

## **Clase 11**

### **Archivos y sistemas de archivos**

#### **11.1 Introducción**

En la mayoría de las aplicaciones, el archivo es el elemento central. Con la excepción de las aplicaciones de tiempo real y algunas aplicaciones especializadas, la entrada a la aplicación se realiza mediante un archivo y, en prácticamente todas las aplicaciones, la salida se guarda en un archivo para un almacenamiento a largo plazo o para su acceso posterior por parte del usuario u otros programas.

Los archivos tienen vida fuera de cualquier aplicación individual que los utilice como entrada y/o salida. Los usuarios desean poder acceder a los ficheros, guardarlos y mantener la integridad de sus contenidos. Con el fin de lograr estos objetivos, prácticamente todos los sistemas operativos proporcionan programas que se ejecutan como aplicaciones privilegiadas. Sin embargo, como mínimo, un sistema de gestión de archivos necesita servicios especiales del sistema operativo; como máximo.

En esta clase consideraremos los elementos básicos de la gestión de archivos, así como los distintos tipos de archivos y principales métodos de compresión.

#### **11.2 Sistemas de archivos**

Desde el punto de vista del usuario, una de las partes más importantes de un sistema operativo es el sistema de archivos. El sistema de archivos proporciona las abstracciones de recursos típicamente asociadas con el almacenamiento secundario (generalmente disco rígido, SSD; también aplica a medios extraíbles).

El sistema de archivos permite a los usuarios crear colecciones de datos, llamadas archivos, con propiedades deseables, tales como las siguientes:

- Existencia a largo plazo. Los archivos se almacenan en disco u otro almacenamiento secundario y no desaparece cuando un usuario se desconecta.
- Compartible entre procesos. Los archivos tienen nombres y pueden tener permisos de acceso asociados que permitan controlar la compartición.
- Estructura. Dependiendo del sistema de archivos, un archivo puede tener una estructura interna que es conveniente para aplicaciones particulares. Adicionalmente, los archivos se pueden organizar en

estructuras jerárquicas o más complejas para reflejar las relaciones entre los mismos.

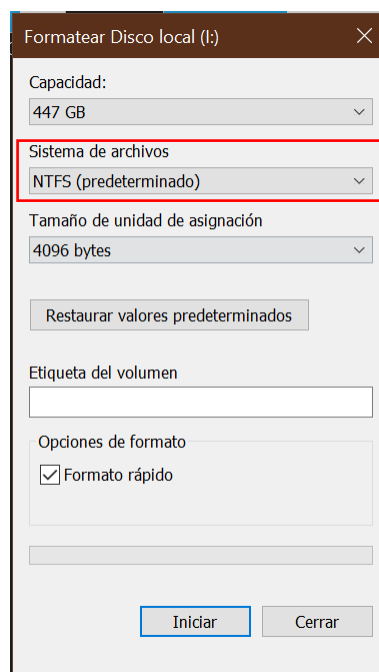
Para lograr su cometido, el sistema operativo utiliza las famosas «carpetas» o «directorios» con el fin de organizar todas las rutas y localizar la información contenida en el disco duro.

Básicamente, el sistema de archivos funciona a manera de una base de datos indexada en donde se encuentra la ubicación física de cada archivo adentro del disco duro de la computadora o cualquier otro dispositivo de almacenamiento, como pueden ser memorias USB, celulares, un servidor de red e incluso hasta cajeros automáticos y la computadora del auto.

Por regla general, sea cual sea el dispositivo de almacenamiento, la información guardada en él se organiza en directorios o carpetas, mismas que pueden contener subcarpetas y otros archivos.

En conclusión, el sistema de archivos de un sistema operativo es el que organiza todos los datos contenidos en el almacenamiento secundario de una forma determinada.

Cada unidad de almacenamiento tiene un sistema de archivos que es impuesto tras formatearla, pues habrán notado que al formatear un disco o pendrive nos da la opción de seleccionar su sistema de archivos, como se muestra, por ejemplo, en la siguiente imagen:



En esta clase repasaremos los siguientes sistemas de archivos:

- NTFS (New Technology File System).
- APFS (Apple File System) y HFS+ (Hierarchical File System plus).
- EXT4 (Fourth Extended Filesystem)
- FAT32.

