



Sistemas de almacenamiento de la información

Francisco Javier González Brime



Sistemas de almacenamiento de la información

1

Información: conjunto organizado de datos procesados que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que recibe dicho mensaje.

La información se puede almacenar en dos estructuras de datos: ficheros y sistemas de bases de datos.

Ficheros

Estructuras de datos creadas por los S.O. para almacenar datos. Constan de nombre y extensión.

Cualquier usuario puede tener acceso a los ficheros y realizar operaciones sobre ellos. Las operaciones más comunes son:

- Creación
- Consulta
- Mantenimiento
- Borrado

Sistemas de almacenamiento de la información

2

Clasificaciones

Pueden ser legibles por el usuario o ilegibles.

Legibles: de texto plano (ASCII).

- Configuración
- Código fuente
- Páginas web
- Etc

Ilegibles: binarios (contienen información codificada)

- Imagen
- Video
- Sonido
- Comprimidos
- Ejecutables
- Etc

Sistemas de almacenamiento de la información

3

Clasificación:

Según su organización:

- Secuencial.
- De acceso directo o aleatorio.
- Indexados.

Según su utilidad:

- Maestros o principales.
- De movimientos.

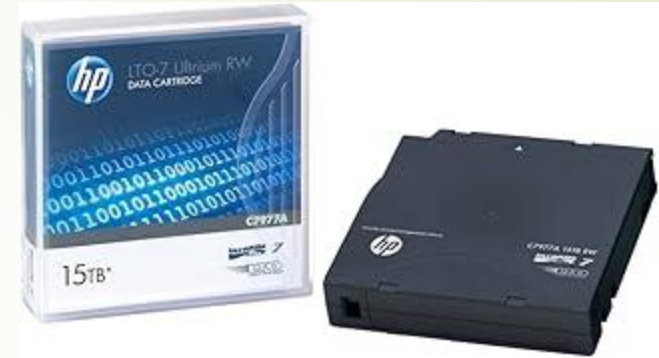
Sistemas de almacenamiento de la información

4

Tipos de soporte de almacenamiento:

Secuencial.

- Para acceder a un dato hay que recorrer todo el contenido del soporte previo a dicho dato.
- Ejemplo: cintas magnéticas.



Directo

- Se puede acceder directamente a un dato sin tener que recorrer todos los anteriores.
- Ejemplo: disco duro



Sistemas de almacenamiento de la información

5

Sistemas de Bases de Datos.

Inicialmente las primeras empresas y organizaciones trabajaban con sistemas de ficheros.

Cuando se incrementó la cantidad de información surgieron problemas:

- Integridad
- Duplicidad de información
- Seguridad
- Accesos concurrentes
- ...

Las empresas y organizaciones comenzaron a organizar la información mediante un sistema más ordenado basado en la centralización de la gestión y la organización de los datos en forma de **bases de datos**.

Sistemas de almacenamiento de la información

6

Sistemas de Bases de Datos.

Principales problemas relacionados con el uso de ficheros:

- Separación y aislamiento de los datos
- Duplicación de datos
- Dependencia de datos
- Formatos de ficheros incompatibles
- Consultas fijas y proliferación de programas de aplicación
- Control de concurrencia
- Autorizaciones
- Catálogo



Sistemas de almacenamiento de la información

7

Sistemas de Bases de Datos.

Una **base de datos** es un conjunto de datos almacenados entre los que existen relaciones lógicas y ha sido diseñada para satisfacer los requerimientos de información de una empresa u organización.

- La base de datos se utiliza al mismo tiempo por muchos equipos y usuarios.
- Los datos están centralizados y organizados (se minimiza la redundancia y se facilita la gestión).
- La base de datos se comparte por toda la organización.
- La base de datos contiene los datos de la organización y una descripción de los mismos (metadatos)

Sistemas de almacenamiento de la información

8

Arquitectura de sistemas de Bases de Datos.

El comité **ANSI-SPARC** propuso en 1975 un estándar para la creación de sistemas de bases de datos basado en una **arquitectura de tres niveles**, con el objetivo de separar en niveles de abstracción el esquema de una base de datos.

Independencia lógica

Nivel externo o vista de usuario

Independencia física

Nivel lógico o global

Nivel interno o físico

Sistemas de almacenamiento de la información

9

Arquitectura de sistemas de Bases de Datos.

