

WUOLAH



cg_enri

www.wuolah.com/student/cg_enri



6236

Tema4sistemasarchivos.pdf

Tema 4 Sistemas de archivos



1º Fundamentos del Software



Grado en Ingeniería Informática



Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Universidad de Granada



Escuela de **LÍDERES**

Master en Marketing Digital

Rafael García Parrado
Director de
Marketing Digital



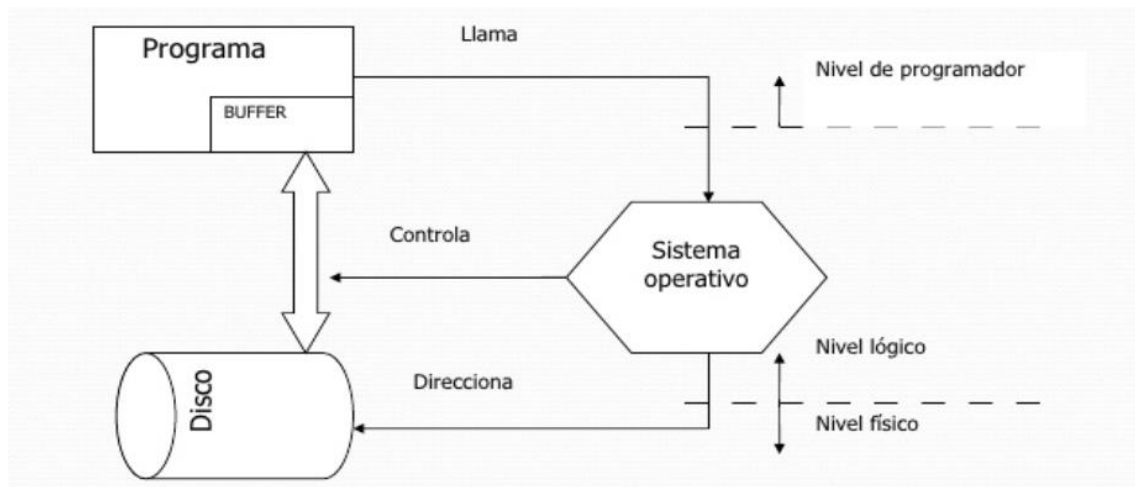
Tema 4: Sistemas de archivos

Concepto de archivo

- Un **archivo** es un conjunto de información sobre un mismo tema, tratada como una unidad de almacenamiento en memoria secundaria y organizada de forma estructurada para la búsqueda de un dato individual. Un archivo está compuesto por registros homogéneos que contiene la información correspondiente a un elemento individual.
- Un **registro** es una estructura dentro del archivo que contiene la información correspondiente a un elemento individual. Se puede dividir en **campos**, son datos que presentan una información unitaria o independiente.

Operaciones sobre archivos

- El SO puede aludir a un archivo mediante un nombre
- Todo archivo tiene asociados atributos: tamaño, fecha, permisos, ...etc.
- El SO proporciona órdenes para operar sobre archivos: copiar, pegar, ...etc.



- El SO transporta, al acceder a un archivo, una cantidad fija de información.
- El SO transforma la dirección lógica usada en los programas de usuario en dirección física con la que se direcciona en el dispositivo.
- Un archivo es una estructura de datos externa al programa que lo usa; en las operaciones de lectura/escritura se transfiere información a/desde un buffer en memoria principal asociado a las operaciones de entrada/salida sobre un archivo.



INEAF
BUSINESS SCHOOL

Es el momento
DE CRECER

Master en Asesoría Fiscal de Empresas



Clasificación de archivos según el tipo de registros

- Longitud fija
- Longitud variable:
 - Con delimitador: un determinado carácter llamado delimitador marca el fin de un registro (suelen usarse como delimitadores el salto de línea, nulo, etc.).
 - Con cabecera: cada registro contiene un campo inicial que almacena el número de bytes del registro.
- Longitud indefinida: el sistema operativo no realiza ninguna gestión sobre la longitud de los registros ya que el archivo no tiene realmente ninguna estructura interna.
 - En cada operación de lectura/escritura se transfiere una determinada subcadena del archivo y será el programa de usuario quien indique al SO el principio y final de ese registro.

