



**Escuela Superior  
de Ingeniería y Tecnología**  
Universidad de La Laguna

# Administración y Diseño de Bases de Datos:

Estado del Arte en Gestión y Almacenamiento  
de Big Data

Jaime Pablo Pérez Moro ([alu0101278919@ull.edu.es](mailto:alu0101278919@ull.edu.es))  
Sheyla Ruiz-Gómez Ferreira ([alu0101124445@ull.edu.es](mailto:alu0101124445@ull.edu.es))  
Saúl Pérez García ([alu0101129785@ull.edu.es](mailto:alu0101129785@ull.edu.es))  
Andrea Calero Caro ([alu0101202952@ull.edu.es](mailto:alu0101202952@ull.edu.es))  
Alejandro Martín de León ([alu0101015941@ull.edu.es](mailto:alu0101015941@ull.edu.es))  
Anabel Díaz Labrador ([alu0101206011@ull.edu.es](mailto:alu0101206011@ull.edu.es))



## Índice:

<b>1. Introducción.</b>	<b>1</b>
<b>2. Historia</b>	<b>1</b>
<b>3. Sistemas gestores de bases de datos</b>	<b>2</b>
3.1. Tipos de sistemas gestores de bases de datos.	2
3.1.1. Open Source	2
3.1.2. Propietario	3
3.2. Funciones del sistema gestor de bases de datos.	4
<b>4. Arquitectura de los sistemas gestores.</b>	<b>4</b>
4.1. Arquitectura ANSI-SPARC.	4
4.2. Componentes de un SGBD.	4
<b>6. Fabricantes de bases de datos.</b>	<b>5</b>
<b>7. Big Data</b>	<b>6</b>
7.1. Big Data y Sistemas Gestores de bases de datos	6
7.2. Big Data y videojuegos	7
<b>8. Data Science.</b>	<b>7</b>
<b>9. Modos de comercialización.</b>	<b>8</b>
<b>10. Conclusión.</b>	<b>8</b>
<b>11. Referencias.</b>	<b>10</b>



# 1. Introducción.

Las **bases de datos** son una herramienta imprescindible en la sociedad de la información en la que vivimos actualmente. Estas, podrían definirse como conjunto de datos almacenados accesibles y manipulables, las cuales ayudan en el proceso de toma de decisiones, así como en la puesta en marcha de diferentes acciones pertenecientes a una gran variedad de ámbitos.

## 2. Historia

El concepto de bases de datos se empezó a forjar en 1884 cuando se buscaba una solución para resolver la problemática que surgía a la hora de elaborar de forma manual los censos de población. Herman Hollerith creó una **máquina tabuladora automática** de tarjetas perforadas, lo que supuso un importante avance que redujo notablemente la duración de la elaboración del censo, convirtiéndose además en el primer tratamiento automático de la información de la historia.

No fue hasta la década de los sesenta, cuando se dio inicio a las primeras generaciones de **bases de datos de red** y las **bases de datos jerárquicas**, capaces de guardar estructuras de datos en listas y árboles.

En los siguientes años, se definió el **modelo relacional** que culminó con el nacimiento de un **sistema de gestión** de bases de datos relacional llamado **Oracle**. Además, IBM creó un nuevo software de base de datos, "System R", y para gestionar los datos almacenados en este, se creó el **lenguaje SQL** que en los años ochenta comenzó a ser el estándar de la industria. Este cambio vino influenciado fundamentalmente como consecuencia de la sencillez a nivel de programación que las bases de datos relacionales ofrecían, lo que hizo que pudieran competir con las bases jerárquicas y de red.

En la década siguiente, el avance principal fue la aparición de las **bases de datos orientadas a objetos**, capaces de gestionar datos complejos en los campos donde las bases de datos relacionales no podían progresar de manera eficiente.

Finalmente, en el siglo XXI, se pueden encontrar varios tipos de bases de datos en función de su **estructura**, su **contenido** o su **utilidad**. En adición, atendiendo a la manera de ordenar la información se puede distinguir entre **estáticas** y **dinámicas**. Se prevé que en los próximos años las bases de datos evolucionen hasta ser completamente desplegadas en la nube, potenciando aquellas que son dedicadas a la analítica.



## 3. Sistemas gestores de bases de datos

Un **Sistema Gestor de Bases de Datos (SGDB)** es una colección de programas cuyo objetivo es servir de interfaz entre la base de datos, el usuario y las aplicaciones. Es un sistema que permite la creación, gestión y administración de bases de datos, así como la elección y el manejo de las estructuras necesarias para el almacenamiento y búsqueda de información del modo más eficiente posible.

