



# Bases de datos

UNIDAD 1

CONCEPTO DE BASE DE DATOS: EL SISTEMA DE GESTOR DE BASE DE DATOS (DBMS)

# Datos, información, conocimiento

---



# Datos

---

- Los datos son la **mínima unidad semántica**.
- Por sí solos son irrelevantes como apoyo a la toma de decisiones.
- Un número telefónico o un nombre de persona, son datos.
- Las tecnologías de la información han aportado mucho a la gestión de los datos.
- Las organizaciones deben recopilar datos de empleados, clientes, proveedores, inventario, etc.
- ¿Tienen utilidad por sí solos los datos?



# Información

---

La información se puede definir como un conjunto de **datos procesados** y que tienen un **significado** (relevancia, propósito y contexto), y que por lo tanto son de **utilidad para quién debe tomar decisiones**, al disminuir su incertidumbre.



# Conocimiento

- Es una **mezcla** de *experiencia*, valores, *información* y *know-how* que sirve como marco para la incorporación de nuevas experiencias e información, y es **útil para la acción**.
- Se origina y aplica en la mente de los conocedores.
- Se encuentra en documentos, pero también en rutinas *organizativas*, *procesos*, *prácticas*, y normas.



# Conclusión

---

- Conforme las organizaciones invierten y se hacen más dependientes de los **sistemas de información**, los procesos de *recopilación, gestión y utilización* de **datos** se hace más importante para el *éxito operacional*.
- Los **datos** sólo son **valiosos** si somos capaces de acceder y extraer **significado** de ellos, es decir, **información**, y combinarlo con nuestra *experiencia* para constituirse en **conocimiento** útil para *tomar buenas decisiones*.
- No podremos extraer tal conocimiento sin **organizarlos, almacenarlos y analizarlos** de forma efectiva.

# Estructuras de datos



# Sistemas de ficheros

---

## DEFINICIÓN

Un **sistema basado en ficheros** (SF) es un conjunto de programas informáticos que permiten al usuario almacenar, consultar y modificar datos. Dichos datos se almacenan en ficheros diseñados para una determinada aplicación.

**Cada programa** define y maneja **sus propios datos**.

Por ejemplo, esta presentación se preparó con un programa que almacenó un fichero pdf. Para abrirlo necesitamos un visor de documentos pdf.



# Sistemas de ficheros

---

## ORIGEN

Los SF surgieron al informatizar el manejo de archivadores manuales para acceder más eficientemente a los datos almacenados.

Un SF sigue un **modelo descentralizado**: cada departamento almacena y gestiona sus datos mediante programas de aplicación escritos *especialmente* para él.

Estos programas son **totalmente independientes** entre un departamento y otro, para introducir datos, mantener los ficheros y generar los informes que el propio departamento necesita.

# Sistemas de ficheros

---

Cuando se trabaja con visión exclusiva de sistema de ficheros, los departamentos no comparten información ni aplicaciones, **los datos comunes deben estar duplicados** en cada uno.

Esto puede originar **inconsistencias** en los datos. Por ejemplo, el domicilio de un cliente puede diferir entre departamentos si el domicilio cambia y sólo uno de los departamentos lo sabe.

Además, cuando los datos se separan en distintos almacenes de ficheros se complica el acceso, el programador de aplicaciones debe **sincronizar** el procesamiento de los distintos ficheros implicados.

# Sistemas de ficheros

---

## INCONVENIENTES

**Separación** y **aislamiento** de datos.

**Duplicación** de datos (se desperdicia capacidad de almacenamiento y se puede **perder la coherencia** de datos).

**Dependencia** de los datos (**estructura física** de los datos codificada en cada programa de aplicación, lo que dificulta los cambios en la misma).

**Incompatibilidad** entre **formatos** de ficheros (la estructura de cada uno puede depender de la aplicación particular).

**Rigidez** en las **consultas**, que quedarían fijas (los SF devuelven una información concreta, no permiten recuperarla según criterios diferentes a los pre-establecidos).

# Bases de datos

---

## DEFINICIÓN

Una **Base de Datos** es un conjunto de datos que modela *hechos* y *objetos* de una parte de la realidad y sirven de soporte a una serie de *aplicaciones informáticas*. Dichos datos deben estar almacenados físicamente en *soporte informático* y deben estar relacionados entre sí mediante una determinada estructura lógica.

