



# Tema 5: Organización de los Datos

## Informática

**Departamento de Informática**

**Israel González Carrasco**

**(israel.gonzalez@uc3m.es)**

**María Belén Ruiz Mezcua**

**(mbelen.ruiz@uc3m.es)**

# Contenido

---

1. Objetivos.
2. Introducción.
3. Tipos de Organización.
4. Tratamiento de Información no Estructurada.
5. Tratamiento de Información Estructurada.

# Objetivos

- Comprender la necesidad y utilidad de organizar y estructurar la información para almacenarla y recuperarla.
- Reconocer los diferentes tipos de datos, bases de datos y sistemas de recuperación con las que se pueden manejar la información.
- Ser capaces de entender y manejar una base de datos en Microsoft Access a nivel usuario.

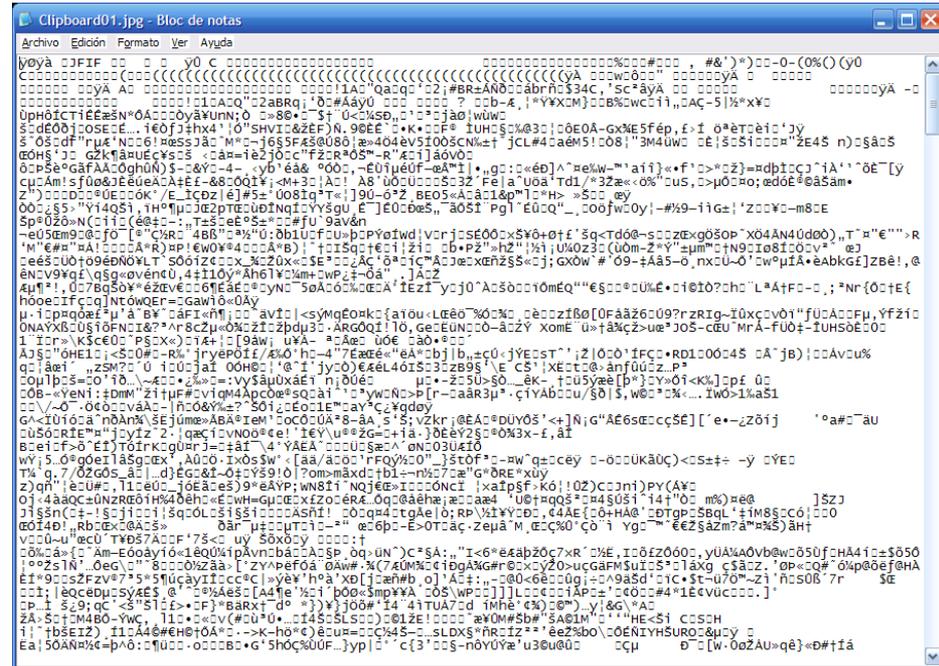
# Introducción

- Una organización de datos es una estructura física-lógica que permite realizar operaciones computacionales (editar, guardar, actualizar, etc.) sobre un contenido de información.
  - Estructura física; describe la manera física (bytes) de almacenar los datos en un soporte (como se almacenan los datos en el soporte).
  - Estructura lógica; describe la manera lógica de representar la información a los usuarios (como ve el usuario la información).

# Introducción (II)

## Representación lógica

## Almacenamiento físico



# Tipos de Organización

- Organización no estructurada
  - Utiliza información no estructurada contenida en libros, artículos informes.
  - Es tan grande la variedad de información que es muy difícil saber que es lo que nos va a interesar en el caso de una búsqueda y sería imposible determinar un campo por cada tema.
- Organización estructurada
  - Utiliza información estructurada o datos definidos (facturas, recibos de clientes, etc.).
  - El diseño de una buena base de datos deberá reconocer con método y exactitud los datos que se van a utilizar, antes de su implementación.
  - Esta organizada mediante campos.

# Tipos de Organización (II)

- Tratamiento de datos formateados (organizados).
  - Sistemas orientados a proceso.
  - Sistemas orientados a datos.
    - ✓ Sistema de Gestión de Datos Relacionales.
  - Tratamiento de datos no formateados (No organizados)
    - Documentos.
    - Sistemas de Recuperación de Información.
      - ✓ Sistemas de Gestión de Bases de Datos Documentales.



# Tratamiento de la Información Estructurada

## Índice

- Sistemas orientados al proceso.
- Sistemas orientados a los datos.
- Bases de Datos Relacionales (BDR).
- Sistema Gestor de Bases de Datos Relacionales (SGBDR).

