Unidad 1: Introducción

BBDD01, Sesión 1:
Introducción a las bases de datos.

Oscar Gutierrez Blanco
Jose A. Gutierrez de Mesa
Javier Albert Segui

Universidad de Alcalá

INDICE

- Introducción
- Niveles de abstracción
- Instancias y esquemas
- Modelos de datos
- Lenguajes de un SGBD
- Estructura de un SGBD
- Usuarios de un BD
- Sistemas informáticos
- Proceso de diseño en BBDD

Referencias: Silberschatz 4ª Ed. pp 1-16

Introducción

- BD: Colección de datos interrelacionados con el objetivo de <u>integrar</u> y <u>compartir</u>
- SGBD: Conjunto de programas que acceden a los datos
- SGBD + BD contienen información sobre una parcela de la realidad
- SGBD proporciona un entorno conveniente y eficiente para usar
- Aplicaciones Bases de Datos:
  - Banca. Transacciones
  - Líneas Aéreas: reserva de billetes
  - Universidades: registros, calificaciones
  - Ventas: productos, clientes, proveedores, etc.
  - Fabricación: inventario, producción, etc.
  - Recursos Humanos: salarios, registros empleados, etc.
- Bases de Datos toca todos aspectos de nuestras vidas.

Introducción

1950-1980? ⇒ Datos sobre ficheros

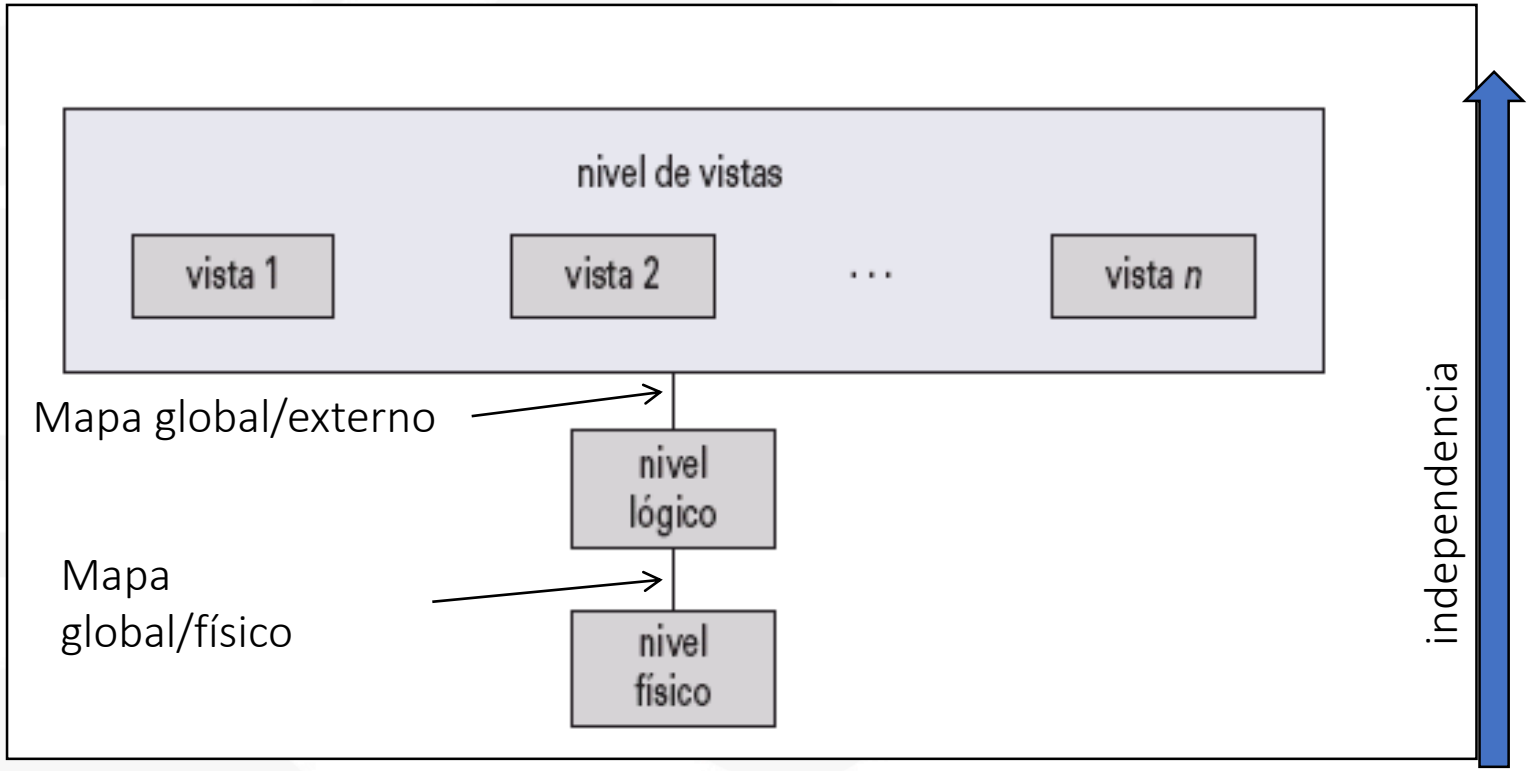
Características:

- Redundancia e Inconsistencias de los datos: Múltiple formatos y duplicidad de información
- Dificultad para acceder a los datos: escribir un nuevo programa para una nueva tarea
- Aislamiento de Datos: múltiple formatos y ficheros
- Problemas de integridad (contenidas en el programa) :
  - Restricciones de integridad (balance de una cuenta → 0)
  - Difícil añadir restricciones o cambiarlas
- Atomicidad de las modificaciones: fallos pueden producir inconsistencias
- Concurrencia de múltiples usuarios: necesario para incrementar rendimiento y controlar el acceso
- Problemas de seguridad

Solución ⇒ Utilización de Sistemas de Bases de Datos

Niveles de abstracción

- Arquitectura de tres niveles: Abstracción de datos



Universidad de Alcalá

Niveles de abstracción

- Nivel Físico ⇒ describe cómo se almacena un registro (ejemplo: cliente). Depende del SGBD
- Nivel Lógico ⇒ describe los datos almacenados en la base de datos y las relaciones entre ellos. Esquema global de BD

```
type cliente = record
                 nombre-cliente : string;
                 id-cliente : string;
                 calle-cliente : string;
                 ciudad-cliente : string;
               end;
```

- Nivel Vistas o externo ⇒ ocultan detalles de tipo de datos. Programas de aplicación se escriben a este nivel
  - Cada usuario o grupos de usuarios tiene su propia vista
  - Pueden ocultar información (ejemplo salario) para determinados usuarios

Universidad de Alcalá

Niveles de abstracción

- Los niveles proporcionan independencia de datos: No cambia el esquema sino el mapa entre dos niveles.
- Ejemplos prácticos:
  1. Se desea migrar una BD (con un diseño en tres niveles: Diccionario de Datos, conceptual E/R, lógico relacional, físico Oracle) de Oracle a SQL Server.
     - No cambia: modelo conceptual, modelo lógico
     - Cambia: modelo físico, regenerándolo para cumplir las reglas y peculiaridades del nuevo SGBD.
  2. Una nueva aplicación requiere un nuevo campo → Una tabla en el modelo global tiene un atributo más.
     - Se crea un esquema externo nuevo para esa aplicación, conteniendo ese nuevo campo
     - El resto de las aplicaciones no se ven afectadas ya que sus esquemas externos no han cambiado.

