

BASES DE DATOS

Tema 1

Introducción a las Bases de Datos

UCLM- E.S. de Informática

Coral Calero, Marcela Genero, Francisco Ruiz, María Ángeles Moraga

Objetivos

- Dar a conocer la importancia de la **información**.
- Presentar los principales componentes de un **sistema de información**.
- Señalar las diferencias entre los sistemas de ficheros y las **bases de datos**.
- Analizar el concepto de **SGBD**.
- Dar a conocer las principales funciones de un SGBD.

Contenido

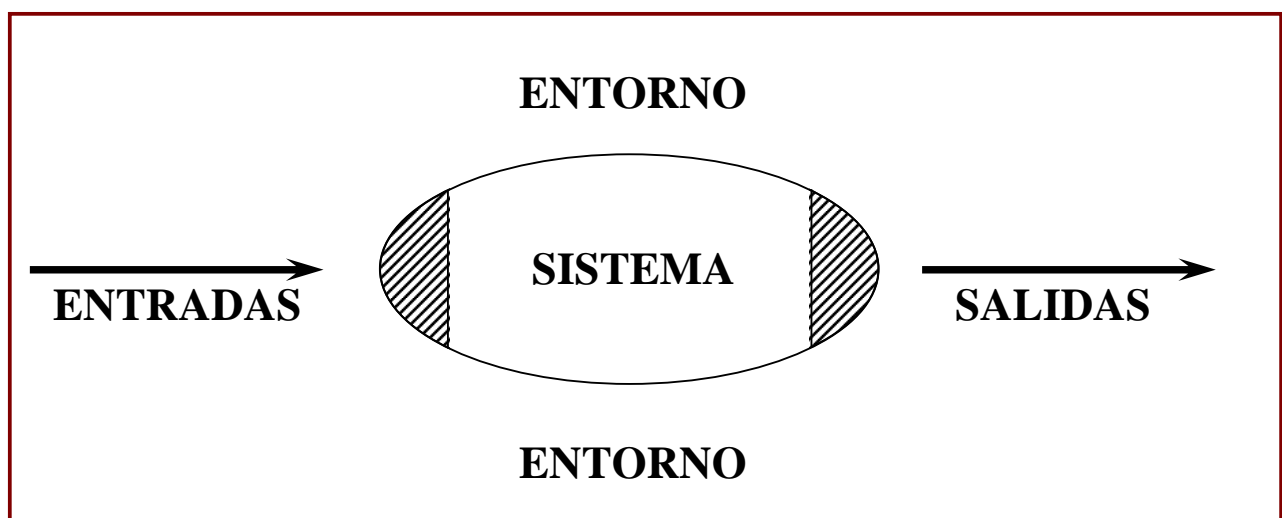
- Sistemas de información
 - Concepto
 - Componentes
 - Clases
- BD y sistemas de BD
 - Concepto de BD
 - Concepto de sistema de BD
 - Sistemas de ficheros versus BD
 - Distintos niveles de abstracción
 - Objetivos y características de un sistema de BD
 - Componentes de un sistema de BD
- SGBD
 - Concepto y funciones
 - Independencia de datos física y lógica
 - Evolución
 - Arquitectura ANSI/X3/SPARC
 - Tipos de usuarios
- Lenguajes de datos

Bibliografía

- Básica
 - Piattini et al. (2006)
 - Caps. 1 y 2
- Complementaria
 - Connolly y Begg (2002)
 - Caps. 1 y 2
 - Elmasri y Navathe (2004)
 - Caps. 1 y 2

Un sistema puede definirse como un **conjunto** de cosas que ordenadamente **relacionadas** entre sí contribuyen a un determinado **objetivo**.

- Los sistemas están limitados, natural o artificialmente. Todo lo que está situado fuera de los límites del sistema se denomina **entorno**. El sistema utiliza elementos del entorno, **entradas**, para elaborar los productos o servicios que se devuelven al entorno, **salidas**.



Los sistemas pueden ser **naturales** o **artificiales**, si son debidos al hombre. Un sistema de información es un sistema artificial.

Concepto

- Toda organización necesita, para su funcionamiento, un conjunto de informaciones que se han de transmitir entre sus elementos y desde y hacia el exterior del sistema.
- Una parte de esta comunicación que se realiza por medio de contactos interpersonales entre los empleados, es el **sistema de información informal**.
- Cuando se trata de organismos complejos, este tipo de flujo de información es insuficiente y costoso, por lo que se hace necesario un **sistema de información formal u organizacional**.

Concepto

- Un Sistema de Información toma **datos** del entorno y devuelve como resultado la **información** que la organización requiere para su gestión y toma de decisiones.
- La organización debe marcar objetivos y directrices que permitan regular al SI.
- Se denomina **Sistema Objeto** a la parte de la organización de la cual se nutre el SI y a la cual revierten sus resultados.
- Los SI son sistemas eminentemente **dinámicos**.

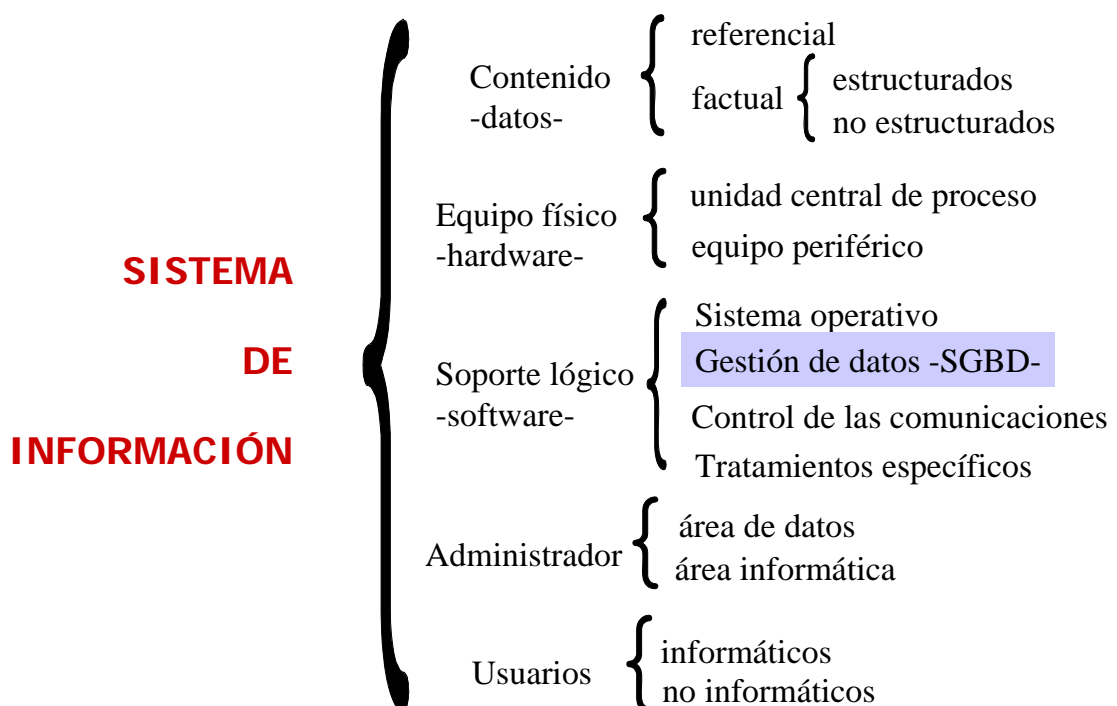
Sistemas de Información

Definiciones

- " *Sistemas de Información son sistemas que **suministran información***", Langefors (1977).
- " *Un Sistema de Información puede ser definido como una **colección de personas, procedimientos y equipos** diseñados, contruidos, operados y mantenidos para recoger, registrar, procesar, almacenar, recuperar y visualizar información*", Teichroew (1976).
- " *Un SI es un **conjunto de elementos**, ordenadamente **relacionados entre sí** de acuerdo con unas ciertas reglas, **que aporta** al sistema objeto (es decir a la organización a la cual sirve y que le marca las directrices de funcionamiento) **la información necesaria** para el cumplimiento de sus fines, para lo cual tendrá que recoger, procesar y almacenar datos, procedentes tanto de la misma organización como de fuentes externas, facilitando la recuperación, elaboración y presentación de los mismos*", Piattini et al. (2006).

