se denomina **motor** y es el encargado de gestionar los accesos a los datos

## 1.8.2. COMPONENTES

físicos y ofrecer un mecanismo de abstracción a los niveles superiores que requieren de información demandada por los usuarios de la base de datos.

Este componente es el encargado de proporcionar una interfaz entre los datos almacenados y los programas de aplicación que los manejan. Es un intermediario entre el usuario y los datos. Es el encargado de garantizar la privacidad, seguridad e integridad de los datos, controlando los accesos concurrentes e interactuando con el sistema operativo.

- **Herramientas de la base de datos.** Son un 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, etc.
- **Usuarios de la base de datos.** En los SGBD existen diferentes perfiles de usuario, cada uno de ellos con una serie de permisos sobre los objetos de la BD. Generalmente existirán:

## 1.8.2. COMPONENTES (PERFILES)

- **Usuario final:** interactúa con los datos pero a través de alguna aplicación o servicio. Podemos hablar de tres tipos de usuarios:
  - o **Expertos/as.** Utilizan el lenguaje de manipulación de datos (DML) para acceder a la base de datos. Son usuarios que utilizan la base de datos para una gestión avanzada.
  - o **Habituales.** Utilizan las aplicaciones creadas por los desarrolladores para realizar diversas tareas: consultar, actualizar los datos... Son los que trabajan en la empresa a diario con estas herramientas y el objetivo fundamental de todo el desarrollo de la base de datos.
  - o **Ocasionales.** Son usuarios que utilizan a veces la base de datos a través de una aplicación para trabajar con ciertos datos.
- **Analista/Ingeniero de datos:** perfiles de alto nivel encargados de diseñar las arquitecturas de datos (tanto a nivel lógico como físico)



## 1.8.2. COMPONENTES (PERFILES)

- **Desarrollador:** desarrolla aplicaciones que explotan los datos, creando consultas y operaciones de inserción, borrado o modificación de los datos existentes. Un desarrollador web o multiplataforma interactúa con una o más bases de datos en sus aplicaciones.
- **Administrador de la base de datos (DBA - Database Administrator):** persona o grupo de personas responsables del control del sistema gestor de base de datos. Las principales tareas de un administrador son:
  - Codificación del esquema lógico y físico de la BD.
  - Codificación de las vistas de usuario.
  - Asignación y edición de permisos de todos los usuarios del sistema.
  - Mantenimiento y seguimiento de la seguridad.
  - Mantenimiento general del SGBD.

### 1.8.3 ARQUITECTURA

Desde comienzos de los años setenta, diversos grupos de informáticos se han ocupado del tema de la estandarización de las bases de datos. La estandarización en el campo de los SGBD es un elemento clave para lograr **sistemas abiertos**.

El grupo **ANSII/X3/SPARC**, a mediados de los años 70, definió una arquitectura en 3 niveles diferenciando tres esquemas, con el objetivo de separar a las aplicaciones de los datos, permitir que los usuarios tuvieran diferentes vistas, y utilizar un catálogo para almacenar el esquema de la base de datos.