Una posibilidad añadida: disponer de un **campo clave** (o llave) que permita localizar rápidamente un registro.

Concepto de directorio

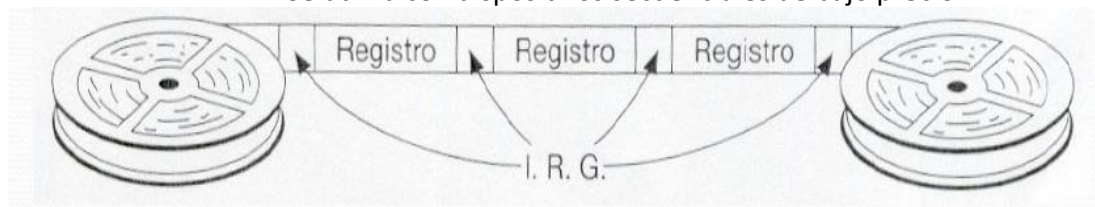
Es un archivo que permite agrupar archivos. Se utilizan diferentes conceptos dentro de los directorios:

- Directorio actual
 - Home
 - Rutas absolutas y relativas
 - Lista de búsqueda
 - Enlace duro, enlace simbólico
- Los maneja el SO

Organización de archivos

Organización secuencial

- Los registros se almacenan de forma contigua
 - Es adecuada si:
 - + Se quieren leer los registros tal y como están ordenados
 - + Se quiere leer la mayoría de los registros
 - Ventajas:
 - + Aprovecha espacio
 - + Fácil de utilizar
 - + Se utiliza con dispositivos secuenciales de bajo precio



Operaciones sobre archivos

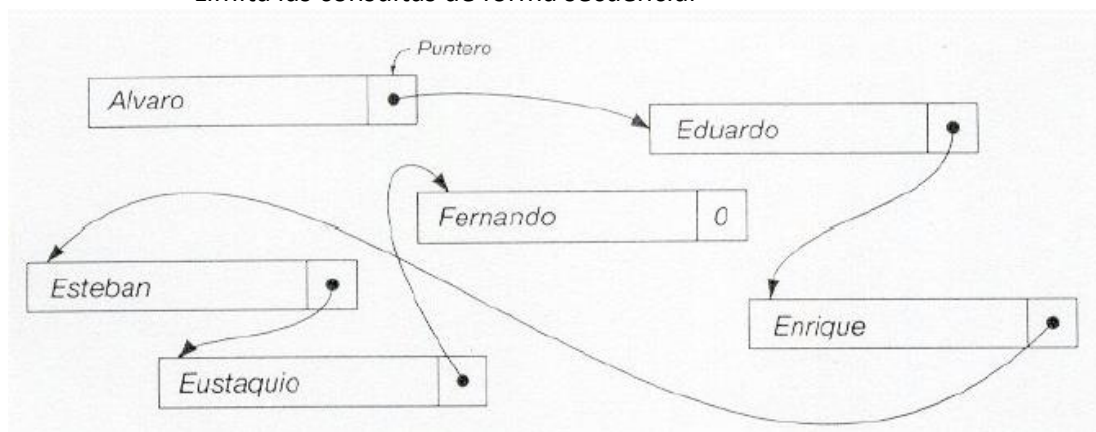
- Añadir: al final del archivo

- Consultar: leer registro $n \rightarrow$ leer $n-1$ registros anteriores
- Modificación: si no se aumenta la longitud
- Eliminación: solo borrado lógico (no se puede leer)
- Inserción: un archivo nuevo con las actualizaciones

Organización secuencial encadenada

Con cada registro se almacena un puntero con la dirección física del siguiente. El último tiene una marca especial:

- Ventajas:
 - Facilidad de inserción y eliminación
- Inconvenientes:
 - Limita las consultas de forma secuencial



