TEMA 2: PERIFÉRICOS: CONECTIVIDAD Y ADMINISTRACIÓN. ELEMENTOS DE IMPRESIÓN. ELEMENTOS DE ALMACENAMIENTO. ELEMENTOS DE VISUALIZACIÓN Y DIGITALIZACIÓN.

ÍNDICE:

1. Periféricos
   - 1.1. Introducción
   - 1.2. Conexión de periféricos al computador
   - 1.3. El teclado
   - 1.4. El ratón
2. Elementos de digitalización
   - 2.1. El escáner
   - 2.2. Reconocimiento de caracteres
   - 2.3. Almacenamiento de ficheros de imágenes
   - 2.4. El lápiz óptico
   - 2.5. Tabletas digitalizadoras
3. Elementos de impresión: impresoras
   - 3.1. Introducción
   - 3.2. Clasificación de las impresoras
   - 3.3. Métodos de impresión
   - 3.4. Memoria de las impresoras
   - 3.5. Conexión de impresora
   - 3.6. Lenguajes de descripción de página y formatos de impresión
   - 3.7. Plotter
4. Elementos de visualización: monitor
   - 4.1. Introducción
   - 4.2. Tamaño de la pantalla y proporción, medición, resolución máxima y colores
   - 4.3. Clasificación de monitores
   - 4.4. Pantalla táctil
   - 4.5. Tarjetas gráficas
5. Elementos de almacenamiento
   - 5.1. Disco Duro
   - 5.2. Memoria Flash y unidades de estado sólido (SSD)
     - 5.2.1. Memoria flash
     - 5.2.2. Unidad de estado sólido (SSD)
   - 5.3. CD, DVD, Blu-ray
   - 5.4. Cinta magnética
   - 5.5. Sistemas RAID
     - 5.5.1. Introducción
     - 5.5.2. Conceptos básicos
     - 5.5.3. Niveles RAID
     - 5.5.4. Implementación en Linux
     - 5.5.5. JBOD (Just a Bunch of Disks)
   - 5.6. Arquitecturas de almacenamiento
     - 5.6.1. Introducción
     - 5.6.2. Discos internos
     - 5.6.3. Direct attached storage (DAS)
     - 5.6.4. Network attached storage (NAS)

5.6.5. Storage area network (SAN)
5.7. Arquitecturas de respaldo (backup)
5.7.1. Introducción
5.7.2. Políticas de backup
5.7.3. Tendencias

//LOS PUNTOS 5.5 AL 5.7 INCLUYEN TAMBIÉN EL CONTENIDO DE ESTA PARTE DEL BLOQUE IV-TEMA 2 (TEMARIO OFICIAL: "…Sistemas de almacenamiento y su virtualización. Políticas, sistemas y procedimientos de backup y su recuperación. Backup de sistemas físicos y virtuales…", POR TANTO, CUANDO VEAMOS ESE TEMA (BLOQUE IV-TEMA 2) ESTOS PUNTOS LOS DAMOS POR VISTOS//

***A DESTACAR:**

1. **Tipos de periféricos: entrada, salida, almacenamiento y de E/S. Concepto y tipos de teclado y ratón.**
2. **Escáneres: de mano, planos, de rodillo, cenitales, especiales. Reconocimiento óptico de caracteres (OCR), inteligente-imágenes (IMR), marcas (OMR), código de barras (OBR).**
3. **Impresoras: de impacto (matriciales y margarita), inyección de tinta, láser, sublimación de tinta, tinta sólida, cera térmica.**
4. **Monitores: características y tipos, sobretodo OLED, LED, AMOLED E IPS.**
5. **Cintas y DVD: características, tipos y velocidades. LTO fundamentalmente.**
6. **Discos duros: características, niveles RAID!!!, discos SSD!!!. Conceptos de DAS, NAS Y SAN y arquitecturas de backup: tipos (completa, diferencial e incremental),rotación abuelo-padre-hijo, tecnologías (DEDUPLICACIÓN Y AUTO-TIERING!!!).**

NOTA:
Recomiendo estudiar este tema por el esquema del tema, lo que está ahí es lo más importante y lo que deberías llevar bien preparado. En cuanto al tema (este documento) si estaría bien darle una lectura, sin llegar a memorizar los datos que ponen, excepto los que están en el esquema.
De este documento sí que recomiendo estudiar por este documento lo referente al disco duro (partes) y conectores usados, los discos SSD y lo de RAID y sus niveles imprescindible!!! Y las arquitecturas de almacenamiento (DAS, NAS, SAN), tipos de copias de seguridad (incrementales y diferenciales) también es importante. Es decir, puntos 5.1, 5.2, 5.5, 5.6 y 5.7

1. Periféricos

1.1. Introducción

Se denominan periféricos tanto a las unidades o dispositivos a través de los cuales el ordenador se comunica con el mundo exterior, como a los sistemas que almacenan o archivan la información, sirviendo de memoria auxiliar de la memoria principal.
Se entenderá por periférico a todo conjunto de dispositivos que, sin pertenecer al núcleo fundamental de la CPU-Memoria Central, permitan realizar operaciones de E/S, complementarias al proceso de datos que realiza la CPU.

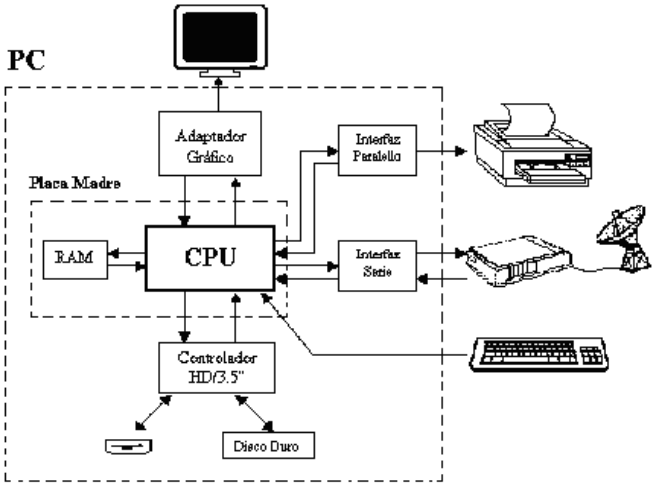


Figura 1. El PC y sus periféricos

El procesador solo se comunica con la memoria principal, y las características de esta son:

- Su reducida capacidad para almacenar datos.
- Su volatilidad, pues al desconectar el ordenador de la red eléctrica se borra su contenido.
- Aunque la memoria puede considerarse también como una unidad de entrada / salida, hay una característica que la diferencia del resto de unidades: la información almacenada no es directamente entendible por el usuario, pues suele estar codificada, comprimida e incluso a veces encriptada por motivos de seguridad. Esto hace que la interacción con el usuario no sea directa como en el resto de unidades.
- Los dispositivos de E/S transforman la información externa como una pulsación en una tecla en impulsos eléctricos que son codificados enviados hacia el procesador para su interpretación, procesamiento y almacenamiento de forma automática. Estas señales eléctricas se codifican según unos códigos como el ASCII y la CPU recibe la información codificada siempre en binario.
- Otros dispositivos como las impresoras realizan el paso contrario, unos valores binarios que codifican un determinado carácter son convertidos en una serie de marcas gráficas sobre un papel a fin de que el usuario lea o vea la información requerida.

Partes de los periféricos.

Todos los periféricos suelen tener dos partes claramente diferenciadas, a saber: una parte mecánica y otra parte electrónica.

- La parte **mecánica** está formada por dispositivos electromecánicos (conmutadores manuales, reles, motores, electroimanes, etc., que son controlados por elementos eléctricos.

- La parte **electrónica** se encarga de controlar las ordenes que llegan de la CPU para la recepción o transmisión de datos, y de generar las señales de control para manejar adecuadamente la parte mecánica del periférico. En la parte electrónica es común usar elementos optoelectrónicos que actúan como detectores o generadores de la información de entrada y salida, respectivamente. También estos elementos se usan como detectores de posición de los elementos mecánicos móviles del periférico. Juegan un papel importante los conversores analógicos/digitales.

Los dispositivos de E/S transforman la información externa en señales codificadas, permitiendo su transmisión, detección, interpretación, procesamiento y almacenamiento de forma automática.
Los periféricos se pueden clasificar en tres tipos: en unidades de entrada, unidades de salida y unidades de memoria masiva o almacenamiento.
En el cuadro siguiente se puede ver una clasificación completa de los periféricos más usuales.

Unidades de entrada.

- Teclado. - Ratón ("*mouse*"). - Dispositivos de captura directa de datos: o Lectora de banda magnética. o Detectores ópticos: de marcas, barras impresas, escáneres, caracteres impresos, cámaras digitales, ... o Detectores de caracteres impresos (OCR). - Unidades de reconocimiento de voz. - Lápiz óptico. - Palanca manual de control ("*joy-stick*"). - Digitalizador o tableta gráfica. - Webcam.	

Unidades de salida.

forjaTIC

- Monitores de visualización: tubo de e- y TFT. - Impresoras. - Sintetizador de voz. - Visualizadores (“*displays*”). - Registrador gráfico (“*plotter*”). - Proyector. - GPS.	

*NOTA: WEBCAM NO ES DE SALIDA!

Unidades de memoria masiva auxiliar o almacenamiento.

- Discos magnéticos. - Cintas magnéticas. - Discos ópticos y magnetoópticos. - Memoria de silicio (PenDrive).	

Unidades mixtas de E/S.

- Terminal interactivo teclado-pantalla. - Router. - Modem. - Impresora Multifunción. - Cascos de realidad virtual. - Pantalla táctil.	

1.2. Conexión de periféricos al computador

Los periféricos se conectan a la CPU a través de grupos de hilos que se conoce como **buses**. En el interior del computador el bus transmite la información de los datos en paralelo.
El bus que conecta la CPU con los otros elementos del procesador se conoce como **bus local** o bus de la CPU. Es un bus muy rápido y conecta la CPU con las tarjetas de la placa base y los controladores de los dispositivos externos.

Las conexiones entre los periféricos y los controladores o tarjetas de la placa base se realizan a través de un bus más general llamado **bus del sistema.** También suele conectar algunas ampliaciones de memoria.
Algunos periféricos requieren un bus especializado que se adapte a su velocidad de transferencia, sus niveles de tensión, la naturaleza de sus señales de control y otros requerimientos. A estos buses se les llama **bus de entrada/salida** o bus de expansión.
Por todo lo vistos, los computadores grandes al disponer de varios tipos de buses requieren de dispositivos adaptadores o de interconexión entre buses.
Los procesadores suelen tener unas ranuras de expansión (6 normalmente) sobre la placa base que están conectadas al bus del sistema y que permiten conectar una serie de dispositivos a este bus a través de tarjetas de circuito integrado y que permiten conectar varios dispositivos a la CPU, como por ejemplo tarjetas digitalizadoras de imágenes, aceleradores gráficos...
Todos los buses poseen unas especificaciones normalizadas, como son:
- protocolos de transmisión de datos,
- velocidades y temporización de las transferencias,
- anchuras de los sub-buses,
- y sistema físico de conexión (conectores estandarizados).

Los Buses normalizados más conocidos (muchos ya vistos en el Tema 1 anterior) son:

- **S-100 Bus (IEEE 696).** Puede considerarse como el primer bus normalizado para microcomputadores, siendo introducido por Atari para su computador 8080 (sistema de 8 bits). En total disponía de unos 100 hilos.
- **CAMAC (o IEEE 583).** El bus CAMAC ("*Computer Automated Measurement and Control*") Se introdujo para interconectar instrumentos de medida nucleares en 1969.
- **GPIB (o IEEE 488).** El GPIB (*"General Purpose Interface Bus")* fue ideado por Hewlett Packard (1965 a 1975) usa 24 hilos, 8 de los cuales son para datos y el resto para señales de control.
- **Multibus (o IEEE 796).** Bus de 16 bit de datos introducido por Intel. En la actualidad hay una versión mejorada llamada Multibus-II (IEEE 1296) para transferir datos de 32 bits.
- **ISA Bus.** El bus ISA *("Industrial Standard Architecture"*), es el bus introducido con el IBM-PC. Tiene 64 hilos de los cuales 8 son para datos.
- **ISA AT Bus.** Fue introducido con los IBM-AT (80286). Ideado para arquitecturas de 16 bits, posee subdirecciones de 24 bits (direcciona hasta 16 Mbytes) y es compatible, como no, con su antecesor de 16 bits.
- **MCA.** *("Micro-Channel Architecture"*) fue introducido por IBM en 1987 en sus equipos PS/2. Es un bus para arquitecturas de 32 bits y es 10 veces más rápido que el ISA AT, llegando a transferir hasta 20 Mbits/seg.
- **EISA** ("*Extendet Industry Satandard Architecture"*). Es un bus ideado por 9 fabricantes de ordenadores, para arquitecturas de 32 bits. Posee velocidad de transferencia de 33Mbits/seg. Es compatible con el bus ISA. Este bus puede solo ser controlado por microprocesadores 80386, 80486 o superiores, y es autoconfigurable.
- **SCSI** *("Small computer System Interface")* es un estándar universal para conexiones paralelas a periféricos. Suele utilizarse para unidades de discos magnéticos y ópticos. Admite hasta 7 dispositivos y fue ideado para entornos UNIX y Macintosh. Permite velocidades de transferencia de 5 Mbits/seg hasta 400 Mbits/seg. En la actualidad se está desarrollando el SCII-3, de 32 bits, que podrá admitir hasta 32 periféricos conectados a gran distancia por fibra óptica (comunicación serie).
- **Futurebus+** (IEEE 896.1 e IEEE 896.2). Es una normalización proyectada para equipos de muy altas prestaciones, que puede considerarse como una evolución de las normas Multibus II y VME. Diseñado para arquitectura de 64 bits. Permite la construcción de sistemas multiprocesador (de hasta 32 procesadores) compartiendo memoria.

- **USB** o Bus Serie Universal, es un estándar de 1995 que define un bus para conectar periféricos al ordenador. Puede llegar a conectar hasta 127 dispositivos con una conexión de tipo estrella. El estándar incluye la transmisión de energía eléctrica al dispositivo conectado.

Hay otros tipos de bus, pensados para estar conectados directamente a la CPU, son los **buses locales.**
Los buses locales más conocidos son:

- **VL-bus**, ideado por VESA para varios fabricantes.
- **PCI** ("Peripheral Component Interconected"), creado para Intel.
- **Quick-ring**, propuesto por Apple.
- **VME**, IEEE 1014 ("*Versa Module Europe*") aparece en 1981 como sucesor delVersabus, utilizado por Motorola para la serie 68000. Es el más sencillo y el de menos prestaciones de todos los buses de sistemas, sin embargo su aceptación se debe a que es posible encontrar más de 200 módulos comerciales para este tipo de bus. Sus aplicaciones se centran en los campos de sistemas industriales en tiempo real, aplicaciones militares e investigación. Es un bus multiproceso, lo que significa que varios procesadores pueden compartirlo y puede llegar a trnasmitir hasta 40 Mbytes/seg con datos de 32 bits.
- **NuBus**, o IEEE 1196, de 96 hilos está diseñado para arquitecturas de 32 bits y usa en los Apple Macintosh.
- **M-bus** y **S-Bus**, utilizados por las estaciones de SUN.

1.3. El teclado

Los teclados son periféricos similares a las máquinas de escribir (Cristopher L. Sholes en 1868 diseñó la primera máquina de escribir comercial y en 1873 el teclado "QWERTY" que se usa en la actualidad), con sus teclas dispuestas en una distribución que se conoce como QWERTY, y que se basa en la idea de que las teclas más utilizadas (en inglés) estén lo más separadas posibles entre si y además la mano izquierda se use en un 60% de las ocasiones (40% la mano derecha, más hábil.

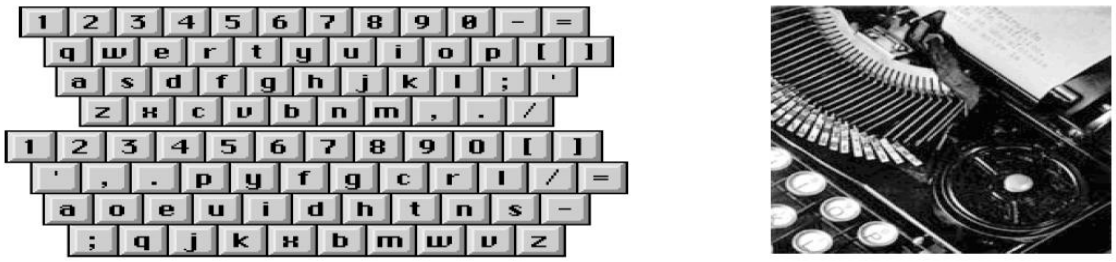


Figura 2. Teclados QWERTY Y Dvorak. Y al lado, detalle de máquina de escribir tradicional.

Al pulsar una tecla se cierra un conmutador que hay en el interior del teclado, y mediante el chip 8742 se detecta que tecla se ha pulsado para que posteriormente una lógica circuitos codificadores llamados **controladores de teclado (8042)** generen el código correspondiente al carácter seleccionado (ASCII, por ejemplo), y envien una interrupción a la CPU para avisar de que tecla o teclas han sido pulsadas. Normalmente, el programa gestor de teclado hace un "**eco**" del carácter pulsado visualizandolo en el monitor, pero eso no quiere decir que ese carácter haya sido enviado a la CPU.
Su funcionamiento detallado es como sigue: los teclados poseen internamente una matriz de teclas y un circuito integrado (el 8742) que se encarga de supervisar esta matriz y detectar las pulsaciones. Cuando se detecta una pulsación, o la liberación de una tecla, se escribe su código

en un *buffer* interno del teclado y a continuación el teclado transmite este código al controlador que se encuentra en la placa base del procesador vía serie a través del cable del teclado. Al pulsar se genera una interrupción o llamada que detecta el 8742 y genera el código ASCII de la tecla pulsada (con el bit-7 a "0") y al soltar se genera el mismo código pero con el bit-7 activado a "1". Por ejemplo, si se pulsa la 'A' se generará una INT 9 y aparecerá en el puerto del teclado (60h) el byte "1Eh", al soltar la 'A' se generará otra INT 9 y se podrá leer el byte "9Eh" del puerto del teclado. Este controlador 8742 incluido en el teclado posee las siguientes funciones:
a) Prevenir falsas repeticiones (rebotes) o pulsaciones incorrectas.
b) Traducir la tecla pulsada en un código único, llamado "*scan-code*". Suele ocupar un byte.
c) Repetir un carácter si se ha pulsado durante cierto tiempo (del orden de un segundo como mínimo).
d) Detectar las pulsaciones simultáneas de varias teclas.

El conector de teclado posee 5 hilos para transmitir las señales:

señal de reloj, datos, reset , tierra y alimentación

Ello permite establecer una comunicación bidireccional entre la CPU y el teclado. La línea de reset sirve para inicializar el teclado.

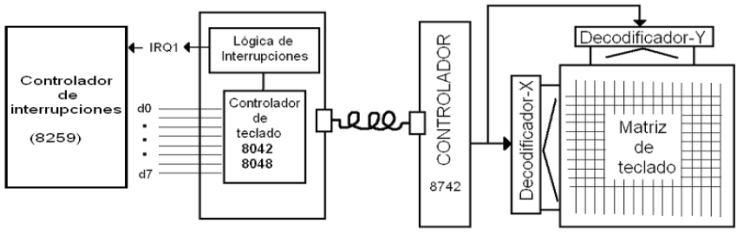


Figura 3. Esquema gráfico de conexión de un teclado.

Cuando se produce una pulsación, el controlador transfiere un código al ordenador ("make-code") que genera una interrupción para que la rutina de tratamiento lea el código enviado. Es el driver de teclado el que en estas interrupciones se encarga de combinar los códigos para generar los caracteres (mayúsculas, minúsculas, teclas de función, control,...). Las teclas expandidas -las que han sido añadidas al teclado estándar de 83/84 teclas- o las combinaciones de teclas como ALT-teclado_numérico, tienen un comportamiento especial, ya que pueden generar hasta 4 interrupciones consecutivas (con un intervalo de unos 1,5 milisegundos, ó 3 ms en los códigos dobles que convierte en uno el 8042). En la figura 1.4 se observa un diagrama similar de cómo funciona el teclado de un ordenador.

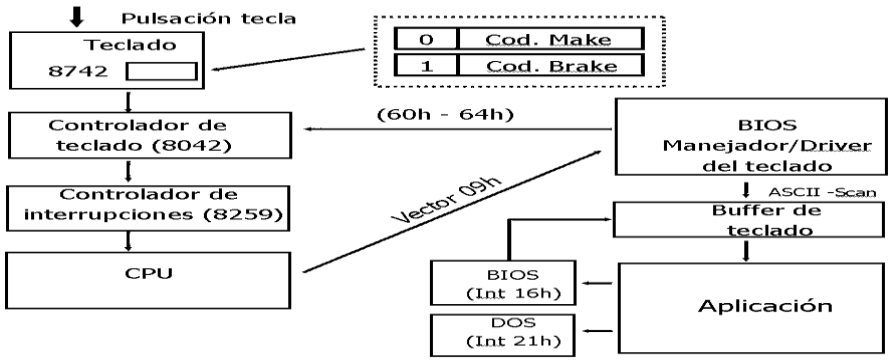


Figura 4. Diagrama del esquema de conexión de un teclado.

forjaTIC

Según las normas ANSI (*American National Standars Institute*) los teclados deben contener los siguientes tipos de letras:

- **Teclado principal**: Caracteres alfabéticos, numéricos y especiales, como una máquina de escribir más o menos.
- **Teclas de gestión de imagen**: sobre la pantalla aparece un "*cursor*" que indica donde va a aparecer el siguiente carácter a teclear, pues bien, deben haber unas teclas que nos permitan mover y situarar este cursor en la posición que nosotros elijamos. Además estas teclas de gestión de imagen deben permitirnos borrar un carácter o una línea completa. Incluso deben permitirnos modificar el contenido de la memoria intermedia antes de que su contenido sea enviado a la CPU.
- **Teclado numérico**: son teclas que contienen los caracteres numéricos y operaciones básicas sobre ellos (*+,-,*,/*), además del punto decimal. En el teclado de la figura 1.3 se puede observar estos caracteres a la derecha del teclado.
- **Teclas de funciones**: son teclas cuya función puede ser predefinida por el usuario mediante los programas que este desarrolle o lo están por los programas que estamos utilizando en cada momento. Son de F1 a F12. Hoy en día exinten además multiples teclas programables por los programas o el usuario.
- **Teclas de funciones locales**: controlan funciones propias del terminal, como la impresión del contenido de la pantalla, teclas de escape, avance de página, etc..

Tipos de teclado

- Según su tecnología:
    - **Teclados mecánicos** constan de una serie de teclas con unos interruptores mecánicos colocadas encima de unos muelles, que son los que hacen retornar las teclas a la posición original, de modo que al ser pulsadas éstas hacen contacto con unas terminaciones metálicas del circuito impreso del propio teclado, cerrando así el circuito, y volviendo a abrirlo al dejar de pulsar por el efecto de retorno del muelle. El contacto establecido entre los terminales metálicos de las teclas y el del circuito impreso determina la señal diferenciada.
    - **Teclados de membrana** se componen de cuatro capas: la inferior tiene una serie de pistas conductores impresas; encima de ella, se coloca una capa de separación con agujeros justo debajo de cada una de las teclas; encima de esta se coloca una capa conductora con pequeñas montañitas debajo de cada una de las teclas y en cada montañita un conector metálico; encima de éstas se coloca una capa de goma para producir el efecto de retorno a la posición inicial. Cuando pulsamos una tecla, lo que hacemos es poner en contacto las dos capas conductoras (la primera con el circuito y la tercera con los conectores) haciendo que el circuito se cierre, y la membrana de goma hace que se separen las capas al impulsar la tecla hacia su posición inicial (similar al mando a distancia del TV).
    - **Otras tecnologías**: de cúpula de goma o de contacto metálico.

- Según su forma física:
    - **Teclados ergonómicos** Se basan en el principio que dividiendo el teclado principal colocando en ángulo cada una de las mitades, los codos descansan en una posición mucho más natural, y cambiando la curvatura del teclado y añadiendo un pequeño "reposamuñecas", el ángulo de escritura es mucho más cómodo para el usuario.



Figura 5. Teclado ergonómico

- **Teclados programables** El propio teclado lleva un microcontrolador interno (que se comunica con la controladora de teclado de la placa base) con un programa integrado que interpreta las señales producidas al cerrarse el circuito cuando dos terminales (tecla y circuito integrado) entran en contacto. Este programa reside en una ROM, la cual puede almacenar muchos otros datos, además del código del teclado (país) y la posición de las teclas, pero para interpretarlos, se ha de instalar un driver o controlador del dispositivo que interprete las señales**.** Un Driver (conductor) es suministrado por el fabricante y tiene como función asegurar la compatibilidad, el buen funcionamiento y el uso de las características que dicho dispositivo ofrece. Un driver normalmente está desarrollado para un sistema operativo específico y no funcionará bajo otros entornos.
  La mayoría de los teclados que se venden actualmente tienen teclas específicas para WINDOWS 9x y NT, que son interpretadas por el propio sistema operativo sin un driver adicional, pero existen teclados desde los cuales podemos manejar parámetros concernientes al sonido, la reproducción de CDs musicales, etc. Recientemente han aparecido en el mercado teclados con teclas adicionales programables sin una función específica, a las que nosotros podremos asignar la ejecución de nuestras aplicaciones favoritas, el guardado de documentos, impresión, etc.
- **Teclado multimedia** Añade teclas especiales que llaman a algunos programas en el computador, a modo de acceso directo, como pueden ser el programa de correo electrónico, la calculadora, el reproductor multimedia, etc.
- **Teclado flexible** Estos teclados son de plástico suave o silicona que se puede doblar sobre sí mismo. Durante su uso, estos teclados pueden adaptarse a superficies irregulares, y son más resistentes a los líquidos que los teclados estándar. Estos también pueden ser conectados a dispositivos portátiles y teléfonos inteligentes. Algunos modelos pueden ser completamente sumergidos en agua, por lo que hospitales y laboratorios los usan, ya que pueden ser desinfectados.



Figura 6. Teclado flexible de Apple

- **Teclado en pantalla** Hoy en día existen también los teclados en pantalla, también llamados teclados virtuales, que son (como su mismo nombre indica) teclados representados en la pantalla, que se utilizan con el ratón o con un

dispositivo especial (podría ser un joystick). Estos teclados se utilizan mayoritariamente en los dispositivos móviles como teléfonos, tabletas y PDA. En el mercado hay una gran variedad de teclados.

- **Teclados de proyección** Existen teclados de proyección, de igual tamaño que un teclado estándar pero que utilizan láser. Se pueden conectar por USB, Bluetooth o Wi-Fi.

- Según su conexión:
  - **Teclado con conector Paralelo** Obsoleto.
  - **Teclado con conector DIN o PS/2** El conector DIN fue usado por pocas computadoras, y fue remplazada rápidamente, más que nada por su tamaño. El PS/2, que también se le denomina «Mini DIN», es parecido al conector DIN, pero con un tamaño más reducido, y se sigue utilizando hoy en día pero cada vez en menor medida (siendo remplazado por el USB en gran parte). Se caracterizan por ser redondos y contener varios pines en su interior.



Figura 7. Conectores PS/2 para ratón (verde) y teclado (morado)

  - **Teclado con conector USB** Aunque los teclados USB comienzan a verse al poco de definirse el estándar USB, es con la aparición de la Apple iMac, que trae tanto teclado como mouse USB de serie cuando se estandariza el soporte de este tipo de teclado. Además tiene la ventaja de hacerlo independiente del hardware al que se conecta. Si queremos conectar a nuestro equipo un *teclado USB,* primero debemos tener una BIOS que lo soporte y en segundo lugar debemos tener instalado el sistema operativo con el "Suplemento USB". Un buen teclado USB debe tener en su parte posterior al menos un conector USB adicional para poderlo aprovechar como HUB y poder conectar a él otros dispositivos USB como ratones, altavoces, etc.



Figura 8. Convertidor PS/2 a USB y HUB USB

  - **Teclado inalámbrico** Suelen ser teclados comunes donde la comunicación entre el computador y el periférico se realiza a través de rayos infrarrojos, ondas de radio o mediante Bluetooth. Los primeros modelos empleaban infrarrojos, pero ya lo habitual es la radiofrecuencia, cuya gran virtud es que no necesitamos tener una línea visual entre el teclado y el sensor, lo que nos da una extraordinaria movilidad. También existen ratones inalámbricos, que a menudo podemos comprar en un kit junto al teclado. El gran problema de estos dispositivos ha sido siempre el gasto de pilas. No obstante, los nuevos modelos están reduciendo enormemente el consumo de electricidad (algunos ya ofrecen varios meses de

autonomía) y, en cualquier caso, siempre podemos recurrir a las pilas recargables. Estos teclados en vez de enviar la señal mediante cable, lo hacen mediante infrarrojos, y la controladora no reside en el propio teclado, sino en el receptor que se conecta al conector de teclado en el PC.



Figura 9. Teclado y ratón inalámbricos

Distribuciones de teclado más usuales:

TIPO TECLADO	PAISES/USO
QUERTY	Latinos, anglosajones, escandinavos y mas
QUERTZ	Germanohablantes
AZERTY	Francófonos
HCESAR	Portugués
JCUKEN	Ruso-cirílico
DVORAK	Teclado simplificado alternativa a QUERTY
COLEMAK	Alternativa a QUERTY y DVORAK. Inglés, para facilitar la escritura colocando las letras de mayor frecuencia bajo los dedos más fuertes

Figura 10. Distribuciones de teclado

1.4. El ratón

Fue diseñado por Douglas Engelbart y Bill English durante los años 60 en el Stanford Research Institute, en pleno Silicon Valley en California. La primera maqueta se construyó en madera, y se patentó con el nombre de "X-Y Position Indicator for a Display System". En San Francisco, a finales de 1968 se presentó públicamente el primer modelo oficial.
La función principal del ratón es transmitir los movimientos de nuestra mano sobre una superficie plana hacia el ordenador. Allí, el software denominado ***driver*** se encarga realmente de transformarlo a un movimiento del puntero por la pantalla dependiendo de varios parámetros.
En el momento de activar el ratón, se asocia su posición con la del cursor en la pantalla. Si desplazamos sobre una superficie el ratón, el cursor seguirá dichos movimientos. Es casi imprescindible en aplicaciones dirigidas por menús o entornos gráficos, como por ejemplo Windows.

Tipos o modelos:



Figura 11. Distintos tipos de ratón

- Según su mecanismo:
  - **Mecánico** Tienen una gran esfera de plástico o goma, de varias capas, en su parte inferior para mover dos ruedas que generan pulsos en respuesta al movimiento de éste sobre la superficie. Una variante es el modelo de Honeywell que utiliza dos ruedas inclinadas 90 grados entre ellas en vez de una esfera. La circuitería interna cuenta los pulsos generados por la rueda y envía la información a la computadora, que mediante software procesa e interpreta.
  - **Óptico** Es una variante que carece de la bola de goma que evita el frecuente problema de la acumulación de suciedad en el eje de transmisión, y por sus características ópticas es menos propenso a sufrir un inconveniente similar. Se considera uno de los más modernos y prácticos actualmente. Puede ofrecer un límite de 800 ppp, como cantidad de puntos distintos que puede reconocer en 2,54 centímetros (una pulgada); a menor cifra peor actuará el sensor de movimientos. Su funcionamiento se basa en un sensor óptico que fotografía la superficie sobre la que se encuentra y detectando las variaciones entre sucesivas fotografías, se determina si el ratón ha cambiado su posición. En superficies pulidas o sobre determinados materiales brillantes, el ratón

óptico causa movimiento nervioso sobre la pantalla, por eso se hace necesario el uso de una alfombrilla de ratón o superficie que, para este tipo, no debe ser brillante y mejor si carece de grabados multicolores que puedan «confundir» la información luminosa devuelta.
    - **Láser** Este tipo es más sensible y preciso, haciéndolo aconsejable especialmente para los diseñadores gráficos y los jugadores de videojuegos. También detecta el movimiento deslizándose sobre una superficie horizontal, pero el haz de luz de tecnología óptica se sustituye por un láser con resoluciones a partir de 2000 ppp, lo que se traduce en un aumento significativo de la precisión y sensibilidad.
    - **Trackball** En concepto de trackball es una idea que parte del hecho: se debe mover el puntero, no el dispositivo, por lo que se adapta para presentar una bola, de tal forma que cuando se coloque la mano encima se pueda mover mediante el dedo pulgar, sin necesidad de desplazar nada más ni toda la mano como antes. De esta manera se reduce el esfuerzo y la necesidad de espacio, además de evitarse un posible dolor de antebrazo por el movimiento de éste. A algunas personas, sin embargo, no les termina de resultar realmente cómodo. Este tipo ha sido muy útil por ejemplo en la informatización de la navegación marítima.
    - **Multitáctil** El principal ejemplo de los ratones táctiles es el Magic Mouse de Apple, y están diseñados con una carcasa superior de una pieza. Su superficie es lisa, sin botones, ya que se trata de una zona multitáctil en el que todo el ratón hace de botón y lo puedan usar tanto los diestros como los zurdos.
- Según su conexión:
    - **Cableado** (PS/2 (puerto) y USB) Es el formato más popular y más económico, sin embargo existen multitud de características añadidas que pueden elevar su precio, por ejemplo si hacen uso de tecnología láser como sensor de movimiento. Actualmente se distribuyen con dos tipos de conectores posibles, tipo USB y PS/2; antiguamente también era popular usar el puerto serie.
      Es el preferido por los videojugadores experimentados, ya que la velocidad de transmisión de datos por cable entre el ratón y la computadora es óptima en juegos que requieren de una gran precisión.
    - **Inalámbrico** En este caso el dispositivo carece de un cable que lo comunique con la computadora, en su lugar utiliza algún tipo de tecnología inalámbrica. Para ello requiere un receptor que reciba la señal inalámbrica que produce, mediante baterías, el ratón. El receptor normalmente se conecta a la computadora a través de un puerto USB o PS/2. Según la tecnología inalámbrica usada pueden distinguirse varias posibilidades (estas tecnologías se estudiarán más a fondo en el Tema 6 del Bloque IV):
        - **Radiofrecuencia** La Radio Frecuencia (RF) es el tipo más común y económico de este tipo de tecnologías. Funciona enviando una señal a una frecuencia de 2.4 GHz, popular en la telefonía móvil o celular, la misma que los estándares IEEE 802.11b y IEEE 802.11g. Es popular, entre otras cosas, por sus pocos errores de desconexión o interferencias con otros equipos inalámbricos, además de disponer de un alcance suficiente: hasta unos 10 metros.
        - **Infrarrojo** La tecnología infrarroja (IR) utiliza una señal de onda infrarroja como medio de transmisión de datos, popular también entre los controles o mandos remotos de televisiones, equipos de música o en telefonía celular. A diferencia de la anterior, tiene un alcance medio inferior a los 3 metros, y tanto el emisor como el receptor deben estar en una misma línea visual de contacto directo ininterrumpido para que la señal se reciba correctamente. Por ello su éxito ha sido menor, llegando incluso a desaparecer del mercado.
        - **Bluetooth** La tecnología Bluetooth (BT) es la más reciente como transmisión inalámbrica (estándar IEEE 802.15.1), que cuenta con cierto éxito en otros

dispositivos. Su alcance es de unos 10 metros o 30 pies (que corresponde a la Clase 2 del estándar Bluetooth).

2. Elementos de digitalización

2.1. El escáner

Los **escáneres** ópticos son unos dispositivos de entrada para el ordenador que utilizan un haz luminoso para detectar los patrones de luz y oscuridad (o los colores) de la superficie del papel, convirtiendo la imagen en señales digitales que se pueden manipular por medio de un *software* de tratamiento de imágenes o con reconocimiento óptico de caracteres. La necesidad de incorporar diagramas, fotografías y gráficos ha hecho de la autoedición la aplicación predominante de los escáneres. El scanner nace en 1984 cuando Microtek crea el MS-200, el primer scanner blanco y negro con una resolución de 200dpi. Luego en 1989 aparece el primer scanner a color.
Transforma las imágenes a formato digital, es decir en series de 0 y de 1, pudiendo entonces ser almacenadas, retocadas o impresas o ser utilizadas para ilustrar un texto. Si el documento que se desea escanear o digitalizar es un texto, por medio de programas de reconocimiento de caracteres, también llamados por las siglas inglesas **OCR** *("Optical Character Recognition"*), es posible reconstituirlo y convertirlo en texto reconocible por el ordenador, pudiendo ser corregido o añadir texto nuevo, es decir, nos evita tener que teclearlo.
Un escáner se compone de dos piezas básicas: la primera de ellas es el **cabezal de reconocimiento óptico**, la segunda es un simple mecanismo de avance por debajo de un cristal que hace las veces de soporte para los objetos que se van a escanear. En principio, el cabezal de reconocimiento óptico realiza un muestreo del objeto en sí, reconociendo un determinado número de puntos por pulgada y a cada uno de estos puntos le asigna un valor en función del número de bits del proceso: 1 bit sería 1 color (negro o blanco), 2 bits serían 4 colores, 8 bits serían 256 colores y así sucesivamente hasta llegar a los 32 bits (color verdadero).