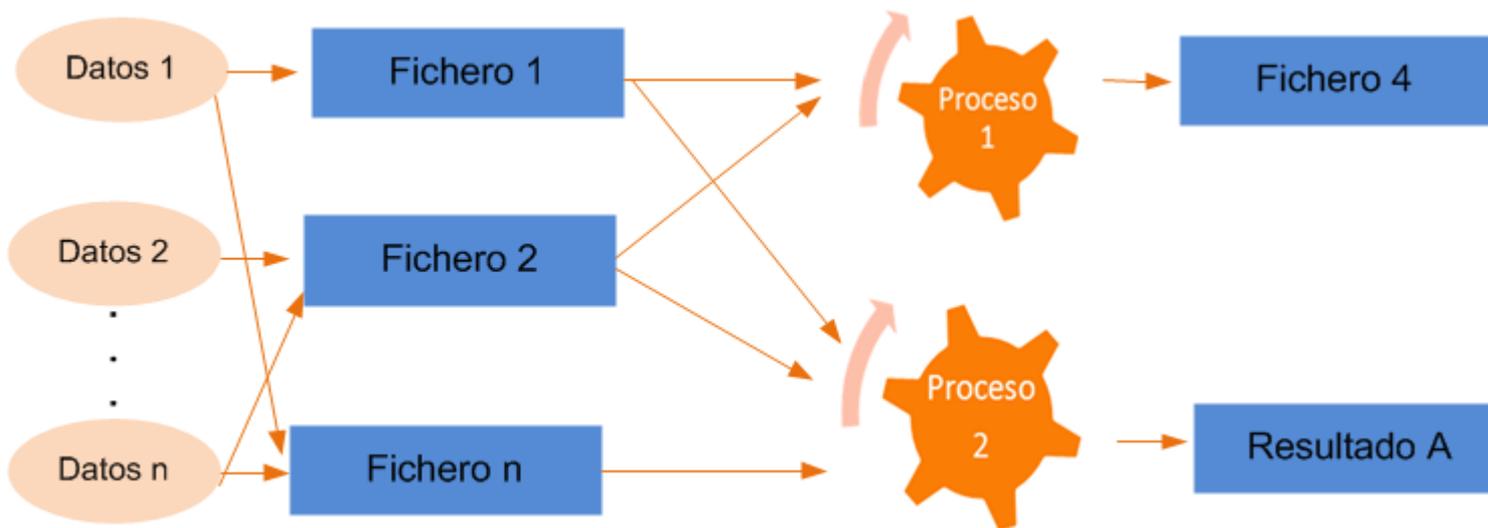
# Tratamiento de la Información Estructurada

## Sistemas orientados al proceso

- Son los sistemas tradicionales.
- En ellos se pone más énfasis en los tratamientos que en los datos.
- Los datos se encuentran repetidos en diversos ficheros tratados de forma independiente por cada aplicación.
- En cada Programa hace falta definir con que datos trabaja y en que fichero están.

# Tratamiento de la Información Estructurada

## Sistemas orientados al proceso (II)



# Tratamiento de la Información Estructurada

## Sistemas orientados al proceso (III)

- Ejemplo de Sistemas orientados al proceso.
  - Departamento de RRHH utiliza el fichero *Selección*.
  - Campos:

DNI	Fecha Ingreso	Nombre	Formación	Dirección	Especialidad	Puesto	Nivel de conocimiento
-----	---------------	--------	-----------	-----------	--------------	--------	-----------------------

- En una reunión posterior se decide informatizar la *Nómina* de los trabajadores.

*¿Qué soluciones puede adoptar la empresa?*

# Tratamiento de la Información Estructurada

## Sistemas orientados al proceso (IV)

- Posibles Soluciones:
  - Solución 1:
    - ✓ Crear una aplicación para la gestión de las *Nóminas* (con su fichero correspondiente) independiente de la de *Selección*.
  - Solución 2:
    - ✓ Modificar el fichero *Selección* existente para que puedan usarlo ambas aplicaciones.
  - Solución 3:
    - ✓ Crear un nuevo fichero que contenga información específica de la aplicación de gestión de las *Nóminas*.

# Tratamiento de la Información Estructurada

## Sistemas orientados al proceso (V)

- Comentarios:
  - Solución 1:
    - ✓ Esta solución producirá datos duplicados.
    - ✓ Conjuntos de datos iguales para aplicaciones distintas
  - Solución 2:
    - ✓ Implica modificar la aplicación de Selección.
  - Solución 3:
    - ✓ En el nuevo fichero hay que incluir “enlaces” al primer fichero para usar los datos comunes.
    - ✓ Al existir dos ficheros el proceso es más lento.

# Tratamiento de la Información Estructurada

## Sistemas orientados al proceso (VI)

- Problemas:
  - Dificultad de mantenimiento.
  - Excesiva dependencia del soporte físico.
  - Dependencia de los datos respecto de los programas.
  - Dificultad para montar sistemas de información orientados a la toma de decisiones.
  - Los ficheros están muy orientados a cada aplicación.

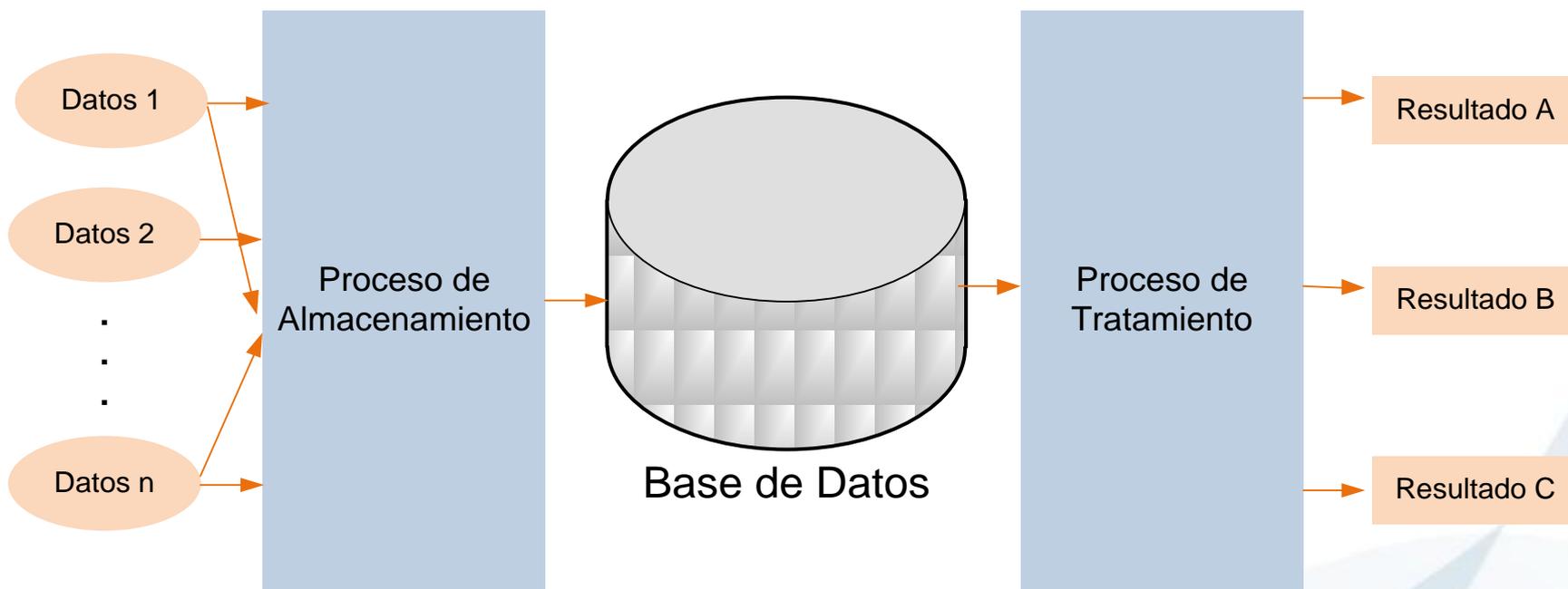
# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos

- Tratan de resolver los problemas del enfoque tradicional (Sistemas orientados al proceso).
- Se fundamentan en la utilización de una Base de Datos que sustituye todos los ficheros por una única colección de datos que puede ser utilizada por todas las aplicaciones.

# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (II)



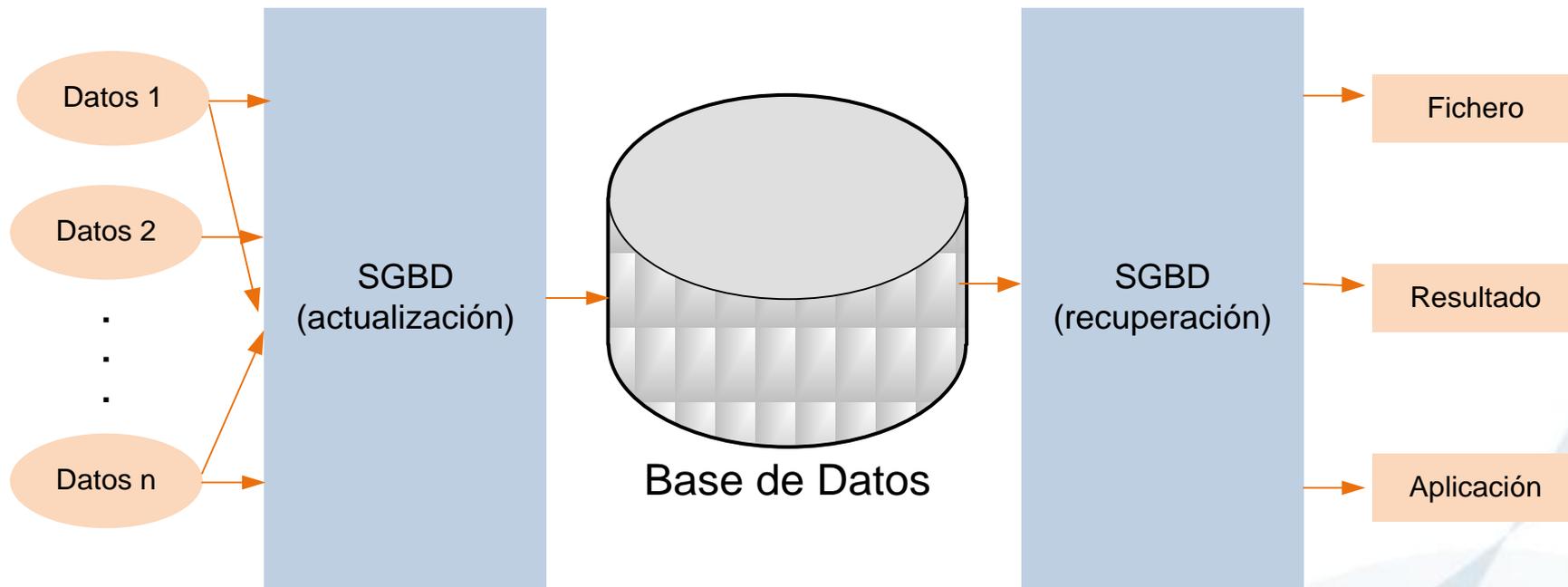
# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (III)

- Conceptos Básicos:
  - Bases de Datos
    - ✓ Una base de datos es un conjunto de información estructurada en registros y almacenada en un soporte electrónico legible desde un ordenador.
    - ✓ Cada registro constituye una unidad autónoma de información que puede estar a su vez estructurada en diferentes campos o tipos de datos que se recogen en dicha base de datos.
  - Sistema Gestor de Bases de Datos (SGBD).
    - ✓ Surgieron para sistematizar el tratamiento de grandes volúmenes de información con alto grado de homogeneidad y compartida por varios usuarios.

# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (IV)



# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (V)

- Base de datos:
  - Colección o depósito de datos integrados.
  - Con redundancia controlada.
  - Su estructura refleja las interrelaciones y restricciones del mundo real.
  - Los datos son independientes de la aplicación o del usuario.
  - Los datos tendrán definición y descripción únicas (y almacenada con ellos).
  - Los procedimientos que involucre tendrán que preservar su integridad, respetando además unas normas de disponibilidad y confidencialidad.

# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (VI)

- Conjunto coordinado de herramientas que proporciona los medios necesarios para interactuar con la BD a todos los niveles.
  - Herramientas: programas, procedimientos, lenguajes, etc.
  - Interaccionar con la base: describir, recuperar y manipular datos almacenados en la base, preservando su integridad, confidencialidad, y seguridad.
  - A todos los niveles: usuario, programador, analista, etc.
- Funciones esenciales de un SGBD:
  - Descripción.
  - Manipulación.
  - Utilización.

# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (VII)

- Un SGBD debe disponer de lenguajes y procedimientos que posibiliten la interacción con la BD.
  - LDD (Lenguaje de Definición de Datos).
    - ✓ Para la definición de las estructuras de datos a todos los niveles (externo, conceptual e interno).
  - LMD (Lenguaje de Manipulación de Datos).
    - ✓ Permite la ejecución de operaciones de recuperación y actualización sobre un conjunto de registros (identificado a través de un criterio de selección), indicando las estructuras externas sobre las que se actúa.
  - LCD (Lenguaje de Control de Datos).
    - ✓ El administrador de la BD utiliza este lenguaje para especificar los aspectos de seguridad física así como de protección frente a accesos no permitidos.
  - Procedimientos de administración: copias de seguridad, estadísticas, carga de datos, ...

# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (VIII)

- Niveles de Abstracción en una BD.
  - En un sistema de almacenamiento de datos, se distinguen siempre los niveles lógico (dato-usuario) y físico (dato-soporte).
  - Esta separación permite solventar varias problemáticas.
  - La idea de Base de Datos permite abstraer aún más, apareciendo un tercer nivel: lógico global (aúna los niveles lógicos de todos los usuarios).
  - Este nivel contendrá (esquema lógico global):
    - ✓ Descripción de datos e interrelaciones entre ellos
    - ✓ Restricciones de integridad y confidencialidad

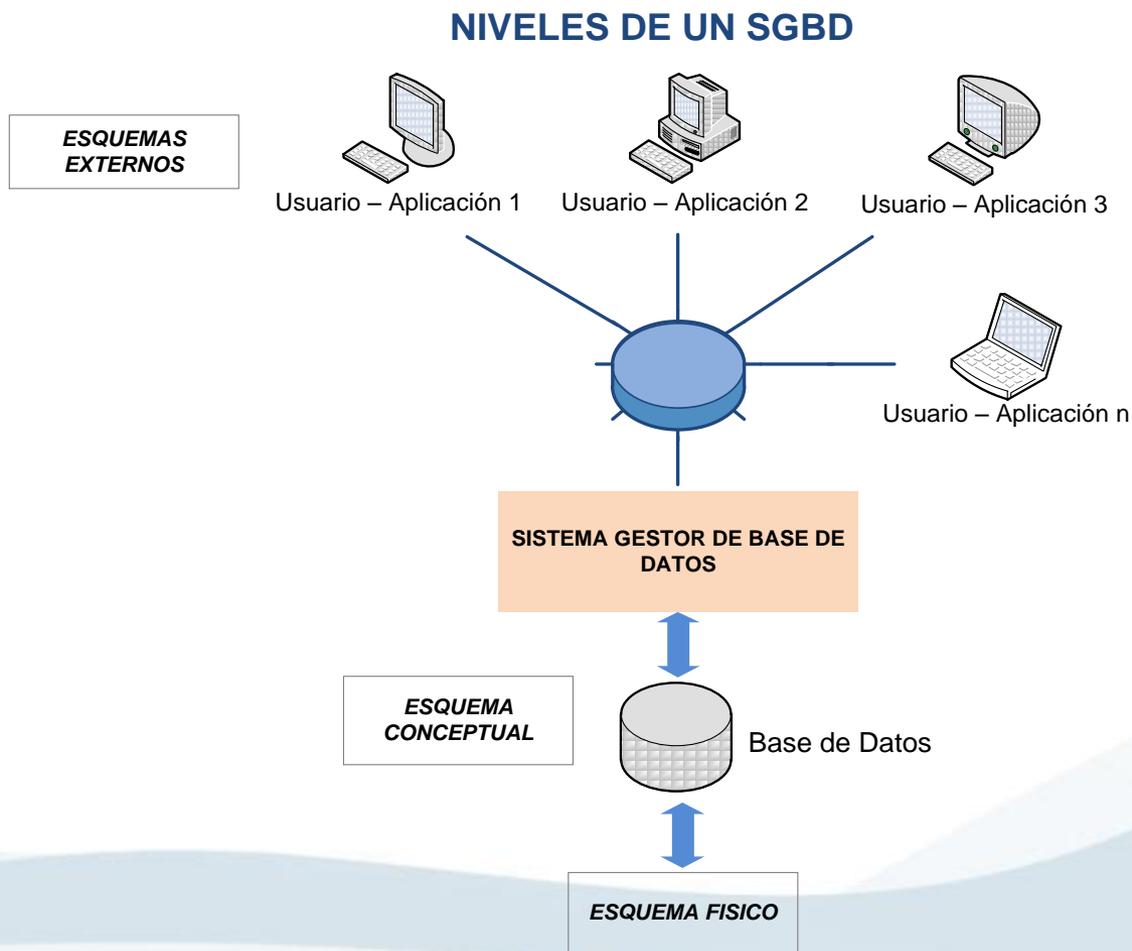
# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (IX)

- Niveles de Abstracción en un SGBD:
  - Arquitectura a tres niveles: Interno, Conceptual y Externo.
  - Nivel Interno:
    - ✓ Es el nivel más cercano a la máquina.
    - ✓ Describe cómo y dónde se almacena un conjunto determinado de datos.
  - Nivel Conceptual:
    - ✓ *Proporciona un eslabón intermedio entre la visión de la BD que se ofrece al usuario y el nivel de almacenamiento físico.*
    - ✓ *Es una representación del contenido total de la BD.*
  - Nivel Externo:
    - ✓ Es el nivel más cercano al usuario.
    - ✓ Cada usuario está interesado en un subconjunto concreto de la Base de Datos Global.

# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (X)



# Tratamiento de la Información Estructurada

## Sistemas orientados a los datos (XI)

- Existen varios modelos de SGBD:
  - Definen como se definen las estructuras y operaciones que se permiten en ellas.
- Modelo Relacional:
  - Basado en la noción matemática de Relación.
  - Propuesto por Codd.
  - El mas extendido en la actualidad
  - Ejemplos de SGBD relacionales comerciales:
    - ✓ Microsoft Access.
    - ✓ Oracle.
    - ✓ Microsoft SQL Server.
    - ✓ BD2.
    - ✓ MySQL.
    - ✓ Informix.

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR)

- Colección o depósito de datos integrados, almacenados en soporte secundario cuyos datos son compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de ellos.
- Los procedimientos de actualización y recuperación facilitarán el conjunto de los datos.
- Una base de datos relacional permite almacenar, relacionar y acceder a la información de la forma más estructurada posible.
- La información aparece estructurada en campos de longitud fija.

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR) (II)

- Es un modelo de datos extremadamente simple y claro, que también ha resultado potente para la mayor parte de las aplicaciones de BDs.
- Sus principales características son:
  - Independencia de datos y programas.
  - Integración de diferentes aplicaciones.
  - Escasa o nula redundancia.
  - Acceso múltiple.
  - Seguridad.

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Terminología

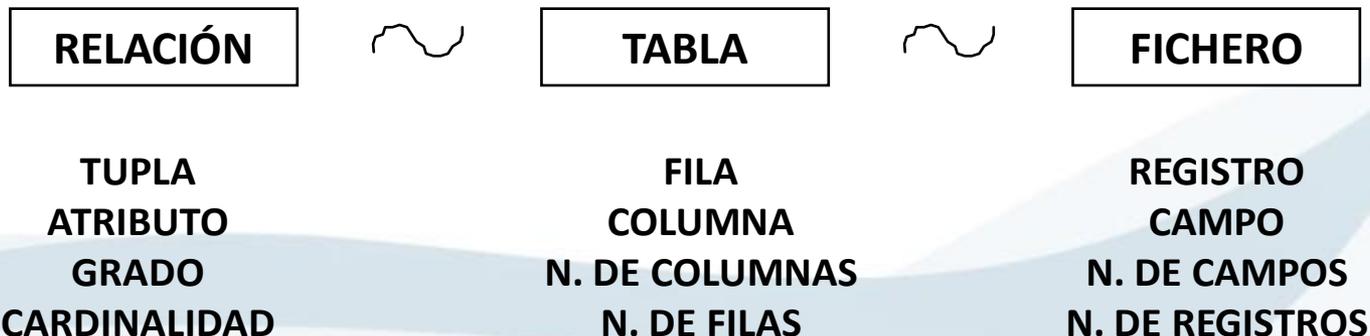
- Los elementos más importantes de las bases de datos relacionales son las tablas (**relaciones**), asociaciones entre ellas (**interrelaciones**), filas (**tuplas**), columnas (**atributos o campos**) y consultas.
- Una **relación** es una colección de datos acerca de un tema en particular.
  - Los datos de la relación se representan en formato tabular en tuplas (**columnas y filas**).
  - Agrupan un conjunto de **atributos** o campos.

Atributo 1	Atributo 2	.....	Atributo n	
XXXXXXXX	XXXXXXXX	XXXXXXXX	XXXXXXXX	Tupla 1
XXXXXXXX	XXXXXXXX	XXXXXXXX	XXXXXXXX	Tupla 2
XXXXXXXX	XXXXXXXX	XXXXXXXX	XXXXXXXX	.
XXXXXXXX	XXXXXXXX	XXXXXXXX	XXXXXXXX	.
XXXXXXXX	XXXXXXXX	XXXXXXXX	XXXXXXXX	Tupla n

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Terminología (II)

- Una consulta son preguntas acerca de los datos en su base de datos.
  - Los datos fuente pueden provenir de una o varias tablas.
  - La consulta reúne la información solicitada.
  - No almacenan la información en un fichero.
- Comparación de la terminología



# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Claves

- Clave de una relación:
  - Una clave candidata de una relación es un conjunto de atributos que identifican unívoca y mínimamente cada tupla de la misma.
- Tipos de Claves:
  - Clave primaria (PRIMARY KEY o PK) permite declarar un atributo o conjunto de atributos como la clave primaria de una tabla.
    - ✓ Conjunto no vacío de atributos que identifican unívoca y mínimamente cada tupla.
    - ✓ Los valores de la clave primaria han de ser distintos y siempre tiene que tener algún valor
    - ✓ Ejemplo: La tabla Alumno está identificada por el atributo NIA (clave primaria).

