

CAPÍTULO 6

Memoria externa

6.1. Discos magnéticos

- Mecanismos de lectura y escritura magnética
- Organización y formato de datos
- Características físicas
- Parámetros para medir las prestaciones de un disco

6.2. RAID

- Nivel 0 del RAID
- Nivel 1 del RAID
- Nivel 2 del RAID
- Nivel 3 del RAID
- Nivel 4 del RAID
- Nivel 5 del RAID
- Nivel 6 del RAID

6.3. Memoria óptica

- Discos compactos
- Disco digital versátil

6.4. Cinta magnética

6.5. Lecturas y sitios web recomendados

- Sitios web recomendados

6.6. Palabras clave, preguntas de repaso y problemas

- Palabras clave
- Preguntas de repaso
- Problemas

PUNTOS CLAVE

- Los discos magnéticos siguen siendo el componente más importante de la memoria externa. Tanto los extraíbles como los fijos, o duros, los discos se usan tanto en los PC, como en computadores grandes y supercomputadores.
- Para conseguir mayores prestaciones y disponibilidad, un esquema de servidores y sistemas grandes extendido es la tecnología RAID de discos. RAID se refiere a una familia de técnicas para utilizar varios discos como un conjunto de dispositivos de almacenamiento de datos en paralelo, con redundancia para compensar los fallos de disco.
- Las técnicas de almacenamiento óptico se han convertido en algo cada vez más importante en los computadores. Mientras que el CD-ROM se ha usado ampliamente durante muchos años, tecnologías más recientes, como el CD reescribible y las unidades de almacenamiento magnético-ópticas, están siendo cada vez más importantes.

En este capítulo se examinan distintos sistemas y dispositivos de memoria externa. Comenzamos con el dispositivo más importante, el disco magnético. Los discos magnéticos son la base de las memorias externas en casi todos los computadores. En la siguiente sección se examina el uso de conjuntos de discos para conseguir mayores prestaciones, concretamente la familia conocida como RAID (*Redundant Array of Independent Disks*, conjunto redundante de discos independientes). La memoria óptica externa es un componente cada vez más importante de muchos computadores, y se examinará en la tercera sección. Al final, se describen las cintas magnéticas.

6.1. DISCOS MAGNÉTICOS

Un disco magnético es un plato circular construido con un material no magnético, llamado sustrato, cubierto por un material magnetizable. Tradicionalmente, el sustrato es aluminio o una aleación de aluminio. Recientemente, se han utilizado sustratos de cristal. Los sustratos de cristal tienen una serie de ventajas, entre las cuales se encuentran:

- Mejora en la uniformidad de la superficie magnética para incrementar la fiabilidad del disco.
- Reducción significativa de los defectos en toda la superficie lo que ayuda a reducir los errores de lectura/escritura.
- Capacidad para soportar grabaciones de gran proximidad (*Fly heights*, que se describirán posteriormente).
- Mejor rigidez para reducir la dinámica del disco.
- Mayor capacidad para resistir golpes y daños.

MECANISMOS DE LECTURA Y ESCRITURA MAGNÉTICA

Los datos se graban y después se recuperan del disco a través de una bobina, llamada **cabeza**; en muchos sistemas, hay dos cabezas, una de lectura y otra de escritura.

Durante una operación de lectura o escritura, la cabeza permanece quieta mientras el plato rota bajo ella.

El mecanismo de escritura se basa en el hecho de que un flujo eléctrico atravesando una bobina crea un campo magnético. Se envían pulsos eléctricos a la cabeza de escritura, y se graban los patrones magnéticos en la superficie bajo ella, con patrones diferentes para corrientes positivas y negativas. La propia cabeza de lectura está hecha de un material fácilmente magnetizable y tiene forma de *donut* rectangular con un agujero a lo largo de un lado y varias vueltas de cable conductor a lo largo del lado opuesto (Figura 6.1). Una corriente eléctrica en el cable induce un campo magnético a lo largo del agujero, que magnetiza una pequeña área del medio grabable. Cambiando la dirección de la corriente, cambia el sentido de magnetización del medio de grabación.

