

# Ciencia de Datos

- Módulo 1

¿Qué son las bases de datos?  
¿Cual es la diferencia entre SQL y NoSQL?



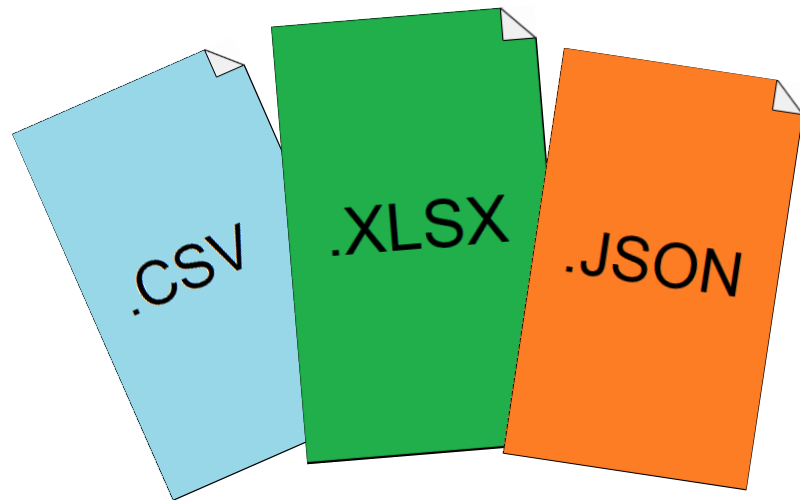
## ¿Qué son las bases de datos?

Existen varios tipos de archivos de los que podemos leer o escribir datos de forma sistematizada.

Los más comunes son:

- CSV
- XLSX
- JSON
- XML

Estos archivos guardan datos, pero no son bases de datos.



## ¿Qué son las bases de datos?

Una base de datos es un conjunto de datos almacenado sistemáticamente para su posterior análisis.

Cuando nos referimos a una base de datos, hablamos de una colección de datos relacionados y de cómo están organizados, y de cómo accedemos a ellos.



Las bases de datos son representadas con este símbolo.



## ¿Qué son las bases de datos?

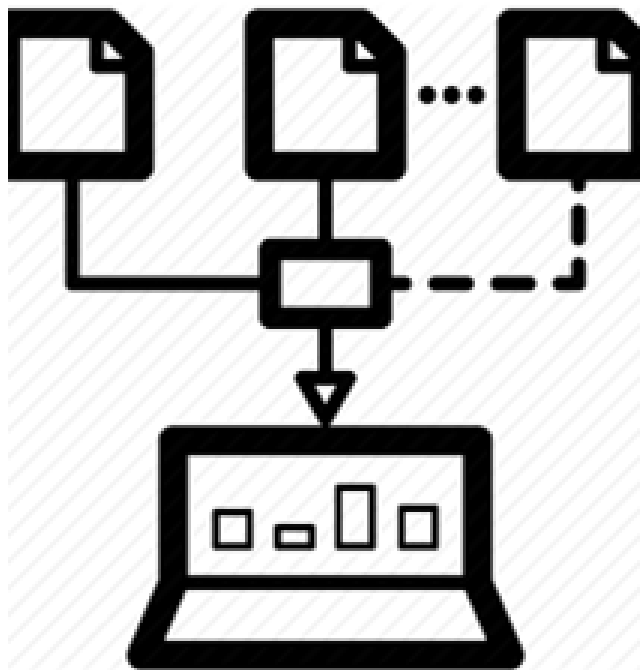
Generalmente se confunde el término “base de datos” con los DBMS (DataBase Management System), estos últimos son el software usado para que los usuarios puedan interactuar con la base de datos. Ejemplos de DBMS: PostgreSQL, MongoDB, MySQL, Oracle, Google BigQuery, etc.



## ¿Qué son las bases de datos?

Las bases de datos:

- Tiene una fuente de la cual derivan los datos.
- Se diseña, construye y puebla con datos para un propósito específico.
- Está dirigida a un grupo de usuarios.



## Bases de datos relacionales o SQL

Son aquellas que mantienen estructuras de “tablas”, es decir, poseen filas y columnas, y las tablas se relacionan entre sí. Se les llaman SQL (lenguaje de consulta estructurado) pues se usa este lenguaje para manipularlas.

PostgreSQL, MySQL, Oracle, Microsoft SQL Server, son algunos ejemplos de estas.

