



Unidad 1

Sistemas de almacenamiento de la información

Contenidos

■	1. Introducción	13
■	2. Ficheros	14
■	3. Bases de datos	16
■	4. Sistemas gestores de base de datos: Funciones, componentes y tipos	18
■	5. Bases de Datos Centralizadas y Distribuidas	21
■	6. Legislación sobre Protección de Datos	23
■	7. Big Data: introducción, análisis de datos, inteligencia de negocios.	25
■	8. Ejercicio	26
■	9. Test de conocimientos	27

1. Sistemas de almacenamiento de la información

1. Introducción

1.1. La información y los datos

En el mundo digital, estamos rodeados de **datos** en diversas formas: cifras, texto, imágenes, sonidos, entre otros. Sin embargo, los datos por sí solos no tienen un significado claro; es necesario organizarlos y darles un contexto para que se conviertan en **información** útil.

Por ejemplo, si encontramos una serie de números como *15, 28, 35, 42, 50*, estos datos no nos dicen nada concreto. Pero si los organizamos de la siguiente manera:

Día	Temperatura (°C)
Lunes	15
Martes	28
Miércoles	35
Jueves	42
Viernes	50

Ahora estos números tienen un contexto: representan la **temperatura** diaria a lo largo de la semana. Gracias a esta organización, podemos observar la evolución de la temperatura y detectar patrones, como un aumento progresivo en los valores, lo cual se convierte en información valiosa.

Este ejemplo ilustra cómo la **organización y el contexto** transforman los datos crudos en información que podemos entender y utilizar para tomar decisiones, como prever que el fin de semana será caluroso.

1.2. Gestión de la Información

En la era digital, el volumen de datos generados crece exponencialmente. La **gestión de la información** es esencial para garantizar que esta avalancha de datos pueda ser organizada, analizada y utilizada de manera eficiente. No basta con almacenar datos; necesitamos herramientas y sistemas que nos permitan procesarlos y convertirlos en conocimiento práctico.

La gestión de la información implica:

- **Organización:** Clasificar y estructurar los datos para su fácil acceso.
- **Extracción:** Identificar y recuperar datos relevantes según las necesidades.
- **Distribución:** Compartir información con los usuarios adecuados de manera segura.
- **Análisis:** Procesar datos para identificar tendencias, tomar decisiones y generar nuevos conocimientos.



1.3. El Rol de la Informática en la Gestión de Información

Aquí es donde la informática desempeña un papel crucial. Consta de dos componentes principales:

- **Hardware:** Proporciona los medios físicos para almacenar y procesar grandes volúmenes de datos.
- **Software:** Facilita la organización, análisis, presentación y distribución de la información.

Por ejemplo, una empresa puede utilizar sistemas informáticos para:

- Almacenar información sobre ventas, clientes y productos.
- Analizar datos históricos para prever tendencias futuras.
- Compartir informes y resultados con distintos departamentos de manera eficiente.

En resumen, la informática transforma la gestión de datos y la convierte en un proceso eficiente y escalable, permitiendo aprovechar al máximo la información disponible para la toma de decisiones y la optimización de procesos.

2. Ficheros

Los **ficheros** son una parte esencial en la informática, ya que actúan como contenedores de información digital. Imaginemos un disco duro como una gran biblioteca, y los ficheros como los libros que la componen. Cada fichero almacena un tipo específico de información, como texto, imágenes, música o programas.

2.1. Concepto de Fichero

Un fichero es una unidad de almacenamiento organizada de forma lógica dentro de un sistema informático. Es una colección de **datos** o **información** que se guarda en un dispositivo de almacenamiento, como un disco duro, una memoria USB o un CD. Cada fichero se identifica por un **nombre único** y una **extensión** que señala el tipo de contenido que almacena. Por ejemplo, "documento.docx" es un fichero de texto y "imagen.jpg" es un fichero de imagen.

2.2. Tipos de Ficheros

Los ficheros se pueden clasificar según el tipo de información que contienen. Los principales tipos son:

- **Ficheros de texto:** Almacenan texto sin formato o con formato básico, como cartas, informes o código fuente. Ejemplos: .txt, .docx, .rtf.

- **Ficheros de imagen:** Contienen imágenes en diferentes formatos, cada uno con sus propias características de compresión y calidad. Ejemplos: .jpg, .png, .gif.
- **Ficheros de audio:** Guardan grabaciones de sonido en diversos formatos. Ejemplos: .mp3, .wav, .ogg.
- **Ficheros de vídeo:** Almacenan secuencias de vídeo en distintos formatos. Ejemplos: .mp4, .avi, .mov.
- **Ficheros ejecutables:** Contienen programas que pueden ser ejecutados por el sistema operativo. Ejemplos: .exe (Windows), .app (macOS).
- **Ficheros comprimidos:** Contienen ficheros comprimidos para reducir su tamaño y facilitar su almacenamiento o transmisión. Ejemplos: .zip, .rar, .7z.
- **Ficheros de base de datos:** Almacenan datos organizados en un formato específico que puede ser manipulado y consultado por aplicaciones de bases de datos. Ejemplos: .mdb (Microsoft Access), .db (SQLite).

2.3. Organización de Ficheros

Para mantener los ficheros organizados, se utilizan las **carpetas** o **directorios**. Las carpetas son como contenedores que agrupan ficheros y otras carpetas en una estructura jerárquica. Esta organización permite una fácil navegación y búsqueda de la información almacenada.

