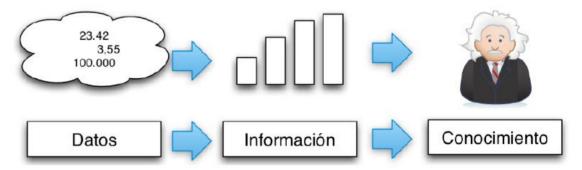
UD1.- Sistemas de almacenamiento de la información

ÍNDICE

- Sistemas de archivos frente a bases de datos
- Abstracción de la información. Arquitectura de una base de datos
- Independencia de datos
- Diseño de una base de datos
- Sistemas de gestión de bases de datos (SGBD)
- Clasificación de los SGBD

CONCEPTOS BÁSICOS

- INFORMÁTICA: Ciencia que estudia el tratamiento automático de la información.
- DATO: Palabras, números, valores usados para construir información, y que por si solos no tienen ningún sentido.
 - Pérez
 - 12/2/2021 18:23:11



- INFORMACIÓN: Dato con significado.
 - Mi primer apellido es Pérez.
 - El 12/2/2021 a las 18:23:11 se apagó el servidor por última vez



CONCEPTOS BÁSICOS

- Las bases de datos (BD) están presentes en las acciones que realizamos día a día:
 - Comprar entradas para el cine, reservar un hotel o un vuelo.
 - Matricularse en el instituto.
 - Usar Whatsapp, Instagram, Twitter...
- Algunos de estos ejemplos serían aplicaciones de bases de datos tradicionales, almacenando información textual o numérica.
- Otros pueden almacenar muy diversos tipos de información:
 - Bases de datos **multimedia** (fotos y vídeos).
 - Sistemas de información geográfica (GIS): Mapas.
 - Data warehouses: Información de grandes volúmenes de datos para toma de decisiones empresariales.
 - **Tiempo real** y bases de datos activas, para control industrial.



CONCEPTOS BÁSICOS

- La mayoría de las aplicaciones informáticas:
 - Requieren el manejo de datos
 - Estos datos deben ser almacenados

■ ¿Cuál es el propósito?

 Diseñar sistemas para <u>almacenar y gestionar</u> datos que proporcionen a los usuarios información de interés sobre un dominio determinado, cuyo análisis posterior permita sacar conclusiones y repercuta en una toma de decisiones mejor informada.



INTRODUCCIÓN

- Una base de datos es un conjunto de datos relacionados almacenados en un soporte informático.
 - Cada BD se diseña para satisfacer los requisitos de información de una empresa u organización: un instituto u hospital...
- Antes de las BBDD se trabajaba con sistemas de archivos o ficheros.
- Ambos son parte importante en los sistemas de información:
 - En informática, un sistema de información es cualquier sistema computacional que permite obtener, almacenar, gestionar, procesar, transmitir o recibir datos, para satisfacer una necesidad de información.



SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

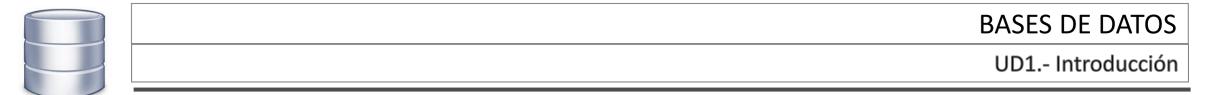
- Un **sistema basado en ficheros** es un conjunto de programas informáticos que permiten al usuario almacenar, consultar y modificar datos. Dichos datos se almacenan en **ficheros** diseñados para una determinada aplicación.
- Cada programa define y maneja sus propios datos
- Los sistemas de ficheros surgieron al informatizar el manejo de los archivadores manuales para proporcionar un acceso más eficiente a los datos almacenados en los mismos.
- Ejemplo: en el entorno empresarial donde ...
 - Cada departamento almacena y gestiona sus propios datos mediante una serie de programas de aplicación escritos especialmente para él.
 - Los departamentos no comparten información ni aplicaciones, por lo que los datos comunes deben estar duplicados en cada uno de ellos.
 - Esto puede originar inconsistencias en los datos.



- Un **fichero o archivo** es una secuencia de números binarios que organiza información relacionada a un mismo aspecto.
- Las aplicaciones pueden usar ficheros individuales, aunque su uso es cada vez menor.