El mecanismo tradicional de lectura se basa en el hecho de que un campo magnético en movimiento respecto a una bobina, induce una corriente eléctrica en la bobina. Cuando la superficie del disco pasa bajo la cabeza, en esta se genera una corriente de la misma polaridad que la que produjo la grabación magnética. La estructura de la cabeza de lectura es, este caso, esencialmente la misma que la de escritura y, por tanto, se puede usar la misma cabeza para ambas operaciones. Estas cabezas únicas se usan en disquetes y discos duros antiguos.

Los discos duros de hoy usan un mecanismo diferente para la lectura, siendo necesaria una cabeza de lectura separada posicionada, por conveniencia, cerca de la cabeza de escritura. La cabeza de lectura consiste en un sensor magnetoresistivo (MR) parcialmente blindado. El MR tiene una

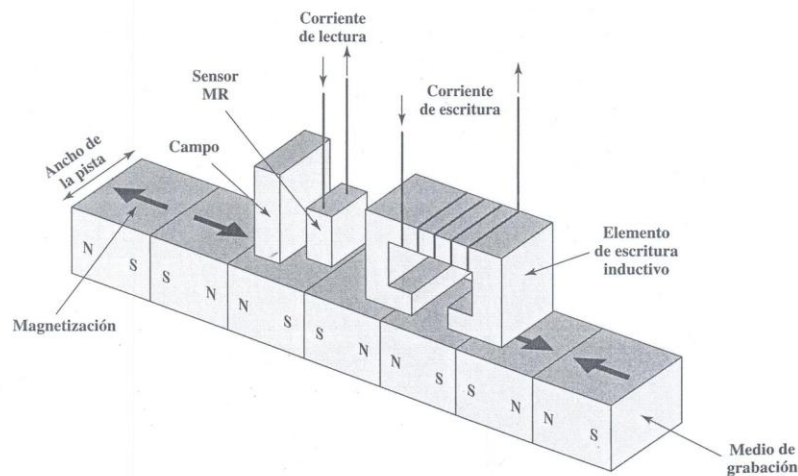


Figura 6.1. Cabeza de (escritura por inducción) / (lectura magnetoresistiva).

resistencia eléctrica que depende de la dirección de la magnetización del medio que se mueve bajo él. Haciendo pasar un corriente a través del sensor MR, los cambios de la resistencia se detectan como señales de tensión. El diseño del MR permite operar a altas frecuencias, lo que lo que equivale a grandes densidades de almacenamiento y de velocidad de funcionamiento.

ORGANIZACIÓN Y FORMATO DE LOS DATOS

La cabeza es un dispositivo relativamente pequeño, capaz de leer o escribir en una zona del plato que rota bajo ella. Esto da lugar a que los datos se organicen en un conjunto de anillos concéntricos en el plato, llamados **pistas**. Cada pista es del mismo ancho que la cabeza. Usualmente hay cientos de pistas por superficie.

En la Figura 6.2 se puede ver la disposición de los datos. Las pistas adyacentes están separadas por **bandas vacías**. Esto previene, o por lo menos minimiza, los errores debidos a desalineamientos de la cabeza o simplemente a interferencias del campo magnético.

Los datos se transfieren al y desde el disco en sectores (Figura 6.2). Normalmente hay cientos de sectores por pista, y estos pueden tener una longitud variable o fija. En la mayoría de los sistemas de hoy se utilizan sectores de longitud fija, siendo 512 bytes el tamaño casi universal de un sector. Para evitar imposiciones de precisión ilógicas del sistema, los sectores adyacentes se separan con intrapistas (intersectores) vacías.

