

BBDD

Bases de Datos – UD01 - Almacenamiento de la información



Objetivos

- Analizar los sistemas lógicos de almacenamiento y sus características.
- Identificar los distintos tipos de bases de datos según el modelo de datos utilizado y según la ubicación de la información.
- Entender y resumir las principales funciones de un sistema gestor de base de datos.
- Saber clasificar los sistemas gestores de bases de datos.
- Reconocer la utilidad de las bases de datos distribuidas.
- Entender políticas de fragmentación de la información.
- Identificar la legislación vigente sobre protección de datos.
- Reconocer los conceptos de Big Data y de la inteligencia de negocios.



Ficheros – Conceptos básicos

Los equipos informáticos necesitan almacenar y registrar multitud de tipos de información (música, datos contables, videos, páginas webs, partidas de videojuegos...). Toda esa información se almacena en dispositivos de almacenamiento del ordenador (discos duros, tarjetas de memoria, pen drives...). Para poder organizar toda esa información se utilizan los **ficheros o archivos**.

Los **ficheros** o **archivos** representan un conjunto de información almacenada bajo una misma agrupación lógica (formato) en el sistema operativo. Son por tanto, colecciones ordenadas de datos que tienen entre sí una relación y que se almacenan de forma permanente en un dispositivo de memoria no volátil.

Características básicas de un fichero o archivo:

- **Nombre:** identifica al archivo.
- **Extensión:** indica al sistema operativo qué estructura tiene el fichero y por lo tanto que programa o programas deben ser usados para su leer su contenido.
- **Contenido:** la información almacenada en el archivo, pueden ser de texto o de datos binarios.

La combinación del nombre y extensión deben de ser únicos para todos los ficheros localizados dentro del mismo directorio.



Ficheros – Conceptos básicos

Los ficheros se clasifican según su contenido (texto o binario) y según su organización (secuencial, directa, indexada).

Tipos de ficheros según contenido:

- **Planos:** Están compuesto por texto y son legibles directamente por seres humanos. Se llaman también **ASCII**. La gran ventaja de este tipo de archivos es que son estándar en todo tipo de sistemas (Windows, MacOS, Unix, etc.)
- **Binarios:** no son de texto y requieren que el programa que los lea reconozca su formato. Ejemplo ejecutables, de aplicaciones o multimedia. Típicamente los archivos que componen una base de datos son binarios.

Ficheros – Conceptos básicos

Tipos de ficheros según su organización:

- **Acceso secuencial:** sus registros están almacenados de forma contigua, de manera, que la única forma de acceder a él, es leyendo un registro tras otro desde el principio hasta el final. Se pueden almacenar en dispositivos de memoria secundaria secuenciales (ej. cintas magnéticas).
- **Acceso directo:** se puede acceder a un registro indicando la posición relativa del mismo dentro del archivo. Debe almacenarse en dispositivos de memoria secundaria de acceso directo (ej. discos duros)
- **Acceso indexado:** Se basan en la utilización de índices, que permiten el acceso a un registro del fichero de forma directa, sin tener que leer los anteriores. Existirá una zona de registros en la que se encuentran los datos del archivo y una zona de **índices**, que contiene una tabla con las claves de los registros y las posiciones donde se encuentran los mismos. La tabla de índices estará ordenada por el campo clave (ej. índice de un libro)





Ficheros – Conceptos básicos

Las **operaciones** disponibles sobre ficheros son:

- **Creación**
- **Borrado**
- **Apertura**
- **Cierre**
- **Lectura**
- **Escritura**

Los ficheros **estructurados**: están formados por **registros** lógicos que contienen datos relativos a un mismo elemento y que están compuestos por **campos** que contienen cada una de las informaciones elementales que forman un registro.



Ficheros – Actividad

1. A continuación se muestra el contenido del fichero “AgendaTelefonica.dat”. La primera línea es una línea de cabecera, que sólo proporciona información.