Niveles de abstracción

Independencia de datos:

- **Lógica:** es la capacidad de modificar el esquema conceptual sin tener que alterar los esquemas externos ni los programas de aplicación (solo el mapa)
- **Física:** es la capacidad de modificar el esquema interno sin tener que alterar el esquema conceptual
- Las aplicaciones dependen del esquema externo → no cambian
- La interfaz (o mapa) entre niveles esta definida para que cambios en una parte, no influyan en otras
- **Ojo**: La independencia es sólo de arriba abajo (si hay que incluir un campo en el nivel externo, afecta al global y al físico)

Base de datos, esquema e instancia

- Conceptos equivalentes a variable, tipo y valor en lenguajes de programación
- BD ⇒ contenedor de información, al igual que una variable
  - Lo deseable es **la** (una) BD de **la** empresa, pero un economista que realiza la nómina de distintas empresas tendría una BD para cada una de ellas

- Esquema ⇒ Estructura lógica de una base de datos
  (recoge restricciones en el mundo real)
  - Equivalente al tipo de una variable en un programa
  - Esquema físico ⇒ cómo almacenar los datos en el disco
  - Esquema lógico ⇒ qué datos son relevantes en el universo del discurso

- Instancia ⇒ el contenido de una base de datos en un instante determinado (conjunto de datos concretos que almacena)
  - Análogo al valor de una variable
  - También llamada ocurrencia, ejemplar, estado (de la BD) o instantánea
  - El back-up contiene una instancia de una BD
  - El Universo Discurso (UD) evoluciona en el tiempo, y con él la BD, generando instancias distintas
  - Virtualmente el número de instancias puede ser infinito.
    (buena discusión de esto en Piattini2006, pag 100)

Universidad de Alcalá

Instancias

- Cuando se crea una base de datos, sólo se está definiendo su **esquema**, aún no tiene datos, por lo que se dice que tiene un **estado vacío** o es **instancia vacía**
- Cuando se cargan los datos iniciales (tablas con datos "maestros") ⇒ **Estado inicial** o **instancia inicial**
- Los estados pueden ser infinitos debido a la evolución del sistema, pero siempre satisfaciendo las restricciones del UD
- El SGBD (Sistema Gestor de Base de Datos) se encarga de que sólo se almacenen estados ó instancias válidos, ajustados a las reglas del esquema proporcionado
- Si las reglas son erróneas ⇒ Base de Datos contendrá estados o instancias imposibles en el mundo real (por fallos en el diseño). Ejemplo: (Sexo: Varón, nº embarazos=2)

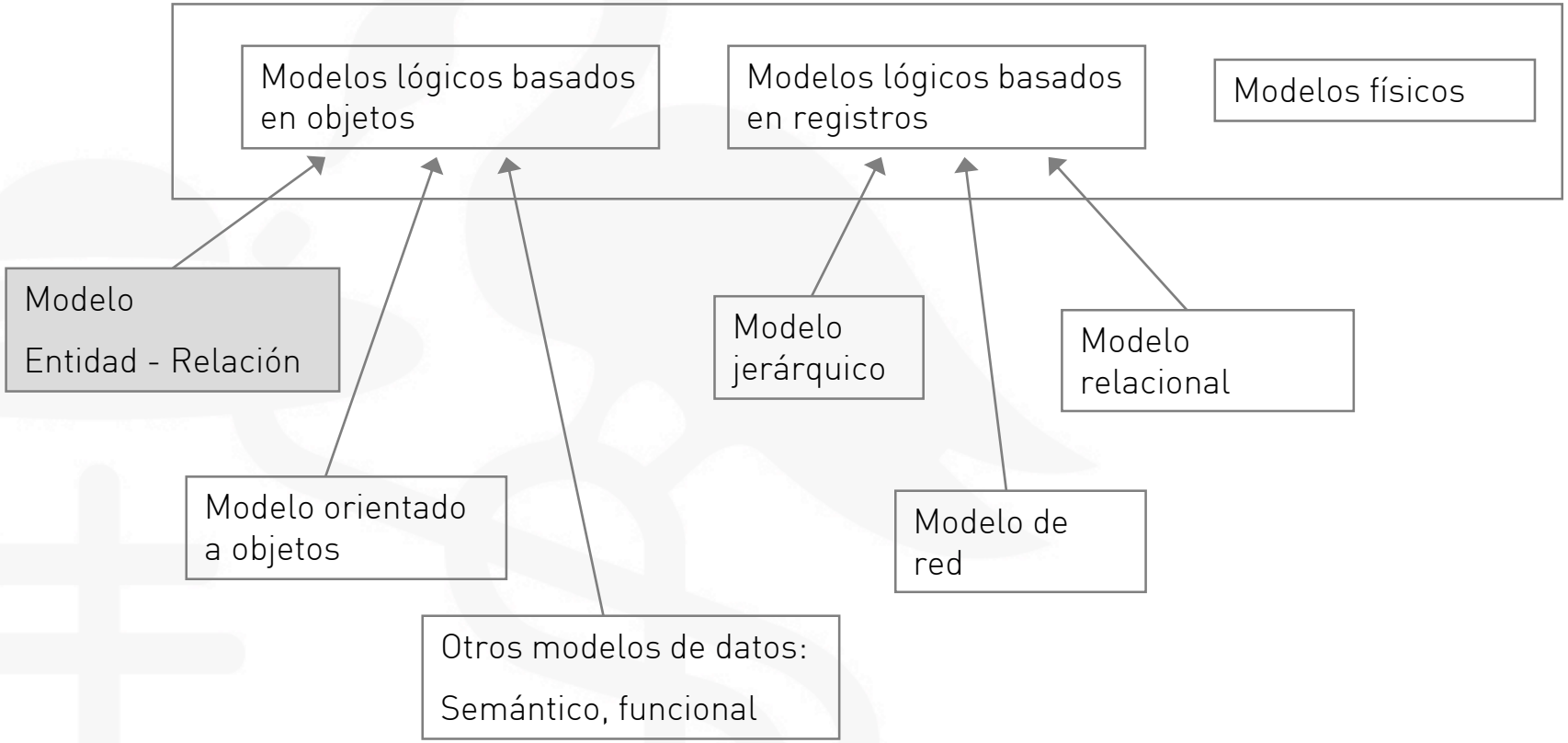
Diccionario de datos

- La validación de las instancias se lleva a cabo mediante el esquema, almacenado en una sección de la BD
- llamada **Meta-Base de Datos**, o **Base de Datos del Sistema**, o **Catálogo del Sistema**
- Actúa como catálogo del sistema, permitiendo al SGBD saber qué reglas debe aplicar
- Metadatos $\Rightarrow$ **Intensión** de la Base de Datos
- Instancia determinada $\Rightarrow$ **Extensión del esquema** de la Base de Datos.