Un bit cercano al centro de un disco girando, pasa por punto fijo (como la cabeza de lectura-escritura) más despacio que un bit más externo. Por tanto, debe haber alguna forma de compensar la

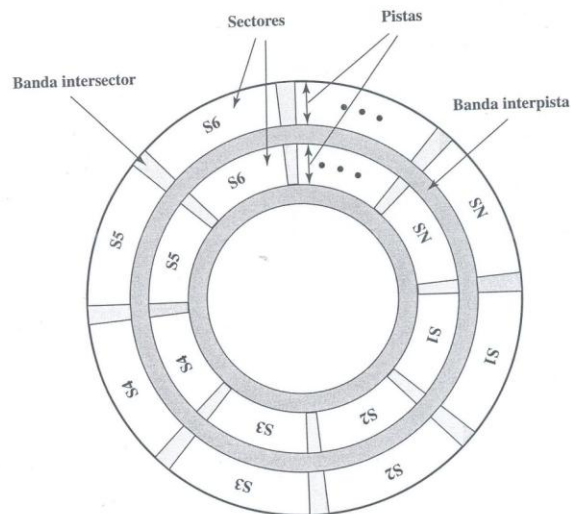


Figura 6.2. Organización de los datos en el disco.

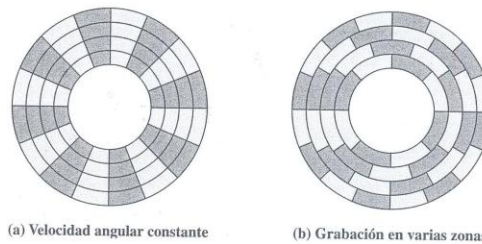


Figura 6.3. Comparación de los métodos de organización de un disco.

variación de la velocidad de forma que la cabeza pueda leer todos los bits a la misma velocidad. Esto se puede hacer incrementando el espacio entre bits de la información grabada en los segmentos del disco. La información se puede escanear a la misma velocidad rotando el disco a una velocidad fija, conocida como **velocidad angular constante** (*constant angular velocity, CAV*). La Figura 6.3a muestra la estructura de un disco que usa CAV. El disco se divide en una serie de sectores en forma de trozo de tarta y en una serie de pistas concéntricas. La ventaja de usar CAV es que los bloques individuales de datos se pueden direccionar directamente con la pista y sector. Para mover la cabeza desde su actual posición a una dirección específica, solo hay que mover ligeramente la cabeza a la pista específica y esperar a que el sector se sitúe bajo la cabeza. La desventaja de CAV es que la cantidad de datos que se puede almacenar en las pistas más externas es solo la misma que la de las pistas internas.

Debido a que la **densidad**, en bits por pulgada, aumenta a medida que nos movemos desde la pista más externa a la más interna, la capacidad de almacenamiento de un disco con un sistema CAV sencillo viene limitada por la máxima densidad de grabación que se puede llevar a cabo en la pista más interna. Para aumentar la capacidad, los discos duros modernos utilizan una técnica conocida como **grabación en varias zonas** (*multiple zone recording*), en la que la superficie se divide en varias zonas concéntricas (usualmente 16). Dentro de una zona, el número de bits por pista es constante. Las zonas más lejanas del centro contienen más bits (más sectores) que las zonas próximas al centro. Esto permite capacidades de almacenamiento mayores a expensas de una circuitería de alguna forma más compleja. Como la cabeza del disco se mueve de una zona a otra, la longitud (a lo largo de la pista) de los bits individuales cambia, provocando un cambio en el tiempo de lectura y escritura. La Figura 6.3b sugiere la naturaleza de la grabación en varias zonas; en esta figura, cada zona es una sola pista.

Algún procedimiento es necesario para situar las posiciones del sector en una pista. Claramente, debe haber algún punto de comienzo de la pista y una manera de identificar el principio y el fin de cada sector. Estos requisitos son gestionados mediante datos de control grabados en el disco. Por tanto, el disco se graba con un formato que contiene algunos datos extra usados solo por el controlador del disco y no accesibles al usuario.

En la Figura 6.4 se muestra un ejemplo del formato de grabación de un disco. En este caso, cada pista contiene treinta sectores de longitud fija de 600 bytes cada uno. Cada sector contiene 512 bytes de datos más información de control útil al controlador del disco. El campo ID es un identificador único o dirección usado para localizar un sector particular. El byte SINCRO es un patrón de bits especial que delimita el comienzo del campo. El número de pista identifica una pista en una super-