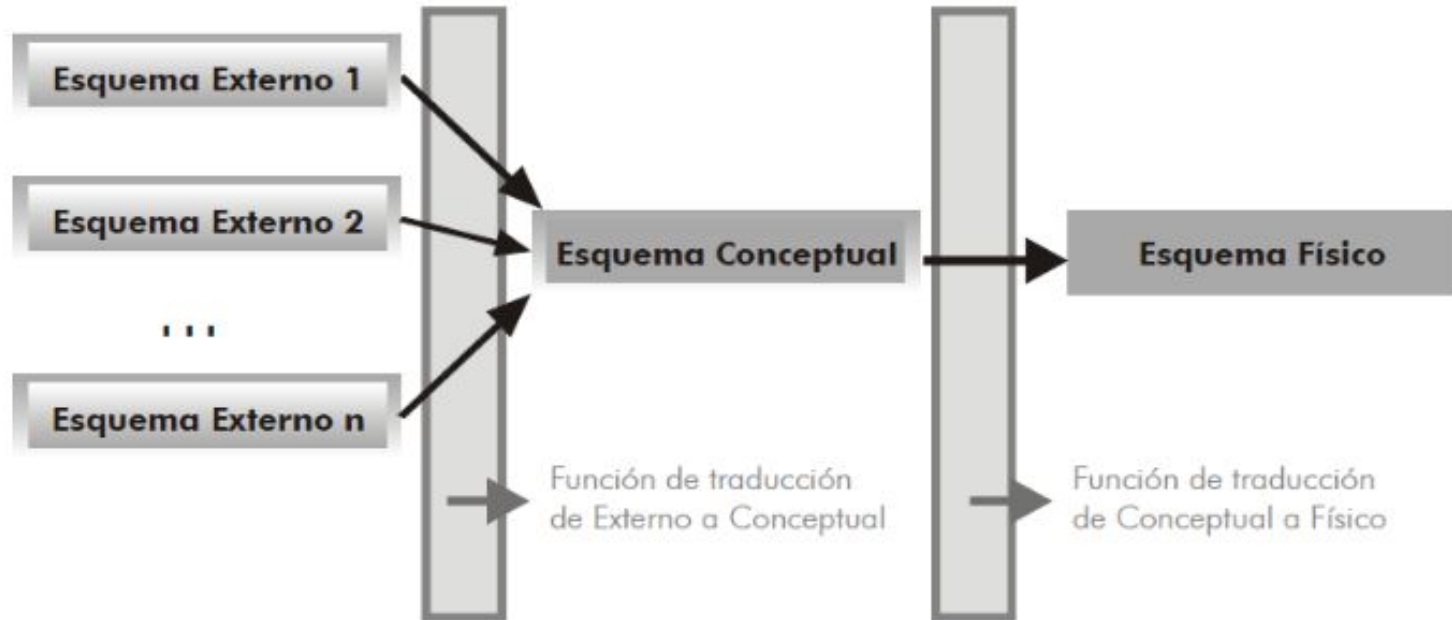
¿Y cómo se consigue? Si queremos independencia entre los datos y las aplicaciones, necesitamos que cada programa que accede a la base de datos manipule sólo ciertos datos y estructura, de manera que cada programa posea una visión de la base de datos, lo que se conoce como **esquema externo**. La unión de todos los datos y sus relaciones forman el llamado **esquema conceptual**. Finalmente, **el esquema físico** representa el almacenamiento de los datos y sus formas de acceso.

### 1.8.3 ARQUITECTURA

- **Esquema físico o nivel físico.** Representa como se almacenan realmente los datos en los dispositivos de almacenamiento, es decir, ficheros que se utilizan, registros que forman esos ficheros, como están organizados esos ficheros, directorios o carpetas, unidades de disco, servidores... Esta visión sólo la requiere **el administrador/a**.
- **Esquema conceptual o nivel conceptual.** Se trata de un esquema teórico de los datos, es decir, describe qué datos se almacenan y qué relaciones existen entre esos datos. Es el primer paso a realizar al crear una base de datos. Cuando hablamos del diseño de una BD nos estamos refiriendo a este esquema. El esquema conceptual lo realiza **diseñadores/as o analistas**. Y luego es utilizado por el desarrollador de aplicaciones.
- **Esquema externo.** Se trata de la visión de los datos que se les ofrece a los **usuarios finales** (tablas, vistas, formularios, informes, ...). Estas vistas las obtienen los desarrolladores de aplicaciones, de modo que el usuario no conoce las

### 1.8.3 ARQUITECTURA

relaciones entre los datos, como tampoco conoce todos los datos que realmente se almacenan. Así, cada usuario puede ver una base de datos diferente según sea el nivel de acceso que se le haya concedido



*Esquemas en tres niveles*

## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

### SEGÚN SU FUNCIONALIDAD.

- **Manipulación de datos:** son las más comunes, de propósito general, como Oracle, SQL Server, Access, etc.
- **Almacenes de datos** (data warehouses): permiten guardar grandes cantidades de información con el objeto de analizarla (no para su manipulación), generalmente con técnicas de minería de datos (data mining), con el objeto de ayudar a la toma de decisiones.
- **Bases de datos orientadas a aplicaciones específicas:**
  - **Bases de datos geográficas:** Guardan mapas e información asociada con ellos en los denominados sistemas de información geográfica (GIS, Geographic Information System) como por ejemplo ARC/INFO desarrollado por el ESRI (Environmental/ System Research Institute) inicialmente en 1981.

## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

- **Bases de datos para el diseño asistido por ordenador (CAD):** que guardan información como los diseños de circuitos integrados o de edificios.
- **Bases de datos multimedia.** La tendencia en SGBD es proporcionar esta capacidad de manipular imágenes, vídeos, sonidos...
- **Bases de datos documentales** (o "gestores de conocimiento"): guardan textos sobre los que se puede hacer una búsqueda libre.
- **Bases de datos para recuperación, de información distribuida en internet.** Hacen uso de motores que continuamente visitan páginas web. Tienen motores de búsqueda como en el caso de Google o Yahoo.
- **Otros muchos usos:** Militares, videojuegos, deportes, etc.

ACTIVIDAD 4

## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

### SEGÚN SU UBICACIÓN

Se pueden clasificar en dos grandes categorías:

- **SGBD Centralizados:** todos los datos se almacenan en un único servidor o ubicación física. Se pueden distinguir dos tipos principales en función de la forma en que los usuarios interactúan con los datos:
  - **Monousuario:** solo un usuario a la vez tiene acceso a la base de datos. Este tipo de bases de datos suelen ejecutarse en computadoras personales y están diseñadas para aplicaciones simples o individuales que no requieren interacción simultánea de múltiples usuarios.(como suele emplearse Access)
  - **Cliente/Servidor:** múltiples usuarios acceden a la base de datos a través de una red. Un servidor central alberga la base de datos, y los usuarios interactúan con ella mediante aplicaciones o terminales que actúan como clientes.  
La base de datos se encuentra en un servidor central y los usuarios acceden a

## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

ella desde terminales o equipos clientes a través de la red (LAN, WAN, etc.).

### **Características:**

- Permite que varios usuarios realicen consultas y actualizaciones de manera concurrente.
- Se gestionan problemas de concurrencia y bloqueo, ya que muchos usuarios pueden acceder simultáneamente a la base de datos.

### **Ventajas:**

- Soporta acceso multiusuario, lo que es adecuado para empresas o entornos con múltiples usuarios que necesitan acceder a la misma información.
- Mejora en la eficiencia y la seguridad, ya que los datos son gestionados por un servidor especializado.
- Escalable: se puede agregar más capacidad según el número de usuarios.

### **Desventajas:**



## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

- Requiere una infraestructura de red y un servidor para gestionar los datos.
- Complejidad en la administración de la base de datos y en la resolución de conflictos de concurrencia.
- Si el servidor falla, todos los usuarios pierden acceso a la base de datos.
- **SGBD Distribuido:** los datos se almacenan en varios servidores o ubicaciones físicas diferentes. Estos servidores pueden estar dispersos geográficamente, pero las bases de datos trabajan en conjunto, lo que permite que los usuarios accedan y gestionen los datos como si estuvieran en una única base de datos.

### **Características:**

- Los datos están distribuidos en diferentes ubicaciones físicas.
- Puede haber replicación de datos (copias de los mismos datos en varios servidores) o partición de datos (cada servidor tiene una parte específica de los datos).
- La base de datos parece ser una sola para los usuarios, a pesar de estar

## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

- distribuida.

### **Ventajas:**

- Mayor disponibilidad de los datos, ya que el sistema sigue funcionando aunque uno de los servidores falle.
- Acceso más rápido a los datos para los usuarios cercanos a las ubicaciones donde se almacenan.
- Mejora en la escalabilidad: es más fácil agregar servidores adicionales para manejar más datos o usuarios.