Un **SGDB** permite definir los datos a distintos niveles de abstracción y manipular dichos datos garantizando la seguridad e integridad de estos. Se compone de un lenguaje de definición de datos, de un lenguaje de manipulación de datos y de un lenguaje de consulta.

### 3.1. Tipos de sistemas gestores de bases de datos.

Los **Sistemas Gestores de Bases de Datos** pueden ser **clasificados** según diferentes características:

Según la variabilidad de la base de datos:

- Bases de datos estáticas.
- Bases de datos dinámicas.

Según el contenido:

- Bases de Datos Bibliográficas.
- Bases de datos de texto completo.
- Directorios.

Según los modelos de bases de datos, como por ejemplo:

- Bases de datos jerárquicas.
- Base de datos de red.
- Bases de datos relacionales.
- Base de datos orientadas a objetos.

Según el lenguaje utilizado:

- SQL.
- NoSQL.

Además de estas clasificaciones, los Sistemas Gestores de Bases de Datos también pueden ser clasificados según su **licencia**:

#### 3.1.1. Open Source

El término **Open Source** empezó a ser usado a partir de 1998, éste se refiere a todo aquél Software que puede ser accesible al **público**, es decir, puede ser distribuido, modificado, copiado o usado por cualquier persona. Este tipo de software está protegido por



**copyleft**, lo cuál impide a sus propios distribuidores colocar restricciones sobre el uso de su producto. A diferencia del Software Libre, no siempre tiene por qué ser gratuito.

Como principales **ventajas** respecto al Software Propietario tiene:

- Bajo costo de adquisición y uso.
- Innovación tecnológica.
- Menores requisitos de hardware.
- Independencia del proveedor.

Por otro lado, la principal **desventaja**, consiste en que no existen compañías únicas que respalden la tecnología.

Dentro de los Sistemas Gestores de Bases de Datos se encuentran como principales **Softwares Open Source** los siguientes:

- MySQL.
- PostgreSQL.
- MariaDB.

### 3.1.2. Propietario

El **Software Propietario** es el software del cuál no existe una forma libre de acceso a su código fuente, el cuál se encuentra a disposición de su desarrollador y no se permite su libre modificación, adaptación o incluso la lectura por parte de terceros. La persona física o jurídica al poseer los **derechos de autor**, tiene la posibilidad de controlar y restringir los derechos del usuario sobre su programa.

Como principales **ventajas** respecto al Software Open Source se encuentran:

- Control de calidad.
- Personal capacitado para soporte.
- Uso comúnmente extendido.

Por otro lado, como principales **desventajas** se encuentran:

- Dependencia de proveedores.
- Imposibilidad de la modificación del mismo.

Dentro de los Sistemas **Gestores de Bases de Datos** se encuentran como principales Softwares Propietarios los siguientes:

- Microsoft Azure.
- Oracle.
- SQL Server.



## 3.2. Funciones del sistema gestor de bases de datos.

A pesar de la gran variedad de sistemas gestores que existen en el mercado, podemos identificar una serie de funciones comunes a todos ellos:

- Recuperar y modificar los datos almacenados en los ficheros internos de la base de datos de forma transparente para el usuario.
- Proporcionar un catálogo o diccionario de datos.
- Garantizar la integridad de los datos.
- Gestionar los accesos concurrentes.
- Recuperación de datos.

## 4. Arquitectura de los sistemas gestores.

### 4.1. Arquitectura ANSI-SPARC.

Existe un modelo de arquitectura estándar en el que se basan la mayoría de gestores de bases de datos comerciales de hoy en día, aunque este no sea un modelo estándar de manera formal.

La arquitectura en cuestión se denomina ANSI-SPARC (American National Standards Institute, Standards Planning And Requirements Committee), el cual es un diseño abstracto de un SGBD propuesto por primera vez en 1975.

Es una arquitectura que consta de tres niveles: **Nivel físico**, **Nivel conceptual**, **Nivel externo**. Dichos niveles, tienen como objetivo separar la vista de los usuarios, para así ocultar la complejidad en sí de la base de datos.

Considerando el punto anterior, este tipo de arquitectura tiene como objetivo:

- Agrupar los tres niveles de abstracción: físico, lógico y externo.
- Permitir la independencia física y lógica de los datos.
- Garantizar la consistencia de los datos.
- Ofrecer seguridad de acceso.
- Gestión de transacciones: conjunto de operaciones relacionadas tratadas como una sola.
- Permitir la concurrencia de usuarios garantizando la integridad de datos.