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Claves (II)

- Tipos de Clave:
  - Clave de unicidad (UNIQUE) nos permite definir claves alternativas (que pueden ser en algún momento PRIMARY KEY).
  - Clave ajena (FOREGEIN KEY o FK) es una clave que referencia una clave de una tabla con otra tabla (crear una interrelación).
    - ✓ Se denomina clave ajena de una relación R2 a un conjunto de atributos cuyos valores han de coincidir con los valores de la clave primaria de una relación R1 (R1 y R2 no son necesariamente distintas).

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Restricciones del modelo

- Restricciones inherentes: impuestas por el propio modelo
  1. En una relación no puede haber dos tuplas iguales (obligatoriedad de clave primaria).
  2. El orden de las tuplas y el de los atributos no es relevante.
  3. Cada atributo sólo puede tomar un único valor del dominio sobre el cual está definido (no hay grupos repetitivos).
  4. Ningún atributo que forme parte de la clave primaria de una relación puede tomar un valor nulo (regla de integridad de entidad).
- Restricciones semánticas (o de usuario): facilidades ofrecidas por el modelo para representar la semántica del mundo real.

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Restricciones del modelo (II)

- Regla de Integridad referencial:
  - Lo valores que tome la clave ajena en la tabla hija han de ser los mismos que tiene en la tabla padre.
  - La clave primaria de la tabla padre ha de tener un campo en la tabla Hija.
  - Clave ajena del mismo tipo datos y que contenga los mismos valores.

# Tratamiento de la Información Estructurada

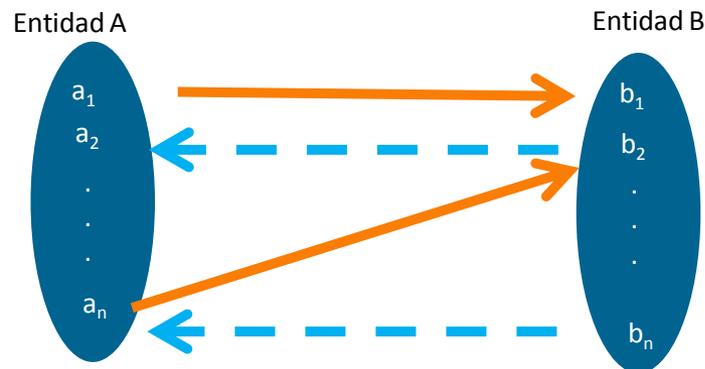
## Bases de Datos Relacionales (BDR). Interrelaciones

- Asociaciones entre tablas:
  - Asociaciones definidas entre los esquemas de relación. No existen vinculaciones físicas, sólo lógicas.
  - Se distinguen distintos tipos según el número máximo de tuplas que intervienen por parte de cada relación.
  - A este número se le denomina cardinalidad de la relación.
  - Pueden ser 1:1, 1:N o N:M.

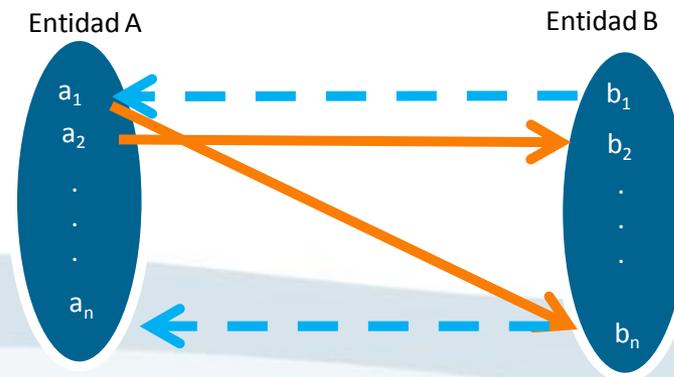
# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Interrelación 1:N

- Correspondencia 1:1



- Correspondencia 1:N



# Tratamiento de la Información Estructurada

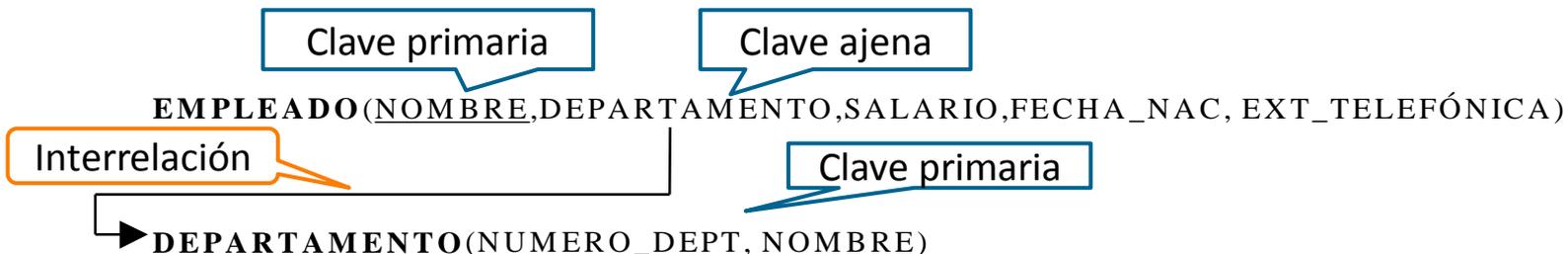
## Bases de Datos Relacionales (BDR). Interrelación 1:N (II)

- *Ejemplo: Empleados asociados a un Departamento de una empresa.*
  - *Cardinalidad 1:N.*
    - ✓ *Un departamento puede tener varios empleados y un empleado solo pertenecer a un departamento.*
  - **Regla de Integridad referencial.**
    - ✓ Lo valores que tome la clave ajena en la tabla hija han de ser los mismos que tiene en la tabla padre.
    - ✓ Los valores que toma la clave ajena DEPARTAMENTO de la tabla EMPLEADO (11,13,14) son los que tiene la clave principal NUMERO\_DEPT en la tabla DEPARTAMENTO.

