

# Tema 3



Dpto. de Métodos Cuantitativos e Informáticos  
Facultad de Ciencias de la Empresa. UPCT



## Tema 3

### El almacenamiento de la información: Ficheros y GBD

*Herramientas de Informática de Gestión (2ºB LADE)*



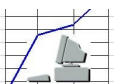
Dpto. de Métodos Cuantitativos e Informáticos  
Facultad de Ciencias de la Empresa. UPCT



## Información

- La información es la materia prima de toda organización
- Información de calidad = Información útil
- procesar la información: eficacia y eficiencia
- Burótica: Conjunto de fases de producción, distribución y explotación de la información a todos los niveles de la organización (Jean Martineau)

*Herramientas de Informática de Gestión (2ºB LADE)*



# Tema 3



## Administración de datos

- Método tradicional de archivo
  - Desventajas
    - Dependencia entre programa y datos
    - Redundancia de datos
    - Integridad de los datos
- Bases de datos
  - Gestores de bases de datos (DBMS)
    - Consultas: Solicitud de datos de campos específicos
    - Seguridad: Diferentes vistas para los usuarios según su perfil/permisos



## Administración de datos

Position number:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39		
Record 1:	0	2	8	3	4	5	6	7	7	R	a	p	a	p	o	r	t					J	o	s	e	p	h	i	n	a	0	2	2	7	6	5	2	3	3	0	
Record 2:	3	7	6	3	4	3	4	5	5	J	o	h	n	s	o	n					K	e	v	i	n				1	2	0	3	6	1	2	3	3	0			
Record 3:	6	7	7	5	4	0	0	9	8	S	t	e	p	h	a	n	o	p	u		A	n	g	e	l	o			1	0	1	0	6	4	2	3	3	1			
.																																									
.																																									
.																																									

Social Security number      Last name      First name      Date of birth      Department number

Presentación de un archivo personal en un archivo tradicional



# Tema 3



## Administración de datos

<i>STUDENT RECORD INFORMATION</i>		
<i>REGISTRAR'S OFFICE</i>	<i>FINANCIAL AID OFFICE</i>	<i>HOUSING OFFICE</i>
last name	last name	last name
first name	first name	first name
SSN	SSN	SSN
home address	home address	home address
major	loan amount	campus address
grades		monthly rent

Información en el registro de un estudiante almacenada en tres lugares diferentes



## Vocabulario

- **Dato (Data)**  
Es el hecho, noción o instrucción representados bajo un fórmula convencional, apropiada para su comunicación, interpretación o tratamiento por el hombre o por medios automáticos. Mínima cantidad de información no elaborada, que no tiene sentido por sí misma, almacenada en un dispositivo magnético y que se puede utilizar para realizar cálculos o para la toma de decisiones. Ejemplo: Antonio Pérez.
- **Información**  
Es la asignación que el hombre atribuye a los datos por medios de convenciones conocidas y utilizadas en su representación. La información se encuentra contenida en "documentos" que pueden contener una o varias unidades de información. Ejemplo: Antonio Pérez es un alumno universitario.
- **Estructura de datos (Data Structure)**  
Conjunto de datos relacionados entre sí, en el cual se pueden llevar a cabo una serie de operaciones previamente definidas: búsqueda, acceso, clasificación, etc. Ejemplo: Pilas, colas, listas, matrices, etc.



# Tema 3



## Vocabulario

- **Campo (Field)**  
Es la unidad más pequeña de datos que se puede utilizar para describir una unidad de información. En bases de datos, espacio reservado para un dato dentro de un registro.
- **Campo clave (Key Field)**  
Campo de una base de datos, que se utiliza para la clasificación, localización o acceso a los distintos datos de la misma. Una base de datos puede tener varios campos clave. Ejemplo: DNI, número de expediente, etc.
- **Registro (Record)**  
Datos de un fichero compuesto por campos relacionados.
- **Índice (Index)**  
Tabla utilizada para definir la posición de un registro; se basa en los campos clave.