Un sistema de archivos jerárquico permite crear subcarpetas dentro de carpetas principales, lo que mejora la organización y el acceso a los ficheros. Este sistema es utilizado por la mayoría de los sistemas operativos modernos.

2.4. Acceso a la Información en Ficheros

Existen diferentes maneras de acceder a la información almacenada en un fichero. Las principales son:

- **Acceso secuencial:** Los datos se leen de manera consecutiva, desde el principio hasta el final del fichero, o hasta que se encuentra la información buscada. Es simple, pero menos eficiente para ficheros grandes.
- **Acceso indexado:** Utiliza un **índice** que contiene la ubicación de cada registro dentro del fichero, lo que permite acceder a los datos rápidamente. Similar al índice de un libro, facilita encontrar información específica.
- **Acceso directo:** Se utiliza una fórmula matemática para calcular la posición exacta de los datos dentro del fichero, lo que permite acceder directamente a la información deseada sin necesidad de leer datos previos.

El acceso indexado y directo son más comunes en sistemas de bases de datos, donde la eficiencia es clave para la búsqueda y recuperación de grandes volúmenes de datos.

3. Bases de datos

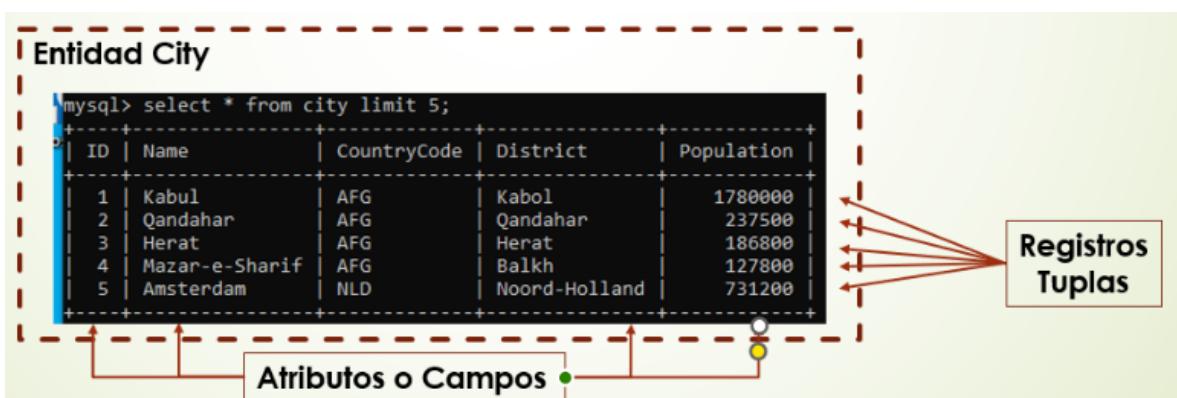
3.1. Conceptos, usos y tipos según el modelo de datos, la ubicación de la información.

Una **base de datos** es un conjunto de datos interrelacionados que pueden ser compartidos por múltiples usuarios, garantizando la máxima **integridad** (medidas de seguridad para mantener los datos correctos) y mínima **redundancia** (evitando repeticiones innecesarias de datos).

En el contexto de las bases de datos, es importante entender los siguientes conceptos:

- **Entidades:** Son los elementos de los cuales se desea almacenar datos, como personas, empleados, coches, domicilios, etc.
- **Registros:** Son cada uno de los elementos dentro de una entidad, por ejemplo, cada empleado, cada factura, cada transacción, etc.
- **Atributos o campos:** Son las características de cada registro dentro de una entidad, como el nombre de un empleado o el número de una factura.

En comparación con otros sistemas de almacenamiento de información como los **ficheros**, las bases de datos presentan numerosas ventajas. La principal diferencia es que en una base de datos, los elementos que la componen están **relacionados** entre sí, lo que permite integrar datos de diferentes archivos en un solo listado, mejorando la **coherencia** y **eficiencia** en la gestión de la información.



3.2. Características Fundamentales de las Bases de Datos

Las bases de datos se distinguen por las siguientes características:

- **Integridad:** Los datos y las relaciones guardadas en la base de datos se mantienen correctos en todo momento. Por ejemplo, no se puede guardar una transacción sin un cliente asociado.
- **Minimización de redundancias:** Cada dato solo se repetirá en diversas tablas si es necesario para gestionar la relación entre ellas, evitando duplicidades.
- **Facilidad de compartición:** Permiten compartir información entre distintos usuarios y gestionar aspectos como nombres, contraseñas y permisos de acceso.
- **Independencia de aplicaciones:** Una misma base de datos puede ser utilizada por diversas aplicaciones, sin importar el lenguaje de programación o sistema operativo, lo que las hace muy versátiles.

