

# Bases de Datos

TEMA 1

OLIVER BITICA – 1ºDAM

# Contenido

Leyenda de Marcadores .....	2
Esquema Detallado sobre Almacenamiento de Información y Bases de Datos.....	2
Parte I: Sistemas de Almacenamiento Basados en Ficheros .....	3
1. Conceptos Fundamentales de Ficheros .....	3
2. Clasificación de Ficheros por su Función .....	4
3. Soportes y Métodos de Acceso .....	5
4. Parámetros de Utilización de Ficheros .....	5
Parte II: Fundamentos de Bases de Datos .....	6
1. Introducción y Conceptos Clave.....	6
2. Usos y Usuarios de las Bases de Datos .....	6
3. Modelos de Bases de Datos.....	6
4. Sistemas Gestores de Bases de Datos (SGBD) .....	8
5. Arquitecturas de Almacenamiento .....	11

## Leyenda de Marcadores

- **Amarillo**: Para conceptos fundamentales, definiciones y puntos clave.
  - **Verde**: Para ventajas, objetivos, propósitos y características positivas.
  - **Azul**: Para tipos, clasificaciones, componentes, estructuras y ejemplos.
  - **Rojo/Salmón**: Para problemas, inconvenientes o limitaciones.
  - **Gris**: Para tecnologías específicas, nombres propios, estándares y secciones marcadas como “Contenido Prioritario”.
- 

# Esquema Detallado sobre Almacenamiento de Información y Bases de Datos

Este documento presenta un esquema jerárquico y detallado sobre los conceptos fundamentales del almacenamiento de información, desde los sistemas basados en ficheros hasta las arquitecturas de bases de datos modernas. El contenido se ha estructurado para **facilitar la comprensión, el estudio y la planificación de proyectos**, con un enfoque particular en los temas clave señalados como prioritarios.

# Parte I: Sistemas de Almacenamiento Basados en Ficheros

Esta sección aborda los métodos tradicionales de almacenamiento de datos, que sentaron las bases para los sistemas de información posteriores, centrados en la organización y gestión mediante ficheros.

## 1. Conceptos Fundamentales de Ficheros

- **Introducción:** El almacenamiento y gestión de datos es una constante en la vida cotidiana. La evolución desde los archivos en papel ha sido impulsada por la necesidad de un tratamiento más eficiente de la información. La primera informatización en los años setenta se centró en trasladar los datos del papel a un formato digital para un acceso más rápido.
- **Definición:** Un fichero es el elemento que permite el almacenamiento de datos de forma permanente en dispositivos de memoria masiva para su posterior recuperación, consulta y procesamiento.
- **Estructura Jerárquica:**
  - **Fichero:** Colección de datos relacionados.
  - **Registro:** Contienen datos relativos a un mismo elemento u objeto (p. ej., datos de un usuario).
  - **Campo:** Divisiones dentro de un registro que contienen información elemental (p. ej., nombre o email del usuario).
- **Transferencia de Datos:**
  - **Bloque:** Cantidad de información transferida entre el soporte de almacenamiento y la memoria principal en una única operación de lectura/escritura.
  - **Empaquetado (Blocaje):** Operación de agrupar varios registros lógicos en un bloque.
  - **Factor de Blocaje:** Número de registros que caben en un bloque.

## 2. Clasificación de Ficheros por su Función

Los ficheros se pueden clasificar según la función que desempeñan dentro de una aplicación.

Tipo de Fichero	Subtipo	Descripción	Ejemplo
Permanentes	Maestros	Contienen el estado actual de los datos que pueden ser modificados por la aplicación.	Fichero de datos de usuarios de una plataforma educativa.
	Constantes	Incluyen datos fijos que no suelen ser modificados y se acceden para consulta.	Fichero con códigos postales.
	Históricos	Almacenan datos que fueron actuales en un momento anterior; se usan para reconstruir situaciones.	Fichero con los usuarios dados de baja de una plataforma.
Temporales	Intermedios	Almacenan resultados de una aplicación que serán utilizados por otra.	
	De maniobra	Guardan datos de una aplicación que no caben en memoria principal por falta de espacio.	
	De resultados	Acumulan datos que serán transferidos a un dispositivo de salida.	