Nombre	Dirección	CP	Teléfono
Ana	Calle Perú, 5	06010	999255124
Javier	Calle Faisán, 3		989525289
Lourdes	Calle Perdiz, 7	06259	929444548

- ¿Crees que es un archivo de acceso directo o secuencial?
- ¿Podrías identificar cuántos registros tiene?
- ¿Cómo sabes dónde empieza un registro y acaba el siguiente?
- ¿Cuántos campos tiene el registro de Javier?
- ¿Cómo sabes dónde empieza un campo y empieza el siguiente?

Ficheros – Actividad

2. A continuación se muestra el contenido del fichero “AgendaTelefonica.csv”. La primera línea es una línea de cabecera, que sólo proporciona información.

```
Nombre;Dirección;CP;Teléfono  
Ana;Calle Perú, 5;06010;999255124  
Javier;Calle Faisán, 3;;989525289  
Lourdes;Calle Perdiz, 7;06259;929444548
```

Los campos, ¿son de longitud fija o variable?

¿Cómo sabes dónde empieza un registro y acaba el siguiente?

¿Cuántos campos tiene el registro de Javier?

¿Cómo sabes dónde empieza un campo y empieza el siguiente?

¿Qué instrucciones darías a una persona para que fuera capaz de procesar la información de este archivo?



Ficheros – Actividad

3. A continuación se muestra el contenido del fichero “AgendaTelefonica.mix”. El tipo de letra utilizado hace que todos los caracteres ocupen lo mismo.

```
03-13-09AnaCalle Perú, 5999255124  
06-15-09JavierCalle Faisán, 3989525289  
07-15-09LourdesCalle Perdiz, 7929444548
```

¿Cómo sabes dónde empieza un registro y acaba el siguiente?

¿Cuántos campos tiene el registro de Javier?

¿Qué son esos números al inicio de cada registro o para qué valen?

¿Cómo sabes dónde empieza un campo y empieza el siguiente?

¿Qué instrucciones darías a una persona para que fuera capaz de procesar la información de este archivo?

Ficheros – Actividad

4. A continuación se muestran dos ficheros que juntos representan la información de entradas de la agenda telefónica. Se ha añadido al principio de cada registro un campo adicional “Grupo” y el “DNI”. Se han eliminado también los saltos de línea. Después de “Amigos” hay cuatro espacios, después de “Trabajo” hay tres espacios y después de “Familia” hay tres espacios.

Agenda.inf	Amigos	11111111x	Ana	Calle Perú, 5	999255124	Amigos
	22222222k	Javier	Calle Faisán, 3	989525289	Trabajo	
	33333333z	Lourdes	Calle Perdiz, 7	929444548	Familia	
	44444444p	Ana	Calle Sol, 27	933111111		

Agenda.ind	11111111x	0
	22222222k	60
	33333333z	120
	44444444p	180

Los campos de “Agenda.inf”, ¿son de longitud fija o variable?

¿Cómo sabes dónde empieza un registro y acaba el siguiente en “Agenda.inf”?

¿Cuántos campos tiene el registro de Javier en “Agenda.inf”?

¿Para qué sirve el archivo “Agenda.ind”?

¿Por qué en “Agenda.ind” uso el DNI en vez del nombre de la persona?

¿Podría leer la información de “Agenda.inf” sin tener “Agenda.ind”?



Bases de Datos – Actividad

Buscad los siguientes términos en internet, relativos a bases de datos hasta que seáis capaces de sintetizar en una frase corta una definición.

Aunque no sepáis lo que significa, si buscáis la definición de cada término en tres o cuatro o más sitios distintos, te harás una idea de la “esencia” de cada término.

- *Dato*
- *Base de datos*
- *Campo*
- *Registro*
- *Tipo*
- *Campo clave*
- *Tabla de una base de datos*
- *Consulta a una base de datos*
- *Informes*
- *Transacciones*
- *Sistema Gestor de Bases de Datos (SGBD)*
- *SQL*

Bases de Datos – Conceptos básicos

Una **base de datos** es un conjunto de información almacenada en ficheros y relativa a un mismo concepto o área de conocimiento o negocio.