### **Desventajas:**

- Mayor complejidad en la sincronización y mantenimiento de los datos entre diferentes ubicaciones.
- Puede haber inconsistencias si los datos no se sincronizan correctamente.
- Requiere una infraestructura de red confiable para garantizar la comunicación entre los servidores distribuidos.

## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

### SEGÚN EL NÚMERO DE USUARIOS A LOS QUE DA SERVICIO.

- **Monousuario:** Los sistemas monousuarios sólo atienden a un usuario a la vez.
- **Multiusuarios:** Los sistemas multiusuario, entre los que se encuentran la mayor parte de los SGBD, atienden a varios usuarios al mismo tiempo.

### SEGÚN EL MODELO LÓGICO SOPORTADO.

- **En red:** El modelo es un grafo dirigido. Los esquemas conceptuales son redes con nodos enlazados por arcos dirigidos y nominados. Por ej., DMS1100 de Univac, IDS (Integrated Data Store) de General Electric , IDMS, DMS 1100.
- **Jerárquico:** IMS. Fue el primer SGBD jerárquico, inicialmente desarrollado por Rockwell e IBM en 1969 para el programa Apollo de la NASA, Mark IV, desarrollado por Informatics, System 2000 de MRI Systems Corporation.
- **Relacional:** basadas en el concepto matemático de relación de la teoría de conjuntos.. Por ej. Oracle, DB2, Mysql, Sybase, Informix, SQL Server, Postgresql Access, etc.

## 1.9 CLASIFICACIÓN DE LOS SISTEMAS DE GESTIÓN DE BASES DE DATOS

- **Orientado a objetos:** Objectivity/DB, de Objectivity, O2, Jasmine, inicialmente desarrollado por Fujitsu.
- **Mixto u objeto-relacional:** Illustra, de Informix (posteriormente IBM). Integraba los modelos relacional y orientado a objeto (El producto está obsoleto), OpenODB de Hewlett Packard, es relacional, con frontal orientado a objeto, UniSQL/X, de UniSQL es un SGBD orientado a objetos con SQL.
- **NoSQL:** MongoDB, Redis, Cassandra, Neo4j, es el sistema NoSQL orientado a gráficos más populares, se utiliza para representar redes sociales, recomendaciones y análisis de rutas.



## 1.10 LEGISLACIÓN

Las bases de datos y su uso por las empresas y otras instituciones están reguladas por el **Reglamento Europeo de Protección de Datos (RGPD) 2016/679**, el cual establece un marco para la protección de los datos personales y los derechos de los individuos en relación con sus datos personales. Y también con la **Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD) 3/2018**, que adapta y complementa el **RGPD**.

Los aspectos clave de la LOPDGDD son:

- **Adaptación del RGPD:** la LOPDGDD adapta el marco europeo a la realidad española, especificando ciertos aspectos del reglamento europeo.
- **Derechos digitales:** la LOPDGDD no solo trata sobre la protección de datos personales, sino que también incluye una serie de derechos digitales, como el derecho a la desconexión digital en el ámbito laboral y el derecho al testamento digital.
- **Principios del tratamiento de datos:** refuerza principios como la transparencia, la

## 1.10 LEGISLACIÓN

lealtad, la minimización de datos, la exactitud, la limitación del plazo de conservación, la integridad y la confidencialidad.

- **Consentimiento:** especifica las condiciones para obtener el consentimiento válido de los interesados, incluyendo el consentimiento explícito para datos sensibles.
- **Autoridad de control:** la [Agencia Española de Protección de Datos \(AEPD\)](#) es la autoridad encargada de velar por el cumplimiento de la normativa de protección de datos en España, supervisar y sancionar en caso de infracción, representando los intereses de los ciudadanos.