3.3. Tipos de bases de datos

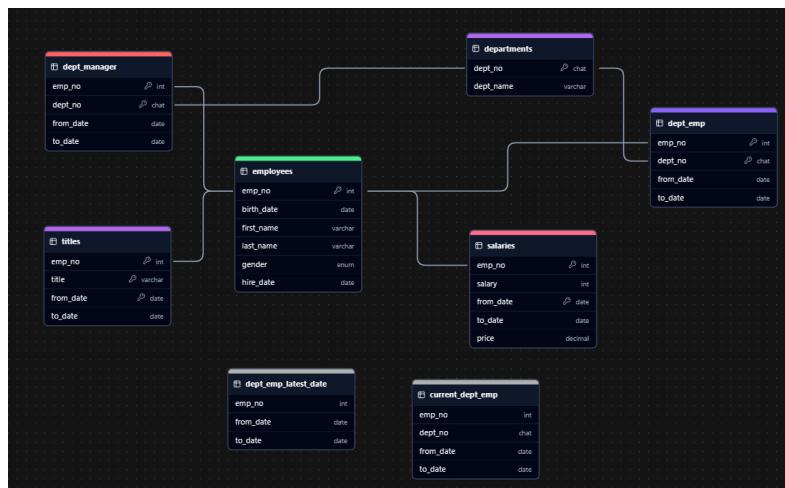
Según el modelo de datos

- **Modelo Jerárquico (obsoleto):** Las relaciones entre entidades se organizan en una estructura jerárquica, donde las entidades de nivel inferior dependen de las de nivel superior. Este modelo presenta limitaciones de flexibilidad.

- **Modelo en Red (obsoleto):** Evolución del modelo jerárquico, permitiendo que una entidad de nivel inferior dependa de varias entidades de nivel superior, pero sigue siendo más complejo que otros modelos.
- **Modelo Relacional (más utilizado):** En este modelo, las relaciones no están organizadas jerárquicamente. Las tablas están relacionadas por campos comunes, lo que permite una mayor flexibilidad y capacidad de integración de datos.
- **Modelo Orientado a Objetos:** Los registros contienen objetos de una clase, permitiendo una mayor complejidad y flexibilidad, especialmente en aplicaciones programadas bajo el paradigma orientado a objetos.
- **Modelo Documental:** Utilizado cuando los datos no siguen una estructura rígida de registros y campos, sino que consisten en documentos con estructuras jerárquicas y complejas, como **JSON**, **BSON** o **XML**.

3.4. Modelo Relacional

En el **modelo relacional**, los datos se organizan mediante una colección de **tablas**. Cada fila, o **tupla**, representa un registro individual, mientras que las columnas definen los **atributos** de esos registros. Las relaciones entre tablas se realizan mediante campos comunes, lo que facilita la **consulta** y la **integración** de la información de manera eficiente.



3.5. Modelo de Datos Orientados a Objetos

En este modelo, los datos se almacenan como **objetos**, siguiendo el paradigma de la **orientación a objetos**. Cada objeto contiene tanto **atributos** como **métodos**. Las características clave de este modelo son:

- **Encapsulación:** Permite ocultar la información al resto de los objetos, protegiendo los datos de accesos incorrectos.
- **Herencia:** Los objetos pueden heredar atributos y comportamientos de otros objetos situados en niveles superiores de una jerarquía.
- **Polimorfismo:** Permite que un único método sea aplicable a diferentes tipos de objetos.

3.6. Modelo Documental

El modelo **documental** organiza los datos en **documentos** flexibles y auto-descriptivos, generalmente en formatos como **JSON**, **BSON** o **XML**. A diferencia del modelo relacional, que utiliza tablas, este modelo permite almacenar datos **semi-estructurados** y complejos, lo que resulta en una mayor flexibilidad para manejar información cambiante.

Características principales de los documentos:

- **Unidad de datos independiente:** Cada documento contiene pares de clave-valor, y puede tener una estructura interna flexible.

- **Estructura variable:** Los documentos en una misma colección pueden tener diferentes estructuras, lo que permite adaptarse a necesidades cambiantes.

Colecciones:

- Los documentos se agrupan en **colecciones**, que son equivalentes a las **tablas** en el modelo relacional.

Flexibilidad:

- Permite realizar **cambios** en la estructura de los documentos sin necesidad de modificar esquemas rígidos, adaptándose fácilmente a nuevos tipos de datos.

Consultas complejas:

- Facilita **consultas complejas** sobre los datos, aprovechando la estructura jerárquica de los documentos.

Escalabilidad:

- Este modelo es especialmente adecuado para aplicaciones que requieren **escalabilidad**, soportando grandes volúmenes de datos y distribuciones de información a través de **servidores**.

```

    "id": "123",
    "nombre": "Juan Pérez",
    "email": "juan.perez@example.com",
    "dirección": {
        "calle": "123 Calle Principal",
        "ciudad": "Ciudad Ejemplo",
        "código_postal": "12345"
    },
    "teléfonos": ["123-456-7890", "098-765-4321"]
  
```

3.7. Según la ubicación de la información:

Las bases de datos también se pueden clasificar según su **ubicación**:

- **Centralizadas:** Todos los datos se almacenan en un único **servidor** central, facilitando el control y la administración.
- **Distribuidas:** Los datos se distribuyen en varios **servidores** interconectados, lo que mejora la **resiliencia** y el **rendimiento**.
- **En la nube:** Los datos se almacenan en **servidores en la nube**, accesibles desde cualquier lugar con una **conexión a Internet**, proporcionando una gran **flexibilidad** y escalabilidad.

4. Sistemas gestores de base de datos: Funciones, componentes y tipos

En el mundo digital actual, la **información** es un activo invaluable. Las empresas, organizaciones e individuos dependen de la capacidad de almacenar, organizar y acceder a grandes volúmenes de datos de manera eficiente. Aquí es donde entran en juego los **Sistemas Gestores de Bases de Datos (SGBD)**, un conjunto de programas que actúan como el cerebro detrás de las bases de datos, permitiendo no solo almacenar información, sino también gestionarla, protegerla y ponerla a disposición de quienes la necesitan.