### 3. Soportes y Métodos de Acceso

- **Tipos de Soportes Físicos:**

- **Soportes Secuenciales:** El acceso a los datos se realiza de **forma lineal, desde el principio hasta la posición deseada**. Se utilizan principalmente para **copias de seguridad**. (*Ejemplo: Cintas magnéticas.*)
- **Soportes Direccionables (de Acceso Directo):** Permiten **posicionarse directamente en la ubicación del dato de interés** sin recorrer la información previa. Son los **más empleados en la actualidad**. (*Ejemplos: Discos magnéticos, discos ópticos, SSD.*)

- **Organizaciones de Ficheros (Métodos de Acceso):**

- **Ficheros Secuenciales:** Los registros **se almacenan de forma contigua y se leen uno tras otro desde el inicio**. Utiliza una marca de fin de fichero (EOF).
- **Ficheros de Acceso Directo (o Aleatorio):** Permite **acceder a un registro directamente** indicando su posición relativa o, más comúnmente, **a través de una clave**. Requiere **dispositivos de acceso directo**.
- **Ficheros Indexados:** Se basa en el **uso de índices, que son tablas que relacionan la clave de un registro con su posición física** en el fichero. Consta de una **zona de datos** y una **zona de índices**.
- **Ficheros Secuenciales Indexados:** **Combina las organizaciones secuencial e indexada**. La zona de datos se divide en segmentos, y el índice apunta al inicio de cada segmento. La búsqueda es **directa hasta el segmento y secuencial dentro de él**.
- **Ficheros con Acceso Calculado (Hash):** Utiliza una **función de hashing para convertir la clave directamente en su dirección física**. El problema principal son las **colisiones** (sinónimos), que se resuelven con técnicas como **zonas de excedentes**.

### 4. Parámetros de Utilización de Ficheros

Estos parámetros ayudan a determinar qué tipo de organización de fichero es más adecuada.

- **Capacidad:** Espacio total que ocupa el fichero.
- **Actividad:** Frecuencia de consultas y modificaciones. Se mide con la **Tasa de Actividad** y la **Frecuencia de Acceso**.
- **Volatilidad:** Frecuencia de inserciones y borrados. Se mide con la **Tasa de Renovación** y la **Frecuencia de Renovación**.
- **Crecimiento:** Variación de la capacidad del fichero, medida por la **Tasa de Crecimiento**.

## Parte II: Fundamentos de Bases de Datos

Esta sección se centra en la definición, modelos, arquitecturas y tipos de los Sistemas Gestores de Bases de Datos (SGBD), destacando los temas de mayor relevancia.

### 1. Introducción y Conceptos Clave

- **Definición de Base de Datos:** Una **colección de datos relacionados lógicamente** entre sí, con una definición y descripción comunes, estructurados de una determinada manera, almacenados con **mínima redundancia** y que permiten un **acceso eficiente**.
- **Componentes Elementales:**
  - **Entidad:** Objeto del mundo real o abstracto sobre el que se almacena información (ej. un cliente, un producto, un doctor).
  - **Atributo:** Propiedad o característica de una entidad (ej. nombre, precio, raza, color).
  - **Metadatos:** **Datos que describen los datos** (su estructura, tipos, relaciones). Se almacenan en el diccionario de datos o catálogo y permiten la **independencia entre la lógica del programa y el almacenamiento físico**.

### 2. Usos y Usuarios de las Bases de Datos

- **Tipos de Usuarios:**
  - **Administrador (DBA):** Responsable de la implementación física, seguridad y rendimiento.
  - **Diseñadores:** Definen la estructura lógica y los procesos de la base de datos.
  - **Programadores de Aplicaciones:** Crean los programas que interactúan con la BD.
  - **Usuarios Finales:** Consumen la información a través de las aplicaciones.
- **Ámbitos de Aplicación:** Las bases de datos son **omnipresentes** en sectores como la banca, líneas aéreas, telecomunicaciones, medicina, justicia, etc.

### 3. Modelos de Bases de Datos

Un modelo de datos **define la estructura lógica de una base de datos**.