Modelos de datos

- Son una colección de herramientas conceptuales para describir:
  - Datos
  - Relaciones entre datos
  - La semántica de los datos
  - Restricciones de los datos
- Modelo Entidad-Relación
- Modelo Relacional
- Otros modelos
  - Modelo orientado a objetos
  - Modelo relacional orientado a objetos
  - Modelos de datos semi-estructurados
  - Modelos viejos: en red y jerárquico

Modelos de datos



Modelo entidad relación (MER)

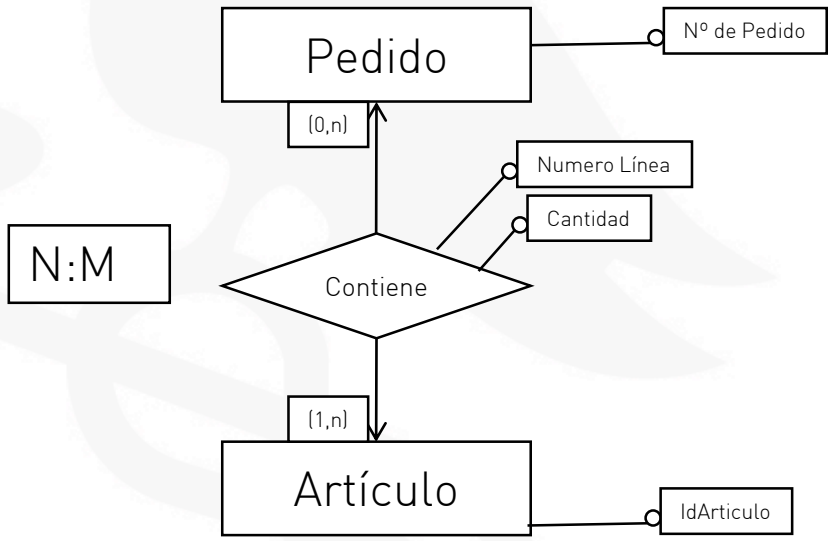
- Modelo E-R (del mundo real)
  - Entidades (objetos): clientes, cuentas
  - Atributos (características): nombre, apellidos, dni
  - Relaciones entre entidades: cuenta A-101 pertenece a Javier
    - Relación impositora asocia clientes con cuentas

- Usado ampliamente para el diseño de bases de datos
  - Diseño de bases de datos con modelo E-R se convierte usualmente en el modelo relacional que es usado para procesamiento y almacenamiento
  - Sencillo, fácil de entender, usado incluso para hablar con nuestro cliente

- El esquema global se representa mediante un diagrama.

Modelos de datos

- Ejemplo (parcial) de esquema en el modelo entidad relación.



Universidad de Alcalá

Modelo relacional

- Se utilizan tablas para los datos y relaciones

Ejemplo de instancia en el modelo relacional

Atributo

Ocurrencia

id-cliente	*nombre-cliente*	*calle-cliente*	*ciudad-cliente*
19.283.746	González	Arenal	La Granja
01.928.374	Gómez	Carretas	Cerceda
67.789.901	López	Mayor	Peguerinos
18.273.609	Abril	Preciados	Valsaín
32.112.312	Santos	Mayor	Peguerinos
33.666.999	Rupérez	Ramblas	León
01.928.374	Gómez	Carretas	Cerceda

Ejemplo de Base de datos relacional

La relación entre tablas se establece por el valor contenido

id-cliente	*nombre-cliente*	*calle-cliente*	*ciudad-cliente*
19.283.746	González	Arenal	La Granja
01.928.374	Gómez	Carretas	Cerceda
67.789.901	López	Mayor	Peguerinos
18.273.609	Abril	Preciados	Valsaín
32.112.312	Santos	Mayor	Peguerinos
33.666.999	Rupérez	Ramblas	León
01.928.374	Gómez	Carretas	Cerceda

(a) La tabla *cliente*

número-cuenta	*saldo*
C-101	500
C-215	700
C-102	400
C-305	350
C-201	900
C-217	750
C-222	700

(b) La tabla *cuenta*

id-cliente	*número-cuenta*
19.283.746	C-101
19.283.746	C-201
01.928.374	C-215
67.789.901	C-102
18.273.609	C-305
32.112.312	C-217
33.666.999	C-222
01.928.374	C-201

(b) La tabla *impositor*

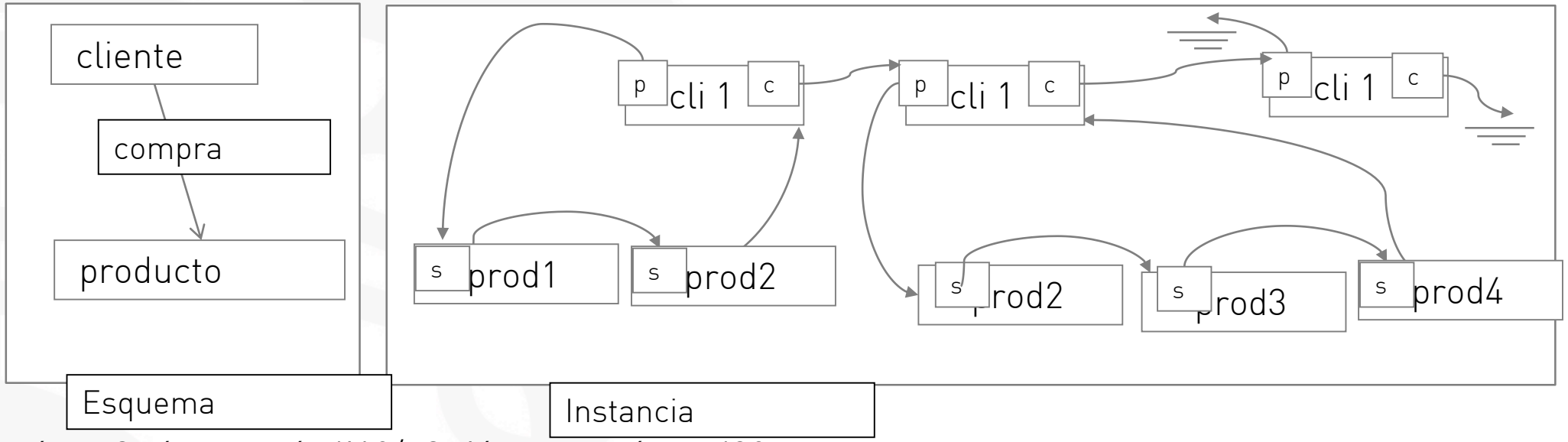
¿Saldo del Cliente "Santos"?=750

Modelo de datos

- Orientado a Objetos: similar al modelo E-R.
  - Colección de Objetos ⇒ valores almacenados en variables + métodos que operan sobre ellos.
  - Los que tienen iguales valores y métodos ⇒ Se agrupan en clases
  - Se accede a los valores por medio de métodos
  - Cada objeto tiene identidad única independiente de los datos⇒ dos objetos con valores iguales son diferentes.