TIPOS DE FICHEROS:

- Secuenciales
- De acceso directo o aleatorio
- Secuenciales encadenados
- Secuenciales indexados
- Indexado-encadenados



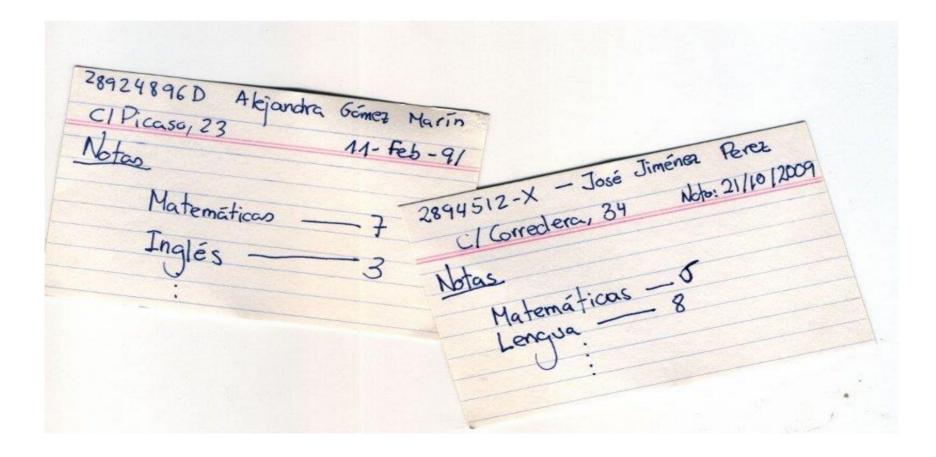
¿Cómo se ha ido guardando la siguiente información a lo largo de los últimos años?

Queremos guardar los datos de los alumnos de nuestro instituto, así como las asignaturas de las que están matriculados y las notas que han obtenido en cada asignatura.



SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

En los años 60, usando papel y "boli" ...







En los años 70 y 80 se utilizaban **ficheros de texto** donde se guardaba la información:

alumnos.tx DNI	t NOMBRE		DIRECCIÓN	FECHA NTO
	José Jiménez Perez Alejandra Gómez Marín	_	•	21-10-90 11-02-91

asignaturas.txt
DNI NOMBRE ASIGNATURA NOTA

2894512X José Jiménez Perez Matemáticas 5
2894512X José Jiménez Perez Lengua 8
....
28924896D Alejandra Gómez Marín Matemáticas 7
28924896D Alejandra Gómez Marín Inglés 3



SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

A partir de los años 90 mediante **bases de datos**. Un SGBD guarda la información en las siguientes tablas:

```
Alumnos (DNI, Nombre, Dirección, Fecha nacimiento)
```

2894512X José Jiménez Perez C/ Corredera,34 21-10-90 28924896D Alejandra Gómez Marín C/ Picaso, 23 11-02-91

• • •

Asignaturas (Código, Nombre)

001 Matemáticas

002 Lengua

003 Inglés

Notas(DNI, Código_asignatura, nota)

2894512X 001 5

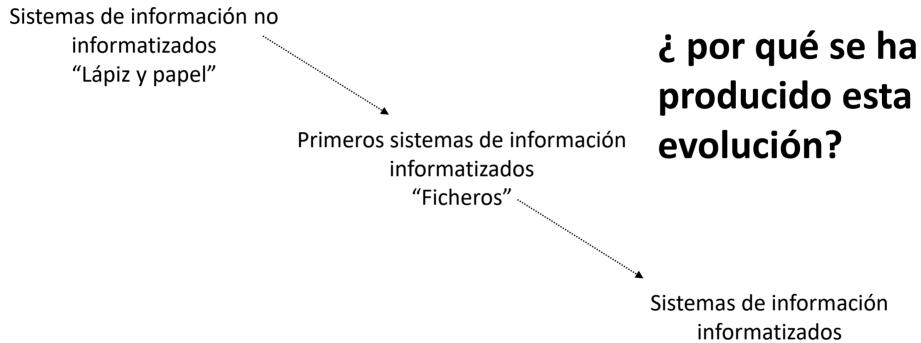
2894512X 002 8

28924896D 001 7

28924896D 003 3



SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS



Bases de Datos

SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

Dificultad en el acceso a los datos

- Se busca un acceso eficiente y rápido.
- Queremos obtener la siguiente información:

Se quiere conocer el número de alumnos de más de veinticinco años y con nota media superior a siete que están matriculados actualmente en la asignatura Bases de datos I.

- ✓ S.I. Sin informatizar: Obtener está información puede requerir mucho tiempo y mucho trabajo, además hay que realizando cálculos (media,...) e ir mirando alumno por alumno.
- ✓ **S.I. Con ficheros**: Podemos crear un programa que vaya obteniendo la información del fichero vaya realizando los cálculos y nos de los resultados.
- ✓ S.I. Con base de datos: Esta consulta es trivial usando un lenguaje de consulta de datos.



SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

Flexibilidad en los cambios

• Si las necesidades del sistema de información cambian, ¿cómo se comporta cada uno de nuestros tres modelos?

Queremos guardar el nombre del profesor que imparte cada asignatura.

- ✓ S.I. Sin informatizar: Tenemos que ir escribiendo el nombre de profesor en cada ficha.
- ✓ S.I. Con ficheros: tendríamos que cambiar el fichero de notas.txt e ir escribiendo una columna más, mucho trabajo.
- ✓ S.I. Con bases de datos: simplemente habría que añadir un atributo a la tabla asignaturas, con lo que sólo se escribiría una vez el nombre del profesor de cada asignatura.

SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

Control de la redundancia

- La redundancia es la cantidad de datos repetidos en la información guardada.
- El objetivo es reducir todo lo posible la redundancia.
- La **inconsistencia** de los datos se produce cuando un dato redundante es diferente en dos o más sitios.

√ ¿Cuál de los modelos presentados crees que tiene menos redundancia?



Integridad

- La integridad se refiere a la validez de los datos.
- La información almacenada debe ser coherente y veraz.
- ¿Qué ocurriría en cada uno de los modelos presentados en los siguientes casos?
 - Una persona se ha mudado y cambia su dirección.
 - Nos hemos equivocado a introducir los datos de una persona y tenemos que cambiar el nombre.
 - Cambiamos el nombre de una asignatura.
 - Desaparece una asignatura del plan de estudio.
- Las bases de datos aseguran automáticamente la **integridad de los datos**, sin que el usuario tenga que realizar ninguna operación.
 - ¿Podemos registrar la nota de un alumno que no existe?

SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

Concurrencia

- La concurrencia se refiere a la posibilidad de que varios usuarios trabajen simultáneamente.
- Ejemplo: Tres administrativos que están trabajando con la información que tenemos guardada.
- ✓ S.I. Sin informatizar: Si por ejemplo las fichas en papel están encuadernadas, es complicado que varias personas puedan trabajar al mismo tiempo con la información.
- ✓ **S.I. Con ficheros**: Si tenemos a los tres administrativos con programas que leen y modifican los ficheros de textos, puede ocurrir que en un determinado momento una de ellas este leyendo un dato incorrecto.
- ✓ **S.I. Con base de datos**: Existe el concepto de **transacción**, por el que se asegura que la información va a ser siempre consistente.

SISTEMAS DE ARCHIVOS FRENTE A BASES DE DATOS

Seguridad

- Estamos trabajando con datos sensibles, que no todo el mundo puede tener acceso a ellos.
- Sólo determinadas personas deben poder acceder a algunas informaciones: datos personales, historial médico, historial policial, etc...

√ ¿Cómo de seguro es cada una de los modelos que hemos estudiado?



Una base de datos es un conjunto exhaustivo, no redundante de datos, estructurados, organizados independientemente de su utilización y su implementación en máquina, accesibles en tiempo real y compatibles con *usuarios concurrentes* con necesidad de información diferente.

- **Ventajas.** Conjunto exhaustivo **no redundante** de datos. Datos organizados **independientemente de su utilización y su implementación en máquina**. Datos compatibles con usuarios concurrentes, y con necesidad de información diferente.
- Inconvenientes. Implantación costosa y difícil. Necesidad de personal cualificado.



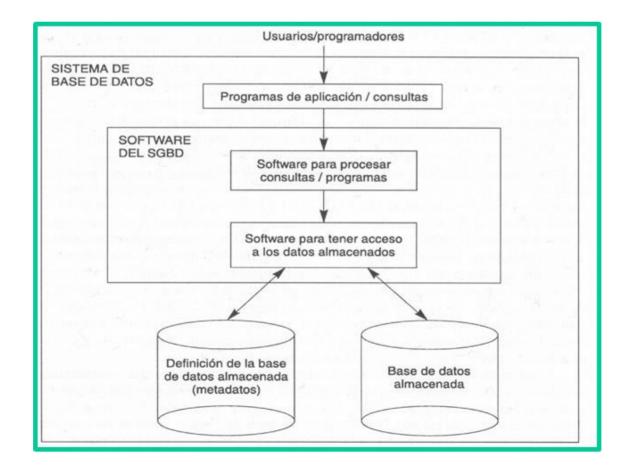
Podemos acceder a nuestra BD desde un programa en Haskell ...