### 3.1. Modelo Relacional (Contenido Prioritario - Sección 6.3)

- **Contexto:** Desarrollado por Codd en 1970, es el **modelo más extendido** en la actualidad.
- **Estructura Lógica:** La información se organiza en **tablas bidimensionales**, también conocidas como relaciones.
- **Terminología Clave:**
  - **Tabla (Relación):** Estructura principal que contiene los datos.
  - **Registro, Entidad o Tupla:** Cada fila de la tabla.
  - **Campo o Atributo:** Cada columna de la tabla.
  - **Dominio:** Conjunto de valores que puede tomar un atributo.
  - **Clave:** Atributo o conjunto de atributos que **identifica de forma única a una tupla**.
- **Requisitos Fundamentales de las Tablas:**
  1. Todos los registros son del mismo tipo.
  2. La tabla solo puede tener un tipo de registro.
  3. No existen campos o atributos repetidos.
  4. No contienen tuplas duplicadas.
  5. No existe un orden específico en el almacenamiento de los registros.
- **Lenguaje de Consulta:** El lenguaje habitual para consultas es **SQL** (Structured Query Language).

### 3.2. Modelo Orientado a Objetos (Contenido Prioritario - Sección 6.4)

- **Definición:** Define una base de datos en términos de **objetos, sus propiedades y sus operaciones**.
- **Conceptos Fundamentales:**
  - **Objeto:** Unidad básica que **combina estructura (datos) y comportamiento (operaciones)**.
  - **Clase:** Plantilla que define la estructura y comportamiento de un conjunto de objetos similares. Las clases se organizan en jerarquías.
  - **Método:** Procedimiento o función asociada a una clase que define sus operaciones.
  - **Encapsulación:** Propiedad que **oculta la información interna del objeto**, impidiendo accesos incorrectos y permitiendo la interacción solo a través de sus métodos.
  - **Herencia:** Mecanismo que permite a una clase (subclase) **heredar propiedades y métodos de otra (superclase)**, fomentando la reutilización.
  - **Polimorfismo:** Propiedad de una operación que le permite ser **aplicada a distintos tipos de objetos**.
- **Relevancia:** Este modelo es considerado el **fundamento de las bases de datos de tercera generación** y cubre limitaciones del modelo relacional.

### 3.3. Otros Modelos de Bases de Datos

- **Modelo Jerárquico:** Organiza los datos en una estructura de árbol (padre-hijo), donde un hijo solo tiene un padre.
- **Modelo en Red:** Similar al jerárquico, pero permite que un nodo hijo tenga múltiples padres.
- **Objeto-Relacional:** Híbrido que combina la simplicidad del modelo relacional con las capacidades del modelo orientado a objetos.
- **Orientado a Documentos:** Almacena datos en documentos semi-estructurados (p. ej., JSON, XML). Ejemplos: MongoDB, CouchDB.
- **Multidimensional:** Estructura los datos en cubos (dimensiones) para análisis complejos (OLAP).
- **Transaccional:** Optimizado para la velocidad en el procesamiento de un gran volumen de transacciones.

## 4. Sistemas Gestores de Bases de Datos (SGBD)

- **Definición:** Conjunto de programas y herramientas que permite a los usuarios crear y mantener una base de datos, facilitando la definición, construcción, manipulación y compartición de la misma.

### 4.1. Funciones del SGBD (Contenido Prioritario - Sección 7.1)

El SGBD desarrolla tres funciones fundamentales a través de lenguajes específicos:

- **Función de Descripción o Definición:**
  - **Propósito:** Permite al diseñador crear las estructuras de la base de datos (tablas, vistas, índices) y definir las relaciones y restricciones.
  - **Lenguaje:** Se realiza mediante el Lenguaje de Descripción de Datos (DDL - Data Definition Language).
- **Función de Manipulación:**
  - **Propósito:** Permite a los usuarios añadir, suprimir, modificar o consultar los datos de la base de datos.
  - **Lenguaje:** Se implementa a través del Lenguaje de Manipulación de Datos (DML - Data Manipulation Language).
- **Función de Control:**
  - **Propósito:** Permite al administrador establecer mecanismos de protección (seguridad), integridad y control de acceso concurrente. Incluye la gestión de copias de seguridad, recuperación ante fallos y auditoría.
  - **Lenguaje:** Se gestiona con el Lenguaje de Control de Datos (DCL - Data Control Language).
- **SQL (Structured Query Language):** Es el lenguaje estandarizado que proporciona sentencias para realizar las tres funciones (DDL, DML y DCL).