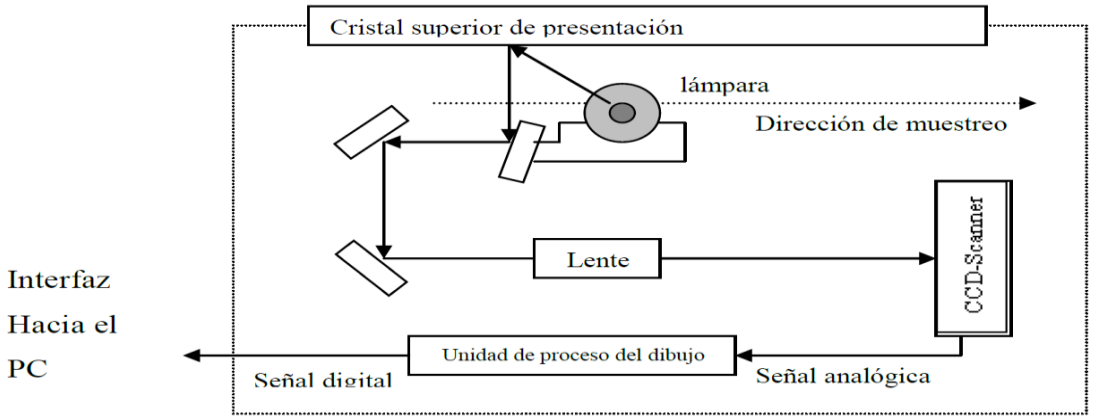


Figura 12. Diagrama de bloques del escáner

Los elementos CCD están colocados en una sola fila de forma que a cada elemento le corresponde un píxel de cada una de las filas de puntos que forman la imagen.
Los escáneres en color suelen disponer de 3 fuentes de luz (rojo, verde y azul), y utilizando simultáneamente cada una de ellas, se pueden formar las imágenes en color.

Funcionamiento del escáner
El documento se ilumina línea a línea por una fuente de luz fluorescente o incandescente. La luz reflejada incide a través de una lente sobre un fotosensor denominado CCD ("charge coupled device").

EL CCD es una tabla lineal de elementos fotoeléctricos o detectores de luz, cuyo número suele oscilar entre 2.000 y 4.000. Cada elemento proporciona un voltaje proporcional a la cantidad de luz que cae sobre él. Un punto negro del documento absorbe la mayor parte de luz, permitiendo que muy poca se refleje en el CCD, generándose una salida de bajo voltaje. Un punto blanco refleja la mayor parte de luz, dando como resultado una salida de alto voltaje. Los niveles de gris (o colores) causan igualmente voltajes proporcionales generados por los elementos CCD.
Para conseguir un escáner en color, el procedimiento descrito es repetido tres veces utilizando cada vez un filtro de color distinto, o bien son los tres chips o captores CCD los que analizan los tres haces luminosos separados previamente por un prisma y filtros rojos, verdes y azules
Estos niveles de tensión eléctrica, analógicos, deben convertirse en valores digitales. Un conversor analógico-digital (ADC) convierte el voltaje de salida de cada elemento del CCD en una trama de bits adecuada, que representa la intensidad de la luz reflejada. Con cualquier tipo de escáner, una vez convertidos los valores analógicos en digitales, la información resultante puede almacenarse localmente en una RAM para un procesado posterior (frecuentemente en aplicaciones OCR, donde el escáner puede disponer de CPU y Memoria de modo que envía al ordenador ya un fichero ASCII). La mayoría de los escáneres, sin embargo, no preprocesan la información obtenida de la digitalización de las imágenes, sino que ésta se envía directamente al ordenador. Así, una vez recibidos los datos de la imagen digitalizada, el ordenador puede procesar la información, mediante software OCR, de edición de gráficos u otras utilidades.

Resolución
Es el parámetro más importante de un escáner, es el grado de finura con el que se puede realizar el análisis de la imagen.
La resolución de un escáner es el número de puntos que puede leer para cada pulgada lineal del documento. Las resoluciones se miden en puntos por pulgada (ppp o dpi). Esta es la resolución óptica o real del escáner. Así, cuando hablamos de un escáner con resolución de "300x600 ppp" nos estamos refiriendo a que en cada línea horizontal de una pulgada de largo (2,54 cm) puede captar 300 puntos individuales, mientras que en vertical llega hasta los 600 puntos; como en este caso, generalmente la resolución horizontal y la vertical no coinciden, siendo mayor (típicamente el doble) la vertical.
Esta resolución óptica viene dada por el CCD y es la más importante, ya que implica los límites físicos de calidad que podemos conseguir con el escáner. Por ello, es un método comercial muy típico comentar sólo el mayor de los dos valores, describiendo como “un escáner de 600 ppp” a un aparato de 300x600 ppp o “un escáner de 1.200 ppp” a un aparato de 600x1.200 ppp. Cuanta mayor sea la resolución, más calidad tendrá el resultado.
Tenemos también la resolución interpolada; consiste en superar los límites que impone la resolución óptica (300x600 ppp, por ejemplo) mediante la estimación matemática de cuáles podrían ser los valores de los puntos que añadimos por software a la imagen. Por ejemplo, si el escáner capta físicamente dos puntos contiguos, uno blanco y otro negro, supondrá que de haber podido captar un punto extra entre ambos sería de algún tono de gris. De esta forma podemos llegar a resoluciones absurdamente altas, de hasta 9.600x9.600 ppp, aunque en realidad no obtenemos más información real que la que proporciona la resolución óptica máxima del aparato.
Por último está la propia resolución de escaneado, aquella que seleccionamos para captar una imagen concreta, que puede ser incluso menor que la resolución óptica. Su valor irá desde un cierto mínimo (típicamente unos 75 ppp) hasta el máximo de la resolución interpolada. En este caso el valor es siempre idéntico para la resolución horizontal y la vertical, ya que si no la imagen tendría las dimensiones deformadas.

Profundidad del color

Este parámetro, expresado en bits, indica el número de tonalidades de color que un píxel puede adoptar; lo normal en la actualidad es un valor de 24 bits por píxeles. Aunque hasta hace poco los escáneres de blanco y negro, tonos de grises o 256 colores eran muy populares, lo cierto es que los 24 bits de color se han convertido en un estándar, lógico si se tiene en cuenta que en la actualidad cualquier tarjeta gráfica es capaz de mostrar esta cantidad de colores. Sin embargo, hay escáneres capaces de utilizar 30 o incluso 36 bits de color, pero la mayoría lo hacen a nivel interno, para disminuir el intervalo entre una tonalidad y la siguiente; posteriormente, lo que envían al PC son únicamente 24 bits.
Por tanto según el número de bits:

- de 1 bits, resultaría una imagen en blanco y negro.
- de 8 bits, se obtendría una imagen de 256 tonos de grises.
- de 24 bits u 8 bits por componente de color (verde, rojo, azul), la imagen puede llegar a ser de 16'7 millones de colores.
- de 30 bits, permite sobrepasar los mil millones de colores..

Conexión

Un escáner puede tener diferentes formas de conectarse al ordenador, cada una con ventajas e inconvenientes. Una conexión por puerto paralelo facilita la instalación, pero es notoriamente más lenta que otras soluciones. La alternativa SCSI es mucho más rápida y fiable, aunque es preciso abrir el equipo, y lidiar con la clásica configuración de la cadena de dispositivos SCSI. La utilidad de la conexión SCSI radica en dos apartados: velocidad y pocos requisitos de microprocesador. Lo primero es fácil de entender: la interfaz SCSI puede transmitir de 5 a 80 MB/s, dependiendo del estándar SCSI en concreto, mientras que el puerto paralelo a duras penas supera 1 MB/s. La otra cualidad de SCSI incide también en la velocidad, aunque de otra forma. No se trata sólo de que se puedan transmitir 10 ó 20 MB/s, sino que además dicha transferencia se realiza sin que el microprocesador realice apenas trabajo; esto permite ir muestreando imágenes mientras realizamos otras tareas, agilizando mucho el trabajo. En un escáner paralelo resulta muy normal que mientras se realiza el muestreo el rendimiento del ordenador baje muchísimo.
La conexión por el puerto USB es lo más común. En general podríamos decir que los escáneres USB se sitúan en un punto intermedio de calidad/precio. La velocidad de transmisión ronda los 1,5 MB/s, algo más que el puerto paralelo pero bastante menos que el SCSI; la facilidad de instalación es casi insuperable, ya que es “Plug and Play”; todos los ordenadores modernos tienen el USB incorporado; y además dejan el puerto paralelo libre para imprimir o conectar otros dispositivos. Se trata, en fin, de una solución claramente enfocada al usuario doméstico u oficinista.

Tecnología del escáner

En el momento existen 2 tecnologías que compiten en la fabricación de Escáner: CIS (“Contact Image Sensor” o Sensores de Imágenes por Contacto), de reciente aparición, y CCD (“Charge-Cuopled Device” o Dispositivo de Carga Acoplada).
El escáner CIS elimina los espejos y las lentes, así como el tubo de rayos catódicos. Además, la proximidad de los sensores con el original conduce a la fabricación de escáneres más delgados. Sin embargo, la tecnología no ha evolucionado lo suficiente para alcanzar la calidad de imagen del escáner CCD.
El corazón de los escáneres CCD es un pequeño chip semiconductor sensible a la luz, que requiere de un sistema de espejos y lentes para colocar la imagen en foco.

Tipos de escáner:

- **ESCÁNER PLANO** Son los llamados escáneres de sobremesa, están formados por una superficie plana de vidrio sobre la que se sitúa el documento a escanear, generalmente opaco, bajo la cual un brazo se desplaza a lo largo del área de captura.
  Están indicados para digitalizar objetos opacos planos (como fotografías, documentos o ilustraciones) cuando no se precisa ni una alta resolución ni una gran cantidad.
  Los escáneres planos son los más utilizados para cubrir necesidades personales y de pequeñas empresas, así como para procesar documentos, fotografías, libros y otros impresos, cuando es pequeña la colección por digitalizar.
  La mayor desventaja de estos escáneres es la limitación respecto al tamaño del documento a escanear, que queda limitado a los formatos DIN-A5 o DIN-A4.



Figura 13. Escáner plano

- **ESCÁNER CENITAL O AÉREO** Un escáner cenital es un tipo de escáner que se utiliza para hacer copias digitales de libros o documentos que, por ser viejos o extremadamente valiosos, se digitalizan con otro tipo de escáner para que no se deterioren.
  El precio de estos escáneres es elevado y suelen utilizarse en museos, archivos y empresas dedicadas a la digitalización de documentos antiguos, con personal especializado.
  Estos escáneres consisten en una cámara montada en un brazo que toma fotos del elemento deseado. Su ventaja principal es que los libros no tienen que ser abiertos completamente (como pasa en la mayoría de los escáneres planos). El escaneo de volúmenes encuadernados se realiza gracias a que la fuente de luz y el sensor CCD se encuentran ensamblados a un brazo de trayectoria aérea.



Figura 14. Escáner cenital

- **ESCÁNER DE TAMBOR** Los escáneres de tambor son los que más fielmente reproducen el documento original, ya que producen digitalizaciones de gran calidad y resolución (hasta 16.000 ppp de resolución óptica). Sus problemas son la velocidad de escaneo (son lentos), no son indicados para documentos de papel quebradizo porque es necesario curvarlo sobre el tambor y requieren un alto nivel de habilidad por parte del operador. Además, son bastante caros.
  Son muy caros, por lo que suelen ser usados exclusivamente por empresas especializadas del sector de las artes gráficas (laboratorios, imprentas, editoriales, etc.).

Utilizan una tecnología diferente a la del CCD. Los originales, normalmente transparencias (aunque se pueden escanear opacos también), se colocan en un cilindro transparente de cristal de gran pureza, que a su vez se monta en el escáner. El tambor gira entonces a gran velocidad mientras se hace la lectura de cada punto de la imagen. La fuente de luz suele ser un láser que se encuentra dentro del tambor (para transparencias) o fuera (para opacos), y el sensor es un tubo fotomultiplicador que recibe la luz de un único punto de la imagen en cada instante. Producen digitalizaciones de alta resolución y buena gama dinámica entre bajas y altas luces, con imágenes en colores primarios, que pueden ser convertidas en CMYK mientras el lector recorre la imagen.



Figura 15. Escáner de tambor

- **ESCÁNER PARA TRANSPARENCIAS O DE RANURA** Los escáneres para transparencias se utilizan para digitalizar diapositivas, negativos fotográficos y documentos que no son adecuados para el escaneado directo. Pueden trabajar con varios formatos de película transparente, ya sea negativa, positiva, color o blanco y negro, de tamaño desde 35 mm hasta placas de 9 x 12 cm. La calidad obtenida es mayor que la que ofrecen los escáneres planos, aunque hay que tener cuidado con la presencia de motas de polvo o rascaduras en las transparencias, que pueden ocasionar la aparición de impurezas en la imagen digitalizada resultante.
  Existen dos modalidades de este tipo de escáneres:
    - Escáneres de 35 mm. Solo escanean negativos y transparencias, pero lo hacen a resoluciones muy altas.
    - Escáneres multiformato. Suelen capturar transparencias y negativos hasta formato medio o hasta formato de placas 4”x 5” o incluso 5”x 7”, tienen una resolución muy alta y un rango dinámico en ocasiones sorprendente, pero frecuentemente no permiten escanear opacos. El uso de medios transparentes por lo general produce imágenes con un buen rango dinámico, pero, dependiendo del tamaño del original, la resolución puede ser insuficiente para algunas necesidades.



Figura 16. Escáner de ranura

- **ESCÁNER DE MANO** Estos escáneres son dispositivos manuales que son arrastrados sobre la superficie de la imagen a escanear. Escanear documentos de esta manera requiere una mano firme, porque una velocidad desigual de exploración produce imágenes distorsionadas, normalmente un indicador luminoso sobre el escáner indica si la exploración fue demasiado rápida. Generalmente tienen un botón "Inicio", el cual es sostenido por el usuario durante la exploración y algunos interruptores para configurar la resolución óptica; y tiene un rodillo, lo que genera un reloj de pulso para sincronización con el ordenador.
  La mayoría de escáneres de mano fueron en blanco y negro, y la luz generada por una serie de LEDs verdes para iluminar la imagen. Un típico escáner de mano también tenía un programa que abría una pequeña ventana a través de la cual se podía ver el documento que se escaneaba. Fueron populares durante la década de los 90 y, por lo general tenían un módulo

de interfaz propietario específico para un determinado tipo de computadora, generalmente un Atari ST o Commodore Amiga.



Figura 17. Escáner de mano

- **ESCÁNER PARA MICROFILM** Los escáneres para microfilm son dispositivos especializados en digitalizar películas en rollo, microfichas y tarjetas de apertura. Puede ser difícil obtener una calidad buena y consistente en un escáner de este tipo, debido principalmente a que suelen tener un funcionamiento complejo, la calidad y condición de la película puede variar y ofrecen una capacidad de mejora mínima. Son escáneres muy caros, existiendo pocas empresas que los fabriquen.



Figura 18. Escáner para microfilm

- **ESCÁNER 3D** Hoy en día existe la posibilidad de utilizar escáneres que captan imágenes en 3 dimensiones. Diversos fabricantes ofrecen productos de estas características que pueden ser utilizados para múltiples utilidades. Desde escanear el cuerpo humano para aplicaciones médicas, hasta digitalizar fachadas de edificios o monumentos.

2.2. Reconocimiento de caracteres

Existen unos procesos que revolucionaron el mundo laboral. Por supuesto hablamos del OCR (por sus siglas en inglés Optical Carachter Recognition), ICR (Intelligent Carachter Recognition) y el OMR (Optical Mark Recognition). El objetivo de estas soluciones es la reducción de tiempo para recabar información y para poder tener un acceso rápido y sencillo a la misma.
El **OCR** es el proceso que analiza los caracteres impresos y determina su forma utilizando patrones de oscuros y claros. Una vez que el escáner o el lector han determinado las formas, éstas se comparan con conjuntos de caracteres definidos para traducirlas a un **texto**. En algunas ocasiones el reconocimiento óptico de caracteres se realiza con lectores especiales, pero lo más frecuente es utilizar un escáner óptico estándar y un software especializado.

El proceso de OCR se inicia con una digitalización en blanco y negro de la página de la cual queremos extraer el texto. Por ejemplo: en una empresa existen cientos y cientos de documentos que conforman el expediente de la plantilla laboral. Si necesitamos corroborar algunos datos debemos ir al archivo para saber por ejemplo si el trabajador alguna vez ha estado afiliado al seguro social. Pues con el OCR este proceso se hace en unos cuantos minutos, siempre y cuando tengamos nuestros documentos ya digitalizados.
El **ICR** es una solución cuyo objeto es el reconocimiento de documentos con **escritura a mano**. Esta es una gran desventaja ya que no todos escribimos de manera uniforme, y no todos escribimos de la misma forma. Sin embargo hay quienes aseguran que esta solución tiene mucho futuro y una gran área de oportunidad. Debido a su margen de error es poco utilizada.
El **OMR** es una solución más específica y más especializada. Su objetivo es el reconocimiento de **marcas**, ya sea en forma de triángulo, en forma de cuadro, forma circular, etc. Por ejemplo: para tratar las respuestas de los cuestionarios realizados en un examen (como el nuestro de TAI). Hasta hace algunos años esto era tarea de nunca acabar, pero hoy en día es mucho más fácil y rápido hacer este proceso. Si utilizamos un cuadernillo de preguntas y una hoja de respuestas con círculos a rellenar, el OMR puede ser configurado para que reconozca estas marcas y las tome como respuesta. Gracias al OMR ya no es necesario tabular cuestionario por cuestionario para poder corregirlos a mano.
OBR: Optical Bar Code Reader, usado para reconocimiento de **códigos de barras**.

2.3. Almacenamiento de ficheros de imágenes

Existen diversos formatos de ficheros de almacenamiento de imágenes digitalizadas dependiendo del software de utilidad con el que trabajemos. La mayoría de escáneres suele aceptar la mayor parte de estos formatos. Otros están limitados a uno o dos formatos, lo cual está bien si estos son los formatos que utiliza su aplicación. El formato de fichero soportado por la mayor parte de los escáneres es el **TIFF**, desarrollado por Aldus y Microsoft. La versión más reciente de la especificación TIFF (5.0) soporta imágenes con escala de grises y en color, además de blanco y negro. Otro formato de fichero soportado por la mayor parte de los escáneres es .**PCX**, utilizado por el popular dibujo de ZSoft, PC Paintbrush. Otros formatos usuales son .**EPS** (Postcript), .**IMG** (formato GEM de Digital) o .**PFF** (utilizado por las impresoras Epson).
Precisamente es el software de utilidad quien determina, al menos en parte, lo que el ordenador hará con las imágenes digitalizadas que recibe. Cualquiera que sea el formato de fichero utilizado, la capacidad de la memoria debe ser alta, no para el proceso de digitalización en sí, sino para la manipulación de las imágenes digitalizadas (normalmente 2M de RAM o más, sobre todo si hay múltiples escalas de gris). Por ejemplo, explorar un documento de 8.5 x 11 pulgadas a 300 ppp con un simple escáner de dos tonos, da como resultado un fichero de 1M. Si se tratase de una imagen con 256 tonos de gris, el resultado sería de 8M.
Los escáneres diseñados estrictamente para OCR a menudo utilizan un programa interno basado en el firmware, y se denominan escáner de texto. Sin embargo, la mayoría de los escáneres se basan en el software de terceros para convertir la imagen del documento explorado en un fichero ASCII. Esto ofrece flexibilidad en la elección del paquete de OCR. Incluso los escáneres que realizan el OCR internamente pueden sacar provecho del software de otros fabricantes, siempre que generen, por ejemplo, ficheros TIFF.

2.4. El lápiz óptico

El lápiz óptico permite marcar un punto en la pantalla de un monitor CRT. Este dispositivo consta, en esencia, de una fotocélula y de un pulsador. Físicamente tiene la forma de una pluma o lápiz grueso, de uno de cuyos extremos sale un cable para unirlo a un monitor. El lápiz

contiene un pulsador, transmitiéndose información hacia el monitor sólo en el caso de estar presionado.



Figura 19. Lápiz óptico

Funcionamiento

La dificultad principal que entrañaba el lápiz óptico es que era el dispositivo de movimiento directo, lo que significa que el cursor debería encontrarse en el lugar mismo donde sé hallarse la punta del lápiz. Para ello, se necesitaba que el ordenador supiese en cada momento dónde apuntaba el lápiz. La cuestión planteada no será tan trivial. A pesar de ello, el problema se resolvió de una forma de lo más sencilla posible, aunque no deja de ser ingeniosa. Para entenderlo, es necesario conocer antes a grandes rasgos cómo funciona la pantalla de un monitor.

La imagen que puede visualizarse en este no se transmite de una sola vez sino que es generada punto a punto, por el adaptador de media localizado en el hardware gráfico del ordenador. Una vez en el monitor, un haz de electrones se encarga de recorrer línea a línea toda la pantalla varias veces por segundo, (más de cincuenta veces por segundo, dependiendo del monitor), con la rapidez suficiente como para engañar al ojo humano, cuando éste quede fijado ante la pantalla. Se trata de una acción que suele denominarse “raster scann” o refresco.

El lápiz óptico aprovecha esta característica, detectando por medio de un sensor colocado en su interior, cuando un punto (o pixel) de la imagen es "refrescado" por el haz. Es el momento en el cual dicho punto emite el brillo algo superior al que suele poseer. Ese destello, convenientemente amplificador por medio de una lente situada en la punta del lápiz, sirve para provocar en impulso de corriente. El impulso se convierte posteriormente a una señal de retorno que se le envía al controlador de vídeo del ordenador.

Una vez allí el controlador de vídeo se limita a medir el tiempo que ha ido transcurriendo desde la última vez que se comenzó a trazar las líneas en la parte superior de la pantalla, y el instante en que recibió la señal proveniente del lápiz.

Con esta información en su poder, calcula sobre qué línea horizontal se encuentra situado el lápiz, y a qué distancia se halla del comienzo de esta. Por último, y mediante una sencilla traslación, todos los últimos datos se convierten y representan en las coordenadas X e Y. Una información que necesita conocer el programa controlador (o driver). Al igual que sucede con los ratones, los lápices ópticos suelen llevar consigo un botón. El botón puede servir tanto para seleccionar opciones, pero también como marcar al ordenador en qué momento quiere activar el lápiz.

Es incómodo mantener el lápiz sobre la pantalla largos períodos de tiempo. Y debido a que si la zona a la que apunta no está iluminada, no genera señal, y por lo tanto no se puede controlar su posición, el lápiz evolucionó hacia su uso como cursor, pero por este motivo está ha sido sustituido progresivamente por el ratón, que es mucho más práctico y cómodo.

Este tipo de dispositivo es más adecuado para aplicaciones con representación visual vectorial que para aquellas que tienen una representación visual que es generada por barrido.

2.5. Tabletas digitalizadoras

Las tabletas digitalizadoras (o digitalizador) son unidades de entrada que generan las coordenadas de un puntero móvil en el plano o incluso en el espacio (modelos tridimensionales). El puntero puede ser un lápiz o un cursor móvil, que un operador puede desplazar siguiendo un dibujo u objeto. El puntero puede disponer de pulsadores para seleccionar la función deseada del computador, por ejemplo, para marcar un punto, solicitar digitalización continua o parar dicha digitalización. Es éste un periférico que nos permite utilizar un ratón como si fuera un bolígrafo. De hecho las tablas digitalizadoras vienen acompañadas de su correspondiente bolígrafo; con el que nosotros escribimos sobre la tabla. Esa información es enviada al ordenador que nos muestra los resultados de nuestras acciones en pantalla.



Figura. 20. Tableta digitalizadora

Todo digitalizador consta de tres elementos:

- Tabla: Donde se ubica el dibujo a digitalizar (puede ser opaca o transparente).
- Mando: Con el que el usuario debe recorrer el dibujo. Este suele tener forma de lápiz o cursor, y está unido al resto del sistema por un cable flexible. En el último caso el cursor tiene una ventana cerrada con una lupa, en cuyo interior se encuentra embebida una retícula en forma de cruz para señalar o apuntar con precisión el punto a digitalizar. El mando puede disponer de uno o varios pulsadores para controlar la modalidad de funcionamiento, forma de transmisión y selección de opciones del programa que gestiona la digitalización.
- Circuitos electrónicos: Controlan el funcionamiento de la unidad.

Dispositivos similares

Algunas pizarras interactivas operan de manera similar a las tabletas digitalizadoras, hay fabricantes que ofrecen paneles de alta resolución y tamaño hasta de 95 pulgadas. Las pizarras interactivas están extendidas en las escuelas de UK, US y México. Las Pantallas Táctiles como las que se encuentran en algunos Tablet PCs y en la videoconsola Nintendo DS se utilizan de manera similar, pero en lugar de medir la señal electromagnética, utilizan una capa sensible a la presión sobre la superficie, de tal manera que no necesitan un lapicero o estilete especial para utilizarlas. Otros dispositivos táctiles son de gran ayuda para personas ciegas o con problemas de visión, como Tactile Talking Tablet o T3.

3. Elementos de impresión: impresoras

3.1. Introducción

Una impresora es un dispositivo periférico del ordenador que permite producir una gama permanente de textos o gráficos de documentos almacenados en un formato electrónico, imprimiéndolos en medios físicos, normalmente en papel, utilizando cartuchos de tinta o tecnología láser (con tóner).
Muchas de las impresoras son usadas como periféricos, y están permanentemente unidas al ordenador por un cable. Otras impresoras, llamadas impresoras de red, tienen una interfaz de red interno (típicamente wireless o ethernet), y que puede servir como un dispositivo para imprimir en papel algún documento para cualquier usuario de la red.
Además, muchas impresoras modernas permiten la conexión directa de aparatos de multimedia electrónicos como las tarjetas CompactFlash, Secure Digital o Memory Stick, pendrives, o aparatos de captura de imagen como cámaras digitales y escáneres. También existen aparatos multifunción que constan de impresora, escáner o máquinas de fax en un solo aparato. Una impresora combinada con un escáner puede funcionar básicamente como una fotocopiadora.
Son diseñadas para realizar trabajos repetitivos de poco volumen, que no requieran virtualmente un tiempo de configuración para conseguir una copia de un determinado documento. Sin embargo, las impresoras son generalmente dispositivos lentos (10 páginas por minuto es considerado rápido), y el gasto por página es relativamente alto.
Para trabajos de mayor volumen existen las imprentas, que son máquinas que realizan la misma función que las impresoras pero están diseñadas y optimizadas para realizar trabajos de impresión de gran volumen, como sería la impresión de periódicos. Las imprentas son capaces de imprimir cientos de páginas por minuto o más.

3.2. Clasificación de las impresoras

En general, las impresoras se pueden dividir en categorías siguiendo diversos criterios.

- Tecnología de impresión:
    - **impresoras de impacto**: se dividen en,
        - impresoras matriciales (pueden subdividirse según el número de agujas que contiene el cabezal de impresión: 9, 18, 24),
        - impresoras de margarita;
    - **impresoras sin impacto**: abarcan todos los demás tipos de mecanismos de impresión, incluyendo:
        - impresoras térmicas,
        - impresoras de inyección o impresoras de chorro de tinta,
        - impresoras láser.
- Formación de los caracteres:
    - **Caracteres con trazo continuo**: los caracteres formados totalmente con trazo continuo (los producidos por una impresora de margarita, por ejemplo).
    - **Caracteres de puntos matriciales**: caracteres matriciales compuestos por patrones de puntos independientes (impresoras: matriciales, de inyección y térmicas).

  Técnicamente, las impresoras láser son matriciales, pero la nitidez de la impresión y el tamaño reducido de los puntos impresos con alta densidad, se puede considerar que los trazos de sus caracteres son continuos.
- Método de transmisión: se refiere al medio utilizado para enviar los datos a la impresora:

- **Paralelo**: transmisión byte a byte.
- **Serie**: transmisión bit a bit.

Muchas versiones de impresoras estaban disponibles en paralelo y en serie, e incluso incorporaban ambas opciones, aumentando la flexibilidad para instalarlas. Actualmente, la tendencia es a favor de las impresoras en serie, a través del estándar USB.

- Método de impresión:
  - **Carácter a carácter**: las impresoras de caracteres son las impresoras:
    - matriciales,
    - de inyección de tinta,
    - térmicas,
    - de margarita.
  - **Línea a línea**: se utilizan frecuentemente en grandes instalaciones (p. e.: centros de cálculo, entornos industriales). Las impresoras de líneas se subdividen en impresoras:
    - de cinta,
    - de cadena,
    - de tambor.
  - **Página a página**: entre las impresoras de páginas se encuentran las electrofotográficas, como las impresoras láser.
- Capacidad de impresión.
  - **Solo texto**: la mayoría de impresoras de margarita y de bola pueden imprimir solamente textos, aunque también existen impresoras matriciales y láser que imprimen solamente caracteres.
  - **Texto y gráficos**: pueden reproducir solamente caracteres previamente grabados, en relieve o en forma de mapa de caracteres interno. Las impresoras de textos y gráficos, reproducen todo tipo de imágenes dibujándolas como patrones de puntos:
    - matriciales,
    - de inyección de tinta,
    - láser.

Velocidad de impresión y calidad del impreso

Los distintos tipos de impresoras se diferencian en la velocidad de impresión y en la calidad del producto impreso.

Las impresoras de caracteres, como las matriciales, imprimen en un rango de velocidad entre 200 y 400 **caracteres por segundo** (cps), que supone de 90 a 180 líneas por minuto (lpm). Las impresoras de línea presentan un amplio rango de velocidades, desde 400 a 2000 **líneas por minuto**. La velocidad de las impresoras de página oscila entre 4 y 800 **páginas por minuto** (ppm) para impresiones en blanco y negro, y la décima parte para la impresión en color.

En entornos de oficinas en los que se empleen formularios en papel continuo o de varias hojas de papel continuo, la impresora más adecuada es la de matriz de puntos, pero si se requiere mayor calidad de impresión se utilizará impresora láser. Las impresoras de inyección de tinta son las preferidas para entornos domésticos, por precio asequible.



Margarita



Matricial



Láser



De Tinta



Térmica



Plotter





Figura 21. Tipos de impresoras

3.3. Métodos de impresión

La elección del motor de compresión tiene un efecto substancial en los trabajos a los que una impresora está destinada. Hay diferentes tecnologías que tienen diferentes niveles de calidad de imagen, velocidad de impresión, coste, ruido y además, algunas tecnologías son inapropiadas para ciertos tipos de medios físicos (como papel carbón o transparencias).
Otro aspecto de la tecnología de impresión que es frecuentemente olvidado es la resistencia a la alteración: la tinta líquida como de una cabeza de inyección de tinta es absorbida por las fibras del papel, y por eso los documentos impresos con tinta líquida son más difíciles de alterar que los que están impresos por tóner o tinta sólida, que no penetran por debajo de la superficie del papel.

Tóner

Las impresoras láser e impresoras térmicas utilizan este método para adherir tóner al medio. Trabajan utilizando el principio de la *xerografía* que está funcionando en la mayoría de las fotocopiadoras: adhiriendo tóner a un tambor de impresión sensible a la luz, y utilizando electricidad estática para transferir el tóner al medio de impresión al cual se une gracias al calor y la presión.
Las impresoras láser son conocidas por su impresión de alta calidad, buena velocidad de impresión y su bajo costo por copia; son las impresoras más comunes para muchas de las aplicaciones de oficina de propósito general. Son menos utilizadas por el consumidor, generalmente debido a su alto coste inicial. Las impresoras láser están disponibles tanto en color como en monocromo.
El advenimiento de láseres de precisión a precio razonable ha hecho a la impresora monocromática basada en tóner dominante en aplicaciones para la oficina. Otro tipo de impresora basada en tóner es la impresora led la cual utiliza una colección de ledes en lugar de láser para causar la adhesión del tóner al tambor de impresión. El tóner (del inglés, *toner*), también denominado tinta seca por analogía funcional con la tinta, es un polvo fino, normalmente de color negro, que se deposita en el papel que se pretende imprimir por medio de atracción electrostática.
Una vez adherido el pigmento, este se fija en el papel por medio de presión o calor adecuados. Debido a que en el proceso no intervienen diluyentes, originalmente se ha denominado xerografía, del griego *xeros* que significa seco.

Inyección de tinta

Las impresoras de inyección de tinta (*Ink Jet*) rocían hacia el medio cantidades muy pequeñas de tinta, usualmente unos picolitros. Para aplicaciones de color, incluyendo impresión de fotos, los métodos de chorro de tinta son los dominantes, ya que las impresoras de alta calidad son poco costosas de producir. Virtualmente todas las impresoras de inyección son

dispositivos en color; algunas, conocidas como impresoras fotográficas, incluyen pigmentos extra para una mejor reproducción de la gama de colores necesaria para la impresión de fotografías de alta calidad (y son adicionalmente capaces de imprimir en papel fotográfico, en contraposición al papel normal de oficina).
Las impresoras de inyección de tinta consisten en inyectores que producen burbujas muy pequeñas de tinta que se convierten en pequeñísimas gotitas de tinta. Los puntos formados son el tamaño de los pequeños pixeles. Las impresoras de inyección pueden imprimir textos y gráficos de alta calidad de manera casi silenciosa.
Existen dos métodos para inyectar la tinta:

1. Método térmico. Un impulso eléctrico produce un aumento de temperatura (aprox. 480 °C durante microsegundos) que hace hervir una pequeña cantidad de tinta dentro de una cámara formando una burbuja de vapor que fuerza su salida por los inyectores. Al salir al exterior, este vapor se condensa y forma una minúscula gota de tinta sobre el papel. Después, el vacío resultante arrastra nueva tinta hacia la cámara. Este método tiene el inconveniente de limitar en gran medida la vida de los inyectores, por eso estos inyectores se encuentran en los cartuchos de tinta.
2. Método piezoeléctrico. Cada inyector está formado por un elemento piezoeléctrico que, al recibir un impulso eléctrico, cambia de forma aumentando bruscamente la presión en el interior del cabezal provocando la inyección de una partícula de tinta. Su ciclo de inyección es más rápido que el térmico.

Las impresoras de inyección tienen un coste inicial mucho menor que las impresoras láser, pero tienen un coste por copia mucho mayor, ya que la tinta necesita ser repuesta frecuentemente. Las impresoras de inyección son también más lentas que las impresoras láser, además de tener la desventaja de dejar secar las páginas antes de poder ser manipuladas agresivamente; la manipulación prematura puede causar que la tinta (que está adherida a la página en forma líquida) se mueva.

Tinta sólida

Las impresoras de tinta sólida, también llamadas de cambio de fase, son un tipo de impresora de transferencia térmica pero utiliza barras sólidas de tinta en color CMYK (similar en consistencia a la cera de las velas). La tinta se derrite y alimenta una cabeza de impresión operada por un cristal piezoeléctrico (por ejemplo cuarzo). La cabeza distribuye la tinta en un tambor engrasado. El papel entonces pasa sobre el tambor al tiempo que la imagen se transfiere al papel.
Son comúnmente utilizadas como impresoras en color en las oficinas, ya que son excelentes imprimiendo transparencias y otros medios no porosos, y pueden conseguir grandes resultados. Los costes de adquisición y utilización son similares a las impresoras láser.
Las desventajas de esta tecnología son el alto consumo energético y los largos periodos de espera (*calentamiento*) de la máquina. También hay algunos usuarios que se quejan de que la escritura es difícil sobre las impresiones de tinta sólida (la cera tiende a repeler la tinta de los bolígrafos), y son difíciles de alimentar de papel automáticamente, aunque estos rasgos han sido significantemente reducidos en los últimos modelos. Además, este tipo de impresora solo se puede obtener de un único fabricante, Xerox, como parte de su línea de impresoras de oficina Xerox Phaser. Previamente las impresoras de tinta sólida fueron fabricadas por Tektronix, pero vendió su división de impresión a Xerox en el año 2000.

Impacto

Las impresoras de impacto o impresoras de golpe se basan en la fuerza de impacto para transferir tinta al medio, de forma similar a las máquinas de escribir, están generalmente limitadas a reproducir texto. En su momento dominaron la impresión de calidad.

Hay dos tipos principales:

1. Impresora de margarita, llamada así por tener los tipos contenidos radialmente en una rueda, de ahí su aspecto de una margarita.
2. Impresora de rueda, llamada así por tener todos los tipos contenidos en una esfera. Es el caso de las máquinas de escribir eléctricas IBM Selectric.

Las impresoras de impacto trabajan con un cabezal en el que hay agujas, estas agujas golpean una cinta, similar al de una máquina de escribir, que genera la impresión de la letra.