Es también un conjunto estructurado de datos que representa entidades y sus interrelaciones, almacenados con la mínima redundancia y posibilitando el acceso a ellos eficientemente por parte de varias aplicaciones y usuarios.

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

Por ejemplo, en una fotografías serían metadatos útiles para su procesamiento posterior: la velocidad de obturación, apertura del diafragma, distancia focal, coordenadas del lugar donde se hizo el disparo o el modelo de cámara utilizado. En una **Base de datos**, la información sobre tablas, columnas, vistas, índices, procedimientos, usuarios...



Bases de Datos – Conceptos básicos

Una **base de datos** constará de los siguientes elementos:

- **Entidades:** objeto real o abstracto con características diferenciadoras de otros, del que se almacena información en la base de datos.
- **Atributos:** son los datos que se almacenan de la entidad. Cualquier propiedad o característica de una entidad puede ser atributo.
- **Registros:** donde se almacena la información de cada entidad (ocurrencias). Es un conjunto de atributos que contienen los datos que pertenecen a una misma repetición de entidad.
- **Campos:** donde se almacenan los atributos de cada registro.

Bases de Datos – usos

Algunos usos comunes de las **base de datos** son:

- **Organismos públicos:** datos ciudadanos, certificados, etc.
- **Banca:** información de clientes, cuentas, transacciones, préstamos, etc.
- **Líneas aéreas:** información de clientes, horarios, vuelos, destinos, etc.
- **Institutos:** información de estudiantes, ciclos, horarios, materias, etc.
- **Transacciones de tarjeta de crédito:** para comprar con tarjetas de crédito y la generación de los extractos mensuales.
- **Telecomunicaciones:** para guardar registros de llamadas realizadas, generar facturas mensuales, mantener el saldo de las tarjetas telefónicas de prepago.
- **Medicina:** información hospitalaria, biomedicina, genética, etc.
- **Justicia y Seguridad:** delincuentes, casos, sentencias, investigaciones, etc.
- **Sistemas de posicionamiento geográfico.**
- **Hostelería y turismo:** reservas de hotel, vuelos, excursiones, etc.
- **Cultura:** gestión de bibliotecas, museos virtuales, etc.

Bases de Datos – usos

¿Quién utiliza las **bases de datos**?

- El **administrador**: encargada de la creación o implementación física de la base de datos, de tomar las decisiones relacionadas con el funcionamiento físico del almacenamiento de información o de establecer las políticas de seguridad y de acceso.
- Los **diseñadores**: encargadas de diseñar cómo será la base de datos identificando los datos, las relaciones entre ellos, sus restricciones conociendo las reglas del negocio a modelar.
- Los **programadores de aplicaciones**: desarrollan los programas de aplicación que servirán a los usuarios finales (consultas de datos, inserción, actualización o eliminación de los mismos).
- Los **usuarios finales**: Son los clientes finales que explotan los datos almacenados en la base de datos.



Bases de Datos – Tipos

Por la capacidad y alcance del sistema gestor:

- **Corporativas:** Usadas por grandes empresas, orientadas a manejar gran cantidad de información, muchos usuarios y una gran carga de trabajo. Requieren mucha inversión en hardware.
- **Ofimáticas:** Orientados a pequeñas empresas, y particulares. Incluyen aplicaciones sencillas para la creación de tablas, consultas e informes.

Por el modelo de BD:

- **Relacionales:** Usan SQL, se basan en el modelo entidad–relación. Datos almacenados lógicamente en tablas.
- **Multidimensionales:** Pueden usar SQL, se basan en el modelo de estrella o copo de nieve. Parten de una tabla de hechos de la que se desprenden distintas tablas de dimensiones.
- **NoSQL:** Permiten almacenar información de forma muy heterogénea (documentales, grafos, clave-valor, orientadas a objetos).
- **Lógicas:** Almacenan reglas que representan procesos deductivos.

Bases de Datos – Tipos

Según la localización de la información:

Bases de Datos Centralizadas.