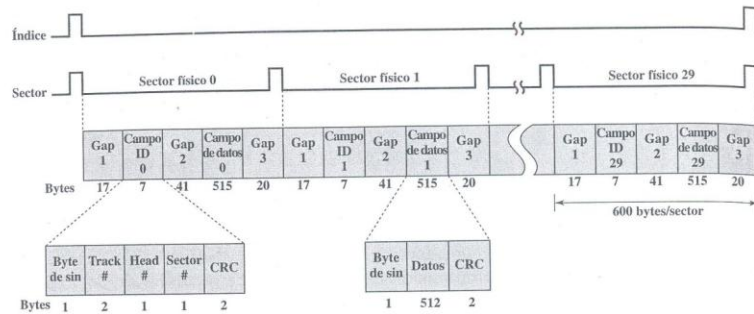


Figura 6.4. Formato de las pistas de un disco Winchester (Seagate ST506).

ficie. El número de cabeza identifica una cabeza, si el disco tiene varias superficies (como acabamos de explicar). El ID y los campos de datos contienen, cada uno, un código de detección de errores.

CARACTERÍSTICAS FÍSICAS

En la Tabla 6.1 se listan las principales características que diferencian los distintos tipos de discos. Primero, las cabezas pueden ser fijas o móviles con respecto a la dirección radial del plato. En un disco de **cabeza fija** hay una cabeza de lectura/escritura por pista. Todas las cabezas se montan en un brazo rígido que se extiende a través de todas las pistas. En un disco de **cabeza móvil**, hay solo una cabeza de lectura/escritura. Como antes, la cabeza se monta en un brazo. Como la cabeza debe poder posicionarse encima de cualquier pista, el brazo debe extenderse o retraerse para este propósito.

El disco mismo, se monta en una unidad de disco, que consta del brazo, un eje que rota el disco y la electrónica necesaria para la entrada y salida de datos binarios. Un **disco no extraíble** está permanentemente montado en la unidad de disco. Un **disco extraíble**, puede ser quitado y sustituido por otro disco. La ventaja de este último tipo es que es posible una cantidad de datos ilimitada con un

Tabla 6.1. Características físicas de los discos.

Movimiento de la cabeza	Platos
Cabeza fija (una por pista)	Un plato
Cabeza móvil (una por superficie)	Varios platos
Portabilidad de los discos	Mecanismo de la cabeza
Disco no extraíble	Contacto (disquete)
Disco extraíble	Separación fija
Caras	Separación aerodinámica (Winchester)
Una cara	
Dos caras	

número limitado de unidades de disco. Además, un disco puede ser utilizado en diversos computadores. Los disquetes y los cartuchos ZIP son ejemplos de discos extraíbles.

En la mayoría de los discos, la cubierta magnetizable se aplica a ambas caras del plato, denominándose estos discos de **doble superficie**. Algunos discos, menos caros, son de **una sola superficie**.

Algunas unidades de disco poseen **varios platos** apilados verticalmente y separados por una distancia de alrededor de una pulgada. Disponen de varios brazos (Figura 6.5). Los discos de varios platos utilizan una cabeza que se mueve, con una cabeza de lectura-escritura para cada superficie del plato. El conjunto de todas las pistas que tienen la misma posición relativa en el plato se denomina **cilindro**. Por ejemplo, todas las pistas sombreadas en la Figura 6.6 pertenecen al mismo cilindro.

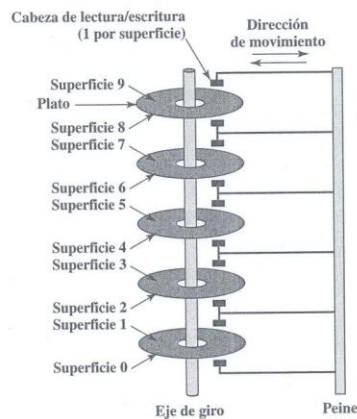


Figura 6.5. Componentes de una unidad de disco.

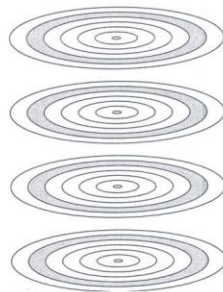


Figura 6.6. Pistas y cilindros.