Imaginemos una **biblioteca** gigante donde cada libro, revista, mapa o fotografía representa un dato. Un SGBD sería el sistema que organiza esa biblioteca, catalogando cada elemento, controlando quién puede acceder a qué información y facilitando la búsqueda y recuperación de cualquier dato de manera rápida y eficiente.

4.1. Características Fundamentales de los SGBD

Los SGBD modernos poseen una serie de características que los hacen esenciales para la gestión de la información:

- **Gestión integral de datos y relaciones:** No se limitan a almacenar datos de forma aislada. Un SGBD establece y mantiene las relaciones entre los diferentes datos, asegurando la coherencia y la integridad de la información. Por ejemplo, en una base de datos de una empresa, el SGBD no solo almacenaría la información de los **empleados** y los **departamentos**, sino que también establecería la relación entre ellos, asegurando que cada empleado esté correctamente asociado a su departamento.
- **Adaptabilidad de la estructura:** Las necesidades de información cambian con el tiempo. Un SGBD permite modificar y expandir la estructura de la base de datos para adaptarse a nuevos requisitos y al crecimiento de la información. Esto significa que se pueden agregar nuevas **tablas**, **campos** o relaciones sin tener que rediseñar todo el sistema.
- **Control de accesos y permisos:** La **seguridad** de la información es crucial. Los SGBD ofrecen herramientas para gestionar **usuarios**, **roles** y **permisos**, asegurando que solo las personas autorizadas puedan acceder a la información y realizar determinadas operaciones. Esto previene accesos no autorizados y protege la información confidencial.
- **Interfaz gráfica para la administración:** Los SGBD suelen contar con interfaces gráficas intuitivas que facilitan la administración y supervisión de la base de datos. Estas interfaces permiten a los administradores realizar tareas como crear usuarios, asignar permisos, realizar copias de seguridad y monitorizar el rendimiento del sistema, sin necesidad de escribir comandos complejos.
- **Manejo de concurrencia e integridad:** En entornos donde múltiples usuarios acceden a la base de datos simultáneamente, los SGBD gestionan la concurrencia, evitando conflictos y asegurando que la información se mantenga consistente. Por ejemplo, si dos usuarios intentan modificar el mismo dato al mismo tiempo, el SGBD se encarga de gestionar la situación para evitar errores o pérdidas de información.
- **Independencia de aplicaciones:** Un SGBD es **independiente** de las aplicaciones que acceden a los datos. Esto significa que diversas aplicaciones, escritas en diferentes lenguajes de programación o que se ejecutan en diferentes sistemas operativos, pueden acceder a la misma base de datos sin problemas.

4.2. Funciones Clave de los SGBD

- **Almacenamiento, acceso y actualización eficiente de datos:** Permiten a los usuarios almacenar, acceder y modificar datos de manera sencilla y eficiente, garantizando un alto rendimiento incluso con grandes volúmenes de información.
- **Aseguramiento de la integridad de los datos:** Mediante la aplicación de reglas y restricciones, los SGBD mantienen la consistencia de la información, previniendo errores e inconsistencias.
- **Seguridad de la información:** Integran sistemas de **seguridad** que protegen la información contra accesos no autorizados, garantizando la **confidencialidad** y la **integridad** de los datos.
- **Soporte para transacciones:** Permiten agrupar varias operaciones en una sola **transacción**, asegurando que todas las operaciones se realicen correctamente o se deshagan en caso de error.
- **Gestión de la concurrencia:** Permiten que múltiples usuarios accedan y modifiquen la base de datos simultáneamente sin comprometer la integridad de la información.

4.3. Roles y Usuarios en un Entorno de Base de Datos

Un sistema de base de datos involucra a diferentes roles con responsabilidades específicas:

- **Administradores del SGBD:** Son los responsables del funcionamiento del SGBD, incluyendo la instalación, configuración, seguridad y optimización del sistema. Se encargan de tareas como la gestión del espacio de almacenamiento, la monitorización del rendimiento y la aplicación de parches de seguridad.
- **Administradores de bases de datos:** Se centran en el diseño físico de la base de datos, su implementación y mantenimiento. Definen cómo se almacenan los datos físicamente, crean índices para optimizar el acceso a la información y se aseguran de que la base de datos funcione de manera eficiente.

- **Diseñadores de la base de datos:** Se encargan del diseño lógico de la base de datos. Identifican los datos que se necesitan almacenar, las relaciones entre ellos y las restricciones que se deben aplicar para mantener la integridad de la información. Su objetivo es crear un modelo de datos que satisfaga las necesidades de la organización.
- **Programadores de aplicaciones:** Desarrollan las aplicaciones que permiten a los usuarios finales interactuar con la base de datos. Estas aplicaciones pueden ser programas de escritorio, aplicaciones web o aplicaciones móviles, y se encargan de presentar la información de forma amigable y permitir a los usuarios realizar operaciones como consultas, inserciones, actualizaciones y eliminaciones de datos.
- **Usuarios finales:** Son las personas que utilizan las aplicaciones para acceder a la información de la base de datos y realizar sus tareas diarias. No necesitan conocer el funcionamiento interno de la base de datos, simplemente utilizan las aplicaciones para consultar, ingresar o modificar datos según sus necesidades.

