

Introducción

(Cap 1 y 2 - Elmasri 5ª edición)

Que es una BD

- ♦ Una BD es un sistema informático que **permite, organiza y administra** el acceso a datos de una manera **eficiente**.
- ♦ Es una **tecnología informática** que permite de una manera eficaz manejar información **estructurada y masiva** en almacenamiento **persistente**.
- ♦ En definitiva es un
 - conjunto de **datos relacionados** entre si y
 - las **técnicas** para **manipular** los mismos.
- ♦ Manipular en el sentido amplio.
- ♦ Los **datos** de una BD, son (i)**hechos con un significado** que (ii)**permiten registrarse**.

¿Qué propiedades tienen las BDs?

- ♦ Las BDs representan algún aspecto del **mundo real** que se suele denominar *universo del discurso*.
- ♦ Están formadas por un conjunto de **datos lógicamente coherentes** y con **cierto orden** (no es una colección aleatoria de datos).
- ♦ Normalmente tienen un **propósito específico**:
 - Cuando una BD se construye y rellena es siempre con un propósito específico.
 - Va dirigida a un grupo de usuarios y la BD dispone de una serie de aplicaciones que usaran esos usuarios
- ♦ Por ejemplo IMDB es una BD dirigida a personas interesadas en el cine que tiene una serie de aplicaciones que te permite buscar información relacionada (al menos para los usuarios finales).

¿Qué es un SGBD?

- ♦ Que un **sistema de gestión** de BD.
- ♦ En definitiva es el conjunto de programas que permiten a los usuarios mantener y crear una BD (ya sea a nivel de usuario final o como administrador)
- ♦ Un SGBD facilita la **definición, construcción y manipulación** de una BD.
- ♦ **Definir** es especificar (la definición de una BD se almacena **en forma de catálogo** o diccionario de la BD, son los **metadatos**). Se pueden definir:
 - Tipos de datos
 - Estructuras de datos
 - Restricciones de datos
- ♦ La **construcción** consiste en guardar los datos en un medio de almacenamiento.

¿Qué es un SGBD?

- ♦ La **manipulación** consiste en:
 - **Consultar** la BD para obtener información o datos.
 - **Actualizar** la BD para reflejar cambios de la misma.
 - **Generar informes** (un informe en una BD es un documento de texto generalmente que muestra los datos de una BD del modo que se especifique).
- ♦ Otras funciones no menos importantes de un SGBD son todos **protocolos** que se realizan en una BD **para ser protegida**:
 - Respecto al mal funcionamiento de HW/SW.
 - Respecto a la seguridad y privacidad de los propios datos que forman la BD.

¿Qué es un sistema de Bases de Datos?