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Interrelación 1:N (III)

### MODELO RELACIONAL



### TABLAS

#### EMPLEADO

NOMBRE	DEPARTAMENTO	SALARIO	FECHA_NAC	EXT_TELEFÓNICA
Pablo Montero	14	220.000	10-11-67	6543
Beatriz Cristoba	13	300.000	20-9-68	6577
J. Luís Martín	11	150.000	25-6-77	6433
Almudena López	13	350.000	4-5-60	6422
Angel Vallejo	14	400.000	15-4-72	6321
Pedro García	11	200.000	12-3-70	6323

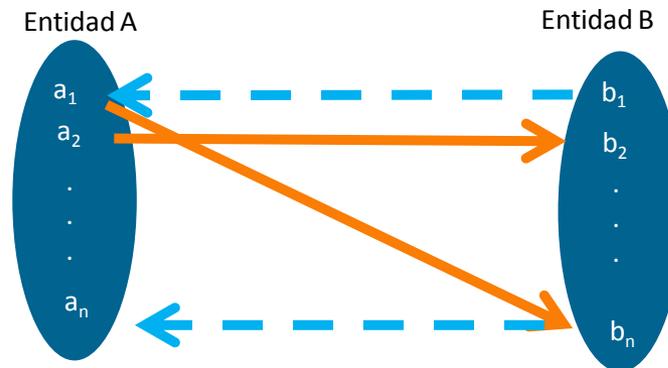
#### DEPARTAMENTO

NUMERO_DEPT	NOMBRE
11	Contabilidad
13	Marketing
14	Informática

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Interrelación N:M

- Correspondencia N:M



- Se crea una nueva tabla (intermedia) entre la entidad A y la entidad B.
- Esta tabla ha de tener como clave principal la que resulta de unir las dos claves principales de las tablas que une.

# Tratamiento de la Información Estructurada

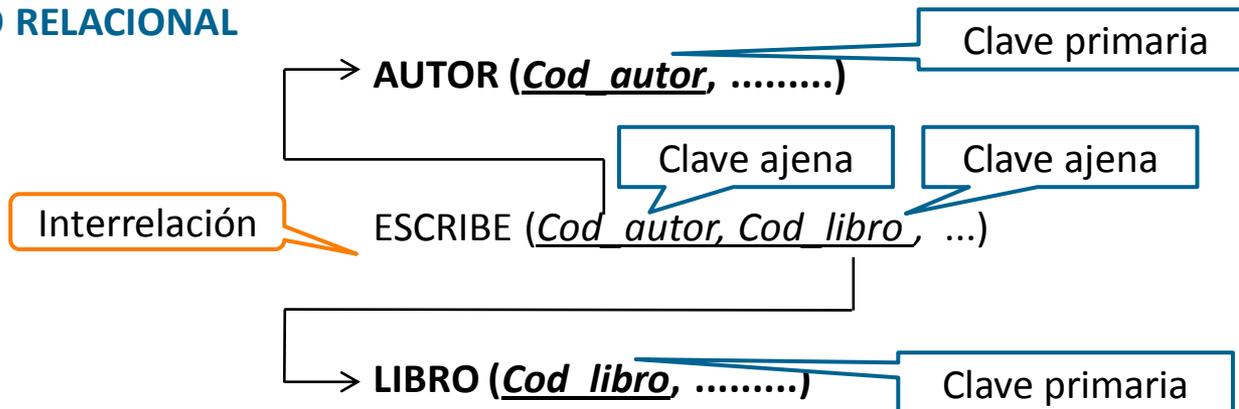
## Bases de Datos Relacionales (BDR). Interrelación N:M (II)

- *Ejemplo: Libros escritos por un Autor.*
  - *Cardinalidad N:M.*
    - ✓ *Un libro puede ser escrito por varios autores y un autor puede escribir varios libros.*
  - Regla de Integridad referencial.
    - ✓ Es necesario crear una nueva tabla (tabla intermedia) ESCRIBE para reflejar la semántica de este tipo de relaciones.
    - ✓ Esta nueva tabla ESCRIBE tiene un clave ajena a cada una de las tablas «padres» (AUTOR y LIBRO).
    - ✓ Los valores de las claves ajenas de la tabla ESCRIBE deben coincidir con los existentes en las tablas «padres».

# Tratamiento de la Información Estructurada

## Bases de Datos Relacionales (BDR). Interrelación N:M (III)

### MODELO RELACIONAL



### TABLAS

#### AUTOR

COD_AUTOR	NOMBRE
1	Arturo Ribagorda
2	Paloma Martinez
3	Elena Castro
4	Dolores Cuadra

#### LIBRO

COD_LIBRO	TITULO
1	Diseño de bases de datos
2	Informática para la empresa y técnicas de programación

#### ESCRIBE

COD_AUTOR	COD_LIBRO
1	2
2	1
3	1
4	1

# Tratamiento de la Información Estructurada

## Sistema Gestor de Bases de Datos Relacionales (SGBDR)

- Herramienta Software que proporciona una interfaz entre los datos almacenados y los programas de aplicación que acceden a éstos.
- Se caracteriza fundamentalmente por permitir una descripción centralizada de los datos y por la posibilidad de definir vistas parciales de los mismos para los diferentes usuarios.
- Conjunto de programas, procedimientos y lenguajes que nos proporcionan las herramientas necesarias para trabajar con una base de datos.
- Incorpora una serie de funciones que nos permite definir los registros, sus campos, sus relaciones, insertar, suprimir, modificar y consultar los datos.