**Propiedad esencial:** Independencia de *datos* (separación entre los datos y las *aplicaciones* informáticas que los manejan).

Cada BD se diseña para satisfacer los requisitos de información de una empresa u organización.

Antes de las BD se trabajaba con **sistemas de ficheros**.

Ambos son parte importante en los **sistemas de información**.

# Bases de datos

---

Una base de datos se puede percibir como un gran **almacén de datos** que se **define** y se **crea una sola vez**, y que se utiliza al mismo tiempo por distintos usuarios.

En una base de datos todos los datos se integran con una **mínima** cantidad de **duplicidad**. De este modo, la base de datos no pertenece a un solo departamento sino que se comparte por toda la organización.

Además, la base de datos no sólo contiene los datos de la organización, también almacena una descripción de dichos datos. Esta descripción es lo que se denomina **metadatos**, se almacena en el diccionario de datos o catálogo y es lo que permite que exista independencia de datos lógica-física.

Representa algún aspecto del mundo real, también denominado dominio.

Está **diseñada**, construida y poblada con datos para un **propósito** específico.

# DBMS: Sistema de gestión de base de datos

---

El sistema de gestión de la base de datos (SGBD) es una aplicación que permite a los usuarios definir, crear, mantener, consultar y controlar el acceso a bases de datos.

- **Definir**: especificar tipos de datos, estructuras, relaciones, restricciones.
- **Crear**: generar la estructura definida, poblar datos iniciales
- **Mantener**: realizar acciones de actualización, inserción, eliminación
- **Consultar**: recuperar información según criterios de selección
- **Controlar el acceso**: garantizar que cada tipo de acción posible solamente es realizada por los grupos de usuarios autorizados.

# DBMS: Sistema de gestión de base de datos

---

En el modelo seguido con los sistemas de bases de datos se da una implementación interna de un objeto y una **especificación externa separada**.

Los usuarios del objeto sólo ven la especificación externa y no se deben preocupar de cómo se implementa internamente el objeto.

Una ventaja de este modelo, conocido como **abstracción de datos**, es que se puede *cambiar* la *implementación* interna de un objeto *sin afectar* a sus usuarios ya que la *especificación externa* no se ve alterada.

# Algunos SGBD

---

- MS **Access** (Microsoft)
- Base (Open/Libre Office)
- dBase IV (Borland)
- DB2 (IBM)
- Oracle (Oracle)
- SQL Server (Microsoft)
- **MySQL** (Oracle)
- **MariaDB**
- Postgre SQL
- **SQLite**
- HSQLDB
- ObjectDB
- **MongoDB**
- **Firestore**
- *Redis*
- Elasticsearch



# DBMS: Componentes o servicios

---

DDL: LENGUAJE DE DEFINICIÓN DE DATOS.

- DEFINE ELEMENTOS DE DATOS, ESTRUCTURAS, RELACIONES Y VALIDACIONES A NIVEL EXTERNO.

DML: LENGUAJE DE MANIPULACIÓN DE DATOS

- PERMITE BUSCAR, AÑADIR, SUPRIMIR Y MODIFICAR DATOS

SQL: LENGUAJE DE CONSULTA ESTANDAR DE LAS BD RELACIONALES

- FUNCIONALIDAD PARA CARGAR ARCHIVOS, REALIZAR COPIAS DE SEGURIDAD, GESTIÓN DE CONCURRENCIA Y ESTADISTICAS DE UTILIZACIÓN.

.

# DBMS: Ventajas

---

- CONTROL DE CONCURRENCIA
- CONTROL DE INTEGRIDAD DE INFORMACIÓN. CONSISTENCIA DE DATOS.
- RESTRICCIÓN DE ACCESOS NO AUTORIZADOS
- COPIAS DE SEGURIDAD Y PLAN DE MANTENIMIENTO
- OPTIMIZADOR DE CONSULTAS

# DBMS: Ventajas

---

El SGBD proporciona acceso controlado a la BD mediante:

Un sistema de **seguridad**, de modo que los usuarios no autorizados no puedan acceder a la base de datos.

Un sistema de **integridad** que mantiene la integridad y la **consistencia** de los datos.

Un sistema de control de **concurrency** que permite el acceso compartido a la base de datos.

Un sistema de control de **recuperación** que restablece la BD tras producirse un fallo del hardware o del software.

Un **diccionario** de datos o *catálogo*, accesible por el usuario, que contiene la descripción de los datos de la base de datos..