### 4.2. Componentes de un SGBD.

Los componentes generales de un SGBD son los siguientes:



- Interfaces externas: medios para comunicarse con el SGBD en ambos sentidos (E/S).
- Intérprete o procesador del lenguaje.
- Optimizador de consultas.
- Motor de la base de datos.
- Mecanismo de almacenamiento.
- Motor de transacciones.
- Gestión y operación de SGBD.

## 6. Fabricantes de bases de datos.

A nivel mundial, los **líderes del mercado** en cuanto a fabricación de bases de datos son:

- **ORACLE** (39.8%).
- **IBM** (31.3%).
- **Microsoft**.

Respecto a empresas españolas que destacan actualmente en Big Data, tenemos:

- **CartoDB**: la más exitosa, permitiendo a sus usuarios la **toma de decisiones en tiempo real** basada en datos cualitativos y cuantitativos, que proporciona herramientas SIG y de mapeo web.

Otros productores son:

- **Epinium**: destacada por su colaboración con marcas como Philips, Nikon, Asus, Acer o Braun que ya han confiado en su tecnología.
- **Inbenta**: trabaja en colaborativo con inteligencia artificial para el procesamiento de lenguaje natural (PLN).
- **Brainsins**: ayuda a sus clientes a mejorar la experiencia de compra, estudiando en todo momento lo que éstos realmente necesitan (en general con tareas de marketing).
- **DatMean**: enfocada en el marketing de audiencias.
- **QMenta**: proporciona una plataforma en la nube para hospitales, centros de investigación, etc.
- **Geoblink**: resulta una solución para ayudar a empresas de Retail a optimizar su red de tiendas y diseñar su estrategia de expansión. Ésta promovida del CEO de Mercadona.



## 7. Big Data

**Big Data** son datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a mayor velocidad (denominando comúnmente estos factores como "las tres V").

Dicho de otro modo, el big data está formado por **conjuntos de datos de mayor tamaño y más complejos**, especialmente procedentes de nuevas fuentes de datos. Estos conjuntos de datos son tan voluminosos que el software de procesamiento de datos convencional sencillamente no puede gestionarlos. Sin embargo, estos volúmenes masivos de datos pueden utilizarse para abordar problemas empresariales que antes no hubiera sido posible solucionar.

### 7.1. Big Data y Sistemas Gestores de bases de datos

Antes del Big Data, las organizaciones disponían de información estructurada en sus aplicaciones, pero con la llegada del Big Data, la información de interés ya no solo se encuentra en el dominio y dentro del perímetro de las organizaciones, sino que está en las redes sociales, estudios estadísticos de terceros... Tratándose generalmente de información no estructurada.

En el Big Data, las BBDD generalmente utilizadas son las **NoSQL** debido a:

- Elevado número de fuentes de datos: internet, IoT, estudios, etc.
- Distinto tipo de datos: estructurados (tablas), no estructurados (documentos, vídeos, etc.), semiestructurados...
- Gran cantidad de datos.
- Alta volatilidad de los datos que cambian constantemente y hay que procesarlos de manera rápida.

Las BBDD usadas en Big Data deben responder a una serie de requisitos, como la capacidad de gestionar volúmenes muy grandes de datos, de distintos tipos (estructurados, no estructurados...), con un coste aceptable (hardware), un buen rendimiento y estar disponibles siempre. Por eso, entre sus cualidades destacan:

- Flexibilidad.
- Escalabilidad.
- Velocidad.
- Alto volumen de demanda.

Los principales gestores de bases de datos empleados en Big Data son:

- MongoDB.
- Cassandra.





- Redis.
- Neo4j.

## 7.2. Big Data y videojuegos

El sector de los videojuegos abarca principalmente el ámbito digital, lo cual impulsa a que las empresas muestren un gran interés en interpretar y acceder al **flujo constante de información** que generan los usuarios. Esta información puede abarcar desde los gustos de los jugadores o el tiempo que pasan jugando, hasta la ubicación desde donde realizan las conexiones.

Con esta información, se pueden crear experiencias enriquecedoras para el usuario y conseguir un producto mucho más personalizado. Además, gracias a la recopilación de datos, se pueden desarrollar algoritmos que sean capaces de medir las habilidades de los jugadores, así como darles a conocer sus puntos débiles, potenciando la comunidad del juego y su continuo desarrollo.

Otro de los usos más relevantes del Big Data en el sector, es su capacidad a través del análisis para determinar si un videojuego tendrá más o menos éxito, así como **predecir** futuras tendencias y apoyar tanto a los procesos de diseño del juego como al marketing.