- La información está almacenada en un solo servidor, que requiere grandes prestaciones.
- La gestión es centralizada, lo que reduce los costes.
- No presentan redundancia y escalan mal si necesitan ser redimensionadas.

Base de Datos Distribuida (BDD) es un conjunto de bases de datos que se encuentran lógicamente relacionadas y se distribuyen en diferentes sitios de forma que necesita una interconexión de red para comunicarse.

- La información está almacenada a lo largo de varios servidores, que no requieren tantas prestaciones como en el caso anterior.
- La gestión es compartida por distintos equipos responsables de los servidores.
- Proporcionan mayor grado de disponibilidad, ya que si falla un servidor, otro puede proporcionar la información. Escalan bien, ya que la estructura distribuida favorece añadir nuevos nodos con facilidad.
- Presentan mayores retos de seguridad, por la interconexión de múltiples.



Bases de Datos – Mejora de rendimiento

Un servidor de base de datos recibe muchas peticiones de lectura y de escritura que pueden sobrecargar el servidor y dificultar el rendimiento del sistema. Para mitigar estos impactos y mejorar el rendimiento, existen diferentes enfoques entre los que destacan:

- La ampliación del servidor de bases de datos
- La replicación de bases de datos
- El particionamiento horizontal

La **ampliación del servidor** es limitada (más capacidad de proceso, más RAM, mejores discos, ...)

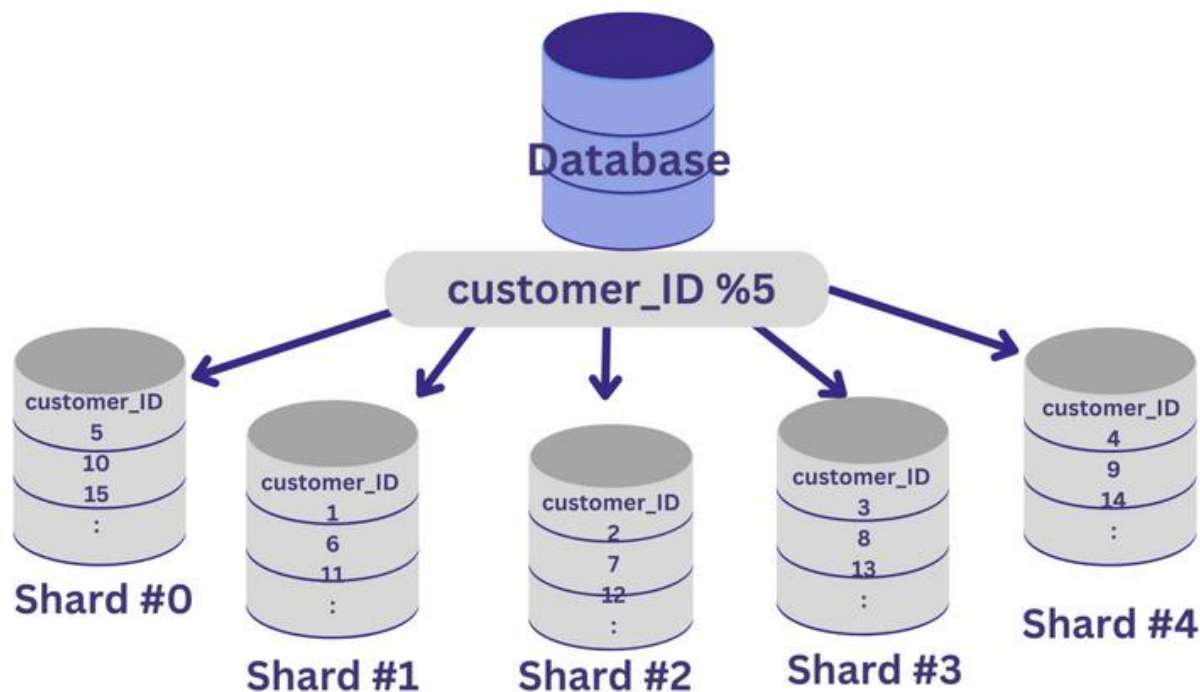
La **replicación de bases de datos** supone disponer de varios nodos cada uno con la base de datos completa “réplicas”. Un nodo ejerce de maestro y suele recibir las peticiones de escritura mientras que existen múltiples réplicas de sólo lectura. Los cambios en el maestro se propagan periódicamente a las réplicas pudiendo ocasionar “lecturas obsoletas”.