Matriz de puntos

En el sentido general, muchas impresoras se basan en una matriz de muchos píxeles o puntos que, juntos, forman la imagen más grande. Sin embargo, el término matriz o de puntos se usa específicamente para las impresoras de impacto que utilizan una matriz de pequeños alfileres para crear puntos precisos. Dichas impresoras son conocidas como matriciales. La ventaja de la matriz de puntos sobre otras impresoras de impacto es que estas pueden producir imágenes gráficas además de texto. Sin embargo, el texto es generalmente de calidad más pobre que las impresoras basadas en impacto de tipos.

Algunas sub-clasificaciones de impresoras de matriz de puntos son las impresoras de alambre balístico y las impresoras de energía almacenada.

Las impresoras de matriz de puntos pueden estar basadas bien en caracteres o bien en líneas, refiriéndose a la configuración de la cabeza de impresión.

Las impresoras de matriz de puntos son todavía de uso común para aplicaciones de bajo costo y baja calidad como las cajas registradoras. El hecho de que usen el método de impresión de impacto les permite ser usadas para la impresión de documentos autocopiativos como los recibos de tarjetas de crédito, donde otros métodos de impresión no pueden utilizar este tipo de papel. Las impresoras de matriz de puntos han sido superadas para el uso general en computación.

Sublimación de tinta

Las impresoras de sublimación de tinta emplean un proceso de impresión que utiliza calor para transferir tinta a medios como tarjetas de plástico, papel o lienzos. El proceso consiste usualmente en poner un color cada vez utilizando una cinta que tiene paneles de color. Estas impresoras están principalmente pensadas para aplicaciones de color de alta calidad, incluyendo fotografía en color, y son menos recomendables para texto. Primeramente utilizadas en las copisterías, cada vez más se están dirigiendo a los consumidores de impresoras fotográficas.

Térmica

Las impresoras térmicas se basan en una serie de agujas calientes que recorren el papel termosensible que al contacto se vuelve de color negro. Por su bajo coste, son muy usadas en los cajeros automáticos y supermercados.

3D

Una impresora 3D es una máquina capaz de realizar réplicas de diseños en 3D, creando piezas o maquetas volumétricas a partir de un diseño hecho por ordenador, descargado de internet o recogido a partir de un escáner 3D. Surgen con la idea de convertir archivos de 2D en prototipos reales o 3D.
Comúnmente se ha utilizado en la prefabricación de piezas o componentes, en sectores como la arquitectura y el diseño industrial. En la actualidad se está extendiendo su uso en la fabricación de todo tipo de objetos, modelos para vaciado, piezas complicadas, alimentos, prótesis médicas (ya que la impresión 3D permite adaptar cada pieza fabricada a las características exactas de cada paciente), etc.

3.4. Memoria de las impresoras

Las impresoras llevan consigo memoria interna. Van desde los 6 KB en las impresoras matriciales hasta como mínimo 2 MB en las impresoras láser.
Actualmente en las láser venden módulos de memoria independientes para ampliar la capacidad de la misma.
La memoria se usa como búfer y como almacenamiento permanente y semipermanente. Además su uso es necesario porque el tratamiento de gráficos vectoriales y el diseño de fuentes en mapa de bits consumen memoria.
El búfer es utilizado para mantener trabajos de impresión activos y la permanencia se utiliza para almacenar el diseño de las fuentes y los datos.
Hay que tener en cuenta que para tratar la impresión de un documento la página tiene que estar enteramente almacenada en memoria. El rendimiento de la memoria depende tanto del sistema operativo como de la configuración del controlador de impresora.
Por ejemplo, la gestión de impresión varía si estamos en un sistema operativo DOS u otro multiplataforma.

3.5. Conexión de impresora

La conexión de la impresora con el computador ha ido evolucionando conllevando a la mejora de rendimiento de impresión y comodidad de usuario.
La forma más antigua de conexión era mediante puerto serie en donde la transferencia se hacía bit a bit, permitía distancias largas con velocidades lentas que no superaban los 19 200 bytes/segundo.
Se elevó hasta la conexión mediante puerto paralelo en la que las transferencias eran byte a byte permitiendo 8 conexiones paralelas consiguiendo una velocidad más rápida entre los ½ MB/segundo hasta los 4 MB/segundo. El inconveniente era la limitación de la distancia del cable que une la impresora con el computador ya que no permite una longitud mayor de 2 metros.

Otra forma de conexión se consiguió poniendo la impresora en red Ethernet mediante conexiones RJ-45 basadas en el estándar IEEE 802.3. Las velocidades conseguidas superan los 10 Mb/segundo basada en el manejo de paquetes. No hay que confundirla con una impresora compartida, ya que las impresoras en red operan como un elemento de red con dirección IP propia.
Otro método de conexión más actual es por medio de puertos USB (*Universal Serial Bus*). La velocidad vuelve a mejorar con 480Mb/segundo con las ventajas que conlleva el puerto USB: compatibilidad con varios sistemas y la posibilidad de usarla en dispositivos portátiles.
Finalmente, la conexión inalámbrica Wi-Fi, mediante el protocolo IEEE 802.11, está siendo la más novedosa. Alcanza 300 Mb/segundo y funciona tanto para impresoras de tinta, láser o multifunción.
Aunque consigue menos velocidad que las conectadas por USB, las wifi proporcionan ventajas tales como la autonomía, la movilidad y libertad del usuario sin la utilización de cables. Para la correcta utilización y evitar accesos no deseados deberemos cifrar la red.

3.6. Lenguajes de descripción de página y formatos de impresión

Un "lenguaje de descripción de página" (PDL) es un medio de codificar cada elemento de un documento para poder así transmitirlo a la impresora para que esta lo imprima. Es el medio que define las características y composición que describirían un documento impreso dentro de un flujo de datos. Hay varios tipos de PDL:

- PostScript
- *Printer Command Language* (PCL), lenguaje de control de impresora
- HPGL, para plóteres

Formatos de definición de caracteres: *Truetype*
Fue creado por Apple para no depender tecnológicamente de los tipos PostScript de Adobe, pero su calidad resultó ser inferior. Fue comprada por Microsoft lo cual ha contribuido a que no llegara a desaparecer. La principal fortaleza de *TrueType* es que ofrece a los diseñadores de fuentes un gran grado de control sobre la forma que sus fuentes se muestran a diferentes tamaños.
El problema con la mayoría de los programas es que no usan normalmente el truetype. En general cargan las fuentes en estilo Postscript y se descartan todas las insinuaciones; esto es una gran pérdida para fuentes con alta calidad. Aparte del diseño de la fuente, hay que tener en cuenta otras dos claves para la calidad de fuente: el perfil del carácter y la insinuación. Solo algunas fundiciones actualmente producen fuentes que exploten al máximo el potencial de insinuación de truetype. Ahora hay aplicaciones que convierten un Type 1 de Postscript en un truetype, pero son los manuscritos mejores que los generados automáticamente.

3.7. Plotter

Los plóteres sirven para hacer impresiones de dibujo de planos de arquitectura, ingeniería, diseño industrial, etc., para la impresión de láminas, pósteres, ampliaciones fotográficas, gigantografías, carteles en rutas, vía pública, señalización, etc. Existen dos clases de plóter según el uso de sus tintas, a base de agua o disolventes. Un caso particular es el plóter de

corte, que corta un medio adhesivo que luego se fijará a otra superficie, desde camisetas a carrocerías.

4. Elementos de visualización: monitor

4.1. Introducción

El **monitor** o **pantalla del ordenador** es el principal dispositivo de salida (interfaz), que muestra datos o información al usuario.
También puede considerarse un periférico de Entrada/Salida si el monitor tiene pantalla táctil o multitáctil.

Parámetros de una pantalla:

- **Píxel**: unidad mínima representable en un monitor. Los monitores pueden presentar píxeles muertos o atascados. Se notan porque aparecen en blanco. Más común en portátiles.
- **Tamaño de punto** o (*dot pitch*): el tamaño de punto es el espacio entre dos fósforos coloreados de un píxel. Es un parámetro que mide la nitidez de la imagen, midiendo la distancia entre dos puntos del mismo color; resulta fundamental a grandes resoluciones. Los tamaños de punto más pequeños producen imágenes más uniformes.
- **Área útil**: el tamaño de la pantalla no coincide con el área real que se utiliza para representar los datos.
- **Ángulo de visión**: es el máximo ángulo con el que puede verse el monitor sin que se degrade demasiado la imagen. Se mide en grados.
- **Luminancia**: es la medida de luminosidad, medida en Candela.
- **Tiempo de respuesta**: también conocido como latencia. Es el tiempo que le cuesta a un píxel pasar de activo (blanco) a inactivo (negro) y después a activo de nuevo.
- **Contraste**: es la proporción de brillo entre un píxel negro a un píxel blanco que el monitor es capaz de reproducir. Algo así como cuantos tonos de brillo tiene el monitor.
- **Coeficiente de contraste de imagen**: se refiere a lo vivo que resultan los colores por la proporción de brillo empleada. A mayor coeficiente, mayor es la intensidad de los colores (30000:1 mostraría un colorido menos vivo que 50000:1).
- **Consumo**: cantidad de energía consumida por el monitor, se mide en Vatio.
- **Ancho de banda**: frecuencia máxima que es capaz de soportar el monitor.
- **Hz o frecuencia de refresco vertical**: son 2 valores entre los cuales el monitor es capaz de mostrar imágenes estables en la pantalla.
- **Hz o frecuencia de refresco horizontal**: similar al anterior pero en sentido horizontal, para dibujar cada una de las líneas de la pantalla.

- **Blindaje**: un monitor puede o no estar blindando ante interferencias eléctricas externas y ser más o menos sensible a ellas, por lo que en caso de estar blindando, o semi-blindado por la parte trasera llevara cubriendo prácticamente la totalidad del tubo una plancha metálica en contacto con tierra o masa.
- **Tipo de monitor**: en los CRT pueden existir 2 tipos, de apertura de rejilla o de máscara de sombra.
- **Líneas de tensión**: son unas líneas horizontales, que tienen los monitores de apertura de rejilla para mantener las líneas que permiten mostrar los colores perfectamente alineadas; en 19 pulgadas lo habitual suelen ser 2, aunque también los hay con 3 líneas, algunos monitores pequeños incluso tienen una sola.

4.2. Tamaño de la pantalla y proporción, medición, resolución máxima y colores

El tamaño de la pantalla es la distancia en diagonal de un vértice de la pantalla al opuesto, que puede ser distinto del área visible cuando hablamos de CRT , mientras que la proporción o relación de aspecto es una medida de proporción entre el ancho y el alto de la pantalla, así por ejemplo una proporción de 4:3 (Cuatro tercios) significa que por cada 4 píxeles de ancho tenemos 3 de alto, una resolución de 800x600 tiene una relación de aspecto 4:3, sin embargo estamos hablando de la proporción del monitor.
Estas dos medidas describen el tamaño de lo que se muestra por la pantalla, históricamente hasta no hace mucho tiempo y al igual que las televisiones los monitores de ordenador tenían un proporción de 4:3. Posteriormente se desarrollaron estándares para pantallas de aspecto panorámico 16:9 (a veces también de 16:10 o 15:9) que hasta entonces solo veíamos en el cine.



Figura 22.Medida de tamaño de la pantalla para TFT.

Medición del tamaño de la pantalla
Las medidas de tamaño de pantalla son diferentes cuando se habla de monitores CRT y monitores LCD:

- Para monitores CRT la medida en pulgadas de la pantalla toma como referencia los extremos del monitor teniendo en cuenta el borde, mientras que el área visible es más pequeña.

- Para monitores LCD la medida de tamaño de pantalla se hace de punta a punta de la pantalla sin contar los bordes.

Los tamaños comunes de pantalla suelen ser de 15, 17, 19, 21 pulgadas. La correspondencia entre las pulgadas de CRT y LCD en cuanto a zona visible se refiere, suele ser de una escala inferior para los CRT, es decir una pantalla LCD de 17 pulgadas equivale en zona visible a una pantalla de 19 pulgadas del monitor CRT (aproximadamente).

Resolución máxima

Es el número máximo de píxeles que pueden ser mostrados en cada dimensión, es representada en filas por columnas. Está relacionada con el tamaño de la pantalla y la proporción.

Los monitores LCD solo tienen una resolución nativa posible, por lo que si se hacen trabajar a una resolución distinta, se escalará a la resolución nativa, lo que suele producir artefactos en la imagen.

Las resoluciones más usadas son:

Estándar	Nombre	Ancho	Alto	% de usuarios de Steam
XGA	eXtended Graphics Array	1024	768	15,37%
WXGA	Widescreen eXtended Graphics Array	1280	800	7,35%
SXGA	Super eXtended Graphics Array	1280	1024	21,01%
WSXGA	Widescreen Super eXtended Graphics Array	1440	900	11,12%
WSXGA+	Widescreen Super eXtended Graphics Array Plus	1680	1050	18,48%

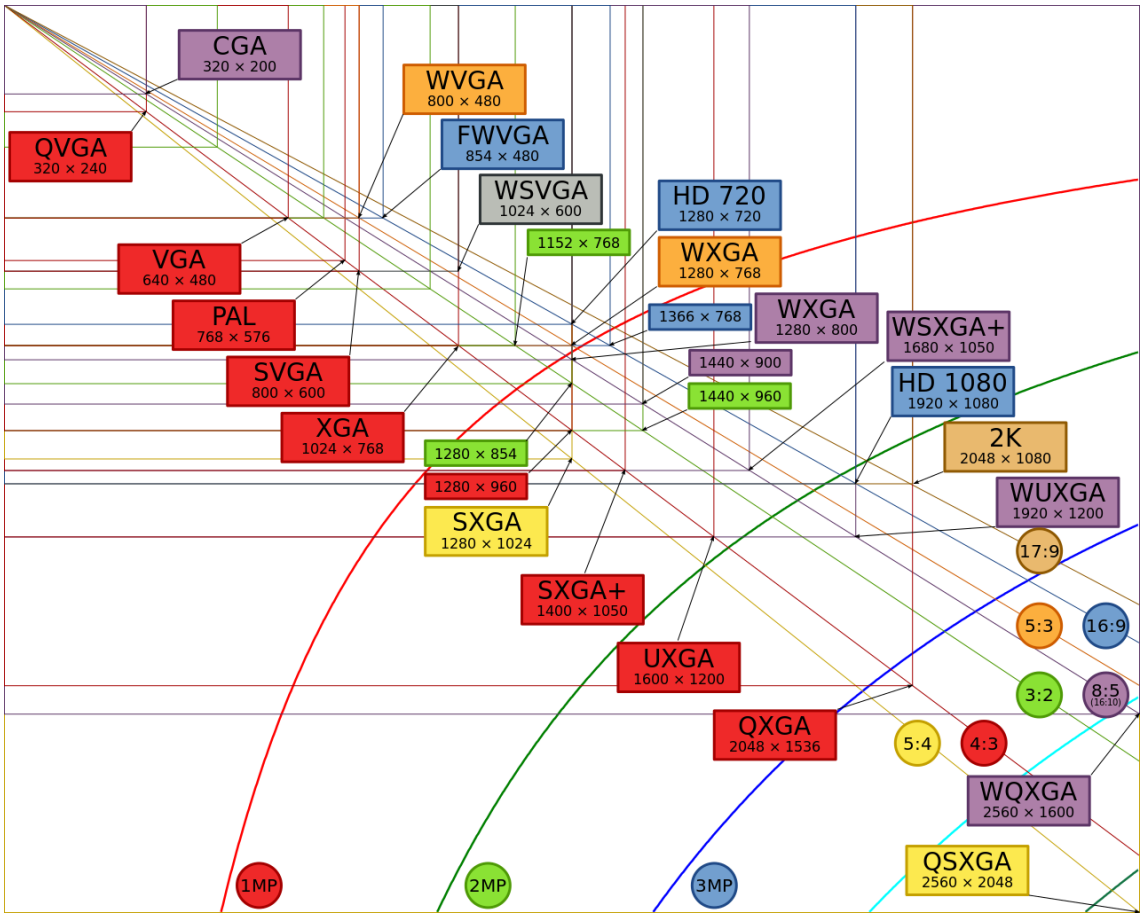


Figura 22. Comparación de resoluciones de vídeo.

Colores

Cada píxel de la pantalla tiene interiormente 3 subpíxeles, uno rojo, uno verde y otro azul; dependiendo del brillo de cada uno de los subpíxeles, el píxel adquiere un color u otro de forma semejante a la composición de colores RGB.

La manera de organizar los subpíxeles de un monitor varía entre los dispositivos. Se suelen organizar en líneas verticales, aunque algunos CRT los organizan en puntos formando triángulos. Para mejorar la sensación de movimiento, es mejor organizarlos en diagonal o en triángulos. El conocimiento del tipo de organización de píxeles, puede ser utilizado para mejorar la visualización de imágenes de mapas de bit usando renderizado de subpíxeles.

La mayor parte de los monitores tienen una profundidad 8 bits por color (24 bits en total), es decir, pueden representar aproximadamente 16,8 millones de colores distintos.

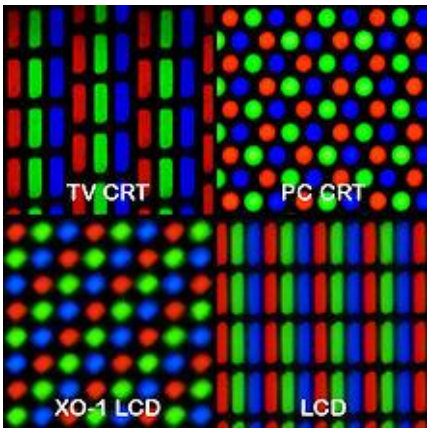


Figura 23. Geometría de los píxeles.

4.3. Clasificación de monitores

- Un monitor puede clasificarse, según la tecnología empleada para formar las imágenes en:
    - Tubo de rayos catódicos o **CRT** (*Cathode Ray Tube*): es una tecnología que permite visualizar imágenes mediante un haz de rayos catódicos constantemente dirigido contra una pantalla de vidrio recubierta de fósforo y plomo. El fósforo permite reproducir la imagen proveniente del haz de rayos catódicos, mientras que el plomo bloquea los rayos X para proteger al usuario de sus radiaciones. Fue desarrollado por William Crookes en 1875. Se emplea principalmente en monitores, televisores y osciloscopios, aunque en la actualidad se está sustituyendo paulatinamente por tecnologías como plasma, LCD, LED.



Figura 24. Tubo de rayos catódicos.

    - Pantalla de cristal líquido o **LCD** (*Liquid Crystal Display*): es una pantalla delgada y plana formada por un número de píxeles en color o monocromos colocados delante de una fuente de luz o reflectora. A menudo se utiliza en dispositivos electrónicos de pilas, ya que utiliza cantidades muy pequeñas de energía eléctrica.

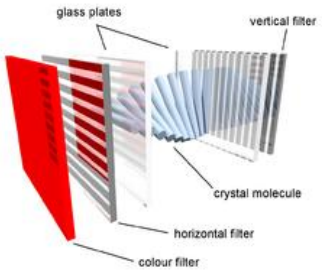


Figura 25. Subpíxel de un LCD de color.

- Pantalla de plasma o **PDP** (*Plasma Display Panel*): es un dispositivo de pantalla plana habitualmente usada en televisores de gran formato (de 37 a 80 pulgadas). Consta de muchas celdas diminutas situadas entre dos paneles de cristal que contienen una mezcla de gases nobles (neón, argón y xenón). El gas en las celdas se convierte eléctricamente en plasma, el cual provoca que una sustancia fosforescente (que no es fósforo) emita luz.

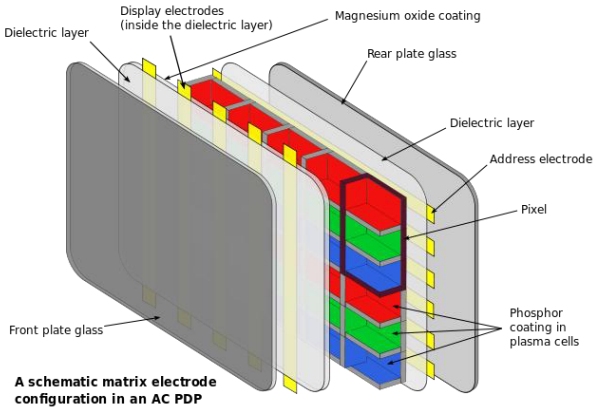


Figura 26. Composición de una pantalla de plasma.

- **TFT LCD** (*Thin Film Transistor*: transistor de películas finas): Las LCD de TFT son un tipo de LCD de matriz activa. Hay varias tecnologías de matriz activa:
  - Twisted nematic (TN), contienen elementos de cristal líquido con desenrollado y enrollado en diversos grados para permitir que la luz pase a través de ellos.
  - In-plane switching (IPS), alinea las celdas de cristal líquido en una dirección horizontal.
  - Vertical alignment (VA), el material de cristal líquido se encuentra en un estado vertical eliminando la necesidad de los transistores extras (como en el IPS).



Figura 27. TFT-LCD de 15" (Pulgadas).

- Pantalla **LED** (*Light Emitting Diode*: diodo emisor de luz), se caracterizan por estar compuesto por diodos emisores de luz o LEDs. Dichos LEDs forman píxeles, lo que permite formar caracteres, textos, imágenes y hasta vídeo, dependiendo de la complejidad de la pantalla y el dispositivo de control.
  - OLED (*Organic Light-Emitting Diode*: diodo orgánico de emisión de luz) es un tipo de diodo que se basa en una capa electroluminiscente formada por una película de componentes orgánicos que reaccionan a una determinada estimulación eléctrica, generando y emitiendo luz por sí mismos.



Figura 28. Sony XEL-1, el primer TV OLED.

  - **AMOLED** (*Active Matrix OLED*: OLED de matriz activa), consiste en un conjunto de píxeles OLED que se depositan o integran en una serie de transistores de película fina (TFT) para formar una matriz de píxeles, que se iluminan cuando han sido activados eléctricamente, controlados por los interruptores que regulan el flujo de corriente que se dirige a cada uno de los píxeles. El TFT continuamente regula la corriente que fluye por cada uno de los píxeles, para así caracterizar el píxel con el nivel de brillo que mostrará. Se utilizan bastante en dispositivos móviles.



Figura 29. Tecnología AMOLED utilizada en móviles.

        - **Super AMOLED** (*Super Active Matrix Organic Light-Emitting Diode*: Súper AMOLED), principalmente para su uso en dispositivos móviles y tablets. Una de las principales diferencias de otras tecnologías de pantallas es que la capa que detecta los toques está integrada en la pantalla, en lugar de ser superpuesta en la parte superior. Comparada con la primera generación de AMOLED, algunas de las ventajas del Super AMOLED son pantallas más brillantes, menor reflexión de la luz solar y menor consumo de energía.
- En tanto, según el estándar, un monitor puede clasificarse en: Monitor numérico, MDA, CGA, EGA, analógico, VGA, SVGA, entre otros.
- En cuanto a los colores que usan los monitores pueden ser:
    - Monitor monocromático
    - Monitor policromático.
- En cuanto a si es solamente un Periférico de salida (S) o Periférico de Entrada/Salida (E/S):
    - Monitor no táctil: S
    - Pantalla táctil (*touch screen*): E/S
    - Multitáctil (*multitouch*): E/S

4.4. Pantalla táctil

Una pantalla táctil (en inglés, touch screen), es una pantalla que mediante un toque directo sobre su superficie permite la entrada de datos y órdenes al dispositivo, y a su vez muestra los resultados introducidos previamente; actúa como periférico de entrada y salida de datos, así como emulador de datos interinos erróneos al no tocarse efectivamente. Este contacto también se puede realizar por medio de un lápiz óptico o de otras herramientas similares. Hay pantallas táctiles que pueden instalarse sobre una pantalla común, de cualquier tipo o denominación (LCD, monitores y televisores CRT, plasma, etcétera).
Las pantallas táctiles se volvieron comercialmente exitosas por su uso en dispositivos de la industria, en ordenadores públicos (como exposiciones de museos, pantallas de información, cajeros automáticos de bancos y otros) donde los teclados y los ratones no permiten una interacción satisfactoria, intuitiva, rápida, o exacta del usuario.
Se han ido haciendo populares desde la invención de la interfaz electrónica táctil, en 1971, por George Samuel Hurst, físico estadounidense. Han llegado a ser comunes en los TPV, en los cajeros automáticos y en los PDA, donde es posible usar un estilete o el dedo para manipular la interfaz gráfica de usuario (GUI) y para introducir datos. La popularidad de los teléfonos inteligentes, de los PDA, de las videoconsolas portátiles o de los navegadores de automóviles está generando la demanda y la aceptación de las pantallas táctiles.

Tipos:
Según la tecnología que usen, hay tres tipos de pantallas táctiles3 de uso habitual:

- Resistivas: Son más baratas y no les afectan el polvo ni el agua salada y, además de ser más precisas, pueden ser usadas con un puntero o con el dedo. Sin embargo, tienen hasta un 15% menos de brillo y son más gruesas, por lo que están siendo sustituidas por otras en los dispositivos móviles que precisan un tamaño y un peso ajustados y mayor brillo en la pantalla por la posibilidad de estar expuestos a la luz directa del sol.

- Capacitivas: Basadas en sensores capacitivos, consisten en una capa de aislamiento eléctrico, como el cristal, recubierto con un conductor transparente, como el ITO (tin-doped indium oxide). Como el cuerpo humano es también un conductor eléctrico, el contacto con la superficie de la pantalla genera una distorsión del campo electrostático de la pantalla, que se mide a través del cambio en la capacitancia (capacidad eléctrica). Pueden utilizarse diferentes tecnologías para determinar en qué posición de la pantalla se hizo el toque. La posición se envía al controlador, para su procesamiento. En este tipo de pantallas, la imagen tiene una mayor calidad, la respuesta es mejor, y algunas permiten el uso de varios dedos a la vez (multitouch). Sin embargo, son más caras y no se pueden usar con el puntero habitual, sino con uno especial para las pantallas capacitivas.
- Onda acústica de superficie: La tecnología de ondas de superficie utiliza ondas ultrasónicas que pasan sobre el panel de la pantalla táctil. Cuando se toca el panel, se absorbe una parte de la onda. Este cambio en las ondas ultrasónicas registra la posición del evento táctil y envía esta información al controlador, para su procesamiento. El panel de pantalla táctil de onda de superficie es el más avanzado de los tres tipos, pero puede dañarse con elementos externos.

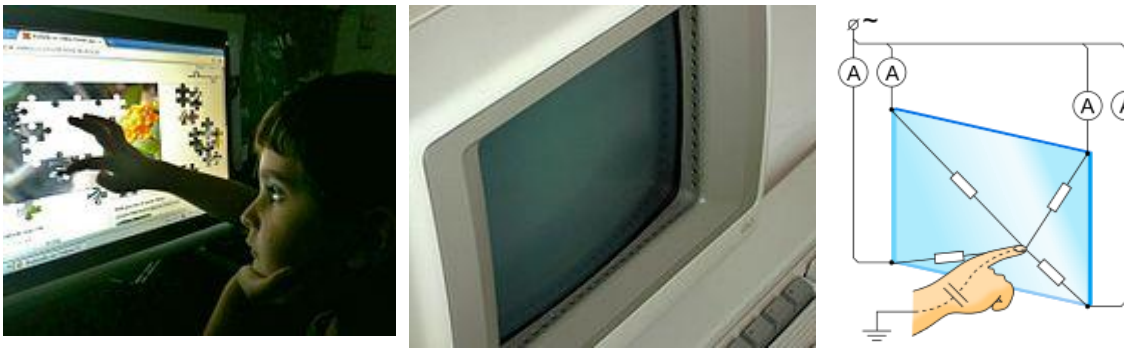


Figura 30. Pantalla táctil, La pantalla de un HP-150 (primer PC de 1983 con pantalla táctil) y La pantalla capacitiva y el principio de su funcionamiento.

Especificaciones HID

Las pantallas táctiles se encuentran definidas dentro de la especificación de dispositivos de interfaz humana (HID) para puerto USB4 como digitalizadores, junto con dispositivos como touchpad y tabletas digitalizadoras, entre otros. Las pantallas táctiles se identifican con el usage ID 04.

Las especificaciones incluyen los campos utilizados para el manejo de este tipo de dispositivos. Algunos de los más interesantes para el manejo de las pantallas táctiles son:

- Tip pressure: Representa la fuerza por un transductor, habitualmente un estilete o también un dedo.
- Barrel pressure: Fuerza que ejerce el usuario en el sensor del transductor, como por ejemplo un botón sensible a la presión en el puntero de manejo.
- In range: Indica que el transductor se encuentra en el área donde la digitalización es posible. Se representa por un bit.
- Touch: Indica si un dedo está tocando la pantalla. El sistema suele interpretarlo como un clic de botón primario.
- Untouch: Indica que el dedo ha perdido contacto con la superficie de la pantalla. Se interpreta como la acción de soltar el botón primario.

- Tap: Indica que se ha realizado un toque con el dedo en la pantalla, levantándolo rápidamente sin prolongar el contacto. Se interpreta como un evento provocado por un botón.

Sistemas operativos y software

Existe una gran variedad de programas de cómputo para el manejo de máquinas con pantallas táctiles que pueden ejecutarse en los principales sistemas operativos tales como, por ejemplo, GNU/Linux, MacOS y Windows. En los tres casos, existen versiones especiales adaptadas para su uso en tabletas y en portátiles táctiles. Android es un sistema operativo de código abierto basado en el núcleo (kernel) de Linux, inicialmente desarrollado por Android Inc. y respaldado por Google. Apple ha desarrollado su sistema iOS para iPad, y Microsoft ha incluido un desarrollo específico en las distintas versiones: desde Windows XP Tablet PC Edition hasta el Windows 8, en el cual el desarrollo táctil no es específico de una versión sino de la versión.
En otro tipo de dispositivos, como las PDA o teléfonos con pantalla táctil, también existen sistemas operativos como: PalmOS, Windows Phone, iOS, Android, BlackBerry OS, WebOS, Symbian OS, MeeGo o Maemo.
Respecto al software específico para pantallas táctiles, al igual que en el caso de otros dispositivos similares como las tabletas digitalizadoras, destacan los programas de reconocimiento de escritura manual, como Inkwell en Macintosh. En el caso de Windows XP Tablet PC Edition, el propio sistema operativo incluye reconocimiento de escritura. También son habituales los programas de dibujo, como por ejemplo Corel Painter, que pueden incluso reconocer la fuerza con la que se está pulsando sobre la pantalla o la inclinación del objeto con el que se está tocando.
Muchas aplicaciones utilizan las características táctiles de los dispositivos y de los sistemas operativos: teclados virtuales, juegos, gestores multimedia y ofimática, entre otros.

4.5. Tarjetas gráficas

Una tarjeta gráfica es una tarjeta de expansión de la placa base del ordenador que se encarga de procesar los datos provenientes de la unidad central de procesamiento (CPU) y transformarlos en información comprensible y representable en el dispositivo de salida (por ejemplo: monitor, televisor o proyector). Estas tarjetas utilizan una unidad de procesamiento gráfico o GPU, que muchas veces se usa erróneamente para referirse a la tarjeta gráfica en sí. También se le conoce como: adaptador de pantalla, adaptador de video, placa de video, tarjeta aceleradora de gráficos, tarjeta de vídeo/video.
Algunas tarjetas gráficas han ofrecido funcionalidades añadidas como sintonización de televisión, captura de vídeo, decodificación2 MPEG-2 y MPEG-4, o incluso conectores IEEE 1394 (Firewire), de mouse, lápiz óptico o joystick.
También pueden estar integradas a la placa madre o a la Unidad central de procesamiento. Casi todas las computadoras con gráficos integrados a la placa madre pueden desactivar la función de dicho chip en la BIOS, y tienen una entrada PCI, o PCI Express (PCI-E) para añadir una tarjeta de mayor rendimiento en lugar de gráficos integrados. Las principales ventajas de los gráficos integrados son el costo, la compacidad, simpleza y bajo consumo de energía. La desventaja en rendimiento de los gráficos integrados es que utilizan recursos de la CPU. Una tarjeta gráfica posee su propia memoria RAM, su propio sistema de ventilación, y reguladores

eléctricos dedicados, con todos los componentes diseñados específicamente para procesar imágenes de vídeo. Mejorar a gráficos dedicados libera trabajo de la CPU y la memoria RAM del sistema, así que no sólo el procesamiento de gráficos será más rápido, también el rendimiento general de la computadora podría mejorar.

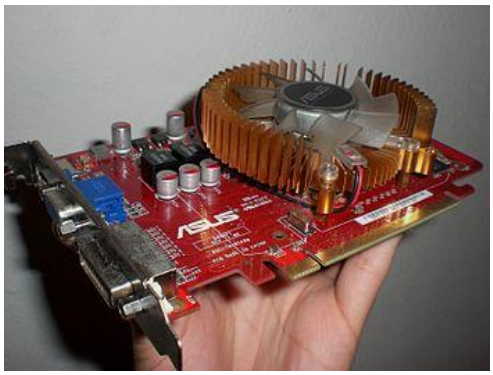


Figura 30. Tarjeta gráfica PCI-Express.

Componentes:

1. GPU

La GPU (acrónimo de «*graphics processing unit*», que significa «unidad de procesamiento gráfico») es un procesador (como la CPU) dedicado al procesamiento de gráficos; su razón de ser es aligerar la carga de trabajo del procesador central y, por ello, está optimizada para el cálculo en coma flotante, predominante en las funciones 3D. La mayor parte de la información ofrecida en la especificación de una tarjeta gráfica se refiere a las características de la GPU, pues constituye la parte más importante de la tarjeta gráfica, así como la principal determinante del rendimiento. Tres de las más importantes de dichas características son la frecuencia de reloj del núcleo, que puede oscilar entre 825 MHz en las tarjetas de gama baja, y 1600 MHz (e incluso más) en las de gama alta, el número de sombreadores y el número de tuberías (de vértices o fragmentos) encargadas de traducir una imagen 3D compuesta por vértices y líneas en una imagen 2D compuesta por píxeles.

Elementos generales de una GPU:

- Sombreadores o *shaders*: es el elemento más notable de potencia de una GPU, estos sombreadores unificados reciben el nombre de *núcleos CUDA* en el caso de NVIDIA y *procesadores de flujos de datos o nucleos en su defecto* en el caso de AMD. Son una evolución natural de los antiguos sombreadores de píxeles (encargados de la rasterización de texturas) y sombreadores de vértices (encargados de la geometría de los objetos), los cuales anteriormente actuaban de forma independiente, un GPU puede contener diferentes cantidades de nucleos segun el tipo de GPU. Los *shaders* unificados son capaces de actuar como de vértices como de píxeles según la carga de trabajo que haya en la GPU, y aparecieron en el 2007 con los chips G90 de NVIDIA (Series 8000) y los chips R600 para AMD (Series HD 2000), antigua ATi, incrementando la potencia drásticamente respecto a sus familias anteriores.
- ROP: se encargan de representar los datos procesados por la GPU en la pantalla, además también es el encargado de los filtros como el de suavizado o *antialiasing*.

2. GRAM

La memoria gráfica de acceso aleatorio (GRAM) son chips de memoria que almacenan y transportan información entre sí, no son determinantes en el rendimiento máximo de la tarjeta gráfica, pero unas especificaciones reducidas pueden limitar la potencia de la GPU.

Una parte importante de la memoria de un adaptador de vídeo es el *z-buffer*, encargado de gestionar las coordenadas de profundidad de las imágenes en los gráficos 3D, es decir, es el espacio de memoria donde se gestiona la profundidad en los gráficos.
Existen memorias gráficas de dos tipos:

- Dedicada: cuando la tarjeta gráfica o la GPU dispone exclusivamente para sí esas memorias, esta manera es la más eficiente y la que mejores resultados da.
- Compartida: cuando se utiliza memoria en detrimento de la memoria de acceso aleatorio (RAM),

Las características de memoria gráfica de una tarjeta gráfica se expresan en tres características:

- Capacidad: la capacidad de la memoria determina el número máximo de datos y texturas procesadas, una capacidad insuficiente se traduce en un retardo a espera de que se vacíen esos datos. Sin embargo es un valor muy sobrevalorado como estrategia recurrente en mercadotecnia para engañar al consumidor, tratando de hacer creer que el rendimiento de una tarjeta gráfica se mide por la capacidad de su memoria. Es una métrica importante en resoluciones grandes (Superiores a 1440p) y monitores múltiples ya que cada imagen toma mucho más espacio en la VRAM.
- Interfaz de memoria: también denominado bus de datos, es la multiplicación resultante del ancho de bits de cada chip por su número de unidades. Es una característica importante y determinante, junto a la frecuencia de la memoria, a la cantidad de datos que puede transferir en un tiempo determinado, denominado ancho de banda. Una analogía al ancho de banda se podría asociar al ancho de una autopista o carriles y al número de vehículos que podrían circular a la vez. La interfaz de memoria se mide en bits.
- Frecuencia de memoria: es la frecuencia a la que las memorias pueden transportar los datos procesados, por lo que es complemento a la interfaz de memoria para determinar el ancho de banda total de datos en un tiempo determinado. Continuando la analogía de la circulación de los vehículos de la autopista, la frecuencia de memoria se traduciría en la velocidad máxima de circulación de los vehículos, dando resultado a un mayor transporte de mercancía en un mismo periodo de tiempo. La frecuencia de las memorias se mide en hercios (su frecuencia efectiva) y se van diseñando tecnologías con más frecuencias, se destacan las adjuntas en la siguiente tabla:

Tecnología	Frecuencia efectiva (MHz)	Ancho de banda (GB/s)
GDDR	166 - 950	1,2 - 30,4
GDDR2	533 - 1000	8,5 - 16
GDDR3	700 - 1700	5,6 - 54,4
GDDR4	1600 - 1800	64 - 86,4
GDDR5	3200 - 7000	24 - 448
HBM	500	512

- **Ancho de banda** (AdB): es la tasa de datos que pueden transportarse en una unidad de tiempo. Un ancho de banda insuficiente se traduce en un importante limitador de potencia de la GPU. Habitualmente se mide en *gigabytes* por segundo (GB/s).

