

1. Datos

Los **datos** son hechos brutos, sin procesar y sin interpretar, que representan eventos o valores específicos. Por sí mismos, los datos suelen carecer de un contexto o significado directo. Pueden ser números, texto, imágenes, sonidos, etc.

2. Información

La **Información** es el resultado de procesar, organizar, estructurar y contextualizar los datos. Cuando los datos se presentan en un contexto significativo y útil, se convierten en información, lo que permite la toma de decisiones y la comprensión.

3. Diferencia entre Datos e Información

Característica	Datos	Información
Naturaleza	Crudos, sin procesar, hechos aislados.	Procesada, organizada, estructurada.
Significado	Mínimo o nulo por sí solo.	Tiene significado y contexto, es útil.
Uso	Materia prima para el procesamiento.	Base para la toma de decisiones y el conocimiento.
Forma	Números, texto, símbolos.	Informes, gráficos, documentos.

4. Base de Datos (Database)

Una **Base de Datos** es una colección organizada y estructurada de datos relacionados que se almacena y se accede de forma electrónica. Permite la gestión eficiente de grandes volúmenes de datos, incluyendo su almacenamiento, recuperación, modificación y eliminación.

5. Tipos de Base de Datos

Los tipos principales se dividen en dos grandes categorías, basadas en cómo estructuran la información:

Tipo	Definición	Estructura
Relacionales (SQL)	Se basan en el modelo relacional, donde los datos se organizan en tablas (relaciones) con filas y columnas, y las tablas se conectan mediante claves (relaciones).	Tabular, con un esquema estricto predefinido.
No Relacionales (NoSQL)	Utilizan diversos modelos de datos alternativos (documentos, clave-valor, grafo) que no requieren un esquema fijo. Diseñadas para manejar grandes volúmenes de datos de alta velocidad y tipos de datos variados.	Flexible (sin esquema fijo), en formatos como JSON, pares clave-valor, etc.

6. Características de las Bases de Datos Relacionales

Las bases de datos relacionales se rigen por el modelo matemático y utilizan **SQL** (Structured Query Language) para su gestión.

- **Estructura Tabular:** Los datos se organizan en **tablas** compuestas por **registros** (filas) y **campos** (columnas).
- **Esquema Estricto:** Requieren que la estructura de la base de datos (tablas, campos y tipos de datos) se defina rigurosamente **antes** de insertar los datos.
- **Integridad de Datos:** Garantizan la **precisión** y la **consistencia** de los datos mediante el uso de **claves primarias** (identificación única) y **claves foráneas** (mantener relaciones).
- **Atomicidad, Consistencia, Aislamiento y Durabilidad (ACID):** Siguen estas propiedades para asegurar que las transacciones sean procesadas de manera fiable.

7. Características de las Bases de Datos No Relacionales (NoSQL)

Las bases NoSQL son más flexibles y están optimizadas para escenarios específicos de gran escala o datos no estructurados.

- **Esquema Dinámico o Flexible:** No requieren un esquema predefinido; puedes añadir datos de diferentes estructuras a la misma colección o almacén.

- **Escalabilidad Horizontal:** Están diseñadas para distribuir la carga de trabajo entre múltiples servidores (sharding) fácilmente, manejando grandes volúmenes de tráfico y datos (**Big Data**).
 - **Modelos Variados:** Incluyen modelos de **Documentos** (ej. MongoDB), **Clave-Valor** (ej. Redis), **Columna Ancha** (ej. Cassandra) y **Grafo** (ej. Neo4j).
 - **Disponibilidad y Tolerancia a Particiones (BASE):** A menudo priorizan la **disponibilidad** y la tolerancia a fallos sobre la consistencia estricta, siguiendo el modelo **BASE** (Básicamente Disponible, Estado Suave, Consistencia Eventual).
-

8. Gestores de Base de Datos Relacionales (RDBMS)

Son el software que permite a los usuarios interactuar con las bases de datos relacionales, definiendo, manipulando y administrando los datos.

- **Populares:**
 - **Oracle Database**
 - **Microsoft SQL Server**
 - **MySQL** (Open Source)
 - **PostgreSQL** (Open Source, robusto)
 - **SQLite** (Ligero, basado en archivo)
-

9. Gestores de Base de Datos No Relacionales (NoSQL)

Software diseñado para almacenar y gestionar datos en formatos distintos al tabular, adaptándose a estructuras flexibles.

- **Populares:**
 - **MongoDB** (Documentos)
 - **Cassandra** (Columna Ancha)
 - **Redis** (Clave-Valor, en memoria)
 - **Neo4j** (Grafo)
 - **Couchbase** (Documentos/Clave-Valor)
-

10. Importancia de las Bases de Datos en Big Data y/o Ciencia de Datos

Las bases de datos son el **fundamento** de las disciplinas de Big Data y Ciencia de Datos, ya que:

- **Almacenamiento y Acceso:** Proporcionan la infraestructura para almacenar los **grandes volúmenes** de datos (**Big Data**) que se van a analizar. Las bases de datos **NoSQL** son cruciales aquí por su escalabilidad horizontal y manejo de datos no estructurados.

- **Organización:** Permiten organizar los datos de manera que sean **consultables** y **recuperables** de forma eficiente (mediante SQL o lenguajes de consulta NoSQL).
- **Integración:** Facilitan la **combinación** de datos de diversas fuentes, creando un conjunto de datos unificado y limpio, esencial para el análisis en **Ciencia de Datos**.
- **Análisis:** Los científicos de datos utilizan consultas complejas (tanto en bases relacionales como NoSQL) para **extraer subconjuntos** de datos, **aplicar modelos** y **obtener *insights*** (información de valor).