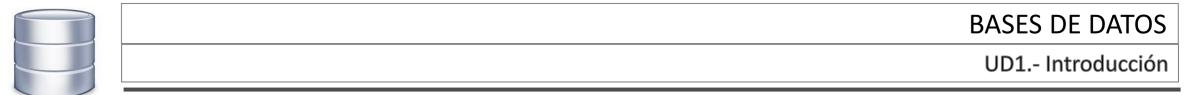


- Una base de datos se puede percibir como un gran **almacén de datos** que se define y se crea una sola vez, y que se utiliza al mismo tiempo por distintos usuarios.
- En una base de datos todos los datos se integran con una **mínima** cantidad de **duplicidad**. La base de datos se **comparte** por toda la **organización**.
- Además, 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.
- Esta diseñada, construida y poblada con datos para un propósito específico.



Cualquier base de datos está formada por un conjunto de ficheros, el SGBD se encarga de los detalles de almacenamiento y gestión.





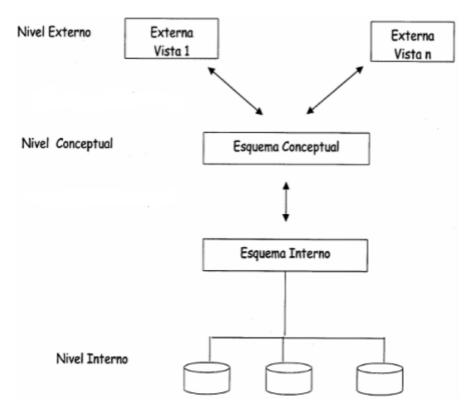
Abstracción de la información. Arquitectura de una Base de datos

Los diferentes **niveles de abstracción** sirven para simplificar la interacción de los usuarios con el sistema.

Niveles de abstracción de la información \rightarrow Son diferentes puntos de vista desde los que se puede OBSERVAR/DEFINIR la base de datos, y ocultar su complejidad.

En cada nivel hay un esquema

Un esquema es una descripción de la base de datos.





Abstracción de la información. Arquitectura de una Base de datos

NIVEL INTERNO

- Describe la forma de almacenar los datos en los dispositivos de almacenamiento.
- Esta visión sólo la requiere el administrador/, para poder gestionar más eficientemente la base de datos.
- El esquema interno se habla de archivos, directorios, unidades de disco,...

> NIVEL CONCEPTUAL

- Describe qué datos son almacenados realmente en la base de datos y las relaciones que existen entre los mismos.
- Tiene un **esquema conceptual** con: entidades, tipos de datos, relaciones y restricciones
- Lo realizan diseñadores. (Diagrama Entidad-Relación y Modelo relacional)

> NIVEL EXTERNO

- Incluye varios esquemas externos o vistas que describen la visión de los datos que poseen los usuarios finales.
- Esa visión es la que se obtiene a través de las aplicaciones.

INDEPENDENCIA DE LOS DATOS

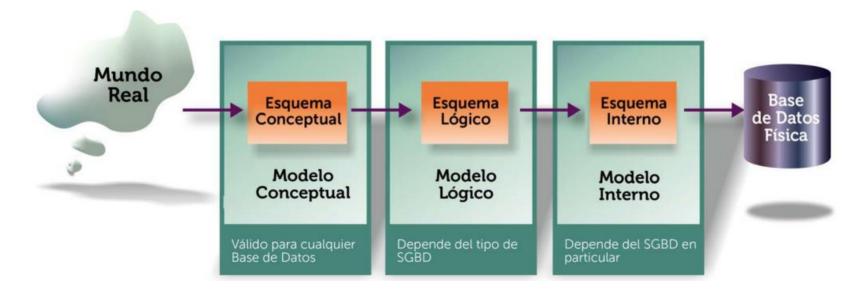
- La separación en niveles proporciona independencia de datos:
 - Independencia de datos es la capacidad para modificar el esquema de un nivel del sistema de la base de datos sin tener que modificar el esquema del nivel inmediato superior.
 - Independencia lógica de los datos es la capacidad de modificar el esquema conceptual sin tener que alterar los esquemas externos ni los programas de aplicación.
 - Independencia física de los datos es la capacidad de modificar el esquema interno sin tener que alterar el esquema conceptual.





DISEÑO DE UNA BASE DE DATOS

- Los modelos de datos permiten la descripción de los datos con diferentes niveles de detalle.
- El DISEÑO de la base de datos implica el diseño de varios esquemas:
 - Fase de diseño conceptual → Esquema conceptual
 - Fase de diseño lógico → Esquema lógico
 - Fase de diseño físico → Esquema interno





SISTEMAS GESTORES DE BASES DE DATOS

Las bases de datos requieren de un Software de Gestión que facilite las operaciones y las interfaces con los usuarios. Esto es lo que se conoce como el **Sistema de Gestión de Bases de Datos (S.G.B.D.).**

➤ **DEFINICIÓN:** Un sistema gestor de bases de datos (también DBMS, Data Base Management System) es el software que permite a los usuarios describir, procesar, administrar y recuperar los datos almacenados en una base de datos.