3. RAMDAC

El convertidor digital-analógico de memoria de acceso aleatorio (RAMDAC) es un conversor de señal digital a señal analógica de memoria RAM. Se encarga de transformar las señales digitales producidas en el ordenador en una señal analógica que sea interpretable por el monitor. Según el número de bits que maneje a la vez y la velocidad con que lo haga, el conversor será capaz de dar soporte a diferentes velocidades de refresco del monitor (se recomienda trabajar a partir de 75 Hz, y nunca inferior a 60). Dada la creciente popularidad de los monitores de señal digital, el RAMDAC está quedando obsoleto, puesto que no es necesaria la conversión analógica si bien es cierto que muchos conservan conexión VGA por compatibilidad.

Salidas:

Los sistemas de conexión más habituales entre la tarjeta gráfica y el dispositivo visualizador (p. ej. monitor o televisor) son:

- VGA: el *video graphics array* (VGA) o *super video graphics array* (SVGA o Súper VGA) fue el estándar analógico de los años 1990; diseñado para dispositivos con tubo de rayos catódicos (CRT); sufre de ruido eléctrico y distorsión por la conversión de digital a analógico y el error de muestreo al evaluar los píxeles a enviar al monitor. Se conecta mediante 15 pines con el conector D-sub: DE-15. Su utilización continúa muy extendida, aunque claramente muestra una reducción frente al DVI.
- DVI: *digital visual interface* (DVI) o interfaz visual digital es sustituta de la anterior, pero digital; fue diseñada para obtener la máxima calidad de visualización en las pantallas digitales o proyectores. Se conecta mediante pines. Evita la distorsión y el ruido al corresponder directamente un píxel a representar con uno del monitor en la resolución nativa del mismo. Cada vez más adoptado, aunque compite con el HDMI, pues el DVI no es capaz de transmitir audio.
- HDMI: la interfaz multimedia de alta definición o *high-definition multimedia interface* (HDMI) es una tecnología propietaria transmisora de audio y vídeo digital de alta definición cifrado sin compresión, en un mismo cable. Se conecta mediante patillas de contacto. Fue ideado inicialmente para televisores, y no para monitores, por eso no apaga la pantalla cuando deja de recibir señal y debe apagarse manualmente en caso de monitores.
- DisplayPort: puerto para tarjetas gráficas creado por VESA y rival del HDMI, transfiere vídeo a alta resolución y audio. Sus ventajas son que está libre de patentes, y por ende de regalías para incorporarlo a los aparatos, también dispone de unas pestañas para anclar el conector impidiendo que se desconecte el cable accidentalmente. Cada vez más tarjetas gráficas van adoptando este sistema, aunque sigue siendo su uso minoritario, existe una versión reducida de dicho conector llamada Mini DisplayPort, muy usada para tarjetas gráficas con multitud de salidas simultáneas, como pueden ser 5.

Otras no tan extendidas: por uso minoritario, por no ser implementadas o por ser obsoletas; son:

- S-Video (*separate video*, vídeo separado): implementado sobre todo en tarjetas con sintonizador TV o chips con soporte de vídeo NTSC/PAL, simplemente se está quedando obsoleto.
- Vídeo Compuesto: analógico de muy baja resolución mediante conector RCA (Radio Corporation of America). Completamente en desuso para tarjetas gráficas, aunque sigue siendo usado para TV.
- Vídeo por componentes: sistema analógico de transmisión de vídeo de alta definición, utilizado también para proyectores; de calidad comparable a la de SVGA, dispone de tres clavijas (*Y*, *Cb* y *Cr*). Anteriormente usado en las PC y estaciones de trabajo de gama alta, ha quedado relegado a TV y videoconsolas.
- DA-15 con conector RGB usado mayoritariamente en los antiguos Macintosh de Apple. Ya no se utiliza.
- Digital TTL con conector DE-9: usado por las primitivas tarjetas de IBM (MDA, CGA y variantes, EGA y muy contadas VGA). Completamente obsoleto.



Figura 31. Salidas de una tarjeta gráfica: HDMI, VGA, DVI-I y S-Video de una tarjeta gráfica.

Interfaces con la placa base

Los sistemas de conexión entre la tarjeta gráfica y la placa base han sido principalmente (en orden cronológico):

1. Ranura MSX: bus de 8 bits usado en los equipos MSX.
2. ISA: arquitectura de bus de 16 bits a 8 MHz, dominante durante los años 1980; fue creada en 1981 para los IBM PC.
3. Zorro II: usado en los Commodore Amiga 2000 y Commodore Amiga 1500.
4. Zorro III: usado en los Commodore Amiga 3000 y Commodore Amiga 4000
5. NuBus: usado en el ordenador Apple Macintosh.
6. *Processor Direct Slot*: usado en los Apple Macintosh.
7. MCA: en 1987, intento de sustitución de ISA por IBM. Disponía de 32 bits y una velocidad de 10 MHz, pero era incompatible con los anteriores.
8. EISA: en 1988, respuesta de la competencia de IBM; de 32 bits, 8,33 MHz y compatible con las placas anteriores.
9. VESA: extensión de ISA que solucionaba la restricción de los 16 bits, duplicando el tamaño de bus y con una velocidad de 33 MHz.
10. PCI: bus que desplazó a los anteriores, a partir de 1993; con un tamaño de 32 bits y una velocidad de 33 MHz, permitía una configuración dinámica de los dispositivos conectados sin necesidad de ajustar manualmente los *jumpers*. PCI-X fue una versión que aumentó el tamaño del bus hasta 64 bits y aumentó su velocidad hasta los 133 MHz.

11. AGP: bus dedicado, de 32 bits como PCI; en 1997, la versión inicial incrementaba la velocidad hasta los 66 MHz.
12. PCI-Express (PCIe): desde 2004, es la interfaz serie que empezó a competir contra AGP; en 2006, llegó a doblar el ancho de banda del AGP. Sufre de constantes revisiones multiplicando su ancho de banda, ya existe la versión 3.0. No debe confundirse con PCI-X, versión de PCI.

Dispositivos refrigerantes

Debido a las cargas de trabajo a las que son sometidas, las tarjetas gráficas alcanzan temperaturas muy altas. Si no es tenido en cuenta, el calor generado puede hacer fallar, bloquear o incluso averiar el dispositivo. Para evitarlo, se incorporan dispositivos refrigerantes que eliminen el calor excesivo de la tarjeta.

Se distinguen dos tipos:

- Disipador: dispositivo pasivo (sin partes móviles y, por tanto, silencioso); compuesto de un metal muy conductor del calor, extrae este de la tarjeta. Su eficiencia va en función de la estructura y la superficie total, por lo que a mayor demanda de refrigeración, mayor debe ser la superficie del disipador.
- Ventilador: dispositivo activo (con partes móviles); aleja el calor emanado de la tarjeta al mover el aire cercano. Es menos eficiente que un disipador, siempre que nos refiramos al ventilador solo, y produce ruido al tener partes móviles.

Aunque diferentes, ambos tipos de dispositivo son compatibles entre sí y suelen ser montados juntos en las tarjetas gráficas; un disipador sobre la GPU extrae el calor, y un ventilador sobre él aleja el aire caliente del conjunto.

La refrigeración líquida es una técnica de enfriamiento utilizando agua en vez de disipadores de calor y ventiladores (dentro del chasis), logrando así excelentes resultados en cuanto a temperaturas, y con enormes posibilidades en overclocking. Se suele realizar con circuitos de agua estancos.

El agua, y cualquier líquido refrigerante, tienen mayor capacidad térmica que el aire. A partir de este principio, la idea es extraer el calor generado por los componentes de ordenador usando como medio el agua, enfriarla una vez fuera del gabinete y luego reintroducirla.



Figura 32. Conjunto de disipador y ventilador.

Alimentación

Hasta ahora la alimentación eléctrica de las tarjetas gráficas no había supuesto un gran problema, sin embargo, la tendencia actual de las nuevas tarjetas es consumir cada vez más energía. Aunque las fuentes de alimentación son cada día más potentes, la insuficiencia

energética se encuentra en la que puede proporcionar el puerto PCIe que sólo es capaz de aportar una potencia por sí solo de 75 W. Por este motivo, las tarjetas gráficas con un consumo superior al que puede suministrar la fuente, PCIe incluyen un conector PCIe que permite una conexión directa entre la fuente de alimentación y la tarjeta, sin tener que pasar por la placa base, y, por tanto, por el puerto PCIe.
Aun así, se pronostica que no dentro de mucho tiempo las tarjetas gráficas podrían necesitar una fuente de alimentación propia, convirtiéndose dicho conjunto en dispositivos externos.

API para gráficos
A nivel de programador, trabajar con una tarjeta gráfica es complicado; por ello, surgieron interfaces de programación de aplicaciones (API) que abstraen la complejidad y diversidad de las tarjetas gráficas. Los dos más importantes son:

- Direct3D: lanzada por Microsoft en 1996, forma parte de la librería DirectX. Funciona sólo para Windows, ya que es privativa. Utilizado por la mayoría de los videojuegos comercializados para Windows. Actualmente van por la versión 12.
- OpenGL: creada por Silicon Graphics a principios de los años 1990; es gratuita, libre y multiplataforma. Utilizada principalmente en aplicaciones de CAD, realidad virtual o simulación de vuelo. Actualmente está disponible la versión 4.3.

OpenGL está siendo desplazada del mercado de los videojuegos por Direct3D, aunque haya sufrido muchas mejoras en los últimos tiempos.

5. Elementos de almacenamiento

Debido a la cantidad de información manejada actualmente, los periféricos de almacenamiento se han convertido casi tan importantes como el mismísimo computador. Estos periféricos se encargan de guardar o salvar los datos de los que hace uso la CPU para que ésta pueda emplearlos una vez que han sido eliminados de la memoria principal, ya que ésta se borra cada vez que se apaga el ordenador. Pueden ser internos, como un disco duro, o extraíbles, como un CD. Los más comunes son: Disco duro, Memorias de estado sólido (SSD), CD, DVD; Blu-Ray, Memoria Flash, Cintas magnéticas.
A continuación se describe con más detalle alguno de ellos.

5.1. Disco Duro

Los discos duros pertenecen a la llamada memoria secundaria o almacenamiento secundario. Un disco duro (en inglés *hard disk*, abreviado con frecuencia *HD* o *HDD*) es un dispositivo que permite almacenar una gran cantidad de datos de forma rápida y segura. Los datos permanecen grabados en su interior permanentemente y pueden ser leídos, borrados y grabados infinidad de veces.
El tamaño de las primeras unidades de disco duro era de 10 ó 20 Megabytes (MB), existiendo hoy en día discos de varios TeraBytes. Estas capacidades tan enormes se deben principalmente a la elevada densidad de grabación.

a) Características.
Los discos duros generalmente utilizan un sistema de grabación magnética digital. En este tipo de disco se encuentra, dentro de la carcasa, una serie de platos metálicos apilados girando a gran velocidad. Los platos están montados sobre un eje central (*spindle*) que gira siempre en el mismo sentido. Para leer y escribir datos en estos platos se usan las cabezas de

lectura/escritura que mediante un proceso electromagnético codifican/ decodifican la información que han de leer o escribir.
La cabeza de lectura/escritura en un disco duro está muy cerca de la superficie del plato. Normalmente, cada plato dispone de dos cabezas de lectura/escritura, una para cada cara.
Los platos de aluminio están recubiertos de una capa de material con propiedades magnéticas en ambas caras. Inicialmente se utilizaba óxido férrico, pero actualmente se suele emplear cobalto por sus mejores cualidades. Estos platos tienen tamaños típicos dependiendo de cada disco duro. La práctica totalidad de ellos incorporan platos de 3,5 pulgadas. Existe un tercer tamaño de 2,5 pulgadas, empleado habitualmente en discos duros para ordenadores portátiles, que se le suele denominar *Microdrive*.
Las principales características son las siguientes:

- Capacidad de Almacenamiento: cantidad de información que se puede grabar o almacenar en un disco duro.
- Velocidad de Rotación: es la velocidad a la que gira el disco duro, más exactamente, la velocidad a la que giran el/los platos (se detalla más adelante) del disco, que es donde se almacenan magnéticamente los datos. La regla es: a mayor velocidad de rotación, más alta será la transferencia de datos, pero mayor será el ruido y mayor el calor generado por el disco duro. Se mide en RPM (Revoluciones Por Minuto).
- Memoria Caché (tamaño del buffer): es una memoria que va incluida en la controladora interna del disco duro, de modo que todos los datos que se leen y escriben a disco duro se almacenan primeramente en el buffer.
- Tasa de transferencia: indica la cantidad de datos que un disco duro puede leer o escribir en la parte más externa del disco o plato en un período de un segundo. Normalmente se mide en Mbits/s.

b) Componentes Físicos:

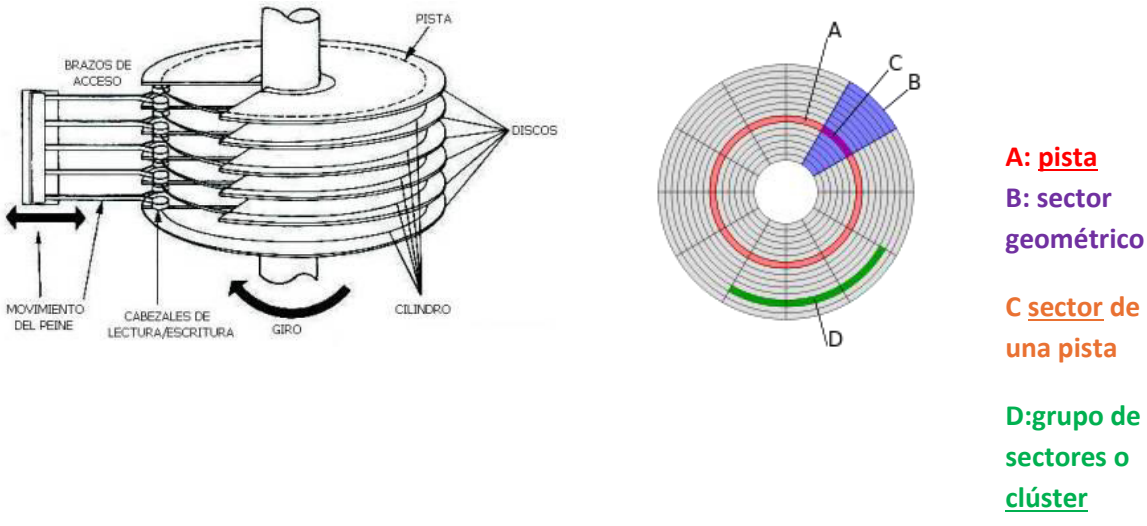


Figura 33. Estructura de un disco duro.

- Los discos (*platters*)

Están elaborados de compuestos de vidrio, cerámica o aluminio pulidos y revestidos por ambos lados con una capa muy delgada de una aleación metálica. Los discos están unidos a un eje y un motor que los hace guiar a una velocidad constante entre las 3600 y 7200 RPM. Convencionalmente los discos duros están compuestos por varios platos, es decir varios discos de material magnético montados sobre un eje central. Estos discos, normalmente, tienen dos

caras que pueden usarse para el almacenamiento de datos, si bien suele reservarse una para almacenar información de control. Cada uno de los lados de un disco es una cara.
Los datos se graban en forma de flujos magnéticos escritos en círculos, alrededor del anillo central del disco. Cada uno de los círculos concéntricos forma una pista y cada pista está dividida en un número igual de segmentos, llamados sectores. Un sector es la mínima unidad de información direccionable en un disco duro. En la mayoría de los discos duros, un sector tiene un tamaño de 512 bytes. Cada pista normalmente está dividida en una cantidad de sectores que oscila entre 100 y 300. La cabeza de lectura/escritura se mueve del borde exterior del disco hacia el anillo central, deteniéndose sobre la pista que contiene la información que necesita el ordenador. Una vez en la posición adecuada, la cabeza espera a que el sector correcto de la pista se encuentre debajo de ella, y a continuación lee o escribe datos a medida que éste pasa por debajo.
Otro concepto importante es la densidad de grabación. Ésta se mide en bits por pulgada (BPI, *Bits Per Inch)* y la densidad de pistas en cada plato se mide en pistas por pulgada (TPI, *Tracks Per Inch).* Multiplicando ambos factores se obtiene la densidad de área, que generalmente se calcula en bits por pulgada cuadrada (BPSI, *Bits Per Square Inch).*
La densidad hace referencia a la separación que existe en los bits de información almacenados en el disco duro. Mientras mayor sea la densidad de grabación, mayor cantidad de datos se podrán almacenar en una misma superficie. Existen dos formas de conseguir una mayor densidad de datos:
- Almacenando los bits más juntos unos de otros.
- Haciendo que la distancia entre pistas sea menor, con lo que se consiguen introducir más pistas en un mismo plato aumentando así la capacidad de almacenamiento.

- Las Cabezas (*Heads*)

Están ensambladas en pila y son las responsables de la lectura y la escritura de los datos en los discos. La mayoría de los discos duros incluyen una cabeza Lectura/Escritura a cada lado del disco, sin embargo, algunos discos de alto rendimiento tienen dos o más cabezas sobre cada superficie, de manera que cada cabeza atiende la mitad del disco reduciendo la distancia del desplazamiento radial.
Las cabezas de Lectura/Escritura no tocan el disco cuando éste gira a toda velocidad; por el contrario, flotan sobre una capa de aire extremadamente delgada (10 millonésima de pulgada).
Esto reduce el desgaste en la superficie del disco durante la operación normal. Cualquier polvo o impureza en el aire puede dañar suavemente las cabezas o el medio. Su funcionamiento consiste en una bobina de hilo que se acciona según el campo magnético que detecte sobre el soporte magnético, produciendo una pequeña corriente que es detectada y amplificada por la electrónica de la unidad de disco. Las cabezas convierten, por tanto, señales eléctricas en señales magnéticas y viceversa, según la operación a realizar sea de escritura o de lectura.
Todos los discos duros actuales incorporan una memoria caché *(buffer)* para agilizar las operaciones de lectura y escritura. Generalmente esta caché suele ser de 32 ó 64 MB.
Las cabezas son un factor fundamental en el rendimiento de un disco duro y en la densidad de grabación.
Las primeras cabezas que se usaron en discos duros eran de ferrita, posteriormente se emplearon cabezas de tipo GAP (de su denominación *Metal-In-Gap)* y de película fina (TF, *Thin Film).* Actualmente todas las cabezas son magnetoresistivas (magnética por la parte de escritura y resistiva por la de lectura), con lo que se soportan densidades de grabación hasta 10.000 veces mayores que las de las primeras unidades.
La última generación de cabezas magnetoresistivas aparecidas en el mercado se denomina GMR *(Giant Magneto Resistive*), lo que permite fabricar discos duros cada vez más pequeños y de mayor capacidad, al estar la información más concentrada en el mismo espacio y ser las cabezas más pequeñas y precisas. Esta tecnología fue desarrollada por IBM.
La pista de aparcamiento de las cabezas se llama LAND ZONE.

- Eje

Es la parte del disco duro que actúa como soporte, sobre el cual están montados y giran los platos del disco.

- Pistas

Un disco está dividido en delgados círculos concéntricos llamados pistas. Las cabezas se mueven entre la pista más externa ó pista cero a la más interna. Es la trayectoria circular trazada a través de la superficie circular del plato de un disco por la cabeza de lectura / escritura. Cada pista está formada por uno o más Cluster.

- Sectores

Un byte es la unidad útil más pequeña en términos de memoria. Los HD almacenan los datos en pedazos gruesos llamados sectores. La mayoría de los HD usan sectores de 512 bytes. La controladora del HD determina el tamaño de un sector en el momento en que el disco es formateado. Algunos modelos de HD le permiten especificar el tamaño de un sector. Cada pista del disco está dividida en 1 ó 2 sectores dado que las exteriores son más grandes que las interiores, las exteriores contienen más sectores.

- Cilindros

El par de pistas en lados opuestos del disco se llama cilindro. Se puede ver como el conjunto de varias pistas. Son todas las circunferencias que están alineadas verticalmente (una de cada cara). Si el HD contiene múltiples discos (sean n), un cilindro incluye todos los pares de pistas directamente una encima de otra (2n pistas). Los HD normalmente tienen una cabeza a cada lado del disco. Dado que las cabezas de Lectura/Escritura están alineadas unas con otras, la controladora puede escribir en todas las pistas del cilindro sin mover el rotor. Como resultado, los HD de múltiples discos se desempeñan levemente más rápido que los HD de un solo disco.

- Tarjeta controladora:

Para controlar el funcionamiento del motor y gestionar la forma en que se graba y lee la información, todos los discos actuales cuentan con una tarjeta controladora integrada en la propia unidad. Dicha controladora se suele alimentar a 5 voltios, mientras que el motor lo hace a 12 voltios. De ahí que el conector de alimentación tenga cuatro cables: uno de 5 y otro de 12 voltios, con sus respectivas masas.

Entre las principales funciones que realiza la tarjeta controladora están las siguientes:

- El control del motor de giro del disco duro, incluyendo el control de la velocidad de giro.
- El control del brazo actuador para la ubicación en la pista correcta.
- El control del tiempo necesario para realizar las operaciones de lectura y escritura.
- El control de la gestión de la energía (en leer y amplificar las señales de bajo voltaje procedentes de las cabezas del disco duro).

Finalmente, la tarjeta controladora incorpora un firmware o software de control. Esto no son más que una serie de rutinas encargadas de ser procesadas por el microprocesador que generalmente todos los discos duros actuales incorporan.

Estas rutinas suelen estar grabadas en una memoria ROM en la propia controladora del disco duro. Muchos discos duros ya incorporan para estas funciones una memoria ROM de tipo flash, que permite su actualización mediante software.

c) Procesos de codificación de la información:

Según los elementos citados, en un disco duro los datos se grabarán en forma de flujos magnéticos escritos en círculos, alrededor del anillo central del disco. Cada uno de los círculos concéntricos forma una pista y cada pista está dividida en un número igual de segmentos, llamados sectores.

La información digital se almacena en forma de cadenas de ceros y unos. En un disco duro, esos ceros y unos que componen la información se almacenan codificados en forma de pulsos magnéticos.
Tipos de codificaciones:

- FM *(Frecuency Modulation* – Modulación de Frecuencia): es la codificación más sencilla. Consiste en la grabación de un cambio de flujo para cada uno, y en omitir el cambio de flujo para cada cero. Este procedimiento se puede realizar con una electrónica de control relativamente simple, pero tiene el inconveniente de que cada bit de datos consume dos cambios de flujo, limitando mucho la capacidad del disco.
- MFM *(Modify Frecuency Modulation* – Modulación de Frecuencia Modificada): en comparación con FM, no necesitaba almacenar tantos cambios por bit, lo que al final redundaba en una mayor capacidad de almacenamiento de información en el disco. Aunque MFM ya se abandonó en los discos duros al usarse métodos mucho más eficientes, en los disquetes éste es el método de codificación que se sigue utilizando.
- RLL (*Run Length Limited* – Longitud de recorrido limitado): este método trabaja con grupos de bits, en lugar de hacer la codificación para bits individuales, operando con dos parámetros: la longitud y el límite del recorrido, de ahí su denominación. La longitud del recorrido es el mínimo espacio entre dos cambios de fase, mientras que el límite del recorrido es la máxima longitud entre dos cambios de fase. Gran parte de los discos duros actuales utilizan alguna variedad de codificación RLL.

d) Interfaces:

La interfaz de un disco duro comprende la forma que usa para relacionarse con los demás elementos que componen un ordenador. Hace también referencia, por tanto, al tipo de conexión que utiliza el disco duro. La evolución ha sido la siguiente:

- ESDI (*EnhancedSmallDeviceInterface*):
  A mediados de los 80's, apareció este estándar para solventar algunos de los problemas que habían surgido con el anterior. Trasladaba algunas funciones de control ya al propio disco duro, incrementando así la tasa de transferencia de datos y produciéndose un ligero aumento de prestaciones en general.
- IDE (*Integrated Drive Electronics*):
  Creado por Western Digital. Su característica más representativa era la implementación de la controladora en el propio disco duro, de ahí su denominación. Igualmente se eliminó la necesidad de disponer de dos cables separados para control y datos, bastando con un cable de 40 hilos desde el bus al disco duro.
  IDE permite transferencias de 4 Megas por segundo, aunque dispone de varios métodos para realizar estos movimientos de datos. La interfaz IDE supuso la simplificación en el proceso de instalación y configuración de discos duros, y estuvo durante un tiempo a la altura de las exigencias del mercado.
  IDE/ATA son unas siglas que responden a *Integrated Drive Electronics /AT Attachment*, y es el estándar en los discos duros actuales, en los cuales, la mayor parte de la circuitería de control del disco duro se integra en la propia unidad. ATA es un conjunto de reglas que deben cumplir todos los dispositivos IDE.
  En la práctica, tanto IDE como ATA viene a hacer referencia al mismo estándar y al mismo concepto. La especificación IDE/ATA es la base original sobre la que se han desarrollado las especificaciones posteriores. La interfaz IDE/ATA original, utilizada en los originales IBM AT, era una interfaz de 16 bits.
  El ATA es una interfaz estandarizada por la ANSI (*American Nacional Standard Institute*) cuyas características principales son las siguientes:

-Dos discos duros. La especificación ATA original sólo da soporte a dos discos duros ambos montados sobre un mismo canal, y configurados como maestro (master) y esclavo (slave) respectivamente.
-Modos de transferencia. El ATA soporta los modos de transferencia PIO (Entrada/Salida Programada) en modo O, 1 y 2. PIO es entrada/Salida programable. Es el componente encargado de ejecutar las instrucciones dirigidas.
-Modos DMA. También soporta la transferencia en modo DMA 0, 1 y 2 y DMA multiword en modo 0. DMA es Acceso Directo a Memoria. Es el componente integrado en un periférico que libera al procesador en la tarea de transferir datos entre dispositivos y memoria. Debido a esto no requiere atención del procesador (a diferencia de PIO).

Inicialmente, el estándar ATA fue desarrollado para ser empleado con discos duros. Pero, posteriormente, se desarrollaron otros dispositivos que hacían uso de la misma interfaz, por lo que surgieron nuevos estándares para darles soporte. Entre ellos se puede destacar:

-ATA/100, ATA/133. Se denominan también UltraDMN100 y UltraDMN133, respectivamente. Hacen referencia a la máxima tasa de transferencia de datos teórica que pueden soportar los discos duros conectados a esa placa base, que viene a ser de 100 y 133 MB/seg respectivamente, de ahí las denominaciones. En función de que un chipset sea más ó menos reciente soporta un modo de transferencia u otro.
-UltraDMA/33 (Ultra ATA-33) totalmente compatible con todos los estándares anteriores.
Es una extensión de la interfaz de disco duro ATA que, teóricamente, permite velocidades de transferencia de datos en ráfagas de 33,3 Megabytes (MB) por segundo. Esto duplica la velocidad del estándar ATA-2/ATA-3 de 16,6 MB por segundo. El uso de Ultra DMA puede provocar un incremento en el rendimiento del 40% sobre el estándar ATA-2/ATA-3. La mejora del rendimiento es el resultado de la mejora en la velocidad de datos de la propia unidad.
-ATAPI (*AT AttachmentPacket Interface*) es un estándar que permite implementar dispositivos con características propias de SCSI sobre interfaces ATA, tales como lectores de CD-ROM y unidades de cinta. Para el correcto funcionamiento de los dispositivos ATAPI se requiere el uso de drivers específicos.

Lo que determina poder usar PIO o DMA es la BIOS. La BIOS juega un papel muy importante en el funcionamiento de un disco duro. Proporciona, además, los servicios de acceso al disco duro mediante una interfaz común para acceder a los discos duros, sean del tipo que sean, lo que ofrece una gran comodidad y sencillez para las aplicaciones. Por otra parte, la BIOS permite la detección y configuración de los discos duros.
Cuando una aplicación accede al disco duro, generalmente lo hace a través de la BIOS, mediante la interrupción 13h (19 en notación decimal). Esta interrupción es estándar y se ha venido utilizando en el PC desde hace mucho tiempo para el acceso a disco duro. El uso de la interrupción 13h necesita que, previamente, se le proporcione determinada información, como la cabeza, el cilindro y el sector exactos a los que se va a acceder para que las rutinas de acceso puedan realizar dicho acceso.
La rutina original de la interrupción soportaba discos de 504 MB o más grandes hasta un límite de 8 GiB utilizando el modo lógico de acceso CHS. Para romper esta barrera, se crearon las extensiones de la INT 13h que utilizan el modo LBA (direcciones de 64 bits) para acceder al disco, soportando así hasta 8 ZiB de espacio (También se puede usar modos LBA de 32 bits ó 48 bits que soportan 2 TiB ó 128 PiB respectivamente).

- EIDE (*Enhanced IDE*):
  Esta interfaz, propuesta también por Western Digital, logra una mejora de flexibilidad y prestaciones. Para empezar, aumenta su capacidad, hasta 8,4 Gigas, y la tasa de transferencia empieza a subir a partir de los 10 Megas por segundo. Permite instalar 4 unidades al mismo tiempo (frente a dos del interfaz IDE). A nivel externo, no existen

prácticamente diferencias con el anterior IDE, en todo caso un menor tamaño o más bien una superior integración de un mayor número de componentes en el mismo espacio.

- SCSI (*Small Computer System Interface*):
  Los discos duros tipo SCSI se han utilizado, sobre todo, debido a que se pueden conectar siete dispositivos por controladora y que la velocidad de transferencia es considerablemente más alta que la de los discos duros IDE. La evolución de la interfaz SCSI ha sido la siguiente:
  - SCSI-1. Corresponde al bus SCSI original. Se implementó en 1986, permitía la conexión de un máximo de 7 dispositivos de 8 bits con una velocidad máxima de 5 Mb/s.
  - SCSI-2. Esta especificación fue una ampliación del SCSI-1. Se terminó de desarrollar en 1992 y es compatible con los dispositivos anteriores. Permite conectar dispositivos tales como unidades de CD-ROM escáneres. Asimismo, se incluyen dos nuevas especificaciones: Fast SCSJ y Wide SCSI que permitían duplicar la velocidad y el número de dispositivos, respectivamente. Las unidades disponen de un bus de 16 bits y se diseñaron nuevos conectores más compactos.
  - SCSI-3. Entre sus nuevas mejoras está el aumento del número de dispositivos hasta 32. También da soporte a canales de fibra y aumenta las prestaciones con la especificación "Ultra".

  Con una interfaz SCSI se pueden conectar hasta 7 dispositivos, o bien 15 si es Wide SCSI. Los dispositivos SCSI pueden ser discos duros, lectores de CD-ROM o DVD-ROM, escáneres, etc.
  En la siguiente tabla se puede ver un resumen de los distintos SCSI:

Norma SCSI	Ancho Bus	Megas/segundo	Dispositivos
SCSI-1	8 bits	3 Megas/s	8
SCSI-2	8 bits	5 Megas/s	8
Fast SCSI-2	8 bits	10 Megas/s	8
Fast/Wide SCSI-2	16 bits	20 Megas/s	16
Ultra SCSI	8/16 bits	20/40 Megas/s	8
Ultra2 SCSI LVD	8/16 bits	40/80 Megas/s	16
Ultra3 SCSI	16 bits	160 Megas/s	16
Ultra 320 SCSI	16 bits	320 Megas/s	16
Ultra 640 SCSI	16 bits	640 Megas/s	16

  Comparando SCSI con IDE se pueden obtener las siguientes diferencias:
  - Los discos duros IDE han representado tradicionalmente al ordenador personal medio, siendo los discos SCSI usados desde el principio por usuarios avanzados en equipos profesionales. De hecho, todos los ordenadores Macintosh de Apple disponen de discos duros SCSI desde el primer momento.
  - Ambos tipos (IDE y SCSI) han sufrido un profundo desarrollo desde sus orígenes pero es preciso reconocer que los del tipo EIDE (*Enhanced* IDE) han ido aproximándose a los SCSI en cuanto a prestaciones se refiere y la diferencia de prestaciones entre ambos tipos de discos actualmente no es sustancial, aproximándose mucho entre sí.
  - Normalmente, los puertos EIDE aparecen integrados en la placa base. Los discos SCSI, por el contrario, precisan una controladora SCSI adicional, aunque algunas placas base de gama alta la integran.

- En circunstancias normales, un solo disco SCSI no aumenta significativamente las prestaciones frente a un EIDE. Sin embargo, el rendimiento se dispara cuando se conectan varios dispositivos SCSI accediendo simultáneamente al mismo bus común (no depende de bus del sistema). Por el contrario en cada puerto EIDE, los dos dispositivos conectados tienen que turnarse en la utilización del bus.
- SCSI es más caro que IDE y algo más complejo de configurar.

- SATA (*Serial* ATA)
  Es una interfaz de transferencia de datos en serie que permite la conexión de dispositivos de almacenamiento y ópticos. Fue diseñada para sustituir a las anteriores ATA y EIDE.
  La primera versión ofrecía transferencias de datos de 150MB/s (frecuencia 1500Mhz), mejorada en las sucesivas versiones de SATA v2.0 con 300MB/s (frecuencia 3000Mhz) y SATA v3.0 (frecuencia 6.000Mhz) con 600MB/s.
  Ofrece nuevas características: reducción del número de cables (tan solo 7), mejor y más eficiente transferencia de datos y conexión/desconexión rápida. Con cada nueva versión han ido aumentando las prestaciones, mejorando la calidad del servicio de video *streaming* y las interrupciones de alta prioridad. Por otra parte, las altas tasas de transferencia de datos requieren un mayor consumo de potencia.
  SATA es una arquitectura "punto a punto", es decir, la conexión entre puerto y dispositivo es directa. Cada dispositivo se conecta directamente a un controlador SATA y, por tanto, disfruta de todo el ancho de banda de la conexión.
  Una variante de esta interfaz (año 2004) es el eSata (*external* SATA) en directa competencia con USB y FireWire. Actualmente, la mayoría de las placas bases han empezado a incluir conectores eSATA. También es posible usar adaptadores de bus o tarjetas PC-Card y CardBus para portátiles que aún no integran el conector.
- SAS (*Serial Attached SCSI*):
  Es una interfaz de transferencia de datos en serie, sucesora del SCSI paralelo, aunque sigue utilizando comandos SCSI para interaccionar con los dispositivos SAS. Proporciona una arquitectura punto a punto más simple y robusta que su antecesora paralela, un mejor rendimiento (de 3Gb a 12 Gb) y mayor escalabilidad.
  Una de las principales características es el aumento de la velocidad de transferencia al aumentar el número de dispositivos conectados, es decir, puede gestionar una tasa de transferencia constante para cada dispositivo conectado y elimina la limitación de 16 dispositivos existente en SCSI.
  Debido a la transferencia de datos en serie, son enviados, uno tras otro (en el orden correcto y en el momento esperado). Con el cableado punto a punto, con cada dispositivo se asegura un canal dedicado para cada uno de ellos (reduciendo la latencia) y, al no compartir bus, la gestión es mucho más sencilla.
  El conector SAS es el mismo que en la interfaz SATA y permite utilizar estos discos duros, para aplicaciones con menos necesidad de velocidad, ahorrando costos. Por lo tanto, los discos SATA pueden ser utilizados por controladoras SAS pero no a la inversa, una controladora SATA no reconoce discos SAS.
  En la tabla adjunta se pueden apreciar las principales diferencias con su antecesor:

	SCSI Paralelo	SAS
Arquitectura	Paralela	Serie
Rendimiento	320Mbs (Ultra SCSI)	3GBs -> 12GBs
Escalabilidad	15 dispositivos	16.000 dispositivos
Compatibilidad	Incompatible con otras interfaces	Compatible con SATA
Máxima longitud cable	12 metros total para todos los dispositivos	8m para cada dispositivo
Conexión en caliente	No	Si

e) Medidas que describen el rendimiento de un disco duro:

- Tiempo de búsqueda: tiempo que tardan las cabezas de lectura/escritura en moverse desde su posición actual hasta la pista donde está localizada la información deseada. Este tiempo varía en cada búsqueda. Actualmente, los discos duros tienen tiempos de búsqueda pista a pista tan cortos como 2 milisegundos, tiempos promedios de búsqueda menores a 10 milisegundos y tiempo máximo de búsqueda (viaje completo entre la pista más interna y más externa) cercano a 15 milisegundos.
- Latencia: cada pista en un disco duro contiene múltiples sectores como ya hemos comentado anteriormente. Una vez que la cabeza de Lectura/Escritura encuentra la pista correcta, las cabezas permanecen en el lugar, e inactivas hasta que el sector pasa por debajo de ellas. Este tiempo de espera se llama latencia.
- Command Overhead: tiempo que le toma a la controladora procesar un requerimiento de datos.
- Transferencia: se refiere al tiempo en el que los datos pueden ser leídos o escritos en el drive, que es afectado por la velocidad de los discos, la densidad de los bits de datos y el tiempo de acceso.

f) Fabricantes:

Western Digital, Maxtor, Samsung, Hitachi, Fujitsu, Quantum, Toshiba, Seagate, Zac.
La nueva generación de discos duros utiliza la tecnología de grabación magnética perpendicular (PMR) que permite mayor densidad de almacenamiento.
También existen discos llamados "Ecológicos" (Green Power), los cuales hacen un uso más eficiente de la energía.
En la actualidad se están imponiendo las **unidades de estado sólido** que tienen el mismo uso que los discos duros y emplean las mismas interfaces, pero no están formadas por discos mecánicos. Estas unidades tienden a categorizarse como "discos" cuando son tipos de memoria ya que los datos no se almacenan sobre superficies cilíndricas ni platos.