El **particionamiento horizontal** permite dividir una tabla única de gran tamaño en varias particiones (o tablas más pequeñas) desde las que podemos leer pero siguen estando en una misma instancia y ubicación.

Bases de Datos – Fragmentación

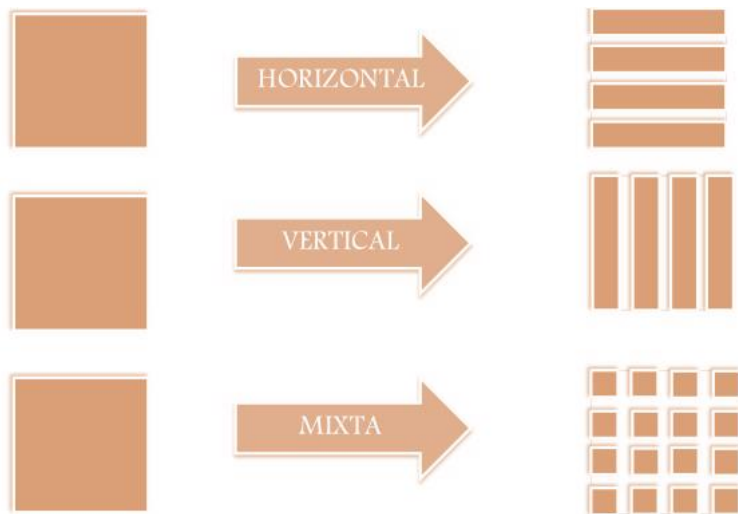
La **fragmentación de bases de datos** persigue mejorar el rendimiento de las mismas dividiendo las bases de datos grandes en varias bases de datos más pequeñas llamadas **fragmentos**, cada una de las cuales contiene un subconjunto único de datos y se ejecuta en un servidor de bases de datos.

La **clave de fragmentación** suele ser una columna o combinación de ellas que determina qué registros van a qué fragmento. (ej. Id de cliente) Es necesario una capa de enrutamiento que dirija las solicitudes entrantes al fragmento oportuno.



Bases de Datos – Fragmentación

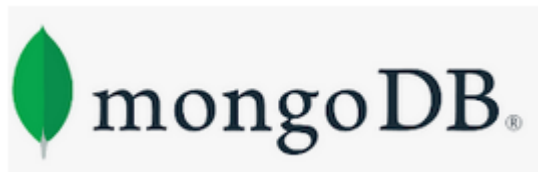
En un modelo de bases de datos distribuido podemos encontrar diferentes técnicas utilizadas para la **fragmentación**:



- **Horizontal:** se basa fundamentalmente en particionar tuplas en subconjuntos. De esta manera, cada subconjunto debe contener aquellos datos que son comunes. Así, cada fragmento, se puede definir como una operación de selección.
- **Vertical:** Se basa en subdividir los atributos en grupos. Todos los fragmentos deben tener la clave y en cada uno de ellos se decide qué atributos adicionales incluir. Así, cada fragmento, se puede definir como una operación de proyección.
- **Mixta:** en este tipo de fragmentación se van a combinar la horizontal con la vertical o viceversa.

SGBDs

Un **Sistema de Gestión de Bases de Datos** (SGBDs) es un programa o software que facilita el diseño, la consulta, uso y actualización de las Bases de Datos (BBDD).