### 11.3.1 NTFS

Se trata de un sistema de archivos que se utiliza desde Windows NT y NT 3.1; en otras palabras, fue introducido en 1993. Está basado en el antiguo sistema que usaba IBM (HPFS), que también fue usado en Mac OS.

Actualmente, es el sistema de archivos que Windows utiliza por defecto en los discos duros cuando instalamos su sistema operativo en un SSD o HDD.

Se caracteriza por no tener límites de tamaño por archivo y es una opción ideal para unidades de almacenamiento con gran capacidad que se vayan a utilizar en Windows.

Teóricamente, NTFS ofrece un tamaño de archivo de 16 EB, que equivalen a 160000000000 GB.

Desde Windows XP este sistema no ha variado demasiado.

Entre sus características más importantes, encontramos las siguientes:

- Compatibilidad con archivos dispersos.
- Cuotas de uso.
- Cifrado de archivo independiente.

NTFS está soportado tanto en escritorio, como en sistemas operativos para servidores. El problema viene con la incompatibilidad que presenta en otros sistemas operativos, como Mac. En OS X podemos leer toda la información de una unidad NTFS, pero no podemos escribir en ella.

Ventajas:

- No tiene límite por archivo.
- Es un sistema ideal para unidades que se utilicen en Windows.

- Desde Mac se puede leer toda la información.
- Es compatible con GNU/Linux.

Desventajas:

- En Mac no podremos escribir en un NTFS, ni es un SO compatible con este sistema. Aunque esto podría verse como una desventaja del MAC OS.
- En Smart TV antiguos, los USB tenían que ser FAT32, pero esto ya casi no sucede y los TV actuales también leen NTFS.

Desde el lanzamiento de Windows XP en el año 2001, NTFS es el estándar obligatorio en los sistemas operativos Windows. Por lo tanto es el sistema de archivos más extendido, dada la popularidad de Windows.

A continuación, una breve cronología del sistema NTFS:

Número de versión NTFS	Lanzamiento	Sistema operativo	Particularidades
1.0	1993	Windows NT 3.1	Primera versión; durante mucho tiempo incompatible con las versiones siguientes
1.1	1995	Windows NT 3.51	Por primera vez permite comprimir datos y controlar el acceso (derechos individuales de acceso a archivos)
1.2	1996	Windows NT 4.0	Incorporación de indicaciones de seguridad para archivos de sistema importantes; también llamada NTFS 4.0
3.0	2000	Windows 2000	Incorporación de varias funciones como la gestión del espacio de almacenamiento y el cifrado basado en el sistema de archivos; también llamada NTFS 5.0
3.1	2001	Windows XP	Ampliación de las entradas de la Master File Table (MFT) mediante números de entrada redundantes para facilitar la recuperación de entradas dañadas; también llamada NTFS 5.1

### 11.3.2 APFS y HFS+

Estos dos sistemas de archivo son propios de los sistemas de Apple.

APFS, o "Sistema de archivos Apple", es una de las nuevas características en Mac OS High Sierra (2017). Está optimizado para unidades de estado sólido (SSD) y otros dispositivos de almacenamiento con memoria flash, aunque también funciona en unidades mecánicas e híbridas.

El sistema APFS mejora la encriptación, el espacio compartido y la redimensión de las carpetas. La diferencia principal de APFS es que se enfoca en SSD o almacenamiento Flash, pero se puede utilizar en HDD igualmente.

APFS es más rápido: el copiar y pegar una carpeta completa es básicamente instantáneo, porque el sistema de archivos apunta básicamente a los mismos datos dos veces. Y las mejoras en los metadatos implican que es muy rápido en cuestiones tales como determinar cuánto espacio ocupa una carpeta en el disco. También hay una serie de mejoras de confiabilidad, por lo que los archivos dañados son mucho menos frecuentes.