## Vocabulario

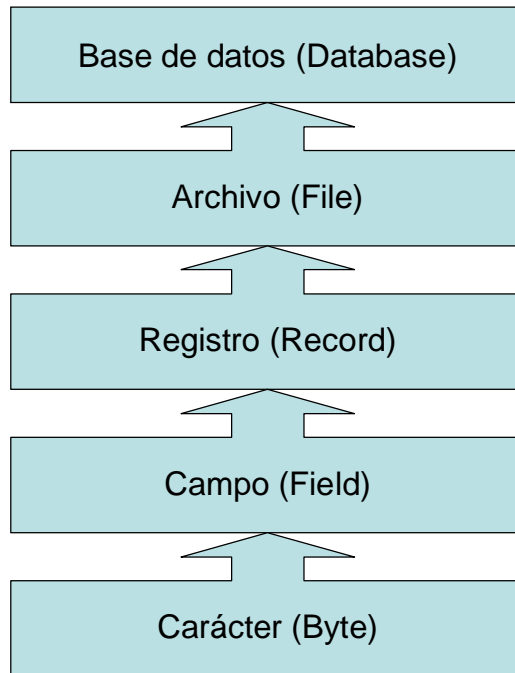
- **Fichero (File)**  
Conjunto organizado de datos que tienen entre sí una relación lógica y que están memorizados en medio adecuado para su posterior recuperación; están compuestos de registros considerados como una sola unidad. También se aplica como sinónimo de documento de texto, hoja de cálculo, base de datos, etc.
- **Base de datos (Database)**  
Conjunto de datos relacionados entre sí, y organizados de tal manera que proporcionan una base para su utilización efectiva: recuperación, modificación, toma de decisiones y creación de informes. Está compuesta por ficheros relacionados entre sí.
- **Gestor de base de datos (GBD) (Data Base Management System)**  
Programa utilizado para construir una base de datos y mantenerla. Hace todo el trabajo: estructuración de archivos, almacenamiento de datos y vinculación de registros.
- **Sistema de gestión de base de datos textual (SGBDT)**  
Almacenamiento, indexación, consulta y modificación de información textual



# Tema 3



## Jerarquía de datos



## Jerarquía de datos

Data Level	Example
Character	223287695 Doe John 1978
Field	223287695 Doe John 1978
Record	200987845 Jewel Mark 1978 223287695 Doe John 1978 249876587 Smith Justin 1978
File	200987845 Jewel Mark 1978 223287695 Doe John 1978 349876587 Smith Justin 1977 410098456 Jones Jose 1976 ...
Database	200987845 Jewel Mark 1978 223287695 Doe John 1978 349876587 Smith Justin 1977 410098456 Jones Jose 1976 ... ACC Dor Avi 9-8776 MKT Jenings Rich 9-8776 FIN Dor Jim 9-8776 ...

Annotations in the table:

- SSN** points to the first column of the Record level.
- Last name** points to the second column of the File level.
- First name** points to the third column of the File level.
- Year of birth** points to the fourth column of the Database level.
- Department** points to the first column of the Database level.
- Campus phone number** points to the fifth column of the Database level.
- Student file** groups the Record and File levels.
- Professor file** groups the last three rows of the Database level.

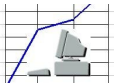


# Tema 3



## Método tradicional vs bases de datos

- Ventajas del método tradicional
  - Simple
  - Eficiente
  - Personalización
- Ventajas de las bases de datos
  - Reducida redundancia de datos
  - Independencia aplicación/datos
  - Mejor control
  - Flexibilidad



## Tipo de archivos según el acceso

- **Fichero secuencial (Sequential File)**
  - Fichero en el que el almacenamiento se produce siguiendo la técnica FIFO. Conforme se introducen los registros se van grabando en el soporte magnético.
  - Para localizar un registro dentro de este tipo de ficheros es necesario seguir una secuencia desde el principio uno a uno; es decir, el tiempo de acceso a un registro es función de su distancia con la posición actual del puntero.
  - La longitud del registro suele ser variable.
  - Se utilizan el casos de no gran volumen de datos y poca actualización, caso de los ficheros denominados descriptores, índices, diarios, etc.

1025	Banco de Santander	250	CajaMurcia	. . . . .	
R0	R1	R2	R3	Ri	Rn







## Tipo de archivos según el acceso

- **Fichero de acceso aleatorio (Random Access File)**

- Fichero en el cual los registros no se encuentran almacenados de forma secuencial en un dispositivo secundario, sino que se almacenan según un algoritmo concreto, por lo cual se puede acceder a cualquier registro sin pasar por los demás.
- El tiempo de acceso es el mismo para todos los registros.
- La longitud de registro debe ser constante, para que su posición quede unívocamente determinada por los dos parámetros de número de registro y longitud del mismo.
- Se utiliza para los ficheros maestros de gran cantidad de datos y alto índice de movimiento.

	C1	C2	C3	C4	C <sub>i</sub>	C <sub>m</sub>
R0	DNI	Apellidos, nombre	Edad	Dirección		
R1						
R2						
...						
...						
R <sub>i</sub>						
...						
...						
R <sub>n</sub>						



## Tipo de archivos según el acceso

- **Fichero indexado (Indexed File)**

- Fichero de acceso aleatorio que además de los datos de información propios de cualquier tipo de fichero, incluye un índice de claves y direcciones que indican la posición exacta de cada registro por medio del campo clave que lo identifique.
- Por tanto consta de uno o más ficheros secuenciales o índices y otro aleatorio que hace de maestro.
- Los índices deben de estar ordenados para una rápida búsqueda y una correcta "gestión de huecos".
- Búsqueda dicotómica.