Tres formas de ver o representar una misma base de datos:

- **Nivel externo.**

Esquemas externos o **vistas de usuario**. Cada esquema describe la parte de la BD que interesa a un grupo de usuarios, ocultando el resto de la BD. Para especificar los esquemas se puede utilizar un modelo conceptual o un modelo lógico.

- **Nivel global.** (¿Qué?)

Describe la estructura de toda la BD para una comunidad de usuarios mediante un **esquema conceptual**. Oculta detalles de almacenamiento y describe entidades, atributos, relaciones y restricciones.

- **Nivel interno.** (¿Cómo?)

Describe la **estructura física** de la BD, describiendo todos los detalles para el almacenamiento de la BD, así como los métodos de acceso. Habla de ficheros, discos, directorios, etc.

Sistemas de almacenamiento de la información

10

Sistemas de Bases de Datos.

Esta arquitectura a tres niveles permite explicar el concepto de independencia de datos:

- **Independencia lógica.** Capacidad de modificar el esquema conceptual sin tener que alterar los esquemas externos ni programas de aplicación.
- **Independencia física.** Capacidad de modificar el esquema interno sin alterar el esquema conceptual.

Sistemas de almacenamiento de la información

11

Modelos de datos.

La base de datos consiste en los datos concretos referentes a un sistema o parte del mundo que hemos modelado.

Esta tarea es sencilla siempre que sean pocos datos. Cuando el volumen crece, se requiere el uso de distintos modelos para facilitar el diseño de la base de datos.

Un **modelo de datos** es una colección de herramientas conceptuales para describir los datos, las relaciones que existen entre ellos y sus restricciones.

El modelo proporciona mecanismos de abstracción para representar una parte del mundo cuyos datos nos interesan. La representación recibe el nombre de **esquema** y el conjunto de datos que representa es la **base de datos**.

Sistemas de almacenamiento de la información

12

Modelos de datos.

La base de datos consiste en los datos concretos referentes a un sistema o parte del mundo que hemos modelado.

- **Conceptuales**



Modelo Entidad-Relación

- **Lógicos tradicionales**

- ❖ Relacional



Modelo Relacional

- ❖ Modelo en Red

- ❖ Modelo Jerárquico

- **Lógicos avanzados**

- ❖ Orientados a Objetos

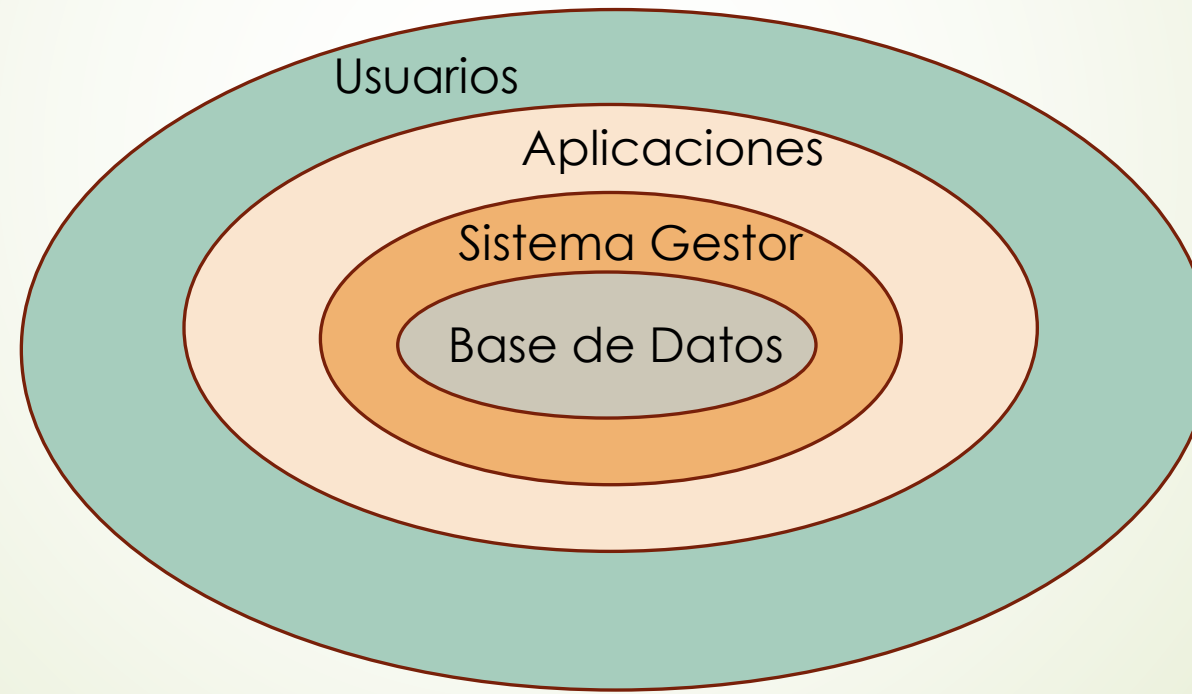
- ❖ Modelos de datos declarativos

Sistemas de almacenamiento de la información

13

Sistemas Gestores de Bases de Datos.