Mac OS Extended, también conocido como HFS Plus o HFS +, es el sistema de archivos utilizado en todas las Mac desde 1998 hasta ahora, aunque en la actualidad le está dando paso al sistema APFS. A partir de Mac OS High Sierra, se utiliza este sistema únicamente en las unidades mecánicas e híbridas, y las versiones anteriores de macOS lo usan de manera predeterminada para todas las unidades.

El sistema HFS+ permite nombrar a un fichero con hasta 255 caracteres. Mejora a HFS usando una tabla de asignación de 32 bits, algo que era necesario cuando se empezaron a estandarizar los archivos de más de 1 GB. Soporta archivos de hasta 8 EB.



### 11.3.3 EXT4

## Ext4 File System

El sistema de archivos ext4 es la cuarta generación de la familia de sistemas de archivos ext. Es el sistema de archivos por defecto Linux.

El sistema ext4 puede leer y escribir en los sistemas de archivos predecesores, los ext2 y ext3.

Es un sistema de archivos del tipo con registro por diario (journaling). El journaling se basa en llevar un registro diario en el que se almacena la información necesaria para restablecer los datos del sistema afectados por un cambio, en caso de que falle.

Con el sistema de archivos ext4 se obtiene una gran reducción del tiempo necesario para recuperar un sistema de ficheros después de una falla. Por lo tanto, podemos decir que sus principales objetivos son la disponibilidad y confiabilidad.

Por ejemplo, en el caso en el que apaguemos nuestro equipo incorrectamente, ext4 se encargará de que al encenderlo lo tengamos igual que lo dejamos antes de apagarlo.

Por otro lado, ext4 (FOURTH EXTENDED FILESYSTEM) es una mejora de ext3 y ext2 que utiliza menos CPU y mejora la velocidad de lectura y escritura. Además, soporta volúmenes de hasta 1024 PiB (PebiByte) (1 PiB =  $2^{50}$  Bytes).

Por último, el sistema ext4 incorpora la capacidad de reservar un área contigua para un archivo, denominada "extents", la cual puede reducir y hasta eliminar completamente la fragmentación de archivos, mejorando el rendimiento al trabajar con archivos de gran tamaño.

#### **11.3.4 Fat32**

El sistema de archivos FAT32 (File Allocation Table 32 bits) fué lanzado en 1996 junto con Windows 95, sin embargo, sus orígenes se remontan al año 1977. En ese año Microsoft desarrolló el estándar File Allocation Table o FAT, el sistema de archivos que sigue siendo hoy en día el más usado en la industria. Gracias a las diferentes versiones de FAT, es posible aplicar el estándar salvando los obstáculos que suponen las diferencias entre sistemas operativos.

FAT32 es el último escalón en la evolución de los estándares FAT convencionales y tiene como predecesores a FAT12 y a FAT16. Más tarde, sin embargo, se desarrolló una especie de derivado de FAT, que fue bautizado como exFAT.

En la actualidad, el sistema FAT32 se sigue utilizando principalmente en los soportes de almacenamiento extraíbles como los pendrive USB, las tarjetas de memoria y los discos duros externos.

Otras veces, es necesario usar este sistema para transferir datos entre soportes antiguos y otros más recientes.

En los discos rígidos o ssd internos actuales de Windows, ya no se usa FAT32, como vimos, ya que desde Windows Vista el sistema de archivos estándar pasó a ser NTFS.

FAT32 aún sobrevive en soportes (especialmente las tarjetas SD o memorias flash) en los que las particiones de pequeño tamaño, de unos 32 GB, no suponen ningún problema, ya que el tamaño y la cantidad de los archivos durante mucho tiempo no alcanzaban tales límites. Sin embargo, el tamaño máximo de cada archivo en el sistema de archivos FAT32 es unos 4 gigabytes, con lo cual seguramente pronto quedará en desuso.

### **11.4 Sistema de particiones**

Si queremos configurar el sistema operativo desde cero o incorporar un nuevo disco a un sistema existente, en primer lugar, siempre hay que formatear el disco duro. Solo entonces el medio de almacenamiento podrá recopilar, registrar y procesar los datos.