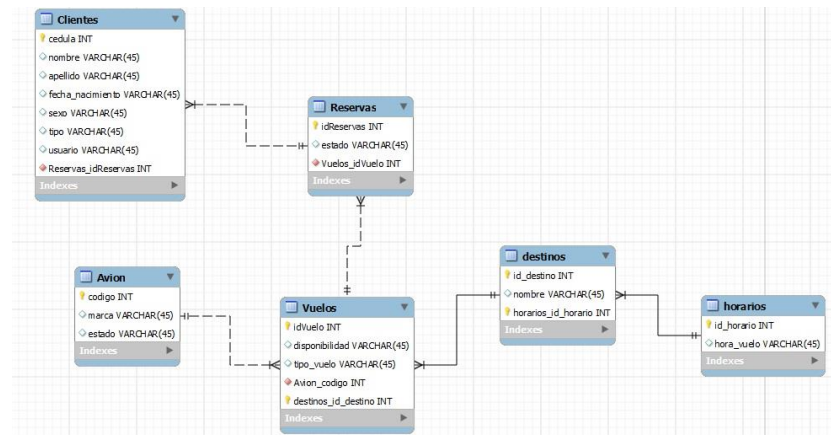


Diagrama de entidad relación, usados para diseñar las bases de datos relacionales, muestra todas las tablas y cómo se relacionan entre sí.

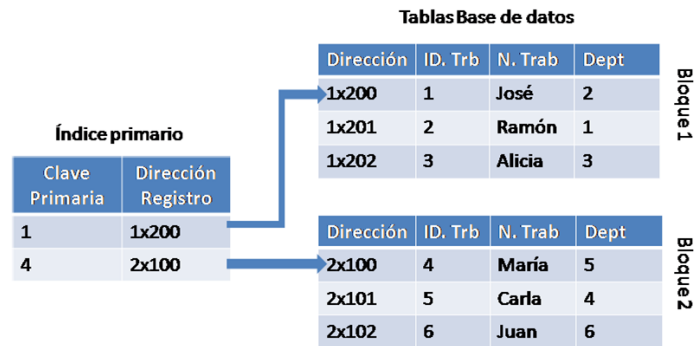


# Clave primaria en bases de datos relacionales

En el diseño de bases de datos relacionales, se llama clave primaria o clave principal a un campo o a una combinación de campos que identifica de forma única a cada fila de una tabla.

Esta clave también se utiliza para relacionar diferentes tablas entre sí, cuando existe un dato en común entre las mismas.

Ejemplos de claves primarias son DNI (asociado a una persona) o ISBN (asociado a un libro).



## Bases de datos no relacionales o NoSQL

Son aquellas que se basan en estructuras llave-valor, documentos u alguna forma de almacenamiento no relacional, suelen ser útiles para guardar todo tipo de información como imágenes, sonidos, videos, texto, etc. MongoDB, Cassandra, Neo4j, Redis, etc. son algunos ejemplos.



Estas bases son las comúnmente usadas para los llamados Data lakes (lagos de datos), y dan origen a los análisis de Big Data.

Fuente:

[https://miro.medium.com/max/605/1\\*NNZshnFMVVpkSbFSc4kFNw.jpeg](https://miro.medium.com/max/605/1*NNZshnFMVVpkSbFSc4kFNw.jpeg)



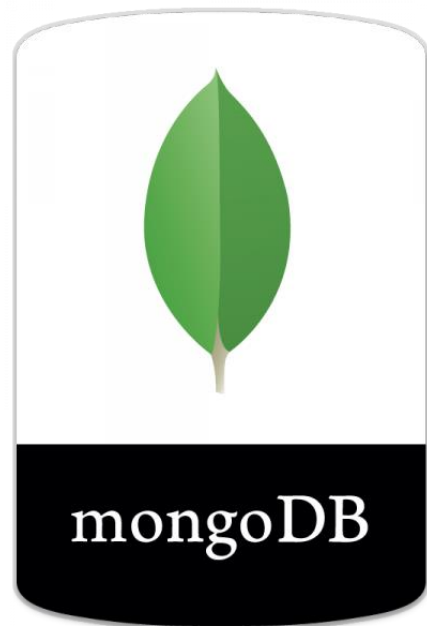


# MongoDB

Dentro de las bases de datos NoSQL, probablemente una de las más famosas sea MongoDB.

MongoDB es una base de datos orientada a documentos. Esto quiere decir que en lugar de guardar los datos en registros, guarda los datos en documentos. Estos documentos son almacenados en BSON, que es una representación binaria de [JSON](#).

Una de las diferencias más importantes con respecto a las bases de datos relacionales, es que no es necesario seguir un esquema. Los documentos de una misma colección (concepto similar a una tabla de una base de datos relacional), pueden tener esquemas diferentes.



## MongoDB - Misma colección, esquemas diferentes

Imaginemos que tenemos una colección a la que llamamos Personas. En la misma colección podemos tener documentos esquemas diferentes, uno con menos campos que el otro, o incluso campos diferentes. Esto que es algo impensable en una base de datos relacional, es algo totalmente válido en MongoDB. Esta flexibilidad hace ideal este tipo de bases de datos para guardar grandes cantidades de datos, sin la necesidad de normalizarlos a un solo esquema.