### REQUERIMIENTOS

Como desarrolladores, debemos tratar los datos personales de los usuarios de las aplicaciones. Estamos hablando de clientes, empleados, proveedores, curriculums, imágenes, etc. Toda esta información también debe cumplir con todos los requerimientos establecidos y los puntos a tener en cuenta son:

## 1.10 LEGISLACIÓN

### 1. Recogida de datos personales

- **Consentimiento:** Obtener el consentimiento explícito e informado de los usuarios antes de recoger sus datos personales. El consentimiento debe ser claro, libre, específico e inequívoco.
- **Finalidad:** Informar a los usuarios sobre la finalidad de la recogida de datos y asegurarse de que solo se utilizan para los fines especificados.

### 2. Información y Transparencia

- **Política de Privacidad:** Desarrollar y mantener una política de privacidad clara y accesible que detalle qué datos se recogen, cómo se utilizan, cómo se protegen y los derechos de los usuarios.
- **Aviso Legal y Términos de Uso:** Proporcionar avisos legales y términos de uso que incluyan información sobre el tratamiento de datos personales.

### 3. Derechos de los usuarios

- **Acceso:** Facilitar a los usuarios el acceso a sus datos personales y obtener una

## 1.10 LEGISLACIÓN

copia de los mismos.

- **Rectificación y Supresión:** Permitir que los usuarios rectifiquen datos inexactos o incompletos y soliciten la supresión de sus datos (derecho al olvido).
- **Portabilidad:** Permitir la portabilidad de los datos a otro proveedor si así lo solicitan los usuarios, en un formato estructurado, de uso común y lectura informatizada, y transmitirlos a otro responsable del tratamiento.
- **Oposición y Limitación:** Facilitar que los usuarios puedan oponerse al tratamiento de sus datos o solicitar la limitación del mismo en determinadas situaciones.

### 4. Seguridad de los Datos

- **Medidas de Seguridad:** Implementar medidas técnicas y organizativas adecuadas para proteger los datos personales contra el acceso no autorizado, pérdida, alteración o destrucción.
- **Cifrado y Anonimización:** Utilizar técnicas de cifrado y anonimización para



## 1.10 LEGISLACIÓN

proteger los datos sensibles.

### 5. Evaluación de Impacto

- **Evaluación de Impacto en la Protección de Datos (DPIA):** Realizar una DPIA si el tratamiento de datos puede implicar un alto riesgo para los derechos y libertades de los individuos, como en el caso de grandes volúmenes de datos o datos sensibles.

### 6. Responsabilidad Proactiva

- **Registro de Actividades de Tratamiento:** Mantener un registro detallado de todas las actividades de tratamiento de datos personales.
- **Designación de un Delegado de Protección de Datos (DPO):** Nombrar un DPO (Data Protection Officer) si la ley lo requiere, especialmente en organizaciones que tratan grandes volúmenes de datos personales o datos sensibles.

### 7. Transferencias Internacionales

## 1.10 LEGISLACIÓN

- **Adecuación de las Transferencias:** Asegurarse de que las transferencias de datos personales fuera del Espacio Económico Europeo (EEE) se realizan a países con un nivel de protección adecuado o mediante mecanismos como las cláusulas contractuales estándar.

### 8. Notificación de Brechas de Seguridad

- **Notificación a la AEPD:** Notificar cualquier brecha de seguridad a la Agencia Española de Protección de Datos (AEPD) dentro de las 72 horas posteriores a su detección.
- **Comunicación a los Afectados:** Informar a los individuos afectados si la brecha de seguridad puede suponer un riesgo elevado para sus derechos y libertades.

### 9. Formación y Concienciación

- **Capacitación Continua:** Formar y concienciar a todo el personal involucrado en el tratamiento de datos personales sobre las obligaciones y mejores prácticas en protección de datos.

### 10. Documentación y Auditoría

- **Documentación Adecuada:** Mantener toda la documentación necesaria para demostrar el cumplimiento con el GDPR y la LOPDGDD.
- **Auditorías Regulares:** Realizar auditorías internas periódicas para verificar el cumplimiento de las políticas y procedimientos de protección de datos.

Cumplir con estas obligaciones no solo ayudará a evitar sanciones, sino que también fomentará la confianza de los usuarios en las aplicaciones web y en la empresa que las desarrolla.