# Arquitectura de un DBMS: ESQUEMA

---

La arquitectura ANSI/SPARC se divide en tres niveles, conocidos como interno, conceptual y externo, respectivamente. Hablando en términos generales:

El **nivel interno** (también conocido como el nivel *físico*) es el que está más cerca del almacenamiento físico; es decir, es el que tiene que ver con la forma en que los datos están almacenados físicamente.

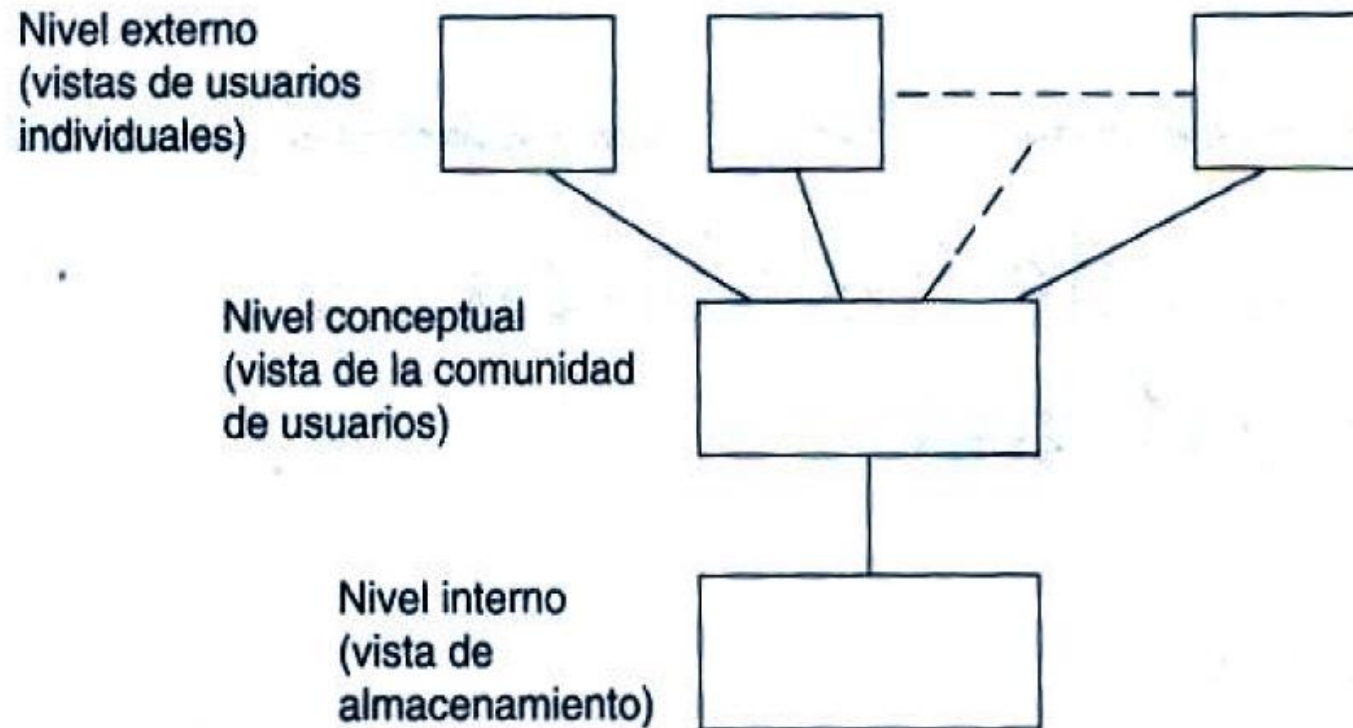
El **nivel externo** (también conocido como el nivel *lógico de usuario*) es el más próximo a los usuarios; es decir, el que tiene que ver con la forma en que los usuarios individuales ven los datos.

El **nivel conceptual** (también conocido como el nivel *lógico de la comunidad*, o en ocasiones sólo como el nivel *lógico*, sin calificar) es un nivel de indirección entre los otros dos.

Observe que el nivel externo tiene que ver con las percepciones de usuarios *individuales*, mientras que el nivel conceptual tiene que ver con la percepción de una *comunidad* de usuarios.