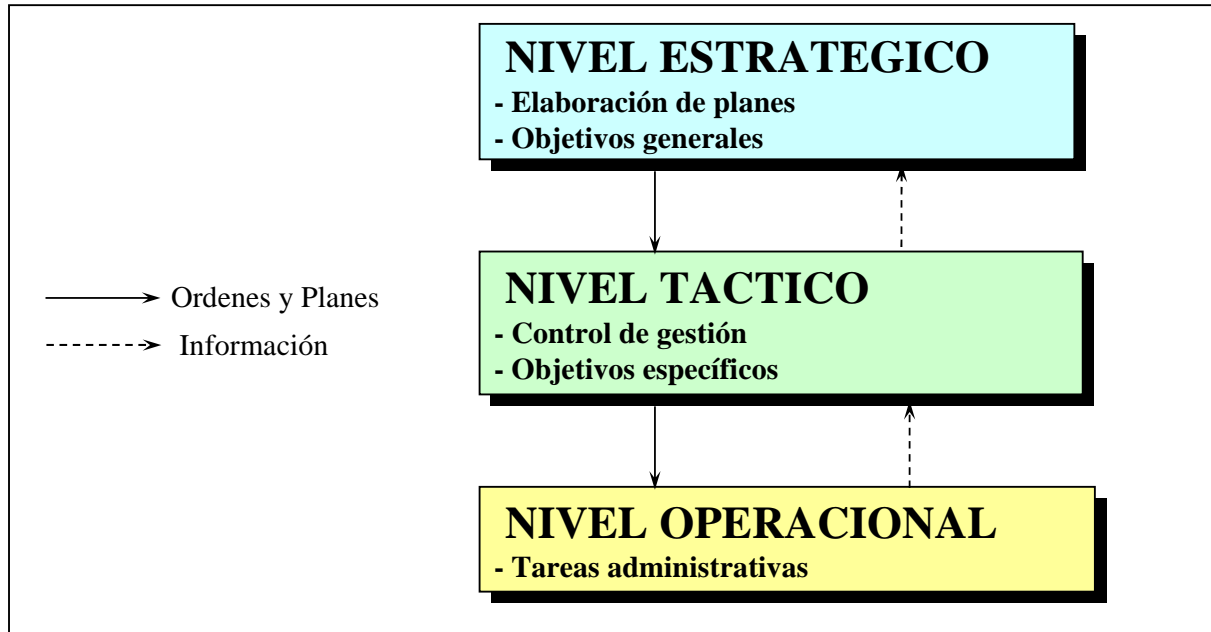
Sistemas de Información

Componentes



Sistemas de Información

Clases – Según niveles organizativos



Sistemas de Información

Clases – Según los tipos de procesos

TIPOS PROCES.	REPETITIVOS	EXCEPCIONALES
FOR MA LI ZA BLES	<p>TAREAS ADMINISTRATIVAS</p> <p><i>Características:</i></p> <ul style="list-style-type: none"> • datos voluminosos • propios, elementales y homogéneos • pocas interrelaciones y simples • muchas salidas normalizadas • procesos sencillos y periódicos • predomina el tratamiento secuencial y por lotes (Ejemplo: nóminas, facturas) 	<p>TAREAS ADMINISTRATIVAS DE EXCEPCION</p> <p><i>Características:</i></p> <ul style="list-style-type: none"> • datos no muy voluminosos • propios, elementales, así como agregados y homogéneos • muchas interrelaciones • pocas salidas normalizadas • procesos complejos pero estructurados • tratamiento no secuencial y, en general, interactivo (Ejemplo: estadísticas, modelos, gestión de personal, etc.)
NO FOR MA LI ZA BLES		<p>AYUDA A LA DECISION</p> <p><i>Características:</i></p> <ul style="list-style-type: none"> • datos muy poco voluminosos • propios y ajenos • agregados y muy heterogéneos • muchas interrelaciones complejas • pocas salidas, con información significativa, oportuna y fácil de interpretar • procesos de difícil o imposible estructuración • tratamiento no secuencial e interactivo (Ejemplo: Creación de una nueva unidad de producción)

- Sistemas de **Gestión**:
 - Procesos formalizables y repetitivos.
- Sistemas de Apoyo a la **Decisión**:
 - Procesos no formalizables y excepcionales.
 - Nuevos Sistemas dirigidos a los directivos:
 - “Decision Support Systems” (DSS), o
 - “Executive Information Systems” (EIS).
 - Avances para dar soporte a la toma de decisiones:
 - Minería de datos (*Data Mining*)
 - Almacenes de datos (*Data Warehouse*).

- “Colección o depósito de datos, donde los datos están lógicamente relacionados entre sí, tienen una definición y descripción comunes y están estructurados de una forma particular. Una base de datos es también un modelo del mundo real y, como tal, debe poder servir para toda una gama de usos y aplicaciones”, (Conference des Statisticiens Européens, 1977).
- “Conjunto estructurado de datos registrados sobre soportes accesibles por ordenador para satisfacer simultáneamente a varios usuarios de forma selectiva y en tiempo oportuno”, (Delobel, 1982).
- “Colección no redundante de datos que son compartidos por diferentes sistemas de aplicación”, (Howe, 1983).
- “Colección integrada y generalizada de datos, estructurada atendiendo a las relaciones naturales de modo que suministre todos los caminos de acceso necesarios a cada unidad de datos con objeto de poder atender todas las necesidades de los diferentes usuarios”, (Deen, 1985).
- “Conjunto de ficheros maestros, organizados y administrados de una manera flexible de modo que los ficheros puedan ser fácilmente adaptados a nuevas tareas imprevisibles”, (Frank, 1988).
- “**Colección de datos interrelacionados**”, (Elsuari y Navathe, 1989).

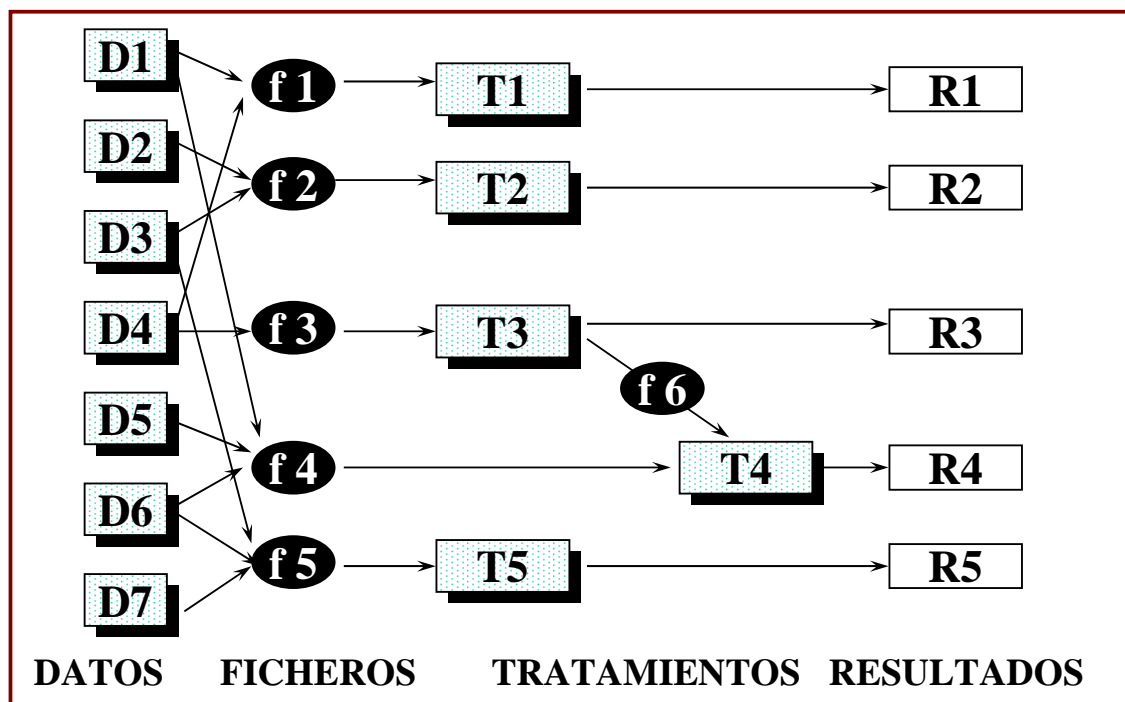
BD y Sistema de BD

Conceptos sobre BD

- Base de Datos (BD):
 - *"Colección o depósito de datos integrados, almacenados en soporte secundario (no volátil) y con redundancia controlada. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de ellos, y su definición (estructura de la base de datos) única y almacenada junto con los datos, se ha de apoyar en un modelo de datos, el cual ha de permitir captar las interrelaciones y restricciones existentes en el mundo real. Los procedimientos de actualización y recuperación, comunes y bien determinados, facilitarán la seguridad del conjunto de los datos", Piattini et al. (2006).*
- Sistema de Gestión de Base de Datos (SGBD)
 - Conjunto de programas que permiten la creación, administración, funcionamiento, uso y mantenimiento de una BD.
- Sistema de Base de Datos (SBD)
 - SGBD + BD + Usuarios