Como vimos, según el sistema operativo el sistema de archivos será NTFS, ext4 o APFS, entre otros.

Sin embargo, otro elemento central del proceso de formateo es la creación o posibilidad de creación de particiones, es decir, la división del espacio disponible en varios sectores independientes, llamados particiones.

Dividir los soportes de datos en varias particiones solo es posible con la ayuda de un esquema que también se conoce como tabla de particiones, o sistema de particiones. En cuanto a opciones, solo existen dos formatos de particiones utilizados en la actualidad: MBR o GPT.

Estos dos modelos son los únicos disponibles para llevar a cabo la partición en todos los sistemas operativos, ya sean Linux, Windows o Mac.

#### **¿Para qué particionar el disco?**

Particionar el disco siempre ha sido lo habitual, aunque no es absolutamente necesario. Al particionar un disco separaremos de manera lógica su espacio en dos o más partes. En el sistema operativo esto se reflejará como si tuviésemos más de un disco.

El hecho de que la partición esté tan extendida es por la serie de ventajas que aporta:

- El sistema operativo y los servicios relevantes del sistema pueden almacenarse en los sectores más rápidos del disco (que suelen ser los primeros en todos los dispositivos) para



beneficiarse de la máxima velocidad de lectura y escritura posible.

- Los datos que apenas se modifican pueden guardarse en su propia partición.
- En general, el sistema y los programas pueden separarse fácilmente mediante las particiones, lo que permite realizar copias de seguridad del sistema o de los datos individuales con mucha más facilidad.

Esto último es muy útil, pues si queremos hacer una copia de seguridad rápida del sistema, podemos crear una imagen de la partición con el sistema operativo y guardarla en un lugar seguro.

### **El sistema de partición MBR**

En 1983, con el lanzamiento del IBM XT y PC DOS 2.0, también se introdujo el llamado registro de arranque principal o MBR (Master Boot Record). Desde entonces, este componente del sistema sirve, por un lado, como administrador de arranque para iniciar sistemas informáticos basados en BIOS (incluida la instalación de estos sistemas) y, por el otro, como una tabla de particiones para asignar el espacio disponible de manera eficaz.

Respecto a la segunda función, el MBR se convirtió rápidamente en el estándar para varios tipos de medios de almacenamiento, como discos duros, memorias USB o tarjetas de memoria. No obstante, en los últimos años, el MBR se ha ido sustituyendo cada vez más por su sucesor oficial, la tabla de particiones GUID o GPT (GUID Partition Table). Este nuevo estándar de las tablas de particiones forma parte de la especificación UEFI, que ha ido superando al BIOS desde el año 2000.

### **El sistema GPT**

El GPT, Tabla de Partición GUID, es el nuevo estándar que viene a reemplazar al MBR, introducido por primera vez como parte de la iniciativa de la UEFI. Comparado con el esquema de partición MBR, es más flexible y tiene mejor compatibilidad con el hardware moderno.

Algunas del GPT en comparación al MBR:

La información sobre las particiones lógicas de la unidad de almacenamiento en el MBR se almacena en un único sitio. Esto lo deja muy expuesto a que, si por algún motivo se corrompe la tabla de particiones (o un virus la infecta), la recuperación de los datos de la unidad es prácticamente imposible. Por el contrario, el sistema GPT distribuye copias de estos datos por todas las particiones. De esa



manera, si una de las particiones se corrompe, puede usar la información de otra para restablecerse.