# Tema 3



## Tipo de archivos según el acceso

	c1	c2
r0	Apellidos + nombre	Núm. registro
r1		
r2		
...	López Mas, Julio	10
...		
n		
...		
m		

Fichero índice ordenado  
(Secuencial)

Fichero maestro  
(Acceso aleatorio)

	C1	C2	C3	C4	Ci	Cm
R0	DNI	Apellidos, nombre	Edad	Dirección		
R1						
R2						
...						
R10	12345678Z	López Mas, Julio				
...						
...						
Rn						



## Modelos de base de datos

- Modelo jerárquico
  - Patrón de árbol invertido
  - Registros padres y registros hijos
  - Los registros hijos sólo tienen un padre



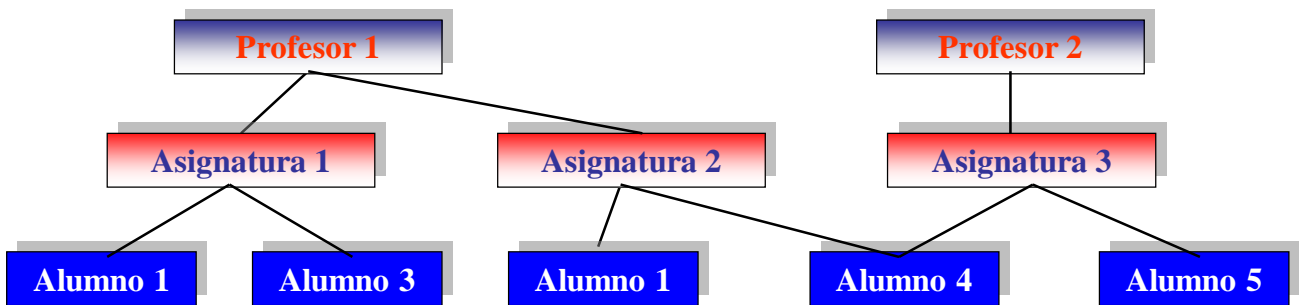


# Tema 3



## Modelos de base de datos

- Modelo en red
  - Capacidad para almacenar un registro una sola vez en la base de datos
  - Un registro hijo puede tener más de un padre
  - Complicadas de construir y mantener



## Modelos de base de datos

- Modelo relacional
  - Consiste en tablas.
  - Los enlaces entre tablas se mantienen con claves externas
  - Ventajas
    - Reducida duplicidad de datos
    - Más fácil de conceptualizar y mantener que las bases de datos jerárquicas y en red
    - Todos los gestores existentes para microordenadores permiten el modelo relacional

FILA	Columna 1 PROFESOR	Columna 2 ASIGNATURA	Columna 3 ALUMNO
1	Profesor 1	Asignatura 1	Alumno 1
2	Profesor 1	Asignatura 1	Alumno 2
3	Profesor 1	Asignatura 1	Alumno 3
4	Profesor 1	Asignatura 2	Alumno 2
5	Profesor 1	Asignatura 2	Alumno 4
6	Profesor 2	Asignatura 3	Alumno 4
7	Profesor 2	Asignatura 3	Alumno 1
8	Profesor 2	Asignatura 3	Alumno 5
9	Profesor 2	Asignatura 3	.....



# Tema 3



## Modelos de base de datos

	DATABASE MODEL		
	HIERARCHICAL	NETWORK	RELATIONAL
CONCEPTUALIZATION	Moderately easy	Difficult	Easy
EASE OF DESIGN	Very difficult	Moderately difficult	Difficult
EASE OF MAINTENANCE	Difficult	Very difficult	Easy
DATA REDUNDANCY	High	Low	High
EASE OF USE	Moderate	Low	High

### Ventajas y desventajas de los modelos de base de datos

*Herramientas de Informática de Gestión (2ºB LADE)*



## Modelos de base de datos

MODELO	PRODUCTO	VENDEDOR	HARDWARE
<b>Jerárquica</b>	Focus	Information Builders International	Mainframe/PC
	IMS	IBM	Mainframe
	Ramis	Online Software International	Mainframe
<b>En red</b>	ADABAS	Software AG of North America	Mainframe
	Image	Hewlett-Packard	Mainframe
<b>Relacional</b>	Access	Microsoft	PC
	DB2	IBM	Mainframe
	dBASE V	Borland International	PC
	EDA/SQL	Information Builders International	PC
	FoxPro	Microsoft	PC
	Ingres	Ask Group	PC
	NOMAD	Must Software International	Mainframe/PC
	Oracle	Oracle	Mainframe/PC
	Paradox	Borland International	PC
	Rbase	Microrim	PC
	SQL/DS	IBM	Mainframe
<b>Objeto-Relacional</b>	SQL Server	Microsoft	PC
	Sybase	Sybase	PC
	ObjectStore	Object Design	PC
	Universal Server	Informix	PC
	Illustra	Informix	PC