## 8. Data Science.

La **ciencia de datos** es un campo que involucra métodos y procesos para extraer conocimiento o una mejor comprensión de los datos, ya sean estructurados o no estructurados.

Los especialistas en la materia, los científicos de datos, tienen distintas habilidades de cara al Big Data como son:

- Recopilar, procesar y extraer valor de las bases de datos.
- Usar la imaginación para comprender, visualizar y comunicar sus conclusiones.
- Tener la capacidad para crear soluciones basadas en datos que aumentan los beneficios reduciendo costos.
- Trabajar en todas las industrias haciendo frente a grandes proyectos que manejan grandes volúmenes de datos a todos los niveles.

Sin embargo, son varios los retos tecnológicos a los que se enfrentan los científicos de datos, como:

- El gran volumen de datos los cuales requieren de nuevas aplicaciones para su integración.



- Almacenamiento físico de los datos, estos requieren de nuevos medios y arquitecturas.
- Problemas de interoperabilidad.
- Limpieza de datos.
- Interpretabilidad de los modelos obtenidos con técnicas de inteligencia artificial.

## 9. Modos de comercialización.

Existen diversas formas para comercializar bases de datos, para principalmente se resumen en dos tipos:

- **Remarketing:** Técnica que consiste en crear anuncios personalizados enfocados a un público que ha visitado previamente una web específica. Dicho técnica, utiliza las 'cookies', las cuales se encargan de recopilar la información y enviarlas a las BBDD. De esta forma, las compañías compran estos listados y envían información a los suscriptores.
- **Vender la propia base de datos,** lo que supone ceder toda la información que contiene a su nuevo propietario. Para ello, es indispensable que el titular de los datos dé su consentimiento y sea informado sobre la finalidad de la cesión de sus datos.

Para seleccionar cuál de las dos técnicas es más viable, se debe llevar a cabo un estudio y emprendimiento comercial a medio/largo plazo, con el propósito de examinar:

- El tipo de información que se va a solicitar a los clientes.
- Utilidad para el negocio.
- La propia estrategia de marketing.
- Investigar y aprovechar la relación entre propietario y cliente.

## 10. Conclusión.

Existen múltiples tipos de **bases de datos**, las cuáles son diferenciadas a través de los tipos de datos que queramos procesar, el tipo de licencia, capacidad... Para poder determinar la más asequible, es necesario considerar el **gestor de bases de datos** que vamos a relacionar con nuestro sistema de almacenamiento. Dicho **SGDB**, tiene como objetivo el servir una interfaz para crear, tramitar y suministrar los datos, de tal forma que la búsqueda y el almacenamiento sea efectuado de manera eficiente.

Por otro lado, la llegada del **Big Data** ha impulsado a un análisis más exhaustivo de la tecnología utilizada, ya que la información actualmente se encuentra descentralizada.



Por ende, es recomendable que las empresas o minoritarios consideren los siguientes aspectos a la hora de utilizar un sistema de almacenamiento:

- Estudio de mercado (a quién venderemos el producto/la información)..
- Arquitectura a utilizar.
- Licencia a adquirir.
- Uso de la propia base de datos.



## 11. Referencias.

1. <https://postcron.com/es/blog/como-crear-una-base-de-datos/>
2. <https://evamunoz.es/puedo-comprar-o-vender-una-base-de-datos/>
3. <https://www.instasent.com/blog/como-hacer-remarketing-en-facebook>
4. <https://www.instasent.com/blog/lee-esto-si-vas-a-comprar-bases-de-datos-como-funciona-y-por-que-no-hacerlo>
5. <https://www.computerworld.es/archive/quien-es-el-lider-del-mercado-de-bases-de-datos#:~:text=A%20la%20cabeza%20se%20encuentra,mercado%20del%2031%2C3%25>
6. <https://marketing4ecommerce.net/empresas-de-big-data-espanolas/>
7. [https://es.wikipedia.org/wiki/Ciencia\\_de\\_datos#Ciencia\\_de\\_datos\\_y\\_Big\\_data](https://es.wikipedia.org/wiki/Ciencia_de_datos#Ciencia_de_datos_y_Big_data)
8. <https://www.techopedia.com/definition/1188/database-engine>
9. <https://www.unir.net/ingenieria/revista/bases-datos-big-data/>
10. [https://es.wikipedia.org/wiki/Arquitectura\\_ANSI-SPARC](https://es.wikipedia.org/wiki/Arquitectura_ANSI-SPARC)
11. <https://www.oracle.com/es/big-data/what-is-big-data/>
12. <https://revistadigital.inesem.es/informatica-y-tics/los-gestores-de-bases-de-datos-mas-usados/>