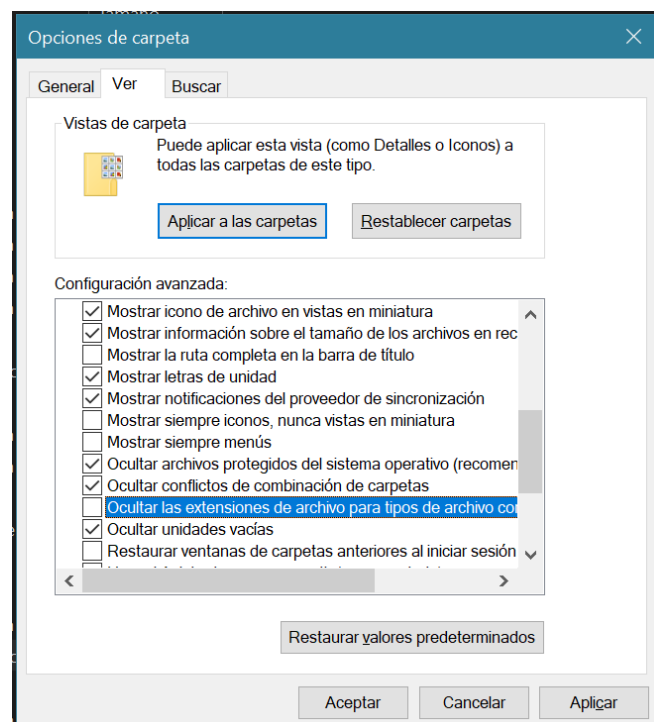
También está el tema de la capacidad de la unidad de almacenamiento. El MBR solo permite hasta 32 bits por cada sector de la unidad de almacenamiento, lo cual acaba limitando el tamaño máximo de estas unidades a los 2 TB. Sin embargo, el GPT permite un máximo de 64 bit por sector, lo que se traduce en que es capaz de ser empleado con unidades de 9,4 ZB. Cada Zettabyte equivale a mil billones de Gigabytes, por lo que en la actualidad el sistema GPT no impone ningún límite real a las unidades de almacenamiento.

Por último, los nuevos sistemas operativos, tales como Windows 11, irán sucesivamente requiriendo para su instalación que el sistema de particiones sea GPT. Afortunadamente se puede convertir de manera sencilla una unidad con sistema MBR a GPT, y si se hace bien no implica riesgo alguno, aunque siempre es mejor realizar una copia de seguridad de los datos más importantes, o bien almacenarlos en la nube.

## 11.5 Tipos de archivo

En este tema solamente mencionaré que los archivos son todos, básicamente, un conjunto de bytes. Sin embargo, según como los interprete la aplicación que los utiliza tendrán una utilidad o significado determinado.

En Windows el tipo de archivo viene determinado por la extensión del mismo, por ejemplo, un archivo .jpeg es de imagen, un archivo .avi es de video o un archivo .exe es un ejecutable. La extensión de los archivos no es visible por defecto en Windows, hay que desactivar la opción de ocultarlas en las opciones de carpeta del explorador de Windows.



Podemos dividir los archivos en dos grandes grupos:

- Los ejecutables.
- Los no ejecutables o archivos de datos.

La diferencia fundamental entre ellos es que los primeros están programados para funcionar por sí mismos y los segundos almacenan información que tendrá que ser utilizada con ayuda de algún programa.

De todos modos, la mayoría, por no decir todos, los programas requieren otros archivos que resultan necesarios aparte del ejecutable.

A continuación, algunas extensiones de archivos comúnmente utilizadas<sup>1</sup>:

Extensión o formato de archivo	Tipo de archivo	Aplicación o Programa que lo ejecuta	Icono habitual
.doc, .docx, .odt	Documento de texto	MSWord, OOWriter (OpenOffice)...	
.xls, .xlsx, .ods	Hoja de cálculo	MSExcel, OOCalc (OpenOffice)...	
.pps, .ppt, .odp	Presentación diapositivas	MSPowerPoint, OOImpress...	
.pdf	Documento publicado	Adobe Reader...	
.jpg, .bmp, .png, .gif	Archivo de imagen	Visor de fotos, Paint...	
.mp3, .wav, .wma	Archivo de audio	Windows Media Player, Audacity...	
.avi, .mp4, .flv, .mov	Archivo de video	Windows Media Player, VLC, Movie Maker...	
.htm, .html	Páginas WEB	Editores WEB...	
.zip, .rar	Archivo comprimido	WinZip, WinRar...	
.exe	Ejecutable	S.O. Windows...	Propio

<sup>1</sup> Créditos al blog: [CocinandoClases: #Un vistazo rápido a los distintos tipos de archivo](#)

## 11.6 Compresión de archivos

Los archivos comprimidos son archivos que contienen uno o varios archivos cuyo tamaño se ha reducido. Puesto que los archivos son de menor tamaño, se pueden almacenar sin ocupar demasiado espacio o transferir por Internet a velocidades más altas.