SGBDs - funciones

Las funciones básicas de los SGBDs son:

- **Manipulación de los datos:** Permiten leer, escribir, modificar y borrar datos.
- **Integridad de los datos:** Fija reglas y restricciones para los datos que almacenan.
- **Seguridad:** Establecen quién puede acceder a la información y a qué nivel.
- **Análisis de rendimiento:** proporcionan herramientas de medición.
- **Concurrencia:** Gestionan el acceso en paralelo de múltiples usuarios.
- **Transacción:** Permiten el correcto tratamiento de operaciones compuestas.
- **Conectividad ODBC:** Proporcionan formas de acceso a programas de terceros.
- **Metadatos:** Información sobre los objetos que componen la base de datos.
- **Restauración:** Incluyen herramientas de salvaguarda de la información.

SGBDs - componentes

Los SGBD son paquetes de software muy complejos que deben proporcionar una serie de servicios que van a permitir almacenar y explotar los datos de forma eficiente.

Desde un punto de vista de los elementos que permiten almacenar, recuperar e interactuar la información, podemos hablar de:

- **Tablas**
- **Consultas**
- **Informes**
- **Formularios**

Desde un punto de la funcionalidad que desempeñan en el gestor de la base de datos:

- **Datos**
- **Metadatos**
- **Lenguajes para la gestión de la BD (DDL, DML, DCL, TCL)**
- **Programas de gestión de los datos**



SGBDs comerciales VS libres

Los **Sistemas de Gestión de Bases de Datos** (SGBDs), al igual que cualquier software, pueden ser comerciales o libres.

Los SGBD comerciales son creados por compañías que cobran por su distribución y soporte, y que tienen un uso restringido.

Son de código cerrado, esto significa que los usuarios no pueden acceder al programa en sí mismo para alterar su comportamiento o funcionalidad.

El software pertenece al fabricante, que tan solo proporciona licencias de uso al cliente que lo compra.

Estas licencias pueden ser de carácter temporal o permanente, y pueden estar ligado al hardware o a las usuarios.

Al comprar el software, el fabricante proporciona mantenimiento y garantía de uso.

Ejemplos: Oracle, TeraData, SQL Server, Informix, HANA, DB2



SGBDs comerciales VS libres

Los **SGBD libres** son creados por organizaciones o particulares que no cobran por su explotación o venta.

Pueden ser de código abierto, esto significa que los usuarios, bajo ciertas condiciones, pueden acceder al código fuente de los programas que forman el SGBD para alterarlo o modificar su funcionalidad.

En función de la BD pueden llegar a permitir estas posibilidades:

- Usar el programa independientemente del propósito, si bien no puede ser con fines comerciales.
- Distribuir copias del mismo
- Modificar el programa en ocasiones
- Acceder al código fuente del programa para estudiarlo, modificarlo o extenderlo.

Ejemplos: PostgreSQL, MySQL, MariaDB, MongoDB, CouchDB

Legislación sobre protección de datos

1. LORTAD: Ley Orgánica 5/1992 de 29 de octubre, fue la primera ley de protección de datos que tuvo por finalidad hacer frente a los riesgos que suponía la recogida y el tratamiento de datos por medios informáticos automatizados. Derogada por la LOPD.

2. LOPD: Ley Orgánica 15/1999 de 13 de diciembre, de Protección de Datos de Carácter Personal equiparaba y convertía el derecho a la protección de los datos personales en un derecho fundamental de las personas ampliando el alcance de la LORTAD para medios informáticos automatizados o no. Derogada por la LOPDGDD.

Establecía una serie de obligaciones sobre las empresas que disponían de datos personales contenidos tanto en ficheros automatizados/informatizados, como en no automatizados o en papel con finalidades tales como, la gestión de personal, proveedores, clientes, campañas de marketing, etc.

3. LOPDGDD: Ley Orgánica 3/2018 de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales, que surge como respuesta al Reglamento UE 2016/679 (**GDPR**) para adaptar al nuevo reglamento el ordenamiento jurídico español.

Legislación sobre protección de datos

GDPR o RGPD: El Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo regula el tratamiento que realizan personas, empresas u organizaciones de los datos personales relacionados con personas en la Unión Europea (UE).