El éxito del SGBD reside en mantener la seguridad e integridad de los datos.

BASES DE DATOS
UD1 Introducción



SISTEMAS GESTORES DE BASES DE DATOS

Los SGBD proporcionan un conjunto coordinado de **programas**, **procedimientos** y **lenguajes** que permiten a los distintos usuarios realizar sus tareas habituales con los datos. Incluye:

- ✓ Herramientas para la creación y especificación de los datos, así como la estructura de la base de datos.
- ✓ Herramientas para administrar y crear la estructura física requerida en las unidades de almacenamiento.
- ✓ Herramientas para la manipulación de los datos de las bases de datos, para añadir, modificar, suprimir o consultar datos.
- ✓ Herramientas de recuperación en caso de desastre.
- ✓ Herramientas para la creación de copias de seguridad .
- ✓ Herramientas para la gestión de la comunicación con la base de datos.
- ✓ Herramientas para la creación de aplicaciones que utilicen esquemas externos de los datos.
- ✓ Herramientas de instalación de la base de datos.
- ✓ Herramientas para la exportación e importación de datos.

SISTEMAS GESTORES DE BASES DE DATOS - FUNCIONES

a) Función de descripción/definición

- Permite definir las tres estructuras de la base de datos en los tres niveles de abstracción:
 - Especificar los elementos de datos que la integran, su estructura y las relaciones que existen entre ellos, las reglas de integridad semántica, los controles a efectuar antes de autorizar el acceso a la BD, etc., así como las características de tipo físico y las vistas lógicas de los usuarios
- Para ello se usa un lenguaje de definición de datos o DDL.

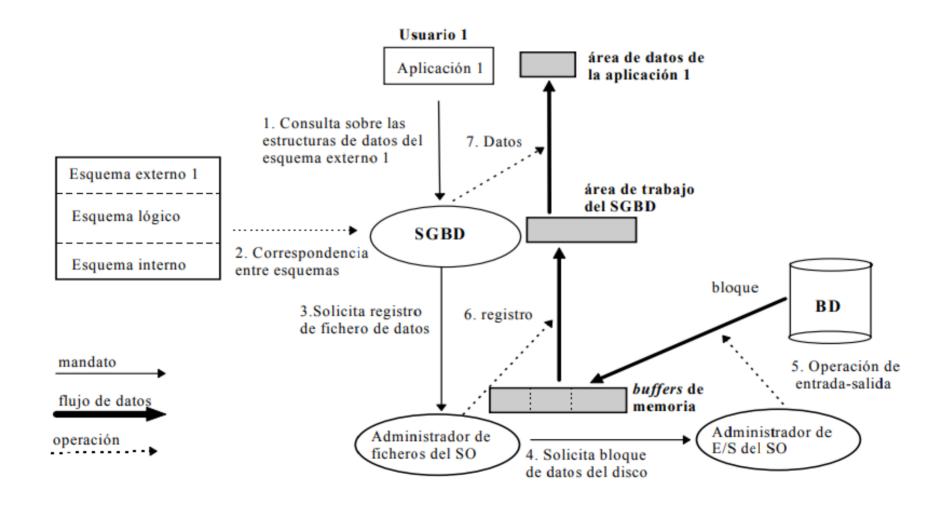
b) Función de manipulación

- Permite a los usuarios buscar, añadir, suprimir o modificar los datos de la base de datos.
- Se llevará a cabo por medio de un lenguaje de manipulación de datos (DML).

c) Función de utilización

- Incorpora interfaces que permiten a los usuario comunicarse con la BD.
- Proporciona un conjunto de procedimientos para el administrador (copias de seguridad, rearranque en caso de caída del sistema, etc.) y de protección frente a accesos no autorizados.
- Esto lo realiza el LCD, Lenguaje de Control de Datos.

SISTEMAS GESTORES DE BASES DE DATOS - INTERACCIÓN DEL SISTEMA OPERATIVO CON EL SGBD





SISTEMAS GESTORES DE BASES DE DATOS - COMPONENTES

Para realizar todas las **funciones** descritas anteriormente, es necesario que el SGBD cuente con una serie de componentes:

- Lenguajes de la base de datos.
- El diccionario de datos.
- > El gestor de la base de datos.
- Usuarios y administradores de la base de datos.