5.2. Memoria Flash y unidades de estado sólido (SSD)

5.2.1. Memoria flash

La **memoria flash** es una evolución de la memoria EEPROM que permite que múltiples posiciones de memoria sean escritas o borradas en una misma operación de programación mediante impulsos eléctricos, frente a las anteriores que sólo permite escribir o borrar una única celda cada vez. Por ello, flash permite funcionar a velocidades muy superiores, cuando

los sistemas emplean lectura y escritura en diferentes puntos de esta memoria, al mismo tiempo.
Sin embargo, todos los tipos de memoria flash sólo permiten un número limitado de escrituras y borrados, generalmente entre 10.000 y un millón, dependiendo de la celda, de la precisión del proceso de fabricación y del voltaje necesario para su borrado.
Este tipo de memoria está fabricado con puertas lógicas **NOR y NAND** para almacenar los 0´s ó 1´s correspondientes. Existen también memorias basadas en ORNAND.
Los sistemas de archivos para estas memorias son JFFS (Journaling Flash File System, Sistema de Ficheros Flash con soporte a Transacciones) originalmente para NOR, evolucionado a JFFS2 para soportar además NAND o YAFFS (Yet Another Flash File System, Sólo otro sistema de ficheros flash), ya en su segunda versión, para NAND. Sin embargo, en la práctica se emplea un sistema de archivos FAT por compatibilidad, sobre todo, en las tarjetas de memoria extraíble.
Estas memorias, como tipo de EEPROM que son, contienen una matriz de celdas con un transistor evolucionado con dos puertas en cada intersección. Tradicionalmente sólo almacenan un bit de información.
Las nuevas memorias flash, llamadas también dispositivos de celdas multinivel, pueden almacenar más de un bit por celda variando el número de electrones que almacenan.
Es necesario destacar que las memorias flash están subdivididas en bloques (en ocasiones llamados sectores) y por lo tanto, para el borrado, se limpian bloques enteros para agilizar el proceso, ya que es la parte más lenta del proceso. Por esta razón, las memorias flash son mucho más rápidas que las EEPROM convencionales, ya que borran byte a byte. No obstante, para reescribir un dato es necesario limpiar el bloque primero para después reescribir su contenido.
La memoria flash NAND usa tecnología flash de celda de único nivel (SLC, del inglés Single-Level Cell) o de celda de varios niveles (MLC, del inglés Multi-Level Cell). NAND SLC almacena un bit por celda y tiene una gran duración (unas 50.000 escrituras por celda). NAND MLC utiliza dos bits por celda (por lo que ofrece una mayor capacidad) pero se desgasta más rápido que NAND SLC (aproximadamente un décimo de la duración de la memoria flash SLC). Las nuevas memorias flash NAND de 3 bits por celda (unas 1.000 escrituras permitidas) y de 4 bits por celda (algunos cientos de escrituras permitidas) están destinadas a aplicaciones con un número muy limitado de escrituras.

5.2.2. Unidad de estado sólido (SSD)

¿Qué es almacenamiento de estado sólido?
El almacenamiento de estado sólido (SSS) es un método de almacenamiento de datos creado mediante dispositivos de circuitos integrados para almacenar datos en lugar de utilizar soportes ópticos o magnéticos móviles. El almacenamiento SSS suele ser no volátil y puede tomar varias formas, como una unidad de estado sólido, una tarjeta de estado sólido o un módulo de estado sólido. Además, SSS incluye opciones de interfaz PATA (anterior), SATA, SAS, canal de fibra o PCIe.

¿Qué es una unidad de estado sólido?
Las unidades de estado sólido (SSD) son dispositivos de almacenamiento de datos que usa memoria no volátil (tales como flash) o memoria volátil, para almacenar datos, en lugar de los platos giratorios encontrados en los discos duros convencionales. Aunque técnicamente no son "discos", estos dispositivos son llamados así porque son usados típicamente como reemplazo para las unidades de disco en situaciones donde los dispositivos convencionales no son prácticos. Al ser inmune a las vibraciones externas, lo hace especialmente apto para su uso en computadoras móviles. Las unidades SSD son compatibles con las interfaces de disco duro tradicionales, como SATA o SAS, y tienen un formato de disco duro familiar, como 3,5; 2,5 o 1,8 pulgadas.

Las primeras memorias estaban basadas en RAM pero en 1995 se presentaron unidades basadas en flash utilizadas en industria militar y aeroespacial ya que presentaban una alta tasa de tiempo medio entre fallos (MTBF), gran capacidad para soportar golpes fuertes, cambios bruscos de temperatura, presión y turbulencias. En 2009, Micron Tecnhology anunció el primer SSD del mundo utilizando la interfaz SATA III.
La arquitectura de las memorias SSD al principio estaba basada en memoria volátil DRAM (incorporando una batería y un sistema de copia de seguridad de almacenamiento). En la actualidad se basan en NAND flash.
Por otro lado, se denomina *Enterprise Fash Drives* (EFD) a aquellas unidades que están diseñadas para aplicaciones que requieren una alta tasa de operaciones por segundo, fiabilidad y eficiencia energetica. El término EFD fue acuñado por EMC en 2008 pero no está normalizado. Cualquier fabricante de SSD puede denominarlas EFD sin que existan unos requisitos mínimos.
IBM está diseñando un dispositivo **"Memoria RaceTrack"** basado en memorias no volatiles que permitirá velocidades hasta cien mil veces superiores a los discos duros tradicionales.

¿En qué se diferencia una unidad flash USB y una unidad de estado sólido?
Tanto unas como otras utilizan memoria flash NAND. No obstante, es la calidad de la NAND usada, así como el controlador y la interfaz, lo que diferencia una unidad flash USB de un dispositivo de almacenamiento de clase empresarial, como los empleados en los servidores blade y los sistemas de almacenamiento externo.

¿En qué se diferencian las unidades SSD y las unidades de disco duro mecánico?
Las unidades SSD actuales son diferentes de las unidades de disco duro en cuanto al almacenamiento de datos. Las unidades SSD son sofisticados dispositivos de almacenamiento que utilizan chips de memoria estáticos, principalmente memoria flash NAND no volátil, en lugar de los discos magnéticos giratorios de las unidades de disco duro. Las unidades de disco duro pueden tomar datos directamente del host y grabarlos en el soporte giratorio. Por el contrario, las unidades SSD no pueden grabar ni un bit de información sin borrar primero y regrabar después grandes bloques de datos a la vez (lo que también se denomina programar y borrar).
Los discos de estado sólido basados en Flash tienen varias ventajas únicas:

- Arranque más rápido.
- Mayor rapidez de lectura, en algunos casos, dos o más veces que los discos duros más rápidos.
- Baja latencia de lectura y escritura, cientos de veces más rápido que los discos mecánicos.
- Lanzamiento y arranque de aplicaciones en menor tiempo.
- Menor consumo de energía y producción de calor.
- Sin ruido, la misma carencia de partes mecánicas los hace completamente silenciosos
- Menor, pero mejorado tiempo de lectura y escritura.
- Seguridad, permitiendo una muy rápida "limpieza" de los datos almacenados.
- Rendimiento determinístico. A diferencia de los discos duros mecánicos, el rendimiento de los SSD es constante y determinístico a través del almacenamiento entero. El tiempo de "búsqueda" constante, y el rendimiento no se deteriora mientras el medio se llena.
- Menor peso y tamaño.

Los discos duros de estado sólido basados en flash tienen también varias desventajas:

- Precio, los precios de las memorias flash son considerablemente más altos que los de los discos convencionales.
- Menor tiempo de vida confiable.

- Menor recuperación, después de un fallo físico los datos son completamente perdidos pues la celda es destruida, mientras que en un disco duro normal que sufre daño mecánico los datos son frecuentemente recuperables usando ayuda de expertos.™
- Vulnerabilidad, contra ciertos tipo de efectos, incluyendo pérdida de energía abrupta, campos magnéticos y cargas estáticas comparados con los discos duros normales (que almacenan los datos dentro de una Jaula de Faraday*).
- Menor capacidad que los discos duros mecánicos.

  * Jaula de Faraday: efecto por el cual el campo electromagnético en el interior de un conductor en equilibrio es nulo, anulando el efecto de los campos externos. Esto se debe a que, cuando el conductor está sujeto a un campo electromagnético externo, se polariza, de esta manera queda cargado positivamente en la dirección en que va el campo electromagnético, y cargado negativamente en el sentido contrario. Puesto que el conductor se ha polarizado, este genera un campo eléctrico igual en magnitud pero opuesto en sentido al campo electromagnético, luego la suma de ambos campos dentro del conductor será igual a 0.
  Se pone de manifiesto en numerosas situaciones cotidianas, por ejemplo, el mal funcionamiento de los teléfonos móviles en el interior de ascensores o edificios con estructura de rejilla de acero. Una manera de comprobarlo es con una radio sintonizada en una emisora de Onda Media. Al rodearla con un periódico, el sonido se escucha correctamente. Sin embargo, si se sustituye el periódico con un papel de aluminio, la radio deja de emitir sonidos: el aluminio es un conductor eléctrico y provoca el efecto jaula de Faraday.
  Este fenómeno, descubierto por Michael Faraday, tiene una aplicación importante en aviones o en la protección de equipos electrónicos delicados, tales como discos duros o repetidores de radio y televisión situados en cumbres de montañas y expuestos a las perturbaciones electromagnéticas causadas por las tormentas.

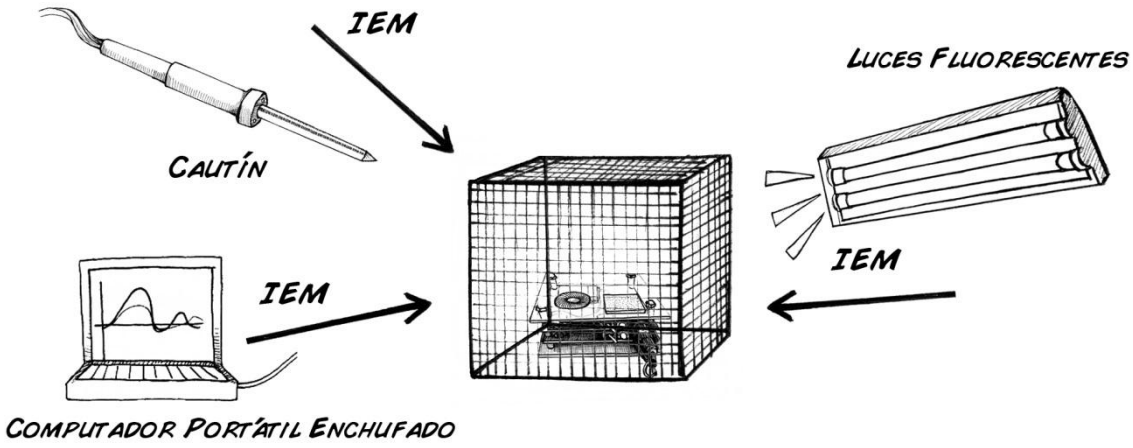


Figura 34. Jaula de Faraday.

¿Qué es la nivelación del desgaste?
La nivelación del desgaste es el proceso que usa un controlador de una unidad SSD para maximizar la vida de la memoria flash. Esta técnica nivela el desgaste en todos los bloques distribuyendo la escritura de datos por los dispositivos de memoria flash.

¿Qué es NVM Express?
NVM Express, NVMe, o Non-Volatile Memory Host Controller Interface Specification (NVMHCI) es una especificación para el acceso a las unidades de estado sólido (SSD) conectadas a través

del bus PCI Express (PCIe). "NVM" es un acrónimo de "memoria no volátil", la cual es utilizada en las unidades de estado sólido (SSD).
Como una interfaz de dispositivo lógica, NVM Express ha sido diseñada desde cero, aprovechando la baja latencia, el paralelismo de los SSD PCI Express, y reflejando el paralelismo de CPUs contemporáneos, plataformas y aplicaciones. Al permitir que los niveles de paralelismo que ofrecen las unidades SSD sean utilizadas en su totalidad por el hardware y el software del host, NVM Express trae varias mejoras de rendimiento.
Los SSDs NVM Express existen tanto en forma de tarjetas de expansión PCI Express, de tamaño estándar, como de unidades de 2,5", que proporcionan una interfaz PCI Express x4 a través de un conector SFF-8639. Los dispositivos de almacenamiento SATA Express y la especificación M.2 para el montado interno de tarjetas de expansión también soportan NVM Express como la interfaz de dispositivo lógica.
Es la sucesora de la especificación AHCI (Advanced Host Controller Interface).

Ventajas:

- Mejora el rendimiento de las unidades de estado sólido (SSD), aprovechando al máximo el ancho de banda y la capacidad de transferencia de las memorias de los mismos.
- Mejora la eficiencia energética en modo hibernación o suspensión.



Figura 35. SSD estándar de 2,5" y SSD NVMe.

5.3. CD, DVD, Blu-ray

CD
Un CD-ROM (Compact Disc - Read Only Memory, Disco Compacto - Memoria de sólo lectura), es un soporte digital óptico utilizado para almacenar cualquier tipo de información (audio, imágenes, vídeo, documentos y otros datos) y que puede ser leído por un ordenador con una lectora de CD. Un CD-ROM es un disco de plástico plano con información digital codificada en una espiral desde el centro hasta el borde exterior. Tiene un diámetro de 12 centímetros y pueden almacenar hasta 80 minutos de audio (o 700 MB de datos).
El disco compacto fue creado por el holandés Kees Immink, de Philips, y el japonés Toshitada Doi, de Sony, en 1979. Al año siguiente, Sony y Philips, que habían desarrollado el sistema de audio digital Compact Disc, comenzaron a distribuir discos compactos, pero las ventas no tuvieron éxito por la depresión económica de aquella época. Entonces, decidieron abarcar el mercado de la música clásica, de mayor calidad, con el lanzamiento del nuevo y revolucionario formato de grabación audio que posteriormente se extendería a otros sectores de la grabación de datos.

El sistema óptico fue desarrollado por Philips, mientras que la lectura y codificación digital corrió a cargo de Sony. Fue presentado en junio de 1980 a la industria y se adhirieron al nuevo producto 40 compañías de todo el mundo mediante la obtención de las licencias correspondientes para la producción de reproductores y discos.

Una lectora de CD es un dispositivo óptico electrónico que permite la lectura de estos CD-ROM mediante el empleo de un haz de un rayo láser y la posterior transformación de estos en impulsos eléctricos que el ordenador interpreta. Está compuesta por los siguientes elementos:

a) Un cabezal, en el que hay un emisor de rayos láser, que dispara un haz de luz hacia la superficie del disco, y que tiene también un fotorreceptor (foto-diodo) que recibe el haz de luz que rebota en la superficie del disco

b) Un motor que hace girar el disco compacto, y otro que mueve el cabezal radialmente. Con estos dos mecanismos se tiene acceso a todo el disco y se regula la velocidad de acceso (CLV o CAV, consultar más adelante)

c) Un DAC (Digital to Analogical Converter), en el caso del CD-audio, y en casi todos los CD-ROM. Es decir, un convertidor de señal digital a señal analógica, la cual es enviada a los altavoces. DAC's también hay en las tarjetas de sonido, que en su gran mayoría, tienen también un ADC (Analogical to Digital Converter), que hace el proceso inverso, de analógico al digital.

Los CD-ROM's están escritos por grabadoras de CD, dispositivo similar a la lectora de CD, con la diferencia que hace lo contrario a la lectora, es decir, transformar impulsos eléctricos en un haz de luz láser que almacenan en el CD datos binarios en forma de pozos (pits) y llanos (lands) Para esto, la grabadora crea los pozos y los llanos cambiando la reflectividad de la superficie del CD. Los primeros son zonas donde el láser quema la superficie con mayor potencia, creando ahí una zona de baja reflectividad. Los segundos, son justamente lo contrario, son zonas que mantienen su alta reflectividad inicial, justamente porque la potencia del láser se reduce.

Uno de los parámetros más importantes para medir la calidad de un CD-ROM es el tiempo de acceso. Este es la cantidad de tiempo que le lleva al dispositivo, desde que comienza el proceso de lectura, hasta que los datos comienzan a ser leídos y se calcula como la conjunción:

a) La latencia: tiempo que tarda en pasar el dato por debajo del cabezal de lectura.

b) El tiempo de búsqueda: tiempo que lleva mover el cabezal de lectura hasta la posición del disco en la que están los datos.

c) El tiempo de cambio de velocidad, con alguno de estos dos tipos:

- CLV (*Constant Linear Velocity*, Velocidad Lineal Constante): la velocidad de giro del motor dependerá de la posición que el cabezal de lectura ocupe en el disco, más rápido cuanto más cerca del centro. Esto implica un tiempo de adaptación para que este motor tome la velocidad adecuada una vez que conoce el punto en el que se encuentran los datos.
- CAV (*Constant Angular Velocity*, Velocidad Angular Constante): velocidad variable según la posición del dato. En este caso la velocidad de rotación es siempre la misma, así que la velocidad de acceso se verá beneficiada por esta característica y será algo menor.

Este tiempo de acceso depende directamente de la velocidad de la unidad de CD-ROM, ya que los componentes de ésta también dependen de ella. La razón por la que el tiempo de acceso es mayor en los CDROM respecto a los discos duros se encuentra en su construcción, ya que su disposición de cilindros reduce considerablemente los tiempos de búsqueda. Por su parte, los CD-ROM no fueron inicialmente ideados para el acceso aleatorio, sino para el secuencial de los CD de audio. Los datos se disponen en espiral en la superficie del disco, y el tiempo de búsqueda es por lo tanto mucho mayor.

Un **CD-R** es un formato de disco compacto grabable (*Compact Disc Recordable*, Disco Compacto Grabable). Se pueden grabar en varias sesiones, sin embargo, la información agregada no puede ser borrada ni sobrescrita, en su lugar, se debe usar el espacio libre que

dejó la sesión inmediatamente anterior. Actualmente las grabadoras llegan a grabar CD-R a 52x, unos 7.800 KB/s. Para muchos ordenadores es difícil mantener esta tasa de grabación y, por ello, las grabadoras tienen sistemas que permiten retomar la grabación ante un corte en la llegada de datos.
La capacidad total de un CD-R suele ser:
- 650 MB
- 700 MB. El más común.
- 800 MB
- 900 MB

Un **CD-RW** es un formato de disco compacto regrabable (*Compact Disc ReWritable*. Disco Compacto Regragable). Se trata de un soporte digital óptico utilizado para almacenar cualquier tipo de información. Este tipo de CD puede ser grabado múltiples veces, ya que permite que los datos almacenados sean borrados. Fue desarrollado conjuntamente en 1980 por las empresas Sony y Philips, y comenzó a comercializarse en 1982.
Hoy en día, tecnologías como el DVD han desplazado en parte esta forma de almacenamiento, aunque su uso sigue vigente.
En el disco CD-RW la capa que contiene la información está formada por una aleación cristalina de plata, indio, antimonio y telurio. Esta aleación presenta una interesante cualidad: cuando es calentada hasta cierta temperatura, al enfriarse se convierte en cristalina; pero si al calentarse se alcanza una temperatura aún más elevada, al enfriarse queda con estructura amorfa. La superficie cristalina permite que la luz se refleje bien en la zona reflectante, mientras que la zona con estructura amorfa absorbe la luz. Por ello el CD-RW utiliza tres tipos de luz:

- Láser de escritura: se usa para escribir. Calienta pequeñas zonas de la superficie para que el
material se torne amorfo.
- Láser de borrado: se usa para borrar. Tiene una intensidad menor que el de escritura con lo que se consigue el estado cristalino.
- Láser de lectura: se usa para leer. Tiene menor intensidad que el de borrado. Se refleja en zonas cristalinas y se dispersa en las amorfas.

DVD
Al comienzo de los años 90, se estaban desarrollando dos estándares de almacenamiento óptico de alta densidad: uno era el *Multimedia Compact Disc* (MMCD), apoyado por Philips y Sony; el otro era el *Super Density Disc* (SD), apoyado por Toshiba, Time-Warner, Matsushita Electric, Hitachi, Mitsubishi Electric, Pioneer, Thomson y JVC.
Philips y Sony abandonaron su formato MMCD y acordaron con Toshiba adoptar el SD, pero con una modificación: la adopción del EFM (*Eight-to-Fourteen Modulation*) Plus de Philips, creado por Kees Immink, que a pesar de ser un 6% menos eficiente que el sistema de codificación de Toshiba (de ahí que la capacidad sea de 4,7 GB en lugar del los 5 GB del SD original), cuenta con la gran ventaja de poseer gran resistencia a los daños físicos en el disco (como arañazos o huellas).
El resultado fue la creación del Consorcio del DVD, fundada por las compañías anteriores, y la especificación de la versión 1.5 del DVD, anunciada en 1995 y finalizada en septiembre de 1996. En mayo de 1997, el consorcio DVD (DVD *Consortium*) fue reemplazado por el foro DVD (DVD Forum).
Un DVD tiene una cuantificación de 24 bits, una velocidad de muestreo de 48MHz y un rango dinámico de 144 dB. El disco puede tener una o dos caras (estos últimos apenas se utilizan), y una o dos capas de datos por cada cara (el número de caras y capas determina la capacidad del disco). Los de capa simple tienen una capacidad de 4,7GB y los de doble capa 8,5GB. Esta capacidad esta expresada en base decimal (la proporcionada por los fabricantes) pero la capacidad real sería, por ejemplo para los de simple cara, de aproximadamente 4,38 gigabytes reales en base binaria o gibibytes (se lo conoce como DVD-5).

El DVD usa un método de codificación más eficiente en la capa física: los sistemas de detección y corrección de errores utilizados en el CD, como la comprobación de redundancia cíclica CRC, la codificación *Reed Solomon - Product Code* (RS- PC), así como la codificación de línea EMF, que fue reemplazada por una versión más eficiente, EFM Plus, con las mismas características que el EFM clásico. Como resultado, el formato DVD es un 47% más eficiente que el CD-ROM, que emplea una tercera capa de corrección de errores.
A diferencia de los discos compactos, donde el sonido se guarda de manera fundamentalmente distinta que los datos, un DVD correctamente creado siempre contendrá datos siguiendo los sistemas de archivos UDF (*Universal Disk Format*), con estándar ISO 9660 publicado en 1986 que define el sistema de archivos de un CD. Igualmente, puede usar Joliet, extensión del estándar ISO 9660 especificado por Microsoft.
Los DVD se pueden clasificar:

- Según su contenido:
  - DVD-Video: películas (vídeo y audio).
  - DVD-Audio: audio de alta fidelidad.
  - DVD-Data: todo tipo de datos.
- Según su capacidad de regrabado:
  - DVD-ROM: sólo lectura, manufacturado con prensa.
  - DVD-R y DVD+R: grabable una sola vez. La diferencia entre los tipos +R y -R radica en la forma de grabación y de codificación de la información: los +R los pozos son 1´s lógicos mientras que en los -R los pozos son 0´s lógicos.
  - DVD-RW y DVD+RW: regrabable.
  - DVD-RAM: regrabable de acceso aleatorio. Lleva a cabo una comprobación de la integridad de los datos siempre activa tras completar la escritura.
  - DVD+R DL (*dual layer*): grabable una sola vez de doble capa.
  - DVD-ROM almacena desde 4,7 GB hasta 17 GB.
- Según su número de capas o caras:
  - DVD-5: una cara, capa simple; 4,7 GB o 4,38 GiB - Discos DVD±R/RW.
  - DVD-9: una cara, capa doble; 8,5 GB o 7,92 GiB - Discos DVD+R DL. La grabación de doble capa permite a los discos DVD-R y los DVD+RW almacenar significativamente más datos, hasta 8,5 GB por disco, comparado con los 4,7 GB que permiten los discos de una capa.
    Los DVD-R DL fueron desarrollados para DVD Forum por *Pioneer Corporation*.
    DVD+R DL fue desarrollado para el DVD+R Alliance por Philips y Mitsubishi Kagaku Media.
    Un disco de doble capa difiere de un DVD convencional en que emplea una segunda capa física ubicada en el interior del disco. Una unidad lectora con capacidad de doble capa accede a la segunda proyectando el láser a través de la primera semitransparente. El mecanismo de cambio de capa en algunos DVD puede conllevar una pausa de hasta un par de segundos.
    Los discos grabables soportan esta tecnología manteniendo compatibilidad con algunos reproductores de DVD y unidades DVD-ROM. Muchos grabadores de DVD soportan la tecnología de doble capa, y su precio es comparable con las unidades de una capa, aunque el medio continúa siendo considerablemente más caro.
  - DVD-10: dos caras, capa simple en ambas; 9,4 GB o 8,75 GiB - Discos DVD±R/RW.
  - DVD-14: dos caras, capa doble en una, capa simple en la otra; 13,3 GB o 12,3 GiB. Raramente utilizado.
  - DVD-18: dos caras, capa doble en ambas; 17,1 GB o 15,9 GiB - Discos DVD+R.

También existen DVD de 8 cm (no confundir con miniDVD, que son CD que contienen información de tipo DVD video) que tienen una capacidad de 1,5 GB.
El DVD Forum creó los estándares oficiales DVD-ROM/R/RW/RAM, y Alliance creó los estándares DVD+R/RW para evitar pagar la licencia al DVD Forum. Dado que los discos

DVD+R/RW no forman parte de los estándares oficiales, no muestran el logotipo “DVD”. En lugar de ello, llevan el logotipo “RW” incluso aunque sean discos que solo puedan grabarse una vez, lo que ha suscitado cierta polémica en algunos sectores que lo consideran publicidad engañosa, además de confundir a los usuarios.
La mayoría de grabadoras de DVD nuevas pueden grabar en ambos formatos y llevan ambos logotipos “+RW”y “DVD- R/RW”. La velocidad de transferencia de datos de una unidad DVD está dada en múltiplos de 1.350 KB/s.
Las primeras unidades DVD leían datos a velocidad constante (CLV). Los datos en el disco pasaban bajo el láser de lectura a velocidad constante. Como la velocidad lineal (metros/segundo) de la pista es tanto mayor cuanto más alejados esté del centro del disco (de manera proporcional al radio), la velocidad rotacional del disco se ajustaba de acuerdo a qué porción del mismo se estaba leyendo.
Actualmente, la mayor parte de unidades DVD tienen una velocidad de rotación constante (CAV). La máxima velocidad de transferencia de datos especificada para una cierta unidad y disco se alcanza, solamente, en los extremos del disco. Por tanto, la velocidad media de la unidad lectora equivale al 50-70% de la velocidad máxima para la unidad y el disco. Aunque esto puede parecer una desventaja, tales unidades tienen un menor tiempo de búsqueda, pues nunca deben cambiar la velocidad de rotación del disco.

Blu-ray DISC (BD)

El Blu-ray es un formato de disco óptico de nueva generación de 12 cm de diámetro (igual que el CD y el DVD) para vídeo de gran definición y almacenamiento de datos de alta densidad. Su capacidad de almacenamiento llega a 25GB a una capa y a 50GB a doble capa, aunque existen de mayor capacidad. Este formato se impuso a su competidor, el HD DVD (abandonado en el año 2008), en la guerra de formatos iniciada para cambiar el estándar DVD.
El disco Blu-ray hace uso de un rayo láser de color azul con una longitud de onda de 405 nanómetros, a diferencia del láser rojo utilizado en lectores de DVD, que tiene una longitud de onda de 650 nanómetros.
Esto, junto con otros avances tecnológicos, permite almacenar sustancialmente más información que el DVD en un disco de las mismas dimensiones y aspecto externo. Blu-ray obtiene su nombre del color azul del rayo láser (*blue ray* significa “rayo azul”).
El DVD ofreció en su momento una alta calidad, ya que era capaz de dar una resolución de 720x480 (NTSC) o 720x576 (PAL), lo que es ampliamente superado por la capacidad de alta definición ofrecida por el Blu-ray, que es de 1920x1080 (1080p).
Además, Blu-ray vino a solucionar dos problemas del DVD:

- En primer lugar, para la lectura en el DVD el láser debe atravesar la capa de policarbonato de 0,6mm en la que el láser se puede difractar en dos haces de luz. Si esta difracción es alta (por ejemplo, si el disco estuviera rayado), impediría su lectura.. Como el disco Blu-ray tiene una capa de tan sólo 0,1mm se evita este problema, ya que tiene menos recorrido hasta la capa de datos; además, esta capa es resistente a las rayaduras.
- En segundo lugar, si el disco DVD estuviera inclinado por igual motivo que el anterior problema, la distorsión del rayo láser haría que leyese en una posición equivocada, dando lugar a errores. En el caso del Blu-ray, gracias a la cercanía de la lente y la rápida convergencia del láser, la distorsión es inferior, pudiéndose evitar los posibles errores de lectura.

Este formato fue desarrollado en conjunto por un grupo de compañías tecnológicas llamado *Blu-Ray Disc Association* (BDA), liderado por Sony y Philips.
Competidores:

- HD-DVD que puede almacenar hasta 30 GB. Se lanzó en 2003 pero se descatalogó en 2008.

- HD-VMD. Tecnología de láser rojo. Utiliza como estándar 4 capas consiguiendo 20 GB, que es comparable a una sola capa de HD DVD (15 GB) o una sola capa de BD (25 GB).

Sucesores:

- HVD o *disco holográfico versátil*, con 6 TB de capacidad. Regrabables al menos durante 5 años.
- Archival Disc (2015- 2016) Está preparado para películas en resolución 4k y se espera capacidad para 300 GB, 500 GB y 1 TB.
- Blu-ray 4K Ultra HD presentado en 2016. Está diseñado para almacenar películas y series de televisión con resolución 4K Ultra HD. Existen tres tipos de discos diferentes, según las capas que se utilicen: 1 capa con 50 GB de tamaño y 82 Mbps de Bitrate, 2 capas con 66 GB de tamaño y 108 Mbps de Bitrate y 3 capas con 100 GB de tamaño y 128 Mbps de Bitrate.
  Además del aumento de resolución, el nuevo formato incorpora nuevas tecnologías de imagen y sonido como HDR / Dolby Vision o Digital Bridge que permite copiar la película completa a un disco duro para disfrutarla en otros dispositivos, o transferir una versión digital a un smartphone o una tablet.

5.4. Cinta magnética

La cinta magnética es un soporte de almacenamiento de información que se graba en pistas sobre una banda plástica con un material magnetizado, generalmente óxido de hierro o algún cromato. El tipo de información que se puede almacenar en las cintas magnéticas es variado, como vídeo, audio y datos.
Estos dispositivos informáticos de almacenamiento masivo de datos son utilizados, principalmente, para respaldo de archivos y para el proceso de información de tipo secuencial. Al almacén donde se guardan estos dispositivos se le denomina cintoteca.
La grabación magnética conlleva registrar sobre la plástica impulsos magnéticos en forma de señales analógicas o digitales por codificar, de forma que la información puede ser accedida repetidamente. Una característica de este medio es que la cabeza codificadora debe estar en contacto directo con la superficie magnética y provocar un movimiento constante para ser leída a través de movimiento mecánico, es decir, manualmente o a través de un motor.
Los métodos de grabación son una forma de clasificar la tecnología y cada cambio ha reutilizado la cinta magnética para aprovechar el medio de almacenamiento:

- Grabación Lineal: con este tipo de grabación de señales analógicas y digitales registradas en la cinta magnética como audio o datos codificados, se puede observar en las primeras cintas la traza de la señal en forma de línea horizontal, de acuerdo a las necesidades de ingenieros en audio video o datos.
  Hubo variantes como la grabación en pistas que comprenden múltiples trazas horizontales grabadas con diferentes cabezas que podían ser reproducidas al mismo tiempo o el Lineal *serpentine* que, en el momento de llegar al final de la cinta, la cabeza graba en dirección opuesta trazando una segunda línea paralela, esto requirió contar con una mayor capacidad de almacenamiento.
- Grabación Transversal: para aprovechar la superficie grabable de la cinta magnética se diseñó un tambor giratorio, con cuatro cabezas de grabación, para almacenar video. Éste requiere mayor cantidad de información almacenada a lo ancho de la cinta, mientras las cinta estaban en constante movimiento, cada pista está ligeramente inclinada. Muy pocos equipos utilizaron este método de grabación, como las grabadoras Mark I y Mark II de Ampex, que nombraron esta tecnología de grabación como "Quádruplex".
- Grabación Helical: para perfeccionar el modo de grabación transversal se redujeron de cuatro cabezas a dos que rotan en diferentes ángulos hacia la dirección en la que la

cinta se transporta. Para lograr una secuencia de grabación continua y mayor superficie de cinta aprovechada, se pueden apreciar las trazas de la cinta grabadas diagonalmente a lo largo de ésta.

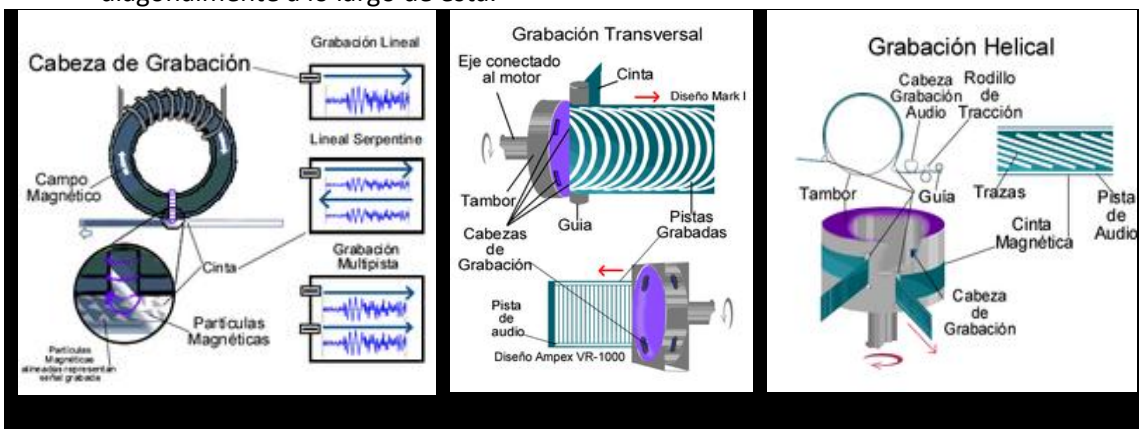
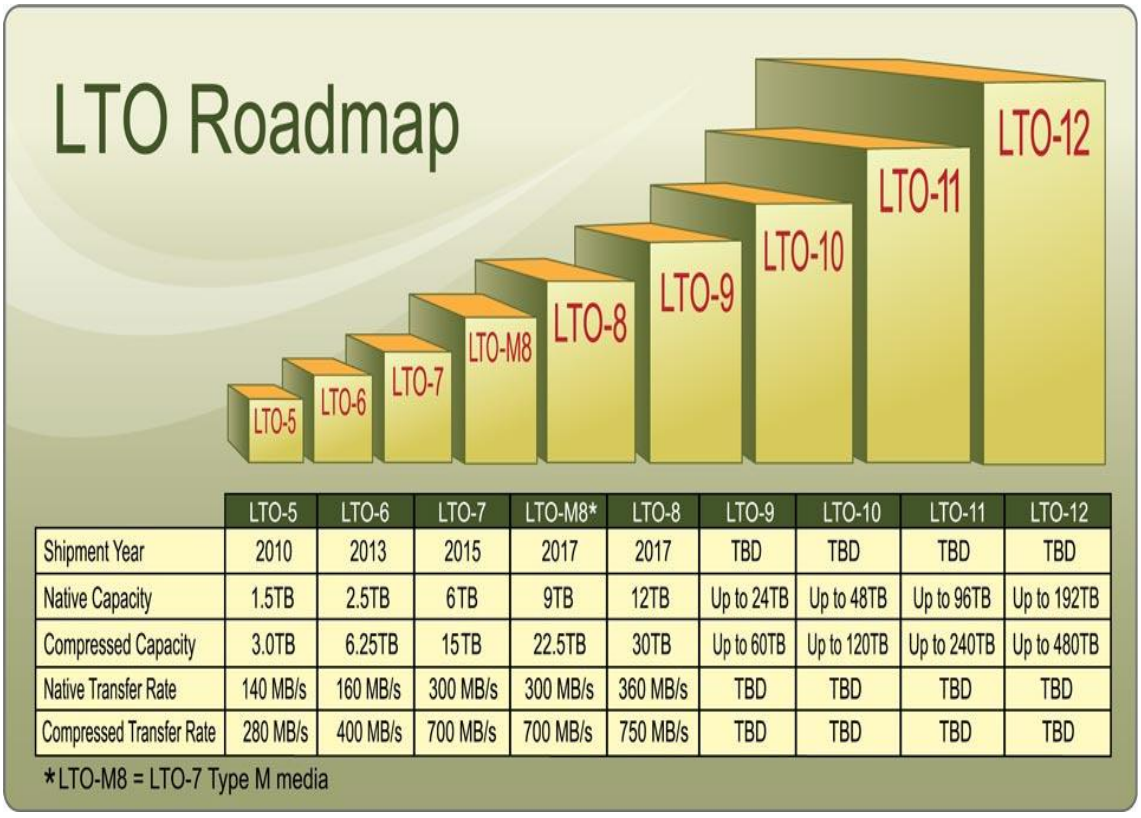


Figura 36. Grabación lineal, transversal y helical.