Exige que las empresas dispongan de un delegado de protección de datos **DPD** responsable de realizar la supervisión y monitorización, de forma independiente y confidencial, de si se está cumpliendo adecuadamente la normativa en materia de protección de datos personales.

<https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32016R0679>



Leyes y reglamento obligan a las empresas a informar a los usuarios sobre el tratamiento de sus datos personales, **garantizar su confidencialidad y seguridad**, y **obtener su consentimiento** para el tratamiento de los mismos y a notificar a las autoridades competentes en caso de que se produzca una **brecha de seguridad** que afecte a los datos personales de los usuarios.



Legislación: Actividad

¿Qué información consideras que puede ser considerada como datos personales?

Legislación: Datos personales

Datos de carácter personal se considera toda información alfanumérica, gráfica, fotográfica, acústica o de cualquier otro tipo concerniente a personas físicas identificadas o identificables.



Ejemplos:

nombre, email, dni, dirección, teléfono, dirección ip, tarjeta crédito ...

Datos especialmente protegidos (intimidad):

- ideología política
- afiliación sindical
- religión y creencias
- expedientes médicos
- orientación sexual
- historial de infracciones

Legislación: Actividad

Accede a alguna de las siguientes páginas de internet y contesta a las siguientes cuestiones:

- <https://elpais.com>
 - <https://www.elmundo.es>
 - <https://www.sport.es>
-
1. ¿Quién es el responsable del tratamiento de tus datos?
 2. ¿Han asignado un DPD (Delegado de Protección de Datos)? ¿Cómo contactar con él/ella?
 3. ¿Indican las finalidades del tratamiento de tus datos personales? Si es así, indicar al menos 2 (las que te parezcan más llamativas)
 4. ¿Qué tipo de datos son recogidos?
 5. ¿Comparte tus datos con terceros?
 6. ¿Qué procedimiento ofrecen para reclamar el borrado de tus datos?
 7. ¿Durante cuánto tiempo conservan tus datos?

Big Data



Big Data es el conjunto de tecnologías creadas para recopilar, analizar y gestionar los datos que generan los usuarios de Internet. Su idea es la de recopilar los datos masivos que son generados en "bruto", y procesarlos para identificar patrones u otro tipo de comportamientos que puedan ayudar a sectores concretos.

Todas nuestras acciones ya sea en internet o haciendo uso de servicios y aplicaciones dejan huella, y aunque sean datos anonimizados y que no nos identifican como persona, sí representan lo que hace un usuario.

Cuando son millones los usuarios, el Big Data es capaz de aprovechar estas grandes cantidades de datos para poder analizarlos y procesarlos intentando beneficiarse de ellos.

Ejemplos:

- Si de repente hay una cantidad creciente de usuarios que empieza a mirar determinado tipo de contenido concreto, puede ser una nueva tendencia y se pueden adelantar a ella y beneficiarse antes de que se convierta en popular.
- Un servicio de streaming puede saber el contenido que más interés despierta en términos generales para generar o sugerir más contenido similar.

Business Intelligence

El **Business Intelligence**, BI o la inteligencia de negocios combina análisis de negocios, minería, visualización, herramientas e infraestructura de datos, además de prácticas recomendadas para ayudar a las empresas a tomar decisiones basadas en los datos que maneja.

Consiste en usar estos datos para impulsar el cambio, eliminar las ineficiencias y adaptarse rápidamente a los cambios del mercado o la demanda.