### OBLIGACIONES

Así pues, en base a los requerimiento de la **LOPDGDD**, un responsable de la gestión de una base de datos debe cumplir con las siguientes obligaciones:

- Mantener las BBDD actualizadas, para reflejar la situación actual del titular de los datos.
- No utilizar los datos de manera que se excedan los fines para los que fueron recabados.

## 1.10 LEGISLACIÓN

- Obtener el consentimiento previo al tratamiento de los datos. Para esto, se debe informar al usuario de sus derechos, para qué fines serán usados sus datos, quiénes serán los responsables de la gestión, y cuáles serán los plazos en los que serán usados.
- Los datos sensibles no deben ser recogidos a menos que sea estrictamente necesario, por lo que información como la ideología, religión o creencias son datos que no deberían pedirse. Cuando así suceda se debe advertir a la persona de que está en su derecho a no dar ese tipo de datos.
- Si la base de datos es pública, el titular de los datos tiene derecho a oponerse al tratamiento de sus datos.
- El titular debe tener garantizado el acceso, rectificación, cancelación y derecho a la oposición al uso de sus datos para los fines establecidos por la empresa. Para ello, se debe ofrecer un medio de contacto, además de especificar quién está a cargo de esos datos.

## 1.10 LEGISLACIÓN

- Cuando surge una reclamación por violación de estos derechos, es el responsable del tratamiento (la empresa) quien tiene la carga de la prueba, es decir, quien debe constatar su inocencia mediante la aportación de pruebas, y no es el titular de los datos quien tiene que demostrar que sus derechos han sido violados.
- Las bases de datos están regidas por el secreto profesional, por lo tanto, la información debe ser guardada tanto durante el tiempo en el que el titular autorizó su uso, como posteriormente.
- El secreto profesional afecta a cualquier persona que haga uso de los datos, incluidos terceros a los que se les cedan estos datos para fines determinados.
- El responsable de los datos deberá inscribir los ficheros en la Agencia Española de Protección de Datos. Previamente, es necesario notificar de este procedimiento al titular de los datos.

En definitiva, la LOPDGDD obliga a las empresas que gestionan una base de datos a cumplir con todas las obligaciones y adoptar todas las medidas de seguridad que

## 1.10 LEGISLACIÓN

garanticen que los datos de una persona solo se utilicen bajo su conocimiento y para los fines que el individuo autorice.

### FALTAS

## ACTIVIDAD 5

El incumplimiento de estas obligaciones da origen a faltas, leves en algunos casos, pero que en numerosas ocasiones son graves.

El reglamento **RGPD** contempla multas significativas a las empresas que lo incumplan. Las empresas podrán enfrentarse a multas de hasta 20 millones de euros o del 4% de su facturación anual del año anterior. Las multas podrán venir acompañadas de cualquier medida correctiva que se estime oportuna (como advertencias o reprimendas).

En España el organismo que vela por el cumplimiento de la **LOPDGDD** es la **AEPD** y es ella también la que establece los grados de infracción y las sanciones.

La información en la que basa sus sentencias son, entre otros: el tiempo que lleva cometiéndose la infracción, el volumen de información tratada de manera fraudulenta, los beneficios obtenidos de las infracciones, el grado de intencionalidad, el daño causado con sus acciones, etc.

## 1.11 BIG DATA

La recolección masiva de datos forma parte de lo que se conoce como **Big Data**. El tamaño de estos almacenes de datos es tan grande que los procesos de análisis en busca de patrones ocultos requieren de técnicas muy avanzadas en su análisis. El objetivo es más desafiante: la obtención de conocimiento en tiempo real, permitido por la propia expansión de internet, avances en las tecnologías de procesamiento de información y el almacenamiento en la nube. Esta unión de almacenes de datos gigantes y algoritmos de análisis en tiempo real es el Big Data.