Modelo de datos

- Modelo jerárquico:
  - Los registros se organizan como colecciones de árboles
  - Establece relaciones de 1 a n

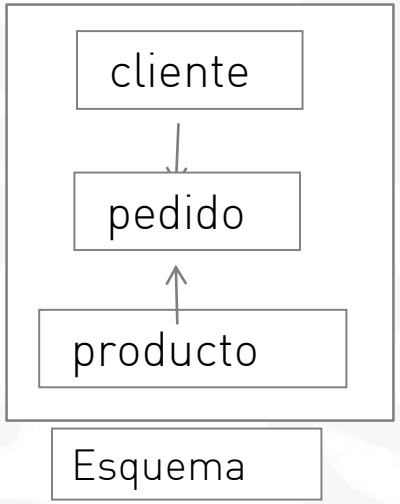


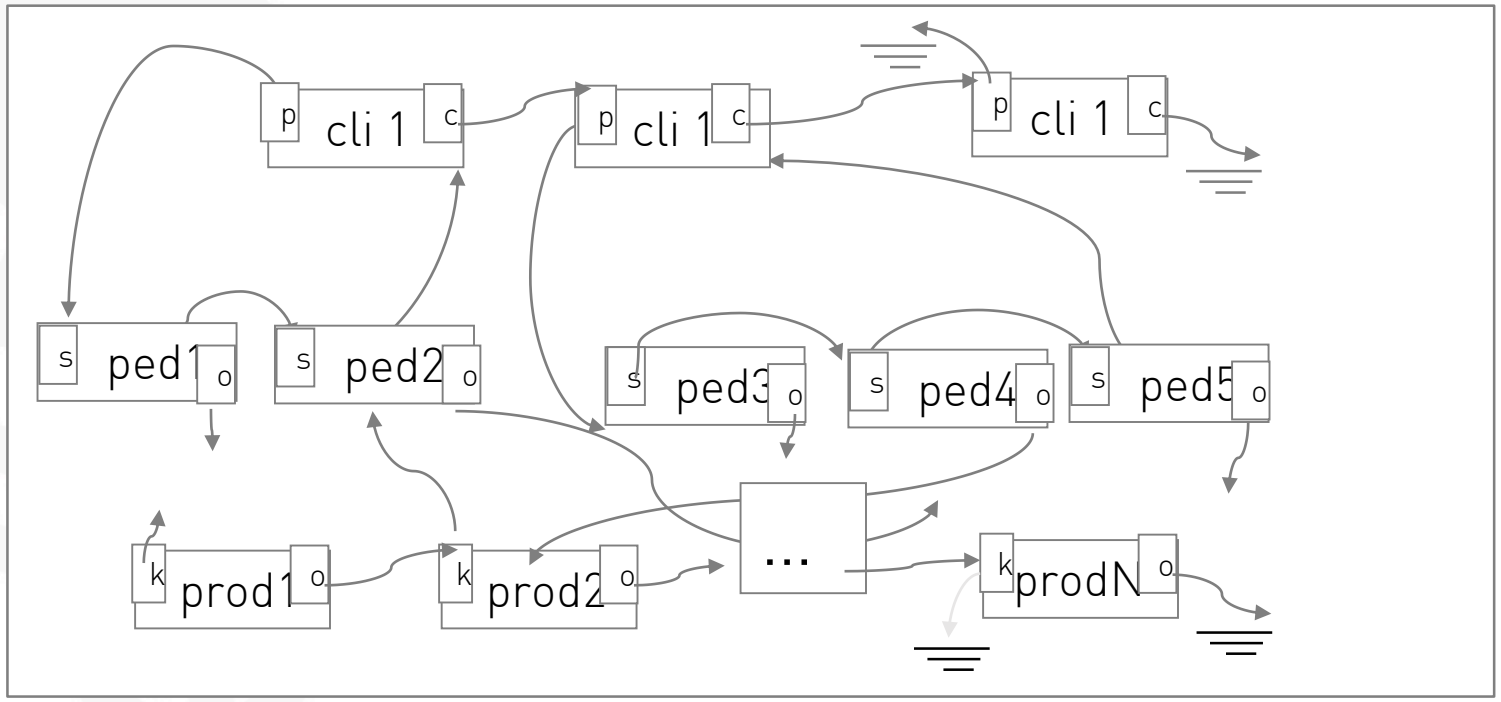
  - Asimetría: ¿Qué compró cli1?/¿Quién compró prod2?
  - Problemas: ¿dónde almaceno un producto aun no comprado?

Universidad de Alcalá

Modelo de datos

- Modelo de red
  - Se establecen grafos dirigidos, o mas bien diversos árboles
  - Establece relaciones de n a m, mediante un nexo (pedido en el ej)





Instancia

- *Simétrico*
- *Cada ped 1 y sólo 1 "padre" de cada tipo*

Universidad de Alcalá

Lenguajes de un SGBD (LDD)

- SGBD ⇒ programarlo para introducir el esquema de definición de la BD
  - LDD (Lenguaje de definición de datos) Se puede dividir en:
    - LDA (Lenguaje de definición de Almacenamiento): utilizado sólo para crear el esquema
    - LDV (Lenguaje de definición de vistas)

    Ej:

```
create table cuenta
        (número-cuenta char(10),
         saldo integer)
```

- Genera el esquema de la base de datos ⇒ Metadatos (datos sobre los datos)
- Se guarda en el **catálogo** de la Base de Datos

Universidad de Alcalá

El **catálogo** de un SGBD relacional

- Almacena el esquema de la base de datos
  - Nombre de las relaciones
  - Nombre de atributos
  - Nuevos dominios
  - Restricciones:
    - Dominio de datos
    - Claves candidatas y primarias
    - Claves extranjeras o foráneas
    - Valores NULL/NOT NULL
  - Vistas
  - Estructura de almacenamiento
  - Índices y métodos de acceso
  - Autorización: Usuarios/Permisos/Datos
  - En sistemas avanzados también almacena:
    - Funciones de usuario
    - Operadores
    - Estadísticas para la gestión del SGBD
    - Disparadores (triggers)

El catálogo de un SGBD relacional

- Es una BD sobre la BD
- Ejemplo: relación de catálogo que describe esquemas de relación.