BD y Sistema de BD

Sistemas de Ficheros vs Sistemas de BD

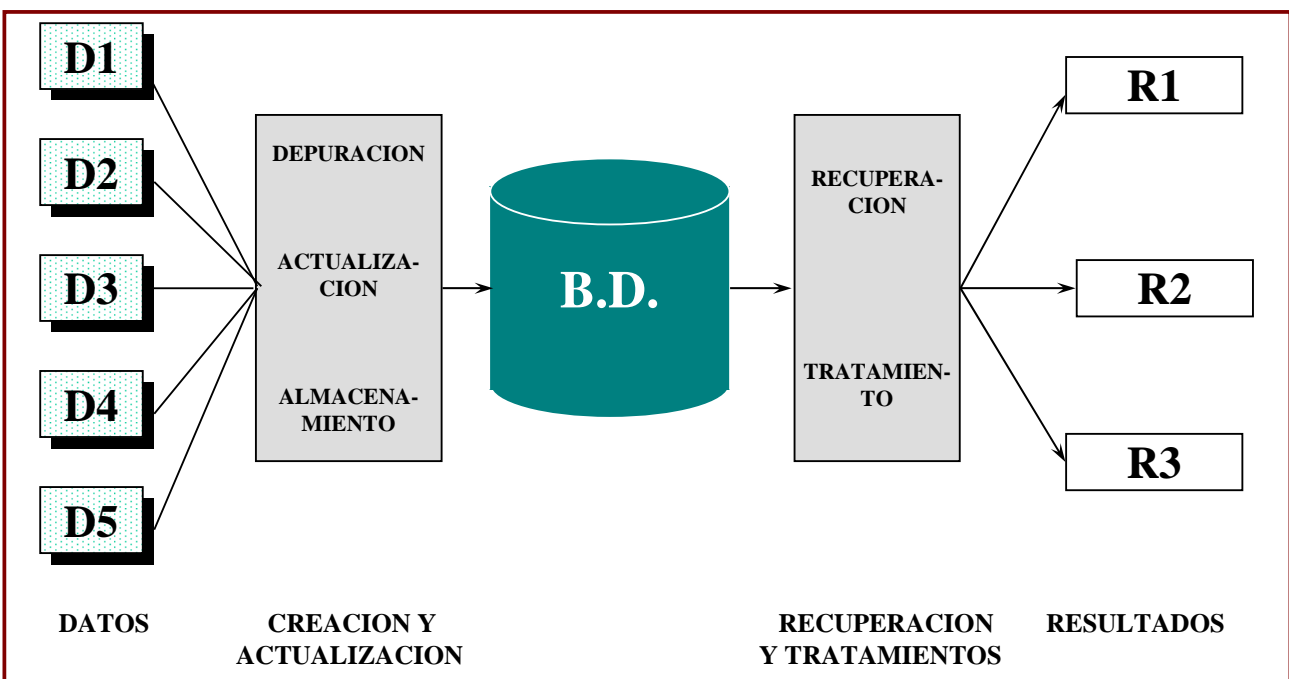


Sistemas clásicos orientados a los procesos

Sistemas de Ficheros vs Sistemas de BD

- Inconvenientes de los sistemas orientados a **procesos**
 - Ocupación inútil de memoria secundaria
 - Aumento en los tiempos de proceso
 - Inconsistencias
 - Dependencia de los datos respecto al soporte físico (falta de flexibilidad frente a cambios)
 - No son apropiados para sistemas de ayuda a la toma de decisiones
- Solución:
 - El enfoque de las **BD**, donde **los datos se almacenan una sola vez**, con independencia de los tratamientos.

Sistemas de Ficheros vs Sistemas de BD



Sistemas orientados a los datos (con BD)

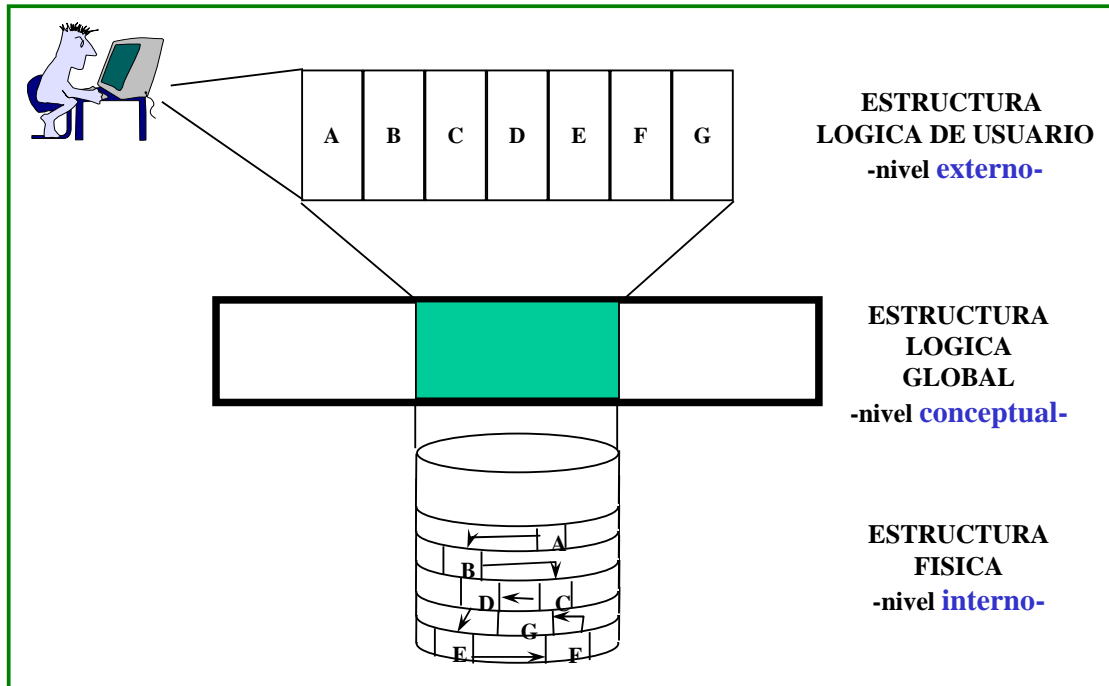
Ventajas de los Sistemas de BD

- Respecto de los **datos**:
 - Independencia de éstos respecto de los tratamientos y viceversa.
 - Mejor disponibilidad de los mismos.
 - Mayor eficiencia en la recogida, codificación y entrada en el sistema.
- Respecto de los **resultados**:
 - Mayor coherencia.
 - Mayor valor informativo.
 - Mejor y mas normalizada documentación, la cual está integrada con los datos
- Respecto de los **usuarios**:
 - Acceso más rápido y sencillo de los usuarios finales.
 - Más facilidades para compartir los datos por el conjunto de los usuarios.
 - Mayor flexibilidad para atender a demandas cambiantes.

Inconvenientes de los Sistemas de BD

- Instalación costosa
- Personal especializado
- Implantación larga y difícil
- Falta de rentabilidad a corto plazo
- Escasa estandarización
- Desfase entre teoría y práctica

Distintos niveles de abstracción



Los tres niveles de abstracción de los sistemas de BD

SGBD

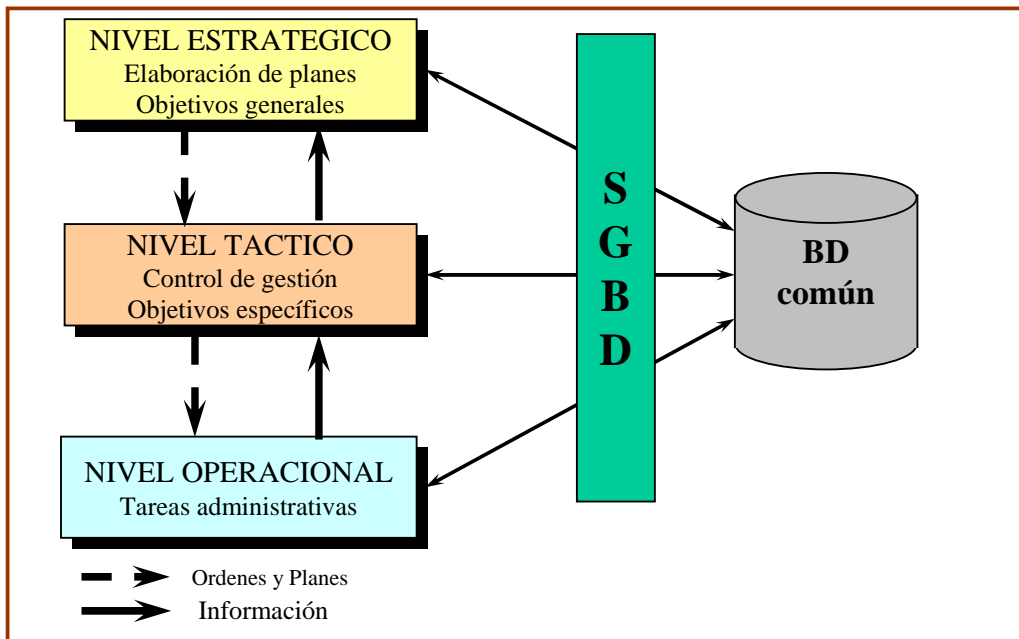
Concepto y Funciones

- Son sistemas **software** complejos
 - Conjunto coordinado de programas, procedimientos, componentes, etc. que suministra a los distintos tipos de usuarios los medios necesarios **para describir y manipular los datos** almacenados en las BD, garantizando su seguridad.
- Operaciones típicas sobre una BD
 - Sobre el conjunto de una BD:
 - Creación
 - Reestructuración
 - Consulta a la totalidad
 - Sobre datos concretos:
 - Inserción
 - Borrado
 - Modificación
 - Consulta selectiva

} Actualización

Concepto y funciones

El **SGBD** actúa como **interfaz** entre la **BD** y los distintos niveles de gestión de la **organización**. Integra los distintos subsistemas, atendiendo a las necesidades de los usuarios en los tres niveles.