SISTEMAS GESTORES DE BASES DE DATOS - COMPONENTES

Los lenguajes de la base de datos deben permitir:

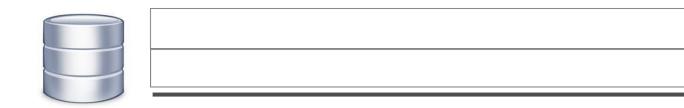
- Crear la estructura de la base de datos, incluyendo todos los objetos que puede incluir la misma (tablas, vistas, usuarios, procedimientos, funciones, triggers, etc.). DDL
- Consultar y manipular la información almacenada en la base de datos. DML
- Asignar privilegios a usuarios, confirmar o abortar transacciones, etc. DCL.
- En algunos casos, también incluyen un lenguaje de cuarta generación (4GL) para RAD (desarrollo rápido de aplicaciones). Ej: Asistentes de Access, Oracle Developer Suite.



SISTEMAS GESTORES DE BASES DE DATOS - COMPONENTES

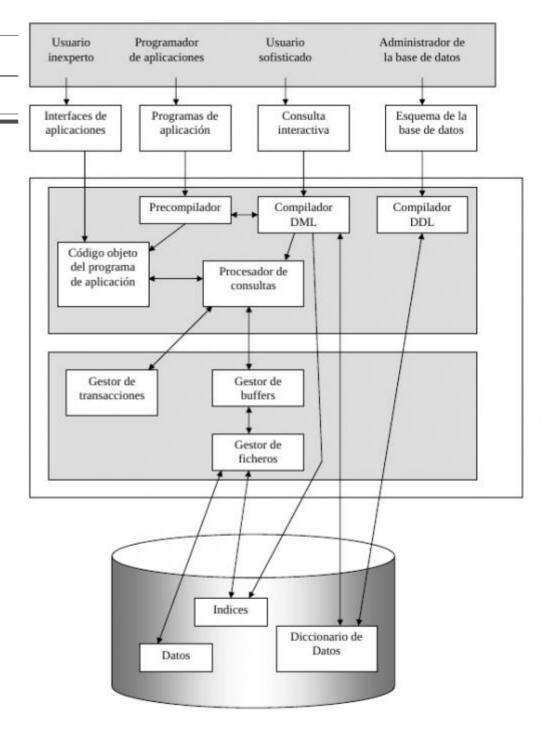
El **diccionario de datos** contiene los metadatos (*datos acerca de los datos*) de la base de datos, es decir:

- La definición de todos los objetos existentes en la base de datos: tablas con sus columnas, vistas, procedimientos, triggers, índices, etc...
- La ubicación física de los objetos y el espacio asignado a los mismos.
- Los privilegios y roles asignados a los usuarios.
- Las restricciones de las tablas.
- Información de auditoría.
- Estadísticas de uso de la base de datos.



El gestor de base de datos o monitor, es un componente software encargado de garantizar el correcto, seguro, íntegro y eficiente acceso y almacenamiento de los datos.

Puede verse el gestor de la base de datos como un intérprete entre el usuario y los datos.



SISTEMAS GESTORES DE BASES DE DATOS — ADMINISTRADORES Y USUARIOS

Administrador de la base de datos (ABD)

Tiene el control centralizado de la base de datos y es el responsable de su buen funcionamiento. Es el encargado de autorizar el acceso a la base de datos, de coordinar y vigilar su utilización y de adquirir los recursos software y hardware que sean necesarios.

- **Definir** el **esquema lógico** de la base de datos mediante sentencias del DDL.
- **Definir** el **esquema físico** de la base de datos.
 - Conjunto de sentencias en DDL que definen las estructuras de almacenamiento y los métodos de acceso a la información.
- Definir las vistas externas o de usuario de la base de datos.
- Concesión de privilegios a usuarios.
- Modificaciones de los esquemas.





SISTEMAS GESTORES DE BASES DE DATOS — ADMINISTRADORES Y USUARIOS

Los diseñadores se encargan de identificar los datos que se almacenarán en la base de datos y de elegir las estructuras apropiadas para almacenar dichos datos. Suelen formar parte del personal de ABD.

Los **operadores y personal de mantenimiento** forman parte del personal del ABD y son los responsables del funcionamiento y mantenimiento reales del entorno software y hardware del sistema de base de datos.



SISTEMAS GESTORES DE BASES DE DATOS — ADMINISTRADORES Y USUARIOS

La base de datos existe para que los usuarios finales la utilicen:

- Usuarios finales ocasionales.
 - Accede ocasionalmente y utilizan un lenguaje de consulta de bases de datos avanzado para especificar sus solicitudes.
 - Suelen ser gerentes de nivel medio o alto.
- Usuarios finales simples o paramétricos.
 - Realizan consultas y actualizaciones constantes de la base de datos, utilizando tipos estándar de consultas y actualizaciones.
- Usuarios finales avanzados.
 - Personal altamente cualificado que implementan sus aplicaciones de forma que cumplan sus complejos requerimientos.
- Usuarios finales autónomos.
 - Mantienen bases de datos personales mediante la paquetes de programas comerciales.