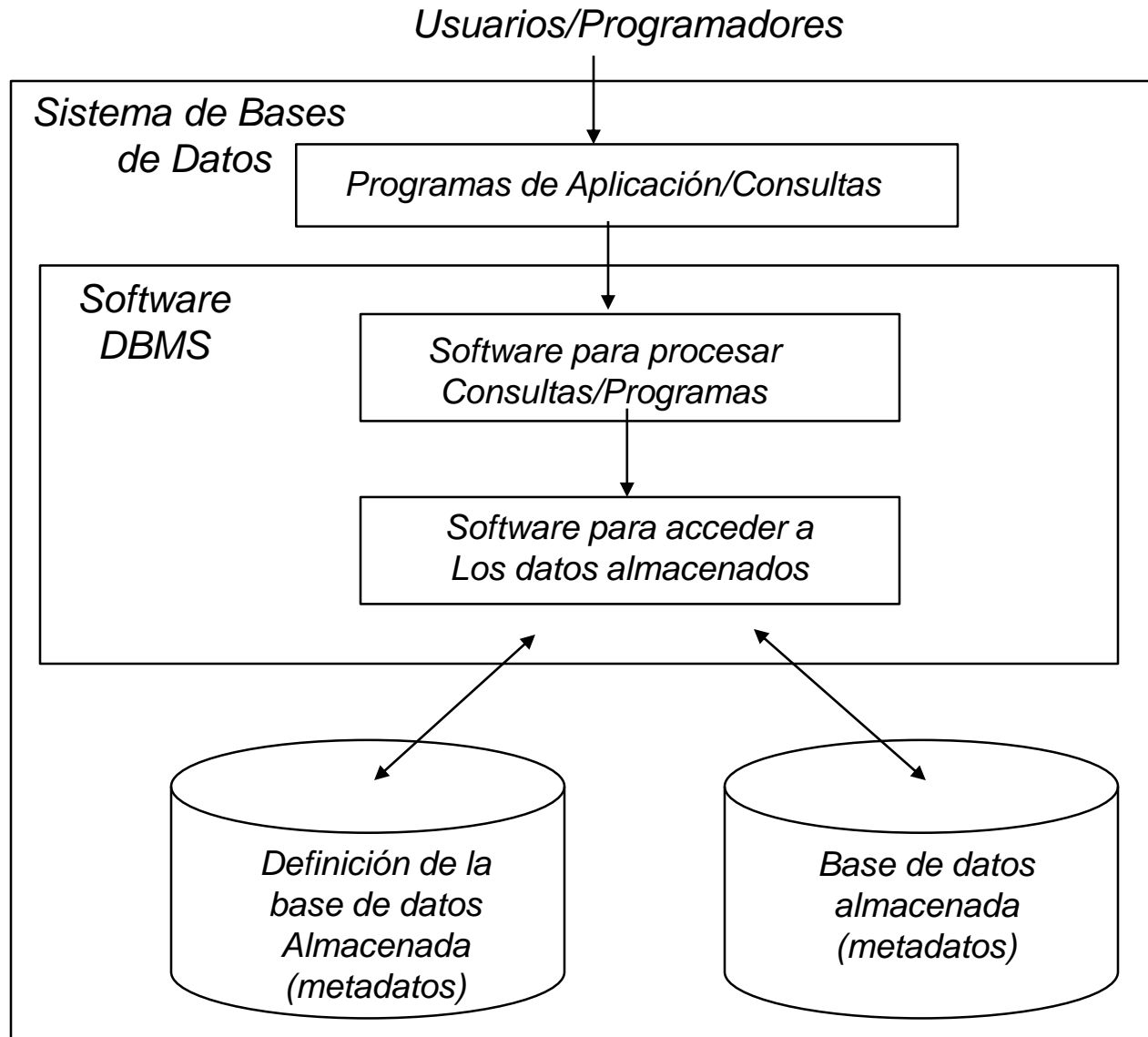
Sistema de Bases de Datos =

BD

+

SGBD (o DBMS en inglés)

Entorno simplificado de un sistema de base de datos



Un primer ejemplo

Base de datos que almacena la información de estudiantes y cursos

- ♦ **Tipos de datos:** estudiantes, identificadores de estudiantes, cursos, profesores, calificaciones,...
- ♦ **Estructuras:**
 - Los estudiantes tienen un número, nombre, departamento,...
 - Los cursos tienen nombre del curso, número, horas, ...
 - Los informes de calificaciones tienen número de estudiante, curso, nota, ...
 - Relaciones: muy variadas
 - Los estudiantes tienen un número asociado, clase, especialidad, los cursos tienen título del mismo, identificador del curso, ...
- ♦ **Funcionalidades:**
 - Buscar numero de matriculados en un curso, expedientes de estudiantes, informes de calificaciones,...
 - Ver / añadir estudiantes, cursos, ...

BD que almacena estudiantes y cursos

ESTUDIANTE

Nombre	NumEstudiante	Clase	Especialidad
Luis	17	1	CS
Carlos	8	2	CS

CURSO

NombreCurso	NumeroCurso	Horas	Departamento
Introducción a la computación	CC1310	4	CC
Estructura de datos	CC3320	4	CC
Matemática discreta	MAT2410	3	MAT
Bases de datos	CC3380	3	CC

SECCION

IDSeccion	NumCurso	Semestre	Año	Profesor
85	MAT2410	Otoño	04	Pedro
92	CC1310	Otoño	04	Ana
102	CC3320	Primavera	05	Elisa
112	MAT2410	Otoño	05	Antonio
119	CC1310	Otoño	05	Juan
135	CC3380	Otoño	05	Enrique

BD que almacena estudiantes y cursos

INFORME_CALIF

NumEstudiante	IDSeccion	Nota
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PRERREQUISITO