Concepto y Funciones

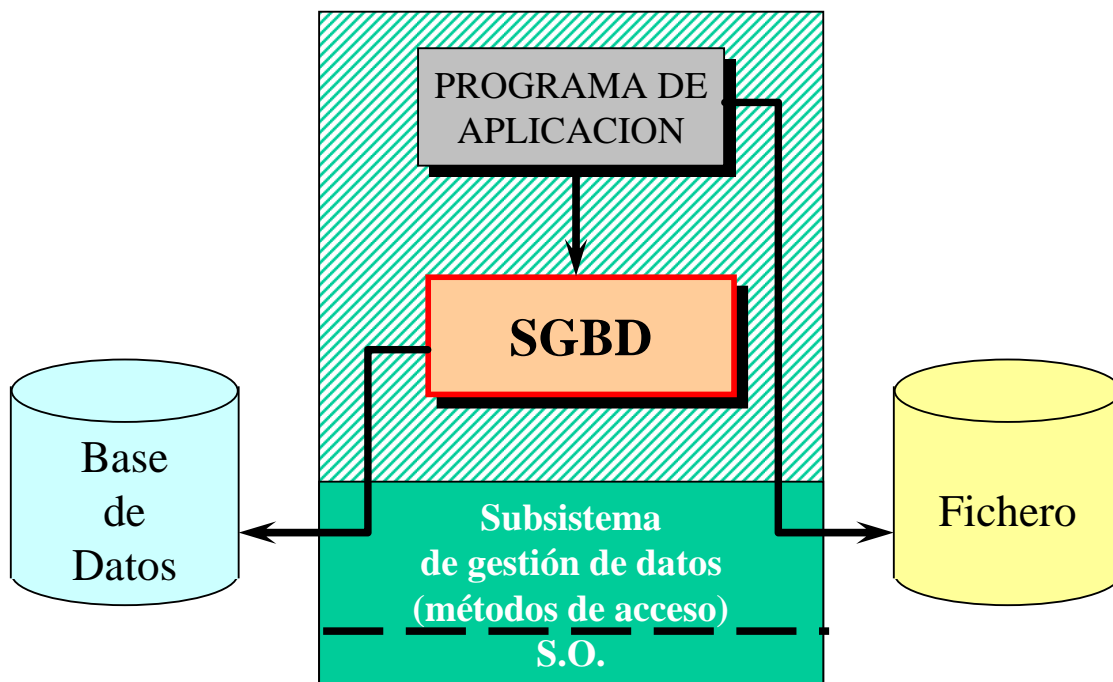
- Funciones esenciales de un SGBD:
 - Función de **Definición** o descripción
 - Función de **Manipulación**
 - Función de **Control**
- Otras funcionalidades:
 - Procedimientos de **Administración**
 - Reorganizaciones
 - Copias de seguridad
 - Cargas de ficheros
 - Estadísticas
 - Generación de listados, etc.
 - **Interfaces** con gestores de transacciones y otros SGBD

Concepto y Funciones

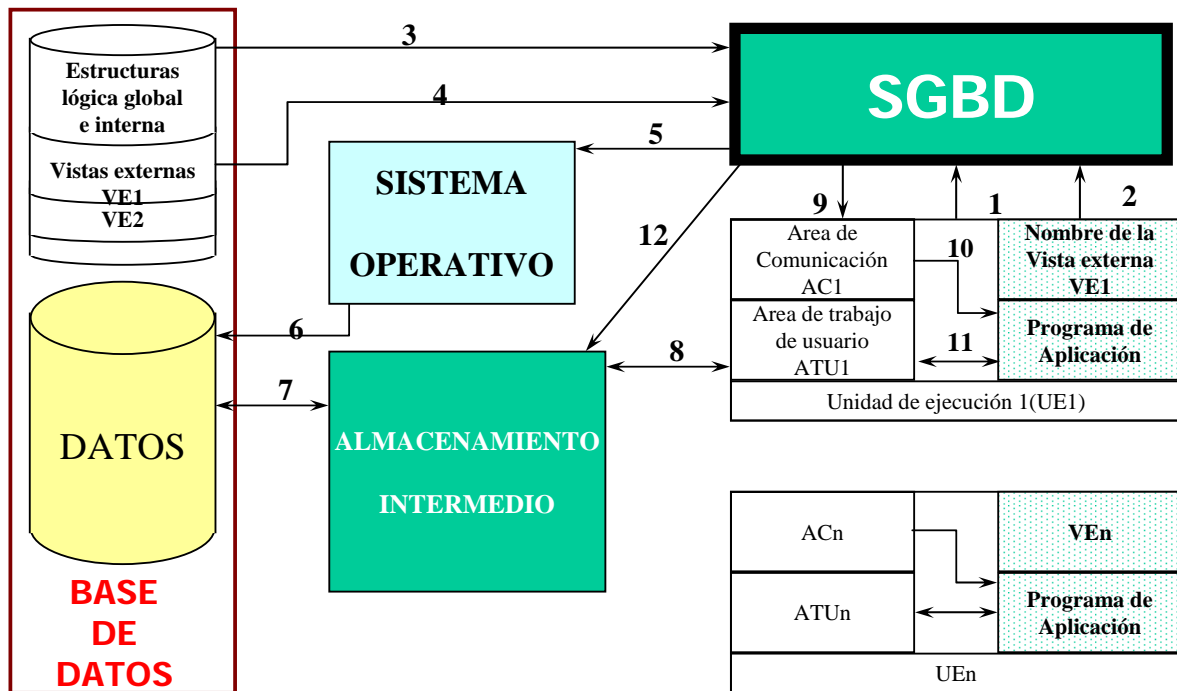
- **Definición**
 - Permite **describir los elementos de datos** con
 - su estructura, sus interrelaciones y sus validaciones.
 - A tres niveles:
 - Externo, Lógico Global e Interno
 - Mediante un **LDD** (Lenguaje de Definición de Datos)
- **Manipulación**
 - Permite manejar los datos
 - Buscar, Añadir, Suprimir y Modificar
 - Mediante un **LMD** (Lenguaje de Manipulación de Datos)
 - Para lo cual es necesario
 - Definir un criterio de selección (el usuario)
 - Indicar la estructura externa a recuperar (el usuario), y
 - Acceder a la estructura física interna (el sistema)
- **Control**
 - Reúne las interfaces de los usuarios, y
 - Suministra procedimientos para el administrador

Funcionamiento – interacción con el S.O.