4.4. Tipos de SGBD: Una Variedad de Opciones

Los SGBD se pueden clasificar de diferentes maneras según diversos criterios:

Por Modelo Lógico

- **Jerárquico:** Organiza los datos en una estructura jerárquica de árbol, donde cada registro tiene un único parent pero puede tener varios hijos. Este modelo es menos flexible que otros y se utiliza en aplicaciones con estructuras de datos bien definidas.
- **En Red:** Permite que un registro tenga múltiples padres y varios hijos, creando una red de relaciones más flexible. Este modelo es más complejo que el jerárquico pero ofrece mayor flexibilidad.
- **Relacional:** Es el modelo más utilizado actualmente. Organiza los datos en **tablas** con filas y columnas, donde las relaciones entre las tablas se establecen mediante **claves**. Ofrece gran flexibilidad y eficiencia en el manejo de la información.
- **Objeto-Relacional:** Combina características de los modelos relacional y orientado a objetos, permitiendo el manejo de **datos complejos** y **tipos de datos** definidos por el usuario.
- **Orientado a Objetos:** Almacena **datos** en forma de objetos, similar a la programación orientada a objetos. Cada objeto contiene datos y **métodos** que operan sobre esos datos.

Por Número de Usuarios

- **Monousuario:** Diseñado para ser utilizado por un solo usuario a la vez. Son adecuados para aplicaciones pequeñas o personales.
- **Multiusuario:** Permiten el acceso simultáneo de varios usuarios a la base de datos. Son esenciales para aplicaciones **empresariales** y sistemas con múltiples usuarios concurrentes.

Por Distribución

- **Centralizados:** Toda la base de datos reside en un único **servidor**. Son más fáciles de administrar pero tienen limitaciones en cuanto a **escalabilidad** y **disponibilidad**.
- **Distribuidos:** La base de datos se distribuye en varios **servidores** interconectados. Ofrecen mayor **escalabilidad**, **disponibilidad** y **tolerancia a fallos**.

Por Tipo de Licencia

- **Comerciales:** Son de pago y suelen ofrecer **soporte técnico**, **actualizaciones** regulares y características avanzadas. Son utilizados por empresas que requieren funcionalidades robustas y soporte profesional.
- **Libres:** Son gratuitos y de **código abierto**, lo que permite su uso, modificación y distribución sin restricciones. Son una buena opción para proyectos con presupuestos limitados o para aquellos que requieren mayor flexibilidad.

Por Tipo de Datos

- **Relacionales Estándar:** Manejan tipos de datos básicos como **números**, **texto** y **fechas**.
- **XML:** Diseñados para trabajar con **datos** en formato **XML**.

- **Objeto-Relacionales:** Extienden el modelo relacional para incluir **tipos de datos complejos** como imágenes, audio o vídeo.
- **Orientados a Objetos:** Soportan tipos de datos **objeto**, que encapsulan datos y métodos.

4.5. Lenguajes de los SGBD: SQL

SQL (Structured Query Language) es el lenguaje estándar para interactuar con bases de datos relacionales. Permite a los usuarios y a las aplicaciones realizar operaciones como consultar, insertar, actualizar y eliminar datos.

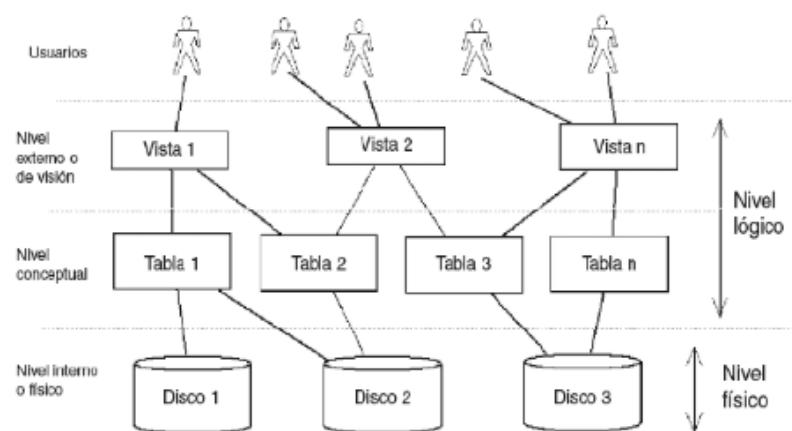
SQL se divide en sublenguajes:

- **DML (Data Manipulation Language):** Para manipular datos (**SELECT, INSERT, UPDATE, DELETE**).
- **DDL (Data Definition Language):** Para definir la estructura de la base de datos (**CREATE, ALTER, DROP**).
- **DCL (Data Control Language):** Para controlar el acceso a los datos (**GRANT, REVOKE**).
- **TCL (Transaction Control Language):** Para manejar **transacciones**.

4.6. Niveles de Abstracción en una Base de Datos

Para facilitar la interacción con la base de datos, se definen diferentes niveles de abstracción:

- **Nivel interno o físico:** Describe cómo se almacenan los datos físicamente en el sistema. Es el nivel más bajo y se ocupa de la organización de los archivos, índices y estructuras de almacenamiento.
- **Nivel conceptual:** Proporciona una visión global de la base de datos, describiendo las entidades, atributos y relaciones sin entrar en detalles de la implementación física.
- **Nivel externo o de visión:** Define vistas personalizadas de la base de datos para diferentes grupos de usuarios. Cada vista muestra solo la información relevante para un grupo de usuarios específico, ocultando los datos que no son necesarios.