Un **sistema de gestión de la base de datos** (SGBD) es una aplicación que permite a los usuarios definir, crear y mantener la base de datos, y proporciona acceso controlado a la misma. Es una herramienta que sirve de interfaz entre el usuario y las bases de datos.

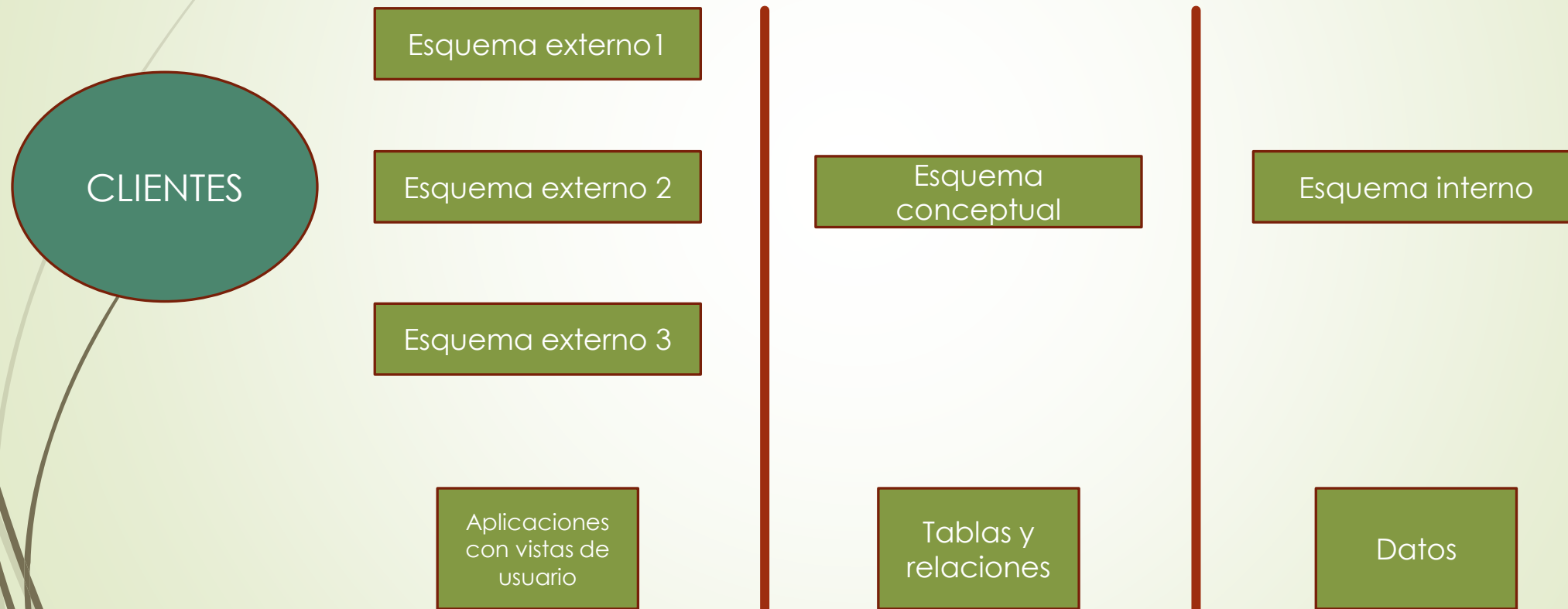


Sistemas de almacenamiento de la información

14

Sistemas Gestores de Bases de Datos.

ANSI estableció los tres niveles de abstracción (externo, lógico o global y físico) como requisito en los SGBD:



Sistemas de almacenamiento de la información

15

Sistemas Gestores de Bases de Datos.

Objetivos:

- Asegurar los tres niveles de abstracción: físico, lógico y externo
- Permitir independencia física y lógica de los datos
- Gestionar la consistencia de los datos
- Ofrecer seguridad de acceso a los datos por parte de usuarios y grupos
- Gestión de transacciones
- Permitir la concurrencia de usuarios sobre los mismos datos mediante bloqueos que mantienen la integridad de los mismos

Sistemas de almacenamiento de la información

16

Sistemas Gestores de Bases de Datos.