Comparación entre el acceso a un fichero y a una BD



Funcionamiento – entorno concurrente



Relaciones entre el SGBD y los programas de aplicación en un entorno concurrente

Funcionamiento – entorno concurrente

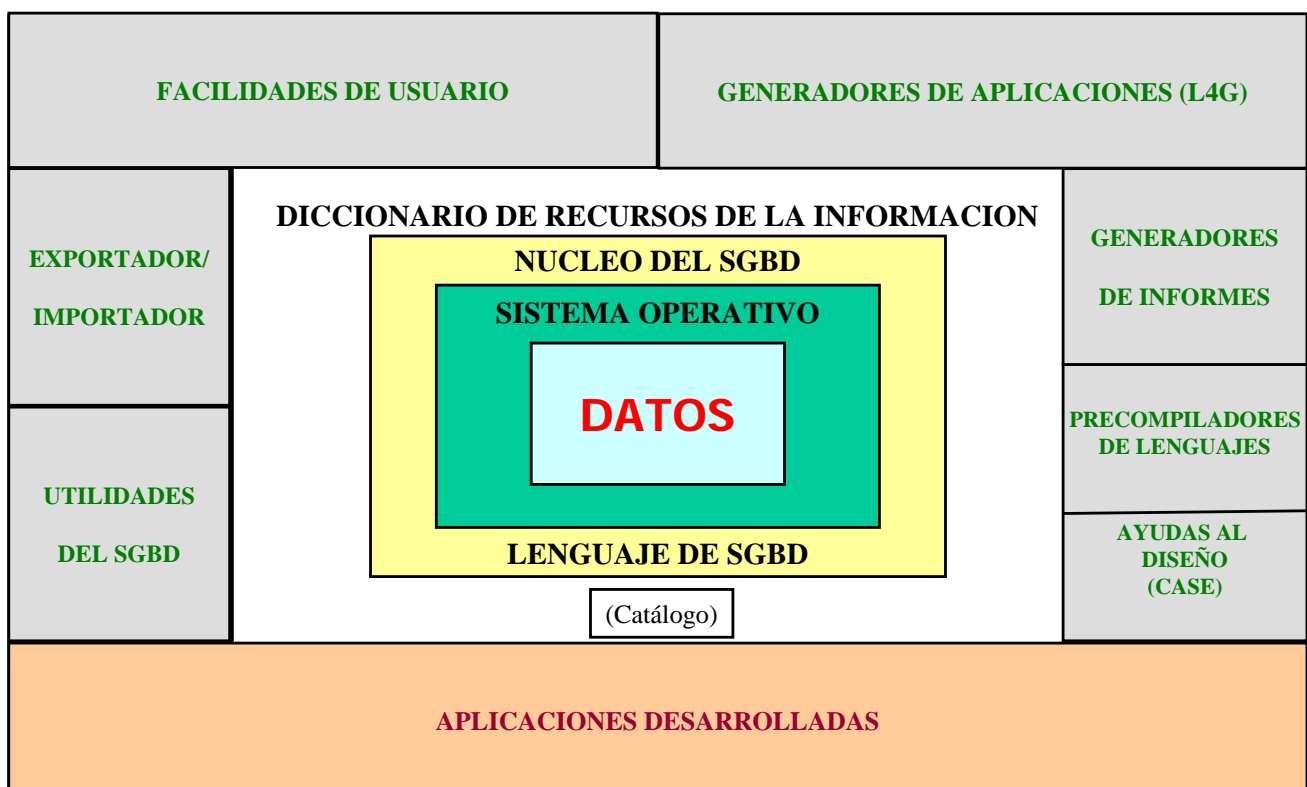
El flujo de datos e instrucciones entre estos elementos es el siguiente:

- Se produce una llamada desde una unidad de ejecución al SGBD (→1); en la llamada se ha de hacer referencia la vista externa implicada (→2).
- El SGBD analiza la llamada y completa los argumentos con la información de la vista externa a la que se ha hecho referencia en la llamada, así como con la correspondiente a la estructura lógica global y la estructura interna con ella relacionadas; esta información se encuentra previamente almacenada en los ficheros del sistema, desde donde pasa al SGBD (→3 y 4)
- Una vez comprobado el derecho del programa de aplicación (PA) a utilizar esta vista, y después de verificar su corrección, el SGBD traduce la llamada convirtiéndola en órdenes a los métodos de acceso del sistema operativo, dirigiéndose a éste (→5).
- El sistema operativo accede al soporte secundario (disco) donde se encuentran almacenados los datos (→6).
- Los datos a recuperar pasan del soporte donde se encuentra almacenada la base de datos al área de almacenamiento intermedio (*buffer*); si se tratase de una inserción o modificación pasarían en sen sentido contrario (→7).
- Los datos son transferidos desde el área de almacenamiento intermedio al área de trabajo del usuario de la unidad de ejecución desde donde se hizo la llamada (→8), o en sentido contrario si se trata de una inserción o modificación, realizándose las correspondientes transformaciones entre las representaciones de los datos.

Funcionamiento – entorno concurrente

7. El SGBD, una vez terminada la operación de manipulación (sea recuperación o actualización); pasa por el área de comunicación los indicadores de estado (→9); en los que se señala si la operación ha acabado satisfactoriamente o no, al tiempo que se dan otras informaciones sobre la operación realizada.
8. El PA revisa el estado de los indicadores que se encuentran en el área de control de la unidad de ejecución desde la que se efectuó la llamada y toma las decisiones oportunas (→10).
9. En el caso de que la operación haya terminado satisfactoriamente, los datos que se encuentran en el área de E/S de la correspondiente unidad de ejecución ya pueden ser utilizados por el PA (→11)

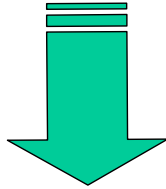
Estructura General



Independencia de Datos

Uno de los principales **objetivos** de las BD es:

***Independencia entre las estructura físicas
y lógicas de los datos***



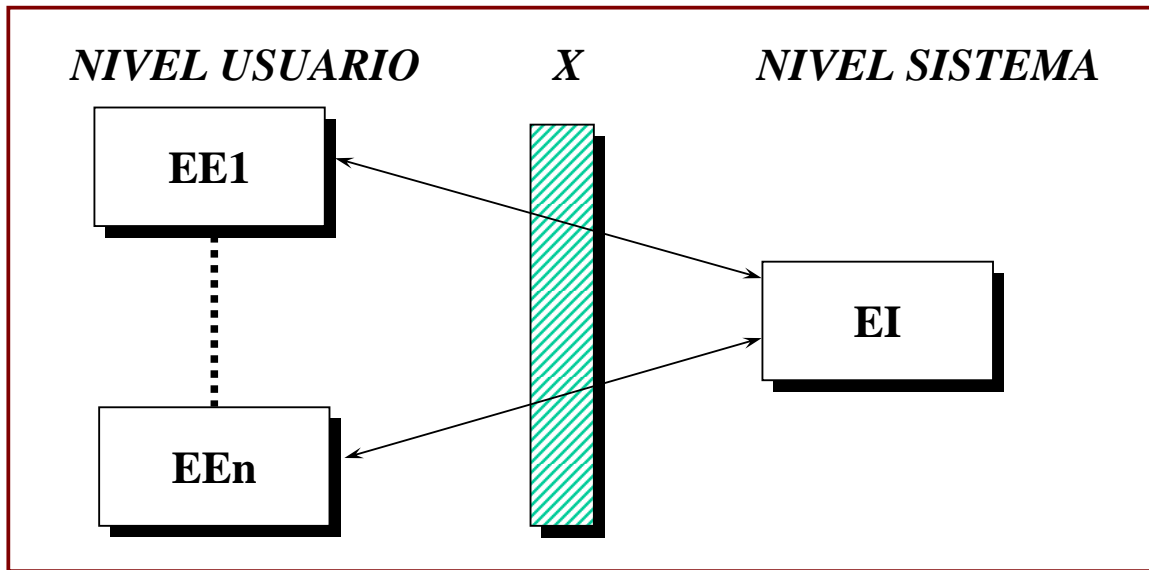
***Independencia entre
datos y aplicaciones***

Independencia de Datos – físico/lógica

- Con la **independencia físico/lógica** se consigue:
 - Que los mismos datos se presenten de distintas formas según las necesidades del usuario.
 - Que el almacenamiento de los datos, su estructura lógica y los programas de aplicación sean independientes.
- Dentro del concepto de independencia físico/lógica podemos distinguir entre:
 - Independencia **de descripción**: que permite separar la definición de los datos a nivel físico y a nivel lógico. Depende de la arquitectura del SGBD.
 - Independencia **de manipulación**: que hace referencia a la de los programas de aplicación con respecto a los caminos de acceso y al soporte físico de almacenamiento. Depende de la arquitectura del SGBD y del modelo de datos.