Tecnologías:

- Travan. Una tecnología de cinta que evolucionó a partir del estándar industrial del QIC. El Travan ofrece más capacidad que los formatos anteriores de QIC. Con capacidades que van de los 400 MB a los 20 GB y alcanzan velocidades de hasta 4 MBps, los formatos Travan son idóneos para el respaldo en máquinas individuales o pequeños servidores.
- DAT (Digital Audio Tape Cinta de audio digital). Las cintas DAT surgieron inicialmente como cintas de grabación de audio digital, pero posteriormente SONY y HP definieron el estándar para grabación de datos. Utilizan cinta de 4mm y grabación helicoidal. El cartucho dispone de dos carretes, de forma que la cinta sale del cartucho para ser accedida y se vuelve a enrollar en el mismo cartucho en el otro carrete, como las cintas de audio o de video. Utiliza 4 cabezas, dos grabadoras y dos lectoras. Las lectoras están situadas detrás de las escritoras de forma que en una escritura verifican lo que la cabeza de escritura ha grabado. Actualmente se dispone de cintas DAT de hasta 80 GB (160 GB con compresión) y velocidades de acceso de hasta 12 MBps. Las cintas deben renovarse tras unas 2.000 regrabaciones.
- DLT (Digital Linear Tape Cinta lineal digital). Tecnología de cintas desarrollada por la Digital Equipment Corporation (DEC). La tecnología DLT utiliza cintas alojadas en cartuchos de un solo carrete. Utiliza grabación lineal, con entre 128 y 208 pistas por cinta según la unidad, permitiendo alcanzar hasta 300 GB por cinta y velocidades de transferencia de 36 MB/s. La tecnología de acceso a la cinta minimiza el desgaste de la misma, pudiendo tener unas 30.000 horas de vida útil cada cinta.
- LTO. El principal de todos ellos es LTO (Linear Tape Open). Presentado por HP, IBM y Seagate. Las unidades LTO estarán destinados a servidores medios/grandes y ambientes de empresa, compite directamente con DLT. Como el nombre indica, LTO es un estándar abierto, ya que por un pequeño pago de derechos cualquier fabricante de unidades puede diseñar y fabricar unidades LTO.

Versiones:



	LTO-5	LTO-6	LTO-7	LTO-M8*	LTO-8	LTO-9	LTO-10	LTO-11	LTO-12
Shipment Year	2010	2013	2015	2017	2017	TBD	TBD	TBD	TBD
Native Capacity	1.5TB	2.5TB	6TB	9TB	12TB	Up to 24TB	Up to 48TB	Up to 96TB	Up to 192TB
Compressed Capacity	3.0TB	6.25TB	15TB	22.5TB	30TB	Up to 60TB	Up to 120TB	Up to 240TB	Up to 480TB
Native Transfer Rate	140 MB/s	160 MB/s	300 MB/s	300 MB/s	360 MB/s	TBD	TBD	TBD	TBD
Compressed Transfer Rate	280 MB/s	400 MB/s	700 MB/s	700 MB/s	750 MB/s	TBD	TBD	TBD	TBD

*LTO-M8 = LTO-7 Type M media

Figura 37. Versiones LTO.

- VXA Variable Speed Architecture. Formato introducido por Ecrix. Están dirigidas a servidores de bajo nivel que pueden haber usado DDS u otros formatos de 8 mm. Versiones:
  - VXA-1: capacidad de 33 GB a una velocidad de transferencia de 3 MB/s.
  - VXA-2: capacidad de 80 GB a una velocidad de transferencia de 6 MB/s.
  - VXA-3: capacidad de 160 GB a una velocidad de transferencia de 12 MB/s.
  - VXA-4: capacidad de 320 GB a una velocidad de transferencia de 24 MB/s.
  - VXA-5: capacidad de 640 GB a una velocidad de transferencia de 48 MB/s.

5.5. Sistemas RAID

5.5.1. Introducción

Las siglas RAID (R*edundant Array of Independent Disks*, conjunto redundante de discos independientes), hacen referencia a un sistema de almacenamiento que utiliza un conjunto de discos duros independientes organizados para que el sistema operativo los vea como un solo disco lógico. Se pueden combinar discos de varios tipos, por ejemplo se pueden combinar grupos de discos duros mecánicos, grupos de dispositivos de estado sólido (SDD) o grupos de ambos . Almacena los datos de forma redundante. Los RAID suelen usarse en servidores y normalmente se implementan con unidades de disco de la misma capacidad.

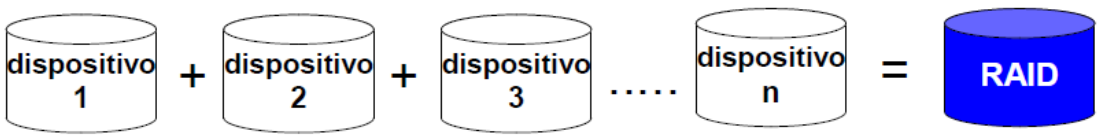


Figura 38. RAID.

Los beneficios de RAID respecto a un único disco duro son: mayor integridad en los datos, mayor tolerancia a fallos, mayor rendimiento y mayor capacidad.
Se mejora el rendimiento ya que permite a varias unidades trabajar en paralelo. La fiabilidad se aumenta mediante dos técnicas: redundancia que implica el almacenamiento de los mismos datos en más de una unidad y la paridad de datos. Esta última consiste en un algoritmo matemático que genera información de paridad, cuando se produce un fallo en una unidad se leen los datos correctos que quedan y se comparan con los datos de paridad almacenados. El uso de paridad es menos costoso que la redundancia ya que no requiere la utilización de un conjunto redundante de unidades de disco.
El funcionamiento del sistema RAID se sustenta en dos elementos: DATA STRIPPING y PARIDAD. En el Data Stripping (Bandeado o creación de bandas), los datos que llegan al RAID, procedentes del servidor, son divididos por la controladora RAID en segmentos, cuyo tamaño depende del bloque que se defina. Posteriormente, esos segmentos son enviados a los diferentes discos que componen el sistema RAID. Por otro lado, en el concepto de Paridad, la controladora RAID genera bits de paridad y los almacena en los discos del RAID, obteniendo así la redundancia de datos. De este modo, si un disco falla los datos pueden ser regenerados.
La tecnología RAID surge ante la necesidad que existe hoy en día de tener un buen sistema de almacenamiento seguro y tolerante a fallos. Asegura la integridad de los datos ante la avería de uno de los discos, asegurando un funcionamiento continuo y permitiendo la substitución de una unidad defectuosa sin necesidad de detener los procesos que se están ejecutando.
El rendimiento general del sistema aumenta ya que pueden funcionar de forma paralela con los diferentes discos del conjunto. Dependiendo del nivel de RAID que escojamos, si uno de los discos del conjunto falla, la unidad continúa funcionando, sin pérdida de tiempo ni de datos. La reconstrucción de los datos del disco que ha fallado se hace de forma automática sin intervención humana. En el caso de algunos sistemas operativos la regeneración de datos se hace desde software, aunque en estos sistemas se pueden usar controladoras RAID que sí regenerarían los datos automáticamente. La capacidad global del disco aumentará, ya que se suman las capacidades de los diferentes discos que componen el conjunto.
Para conseguir mejoras en las prestaciones del RAID, se utiliza memoria cache de lectura, que contiene datos de forma temporal, minimizando así el número de accesos necesarios. En la memoria caché de escritura se almacena un determinado número de bloques de datos, adyacentes a ser escritos, disminuyendo de ese modo los accesos a disco. Un sistema de discos RAID es plenamente multi-usuario, ya que todas las solicitudes de los usuarios pueden ser atendidas simultáneamente.
Existen dos tipos de **Implementaciones**, según la controladora, para los sistemas RAID.

- Las controladoras RAID hardware ofrecen mayores prestaciones y son independientes del sistema operativo, lo cual evita que se pueda generar cualquier tipo de conflicto.
- Por contra, si se emplea un Software RAID para que realice las funciones propias del RAID, se consumirán recursos del servidor. Esto ralentizará su rendimiento. Además, al trabajar con un software específico se depende del sistema operativo, lo cual puede llevar a generar conflictos que deriven incluso en la corrupción de los datos.

¿Qué es mejor: una controladora RAID Interna o Externa? Sin duda es mejor utilizar un sistema RAID externo al servidor. Si el RAID es interno y se produce un fallo, en el propio RAID o en otro componente, se pierden los datos. En cambio, si el RAID es externo, aunque falle el servidor los datos seguirán intactos. Luego se conecta el RAID a otro servidor y el trabajo puede continuar sin apenas perder tiempo. Además, solo el RAID externo permite configuraciones CLUSTER, y al no ser un SISTEMA PROPIETARIO, se puede conectar a cualquier servidor, con independencia de su marca, sin que dé ningún problema ni limite las posibilidades de configuración.
Tipos de Cache: Write Back envían al servidor un mensaje de terminación tan pronto como han escrito los datos en la caché. La controladora escribirá los datos en disco a su

conveniencia. De esta forma se mejoran, en definitiva, las prestaciones en escritura. Write Through: En este segundo caso, la controladora escribe los datos a disco y posteriormente manda el mensaje de finalización al servidor. Este proceso de caché resulta más lento que el anterior.

Tipos de Arreglos:

- Arreglos paralelos: éstos son aquellos en que cada disco participa en todas las operaciones de entrada/salida. Este tipo de arreglo ofrece tasas altísimas de transferencia debido a que las operaciones son distribuidas a través de todos los discos del arreglo y ocurren en forma prácticamente simultánea. La tasa de transferencia será muy cercana, 95%, a la suma de las tasas de los discos miembros, mientras que los índices de operaciones de entrada/salida serán similares a las alcanzadas por un disco individual. En español: un arreglo paralelo accesará sólo un archivo a la vez pero lo hará a muy alta velocidad. Algunas implementaciones requieren de actividades adicionales como la sincronización de discos. Los RAID de niveles 2 y 3 se implementan con arreglos paralelos.
- Arreglos independientes: son denominados así aquellos arreglos en los cuales cada disco integrante opera en forma independiente, aún en el caso de que le sea solicitado atender varios requerimientos en forma concurrente. Este modelo ofrece operaciones de entrada/salida sumamente rápidas debido a que cada disco está en posición de atender un requerimiento por separado. De esta forma las operaciones de entrada/salida serán atendidas a una velocidad cercana, 95%, a la suma de las capacidades de los discos presentes, mientras que la tasa de transferencia será similar a la de un disco individual debido a que cada archivo está almacenado en sólo un disco.

Los niveles 4 y 5 de RAID se implementan con arreglos independientes, mientras que los niveles 0 y 1 pueden ser implementados por cualquiera de las categoríías, sin perjuicio de suelan ser implementados en forma de arreglos independientes. Stripping y mirroring RAID a niveles 0, 1 y 0 & 1 puede ser implementado, tanto en forma de arreglos independientes o paralelos.

5.5.2. Conceptos básicos

- **Striping o particionado:** técnica de segmentación lógica de los datos que luego son accedidos de forma secuencial en diferentes dispositivos de almacenamiento. Provee un mayor rendimiento en la velocidad de acceso a los datos almacenados en múltiples dispositivos. La falla de un dispositivo causa la perdida de todos los datos. Según lo anterior el porcentaje de falla es la suma del porcentaje de falla de cada dispositivo.
- **Mirroring o espejo:** técnica en la cual se crean replicas en tiempo real del volumen lógico de un dispositivo en otros dispositivos físicos. Ofrece alta disponibilidad de los datos debido a las múltiples réplicas de los datos (redundancia). De forma adicional provee mejoras en el rendimiento de acceso a los datos (lectura) en múltiples dispositivos. Implementaciones de esta técnica representan un alto costo debido a la réplica de cada dispositivo.
- **Parity o paridad:** se genera un conjunto de datos de redundancia a partir de dos o más conjuntos de datos primarios aplicando la función Boolena XOR. Con los datos de redundancia se puede reconstruir los datos de algunos de los conjuntos de datos primarios. Aunque no implica duplicar por completo los datos primarios, esta técnica puede ocasionar bajo rendimiento en la velocidad de escritura en un RAID.
- **Modo degradado:** momento en que un grupo de dispositivos falla por cualquier razón en un RAID (con redundancia). Cuando ocurre funcionamiento no óptimo, redundancia comprometida, suele ocurrir en RAID 1, 5, 10, 50 y 60.

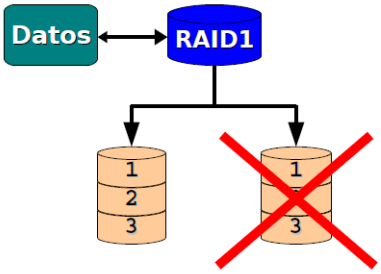


Figura 39. Modo degradado.

- **Hot spare o disco sobrante:** Es un disco que permanece siempre en el sistema esperando a que uno se estropee y él entre directamente en funcionamiento. Esta característica en niveles con soporte de redundancia permite que RAID se recupere de una falla por si mismo. Soporte tanto en RAID Hardware como Software. El RAID pasa a modo degradado mientras hot spare actúa.

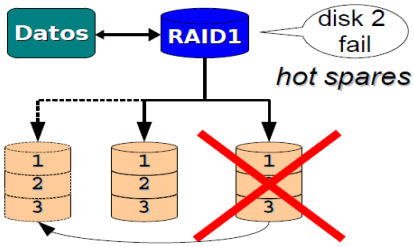


Figura 40. Hot spare.

- **Hot swap:** Son discos que se reemplazan en caliente. Esto quiere decir que si alguno de los discos RAID se estropeara físicamente bastaría con quitarlo y poner otro sin tener que apagar el sistema. Generalmente se usa en caso de que no se tiene espacio físico para instalar dispositivos con soporte hot spares o cuando un dispositivo en modo *hot spare* está activo y se debe reemplazar el dispositivo dañado previendo fallas a futuro.

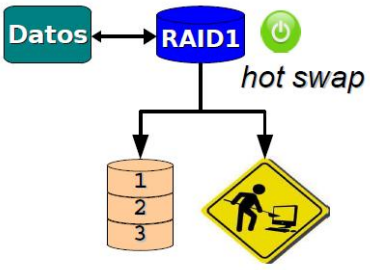


Figura 41. Hot swap.

5.5.3. Niveles RAID

Son los que determinan las diferentes configuraciones RAID posibles, las cuales van a depender de las necesidades de los usuarios y de las posibilidades de las controladoras. Dichos niveles han sido definidos por el RAID ADVISORY BOARD (RAB).
Cada nivel RAID ofrece una combinación específica de **tolerancia a fallos (redundancia), rendimiento y coste**, diseñadas para satisfacer las diferentes necesidades de almacenamiento. La mayoría de los niveles RAID pueden satisfacer solamente uno o dos de estos tres criterios.

Los niveles de RAID más usados son: RAID 0, RAID1 y RAID5. También existen niveles RAID anidados, que son la combinación de los niveles RAID normales: RAID 0+1, RAID 1+0, entre otros.

RAID 0: disk striping

Tiene la transferencia más alta pero sin tolerancia a fallos. Se utiliza normalmente para incrementar el rendimiento. Los datos se dividen en pequeños segmentos (stripes) de tamaño similar y se distribuyen en varias unidades, estos segmentos son intercalados de manera secuencial y repetida. Al no existir redundancia de datos no hay tolerancia a fallos. No se considera RAID propiamente dicho, sino un conjunto de unidades de disco conectadas en paralelo que permiten una transferencia simultánea de datos a gran velocidad.

Se necesita un mínimo de dos unidades de disco para implementar esta solución y la capacidad total de almacenamiento es la suma de la capacidad de todos los dispositivos del grupo, se pueden usar dispositivos de diferentes tamaños, recordando que el dispositivo de menor tamaño limita la cantidad de espacio usado en los demás dispositivos. Por ejemplo, tenemos un disco de 40 GB y otro de 80 GB, si hacemos RAID 0 con esos discos, la capacidad total sería 80 GB (40 GB+ 40 GB), perderíamos 40 GB del disco de 80 GB. Un daño en cualquiera de los dispositivos haría inutilizable el RAID.

Este tipo de solución es válida en aplicaciones que requieran un gran RATIO DE E/S, y en las que se puede aceptar la pérdida de algún dato, recomendable su uso en aplicaciones de tratamiento de imágenes, audio, vídeo y prensa.

*Una página interesante donde ver las capacidades de cada nivel RAID es: https://www.synology.com/es-es/support/RAID_calculator?

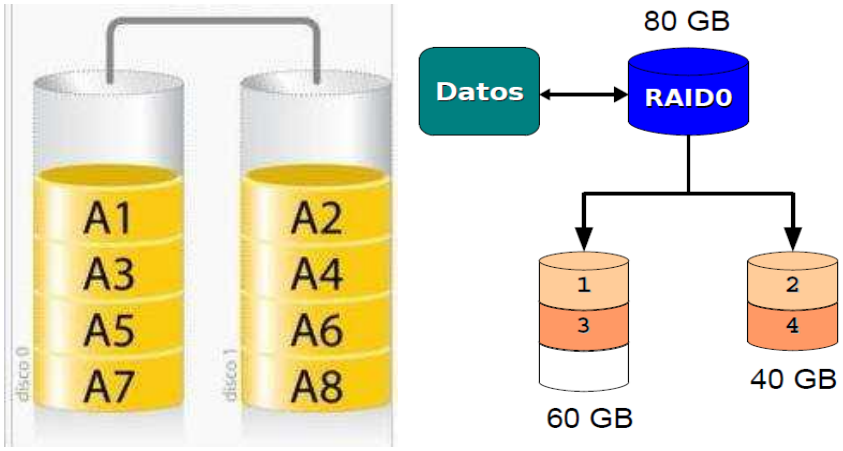


Figura 42. RAID 0.

RAID 1: mirroring

Tiene redundancia. Se basa en la utilización de discos adicionales sobre los que se realiza una copia en todo momento de los datos que se están modificando. Se duplican todos los datos de una matriz en otra, con esto se asegura la integridad en los datos y la tolerancia a fallos. Si el disco primario falla el espejo continúa trabajando. Una vez sustituido el disco averiado, los datos se reconstruyen al 100%. En escritura se pierden prestaciones, al tener que escribir la misma información simultáneamente en dos discos. Por ello, en ocasiones, se utiliza la duplicación de controladoras del disco además de la duplicación de los discos. A este recurso se le denomina DUPLEXING. Dicho recurso resulta caro, ya que requiere instalar en el RAID el doble de la capacidad requerida. En lectura si ofrece mejor rendimiento.

La capacidad total de almacenamiento es igual al tamaño del dispositivo de menor, se necesitan mínimo 2 discos y las unidades se deben añadir a pares.

Su uso se limita a aquellos casos en los que seguridad y continuidad sean fundamentales, como en finanzas, contabilidad, RRHH...

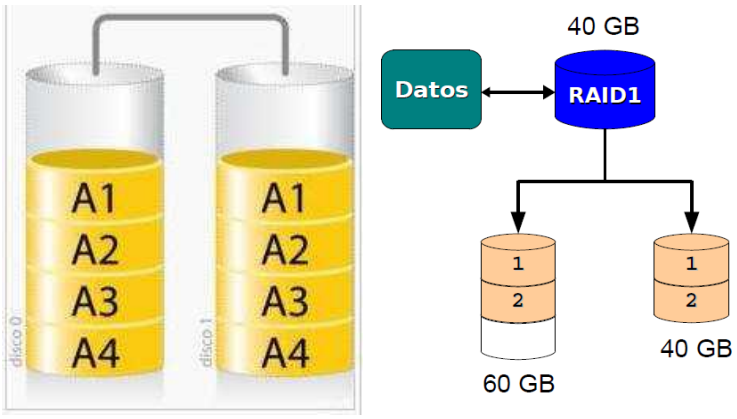


Figura 43. RAID 1.

RAID 2: código Hamming

Divide los datos a nivel de bits en lugar de a nivel de bloques, utiliza código Hamming para la corrección de errores. Los discos son sincronizados por la controladora para funcionar al unísono. Permite tasas de transferencia extremadamente altas, aunque actualmente no se usa, ya que las propiedades de código Hamming restringen configuraciones posibles de matrices para el cálculo de paridad de los discos. La segmentación a nivel de bits crea un alto impacto (lectura/escritura) en los recursos del sistema lo que lo hace inviable a nivel práctico. Necesitaría 39 discos, 32 se usarían para almacenar los bits individuales que forman cada palabra y 7 para la corrección de errores.

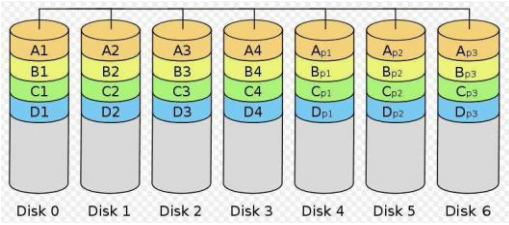


Figura 44. RAID 2.

RAID 3: nivel bytes

Acceso síncrono con un disco dedicado a paridad. Usa división a nivel de bytes. Sirve para disponer de redundancia sin tener que recurrir al mirroring. En el RAID 3 se utilizan X discos de datos y un disco adicional dedicado a la paridad. Los datos son divididos en bytes y cada byte se escribe en uno de los X discos de datos. Aplicando un determinado algoritmo se genera el byte de paridad, que se escribe en el disco de paridad. Toda la información se escribe en los discos de forma paralela. De este modo, la velocidad de transferencia del RAID equivale a la velocidad de transferencia de un disco multiplicada por X. Sin embargo, solo se puede gestionar un E/S a la vez. La recuperación de datos se consigue calculando el O exclusivo (XOR) de la información registrada en los otros discos. Dado que una operación E/S accede a todos los discos al mismo tiempo, el RAID-3 no puede traslapar E/S. Por esta razón, el RAID-3 es mejor para sistemas de un solo usuario con aplicaciones que contengan grandes registros.
Aplicaciones: Producción y distribución de video en streaming, Aplicaciones de imagen, video y Prensa, Servidores de Base de Datos Mono-usuario.
Funcionamiento: Una petición del bloque «A» formado por los bytes A1 a A6 requeriría que los tres discos de datos buscaran el comienzo (A1) y devolvieran su contenido. Una petición simultánea del bloque «B» tendría que esperar a que la anterior concluyese.

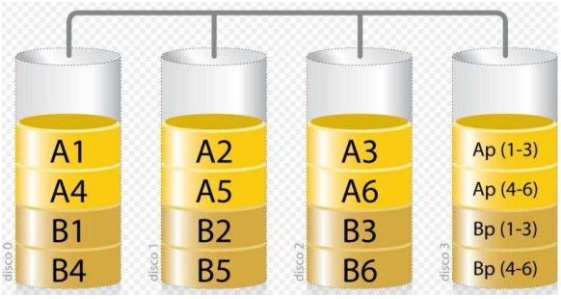


Figura 45. RAID 3.

RAID 4: block striping & dedicated parity

Segmenta los datos a nivel de bloques distribuyéndolos entre los dispositivos del RAID. Es un esquema similar a RAID2 y RAID3 pero la división en bloques evita un alto impacto (lectura/escritura) en los recursos del sistema. Acceso independiente con un disco dedicado a la paridad. Basa su tolerancia a fallos en la utilización de un disco dedicado a guardar la información de paridad calculada a partir de los datos guardados en otros discos.

Se necesita un mínimo de tres unidades. La ventaja frente a RAID 3 es que se puede acceder a los discos de forma independiente.

Una petición del bloque «A1» sería servida por el disco 0. Una petición simultánea del bloque «B1» tendría que esperar, pero una petición de «B2» podría atenderse concurrentemente.

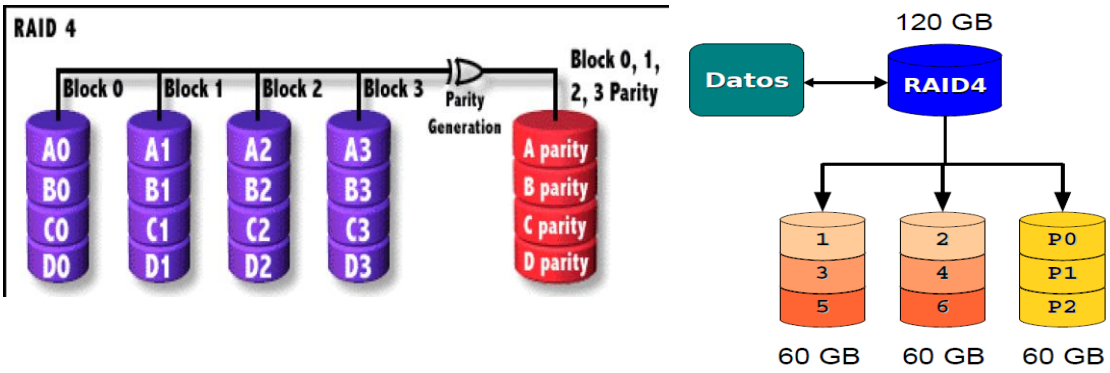


Figura 46. RAID 4.

RAID 5: block striping & distributed parity

Segmenta los datos a nivel de bloques distribuyéndolos entre los dispositivos del arreglo de forma similar a RAID 4. Distribuye los datos de paridad entre todos los dispositivos del arreglo. RAID 4 y RAID 5 proveen redundancia ante la falla de un dispositivo en base a la información de paridad. La información del usuario se graba por bloques y de forma alternativa en todos ellos. Elimina el problema de RAID 4 al no asignar un disco específico para la comprobación, sino que asigna un bloque alternativo de cada disco a esta misión de escritura.

Este es el nivel más eficaz y el de uso preferente para las aplicaciones de servidor en la empresa. Ofrece la mejor relación rendimiento- coste en un entorno con varias unidades. Difiere del RAID 3 en que la información se divide en bloques en vez de en bytes. Debido a ello las lecturas pueden ser independientes, mejorando por tanto el número de transacciones E/S que puede gestionar el RAID de forma simultánea. La otra gran diferencia con el RAID 3 radica en el hecho de que la información de la paridad se reparte entre todos los discos de forma ROTATORIA, aliviando así el cuello de botella que se forma en operaciones simultáneas de E/S. Esto ocurre, en concreto, gracias a que solo se debe leer la paridad de un único disco, como es

el caso del RAID 3. Con lo que resuelve las limitaciones de escritura en RAID-4. Así, todas las operaciones de lectura y escritura pueden superponerse.
Gracias a la combinación del fraccionamiento de datos y la paridad como método para recuperar los datos en caso de fallo, constituye una solución ideal para los entornos de servidores en los que gran parte de E/S es aleatoria, la protección y disponibilidad de los datos es fundamental y el coste es un factor importante. Esta especialmente indicado para trabajar con sistemas multiusuario y también recomendable para aplicaciones que trabajen con ficheros pequeños pero con un gran número de transacciones E/S, como es el caso de las bases de datos relacionales o las aplicaciones de gestión. Aplicaciones: Servidores de archivos y aplicaciones, Servidores de Base de Datos, Servidores www, correo, noticias, Servidores de Intranet.
Se necesita un mínimo de tres unidades de disco para implementarlo, aunque su resultado óptimo de capacidad se obtiene con siete o más unidades.

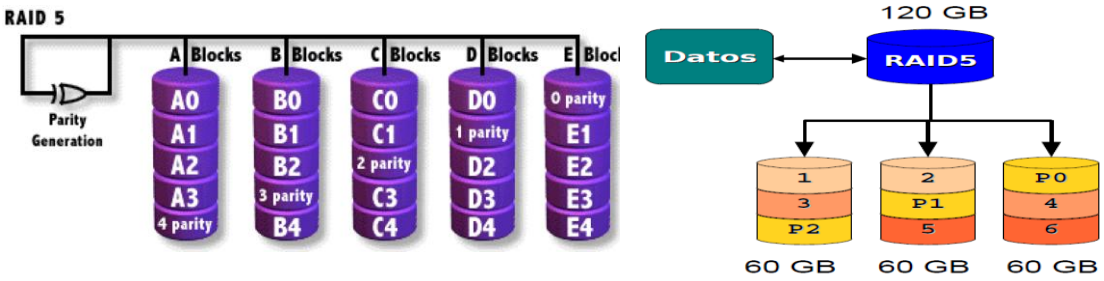


Figura 47. RAID 5.

RAID 6: block striping & doble distributed parity
Acceso independiente con doble paridad. Parecido al RAID 5, pero incluye un segundo esquema de paridad distribuido por los diferentes discos, por lo que ofrece tolerancia extremadamente alta a los fallos y las caídas de disco, ofreciendo dos niveles de redundancia. Su coste de implementación es muy alto por eso casi no se implementa. Un segundo conjunto de datos de paridad lo que provee redundancia ante la falla de dos dispositivos. Hay pocos ejemplos comerciales en la actualidad. Posee un diseño muy complejo, un rendimiento pobre y se necesitan N+2 discos.

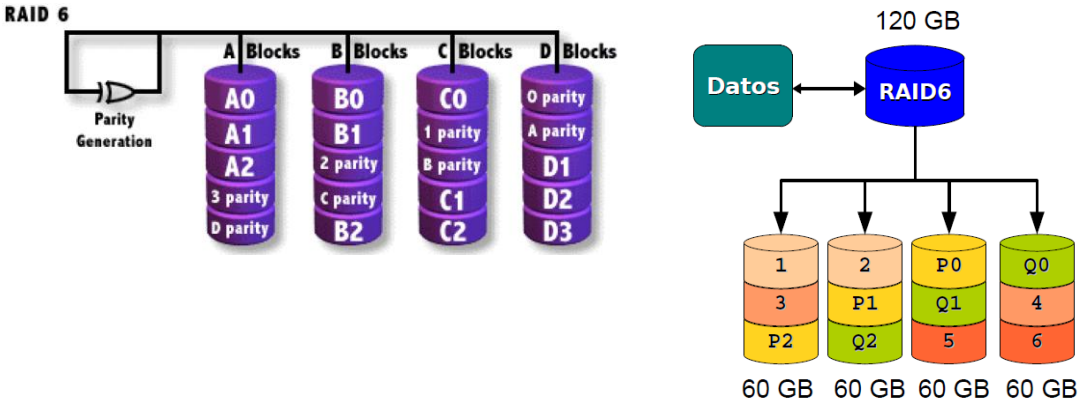


Figura 48. RAID 6.

RAID 7
Este tipo incluye un sistema operativo incrustado de tiempo real como controlador, haciendo las operaciones de caché a través de un bus de alta velocidad y otras características de un ordenador sencillo. Todas las transferencias son asíncronas. Y las E/S están centralizadas

por la caché. Se necesita un disco de paridad exclusivo. El agente SNMP permite su administración remota. Aplicaciones: Sistemas de Tiempo Real (Industriales).

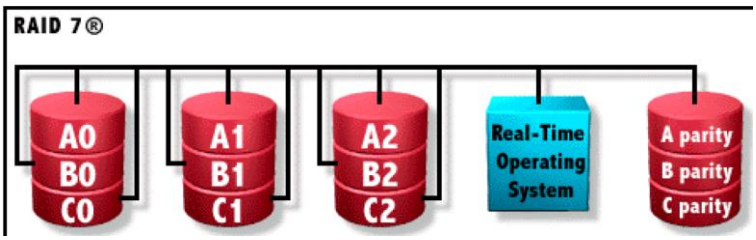


Figura 49. RAID 7.

RAID 0+1: stripping & mirroring

Consiste en la duplicación de los datos en diferentes conjuntos de discos, para un posterior stripping dentro de cada uno de dichos conjuntos. Utiliza un poco de cada uno lo que proporciona velocidad y tolerancia a fallos. Fracciona los datos para mejorar el rendimiento mientras que utiliza un conjunto de discos duplicados para conseguir redundancia de datos. Requiere un mínimo de cuatro unidades, sólo dos de ellas se utilizan para el almacenamiento de datos.

Este nivel está indicado para aplicaciones que necesiten altas prestaciones y un alto nivel de seguridad. Aplicaciones: Aplicaciones de Imagen, Servidores de archivos generales.

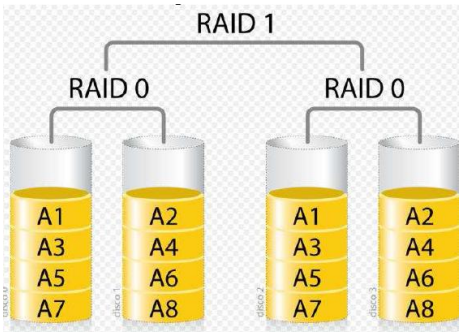


Figura 50. RAID 0+1.

RAID 1+0: mirroring & striping

Este nivel ofrece un conjunto de bandas en el que cada banda es un grupo de discos RAID-1. Esto proporciona mejor rendimiento que el RAID-1, pero a un costo mucho mayor. A pesar de su alto costo es muy utilizado. Combina el rendimiento (velocidad de acceso) de la segmentación de datos (striping) con las propiedades de la redundancia (mirroring). Si un dispositivo falla ambos lados del RAID 10 seguirán funcionando (aunque un lado en modo degradado).

Aplicaciones: Servidores de Bases de Datos con alto rendimiento y tolerancia a fallos.

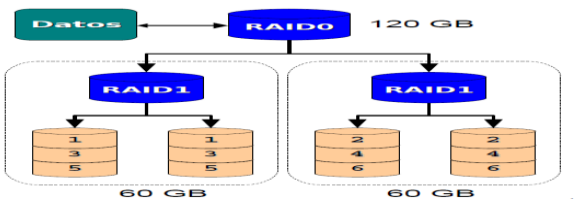


Figura 51. RAID 1+0.

RAID 5+0: striping & parity

En referencia a un RAID 10:

- Menor costo.

forjaTIC

- Rendimiento de lectura más bajo (aunque sigue siendo bueno).
- A nivel de escritura hay mayor rendimiento (velocidad).
- Cada RAID 5 puede soportar la falla de un dispositivo.

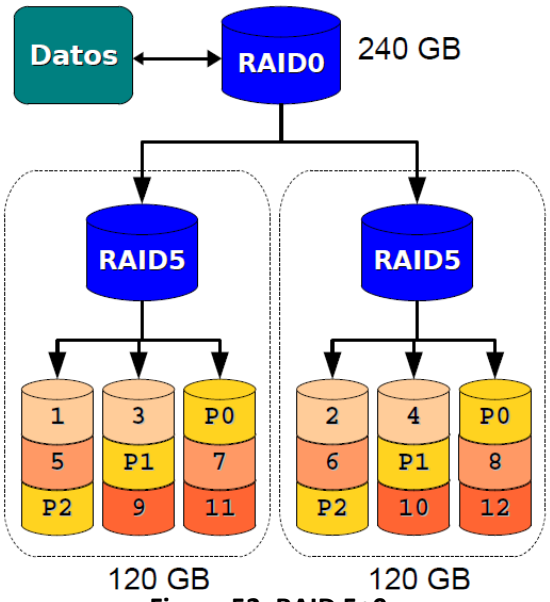


Figura 52. RAID 5+0.

RESUMEN NIVELES:

Nivel	Tipo	Dispositivos	Redundancia	Capacidad	Rendimiento (lectura)	Rendimien (escritura
RAID 0	*Striping (block level)*	N > 1	0	1	N	N
RAID 1	*Mirroring*	N > 2	N - 1	1/N	N	1
RAID 4	*Striping (block level) Parity (dedicated)*	N > 2	1	1 - 1/N	N - 1	N - 1
RAID 5	*Striping (block level) Parity (distributed)*	N > 2	1	1 - 1/N	N - 1	N - 1
RAID 6	*Striping (block level) Double Parity (distributed)*	N > 3	2	1 - 2/N	N - 2	N - 2

Tasa de fallas *(fail rate)*	Posibles Aplicaciones
$1-(1-r)^N$	Almacenamiento de archivos grandes que no requieren redundancia en tiempo real.
r^N	Bases de datos y archivos de bajo contenido dinamico (poca capacidad)
$N(N-1)r^2$	Bases de datos, servidores de archivos, correo electronico, contenido.
$N(N-1)r^2$	Bases de datos, servidores de archivos, correo electronico, contenido.
$N(N-1)(N-2)r3$	Bases de datos, servidores de archivos, correo electronico, contenido (mayor tolerencia a fallas que RAID 5)

N = Cantidad de dispositivos en el grupo de almacenamiento

r = % de error estimado por cada dispositivo del grupo de almacenamiento.