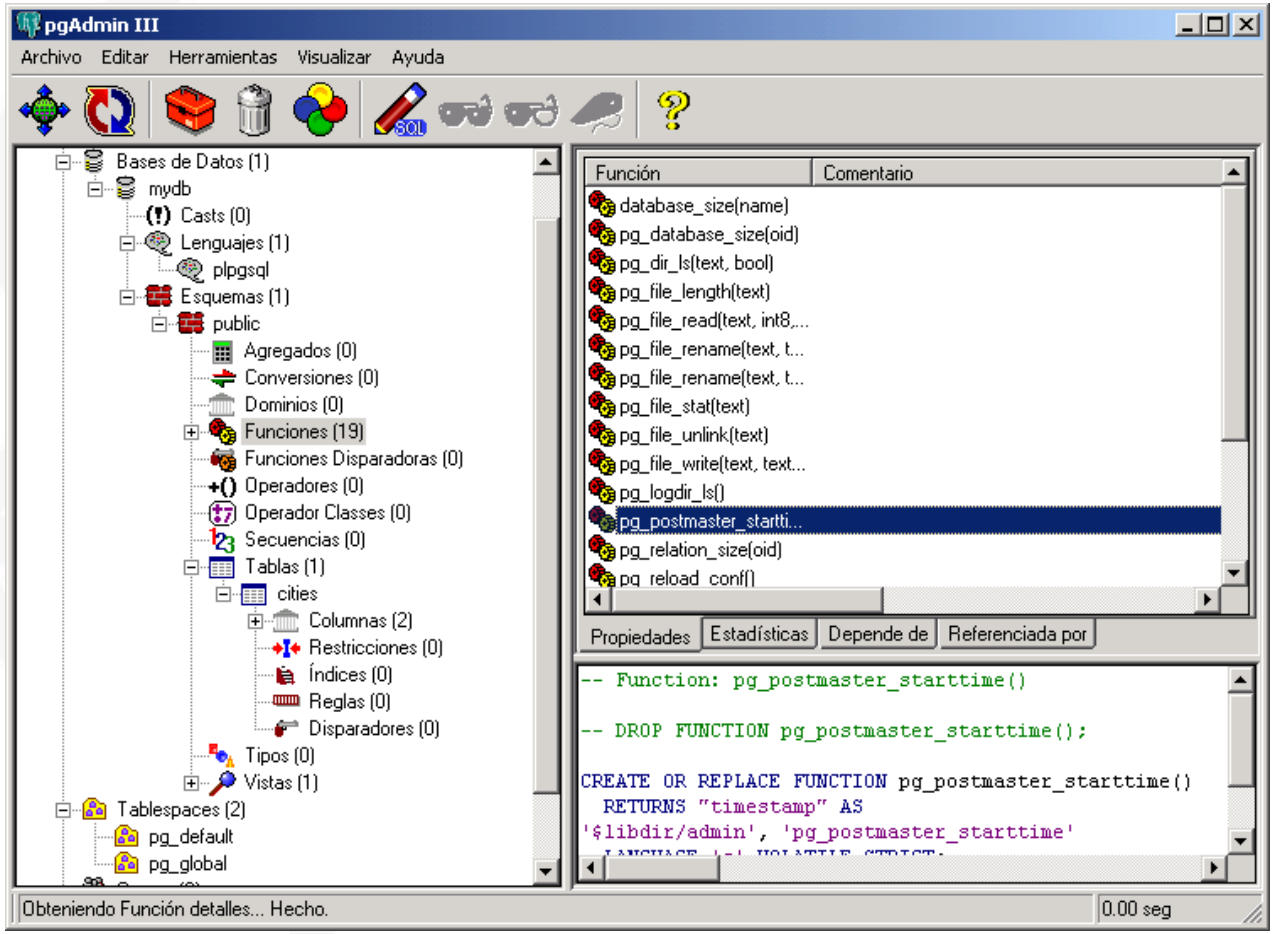
REL_AND_ATTR_CATALOG

REL_NAME	ATTR_NAME	ATTR_TYPE	MEMBER_OF_PK	MEMBER_OF_FK	FK_RELATION
EMPLEADO	NOMBRE	VSTR15	no	no	
EMPLEADO	INIC	CHAR	no	no	
EMPLEADO	APELLIDO	VSTR15	no	no	
EMPLEADO	NSS	STR9	sí	no	
EMPLEADO	FECHA_NCTO	STR9	no	no	
EMPLEADO	DIRECCIÓN	VSTR30	no	no	
EMPLEADO	SEXO	CHAR	no	no	
EMPLEADO	SALARIO	INTEGER	no	no	
EMPLEADO	NSS_SUPERV	STR9	no	sí	
EMPLEADO	ND	INTEGER	no	sí	EMPLEADO
DEPARTAMENTO	NOMBRED	VSTR10	no	no	DEPARTAMENTO
DEPARTAMENTO	NÚMEROD	INTEGER	sí	no	
DEPARTAMENTO	MGRSSN	STR9	no	sí	
DEPARTAMENTO	FECHA_INIC_JEFE	STR10	no	no	EMPLEADO
LOCALIZACIONES_DEPT	NÚMEROD	INTEGER	sí	sí	
LOCALIZACIONES_DEPT	LOCALIZACIÓND	VSTR15	sí	no	DEPARTAMENTO
PROYECTO	NOMBREP	VSTR10	no	no	
PROYECTO	NÚMEROP	INTEGER	sí	no	
PROYECTO	LOCALIZACIÓNP	VSTR15	no	no	
PROYECTO	NUMD	INTEGER	no	sí	
TRABAJA_EN	NSSE	STR9	sí	sí	DEPARTAMENTO
TRABAJA_EN	NP	INTEGER	sí	sí	EMPLEADO
TRABAJA_EN	HORAS	REAL	no	no	PROYECTO
DEPENDIENTE	NSSE	STR9	sí	sí	
DEPENDIENTE	NOMBRE_DEPENDIENTE	VSTR15	sí	no	EMPLEADO
DEPENDIENTE	SEXO	CHAR	no	no	
DEPENDIENTE	FECHA_NCTO	STR9	no	no	

Universidad de Alcalá

El catálogo de un SGBD relacional

- Ej. acceso de usuario mediante herramienta administrativa:



Lenguajes de un SGBD (LMD)

- Lenguajes de un SGBD
  - LMD
    - De Alto Nivel o no procedimental:
      - Típico lenguaje de consulta orientado a conjuntos
      - **Qué** obtener pero no **cómo** obtenerlo
      - Son declarativos
    - De Bajo Nivel o procedimental:
      - Trabajan registro a registro
      - Están integrados en un lenguaje de programación de propósito general (**Lenguaje anfitrión**).
    - Los LMD utilizados de forma independiente se les llama **lenguajes de consulta**
  - **SQL (Select Query Language)** es el lenguaje de consultas más utilizado. **Es un estándar.**
  - Especializados: 4GL

Lenguajes de un SGBD (SQL)

- SQL: ampliamente usado como lenguaje no-procedimental

  Ejemplo: En la BD de la trasparencia 16

  - Encontrar el nombre del cliente cuyo identificador es 19.283.746

```
SELECT cliente.nombre_cliente
FROM cliente
WHERE cliente.id_cliente= '19.238.746'
```