Funciones:

- Incorporar catálogo donde se almacenen las descripciones de los datos y sea accesible por los usuarios (diccionario de datos)
- Garantizar la integridad (todas las operaciones de una transacción se realizan o no se realiza ninguna). Transacción: conjunto de acciones.
- Concurrencia de accesos
- Recuperación de datos (vuelta un estado consistente tras un fallo)
- Integración con algún software de comunicación
- Cumplimiento de restricciones
- Herramientas de administración

Sistemas de almacenamiento de la información

17

Sistemas Gestores de Bases de Datos.

Componentes:

- Lenguajes de datos
 - Lenguaje de definición de datos (DDL, Data Definition Language)
Define y describe los objetos de la base de datos.
 - Lenguaje de control de datos (DCL, Data Control Language)
Encargado del control y seguridad de los datos.
 - Lenguaje de manipulación de datos (DML, Data Manipulation Language)
Inserción, actualización, eliminación y consulta.
- Diccionario de datos
- Objetos: Tablas, vistas, funciones, procedimientos, disparadores o triggers.
- Herramientas para facilitar el trabajo con datos: seguridad, integridad, control, etc
- Optimizador de consultas
- Gestión de transacciones
- Planificador (scheduler). Automatiza procesos.
- Copias de seguridad

Sistemas Gestores de Bases de Datos.

Usuarios:

- **Usuarios finales**

Nivel externo (vistas o porciones de la base de datos)

- **Diseñadores**

Diseño lógico de la base de datos.

- **Programadores**

De aplicaciones, de objetos de la base de datos, de procedimientos almacenados o disparadores.

- **Administradores**

Nivel de abstracción físico.

Sistemas de almacenamiento de la información

19

Sistemas Gestores de Bases de Datos.

Tipos de SGBD:

- Según el modelo lógico en el que se basan
 - Jerárquico
 - En red
 - **Relacional**
 - Objeto-relacional
 - Orientado a objetos
- Número de usuarios
 - Monousuario
 - Multiusuario
- Número de sitios
 - Centralizado
 - Distribuido
- Ámbito de aplicación
 - Propósito general
 - Propósito específico
- Tipos de datos
 - Sistemas relacionales estándar
 - XML
 - Objeto-relacionales
 - Orientados a objetos
- Lenguajes soportados
 - **SQL** estándar
 - NoSQL o nuevo lenguaje de consulta

Sistemas de almacenamiento de la información

20

Sistemas Gestores de Bases de Datos.

Sistemas libres y comerciales

LIBRES

Disponibles de forma abierta en la red (no necesariamente gratuitos) y modificables según licencia GPL

- MySQL (Oracle)
- MariaDB
- PostgreSQL
- Firebird
- SQLite

NoSQL

- MongoDB
- Apache Cassandra
- CouchDB
- DynamoDB (Amazon)

COMERCIALES

Requieren pago de licencias, aunque suelen ofrecer versiones gratuitas limitadas

- ORACLE
- MySQL (Oracle)
- SQL server (Microsoft)
- DB2 (IBM)
- Microsoft Access (Microsoft Office 365)
- ...

Sistemas de almacenamiento de la información

21

Bases de datos centralizadas y distribuidas

Bases de datos centralizadas:

- Software, datos y soportes físicos residen en un único lugar físico.
- Los clientes (aplicaciones, usuarios) acceden al sistema a través de interfaces que se conectan al servidor.

Bases de datos distribuidas:

- Los componentes se distribuyen en distintos computadores comunicados a través de una red.
- Se conocen como Sistemas Gestores de Bases de Datos Distribuidas (SGBDD)

Base de datos distribuida: colección de datos que pertenece lógicamente al mismo sistema pero que se encuentra físicamente almacenada en distintas máquinas conectadas por una red.

Sistemas de almacenamiento de la información

22

BASES DE DATOS CENTRALIZADAS

Software para una arquitectura centralizada.

Software de gestión de datos. Reside en servidor.

Software de interacción con el usuario. Reside en cliente.

Software de desarrollo. Reside en cliente.

Ventajas:

Se evita la redundancia y la inconsistencia de datos.

Pueden aplicarse restricciones de seguridad.

Conservan la integridad.



BASES DE DATOS DISTRIBUIDAS

El diseño implica tomar decisiones sobre la ubicación del software de acceso a los datos, así como los propios datos.