SISTEMAS GESTORES DE BASES DE DATOS — ADMINISTRADORES Y USUARIOS

Analistas de sistemas y Programadores de aplicaciones: no son considerados como usuarios en sí de una base de datos.

- ➤ Los **analistas** de sistemas **determinan los requerimientos** de los usuarios finales y desarrollan especificaciones para transacciones programadas que satisfagan dichos requerimientos.
- ➤ Los **programadores** de aplicaciones **implementan** esas especificaciones en forma de **programas** y luego prueban, depuran, documentan y mantienen estas transacciones programadas.



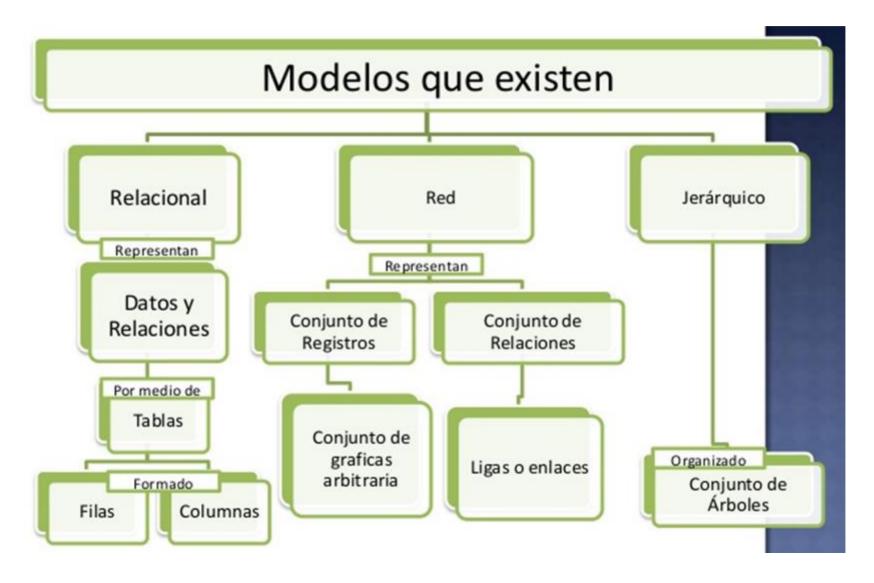


Según el modelo lógico de datos:

Los modelos lógicos empleados en los SGBD comerciales actuales son:

- Modelo en red (obsoleto)
- Modelo jerárquico (obsoleto)
- Modelo relacional.
- Modelo orientado a objetos
- NoSQL

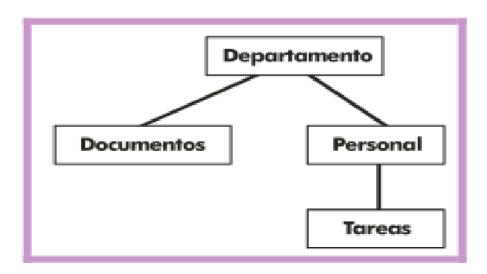
SISTEMAS GESTORES DE BASES DE DATOS - CLASIFICACIÓN



SISTEMAS GESTORES DE BASES DE DATOS - CLASIFICACIÓN

EL MODELO JERÁRQUICO:

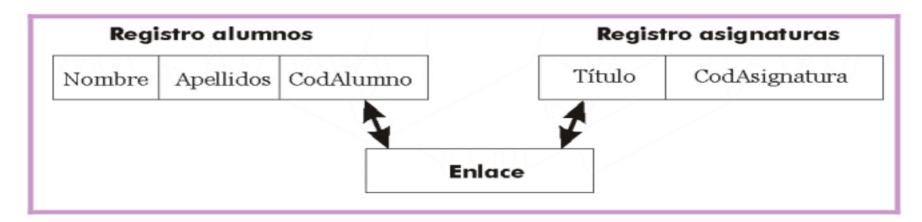
- Era utilizado por los **primeros SGBD**, desde que IBM lo definió para su IMS (Information Management System, Sistema Administrador de Información).
- La información se organiza con una **jerarquía** en la que la relación entre las entidades de este modelo siempre es del tipo padre / hijo. Los datos de este modelo se almacenan en estructuras lógicas llamadas **segmentos**.
- Los segmentos se relacionan entre sí utilizando arcos.