VIP!!! Sobre los niveles RAID, para calcular capacidad:
RAID 0= N*(CAPACIDAD DEL MENOR MIEMBRO RAID)
RAID 1= N/2*(CAPACIDAD DEL MENOR MIEMBRO RAID)
RAID 5= (N-1)*(CAPACIDAD DEL MENOR MIEMBRO RAID)
RAID 6= (N-2)*(CAPACIDAD DEL MENOR MIEMBRO RAID)

5.5.4. Implementación en Linux

Verificando que dispositivos están instalados y como están distribuidos:
 # fdisk /dev/sda
Usando herramientas de RAID en Linux, herramienta **MDADM**:

forjaTIC

 # apt-get install mdadm
Creando un arreglo:
*# mdadm -C /dev/md0 -a yes -l **0** -n **2** /dev/sda10 /dev/sda11*
Dando formato a un arreglo:
 # mkfs.ext3 /dev/md0
Definiendo en el arranque:
 # vim /etc/fstab
/dev/md0 /usr2 ext3 defaults 0 0
 # reboot
Monitoreo de los arreglos:
 # cat /proc/mdstat
 # mdadm –-detail /dev/md0
Chequeando el arreglo:
Simulando una falla en un disco:
 # mdadm /dev/md0 -f /dev/sda1
Restableciendo una falla en un disco:
 # mdadm /dev/md0 -r /dev/sda1
Mas documentación
 # man mdadm

5.5.5. JBOD (Just a Bunch of Disks)

JBOD es el acrónimo de Just a Bunch of Disks (Un mero puñado de discos). En esta configuración, cada unidad se considera un dispositivo independiente, de forma similar a las unidades de disco duro independientes. Todos los discos aparecerán en su ordenador como discos duros individuales, divididos en las particiones creadas por la utilidad de gestión de discos del sistema operativo. Una gran ventaja de JBOD es el uso de la capacidad completa de ambas unidades. Asimismo, en caso de fallo de una de las unidades, no se pierden los archivos almacenados en la unidad correcta. Sin embargo, esta configuración no proporciona ninguna de las ventajas de RAID 0 (velocidad) ni RAID 1 (protección de datos completa).
La capacidad total de una matriz JBOD es la suma de capacidades de cada uno de sus discos, o C = n*d donde: C = capacidad disponible n = número de discos d = capacidad de disco. Por ejemplo, en una matriz JBOD con cuatro unidades de 1000 GB de capacidad cada una, la capacidad total de la matriz sería de 4000 GB: C = (4*1000).

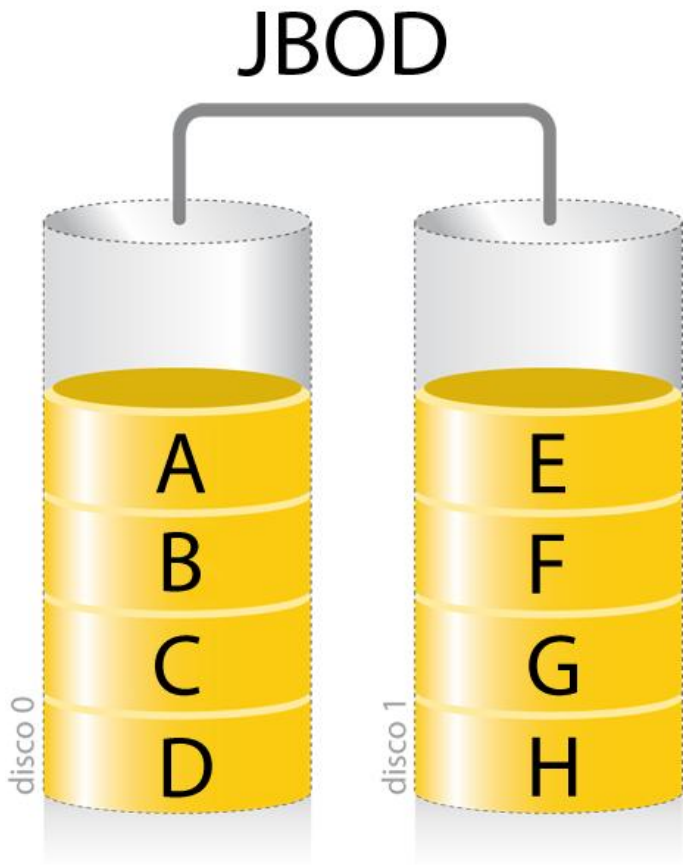


Figura 53. JBOD.

5.6. Arquitecturas de almacenamiento y de respaldo

5.6.1. Introducción

Después de describir la tecnología subyacente a los sistemas de almacenamiento (discos y cintas), en este apartado se tratarán las arquitecturas de almacenamiento y de respaldo del mismo.

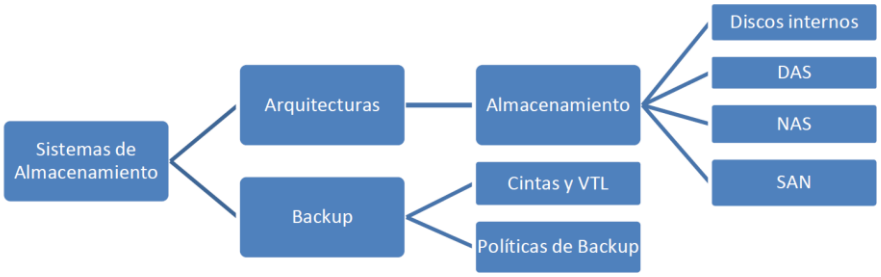


Figura 54. Esquema de los sistemas de almacenamiento.

La primera diferenciación necesaria al referirse a recursos de almacenamiento es la **unidad en la que se sirve este almacenamiento:**

- Si el cliente requiere acceso a disco, éste se le presenta directamente con discos internos, por medio de una cabina dedicada (DAS) o mediante una red específica de almacenamiento (SAN). Este tipo de acceso se denomina también acceso a bloques
- La otra opción es que el cliente trabaje a nivel de archivo. En este caso el sistema de ficheros es creado y gestionado por el sistema operativo del servidor (NAS), que a su vez tendrá

asociado un determinado espacio en discos. El cliente o aplicación, compatible con ese sistemas de ficheros, hará las peticiones de los mismos al NAS.
En el siguiente diagrama se muestran los principales elementos de cada una de las arquitecturas de almacenamiento señaladas, que se desarrollarán a continuación.

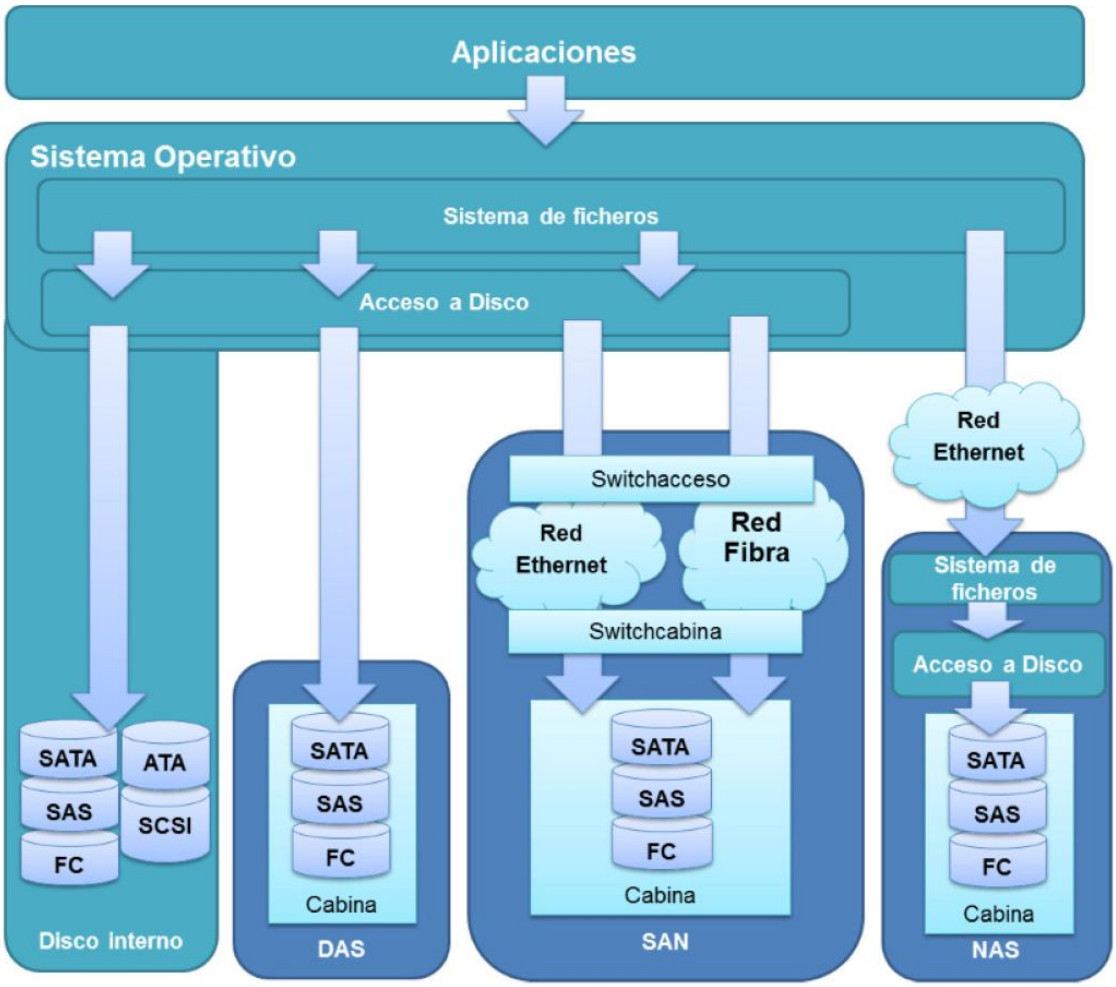


Figura 55. Esquema con los principales elementos de arquitectura de almacenamiento.

5.6.2. Discos internos

El modo más básico de conectividad del almacenamiento a los servidores es el almacenamiento interno, en el que los discos están conectados a un bus interno del servidor. Este modelo de conectividad presenta dos variantes, la conectividad basada en el bus SCSI para sistemas empresariales y la basada en el bus ATA para sistemas domésticos, cada uno de ellos con características distintas, siendo el SCSI más avanzado y consecuentemente más complejo y caro.
Ambos buses son de tipo paralelo, lo que conlleva ciertas dificultades y limitaciones que han llevado a la aparición de conexiones serie en ambos tipos de bus, SAS como evolución de SCSI y SATA como evolución de ATA. Ambas evoluciones modifican las capas bajas de los protocolos pasando a un protocolo serie en lugar de paralelo pero conservan las capas altas, de forma que aseguran la compatibilidad a nivel software.

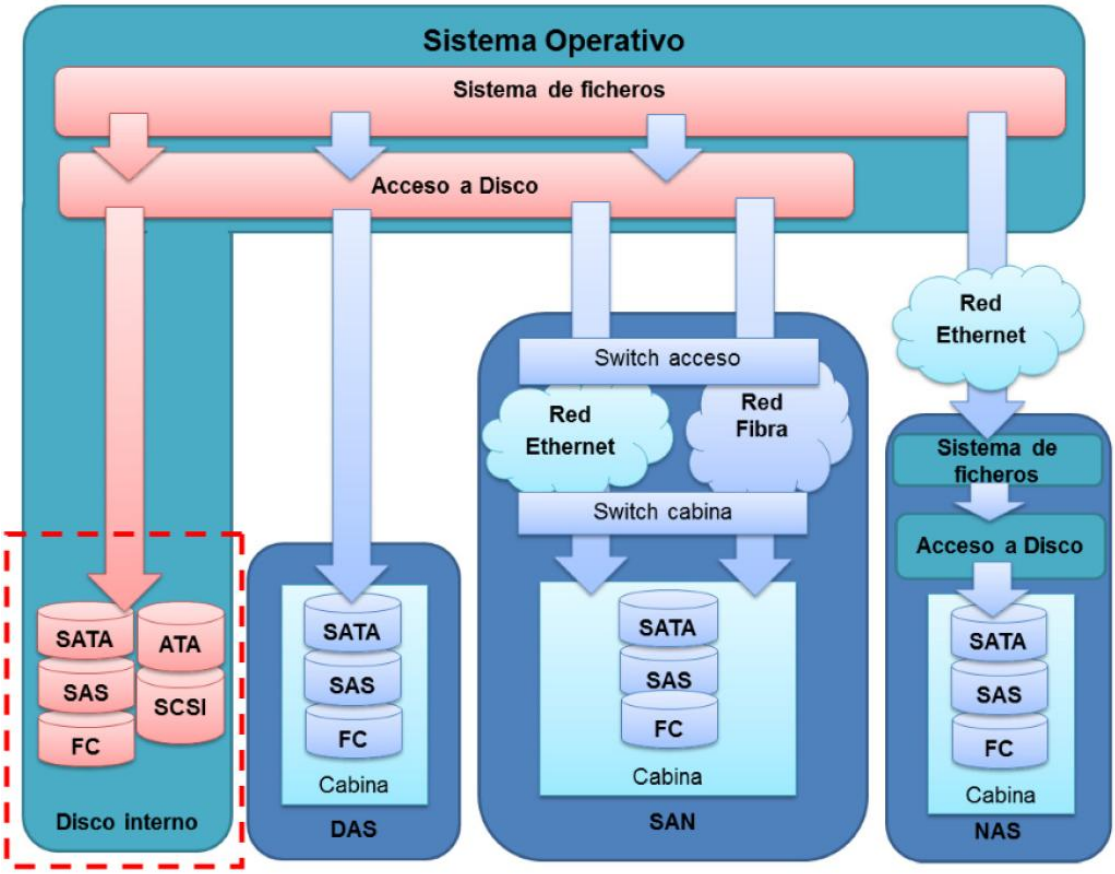


Figura 56. Solución de almacenamiento basada en discos internos.

El almacenamiento interno conlleva una serie de problemas que hacen preferible en un entorno empresarial el almacenamiento externo. Entre los **problemas** de esta opción se encuentran:

- la falta de **escalabilidad**, ya que una vez consumidas las posibilidades de expansión del servidor no es posible ampliar el almacenamiento
- la dificultad de una **ampliación** de capacidad, ya que implica conexiones internas al servidor,
- la dificultad de **protección** de los datos, ya que en general el almacenamiento interno no cuenta con las capacidades de replicación y la redundancia de las cabinas de discos de almacenamiento externo,
- la **infrautilización** de la capacidad, ya que el almacenamiento sobrante en un servidor no puede ser utilizada por otro servidor

Los datos de aplicaciones con un mínimo de criticidad no deberían ubicarse en los discos internos del servidor, quedando éstos relegados a alojar la instalación del sistema operativo y de la aplicación.

Sin embargo, hasta el uso de esta opción se está limitando aún más debido a la virtualización de servidores, puesto que también la instalación reside en una red separada de almacenamiento, en forma de máquina virtual que puede ubicarse en un servidor u otro de manera rápida.

5.6.3. Direct attached storage (DAS)

La conectividad más común entre servidores y almacenamiento está basada en SCSI o en FC, que sustituye las capas bajas de SCSI por el protocolo serie FC y utiliza las capas altas de SCSI para enviar comandos de almacenamiento.
El protocolo SAS también sustituye la capa física paralela de SCSI por un protocolo serie y puede ser utilizado para conectar las cabinas de almacenamiento directamente a los servidores. Desde el punto de vista del software del servidor, al usar siempre las capas altas del protocolo SCSI, el almacenamiento externo es utilizado igual que el almacenamiento interno, sin que las capas de S.O. más allá del driver que gestiona la conectividad distingan entre almacenamiento interno y externo.
El uso del almacenamiento externo dedicado resuelve los problemas de escalabilidad y seguridad del almacenamiento interno, ya que en general las cabinas de discos son ampliables y tienen una capacidad máxima muy superior, al poder albergar muchos más discos. Además presentan características de alta disponibilidad, como redundancia de componentes, niveles de RAID en espejo o con paridad para recuperarse del fallo de un disco, software de copia de datos que permite replicar los datos en tiempo real en otra cabina distinta, etc.

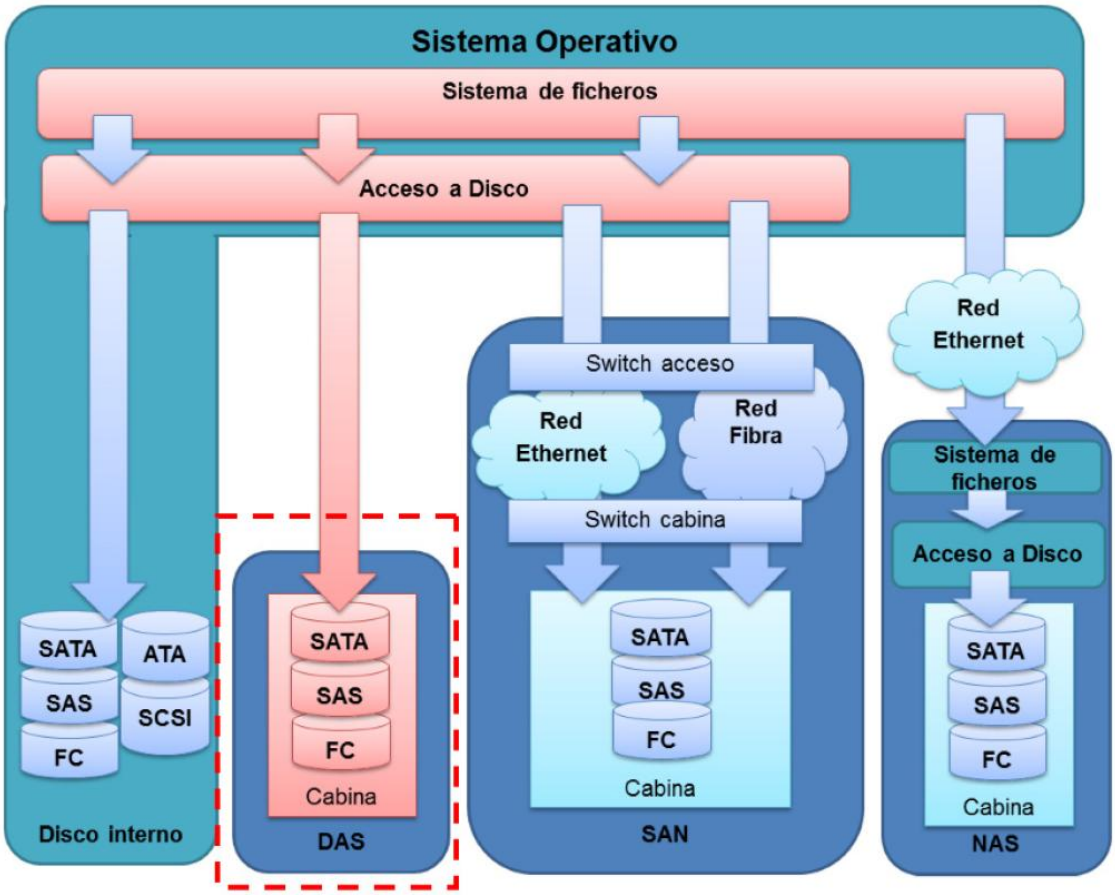


Figura 57. Solución de almacenamiento basada en DAS (discos asignados directamente).

El rendimiento de las cabinas de discos es superior al del almacenamiento interno, ya que las cabinas están diseñadas exclusivamente para ello e incluyen caché interna y controladores sofisticados para gestionar los discos y las colas de I/O que no están presentes en los servidores.
A pesar de solventar algunos de los problemas del almacenamiento interno, el almacenamiento externo dedicado aún presenta algunas deficiencias, como la **infrautilización de la capacidad de almacenamiento**, ya que al estar conectado directamente a un servidor no se puede utilizar la capacidad sobrante en otro servidor o la creación de clústeres, ya que si bien las cabinas pueden permitir la conexión simultanea de varios servidores y la compartición

de discos lógicos entre ellos, la **escalabilidad** de los mismos (número de servidores que forman el clúster) está limitada por la conectividad física de la cabina.

5.6.4. Network attached storage (NAS)

A diferencia de lo visto hasta este punto, en soluciones NAS se realiza acceso a ficheros directamente contra el servidor de ficheros o la cabina de almacenamiento con capacidad NAS donde se gestiona el sistema de ficheros.
El S.O. del servidor "monta" un sistema de ficheros externo a través de alguno de los protocolos destinados a ese fin (NFS, CIFS, Novell, etc) y delega la transformación de fichero a bloque de bits en la red de almacenamiento.

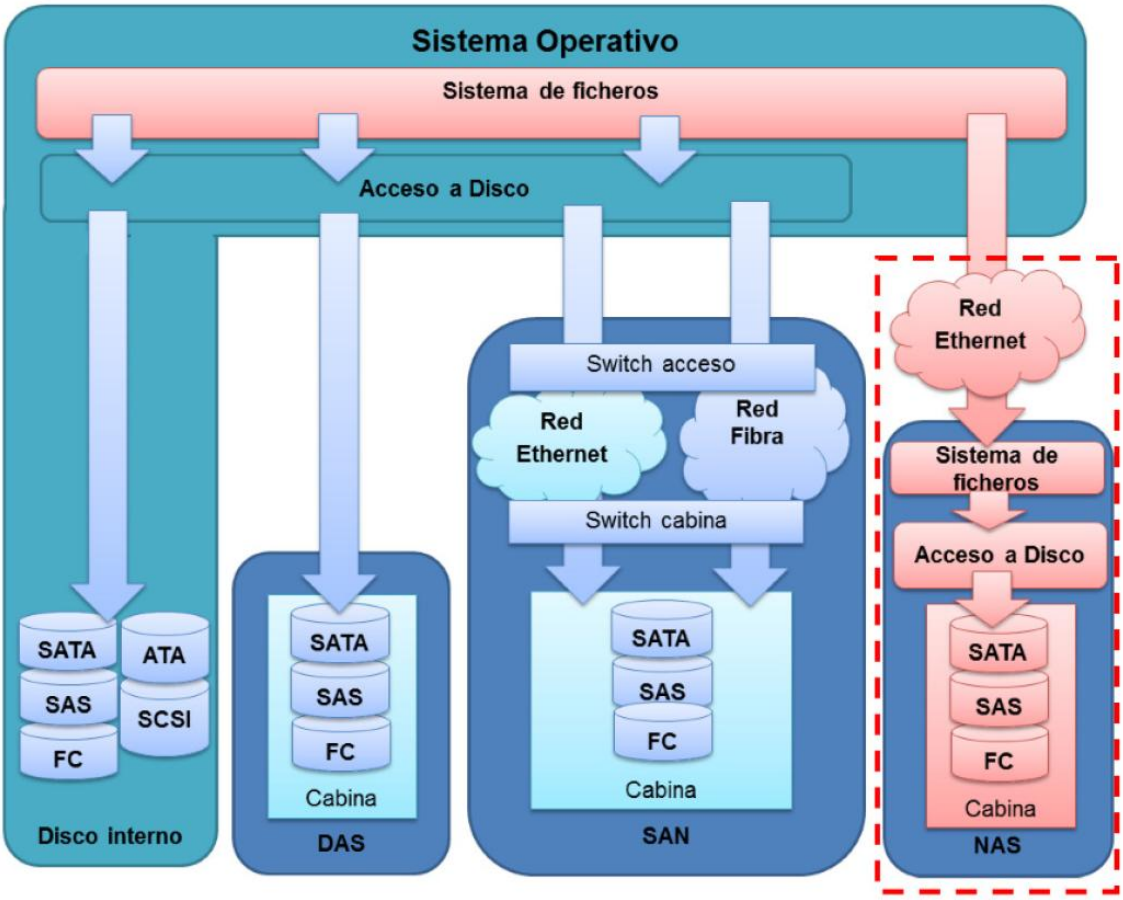


Figura 58. Solución de almacenamiento basada en NAS (servidor de ficheros).

Red LAN
Los sistemas de acceso a ficheros solventan el problema del acceso a los datos desde diferentes orígenes delegando la gestión del sistema de ficheros en el equipo de red que sirve los datos. Los servidores montan los sistemas de ficheros exportados (por ejemplo, unidades de red en sistemas Windows montadas a través de CIFS, montaje por NFS en sistemas UNIX/Linux, etc) que son utilizados por las aplicaciones como si fuesen discos locales pero que para el sistema operativo son discos remotos con el sistema de ficheros delegado en otro sistema.
Los sistemas que actúan de servidores de ficheros pueden ser servidores de propósito general en los que se ejecuta un software (que puede ser parte o no del sistema operativo) que actúa como servidor de ficheros o puede tratarse de dispositivos especiales, denominados NAS, dedicados en exclusiva a servir ficheros y diseñados especialmente para ello.

El servidor de ficheros o NAS tiene como principales funciones la implementación de los protocolos que permiten compartir los ficheros y la gestión de la estructura de ficheros. El almacenamiento final de los datos se realiza lógicamente mediante bloques de bits que por lo tanto pueden residir tanto internamente en el servidor de ficheros como externamente en una cabina de la red SAN.

Red WAN

Los protocolos de compartición de ficheros están diseñados para utilizarse en una red LAN, con gran ancho de banda y baja latencia, por lo que su uso a través de redes WAN donde el ancho de banda es menor y, especialmente, la latencia es mayor presenta problemas que pueden hacer inviable la compartición de ficheros a grandes distancias (por ejemplo con oficinas remotas desde un servidor central consolidado).

Un número elevado de mensajes entre cliente y servidor puede provocar problemas cuando la latencia de la red es elevada, ya que el tiempo necesario para acceder a un fichero será la suma del tiempo de ida y vuelta de todos los mensajes necesarios.

Para evitar estos problemas está surgiendo el concepto de WAFS (Wide Area File Services), basado en la compresión de las tramas de datos de los protocolos CIFS y NFS, la simplificación de ambos protocolos para evitar un excesivo número de mensajes y el uso de caché local en las oficinas remotas.

5.6.5. Storage area network (SAN)

Una SAN es una red dedicada al almacenamiento que está conectada a las redes de comunicación de una organización. Además de contar con interfaces de red tradicionales, los equipos con acceso a la SAN tienen una interfaz de red específica que se conecta a la SAN.

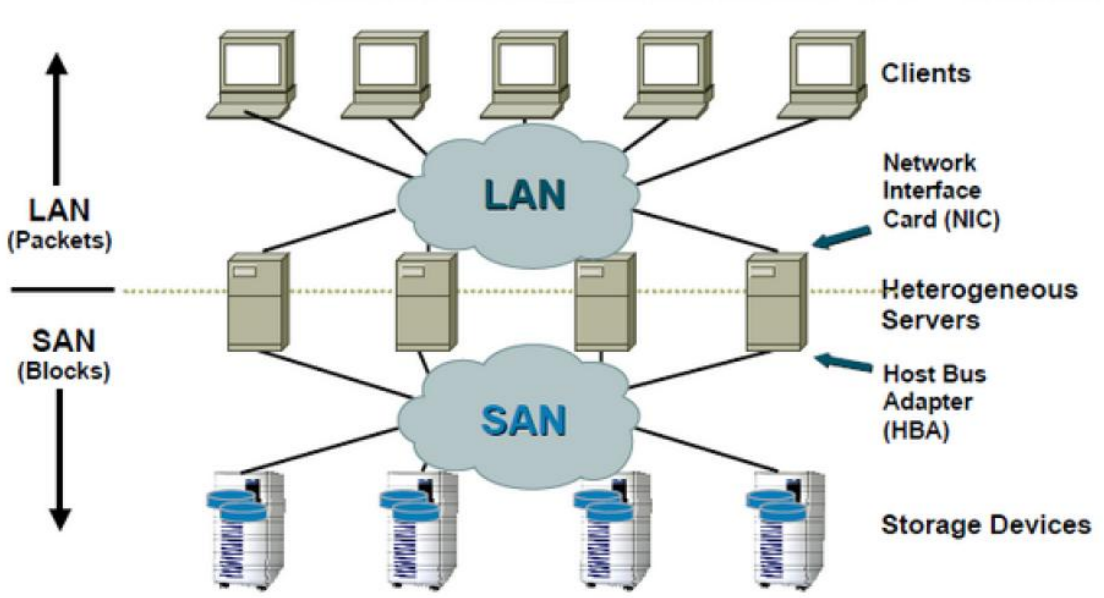


Figura 59. Conexión entre red LAN y SAN.

Se trata de una arquitectura de acceso a bloques en la que los protocolos de acceso a disco que circulan por la red de transporte son los mismos que se utilizan en las conexiones directas con los discos o cabinas dedicadas. De esta forma, los drivers dedicados al acceso a la red de almacenamiento ocultan al resto del S.O. el hecho de que los discos no están conectados internamente al servidor.

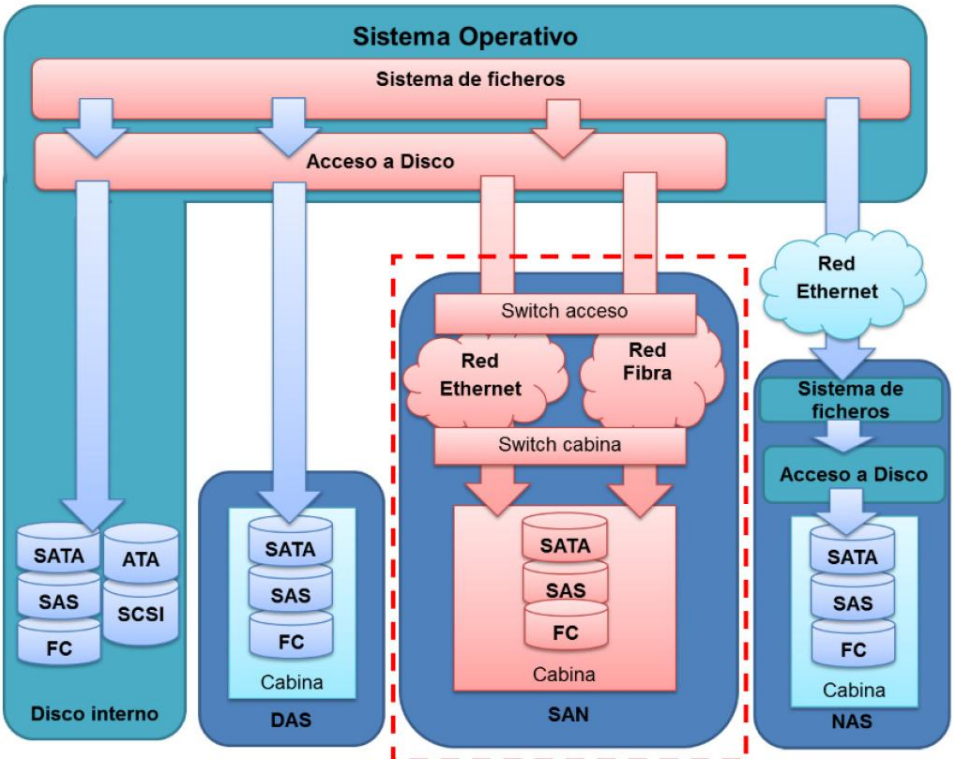


Figura 60. Solución de almacenamiento basada en SAN (red de almacenamiento).

Al tratarse de un acceso por bloques, el sistema de ficheros es creado por el servidor, por lo que servidores independientes no deberían poder acceder a los mismos bloques, ya que ambos tratarían de mantener el sistema de ficheros, causando la corrupción del mismo. Por tanto, este tipo de almacenamiento no permite el acceso de varios servidores a los mismos datos, a no ser que estos formen parte de un clúster de y se coordinen para el acceso a los datos.

Por tanto, con respecto a una solución NAS, las diferencias se muestran en la siguiente ilustración.

NAS vs. SAN	
NAS	**SAN**
Los ordenadores se conectan a los dispositivos de almacenamiento directamente a través de la LAN generalmente mediante los protocolos NFS, CIFS, FTP o TFTP. Entre los clientes y las unidades de almacenamiento se encuentran las llamadas 'cabeceras NAS'.	Al ser una red dedicada conectada a la LAN, los clientes acceden a los dispositivos de almacenamiento a través de servidores mediante protocolos SCSI o Fibre Channel.
Opera a nivel de fichero de datos.	Opera a nivel de bloques de datos
El sistema de ficheros reside en el dispositivo de almacenamiento. El cliente solicita el archivo completo al servidor y lo maneja localmente. Esto permite dar servicio a ordenadores y sistemas operativos diferentes.	El sistema de ficheros reside en el cliente, por lo que –en principio- no permite trabajar con ordenadores y sistemas operativos diferentes. Cualquier solución a este inconveniente añade costes y complejidad.

Figura 61. Diferencias fundamentales entre soluciones NAS y SAN.

La unidad de almacenamiento que provee una red SAN se denomina **LUN** (Logical Unit Number) y es un disco virtual proporcionado por la SAN. El servidor tiene el mismo acceso y derechos sobre la LUN que los que tendría si fuera un disco directamente conectado. Por tanto, puede particionar y formatear el disco en cualquier formato que requiera.
Las SAN se componen de **tres capas**:
• **Capa Cliente**. Esta capa consiste principalmente en Servidores, dispositivos o componentes (interfaz de red, de fibra HBA…) y software (sistemas operativos).
• **Capa Red**. Esta capa la conforman los cables (Fibra óptica) así como los SAN Hubs y los SAN switches como punto central de conexión para la SAN. Otra posibilidad es la utilización de redes basadas en Ethernet
• **Capa Disco**. Esta capa la componen las agrupaciones de discos (Disk Arrays, Memoria Caché, RAIDs) y cintas empleados para almacenar datos.

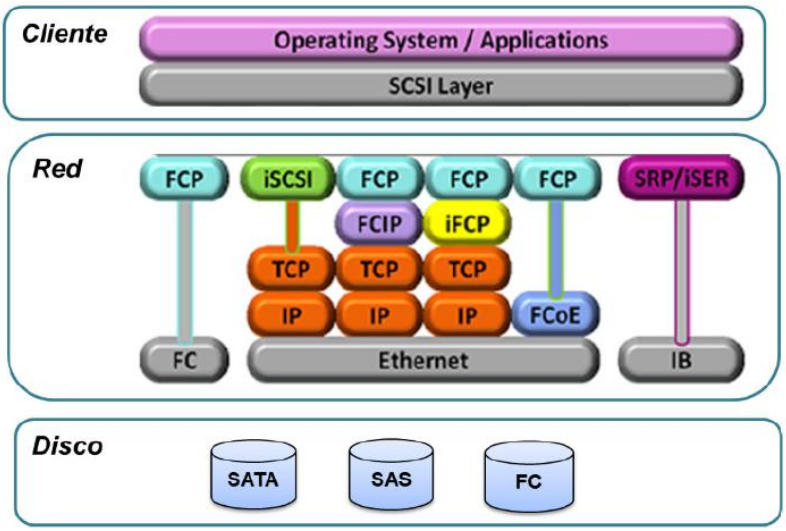


Figura 62. Capas y estándares en soluciones de almacenamiento.

La red de almacenamiento puede ser de varios tipos, lo cual va a originar los diferentes estándares existentes.

SAN fibrechannel (SAN FC)
Las redes SAN FC están formadas por un conjunto de switches que conmutan las tramas FC permitiendo la conectividad de los servidores y los sistemas de almacenamiento conectados a los mismos.

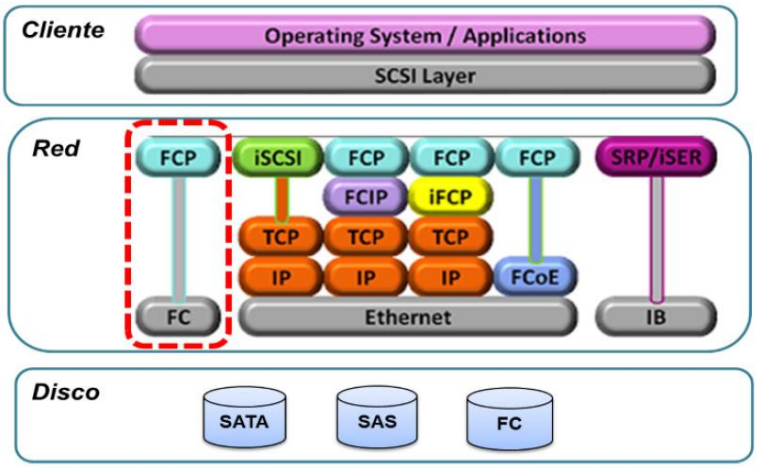


Figura 63. Protocolos utilizados en redes SAN FC.