5. Bases de Datos Centralizadas y Distribuidas

Las bases de datos se pueden clasificar en dos tipos principales según la ubicación de los datos: centralizadas y distribuidas. Cada tipo tiene sus propias características, ventajas y desventajas.

5.1. Bases de Datos Centralizadas

En una base de datos centralizada, todos los datos se almacenan en un único servidor o ubicación física. Todos los usuarios y aplicaciones acceden a la base de datos desde ese punto central.

Ventajas:

- Gestión simplificada: La administración y el mantenimiento se concentran en un único lugar.
- Menor coste: La infraestructura es más simple y requiere menos inversión.
- Mayor coherencia: Es más fácil mantener la coherencia de los datos al estar en un solo lugar.
- Rendimiento optimizado: Se puede optimizar el rendimiento para un único servidor.

Desventajas:

- Punto único de fallo: Si el servidor central falla, toda la base de datos se vuelve inaccesible.
- Escalabilidad limitada: Escalar la base de datos puede ser complejo y costoso.
- Posibles cuellos de botella: El acceso al servidor central puede saturarse si hay muchos usuarios concurrentes.

5.2. Bases de Datos Distribuidas

En una base de datos distribuida, los datos se distribuyen en varios servidores o nodos interconectados. Estos nodos pueden estar en la misma ubicación física o en diferentes ubicaciones geográficas.

Ventajas:

- Alta disponibilidad: Si un nodo falla, los otros nodos pueden seguir funcionando.
- Escalabilidad: Es más fácil escalar la base de datos añadiendo nuevos nodos.
- Rendimiento mejorado: Los usuarios pueden acceder a los datos desde el nodo más cercano, reduciendo la latencia.
- Flexibilidad: Se adapta mejor a entornos con diferentes ubicaciones o necesidades.

Desventajas:

- Gestión más compleja: La administración y el mantenimiento son más complejos al tener que coordinar varios nodos.
- Mayor coste: La infraestructura es más compleja y requiere mayor inversión.
- Riesgo de inconsistencia: Es más difícil mantener la coherencia de los datos al estar distribuidos.

	Centralizada	Distribuida
Ubicación de los datos	Almacenados en un único servidor	Distribuidos en varios servidores
Gestión	Simplificada	Más compleja
Coste	Menor	Mayor
Escalabilidad	Limitada	Alta
Disponibilidad	Un punto único de fallo	Alta disponibilidad
Rendimiento	Puede optimizarse para un único servidor	Puede mejorar al acceder a datos desde el nodo más cercano
Consistencia de datos	Más fácil de mantener	Más difícil de mantener
Ejemplos	MySQL en un solo servidor	MongoDB, Cassandra

5.3. Técnicas de Fragmentación

En una base de datos **centralizada**, todos los datos se almacenan en un único servidor o ubicación física. Todos los usuarios y aplicaciones acceden a la base de datos desde ese punto central.

Ventajas:

- **Gestión simplificada:** La administración y el mantenimiento se concentran en un único lugar.
- **Menor coste:** La infraestructura es más simple y requiere menos inversión.
- **Mayor coherencia:** Es más fácil mantener la coherencia de los datos al estar en un solo lugar.
- **Rendimiento optimizado:** Se puede optimizar el rendimiento para un único servidor.

Desventajas:

- **Punto único de fallo:** Si el servidor central falla, toda la base de datos se vuelve inaccesible.
- **Escalabilidad limitada:** Escalar la base de datos puede ser complejo y costoso.
- **Posibles cuellos de botella:** El acceso al servidor central puede saturarse si hay muchos usuarios concurrentes.

5.4. Tecnología Blockchain

La **tecnología blockchain** es un tipo especial de base de datos distribuida que se caracteriza por su **seguridad, transparencia e inmutabilidad**. Se utiliza principalmente en criptomonedas como **Bitcoin**, pero también tiene aplicaciones en otros campos donde la seguridad y la confianza son esenciales, como en contratos inteligentes y registros descentralizados.

Características de Blockchain:

- **Descentralización:** No hay un único punto de control, lo que reduce los riesgos asociados con fallos de servidores centralizados.
- **Inmutabilidad:** Una vez que los datos se añaden a la cadena de bloques, no se pueden modificar, lo que garantiza la integridad de la información.
- **Transparencia:** Todos los nodos tienen acceso a la misma información, lo que facilita la verificación de las transacciones y aumenta la confianza en el sistema.

En resumen, la elección entre una base de datos centralizada o distribuida depende de las necesidades específicas de cada aplicación. Las bases de datos centralizadas son más simples y económicas, mientras que las bases de datos distribuidas ofrecen mayor **disponibilidad y escalabilidad**. La fragmentación es una técnica útil para optimizar el rendimiento y la disponibilidad en bases de datos distribuidas, y la **tecnología blockchain** ofrece un enfoque único y altamente seguro para la gestión de datos en entornos descentralizados.

6. Legislación sobre Protección de Datos

6.1. El Reglamento General de Protección de Datos (GDPR)

El **GDPR** es una normativa de la **Unión Europea** que establece las reglas sobre cómo se deben recopilar, procesar y almacenar los **datos personales** de los ciudadanos dentro del **Espacio Económico Europeo** (EEE). Su principal objetivo es **proteger la privacidad** de los individuos y darles **control** sobre sus datos personales, garantizando que se gestionen de manera transparente, segura y respetuosa con los derechos fundamentales. El GDPR también establece obligaciones claras para las empresas, incluidas las **notificaciones de violaciones de datos** y las **evaluaciones de impacto** sobre la protección de datos.