EL MODELO EN RED/CODASYL:

- Ha tenido una gran aceptación (apenas se utiliza actualmente).
- El modelo en red organiza la información en registros (también llamados nodos) y enlaces. En los registros se almacenan los datos, mientras que los enlaces permiten relacionar estos datos.
- En este modelo se pueden representar perfectamente cualquier tipo de relación entre los datos, pero su manejo es complicado.





EL MODELO RELACIONAL:

- Propuesto en los años 70 por Codd.
- System R (IBM) se desarrolló para probar la funcionalidad del modelo relacional.
- Grandes aportaciones:
 - Desarrollo de un lenguaje de consultas estructurado, SQL, estándar de los sistemas relacionales.
 - Producción de SGBD relacionales durante los 80, como DB2 y SLQ/DS, de IBM, y Oracle.
- En 1976, Peter Chen presentó el modelo entidad-relación, que es la técnica más utilizada en el diseño de bases de datos.



SISTEMAS GESTORES DE BASES DE DATOS - CLASIFICACIÓN

MODELO DE BASES DE DATOS OO:

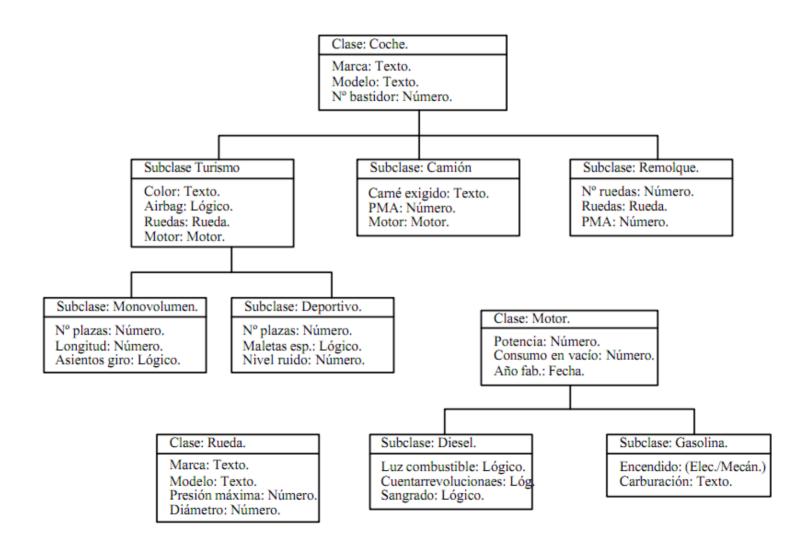
- Desde la aparición de la **programación orientada a objetos** (POO u OOP) se empezó a pensar en bases de datos adaptadas a estos lenguajes.
- Se intenta que estas bases de datos consigan arreglar las limitaciones de las relacionales: tipos definidos por el usuario, disparadores (triggers) almacenables en la base de datos, soporte multimedia...
- Se supone que son las bases de datos de 3G (la primera fueron las bases de datos en red, y la segunda las relacionales). Pero siguen sin reemplazar a las relacionales, aunque son el tipo de base de datos que más está creciendo en los últimos años.
- Su modelo conceptual se suele diseñar en **UML** y el lógico actualmente en ODMG (*Object Data Management Group*, grupo de administración de objetos de datos, organismo que intenta crear estándares para este modelo).



BASES DE DATOS

UD1.- Introducción

SISTEMAS GESTORES DE BASES DE DATOS - CLASIFICACIÓN





BASES DE DATOS NoSQL:

- Bajo este nombre se agrupan las bases de datos pensadas para grabar los datos de manera veloz para así poder atender a miles y miles de peticiones.
- Es el modelo de las bases de datos que se utilizan en los grandes servicios de Internet (como twitter, Facebook, Amazon,...).
- La idea es que los datos apenas necesitan validarse y relacionarse y lo importante es la disponibilidad de la propia base de datos.
- El nombre NoSQL, hace referencia a que este modelo de bases de datos rompe con el lenguaje SQL (el lenguaje de las bases de datos relacionales) para poder manipular los datos con lenguajes de otro.









SEGÚN EL NÚMERO DE USUARIOS:

- Sistemas monousuario sólo atienden a un usuario a la vez.
- Sistemas multiusuario, atienden a varios usuarios al mismo tiempo.

SEGÚN EL NÚMERO DE SITIOS EN LOS QUE ESTÁ:

- Casi todos los SGBD son centralizados: sus datos se almacenan en un solo ordenador. Los SGBD centralizados pueden atender a varios usuarios, pero el SGBD y la base de datos en sí residen por completo en una sola máquina.
- En los SGBD distribuidos la base de datos real y el propio software del SGBD pueden estar distribuidos en varios sitios conectados por una red.