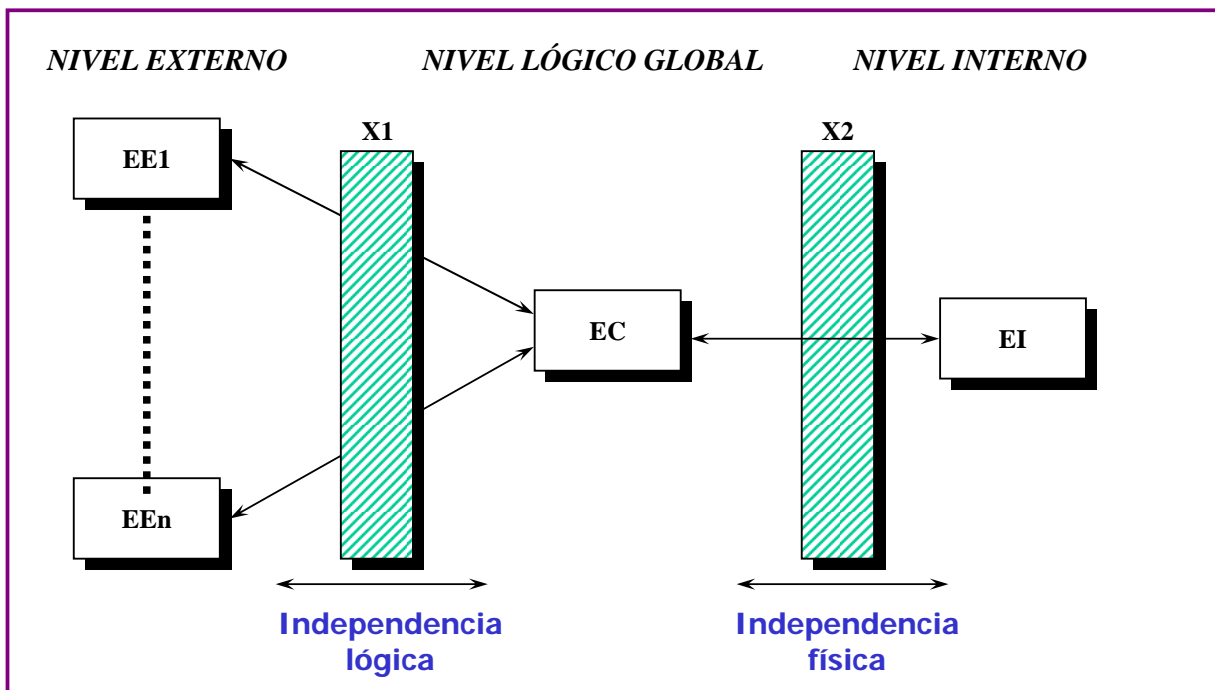
Independencia de Datos – físico/lógica

En una arquitectura de dos niveles



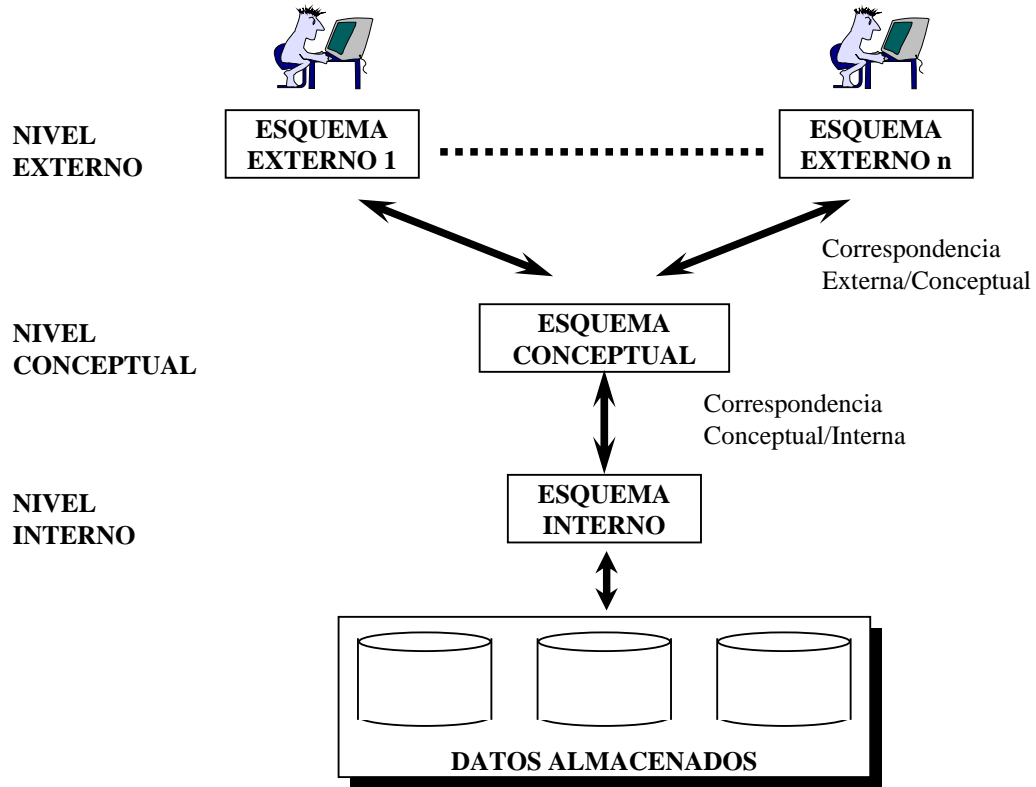
Independencia de Datos – físico/lógica

En una arquitectura de tres niveles

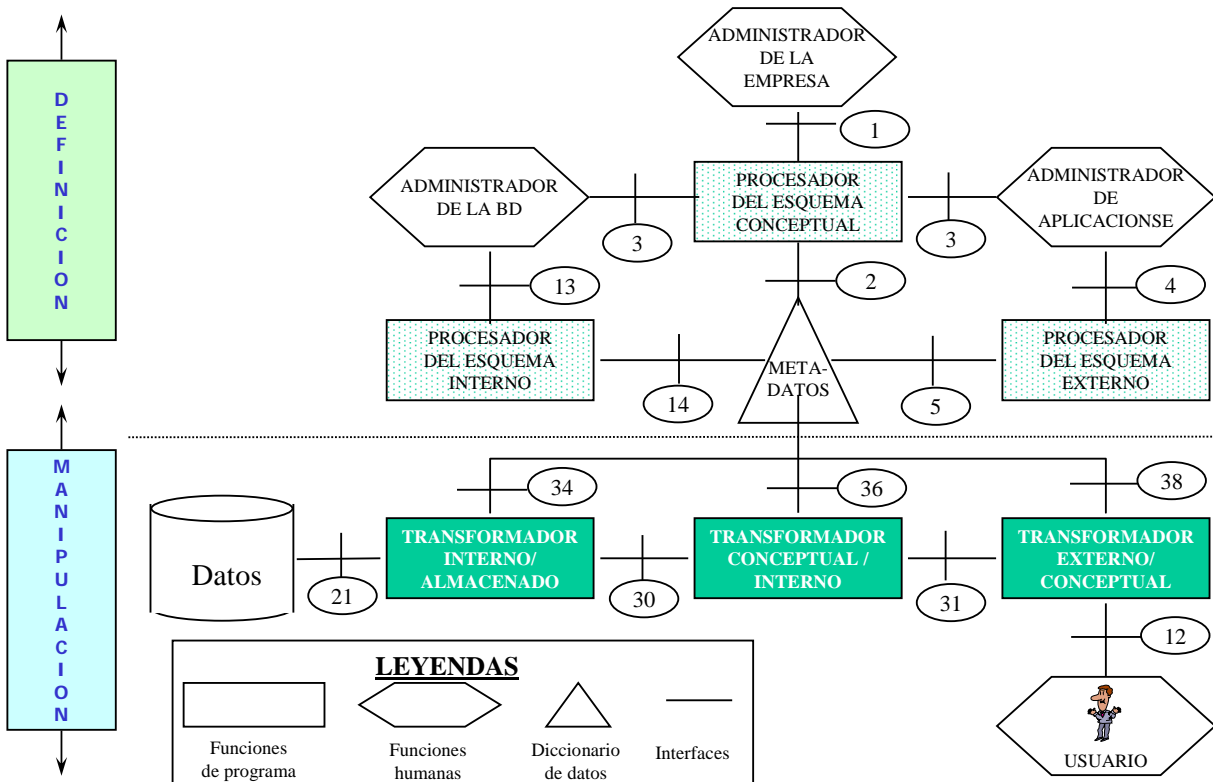


Arquitectura ANSI/X3/SPARC

Es un grupo de estudio del Standard Planning and Reuirements Committee (SPARC) del ANSI (American National Standars Institute) dentro del Comité X3 que se ocupa de ordenadores e informática.