6.2. La Nueva Ley de Protección de Datos (LOPDGDD)

En España, la **Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD)** adapta el **GDPR** al marco legal español. Regula el tratamiento de **datos personales**, tanto automatizado como no automatizado, y establece los **derechos** de los individuos en relación con sus datos, fortaleciendo su **privacidad** y **seguridad** en el entorno digital.

Puntos clave de la LOPDGDD:

- **Consentimiento:** Se requiere el **consentimiento explícito** de la persona para el tratamiento de sus datos personales, y este debe ser **informado, inequívoco y revocable**.
- **Derechos de los usuarios:** Los individuos tienen derecho a **acceder, rectificar, suprimir, oponerse, limitar y portabilizar** sus datos, asegurando un control efectivo sobre la información que comparten.
- **Seguridad de los datos:** Los responsables del tratamiento deben implementar **medidas de seguridad** adecuadas para proteger los datos personales contra **pérdidas, alteraciones o accesos no autorizados**.
- **Limitación de la finalidad:** Los datos solo pueden ser utilizados para la **finalidad específica** para la que fueron recopilados, garantizando que no se procesen de manera incompatible con esa finalidad.
- **Tratamiento de datos sensibles:** Se establecen restricciones más estrictas para el tratamiento de datos sensibles, como aquellos relacionados con **creencias religiosas, origen racial o orientación sexual**, entre otros.

6.3. La "Ley de las Cookies"

En 2020, se introdujo una nueva normativa europea que regula el uso de **cookies** y tecnologías similares en las páginas web. Esta ley exige que las páginas web obtengan el **consentimiento previo e informado** de los usuarios antes de almacenar o acceder a información en sus dispositivos. Además, se debe proporcionar a los usuarios la opción de **gestionar** sus preferencias en cuanto al uso de cookies, mejorando la **transparencia** y el **control** sobre el tratamiento de sus datos personales.

6.4. Implicaciones para los Diseñadores y Administradores de Bases de Datos

Aunque los diseñadores y administradores de bases de datos no son directamente responsables del cumplimiento legal de la empresa, tienen un papel crucial en la **protección de datos** personales y el cumplimiento normativo:

- **Asesoramiento:** Deben asesorar a la empresa sobre la **necesidad de consultar** con expertos legales para garantizar que las prácticas de manejo de datos cumplen con la normativa aplicable.
- **Seguridad:** Es fundamental que colaboren en la implementación de **medidas de seguridad** adecuadas para proteger los datos personales almacenados en las bases de datos, como **cifrado, control de acceso y auditorías regulares**.
- **Diseño de la base de datos:** Al diseñar la estructura de la base de datos, deben asegurarse de que se incorpora la **minimización de datos**, almacenando solo la información estrictamente necesaria y garantizando su **integridad y confidencialidad**.

En resumen, la legislación sobre protección de datos es un tema fundamental en la era digital. Las bases de datos que almacenan información personal deben cumplir con las leyes y regulaciones para proteger la **privacidad** y **seguridad** de los individuos. Los diseñadores y administradores de bases de datos tienen un papel esencial en el cumplimiento de la normativa y deben estar siempre al tanto de los cambios y **exigencias legales** vigentes.

7. Big Data: introducción, análisis de datos, inteligencia de negocios.

7.1. Las 3 "V" del Big Data

Para comprender mejor el concepto de **Big Data**, se suelen identificar tres características clave:

- **Volumen:** La cantidad de datos generados cada segundo es masiva, alcanzando **terabytes, petabytes** e incluso **zettabytes** de información, lo que plantea desafíos tanto en almacenamiento como en procesamiento.
- **Variedad:** Los datos provienen de múltiples fuentes, como **redes sociales, dispositivos móviles, sensores IoT y registros de transacciones**. Además, esta variedad abarca diferentes tipos de datos: **texto, imágenes, audio, video**, y datos estructurados y no estructurados.
- **Velocidad:** Los datos se generan y deben procesarse a una velocidad vertiginosa. Ejemplos clave incluyen la recopilación de datos en **tiempo real de mercados financieros** o el análisis continuo de datos de **sensores en vehículos autónomos**.

7.2. Domando el Diluvio: Análisis de Datos en Big Data

El **Big Data** no se trata solo de almacenar enormes cantidades de datos, sino de **extraer valor** de ellos mediante técnicas avanzadas de **análisis de datos**:

- **Procesamiento de datos:** Antes de realizar cualquier análisis, los datos deben ser **limpiados, transformados y organizados**. Este proceso incluye eliminar duplicados, corregir errores y garantizar que los datos estén en el formato adecuado para el análisis.
- **Modelado predictivo:** Utilizando algoritmos de **aprendizaje automático** (machine learning), es posible crear modelos que **predicen el comportamiento futuro**, como en la **prevención del fraude, la predicción de la demanda o la personalización de la experiencia del cliente**.
- **Análisis descriptivo:** Este enfoque busca **resumir y analizar** datos históricos para identificar **patrones y tendencias**, lo que permite entender mejor fenómenos como el comportamiento de los consumidores o el rendimiento de procesos empresariales.
- **Análisis prescriptivo:** Además de describir y predecir, el análisis prescriptivo recomienda acciones **concretas** a seguir, basándose en los datos. Por ejemplo, un sistema prescriptivo podría sugerir las mejores estrategias de **marketing** o cómo **optimizar una cadena de suministro**.