### Gestores de bases de datos

*Herramientas de Informática de Gestión (2ºB LADE)*





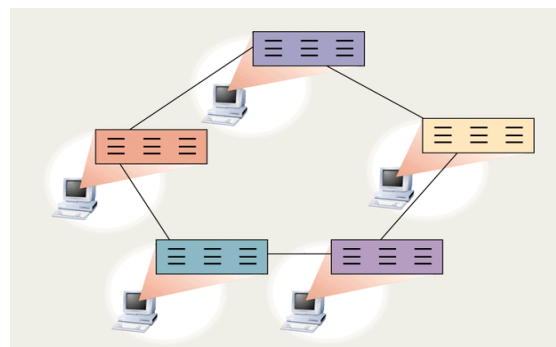
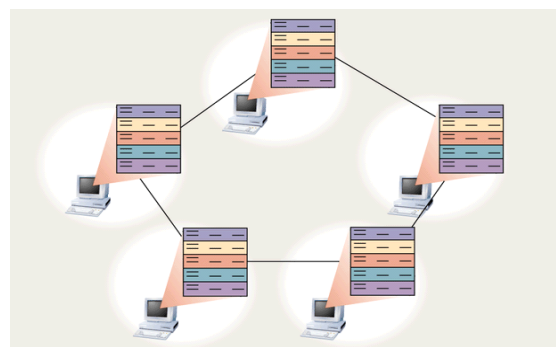
## Structured Query Language

- SQL
  - Estándar internacional como lenguaje de definición de datos y lenguaje de manipulación de datos para bases de datos relacionales
- Ventajas de SQL
  - Los usuarios no necesitan aprender diferentes lenguajes.
  - SQL se puede integrar en lenguajes de tercera generación aumentando la eficiencia y efectividad.
  - Los programadores no tienen que reescribir código ya que las instrucciones SQL son portables.



## Bases de datos distribuidas

- **Duplicación**
  - Hay copias completas de la base de datos en varios lugares
- **Fragmentación**
  - Las partes de la base de datos se almacenan en los lugares donde su acceso es más frecuente





## Data Warehousing

- **Almacén de datos inteligente (Data warehouse)**
  - Colección de datos que dan soporte a la administración para la toma de decisiones
  - No reemplazan a las bases de datos, sino que se nutren de ellas y de datos externos
- **Fases en la construcción de un Data Warehouse**
  - Fase de extracción
  - Fase de depuración
  - Fase de carga

*Herramientas de Informática de Gestión (2ºB LADE)*



## Data Warehousing



*Herramientas de Informática de Gestión (2ºB LADE)*





## Minería de datos

- **Minería de datos (Data Mining)**
  - Combina técnicas estadísticas y de inteligencia artificial para manipular grandes volúmenes de información y localizar patrones y relaciones desconocidas entre datos

DATA-MINING APPLICATION	DESCRIPTION
Market segmentation	Identify the common characteristics of customers who buy the same products from your company
Customer churn	Predict which customers are likely to leave your company and go to a competitor
Fraud detection	Identify which transactions are most likely to be fraudulent
Direct marketing	Identify which prospective clients should be included in a mailing list to obtain the highest response rate
Interactive marketing	Predict what each individual accessing a Web site is most likely to be interested in seeing
Market basket analysis	Understand what products or services are commonly purchased together
Trend analysis	Reveal the difference between a typical customer this month and a typical customer last month

Source: Datamind Corp., quoted in McCarthy, V., "Strike it rich!", *Datamation*, February 1997, p. 47.



## Gestión del conocimiento

- **Conocimiento**
  - Información que tenemos de un tema o información de dónde encontrarla
- **Gestión del conocimiento**
  - Transferir el conocimiento a bases de datos
  - Filtrar y separar el conocimiento más relevante
  - Organizar el conocimiento en bases de datos que
    - Permitan a otros empleados acceder fácilmente al conocimiento
    - Proporcionar el conocimiento específico a los empleados en función de sus necesidades





## Gestión documental

### Las organizaciones precisan hoy en día herramientas informáticas y telemáticas que les proporcionen:

- Más competitividad en la gestión.
- Ciclos más cortos de proceso y eliminación de elementos innecesarios.
- Mejor calidad de servicio.
- Mayor coordinación, comunicación y cooperación.
- Reducción de costes de papel y de almacenamiento físico.

### Para ello existen:

- Archivos Electrónicos de Datos (AED)
- Gestores de Bases de Datos Documentales (GBDD)
- Controladores de Flujos de Trabajo (WorkFlow)
- **Gestor Integral de Documentos (GID): Integra las tres funcionalidades anteriores.**