Finalmente, el mecanismo de la cabeza proporciona una clara clasificación de los discos en tres tipos. Tradicionalmente, la cabeza de lectura/escritura se posiciona a una distancia fija sobre el plato, dejando entre ambos una capa de aire. En el otro extremo está el mecanismo de la cabeza que realmente efectúa un contacto físico con el medio durante la operación de lectura o escritura. Este mecanismo es el que se usa con los **disquetes**, que son pequeños, de plato flexible y es el tipo de disco más barato.

Para entender el tercer tipo de disco, necesitamos comentar la relación entre la densidad de datos y la anchura de la capa de aire. La cabeza debe generar o detectar un campo magnético de intensidad suficiente para escribir y leer correctamente. Cuanto más estrecha es la cabeza, más cercana debe estar a la superficie del plato para funcionar. Esto es deseable, ya que una cabeza más estrecha implica pistas más estrechas y por tanto, mayor densidad de datos. Sin embargo, cuanto más cerca esté la cabeza del disco, mayor será el riesgo de error debido a impurezas o imperfecciones. Los discos Winchester supusieron un avance tecnológico en este sentido.

Las cabezas de los Winchester están montadas en unidades herméticamente cerradas, que están casi libres de contaminación. Fueron diseñados para operar más cerca de la superficie del disco que las cabezas de los discos rígidos anteriores, por tanto permiten una densidad de datos mayor. La cabeza está en el contorno de una hoja de metal aerodinámica que reposa suavemente sobre la superficie del plato cuando el disco no se mueve. La presión del aire generada por el giro del disco es suficiente para hacer subir la hoja encima de la superficie. El sistema sin contacto resultante puede ser diseñado para usar cabezas más estrechas que las de los discos rígidos convencionales, operando más cerca de la superficie de los platos¹.

La Tabla 6.2 muestra los parámetros de los discos de altas prestaciones actuales.

PARÁMETROS PARA MEDIR LAS PRESTACIONES DE UN DISCO

Los detalles de las operaciones de E/S de un disco dependen del tipo de computador, del sistema operativo, de la naturaleza de los canales de E/S y del hardware controlador del disco. En la Figura 6.7 se muestra un diagrama de temporización general de las transferencias de E/S del disco.

Cuando la unidad de disco está funcionando, el disco está rotando a una velocidad constante. Para leer o escribir, la cabeza debe posicionarse en la pista deseada y al principio del sector deseado en la pista. La selección de la pista implica un movimiento de la cabeza, en un sistema de cabeza móvil, o una selección electrónica de una cabeza, en un sistema de cabezas fijas. En un sistema de cabeza móvil, el tiempo que tarda la cabeza en posicionarse en la pista se conoce como **tiempo de búsqueda**. En cualquier caso, una vez seleccionada la pista, el controlador del disco espera hasta que el sector apropiado rote hasta alinearse con la cabeza. El tiempo que tarda el sector en alcanzar la cabeza se llama **retardo rotacional** o latencia rotacional. La suma del tiempo de búsqueda, si lo hay, y el retardo rotacional se denomina **tiempo de acceso**, o tiempo que se tarda en llegar a la posición de lectura o escritura. Una vez posicionada la cabeza, se lleva a cabo la operación de lectura o escritura, desplazándose el sector bajo la cabeza; esta operación conlleva un **tiempo de transferencia de datos**.

¹ Como información de interés histórico, el término Winchester fue usado originalmente por IBM como nombre preliminar para el modelo de disco 3340. El 3340 era un paquete de discos extraíble con las cabezas integradas en el paquete. El término se aplica ahora a cualquier unidad de disco integrada con un diseño de cabezas aerodinámico. El disco Winchester se usa habitualmente en PC y estaciones de trabajo, y se le suele llamar *disco duro*.

Tabla 6.2. Parámetros de las unidades de disco duro.