7.3. Inteligencia de Negocios (BI): Tomando Decisiones Inteligentes

La **Inteligencia de Negocios (BI)** es un conjunto de tecnologías, herramientas y estrategias utilizadas para analizar los datos y tomar decisiones más informadas. La relación entre **Big Data** y **BI** es fundamental, ya que **BI** proporciona el marco para convertir los grandes volúmenes de datos en **información procesable**:

- **Visualización de datos:** Las herramientas de BI permiten la creación de **gráficos, paneles interactivos** y otras visualizaciones, lo que facilita la interpretación de datos complejos de manera rápida y clara. Esto ayuda a los tomadores de decisiones a identificar rápidamente tendencias y patrones significativos.
- **Toma de decisiones basada en datos:** La **toma de decisiones** se basa cada vez más en el análisis exhaustivo de datos. Big Data proporciona la **materia prima** que, a través de BI, se transforma en información útil para **tomar decisiones estratégicas** fundamentadas.
- **Optimización de procesos:** A través del análisis de **datos operativos** y de **mercado**, BI ayuda a identificar áreas de mejora dentro de las operaciones empresariales. Esto puede resultar en la **reducción de costes, la aumento de eficiencia y la mejora de la satisfacción del cliente**.

Herramientas populares como **Qlik Sense, Microsoft Power BI y Tableau** permiten a las organizaciones visualizar y analizar los datos de **Big Data**, facilitando la toma de decisiones más estratégicas y fundamentadas.

En resumen, **Big Data**, el **análisis de datos** y la **Inteligencia de Negocios (BI)** son fundamentales en la era digital. Ayudan a las organizaciones a transformar grandes volúmenes de datos en **conocimiento valioso**, optimizando procesos, mejorando la competitividad y ofreciendo una base sólida para decisiones empresariales más informadas.

8. Ejercicio

Imagina que trabajas como consultor para una empresa que está a punto de lanzar un nuevo proyecto. La empresa necesita seleccionar un Sistema Gestor de Bases de Datos (SGBD) que mejor se ajuste a sus necesidades específicas. Tu tarea es evaluar diferentes SGBD y recomendar el más adecuado para el proyecto en cuestión.

La empresa está considerando un proyecto que involucra una aplicación web con alta concurrencia de usuarios, manejo de grandes volúmenes de datos y requisitos de alta disponibilidad. Además, la empresa valora la facilidad de uso y el costo del SGBD.

Pasos recomendados:

1. Investiga y compara al menos tres SGBD populares. Considera tanto opciones comerciales como de código abierto.
2. Para cada SGBD, proporciona una comparación basada en los siguientes criterios: **Modelo de Datos Soportado: Escalabilidad, Rendimiento, Coste**
3. Basado en el análisis de requisitos y la comparación de SGBD, selecciona el SGBD que consideres más adecuado para el proyecto.
4. Justifica tu elección detallando cómo el SGBD recomendado cumple con los requisitos específicos del proyecto y cómo se compara con las otras opciones en términos de rendimiento, escalabilidad, costo y facilidad de uso.

9. Test de conocimientos

1. ¿Qué modelo de base de datos organiza las relaciones en una estructura jerárquica donde las entidades de nivel inferior dependen de las de nivel superior?

- a) Modelo en Red
- b) Modelo Relacional
- c) Modelo Jerárquico
- d) Modelo Orientado a Objetos

2. ¿Cuál de los siguientes NO es un lenguaje de los SGBD?

- a) DML
- b) DDL
- c) XML
- d) TCL

3. ¿Qué tipo de archivo tiene una estructura simple y almacena datos en texto sin formato?

- a) Fichero Indexado
- b) Fichero de Acceso Directo
- c) Fichero Plano
- d) Fichero Secuencial

4. ¿Cuál de las siguientes NO es una característica clave de la tecnología Blockchain?

- a) Centralización
- b) Inmutabilidad
- c) Seguridad
- d) Transparencia

5. ¿Qué significa GDPR?

- a) Global Data Protection Regulation
- b) General Data Protection Regulation
- c) Government Data Protection Rule
- d) General Database Protection Regulation

6. ¿Cuál de los siguientes NO es un tipo de SGBD según su licencia?

- a) Comercial
- b) Libre
- c) De código abierto
- d) Gratuito

7. ¿Qué tipo de SGBD permite la conexión simultánea de varios usuarios?

- a) Monousuario
- b) Multiusuario

- c) Centralizado
- d) Distribuido

8. ¿Cuál de los siguientes es un ejemplo de SGBD relacional?

- a) MongoDB
- b) ObjectDB
- c) IMS
- d) MySQL

9. ¿En qué nivel de arquitectura de base de datos se definen varios esquemas externos o vistas para diferentes grupos de usuarios?

- a) Nivel interno o físico
- b) Nivel conceptual
- c) Nivel externo o de visión
- d) Nivel lógico

10. ¿Cuál de las siguientes NO es una función principal de un SGBD?

- a) Control de acceso
- b) Gestión de la concurrencia
- c) Compilación de código
- d) Aseguramiento de la integridad.