Ventajas:

Mejoran eficacia al compartir y acceder a la información.

Mejoran rendimiento.

Cada servidor tiene cierto grado de control sobre sus datos.

Si se produce fallo en un nodo, el resto podría seguir funcionando.

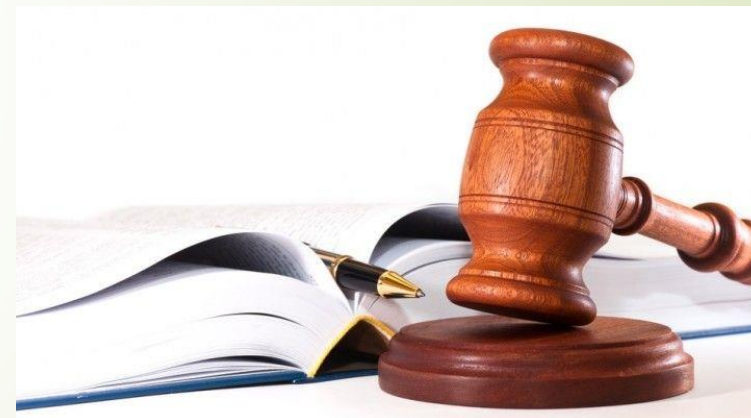


Legislación sobre protección de datos



Legislación sobre protección de datos

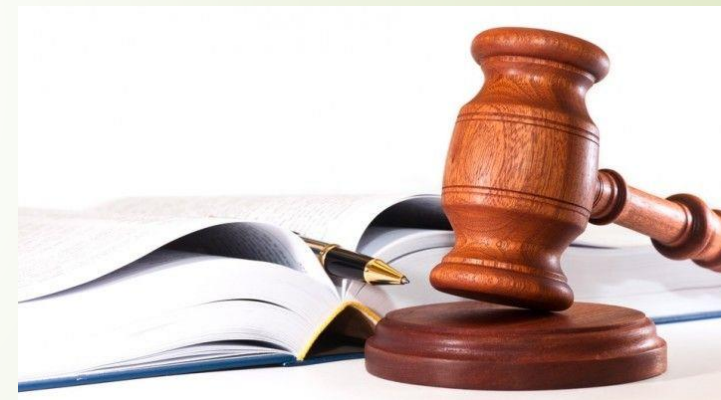
- Artículo 18º 4. Constitución española del año 1978. “La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el ejercicio de sus derechos”
- Ley Orgánica 5/1992. Ley de Regulación del Tratamiento Automatizado de Datos de Carácter Personal (LORTAD)
- Directiva 95/46/CE del Parlamento Europeo y del Consejo.





Legislación sobre protección de datos

- Ley Orgánica 15/1999. Ley de Protección de Datos de Carácter Personal (LOPD)
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo. Reglamento General de Protección de Datos (RGPD). Comenzó a aplicarse el 25 de mayo de 2018.
- Ley Orgánica 3/2018. Ley de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD).