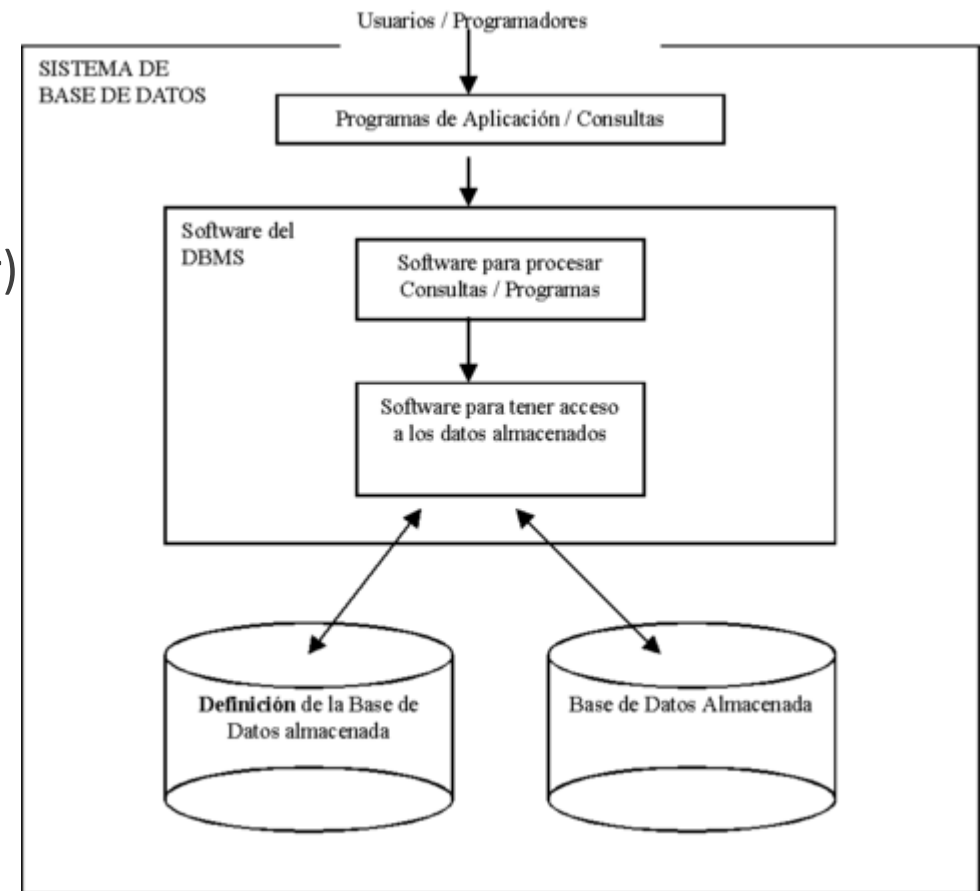
# Arquitectura de un DBMS: Esquema