Generalmente con el mismo programa compresor de archivos se puede comprimir, pero también descomprimir el archivo y devolverlo a su estado original.

Si bien hay varios tipos de técnicas o formatos de compresión de archivos, como ZIP, ARC, ARJ, RAR, CAB, JPEG, MKV, entre tantos otros; veremos en general los fundamentos de la compresión de archivos.

En primer lugar, hay dos tipos principales de compresión de archivo:

- Sin pérdida.
- Con pérdida.

La compresión sin pérdida reduce el tamaño de los archivos sin perder ninguna información. Se utiliza para comprimir la mayoría de los archivos en general.

La compresión con pérdida reduce el tamaño de archivo cortando y eliminando trozos que no son totalmente necesarios para el funcionamiento. Se utiliza para comprimir casos específicos de tipos de archivo, como ser los archivos de imagen o video, pues se puede perder calidad en los mismos pero la imagen o el video sigue siendo funcional, por ejemplo, una imagen compartida por WhatsApp o un video en youtube en donde haya bajado la calidad con respecto al original (lo cual sucede en todos los casos).

### Compresión de archivos sin pérdida

Para que funcione la compresión sin pérdida, debe reducirse el tamaño del archivo sin perder nada.

Esto se consigue mediante el concepto de redundancia. En particular, los compresores de archivos eliminan la redundancia.

La redundancia de datos es una situación en la cual se almacenan los mismos fragmentos de datos en varios lugares, o bien las mismas cadenas de bits dentro de un mismo archivo.

Al eliminar la redundancia, se queda con tan solo una instancia de cada fragmento de datos repetidos, mediante una codificación que nos

permita transmitir la misma información de una manera mucho más compacta, usando menos bits.

Para dar un ejemplo práctico, la compresión de archivos sin pérdida sería como comprimir:

AAABBBBBBCCA AAAA            en            A3B5C2A5

La segunda cadena de caracteres contiene la misma información, pero se ha comprimido para que ocupe menos. De este modo, al descomprimir (abrir/extraer) el archivo, se puede volver a su estado original.

Se utiliza en aquellos casos en los cuales resulta inaceptable perder información.

Por supuesto, la manera en la que los compresores de archivos detectan y codifican las redundancias es mucho más compleja. Pasando desde los ya clásicos sistemas de compresión ZIP o RAR, hasta sistemas sumamente complejos como el FreeArc, capaz de comprimir, por ejemplo, un conjunto de archivos de 120Gb en 60Gb. Por supuesto, todo tiene su contra, y en este caso es el alto uso de CPU y de memoria RAM requerido para la descompresión, llegando a un tiempo de más de dos horas en procesadores actuales.

### **Compresión de archivos con pérdida**

La compresión con pérdida funciona esencialmente de la misma manera, pero tal y como se puede imaginar por su nombre, hace que se pierdan permanentemente algunos datos del archivo original.

Este tipo de compresión se suele utilizar muy a menudo con archivos multimedia tales como vídeos, audio e imágenes, puesto que no suelen sufrir demasiadas consecuencias de la pérdida de datos.

De hecho, es casi seguro que la música y los vídeos que escuchan diariamente se hayan comprimido, y en muchos casos no se nota que falte algún fragmento.

Esto se debe a que los datos que se eliminan son difíciles de percibir.

Los exponentes más claros de este tipo de compresión son los archivos jpeg para imágenes, el mp3 para audio y el H.265<sup>2</sup> para videos.

---

<sup>2</sup> [Qué es y para qué sirve el formato de compresión HEVC o H.265 \(hardzone.es\)](http://hardzone.es)