  - Encontrar el saldo del cliente anterior

```
SELECT cuenta.saldo
FROM impositor, cuenta
WHERE impositor.id_ cliente= '19 238 746'
        AND
impositor.numero_cuenta=cuenta.numero_cuenta
```

id-cliente	*nombre-cliente*	*calle-cliente*	*ciudad-cliente*
19.283.746	González	Arenal	La Granja
01.928.374	Gómez	Carretas	Cerceda
67.789.901	López	Mayor	Peguerinos
18.273.609	Abril	Preciados	Valsaín
32.112.312	Santos	Mayor	Peguerinos
33.666.999	Rupérez	Ramblas	León
01.928.374	Gómez	Carretas	Cerceda

(a) La tabla *cliente*

número-cuenta	*saldo*
C-101	500
C-215	700
C-102	400
C-305	350
C-201	900
C-217	750
C-222	700

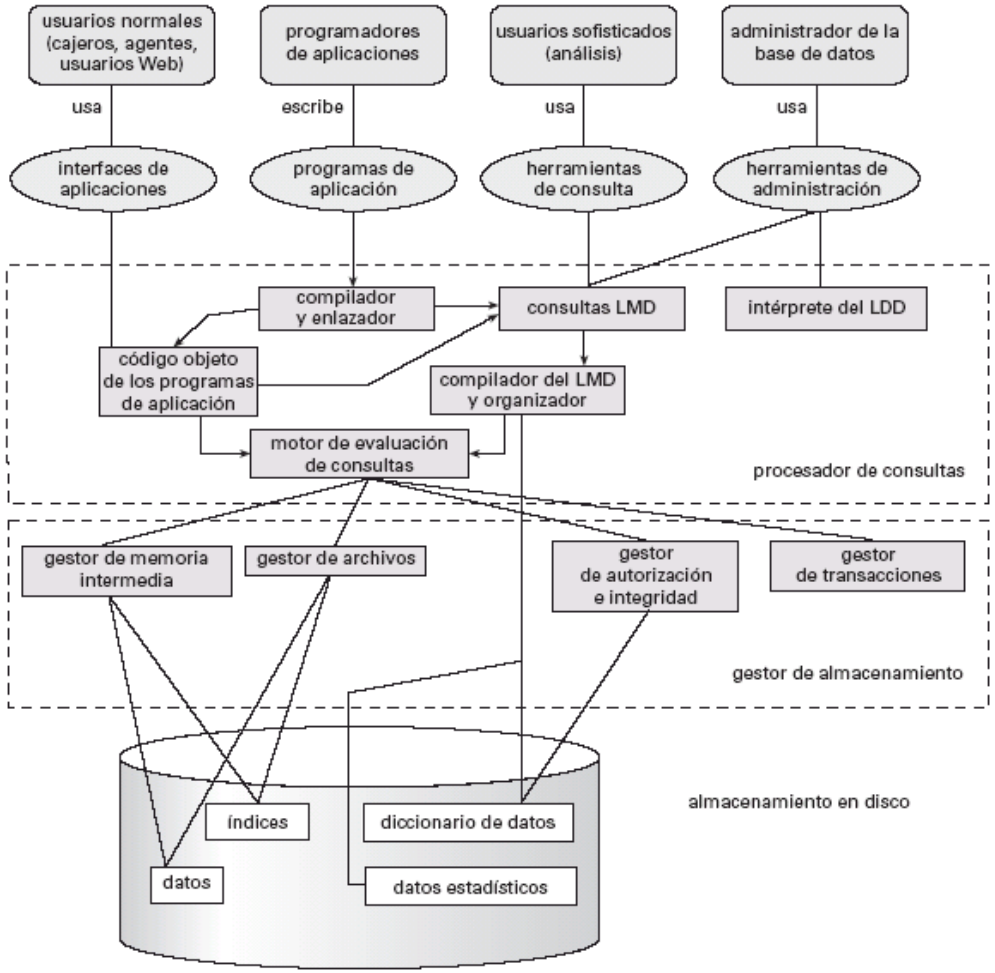
(b) La tabla *cuenta*

id-cliente	*número-cuenta*
19.283.746	C-101
19.283.746	C-201
01.928.374	C-215
67.789.901	C-102
18.273.609	C-305
32.112.312	C-217
33.666.999	C-222
01.928.374	C-201

(b) La tabla *impositor*

- Los programas de aplicación acceden a la base de datos por:
  - Extensiones de un lenguaje de programación convencional (COBOL, PL1..) que permiten embeber SQL
  - Interfaces de aplicación (ODBC/JDBC) que permiten enviar consultas SQL a la base de datos

Esquema de funcionamiento de un SGB



Tipos de usuarios

Procesadores de consultas

Gestor de almacenamiento

Universidad de Alcalá

Estructura de un SGBD a nivel interno

- Gestor de almacenamiento
  - Interface entre los datos y programas de alto nivel y consultas
  - Responsable del almacenamiento, recuperación y actualización de la base de datos
  - Componentes:
    - Gestor autorización e integridad ⇒ satisface las ligaduras de integridad y la autorización de usuarios para acceder
    - Gestor de transacciones ⇒ asegura que la BD quede en estado consistente (correcto) a pesar de fallos en el sistema y transacciones concurrentes (ACID)
    - Gestor de archivos ⇒ gestiona la reserva de espacio en disco y las estructuras de archivos empleadas para la representación de la información almacenada
    - Gestor de memoria intermedia ⇒ trae los datos del disco a la memoria principal y decide qué datos tratar en la memoria cache.

Estructura de un SGBD a nivel interno

- Procesador de consultas
  - Intérprete del LDD ⇒ interpreta las instrucciones LDD y las registra en un conjunto de tablas que tienen metadatos
  - Compilador del LMD ⇒ traduce instrucciones del LMD a instrucciones de bajo nivel que entiende el motor de evaluación de consultas
  - Precompilador del LMD ⇒ convierte las instrucciones del LMD en llamadas a procedimientos normales del anfitrión
  - **Optimizador de consultas** ⇒ obtiene la consulta más eficiente equivalente a la original para ser procesada posteriormente
  - Motor de evaluación de consultas ⇒ ejecuta las instrucciones de bajo nivel generadas por el compilador del LMD.

Herramientas

- Carga de Datos de ficheros existentes
- Herramientas de conversión (importar/exportar)
- Copia de Seguridad (Back-up)
- Reorganización de ficheros
- Control del rendimiento para la supervisión de la base de datos
- Compresión de datos
- Sistema de comunicaciones.