Los equipos pertenecientes a una red FC se identifican mediante un WWN (World Wide Name), dirección de 64 bits asignada unívocamente a cada dispositivo.
Cuando el dispositivo realiza su registro en la red, esta le asigna un identificativo de 24 bits (FCID) que utilizará el dispositivo en sustitución de su WWN cada vez que envíe una trama por la red. De esta forma se optimiza la conmutación de las tramas, al utilizar un identificativo FCID más pequeño que el WWN y asignado por los conmutadores de la red en función de la topología de la misma y no utilizar un WWN asignado al equipo en el proceso de fabricación y que lógicamente no corresponderá con la topología de la red.
Con los 24 bits del FCID se pueden direccionar unos 16 millones de dispositivos en una misma red FC. Los switches dentro de una red SAN se identifican mediante el DomainID, que es asignado de forma dinámica al registrase los equipos en la red.
Al igual que en redes LAN, existen diferentes categorías de switches FC, desde pequeños switches de 8 puertos hasta grandes equipos modulares con cientos de puertos (denominados directores en la terminología de FC) y con distintas características de redundancia, prestaciones, disponibilidad, etc.
La conexión entre dos switches se realiza mediante ISL (Inter Switch Links) a través de los cuales los switches se intercambian información sobre la topología de la red.
Dentro de una red FC existen una serie de servicios básicos para su correcto funcionamiento y que son implementados automáticamente de forma distribuida entre todos los switches que componen la red. Estos **servicios** son los siguientes:

- SNS: Servidor que nombres que asigna los FCID y permite traducir de FCID a WWN.
- RSCN: Notificación de cambios en la topología de la red. Informa al resto de equipos de la aparición o desaparición de equipos en la red.
- Login: Servicio al que se registra un equipo al ser conectado a la red FC.
- Servicio de gestión: Servicio que permite la gestión in-band de la red FC utilizando los mismos enlaces FC por los que se transmite el tráfico.
- Servicio de tiempos: Servidor de tiempos. Está definido en los estándares pero no suele ser implementado en los equipos.

Cuando dos switches de FC se conectan pasan automáticamente a formar parte de la misma red SAN, se intercambian la información necesaria para implementar los servicios anteriores así como información de rutado de tramas basada en FSPF (FabricShortestPathFirst), versión de OSPF (Open ShortestPathFirst) adaptada a las redes FC.
La tecnología FC está basada en estándares, pero cada fabricante implementa sus propias variaciones, lo que hace que en general dos switches de distinto fabricante no sean compatibles en modo nativo. Para hacer compatibles switches de distintos fabricantes existe el mono interoperatividad, consistente en rebajar la funcionalidad de los equipos hasta el cumplimiento del estándar.

Internet SCSI (iSCSI)
iSCSI (*Internet SCSI*) es un estándar que permite el uso del protocolo SCSI sobre redes TCP/IP. iSCSI es un protocolo de la capa de transporte definido en las especificaciones SCSI-3. Otros protocolos en la capa de transporte son SCSI Parallel Interface y canal de fibra.
Este tipo de redes SAN permiten el intercambio de tramas SCSI pero sobre TCP/IP, en lugar de sobre FC.
El objetivo es utilizar la infraestructura LAN existente, u otra dedicada exclusivamente a almacenamiento, aprovechando de esta forma los menores costes económicos y el mayor conocimiento de la tecnología frente a FC, que presenta un coste mayor y cuyo conocimiento está menos extendido entre los recursos humanos.

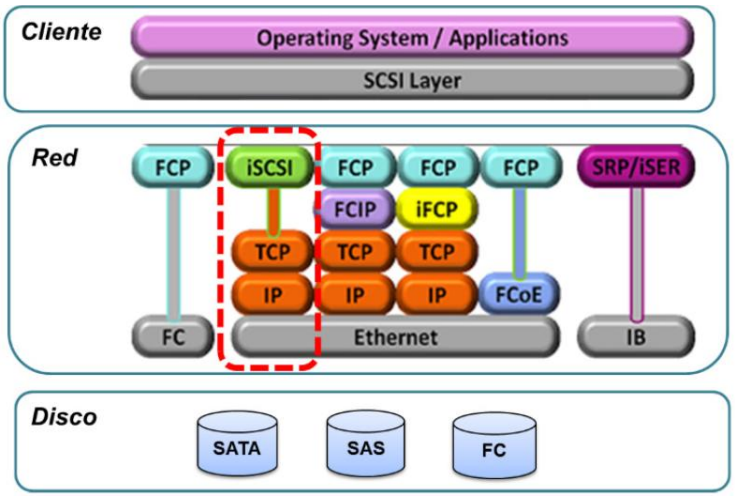


Figura 64. Protocolos utilizados en redes SAN iSCSI.

Al utilizar las capas altas de SCSI igual que FC, a nivel funcional puede ofrecer las mismas capacidades que las redes FC.
Frente al uso de FC, iSCSI presenta la desventaja de utilizar protocolos "genéricos" de transporte, como TCP/IP, que no están tan optimizados para el tipo de tráfico requerido en redes de almacenamiento, por lo que a igualdad de ancho de banda "bruto" para el transporte las redes iSCSI son menos eficientes que las redes FC. Por el contrario, la evolución tecnológica de las redes ethernet, con la aparición de interfaces a 10Gbps antes que las interfaces FC a 8Gbps, puede hacer que esta menor eficiencia se solvente disponiendo de mayor ancho de banda a un precio más económico.
En principio es posible implementar iSCSI en software, haciendo uso de tarjetas NIC Ethernet normales, pero de esta forma aumenta el consumo de CPU de los servidores y disminuye el rendimiento de iSCSI. Para solventar este problema es necesario utilizar tarjetas de red similares a las HBAs (Host Bus Adapter) de FC, que implementen en hardware tanto el protocolo iSCSI como el TCP/IP, liberando al servidor de estas tareas y aumentando el rendimiento, a costa de perder algo de la ventaja competitiva en costes, al ser estas tarjetas más caras que las NIC ethernet normales.

Fibre channel sobre ethernet (FCoE)
Fibre Channel sobre Ethernet (FCoE) permite consolidar una estructura Ethernet para satisfacer las necesidades de las redes de área local y de almacenamiento. FCoE ha sido diseñado sin la capa Protocolo de Internet (IP), no se puede enrutar de forma intrínseca mediante IP.

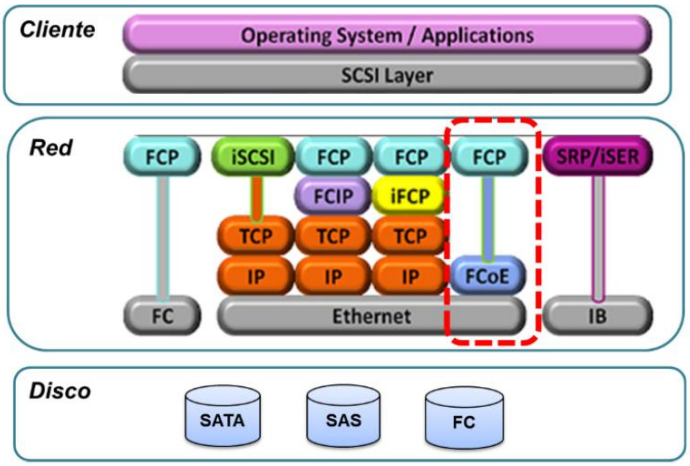


Figura 65. Protocolos utilizados en redes SAN FCoE.

Sin embargo, esto también significa que existen dos opciones de SAN que puede usar en redes Ethernet: FCoE e iSCSI. Los requisitos infraestructurales de FCoE son más rigurosos que los de iSCSI, lo que puede influir a la hora de elegir definitivamente un protocolo. En algunos casos, la elección de un protocolo puede verse decisivamente influida por la infraestructura que admiten los proveedores de las aplicaciones fundamentales.
La opción iSCSI es interesante si los principales objetivos son:

- Reducción de costes, dado que se podrá reutilizar la infraestructura TCP/IP existente.
- Facilidad de uso, dado que la gestión de red resultará más familiar, lo que facilita la configuración y la gestión.

FCoE no utiliza TCP/IP. La gestión es muy similar a la de una SAN FC tradicional, lo que supondrá retos de gestión si no existen infraestructuras previas de una SAN FC. En general, la opción FCoE es interesante si los objetivos son:

- Disponibilidad.
- Máximo rendimiento.

SAN para mainframe (ESCON y FICON)
En el entorno Mainframe de IBM apareció a mediados de los 90 la tecnología **ESCON** como sustituta de la antigua tecnología de conexión paralela por cables de cobre 370.
ESCON introdujo los canales de fibra con un interfaz serie semidúplex con tasas de transferencia de 20 MBs.
La tecnología ESCON permite conexiones punto a punto directas, en la que el almacenamiento estaría dedicado a un servidor, y conexiones punto a punto conmutadas, en las que los distintos servidores y cabinas de almacenamiento están conectadas a un conmutador ESCON Director, que puede establecer de forma dinámica conexiones dedicadas entre ellos. La tecnología ESCON no permite la interconexión de conmutadores, por lo que una red está formada por un único conmutador al que se conectan los servidores y el almacenamiento.
Como evolución de la tecnología ESCON, IBM introdujo la tecnología **FICON** que utiliza como medio de transporte FibreChannel, pero no utiliza en capas superiores SCSI como las redes FC de sistemas abiertos, sino la tecnología propietaria de IBM.
Los canales FICON permiten una tasa de transferencia de 100 MBytes por segundo en canales de 1Gbps y de 200 MBytes por segundo en canales FibreChannel de 2Gbps.
La tecnología FICON permite la interconexión de FICON Director mediante ISLs creando redes SAN con más de un conmutador, pero el camino entre el servidor y los discos solo puede contener dos directores.
Al igual que FC, FICON es una tecnología dúplex, no semidúplex como en el caso de ESCON.

Seguridad en redes SAN
La seguridad en las SAN ha sido desde el principio un factor fundamental, desde su creación se notó la posibilidad de que un sistema accediera a un dispositivo que no le correspondiera o interfiriera con el flujo de información.
Es por ello que se ha implementado la tecnología de **zonificación**, la cual consiste en que un grupo de elementos se aíslen del resto para evitar estos problemas, la zonificación puede llevarse a cabo por hardware, software o ambas, siendo capaz de agrupar por puerto o por WWN (World Wide Name), una técnica adicional se implementa a nivel del dispositivo de almacenamiento que es la **presentación**, consiste en hacer que una LUN (Logical Unit Number) sea accesible sólo por una lista predefinida de servidores o nodos.
A continuación se detallan ambos conceptos:

- **Zonificación (Zoning)**

Las redes SANs de fibrechannel pueden interconectar miles de dispositivos entre sí. Los conmutadores FC tienen la capacidad de segmentar la red SAN en distintas zonas, en base al WWN (World Wide Name) de los dispositivos (cabinas de discos o servidores) de forma que cada dispositivo solo es capaz de comunicarse con el resto de dispositivos que pertenezcan a

su misma zona. Cada zona se comporta como si en la red SAN solo existiesen los dispositivos que forman parte de ella.
Por lo tanto, para que un servidor pueda acceder a una LUN en una cabina, el servidor y la cabina deben estar en la misma zona y adicionalmente la cabina debe permitir el acceso del servidor a dicha LUN.

• **Presentación de LUN (LUN Masking)**
El almacenamiento disponible en las cabinas de una red SAN se divide en discos lógicos o LUNs.
Este concepto pertenece a la interfaz SCSI, cuyas capas altas son transportadas mediante FC, que es el protocolo por excelencia para conectar cabinas de discos y servidores, bien sea directamente o bien mediante redes SAN.
Si una cabina de discos está conectada a varios servidores mediante una red SAN o directamente si la cabina dispone de varios puertos de comunicaciones, podrá ofrecer los discos lógicos, o LUNs, a los servidores que sean configurados.
Si una LUN no es presentada por la cabina a un servidor concreto este no podrá acceder a la misma, no teniendo constancia ni siquiera de su existencia. De esta forma se evita que un servidor pueda acceder a una LUN que pertenece a otro servidor.
En el caso de servidores en clúster, es normal que tengan acceso simultáneo a un conjunto de LUNs comunes, para almacenar la información del clúster. En ese caso la cabina de discos debe permitir el acceso a las mismas LUNs por todos los servidores del clúster.
Un mecanismo adicional de seguridad ofrecido por las redes SAN es el de **portbinding**, mediante el cual se puede limitar la conectividad de un determinado equipo (mediante la identificación de su WWN) a un determinado puerto de un switch dentro del fabric (red SAN). El dispositivo solo podrá acceder a la SAN si se conecta físicamente al puerto permitido y solamente un dispositivo con el WWN indicado podrá conectarse a dicho puerto. De esta forma se evita que un dispositivo no perteneciente a la SAN pueda conectarse a la misma y obtener información.
Se denomina **fabric** al conjunto de elementos de red, incluyendo su programación, que permite fijar las rutas para que los datos que llegan a un nodo de la red SAN se encaminen correctamente al siguiente nodo. Si se configuran varios fabric para que desde un mismo host se acceda a las LUNs que se le han presentado, la solución garantizará alta disponibilidad a nivel de transporte de datos

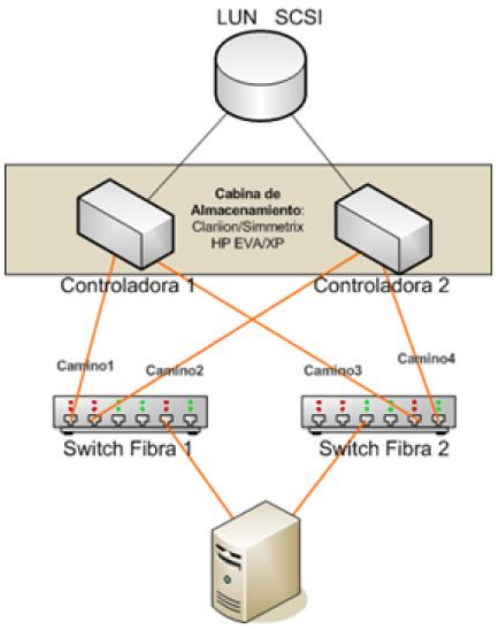


Figura 66. Mecanismos de alta disponibilidad en el acceso a una red SAN.

Mecanismos de replicación

La **replicación** es el proceso de copiar y mantener actualizados los datos en varios nodos de unidades de almacenamiento (base de datos, LUN, cabina entera…). Gracias a este proceso se genera una copia exacta de parte del sistema. Esa parte puede ser desde un archivo hasta una carpeta, una partición, un disco o incluso varios discos

En el caso de una red de almacenamiento, es imprescindible disponer de un mecanismo de este tipo para que la organización no pierda datos en caso de fallos en una de las cabinas.

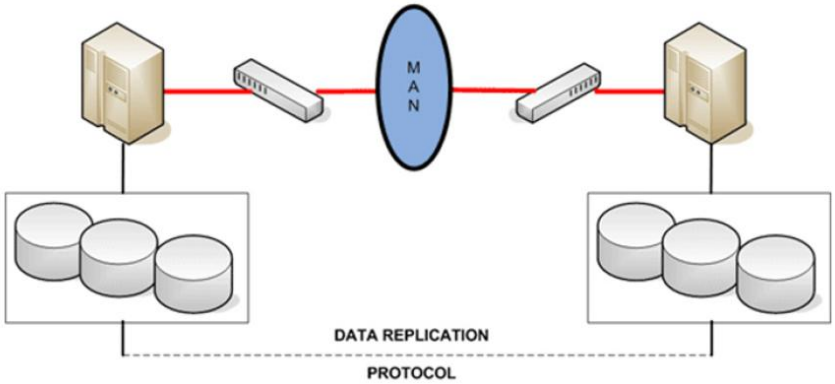


Figura 67. Esquema de replicación de cabinas en una SAN.

Debido a la vital importancia de garantizar la seguridad de los datos, tanto ante desastres como ante errores humanos, sabotajes, etc, se hace necesario mantener **copias de los datos en tiempo real** en distintos sistemas de almacenamiento y en distintas localizaciones físicas, de forma que el fallo de uno de los sistemas no implique la pérdida de los datos.

Las capacidades del almacenamiento en red, tanto SAN como NAS, permiten ofrecer réplica automática de información de forma transparente a los servidores y las aplicaciones.

En función de los requisitos de seguridad, distancia, rendimiento… es posible elegir entre distintos procedimientos para realizar la replicación en tiempo real de los datos.

Dentro de este tipo de réplica, podemos diferenciar entre copia remota y copia local.

• **Copia remota**

Mediante la copia remota se replican los datos que los servidores escriben en los sistemas de almacenamiento en dos o más instalaciones de forma simultánea.

Si se necesita garantizar que los sistemas de almacenamiento remotos son en todo momento una copia de los sistemas locales se debe utilizar copia remota **síncrona**, en la que el servidor no recibe confirmación de la escritura en el sistema de almacenamiento hasta que el dato ha sido escrito tanto en el sistema local como en el remoto.

Cuando un servidor escribe un dato en almacenamiento en SAN este es enviado de forma transparente al servidor tanto a un sistema de almacenamiento local como a uno remoto y solo cuando ha sido escrito en ambos se devuelve al servidor la confirmación de la escritura. En caso de no poderse escribir en ambos se deshace la escritura y se devuelve un error al servidor. De esta forma, en todo momento se puede garantizar que el estado del sistema de almacenamiento local y remoto es el mismo y además coherente con la visión que de ello tiene el servidor.

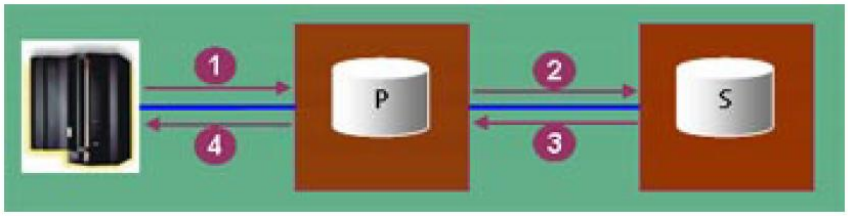


Figura 68. Mecanismo de copia remota síncrona.

forjaTIC

Al tener que esperar a que se produzca la escritura en el sistema remoto, el tiempo de duración de la escritura es el tiempo de acceso al sistema de almacenamiento más la latencia de las líneas de comunicaciones entre el centro local y el remoto. Este hecho, junto con la exigencia de ancho de banda entre ambas localizaciones, hace que la distancia a la que se puede realizar una réplica síncrona sin que afecte significativamente al rendimiento sea limitada, no pudiendo alcanzar distancias de centenares de kilómetros.
Para solventar este problema se puede utilizar copia **asíncrona**, en la que el servidor recibe la confirmación de la escritura en cuanto esta se ha realizado en el centro local y posteriormente la red de almacenamiento se encarga de transmitir los datos de forma asíncrona al centro remoto.
Los mecanismos de copia asíncrona deben garantizar que las escrituras en el centro remoto conservan el orden en el que el servidor las hizo, mediante la definición de grupos de consistencia, de forma que se garantice que el estado del sistema remoto sea una réplica del estado del sistema local pero con un cierto desfase en el tiempo.

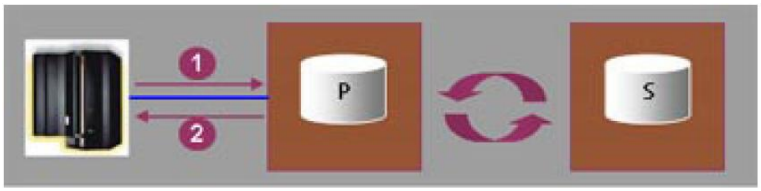


Figura 69. Mecanismo de copia remota asíncrona.

En función del desfase temporal que se pueda tolerar es posible extender la copia asíncrona a grandes distancias, pudiendo incluso situar los centros remotos en continentes distintos.
Una situación posible es el uso simultaneo de copias síncronas situadas a decenas de kilómetros para garantizar la persistencia total de los datos en caso de desastres locales, junto con una copia asíncrona simultánea, que aunque mantenga un desfase temporal con el centro local, permita situar un centro de respaldo a centenares de kilómetros, garantizando la persistencia de los datos frente a desastres de mayor alcance a costa de perder las transacciones de los últimos segundos.

• **Copia local**

Si no se requiere protección frente a fallos completos de los sistemas de almacenamiento, sino que se desea protegerse frente a fallos humanos o del software, o se requiere una copia de los datos para hacer pruebas, es posible utilizar sistemas de copia local.
Estos sistemas permiten disponer, prácticamente sin pérdida de prestaciones y de forma instantánea, de una copia de los datos residiendo en mismo dispositivo de almacenamiento.
Un fallo hardware del sistema de almacenamiento invalidaría ambas copias, pero no así un fallo humano o un fallo del software, ya que ambas copias son lógicamente distintas, siendo una de ellas accedida por los sistemas de producción y estados la otra reservada para pruebas o para recuperación de desastres.

	Discos internos	DAS	NAS	SAN
Cantidad	Capacidad limitada Crecimiento limitado	Capacidad media Crecimiento fácil	Capacidad media Crecimiento fácil	Capacidad elevada Crecimiento fácil
Calidad	Rendimiento bajo Disponibilidad reducida	Rendimiento medio	Rendimiento medio, orientado a servidor de ficheros	Rendimiento alto Disponibilidad mayor Red separada de fibra
Coste	Bajo	Medio	Medio	Alto

Figura 70. Resumen de las arquitecturas de almacenamiento.

5.7. Arquitecturas de respaldo (backup)

5.7.1. Introducción

Los sistemas de backup permiten grabar en cinta los datos en un determinado instante de tiempo. A diferencia de las copias en tiempo real, los datos no se copian en las cintas a medida que se escriben en los sistemas de almacenamiento “on-line”, sino que se realiza una copia completa con la periodicidad que se desee y con las aplicaciones paradas, sin realizar escrituras en los sistemas de almacenamiento.
Frente a los sistemas de réplica en tiempo real presenta la desventaja de tener una imagen de los datos de un instante fijo anterior en el tiempo, pero tiene como principales ventajas la posibilidad de mantener varias copias de varios instantes de tiempo (por ejemplo, todas las noches del último mes) en distintas cintas y poder recuperar la que se desee y además la de representar un estado coherente de los datos a nivel de aplicación, al detener las aplicaciones antes de hacer el backup.
En el caso de la réplica síncrona en tiempo real se garantiza que el estado de los datos en el centro remoto es idéntico al local a nivel de bloque, pero no a nivel de S.O. o aplicación. Un fallo en el centro local es similar a una parada brusca de un servidor respecto a sus discos internos. Todo aquello que las aplicaciones y el S.O. tuviesen en caché y no hubiesen volcado aún al disco se pierde y por lo tanto la imagen del disco tanto local como remoto, a pesar de ser idénticos a nivel de bit, pueden no ser coherentes a nivel de aplicación.
En cambio, si en un backup se paran las aplicaciones (o se fuerza que escriban todos los datos a disco), la copia que se realiza será consistente a nivel de operación de E/S y a nivel de aplicación.
Sin embargo, el tener que mantener las aplicaciones paradas durante la realización del backup, en lo que se denomina ventana de backup, hace que la duración del mismo sea crítica. Una situación típica es la de parar la producción todas las noches durante el tiempo que dure la realización del backup y volcar todos los datos en cinta.
Esta es una de las razones que potencian la existencia de opciones de backup que limitan este problema:

- Por un lado, cada vez más aplicaciones permiten la realización de **backup de sus datos “en caliente”**, sin necesidad de parar sus servicios, aunque siempre existirá un impacto en su rendimiento.
- Por otro lado, existe la posibilidad de hacer **backup a disco** (directamente o utilizando los virtualizadores de cinta). Estos dispositivos reciben el backup como si fuesen cintas pero en realidad lo están enviando a disco, con lo que el tiempo de respuesta es mucho mayor. Cuando acaba el backup se reanuda la producción y a partir de ese momento el virtualizador puede volcar los datos a cinta sin restricciones temporales.

Para la realización automática de backups de dispositivos NAS o de servidores de ficheros existe un protocolo estándar denominado NDMP (Network Data Management Protocol) que permite separación de los caminos de control y de datos, de forma que el backup puede ser gestionado desde un servidor central de backup mientras que los datos viajan directamente del servidor de ficheros a las unidades de cinta, pudiendo pasar directamente por la SAN sin tener que hacer uso de la LAN, como ocurriría en el caso de que el servidor de backup tuviese que recabar los datos de los servidores de ficheros para posteriormente enviarlos a las unidades de cinta.

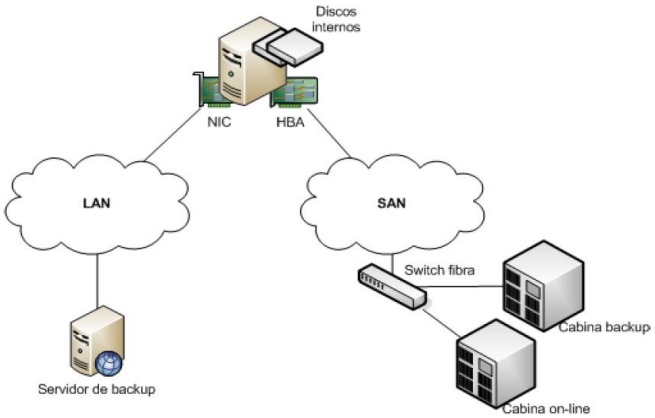


Figura 71. Arquitectura general de solución de backup.

Las cintas presentan un modo secuencial de acceso a los datos frente al modo aleatorio de los discos, por lo que la tecnología de cintas es intrínsecamente más lenta que la de disco, aunque si la aplicación solamente requiere acceso secuencia (por ejemplo, un backup) esa diferencia puede verse reducida.
Las unidades de cinta pueden estar agrupadas en librerías, de forma que un servidor dispone de varias unidades sobre las que cargar los cartuchos. Las librerías manuales están en desuso en los centros de datos, siendo sustituidas por librerías automáticas, o robots, en los que un brazo mecánico selecciona las cintas y las introduce o saca de las unidades de cinta. También está la opción de Librería de Cintas Virtual, que realmente son discos de bajo coste como soporte intermedio para un backup más rápido.

5.7.2. Políticas de backup

La política de backup es la definición de los diferentes aspectos de las copias de seguridad: ¿de qué se debe hacer backup? ¿Cada cuánto se realiza la copia de seguridad? ¿Qué retención deben tener? ¿Dónde se guardan las copias? ¿Cuánto tiempo es aceptable que se pueda tardar en recuperar datos?
En los sucesivos apartados se analizarán los puntos clave que requieren concretarse para estableces una política de backup adecuada a las necesidades de cada sistema.

Tipos de backup
En forma general y a grandes rasgos, existen 4 tipos distintos de backups. Uno de los mecanismos más extendidos para informar de los cambios de los ficheros respecto a algún backup anterior es la utilización de un bit de control. Utilizando éste u otro mecanismo equivalente, se pueden realizar los siguientes tipos de backup:

- **Copia de seguridad completa:** copia de seguridad total de todos los archivos y directorios seleccionados. En el caso de la copia completa, el programa borra el bit de modificado de cada archivo. Es la base para futuras tareas que solo realizan copias de seguridad de los archivos modificados.
- **Copia de seguridad incremental:** el programa examina el bit de modificado y hace una copia de seguridad solo de los archivos que han cambiado desde la última copia de seguridad incremental o normal. Esta tarea borrar el bit de modificado de cada archivo que copia.
  Utilizan la mínima cantidad de cinta y ahorran tiempo, sin embargo, realizar una restauración es un inconveniente
- **Copia de seguridad diferencial:** es lo mismo que una copia de seguridad incremental exceptuando que el programa no elimina el bit de modificación. Requiere más espacio en cinta y tiempo que las incrementales pero su ventaja radica en que cuando se

realiza una restauración se necesitan solo las cintas que contengan la copia de seguridad normal y la más reciente diferencial.

• **Copia de seguridad intermedia:** equivalente a una copia de seguridad normal, excepto que el programa no desactiva el bit de modificado.

Como caso especial de copia normal se puede hacer una copia de seguridad diaria: solo copia los archivos que han sido modificados en el día en que se ejecuta la tarea sin tener en cuenta el estado del bit de modificación. Tampoco borra el bit de modificación. Útiles cuando se quiere realizar una copia de seguridad extra en un día determinado, sin afectar a la estrategia de copia de seguridad establecida.



Figura 72. Tipos de backup.

Rotación de medios

Un esquema de rotación de medios dicta cuantas cintas se usan para realizar las copias de seguridad.

Un esquema popular es el *método del abuelo-padre-hijo,* utiliza tres generaciones de cintas que representan copias de seguridad mensual, semanal y diaria.

• Se realiza una copia de seguridad completa cada mes y se guarda la cinta durante un año (abuelo).

• Se realiza una copia de seguridad completa semanalmente y se guarda durante un mes (padre).

• Las copias de seguridad hijo se realizan diariamente y se guardan durante una semana

(12+4+7=23 cintas). Las tareas diarias pueden ser copias de seguridad completa, incremental o diferencial.



Figura 73. Rotación de medios en backup.

En función de la criticidad de los datos, del espacio y del tiempo disponible, y una vez concretado el tipo de backup y la rotación de medios, se podrá establecer la política de backup correspondiente.

En este momento se ha de tener en cuenta aspectos legales de retención, entorno que se salvaguarda y, en general, criticidad de la información a respaldar

En la figura se muestra un ejemplo de una política de backup en función del entorno (productivo, preproductivo, test…).

PERIODICIDAD	DESARROLLO-PREEXPLOTACIÓN-FORMACIÓN	PRODUCCIÓN	TEST	DIA EJECUCIÓN	TIPO	FRANJA HORARIA
DIARIO	NO	SI (Retención: 10 días)	NO	de Lunes a Viernes	INCREMENTAL	20:00 a 08:00
SEMANAL	SI (Retención: 1 mes)	SI (Retención: 1 mes)	NO	Sábado o Domingo	FULL	00:00 a 24:00
MENSUAL	SI (Retención: 2 meses)	SI (Retención: 3 meses)	SI (Retención. 2 meses)	Sin especificar (Sábado o Domingo)	FULL	00:00 a 24:00
				BAJO PETICIÓN (dia pedido o último dia mes)		20:00 a 08:00
ANUAL	NO	SI (Retención: 1 año)	NO	BAJO PETICIÓN	FULL	20:00 a 08:00
ANUAL 5 AÑOS				POR LEY		
Puntual/Otra	Muy justificada					

Figura 74. Ejemplo de política de backup.

5.7.3. Tendencias

A continuación se detallan las tendencias y novedades dentro la categoría soluciones de almacenamiento y backup:

• **Las nubes híbridas predominarán.** Las organizaciones comenzarán a adoptar soluciones híbridas mayoritariamente en sus infraestructuras, acabando así con las tensiones previas producidas por la elección del entorno cloud más adecuado.

Además, la **nube híbrida** será cada vez más funcional. Las empresas trabajan tanto con almacenamiento en sus propias instalaciones como fuera de ellas (cloud híbrido). Este será un problema mayor ya que los administradores son cada vez más conscientes de lo práctica que es la utilización de la nube para las cargas de trabajo. El almacenamiento en las propias instalaciones se suele destinar a almacenar datos orientados a las transacciones de la propia empresa. Sin embargo, la nube todavía se utiliza sobre todo para los datos inactivos, copias de seguridad y recuperación de los datos en caso de desastre en el sistema, debido a las velocidades de internet.

Sin embargo, es una tendencia que no se impondrá en las empresas cuyos sistemas son críticos, como los bancos, empresas de servicios y compañías financieras. Este tipo de organizaciones, en lugar de confiar sus activos de información a terceros, está construyendo sitios alternativos de almacenamiento -también llamados sites de contingencia- donde mantienen el nivel de operaciones en caso de que ocurra un desastre en el sistema principal.

• **Almacenamiento definido por software.** Durante los próximos años veremos una clara tendencia a abandonar las políticas basadas en los componentes de la infraestructura tradicional para pasar a entornos dominados por componentes de infraestructura virtual (controladores de red y de almacenamiento).

• **Resurgimiento de las SMV.** Las Storage Virtual Machine (máquinas virtuales de almacenamiento) ofrecerán soluciones adaptadas a los nuevos requerimientos, tanto a nivel de movilidad en los datos como en lo que se refiere a rendimiento en aplicaciones. En este sentido, las SMV simplificarán la migración de cargas de trabajo entre clústeres de almacenamiento y proporcionarán la posibilidad de contar con clústeres de almacenamiento altamente disponibles en áreas metropolitanas.

• **40 Gb, la nueva Ethernet.** La siguiente evolución de las redes Ethernet – los 40 Gb –, comenzará a adoptarse en los centros de datos, aportando mayores anchos de banda y propiciando sistemas más ágiles, con mayor capacidad para mover datos entre sistemas.

• **Consolidación del Big Data.** Dentro del ámbito del Big Data, las empresas comenzarán a obtener datos nuevos datos que les permitan saber más sobre el comportamiento de sus consumidores, así como los procesos industriales o naturales por los que pasa. Estas medidas abrirán la puerta a nuevos entornos analíticos que extenderán el conocimiento de las compañías.

• **Tecnologías:**

- Deduplicación: Esta tecnología elimina de la copia todos los bloques de datos que estén repetidos, almacenando una sola copia y estableciendo enlaces a ella. Esta opción también está disponible en cabinas de almacenamiento, pero la ganancia es habitualmente mayor a la hora de realizar backups, ya que los datos se suelen repetir con mayor frecuencia.

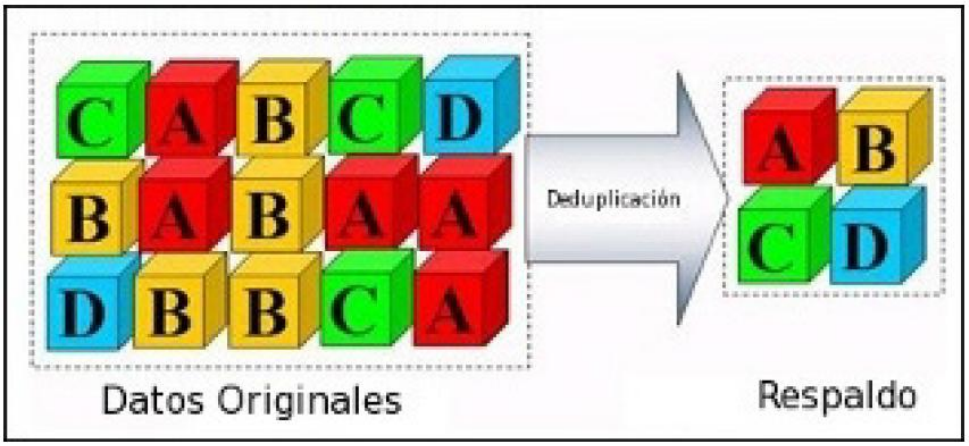


Figura 75. Deduplicación de datos.

- Auto-tiering: consiste en que una cabina de almacenamiento presenta un disco a los servidores y adapta su rendimiento según se lo vaya demandando cada aplicación. Crea una capa virtual que está conformado por discos de una determinada calidad (tier o capa). La carga requerida por una determinada aplicación puede cambiar y es cuando entra en funcionamiento el auto-tiering, un mecanismo en tiempo real que continuamente posiciona los datos en la clase de almacenamiento apropiado según la frecuencia de acceso a los mismos.

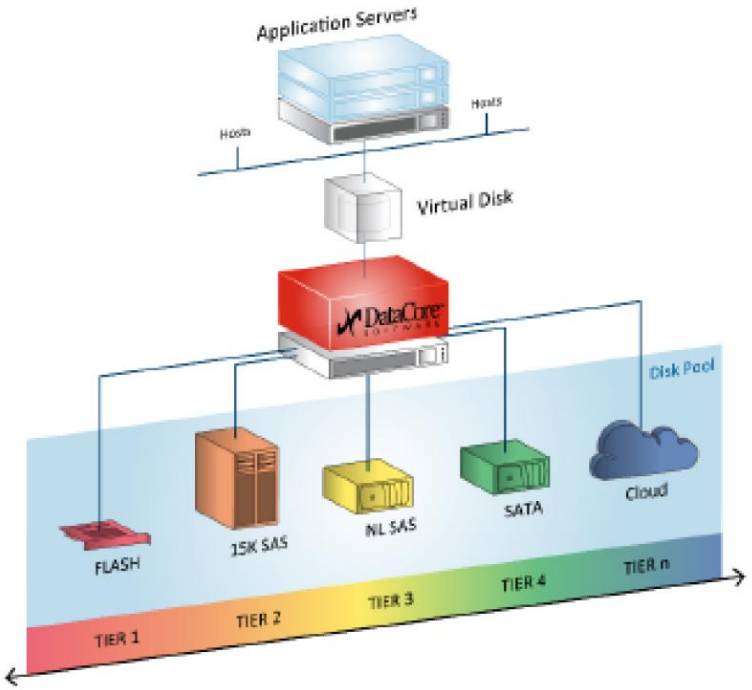


Figura 76. Auto-tiering por parte de una cabina de almacenamiento.