# Tratamiento de la Información Estructurada

## Sistema Gestor de Bases de Datos Relacionales (SGBDR) (II)

- Entre sus funciones:
  - Recuperar la BD, si es dañada
  - Seguridad contra accesos no autorizados
  - Que la base de datos se actualice correctamente cuando varios usuarios la usan concurrentemente
  - Proporciona un catálogo de la descripción de los datos: Diccionario de datos (Relaciones, Usuarios)
  - Integridad : Restricciones y validación de los datos.

# Tratamiento de la Información Estructurada

## Sistema Gestor de Bases de Datos Relacionales (SGBDR) (III)

- Ejemplos de SGBD relacionales comerciales:
  - Microsoft Access.
    - ✓ Integrado en la suite de Office.
  - Microsoft SQL Server.
    - ✓ Integrado en la suite de desarrollo Visual Studio.
  - Oracle.
    - ✓ Líder del mercado.
  - MySQL (en la actualidad de SUN MicroSystems).
    - ✓ Licencia GPL.
  - PostgreSQL
    - ✓ Licencia BSD de código abierto.
  - Informix y DB2 (en la actualidad de IBM).
    - ✓ Poca cuota de mercado.

# Tratamiento de la Información no Estructurada



## Índice

- Informática Documental.
- Base de Datos Documental (BBD ).
- Sistema Gestor de Bases de Datos Documentales (SGBDD)

# Tratamiento de la Información no Estructurada



## Informática Documental

- Trata de la resolución de los problemas que implican el tratamiento de la incertidumbre en la búsqueda de información de documentos.
- Objetivo:
  - Ofrecer herramientas para automatizar el análisis documental como parte fundamental de la documentación.
- Documentación:
  - Disciplina teórica que permite describir de forma científica los documentos.

# Tratamiento de la Información no Estructurada



## Base de Datos Documental (BDD)

- Las BDDs están especialmente concebidas para almacenar y extraer información no estructurada.
- La unidad básica y única para obtener información es el documento.
- Para el acceso de las BDDs se persigue obtener una flexibilidad en la consulta, de forma que el cliente no deba estar restringido a un número limitado de términos de búsqueda.
  - Resolver el problema de incertidumbre en la búsqueda de información

# Tratamiento de la Información no Estructurada



## Base de Datos Documental (BBD) (II)

- Características:
  - Mantienen una base de datos de información textual de longitud variable.
  - Son capaces de recuperar documentos selectivamente a partir de cualquiera de los aspectos de su contenido.
  - Los documentos se representan en la base de datos en forma de registros.
  - Se utilizan descriptores para acceder al contenido y contemplan la gestión de sinónimos.

# Tratamiento de la Información no Estructurada



## Bases de Datos Documentales (BBDs ) (II)

- Funcionamiento interno
  - Disponen de un sistema de indexación por el que el SGBD genera índices en cada campo por cada elemento, es decir cada palabra es indexada.
  - Para palabras posteriores el sistema SGBD lee textos y comprueba si han sido almacenadas en el diccionario con lo cual las almacena una sola vez con un puntero al documento y situación exacta, donde la palabra se repite.
  - La localización de una palabra se hace por consulta en el índice.

# Tratamiento de la Información no Estructurada



## Sistema Gestor de Bases de Datos Documentales (SGBDD)

- Los sistemas de gestión documental permiten la elaboración de diccionarios o índices alfabéticos.
  - Permiten la indexación de textos completos.
- Para ello recurren a diferentes técnicas o heurísticas:
  - Diccionarios terminológicos.
    - ✓ Enuncian y definen los términos propios de una ciencia, una técnica u otra actividad.
  - Tesauro (Thesaurus).
    - ✓ Listado de palabras o términos empleados para representar conceptos.
  - Estrategias de búsqueda.
    - ✓ Conjunto de procedimientos y operaciones que un usuario realiza con el fin de obtener una determinada información.

# Tratamiento de la Información no Estructurada



## Sistema Gestor de Bases de Datos Documentales (SGBDD) (II)

- Tienen en común una serie de funciones especialmente diseñadas para resolver los problemas típicos de un entorno documental:
  - Lenguaje de definición de datos
  - Lenguaje de interrogación de la base de datos
  - Hipertexto
  - Mecanismos de seguridad
  - Facilidades para introducción de información
  - Entorno amigable y facilidad de uso

# Tratamiento de la Información no Estructurada



## Sistema Gestor de Bases de Datos Documentales (SGBDD) (III)

- Lenguaje de definición de datos
  - Permite al diseñador de la base de datos especificar las características estructurales de la misma: Campos que la componen, restricciones y mecanismos de validación asociados a ellos.
- Lenguaje de interrogación de la base de datos
  - Tiene una sintaxis que lo caracteriza, y que precisa cuáles son las frases correctas que se puedan escribir en él.
- Hipertexto
  - Enlaces que permiten moverse por los documentos de múltiples formas, utilizando referencias cruzadas.

# Tratamiento de la Información no Estructurada



## Sistema Gestor de Bases de Datos Documentales (SGBDD) (IV)

- Mecanismos de seguridad
  - Permite establecer diferentes categorías de usuarios en función de los distintos tipos de privilegios que tengan sobre los documentos.
- Facilidades para introducción de información
  - Una vez generada la base de datos, se carga luego la información, que se realiza por:
    - ✓ Entrada directa de la información
    - ✓ Entrada diferida, trabajando sobre un fichero auxiliar
- Entorno amigable y facilidad de uso
  - Lo más importante es que las BDDs sea fácil de utilizar.