Usuarios de las BD

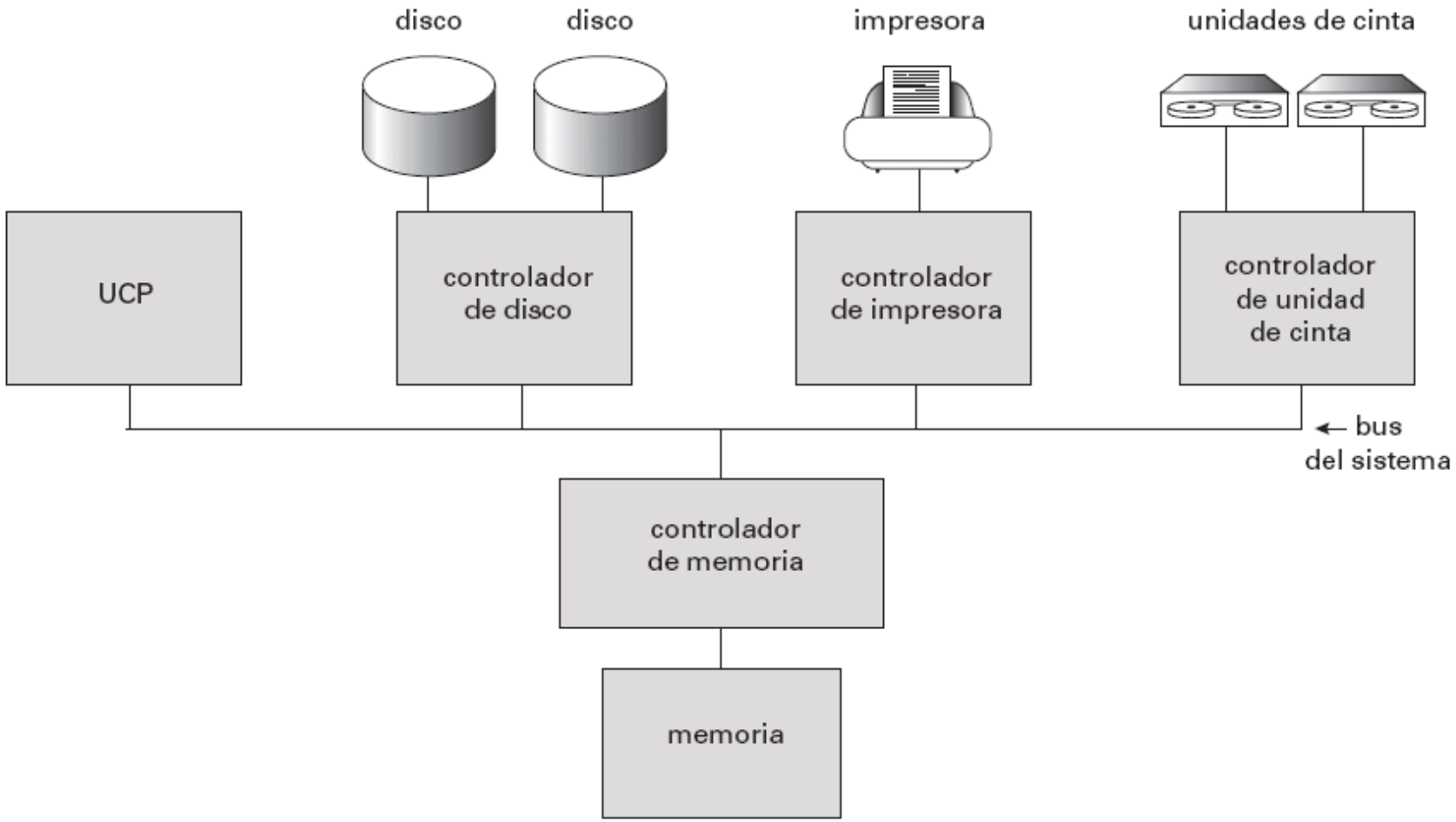
- Según la manera en que interactúen con la BD:
  - Programadores de aplicaciones ⇒ interactúan con el sistema a través de llamadas al LMD sobre otro lenguaje (anfitrión)
  - Usuarios sofisticados ⇒ realizan peticiones usando un lenguaje de consultas
  - Usuarios especializados ⇒ escriben aplicaciones especializadas
  - Usuarios normales ⇒ usuarios no sofisticados que interactúan con el sistema a través de aplicaciones permanentes
    - Oficinistas, clientes que acceden a través de web o puestos de consulta. Ej: reservas aéreas, banca.
  - Administrador de la Base de Datos (ABD) ⇒ control central sobre el sistema. Una o varias personas.
    - Definición del esquema
    - Definición de la estructura y métodos de acceso
    - Modificación del esquema y de la organización física
    - Concesión de la autorización para el acceso a los datos
    - Mantenimiento rutinario: Back-up, espacio en disco, supervisión, etc.

Universidad de Alcalá

Sistemas Informáticos (Centralizado)

- El sistema de la Base de Datos se ejecuta en un único sistema informático, sin interactuar con ningún otro sistema
- Estos sistemas abarcan a los típicos equipos monopuesto, y a los sistemas multipuesto, donde la base de datos está centralizada en el sistema principal
- No suelen ofrecer soluciones excesivamente avanzadas en la gestión de la base de datos.

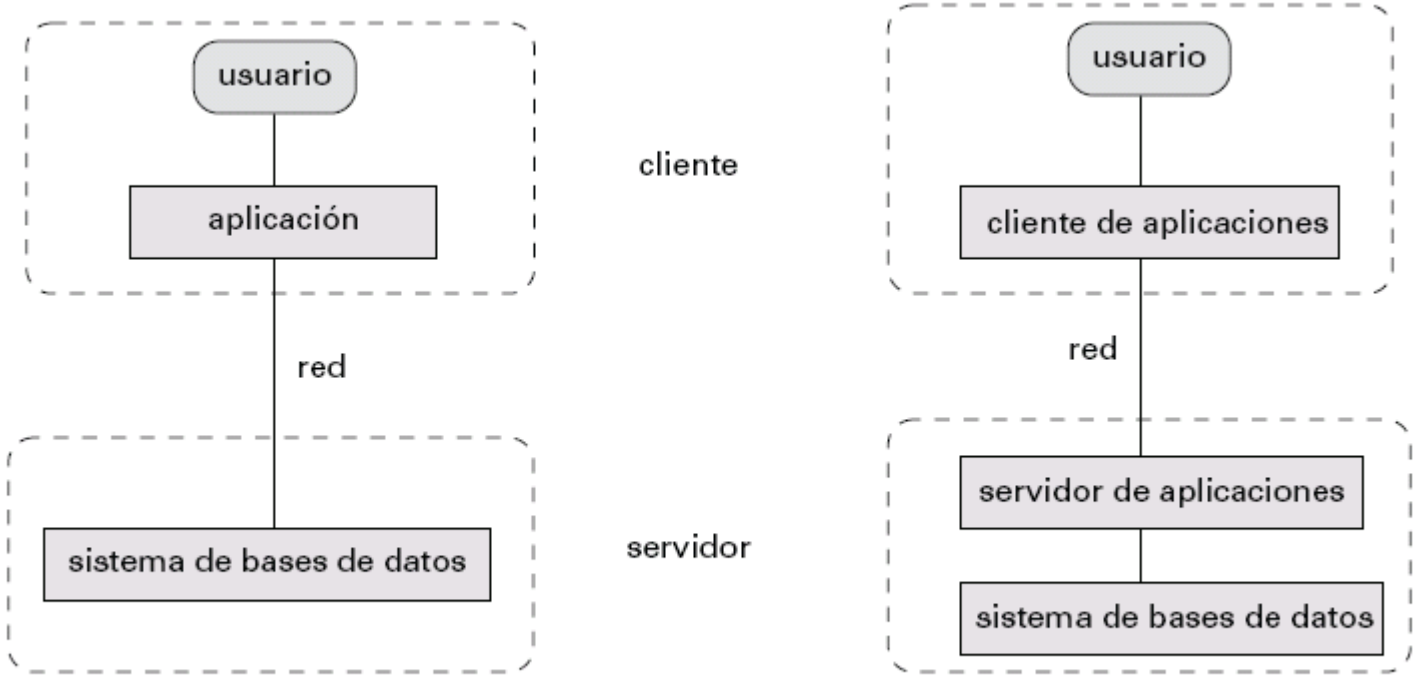
Sistemas Informáticos (Centralizado)