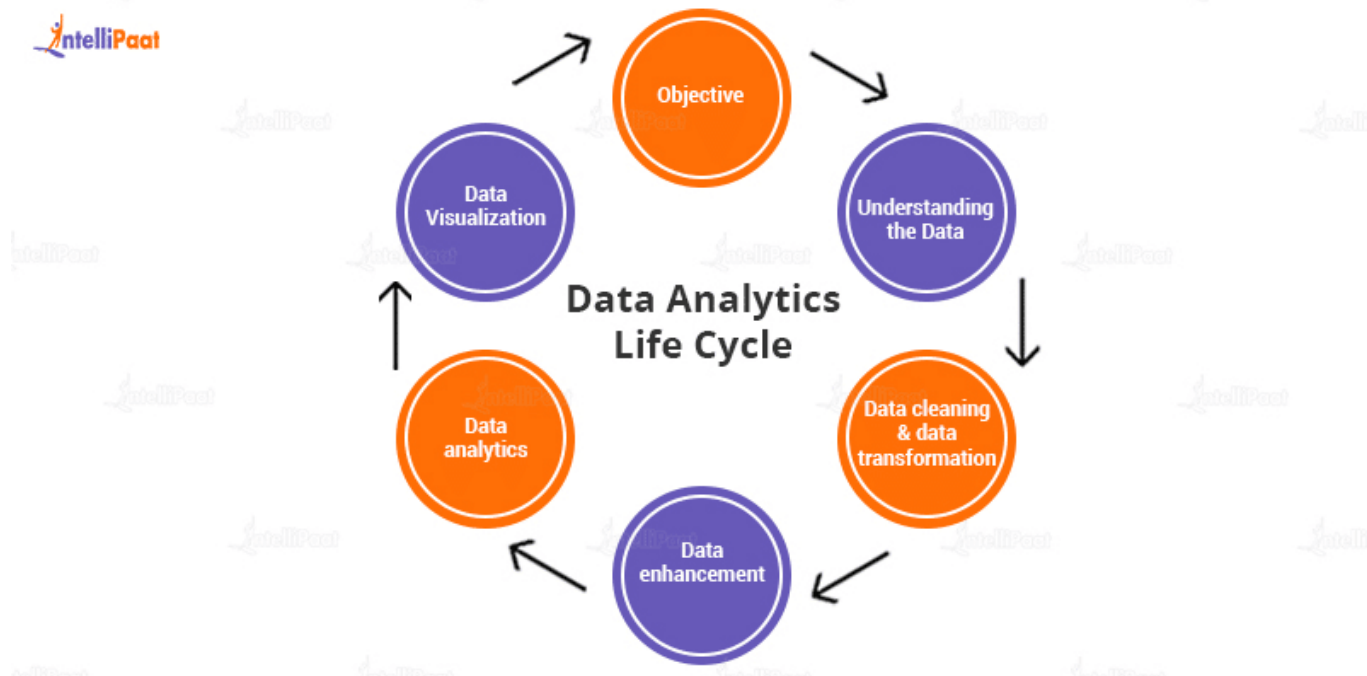
La inteligencia de negocios permite a las organizaciones tomar mejores decisiones, porque ofrece datos actuales e históricos dentro de su contexto empresarial.



Analisis de datos

El **análisis de datos** convierte datos en bruto y no estructurados en un formato claramente entendible por el usuario. El modelado, la transformación y la limpieza de datos, son las principales tendencias de la implementación de análisis de datos dentro de una organización con el objetivo de resaltar información útil, para sugerir conclusiones y apoyo en la toma de decisiones.

El análisis de datos convierte datos sin procesar en información práctica.



Fuente: <https://intellipaat.com/>

Big Data: Actividad

Consulta en internet información sobre algún caso de éxito del uso de Big Data por parte de las empresas.



Algunos casos de éxito

- *Amazon* emplea Big Data para la recomendación de productos personalizados siendo el 35% de sus ventas anuales dichas recomendaciones de productos.
- *Netflix* emplea Big Data recopilando los gustos y hábitos de sus clientes consiguiendo así una tasa de retención de clientes del 95%.
- *Tesla* diseña sus vehículos autónomos a partir de la recopilación y análisis de los datos de sus coches, rutas y conductores.
- *Starbucks* emplea Big Data para ubicar estratégicamente sus tiendas a partir de información geográfica, demográfica o el tráfico.
- *Facebook* emplea Big Data para conseguir que la publicidad y productos que nos ofrecen sea lo más cercana a nuestros gustos.
- *Spotify* predice con éxito los ganadores de los premios Grammy analizando los gustos musicales y preferencias de sus millones de clientes.

Y así un sinfín de empresas...