Los Big Data tienen las siguientes características, las cinco uves:

- **Valor:** aportar valor a la sociedad ayudando a impulsar el desarrollo, la innovación y la competitividad, pero sin olvidar la calidad de vida de las personas.
- **Variedad:** conjugar y combinar cada tipo de información y su tratamiento específico para alcanzar un todo homogéneo.
- **Velocidad:** el análisis de la información disponible debe ser en un tiempo lo más real posible, de lo contrario, su evolución estaría siempre por encima de su tratamiento,

## 1.11 BIG DATA

haciendo de su explotación tecnologías improductivas, pues se alejan de estados reales.

- **Veracidad:** los datos deben ser de calidad y estar siempre disponibles.
- **Volumen:** la información ha de ser lo más completa posible para que la toma de decisión sea óptima.

Los datos almacenados en los Big Data son de tres grandes categorías: los datos estructurados, los no estructurados y un nivel intermedio.

- **Datos estructurados:** la naturaleza de los datos permite estructurar los datos en elementos de almacenamiento como listas, tablas, objetos, etcétera.
- **Datos no estructurados:** no es posible localizar un formato que unifique los datos almacenados por los que se almacenan tal como entran en el sistema. Pueden ser planos, por ejemplo, los textos en redes sociales, o de tipo no texto como una fotografía, un vídeo o un audio.
- **Datos semiestructurados:** no están almacenados en estructuras de datos, aunque



## 1.11 BIG DATA

sí se pueden organizar para su tratamiento. Ejemplos de estos son documentos HTML, XML, JSON o bases de datos NoSQL.

Un ejemplo de producto comercial para la gestión completa de almacenes de datos gigantes es **Apache Hadoop**, que se ha convertido en uno de los mejores entornos de trabajo para programar aplicaciones distribuidas de acceso a tecnologías **Big Data**. Usa un sistema de archivo distribuido denominado HDFS (Hadoop Distributed File System), y está compuesto por nodos de datos y un clúster de datos HDFS.



## 1.12 INTELIGENCIA DE NEGOCIOS

La inteligencia de negocios o Business Intelligence (BI) es un conjunto de arquitecturas técnicas, aplicaciones, datos, estrategias, productos y tecnologías enfocadas a la creación y administración de conocimiento, a través del análisis de los datos almacenados en grandes almacenes de datos y que facilita la toma de decisiones. Por ejemplo, los tableros o Dashboards son interfaces gráficas que permiten leer los datos recopilados de una forma cómoda y flexible a través de gráficas y tablas. Muestran resúmenes, tendencias clave, comparaciones, excepciones en históricos, entre muchos otros, con el fin de tomar decisiones óptimas. Otra herramienta, los informes ad hoc, dan respuestas a preguntas comerciales de interés. Se basan en los esquemas DSS (Sistemas de soporte a la decisión) y se enmarcan como producto comercial. También se usa comercialmente la herramienta denominada descubrimiento de datos. Esta consiste en la búsqueda de patrones en bases de datos gigantes.

Las tendencias en Business Intelligence actuales son:

## 1.12 INTELIGENCIA DE NEGOCIOS

- **Tecnología de Analytics:** software de minería de datos e inteligencia artificial (Data Mining, Deep Learning y Machine Learning). Ejemplos: Apache Spark, Azure, IBM Watson Analytics, Oracle Analytics.
- **Tecnología de Big Data:** software para manejar grandes volúmenes de datos provenientes de diversas fuentes y que se generan a gran velocidad siendo necesario su tratamiento en tiempo real. Ejemplos: Apache Hadoop, AdoreBoard, CrownEdition, Domo, Panoply, Text Emotions, Teradata, Tibco.
- **Tecnología del lenguaje:** software del procesamiento del lenguaje natural (Natural Processing Language, NPL). Ejemplos: Apache Lucene and Soir, Apache OpenNLP, Apache UIMA, Azure, Affectiva, DeepFase, GATE, Natural Language Toolkit, Kairos, SightCorp, SkyBiometry.
- **Tecnologías Integradoras:** integran más de un tipo de tecnología. Ejemplos: SAP BI y Power BI