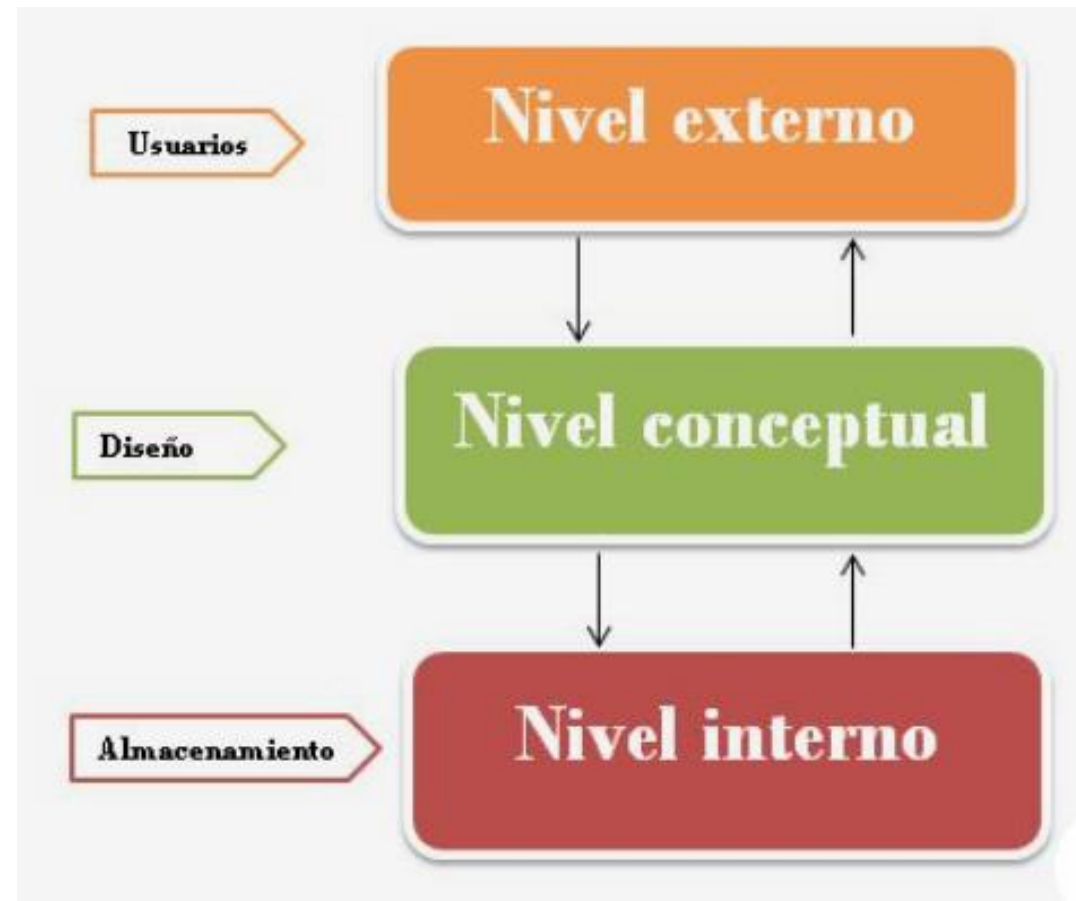
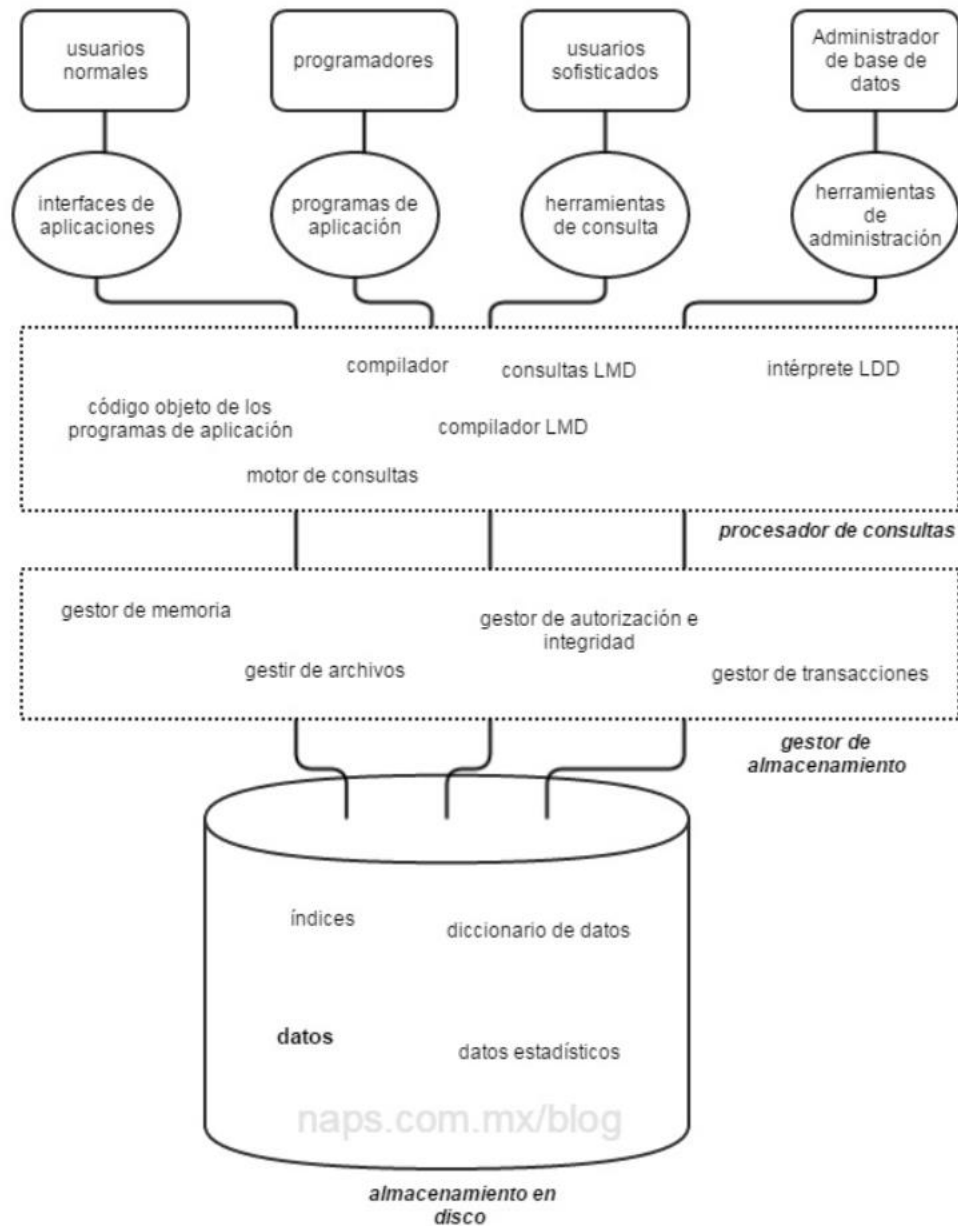
---