Sistemas Informáticos (Cliente-Servidor)

- Muy extendidos en la actualidad
- La base de datos se sitúa en un ordenador, el cuál realiza toda la gestión y almacenamiento de datos. Es el servidor
- Los ordenadores (quizá menos potentes) hacen consultas sobre los datos del servidor ⇒ proporcionando una interface amigable de acceso a datos al usuario, descargando de esta tarea al servidor. Son los clientes
- Sistema cliente-servidor típico: Servidores de páginas Web con conexión a base de datos.

Sistemas Informáticos (Cliente-Servidor)

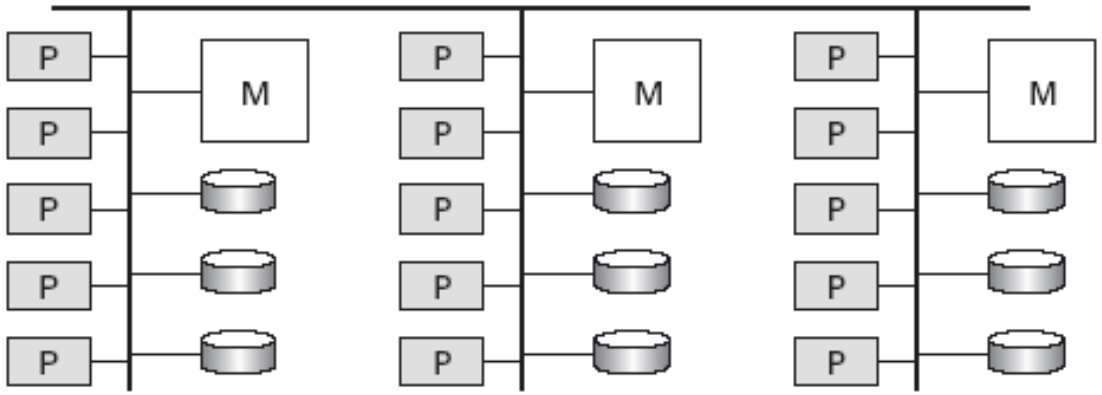


a. arquitectura de dos capas

b. arquitectura de tres capas

Sistemas Informáticos (Paralelos)

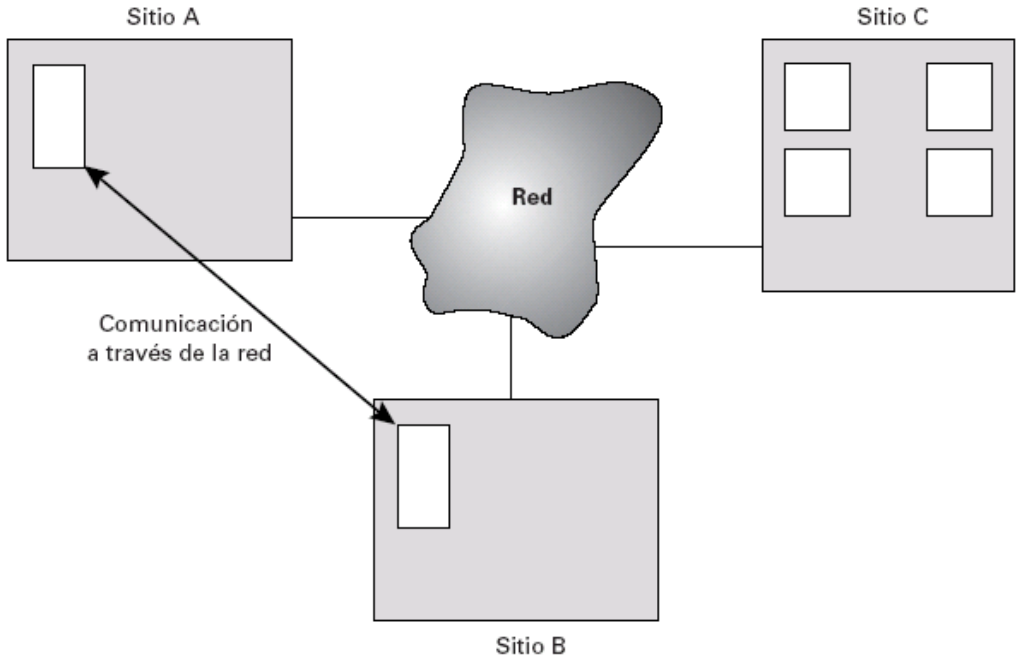
- Utilizan varias CPU y discos en paralelo para optimizar el rendimiento
- Rendimiento:
  - Productividad : nº de tareas completadas en un intervalo de tiempo
  - Tiempo de respuesta : tiempo en completar una tarea.



Sistemas Informáticos (Distribuidos)

- La información se almacena en varios ordenadores

- Dichos ordenadores están conectados entre sí por redes de comunicación.



Clasificación Sistemas de Bases de Datos

- Por Modelo de Datos: relacional, objeto-relacional, jerárquico, red
- Por el nº de usuarios: monousuario, multiusuario
- Nº de sitios en los que está dividido: centralizado, distribuido
- Campo de aplicación:
  - Propósito general
  - Propósito específico : reserva de billetes de líneas aéreas (OLTP)
- Coste.

Proceso de Diseño de las BBDD

- El proceso de diseño de una base de datos, idealmente, es un proceso secuencial e incremental
  - Captura de requisitos: Diccionario de Datos – (Documentos + ME/R Ext)
  - Aproximación inicial (Análisis): Modelo Lógico – (ME/R Ext → MR)
  - Solución específica (Diseño): Modelo Físico – (MR específico del SGBD)
  - Programación: Implementación (Estructuras y consultas) – (SQL)
  - Carga inicial (a veces conocida como carga de ficheros maestros) – (SQL)
  - Pruebas – (SQL, Java, otros lenguajes con acceso al SGBD).
- Problemas:
  - Partimos de un enunciado en lenguaje natural: poco formal, dado a confusiones y ambigüedades, con información incompleta
  - Errores en las primeras fases se magnifican en fases siguientes
  - La base de datos es un ente vivo, cambia y se adapta, y por tanto el diseño también se debe adaptar: A veces hay que volver atrás a cambiar elementos, pero sin perder los datos ya almacenados.