## 4.2. Arquitectura del SGBD (Contenido Prioritario - Sección 7.3)

La arquitectura estándar, basada en el modelo ANSI/SPARC, propone una estructura de tres niveles para simplificar la interacción y proporcionar independencia.

- **Arquitectura de Tres Niveles:**

- **Nivel Interno o Físico:** Describe la estructura física de almacenamiento de la base de datos. Es el nivel más cercano al hardware y detalla cómo se guardan los datos (ficheros, índices, etc.).
- **Nivel Lógico o Conceptual:** Describe la estructura completa de la base de datos para la comunidad de usuarios. Define las entidades, relaciones, operaciones y restricciones, ocultando los detalles de la implementación física.
- **Nivel Externo o de Visión del Usuario:** Describe las diferentes vistas que los usuarios o aplicaciones tienen de la base de datos. Cada vista muestra solo la parte de la base de datos que es relevante para un grupo de usuarios específico.

- **Independencia de Datos:** Esta arquitectura consigue dos niveles de independencia cruciales:

- **Independencia Lógica:** Permite modificar el esquema conceptual sin necesidad de alterar los esquemas externos o las aplicaciones existentes.
- **Independencia Física:** Permite modificar el esquema interno sin que afecte al esquema conceptual.

#### 4.3. Tipos de SGBD (Contenido Prioritario - Sección 7.4)

Los SGBD se pueden clasificar atendiendo a diferentes criterios:

Criterio	Tipos	Descripción
1. Modelo Lógico	Relacional, Orientado a Objetos, Jerárquico, En Red.	Se basa en el modelo de datos subyacente. El relacional es el más utilizado; el orientado a objetos gana terreno.
2. Número de Usuarios	Monousuario, Multiusuario.	Define si el sistema atiende a un solo usuario a la vez o a múltiples usuarios de forma concurrente.
3. Distribución	Centralizados, Distribuidos.	<b>Centralizados:</b> Los datos se almacenan en un solo computador. <b>Distribuidos:</b> Los datos se reparten en varios sitios o nodos.
4. Costo	Comerciales, Libres/Gratuitos.	Clasificación según el modelo de licencia (p. ej., Oracle vs. PostgreSQL).
5. Propósito	Propósito General, Propósito Específico.	<b>General:</b> Utilizable para cualquier tipo de aplicación. <b>Específico:</b> Diseñado para un problema concreto.

#### 4.4. Sistemas Comerciales vs. Sistemas Libres

Tipo	SGBD Ejemplos	Descripción
Comerciales	Oracle, Microsoft SQL Server, DB2, Informix	Productos con licencia que ofrecen soporte y funcionalidades avanzadas.
Libres	MySQL, PostgreSQL, Firebird, SQLite, Apache Derby	Sistemas de código abierto, a menudo gratuitos, con una gran comunidad de soporte.

## 5. Arquitecturas de Almacenamiento

- **Tecnologías de Almacenamiento:**

- **Discos SATA y SCSI:** Interfaces estándar para discos duros.
- **RAID (Redundant Array of Independent Disks):** Tecnología que combina múltiples discos para mejorar el rendimiento, la seguridad o ambos.
- **NAS (Network Attached Storage):** Dispositivo de almacenamiento conectado a una red que permite el acceso centralizado a los datos.
- **SAN (Storage Area Network):** Red de alta velocidad dedicada exclusivamente a conectar servidores y dispositivos de almacenamiento masivo.

- **Bases de Datos Centralizadas:**

- **Arquitectura:** La arquitectura completa del sistema (SGBD, datos) reside en un único computador. No hay múltiples elementos de procesamiento ni mecanismos de intercomunicación como en las bases de datos distribuidas.

- **Bases de Datos Distribuidas (BDD):**

- **Arquitectura:** Conjunto de múltiples bases de datos lógicamente relacionadas que se encuentran distribuidas en diferentes nodos de una red de comunicaciones.
- **Fragmentación:** Técnica utilizada en BDD que consiste en particionar las tablas de la base de datos y distribuir los fragmentos en diferentes nodos de la red para optimizar las consultas y el almacenamiento.