Características	Seagate Barracuda 180	Seagate Cheetah X15-36LP	Seagate Barracuda 36ES	Toshiba HDD1242	Hitachi Microdrive
Aplicación	Servidor de gama alta capacidad	Servidor de altas prestaciones	Servidor básico	Portátil	Dispositivos de bolsillo
Capacidad	181.6 GB	36.7 GB	18.4 GB	5 GB	4 GB
Tiempo de búsqueda mínimo pista-pista	0.8 ms	0.3 ms	1.0 ms	—	1.0 ms
Tiempo de búsqueda medio	7.4 ms	3.6 ms	9.5 ms	15 ms	12 ms
Velocidad del eje de giro	7200 rpm	15K rpm	7200	4200 rpm	3600 rpm
Retardo rotacional medio	4.17 ms	2 ms	4.17 ms	7.14 ms	8.33 ms
Velocidad máxima de transferencia	160 MB/s	522 a 709 MB/s	25 MB/s	66 MB/s	7.2 MB/s
Bytes por sector	512	512	512	512	512
Sectores por pista	793	485	600	63	—
Pistas por cilindro (número de superficies del plato)	24	8	2	2	2
Cilindros (número de pistas en una cara del plato)	24,247	18,479	29,851	10,350	—

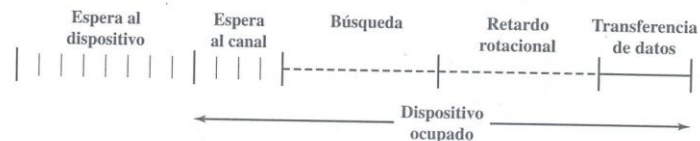


Figura 6.7. Temporizador de la transferencia entre disco y E/S.

Además del tiempo de acceso y de transferencia, hay varios retardos en cola usualmente asociados con operaciones de E/S del disco. Cuando un proceso hace una petición de E/S, primero debe esperar en cola hasta que el dispositivo esté disponible. En ese momento, el dispositivo es asignado al proceso. Si el dispositivo comparte un único canal E/S o un conjunto de canales de E/S con otros discos, entonces puede tener que hacer esperas adicionales para que el canal esté disponible. En este punto se hace la búsqueda para empezar el acceso al disco.

En algunos computadores grandes, se usa una técnica conocida como detección de posición rotacional (RPS, *rotational positional sensing*). Esta funciona de la siguiente forma: cuando se lleva a cabo una orden de búsqueda, el canal es liberado para atender otras operaciones de E/S. Cuando la

búsqueda se ha completado, el dispositivo determina cuándo se rotan los datos bajo la cabeza. Mientras el sector se aproxima a la cabeza, el dispositivo intenta restablecer el camino de comunicación hacia el anfitrión. Si la unidad de control o el canal están ocupados con otra E/S, la conexión puede fallar y el dispositivo debe rotar una vuelta completa antes de que pueda intentar conectarse de nuevo, lo que se denomina una pérdida RPS. Esto supone un retardo extra que se debe añadir a la línea de tiempo de la Figura 6.7.

Tiempo de búsqueda. El tiempo de búsqueda es el tiempo necesario para desplazar el brazo del disco hasta la pista requerida. Este tiempo resulta difícil de precisar. El tiempo de búsqueda está formado por dos componentes clave: el tiempo inicial de comienzo y el tiempo necesario para atravesar las pistas que tienen que cruzarse una vez que el brazo de acceso esté a la velocidad adecuada. El tiempo transversal no es, desgraciadamente, una función lineal del número de pistas, pero incluye un tiempo de espera (tiempo desde que se posiciona la cabeza sobre la pista objetivo hasta que se confirma la identificación de la pista).

Muchas mejoras provienen de componentes más pequeños y ligeros. Hace algunos años, un disco típico tenía 14 pulgadas (36 cm) de diámetro, mientras que hoy el tamaño más normal es de 3,5 pulgadas (8,9 cm), reduciéndose la distancia que tiene que recorrer el brazo. Un tiempo de búsqueda medio típico de un disco actual está entre 100 y 50 ms.

Retardo rotacional. Los discos, que no sean disquetes, rotan a velocidades de 3 600 rpm (para controlar dispositivos como cámaras digitales) en adelante, como 15 000 rpm; esta última velocidad es una revolución cada 4 ms. Por tanto, de media, el retardo rotacional será de unos 2 ms. Las disqueteras normalmente rotan entre 300 y 600 rpm. Por tanto, el retardo medio estará entre los 100 y 50 ms.