# Arquitectura de un DBMS: Esquema

- PERSONAS
- PROGRAMAS: MANEJADOR DE DATOS
  - (SQL, ORACLE, MYSQL)
- MAQUINAS (Soft para procesar, Soft para acceder)
- DATOS: BASE DE DATOS





# Clasificación de un DBMS

## SEGÚN EL MODELO LÓGICO

- JERARQUICO
- EN RED
- RELACIONAL
- ORIENTADO A OBJETOS

## NÚMERO DE USUARIOS

- MONOUSUARIO
- MULTIUSUARIO

## NÚMERO DE SITIOS

- CENTRALIZADO
- DISTRIBUIDO





# Independencia en los datos

---

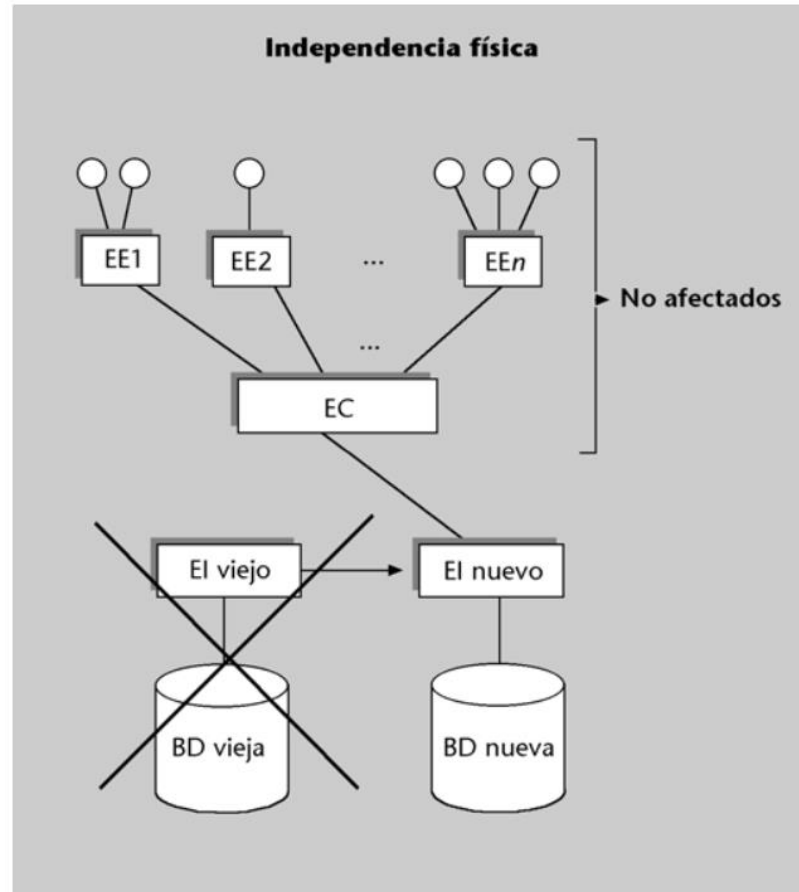
La arquitectura de tres niveles proporciona dos tipos de independencia:

- Física.
- Lógica.

Hay independencia física cuando los cambios en la organización física de la BD no afectan al mundo exterior (es decir, los programas usuarios o los usuarios directos).