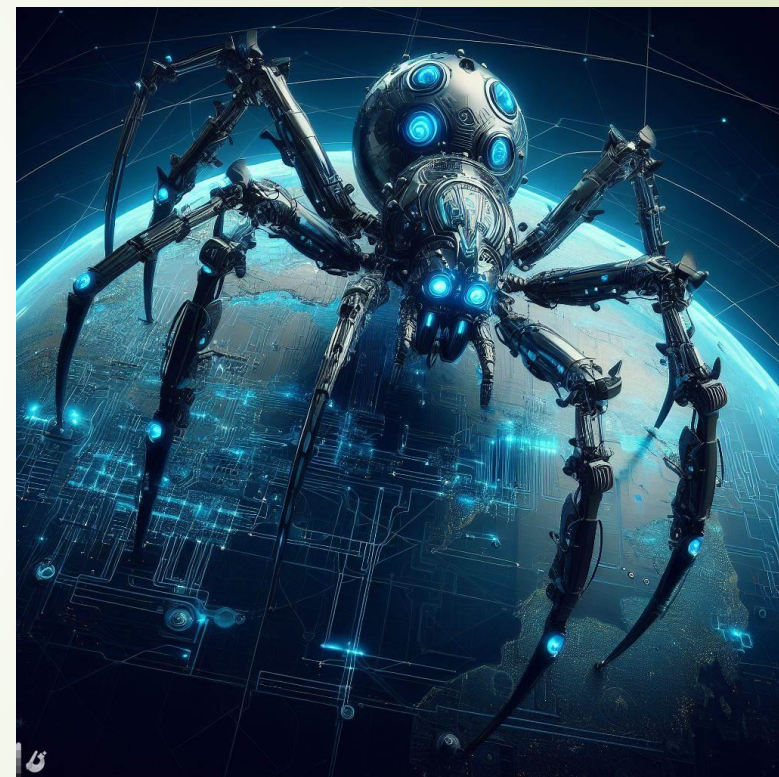
Big Data



Big Data es el término que se usa para describir el conjunto de datos que se generan en Internet con el objetivo de identificar patrones, comportamientos, tendencias o soluciones, y que son tan grandes, variados y rápidos que requieren de tecnologías especiales para su análisis y gestión.

Estos conjuntos de datos pueden provenir de una variedad de fuentes, como las transacciones comerciales, las redes sociales, los sensores y los dispositivos móviles.

El big data puede tener muchas aplicaciones en diferentes sectores como la salud, la banca, la geolocalización y el desarrollo de negocio.





El Big Data se caracteriza por las siguientes características:

- **Volumen:** los conjuntos de datos de big data son muy grandes. Pueden tener millones, miles de millones o incluso billones de registros. Esto podría ser terabytes, petabytes o incluso exabytes de información, a menudo generada en tiempo real.
- **Velocidad:** los datos se generan a una velocidad sin precedentes. Publicaciones en redes sociales, datos de sensores y otras fuentes en tiempo real contribuyen al flujo constante de datos que deben procesarse rápidamente.
- **Variedad:** El big data viene en diversos formatos. Incluye datos estructurados (como bases de datos), datos no estructurados (como texto) y datos semi-estructurados (como archivos XML). La diversidad de tipos de datos agrega complejidad al procesamiento y análisis.



- **Veracidad** se refiere al grado de fiabilidad y calidad de los datos, es decir, si están contrastados y completos. Los conjuntos de datos pueden ser inexactos o incompletos.
- **Valor** se refiere a la información útil que se puede extraer de los datos para tomar mejores decisiones, identificar tendencias y desarrollar nuevos productos y servicios.





El Big Data funciona mediante tres pasos:

- Captación de información de diversas fuentes.
- Almacenamiento y gestión en sistemas seguros y eficientes.
- Análisis mediante algoritmos y programas.

Algunas de las herramientas más usadas para son: Apache Hadoop, Apache Spark, MongoDB, R o Python.



Consideraciones a tener en cuenta sobre Big Data:

- **Almacenamiento distribuido:** es necesario almacenarlos de manera distribuida en sistemas de almacenamiento como Apache Hadoop Distributed File System (HDFS) o Amazon S3.
- **Procesamiento distribuido:** se utilizan frameworks de procesamiento distribuido como Apache Hadoop y Apache Spark. Estos permiten realizar operaciones paralelizadas en clústeres de computadoras.
- **Machine Learning y Big Data:** El análisis de grandes conjuntos de datos puede revelar patrones y tendencias que son difíciles de descubrir con enfoques tradicionales.
- **Seguridad y Privacidad:** Con grandes cantidades de datos, existe un riesgo aumentado de brechas de seguridad y es importante implementar medidas adecuadas para proteger la información.
- **Ecosistema de herramientas:** El ecosistema de Big Data incluye una variedad de herramientas y tecnologías, como bases de datos NoSQL, sistemas de procesamiento de datos en tiempo real (por ejemplo, Apache Kafka), herramientas de visualización de datos , ...



- **Escala horizontal:** el enfoque de Big Data a menudo implica escalar horizontalmente, es decir, agregar más nodos o servidores al sistema para manejar la carga.
- **Gestión de datos no estructurados:** Dado que el Big Data incluye datos en una variedad de formatos, la gestión de datos no estructurados se vuelve crucial. Esto implica trabajar con datos como texto, imágenes, audio, video, etc.
- **Ética en el uso de datos:** Con el aumento de la cantidad de datos, surge la necesidad de considerar cuestiones éticas en su recopilación, uso y análisis. Esto incluye preocupaciones sobre la privacidad, el consentimiento y la equidad.
- **Costos:** Gestionar grandes cantidades de datos puede ser costoso en términos de almacenamiento, procesamiento y ancho de banda.
- **Internet of Things (IoT):** El aumento de dispositivos conectados a Internet genera grandes cantidades de datos en tiempo real, lo que contribuye significativamente al panorama del Big Data.

FIN