Tiempo de transferencia. El tiempo de transferencia hacia o desde el disco depende de la velocidad de rotación del disco de la siguiente forma:

$$T = \frac{b}{rN}$$

donde:

T = tiempo de transferencia

b = número de bytes a transferir

N = número de bytes de una pista

r = velocidad de rotación en revoluciones por segundo

Por tanto, el tiempo de acceso medio total se puede expresar como

$$T_a = T_s + \frac{1}{2r} + \frac{b}{rN}$$

donde T_s es el tiempo de búsqueda medio. Nótese que en una unidad con zonas, el número de pistas es variable, complicándose el cálculo.

Una comparación de tiempos. Con los parámetros definidos anteriormente, veamos dos operaciones de E/S diferentes que ilustrarán el peligro de fiarse de los valores medios. Considérese un disco con un tiempo de búsqueda medio especificado de 4 ms, una velocidad de rotación de 15 000 rpm, y sectores de 512 bytes con 500 sectores por pista. Supóngase que queremos leer un fichero

que consta de 2 500 sectores con un total de 1,28 Mb. Queremos estimar el tiempo total de transferencia.

Primero, supongamos que el fichero está almacenado de la forma más compacta posible en el disco. Es decir, el fichero ocupa todos los sectores de 5 pistas adyacentes (5 pistas \times 500 sectores/pista = 2 500 sectores). Esto se conoce como *organización secuencial*. Ahora, el tiempo para leer la primera pista es el siguiente:

Búsqueda media	4 ms
Retardo rotacional	2 ms
Leer 500 sectores	4 ms
	<hr/> 10 ms

Supongamos que el resto de las pistas se puede leer ahora sin prácticamente tiempo de búsqueda. Es decir, la operación de E/S puede mantenerse con un flujo continuo desde el disco. Entonces, al menos, se necesita considerar un retardo rotacional para cada pista leída. Entonces, cada pista siguiente se lee en $2 + 4 = 6$ ms. Para leer el fichero entero:

$$\text{Tiempo total} = 10 + (4 \times 6) = 34 \text{ ms} = 0,034 \text{ segundos}$$

Ahora calculemos el tiempo requerido para leer los mismos datos utilizando acceso aleatorio en vez de secuencial; es decir, los accesos a los sectores se distribuyen aleatoriamente sobre el disco. Para cada sector tenemos:

Búsqueda media	4 ms
Retardo rotacional	2 ms
Leer 1 sector	0,008 ms
	<hr/> 6,008 ms

$$\text{Tiempo total} = 2.500 \times 6,008 = 15.020 \text{ ms} = 15,02 \text{ segundos}$$

Está claro que el orden en que se lean los sectores desde el disco tiene una repercusión enorme en las prestaciones de E/S. En el caso de acceso a ficheros en los que se lean o escriban varios sectores, se tiene un cierto control sobre la forma en la que los sectores o datos se organizan, y debemos decir algo sobre este tema en el siguiente capítulo. Sin embargo, aún en el caso de un acceso a un fichero, en un entorno de multiprogramación, habrá peticiones de E/S compitiendo por el mismo disco. Entonces, merece la pena examinar maneras en las que las prestaciones de E/S del disco mejoran respecto a las llevadas a cabo con accesos al disco puramente aleatorios. Esto conduce a considerar algoritmos de planificación del disco, que son jurisdicción de los sistemas operativos y están fuera del alcance de este libro (ver [STAL05] para más detalles).

6.2. RAID

Como se dijo anteriormente, el ritmo de mejora de prestaciones en memoria secundaria ha sido considerablemente menor que en procesadores y en memoria principal. Esta desigualdad ha hecho, quizá, del sistema de memoria de disco el principal foco de optimización en las prestaciones de los computadores.

Como en otras áreas de rendimiento de los computadores, los diseñadores de memorias de disco reconocen que si uno de los componentes solo se puede llevar a un determinado límite, se puede

conseguir una ganancia en prestaciones adicional usando varios de esos componentes en paralelo. En el caso de la memoria de disco, esto conduce al desarrollo de conjuntos de discos que operen independientemente y en paralelo. Con varios discos, las peticiones separadas de E/S se pueden gestionar en paralelo, siempre que los datos requeridos residan en discos separados. Además, se puede ejecutar en paralelo una única petición de E/S si el bloque de datos al que se va a acceder está distribuido a lo largo de varios discos.