O sea, habrá independencia física cuando los cambios en el esquema interno no afecten al esquema conceptual ni a los esquemas externos.

# Independencia en los datos



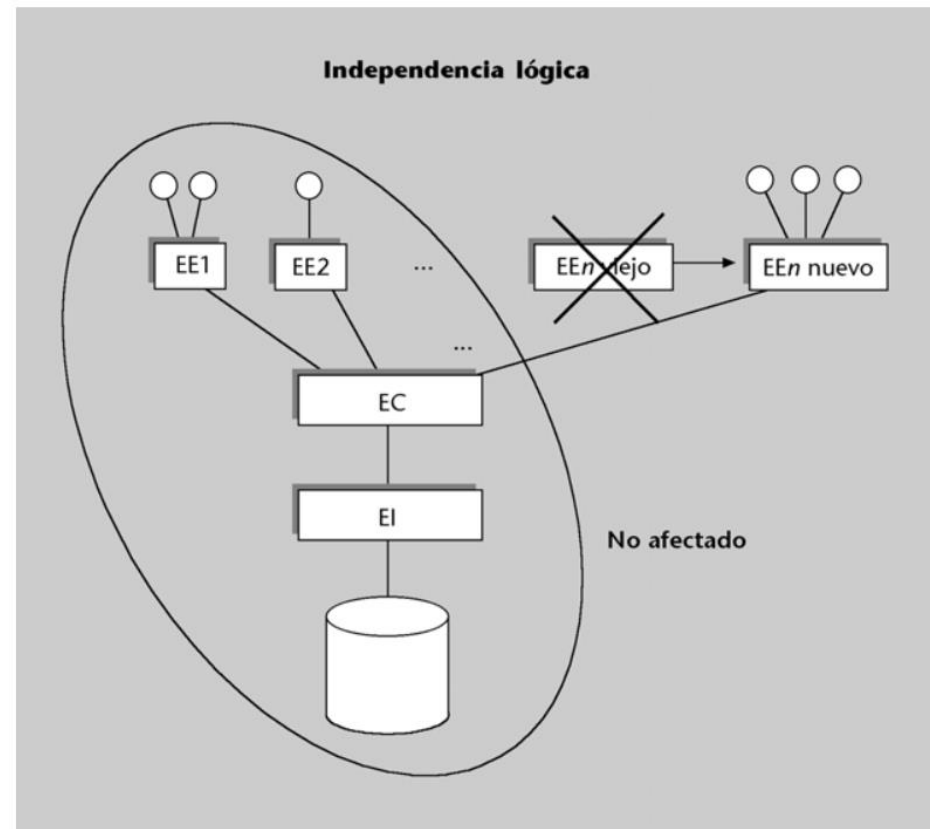
# Independencia en los datos

---

Si hay independencia física de los datos, lo único que variará al cambiar el esquema interno son las correspondencias entre el esquema conceptual y el interno. Obviamente, la mayoría de los cambios del esquema interno obligarán a rehacer la BD real (la física).

Hay independencia lógica cuando los usuarios no se ven afectados por los cambios en el nivel lógico.

# Independencia en los datos



# Funciones del DBA (Administrador de bases de datos)

---

¿Y el **Administrador de Bases de Datos**? Entre sus tareas se encuentran las siguientes:

- Crear el esquema de la base de datos.
- Definir la estructura y el método de acceso.
- Modificar el esquema y la organización física.
- Autoriza el acceso a la base de datos.
- Da mantenimiento de rutina a la B.D., lo que incluye hacer copias de seguridad, asegurarse de que exista suficiente espacio libre, y supervisar el rendimiento del sistema.

# DBMS: DIFERENCIA SISTEMA DE ARCHIVOS

---

SISTEMA DE ARCHIVOS	DBMS
MANEJO DE ARCHIVOS EN TEXTO PLANO	ARCHIVOS COMPARTIDOS
PUEDE HABER DATOS REPETIDOS	NO SE REPITEN DATOS, ESTÁN NORMALIZADOS
CUENTA CON REDUNDANCIA NO CONTROLADA Y ERROR DE REFERENCIAS	INTEGRIDAD REFERENCIAL
POCA SEGURIDAD	SEGURIDAD DEFINIDA POR USUARIO Y NIVELES
COMPONENTES: ARCHIVOS, REGISTROS Y CAMPOS	BASE DE DATOS Y USUARIOS