NumCurso	NumPrerrequisitos
CC3380	CC3320
CC3380	MAT2410
CC3320	CC1310

Un primer ejemplo: construcción

- ♦ Observar la **relación entre los registros**: por ejemplo Luis en estudiantes esta relacionado con dos registros en informe de calificaciones.
- ♦ Fases para la creación del sistema de BD:
 - **Definición de requisitos y análisis**: se documentan y transforman en un **diseño conceptual**. El diseño conceptual se puede representar y manipular mediante herramientas computarizadas. Así la implementación de la BD se puede mantener, modificar y transformar fácilmente (por ejemplo el más sencillo es modelo E-R).
 - El diseño conceptual después se convierte en un **diseño lógico** que se puede expresar en un modelo de datos implementado en el un SGBD comercial. El más conocido y que más se utiliza es el **modelo relacional** que esta basado en relaciones entre datos como su nombre indica.
 - Finalmente está el **modelo físico** donde se proporcionan especificaciones simplemente para el almacenamiento y acceso de la BD (**nivel de implementación**).

Diferencias entre ficheros y una BD

- ♦ Un **sistema de ficheros** como tal permite el acceso a los datos que se encuentran en los ficheros. **Para ser una BD** el sistema **necesita algo más**:
 - Un **modelo de datos físico**: La disposición de los datos en el disco que permita la **indexación y acceso rápido** a todos los datos.
 - Un **modelo de datos lógico**: nos permite el **acceso a datos a nivel alto**, preocupándonos de cómo los datos tienen que estar organizados para describir el dominio que estamos modelando, y sin preocuparnos de cómo se organizan los datos a nivel físico.
 - **Lenguaje de consulta**: lenguaje de alto nivel que nos permite hacer consultas de los datos representados mediante el modelo lógico. Generalmente es un lenguaje declarativo y no procedural. Este lenguaje utiliza de manera automática los algoritmos óptimos y oportunos para acceder a los datos del modelo físico.

Desacople del modelo lógico y físico

- ♦ Un hecho muy importante que produjo el desarrollo de las BDs es que las **características del modelo lógico y físico se desacoplaran y se comunicaran por un lenguaje de consulta.**
- ♦ Se dio por primera vez en 1970 con la introducción del modelo relacional de Codd ([A Relational Model of Data for Large Shared Data Banks", in Communications of the ACM, 1970](#)).
- ♦ Esta separación es muy importante (**divide y vencerás**), ya que el problema de acceso a datos se divide en:
 - Estudio de algoritmos eficientes para almacenamiento de datos en disco y su acceso optimizado (**modelo físico**).
 - Y el diseño de las aplicaciones se centra en generar un buen modelo de datos a alto nivel (**modelo lógico**).
- ♦ Esta separación de actividades **asegura el gran éxito de las BDs** en todos los campos.

Características de la metodología de BD vs ficheros

- ♦ **En programación con ficheros tradicional** cada usuario **define e implementa los archivos necesarios para una aplicación concreta:**
 - Por ejemplo un usuario de la oficina de modificación de notas y un usuario de la oficina de contabilidad utilizan datos comunes, pero cada uno utiliza los que tiene en ese momento.
 - Así los datos de las dos oficinas están relacionados y repetidos en muchos casos.
 - Además ambas oficinas deben mantener datos comunes y se tienen que pasar información.
- ♦ **En una BD esto no pasa**, los datos están centralizados, hay un único almacén de datos.

Características de las BDs vs ficheros

- ♦ **Naturaleza autodescriptiva** del sistema de BD:
 - Se almacena en el catálogo del SGBD (metadatos)
 - El SW de SGBD puede acceder a diferentes BD's solo mirando los catálogos.
 - Así un ejemplo ilustrativo de una parte del catálogo para la BD estudiantes está en la siguiente figura.
- ♦ **Aislamiento entre programas, datos y abstracción de los datos:**
 - Generalmente en el enfoque de ficheros si se introducen cambios en un fichero cambia el programa que accede a los mismos (ver figura con registros de estudiantes).
 - En una BD si queremos un nuevo campo (ej. fecha de nacimiento del estudiante), lo añadimos, se añade al catálogo, actualizamos la BD y todos los programas de acceso a datos siguen funcionando.