Operaciones sobre archivos

- Consulta: consulta secuencial
- Inserción:
 1. Localiza la posición de inserción
 2. Escribe el registro en una zona libre de memoria
 3. Asigna al nuevo registro la dirección física del siguiente

Organización secuencial indexada

Hay dos estructuras:

- Zona de registros: se direccionan los registros del archivo derivada en tramos. Registros ordenados según el valor de una llave.
- Zona de índices: por cada tramo de la zona de registros, contiene:
 - El mayor valor de la llave del tramo
 - Dirección del primer registro del tramo

Estas estructuras se gestionan mediante el SO y un software especial

Se inserta un registro con llave "Gato".

Dir. Física	LLave	Puntero
1	Alba	4
2	Elefante	3
3	León	8
4	Coral	2
5	Mosca	6
6	Ñu	7
7	Pantera	0
8	Morsa	5
9		
10		

Paso 1

Dir. Física	LLave	Puntero
1	Alba	4
2	Elefante	3
3	León	8
4	Coral	2
5	Mosca	6
6	Ñu	7
7	Pantera	0
8	Morsa	5
9	Gato	3
10		

Pasos 2 y 3

Dir. Física	LLave	Puntero
1	Alba	4
2	Elefante	9
3	León	8
4	Coral	2
5	Mosca	6
6	Ñu	7
7	Pantera	0
8	Morsa	5
9	Gato	3
10		

Paso 4

Operaciones sobre archivos

- Consulta: localiza un archivo conocida su llave, no lee necesariamente los registros que se encuentran en ese tramo:
 - Lee las llaves hasta encontrar una mayor o igual
 - Obtiene la dirección del comienzo del tramo
 - Lee secuencialmente el tramo hasta encontrar el registro o uno con llave mayor.
- Inserción: no se pueden insertar registros, solo crea al final de la zona de registros, pero no se podrán consultar.
- Borrado: borrado lógico.
- Modificación: si no aumenta el tamaño.

Organización directa o aleatoria

Archivo escrito sobre un soporte de acceso directo para el cual existe una función de transformación que genera la dirección de cada registro, a partir de un campo que actúa como llave.

Útil para archivos donde los accesos se realizan por llave, accediéndose a registros concretos.



Problemas de los sinónimos

El problema es elegir la función de transformación o el método de direccionamiento; puede pasar:



- Direcciones que no se corresponden con ninguna llave
- Direcciones que se corresponden con más de una llave

Soluciones

- Si se asocia a una llave de dirección ya ocupada, buscar en el archivo hasta una posición libre
- Reservar la zona de desbordamiento para los registros que no se puedan escribir en la zona que les corresponde

Métodos de direccionamientos

- Directo: se utiliza la dirección de la llave (si es numérica), quedan direcciones sin utilizar
- Asociado: cualquier tipo de llave, debe construir una tabla que almacene para cada llave la dirección de registro correspondiente.
- Calculado: la dirección de registro se obtiene transformando la llave, se usa cuando la llave no es numérica. Toma valores en un rango inadecuado.

Operaciones sobre archivos

- Consulta: para leer un registro debe aplicarse a la llave un algoritmo de transformación, devuelve la dirección de registro.
- Borrado: lógico, se puede reutilizar el espacio.
- Modificación e inserción: transformando la llave correspondiente.

Bases de datos

Problemática

- Dificultad de mantenimiento: suele haber archivos duplicados
- Redundancia: si un dato se puede deducir de otro
- Rigidez de la búsqueda; hay que buscar datos, combinando acceso secuencial y directo
- Dependencia del programa: modificar la estructura de un archivo obliga a modificar todos los programas que lo usen
- Seguridad: deficiente en los sistemas de archivos

Una **base de datos** es una alternativa a los sistemas de archivos. Sistema formado por un conjunto de datos y un paquete de software para gestionarlo.