Arquitectura ANSI/X3/SPARC simplificada



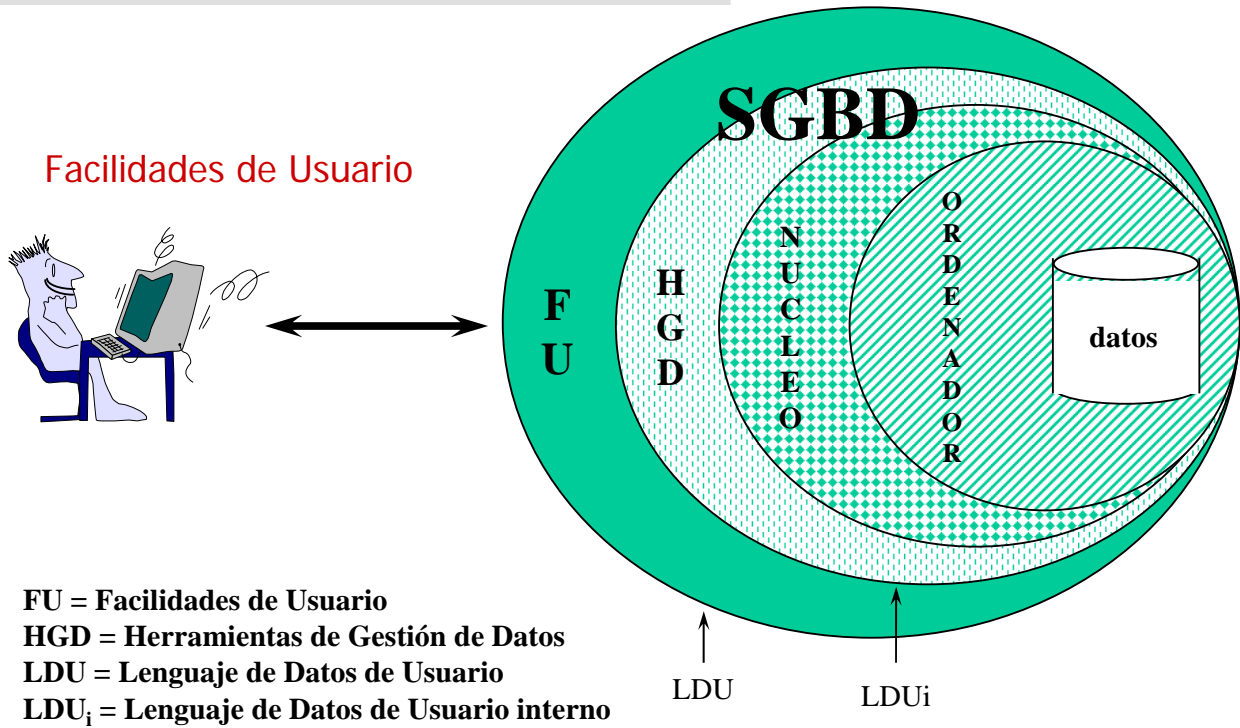
Arquitectura ANSI/X3/SPARC simplificada

- **Definición de la BD:** La parte de definición se facilita por medio de una serie de funciones de programa e interfaces, dando lugar a un conjunto de datos llamados *metadatos* que se almacenan en el diccionario o *catálogo*.
- Una BD se define especificando primeramente el esquema conceptual a través de la interfaz 1, que podría ser un lenguaje de definición del esquema conceptual, o una herramienta CASE integrada. Este esquema conceptual es compilado por el procesador del esquema conceptual y se almacena por medio de la interfaz 2 en la metabase de datos.
- El procesador del esquema conceptual utilizando la interfaz 3, que podría consistir, por ejemplo, en un conjunto de *menús*. Utilizando esta información pueden definirse los esquemas interno y externo a través de las interfaces 4 y 13, que serían controlados por los procesadores correspondientes, y almacenados en la BD a través de las interfaces 5 y 14.

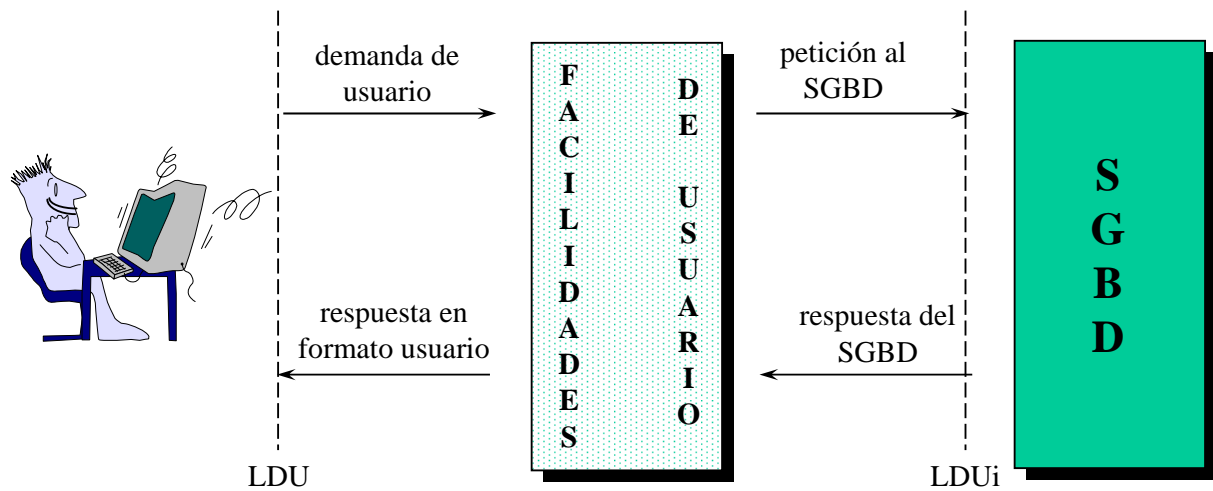
Arquitectura ANSI/X3/SPARC simplificada

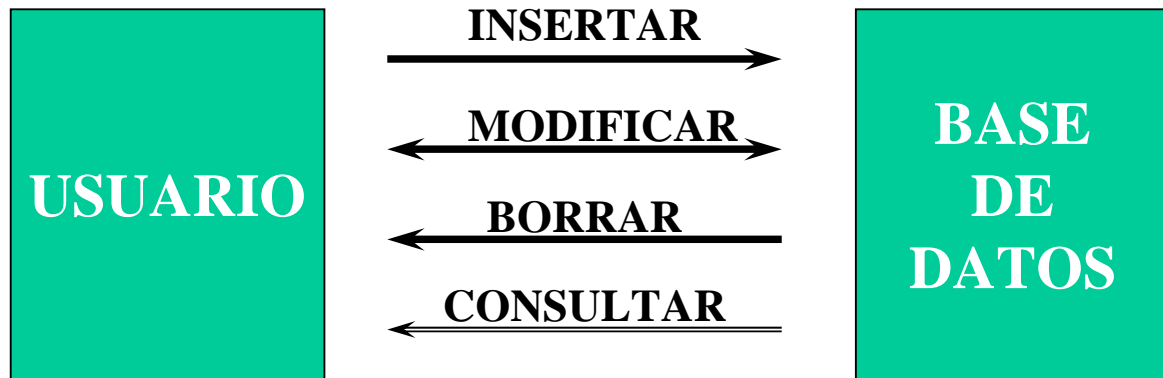
- **Manipulación de la BD:** El usuario puede entonces manipular (insertar, borrar, modificar y recuperar) los datos utilizando la interfaz 12, que podría ser un lenguaje de manipulación, por ejemplo SQL.
- Una petición de los datos por parte del usuario es ejecutada por los transformadores externo/conceptual, conceptual/interno, interno/almacenado, que utilizan los metadatos por medio de las interfaces 38, 36 y 34.
- La solicitud del usuario en la interfaz 12 la convierten los transformadores en peticiones a las interfaces 31, 30 y 21, que devuelven el resultado al usuario. Estas últimas interfaces constituyen la función de vinculación entre los distintos niveles.

SGBD Arquitectura ANSI/X3/SPARC



SGBD Arquitectura ANSI/X3/SPARC





———— ACTUALIZACIONES

===== CONSULTAS

Interacción Usuario/Base de Datos

- a) **Usuarios Informáticos**
 - Diseñadores (lógicos/físicos)
 - Administradores
 - Analistas y programadores
- b) **Usuarios Finales**
 - Esporádicos
 - Habituales

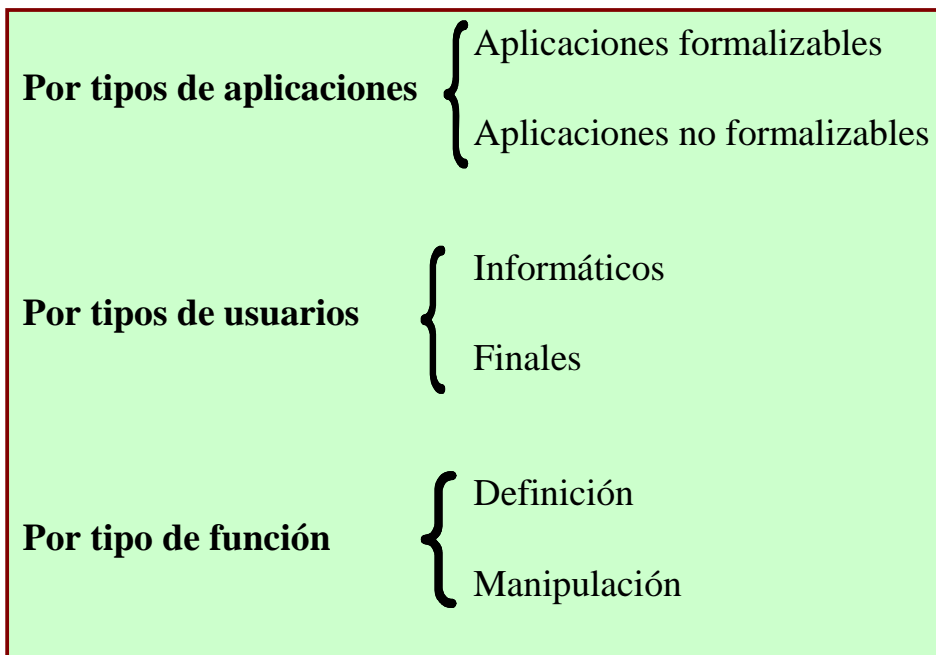
Usuarios informáticos

- **Diseñadores:** Tienen la responsabilidad de identificar los datos que han de estar contenidos en la BD, de acuerdo con las necesidades que les manifiesten los usuarios, así como de determinar las estructuras más apropiadas para conseguir satisfacer estas necesidades
 - **Diseñadores lógicos:** Deben ser capaces de determinar qué tipos de datos han de estar contenidos en la base y conseguir plasmar el punto de vista del conjunto de usuarios en las estructuras lógicas más adecuadas para éstos.
 - **Diseñadores físicos:** Su objetivo es transformar las estructuras lógicas en estructuras físicas que proporcionen la mayor eficiencia de cara a la máquina, minimizando el tiempo de respuesta y el consumo de recursos, es decir, optimizar el ratio coste/beneficio.
- **Administradores:** Su misión es la vigilancia y gestión de los datos. Debe velar por la confidencialidad, disponibilidad e integridad de los datos. Será el responsable de establecer el sistema de autorizaciones de acceso y deberá coordinar y controlar su uso. Tendrá a cargo la gestión del SGBD y otras herramientas relacionadas. Deberá ocuparse del buen funcionamiento de todo el sistema, sin que se produzcan paradas y de modo que se proporcionen los tiempos adecuados de respuesta.
- **Analistas y programadores:** tienen a su cargo el análisis y la programación de las aplicaciones solicitadas por los usuarios.