Ejemplo de catalogo de la BD

Relaciones

NombreRelacion	NumDeColumnas
ESTUDIANTE	4
CURSO	4
SECCIÓN	5
INFORME_CALIF	3
PRERREQUISITO	2

Ejemplo de catalogo de la BD

Columns

NombreColumna	TipoDatos	PerteneceARelacion
Nombre	Carácter (30)	ESTUDIANTE
NumEstudiante	Carácter (4)	ESTUDIANTE
Clase	Entero (1)	ESTUDIANTE
Especialidad	TipoEspecialidad	ESTUDIANTE
NombreCurso	Carácter (10)	CURSO
NumCurso	XXXXNNNN	CURSO
*****	*****	*****
*****	*****	*****
*****	*****	*****
*****	*****	*****
NumPrerrequisitos	XXXXNNNN	PRERREQUISITO

Ejemplo de almacenamiento de un registro de ESTUDIANTE basado en anterior catalogo de BD

Nombre del elemento de datos	Posición inicial en el registro	Longitud en caracteres (bytes)
Nombre	1	30
NumEstudiante	31	4
Clase	35	1
Especialidad	36	4

Características de las BDs vs ficheros

- ♦ Soporte de **varias vistas de datos**:
 - Cada usuario puede necesitar una perspectiva de la BD.
 - Una vista es un subconjunto de la BD, o puede contener datos virtuales derivados de BD que no están explícitamente almacenados. Todo esto es transparente al usuario.
 - Así un ejemplo ilustrativo de una parte del catálogo para la BD estudiantes está en la siguiente figura.
- ♦ **Compartición de datos y procesamiento de transacciones multiusuario.**
 - Las BDs son concurrentes y tienen control de concurrencia: que varios usuarios que actualicen los mismos datos lo hagan de manera controlada (ej, reserva de asientos de vuelo).

Soporte de varias vistas de la BD

Certificado

NombreEstudiante	CertificadosEstudiante				
	NumCurso	Nota	Semestre	Año	IDSeccion
Luis	CC1310	C	Otoño	05	119
	MAT2410	B	Otoño	05	112
Carlos	MAT2410	A	Otoño	04	85
	CC1310	A	Otoño	04	92
	CC3320	B	Privamera	05	102
	CC3380	A	Otoño	05	135

Prerrequisito_curso

NombreCurso	NumCurso	Prerrequisitos
Base de Datos	CC3380	CC3320
		MAR2410
Estructura de Datos	CC3320	CC1310

Roles en el uso de una base de datos

- ♦ **Usuarios finales**

- Interactúan con aplicaciones que acceden a la BD

- ♦ **Usuarios *avanzados***

- Interactúan con la BD en SQL

- ♦ **Programadores de aplicación**

- Interactúan con la BD escribiendo programas

- ♦ **Diseñadores**

- Definen el diseño de la BD

- ♦ **Administradores**

- Mantienen el diseño de la BD
- Gestionan usuarios y permisos de acceso
- Gestionan necesidades de actualización

- ♦ **Desarrolladores de herramientas SGBD**

- Implementan la capa inferior de acceso físico a los datos
- Desarrollan el software y herramientas que dan servicio a todo lo anterior

Problemas de los sistemas de ficheros

- ♦ **Redundancia e inconsistencia** de datos: la misma información puede estar duplicada en diferentes archivos como hemos visto antes.
- ♦ **Dificultad en el acceso de los datos:**
 - Hay que escribir un programa de acceso para cada consulta.
 - Los programas son difíciles de escribir ya que la información esta en varios ficheros y si se añaden nuevos campos hay que cambiar el programa.
- ♦ **Problemas con la integridad:** las restricciones de integridad son propiedades que deben satisfacer los datos y si estos están distribuidos en varios ficheros se puede violar de manera más fácil.
- ♦ **Problemas de atomicidad:** ciertos conjuntos de operaciones tienen que ser atómicas (ocurrir completas o no ocurrir). Esto es difícil de asegurar con archivos.
- ♦ **Anomalías en acceso concurrentes:** Múltiples usuarios a los mismos datos → inconsistencia.
- ♦ **Problemas de seguridad:** Es difícil asegurar que solo accedan unos determinados usuarios a los ficheros.