- Almacena datos redundantes
- Datos independientes a los programas
- Las relaciones se almacenan junto con los datos
- Se pueden acceder a ellos de diversas formas

Requisitos que debe cumplir una base de datos

- Acceso y utilización múltiple
- Flexibilidad
- Seguridad
- Interfaz de alto nivel
- Protección ante errores
- Independencia física y lógica

- Redundancia controlada
- Interrogación directa

Estructura

Uno o más atributos de una entidad es un identificador (determina de forma unívoca a los elementos de la entidad), se almacenan las relaciones que hay entre entidades.

Niveles de abstracción de la información

- Nivel de vista: describe diferentes vistas o subesquemas cada una correspondiente a la parte de la base de datos que interesa a un determinado grupo de usuarios.
- Nivel conceptual: describe el esquema de la base de datos, especifica todos los datos guardados en ella.
- Nivel físico: describe como se almacenan los datos.

Modelo de datos

Describen la estructura de una base de datos.

Clasificación:

- Modelo lógico basado en objetos: estructuración flexible y permite especificar limitaciones de los datos.
- Modelo lógico basado en registros: permite una estructura lógica, pero sin especificar limitaciones.
- Modelo físico de los datos: describe los datos en el nivel de implementación de los sistemas de base de datos.

Modelo entidad-relación

- **Entidad:** objeto que tiene existencia propia, que puede distinguirse de otros del cuál se quiere almacenar información de ciertas características.
- **Conjunto de entidades:** grupo de entidades del mismo tipo que representa la estructura genérica de una entidad de interés. Las entidades pueden pertenecer a más conjunto de entidades.
- **Relación:** asociación entre entidades.
- **Conjunto de relaciones:** grupo de relaciones del mismo tipo que representa la estructura genérica de las relaciones entre conjuntos de entidades.
- **Grado de una relación:** número de tipos de entidad que intervienen en un tipo de relación.
- **Atributo:** unidad básica de información sobre un tipo de entidad o un tipo de relación.

La **cardinalidad** expresa el número de entidades con las que puede asociarse o corresponderse una determinada entidad mediante una relación.

En el caso de un conjunto binario de relaciones R entre los conjuntos de relaciones A y B, los tipos de correspondencias o cardinalidad pueden ser:

- **Uno a uno (1:1):** a cada entidad de A le pueden corresponder una única entidad B y viceversa.

- **Uno a muchos (1:N):** a cada entidad de A le puede corresponder más de una entidad de B, pero cada entidad de B solo puede asociarse con una única entidad de A.
- **Muchos a uno (N:1):** a cada entidad de A le puede corresponder una única entidad de B, pero cada entidad de B puede asociarse con varias entidades de A.
- **Mucho a muchos (N:M):** a cada entidad de A le puede corresponder más de una entidad de B, y a cada entidad de B le puede corresponder más de una entidad de A.

Diagrama entidad-relación

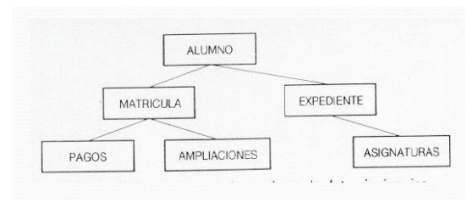
Representación gráfica de la estructura de una base de datos organizada según el Modelo entidad-relación.

Sus componentes principales son:

- **Rectángulos:** representa conjuntos de entidades.
- **Rombos:** representa conjuntos de relaciones.
- **Elipses:** representan atributos (de conjuntos de entidades o de conjuntos de relaciones).
- **Representación de la cardinalidad:** se puede hacer de varias maneras:
 - Con una etiqueta asociada al conjunto de relaciones (solo para relaciones binarias).
 - Poniendo una punta de flecha que señale hacia el conjunto de relaciones que participa de forma “uno” y usando una línea sin punta en el caso de conjuntos de entidades que participan de forma “muchos”.

Modelos lógicos basados en registros: tipos de Bases de Datos

- **Modelo de datos jerárquico:** Permite especificar una **base de datos jerárquica**, donde se establece una relación jerárquica entre los datos en forma de árbol, y no es posible definir relaciones muchos a muchos.