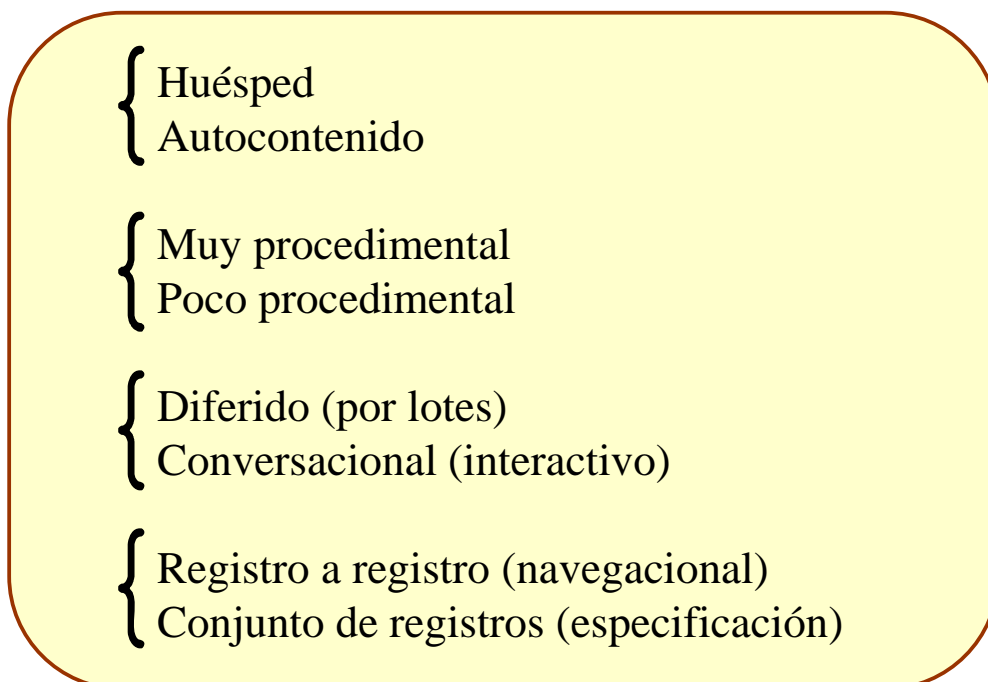
Usuarios finales

- **Habituales:** Suelen hacer consultas/actualizaciones en la base de datos como parte habitual de su trabajo. Utilizan menús previamente preparados por programadores, de forma que se facilite su interrelación con el ordenador. En otros casos pueden usar lenguajes sencillos para el acceso a la BD, así como paquetes de programas.
- **Esporádicos:** Es un tipo de usuarios muy parecido al anterior en la medida en que necesitan el ordenador a fin de que les preste una ayuda en su trabajo, pero en cambio no lo utilizan habitualmente porque el tipo de actividad que realizan no lo exige. Se les tiene que suministrar herramientas sencillas y, en general, potentes.

Lenguajes de los SGBD - clasificación



Lenguajes de Datos - clasificación



Lenguajes de Datos - clasificación

1. **Huésped:** LMD que el programador precisa que se embeba en un lenguaje de programación.
Autocontenido: ofrecen facilidades a los usuarios con pocos conocimientos de programación, para, desde un terminal y en modo interactivo, acceder a la BD y manipular los datos almacenados en ella sin necesidad de apoyarse en un lenguaje de programación.
2. **Muy procedimental, Poco procedimental:** Un LMD es tanto más procedimental cuanto con más detalle sea preciso especificar el procedimiento necesario para acceder a la BDa fin de recuperar o actualizar los datos. En un lenguaje poco procedimental basta con decir *qué* se quiere, sin explicar *cómo* obtenerlo; mientras que si el lenguaje es más procedimental no es suficiente con que se indique el *qué*, sino que es necesario, además precisar el algoritmo.
3. **Diferido, Conversacional:** aunque algunos LMD se utilizan en diferido (tratamiento por lotes), en la actualidad la mayoría de los LMD permiten su uso en modo conversacional o interactivo desde un terminal.
4. **Navegacional:** LMD que recuperan o actualizan los datos registro a registro, y es el programador quien debe indicar el camino que se ha de recorrer, a través de la estructura definida, hasta llegar al registro buscado. Cada sentencia del LMD permite recuperar un único registro.
Especificación: LMD que actúan sobre grupos de registros, de forma que una única sentencia puede dar lugar a la recuperación o actualización del conjunto de registros que cumpla el criterio de selección especificado.

Lenguajes de Datos – clasificación

Procesos formalizables y muy repetitivos:

Se realizan mediante programas, con tratamiento por lotes y periodicidad fija (emisión mensual de nóminas), o tratamiento interactivo (consultas).

Procesos difícilmente formalizables, o poco repetitivos:

No suele ser conveniente escribir programas. Es más conveniente que los resuelva el usuario final con los procedimientos que le proporcione el SGBD (lenguajes de usuario final, menús, etc.).

Usuarios informáticos:

Requieren medios potentes y flexibles. Utilizan lenguajes anfitriones y lenguajes huéspedes.

Usuarios finales:

Medios simples para comunicarse con la BD (lenguajes autocontenidos de sintaxis sencilla, menús, ...).

Lenguajes de Datos – Definición vs Manipulación

Lenguaje de Definición de Datos (LDD) ó Data Definition Language (DDL)

- Suelen ser **autocontenidos** y no necesitan apoyarse en ningún lenguaje de programación.

Lenguaje de Manipulación de Datos (LMD) ó Data Manipulation Language (DML)

- Deben permitir al usuario referirse a un conjunto de datos que cumplan una cierta condición (criterio de selección).
- Un mismo LMD puede actuar como huésped y como autocontenido.
- La mayoría permiten el uso en diferido y en conversacional.
- El **programador** precisa de un LMD embebido y bastante procedimental.
- El **usuario** no informático precisa un LMD autocontenido, muy poco procedimental e interactivo.

Lenguajes de Datos – Huésped vs Autocontenidos

Un lenguaje **huésped** es, en general:

- Procedimental
- Se explota en diferido
- Navegacional

Ejemplo:

DL/1 (IMS de IBM): huésped, muy procedimental, diferido y navegacional

Un lenguaje **autocontenido** es, en general:

- Declarativo
- Conversacional
- Recupera conjuntos de registros

Ejemplo:

SQL: huésped y autocontenido, poco procedimental, recupera conjuntos de registros, actúa en diferido y en conversacional.

```
SELECT nombre, apellido  
FROM persona  
WHERE fecha_nac = "28/11/65"
```


Lenguajes de Datos – Clases (vista general)

Funciones Usuarios		DESCRIPCION			MANIPULACION
		INTERNA	LOGICA GLOBAL	EXTERNA	
ADMINISTRADOR		Lenguaje de descripción de datos internos	Lenguaje de descripción de datos globales	Lenguaje de descripción de vistas externas	PROCEDIMIENTOS: - CREACION - OPTIMIZACION - REORGANIZACION - RECUPERACION - COPIAS SEGURIDAD
	INFORMÁTICO			LLAMADA A UNA VISTA DEFINIDA POR EL ADMINISTRADOR	- LMD (EMBEBIDO) - GENERADOR DE INFORMES - AYUDAS - L4G
NO INFORMÁTICOS	FORMALIZABLE			VISTA EXTERNA DESCRITA EN EL PROPIO PROCEDIMIENTO	- USO DE APLICACIONES - L4G - LENGUAJES PARAMETRICOS PAQUETES
	NO FORMALIZABLE			FACILIDADES INCLUIDAS EN LOS LENGUAJES O EN LOS PAQUETES	- LENGUAJES AUTOCONTENIDOS - PAQUETES