- **Modelo de datos en red:** Permite especificar una **base de datos en red**, donde pueden darse relaciones binarias con cualquier cardinalidad (1:1, 1:N, N:1, N:N) y no es necesario que la estructura tenga forma de árbol.
- **Modelo de datos relacional:** Permite especificar una **base de datos relacional**, que estará formada por tablas. Una **tabla** es una estructura bidimensional formada por una sucesión de registros del mismo tipo. Si se imponen ciertas condiciones a las tablas, se pueden tratar como **relaciones** matemáticas. Las tablas deben cumplir las siguientes condiciones:
 1. Todos los registros de una tabla son del mismo tipo. Para almacenar registros de tipos distintos se usan tablas distintas.
 2. En ninguna tabla aparecen campos repetidos.
 3. En ninguna tabla existen registros duplicados.



4. El orden de los registros en la tabla es indiferente. En cada momento se pueden recuperar los registros en un orden particular.
5. En cada tabla hay una llave, formada por uno o varios campos.

Transformación del Modelo entidad-relación al Modelo de datos relacional

Dado un Diagrama entidad-relación, el paso a tablas o relaciones del Modelo de datos relacional se efectúa como sigue:

- **Conjuntos de entidades:**
 - + Se define una tabla para cada conjunto de entidades.
 - + Para cada atributo se define una columna en la tabla.
- **Conjuntos de relaciones sin atributos propios:**
 - + **Cardinalidad 1:1:** En la tabla de uno de los conjuntos de entidades (el que se considere principal) se añaden las columnas necesarias para albergar los atributos que forman la clave del otro conjunto de entidades.
 - + **Cardinalidad 1:N o N:1:** En la tabla del conjunto de entidades que participa de forma “muchos” se añaden las columnas necesarias para albergar los atributos que forman la clave del otro conjunto de entidades.
 - + **Cardinalidad N:M:** Se define una tabla propia para el conjunto de relaciones y, en ella, se definen las columnas necesarias para albergar los atributos que forman la clave de cada uno de los conjuntos de entidades que relaciona ese conjunto de relaciones.
 - + **Cardinalidad 1:1, 1:N o N:1:** Se puede seguir el mismo enfoque dado para conjuntos de relaciones sin atributos propios, pero añadiendo también las columnas necesarias para albergar los atributos del conjunto de relaciones en la misma tabla del conjunto de entidades donde se añada la clave del otro conjunto de entidades. Esta forma está desaconsejada desde el punto de vista conceptual y por cuestiones de mantenimiento del software.
 - + **Cardinalidad N:M, 1:1, 1:N o N:1 :** Se define una tabla propia para el conjunto de relaciones y, en ella, se definen las columnas necesarias para albergar los atributos que forman la clave de cada uno de los conjuntos de entidades que relaciona ese conjunto de relaciones. En esa misma tabla, además, se definen las columnas necesarias para albergar los atributos propios del conjunto de relaciones.

Sistema de gestión de la base de datos - SGBD ("Data Base Management System" - DBMS)

Conjunto de software destinado a la creación, control y manipulación de la información de una base de datos.

Un SGBD debe permitir la realización de las siguientes **tareas**:

1. **Definición** del esquema de la base de datos y de los distintos subesquemas.
2. **Acceso** a los datos desde algún lenguaje de alto nivel.
3. **Interrogación** (o recuperación de la información) directa en modo conversacional.

4. Organización física de la base de datos y recuperación tras fallos del sistema.

Lenguajes específicos en un SGBD

Las tres primeras tareas se realizan mediante dos lenguajes específicos:

- **Lenguaje de descripción de datos (DDL, "Data Description Language»):** Se usa para la descripción del esquema y de los subesquemas.
- **Lenguaje de manipulación de datos (DML, "Data Manipulation Language»):** Se utiliza para el acceso a la base de datos desde lenguajes de alto nivel o en modo conversacional.

El **sistema de gestión de la base de datos** actúa como intermediario entre los programas de aplicación y el sistema operativo, lo que permite que los programas sean independientes de la estructura física de los datos.