Con el uso de varios discos, hay una amplia variedad de formas en las que se pueden organizar los datos, y en las que se puede añadir redundancia para mejorar la seguridad. Esto podría dificultar el desarrollo de esquemas de bases de datos que se pueden usar en numerosas plataformas y sistemas operativos. Afortunadamente, la industria está de acuerdo con los esquemas estandarizados para el diseño de bases de datos para discos múltiples, conocidos como RAID (*Redundant Array of Independent Disks*, conjunto redundante de discos independientes). El esquema RAID consta de seis niveles² independientes, desde cero hasta cinco. Estos niveles no implican una relación jerárquica, sino que designan métodos diferentes que poseen tres características comunes:

1. RAID es un conjunto de unidades físicas de disco vistas por el sistema operativo como una única unidad lógica.
2. Los datos se distribuyen a través de las unidades físicas del conjunto de unidades.
3. La capacidad de los discos redundantes se usa para almacenar información de paridad que garantice la recuperación de los datos en caso de fallo de disco.

Los detalles de las características segunda y tercera cambian según los distintos niveles RAID. RAID 0 no soporta la tercera característica.

El término *RAID* fue originalmente ideado en un artículo de un grupo de investigación de la Universidad de California en Berkley [PATT88]³. El artículo perfilaba varias configuraciones y aplicaciones RAID e introducía las definiciones de los niveles RAID que todavía se usan. La estrategia RAID reemplaza una unidad de disco de gran capacidad por unidades múltiples de menor capacidad y distribuye los datos de forma que se puedan habilitar accesos simultáneos a los datos de varias unidades mejorando, por tanto, las prestaciones de E/S y permitiendo más fácilmente aumentos en la capacidad.

La única contribución de la propuesta RAID es, efectivamente, hacer hincapié en la necesidad de redundancia. El uso de varios dispositivos, además de permitir que varias cabezas y actuadores operen simultáneamente, consiguiendo mayores velocidades de E/S y de transferencia, incrementa la probabilidad de fallo. Para compensar esta disminución de seguridad, RAID utiliza la información de paridad almacenada que permite la recuperación de datos perdidos debido a un fallo de disco.

A continuación examinaremos cada nivel de RAID. La Tabla 6.3, a partir de [MASS97], proporciona una amplia guía sobre los siete niveles. De ellos, los niveles 2 y 4 no se ofrecen

² Algunos investigadores y compañías han definido niveles adicionales, pero los seis niveles descritos en esta sección son los convenidos universalmente.

³ En este artículo, el acrónimo RAID significaba conjunto redundante de discos baratos (*Redundant Array of Inexpensive Disks*). El término barato se usó para contrastar los discos pequeños de los conjuntos RAID, relativamente baratos, frente a la alternativa de discos únicos, grandes y caros (SLED, *Single Large Expensive Disk*). Hoy, el término SLED está obsoleto, y se usan tecnologías similares tanto para configuraciones RAID como no/RAID. De acuerdo con esto, la industria ha adoptado el término *independiente*, para enfatizar que el conjunto RAID proporciona prestaciones adecuadas y mejoras de seguridad.

Tabla 6.3. Niveles RAID.

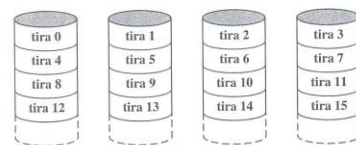
Categoría	Nivel	Descripción	Discos necesarios	Disponibilidad de los datos	Capacidad de transferencia de datos de E/S alta	Velocidad de petición de E/S baja
Estructuras en tiras	0	No redundante	N	Menor que con un solo disco	Muy alta	Muy alta tanto para lectura como para escritura
Estructura en espejo	1	Xxx	$2N, 3N,$ etc.	Xxxx	xxxx	xxx
Acceso paralelo	2	xxxxx	$N + m$	xxx	xxxx	xxxx
	3	xxxx	$N + 1$	xxxxx	xxxxx	xxxx
Acceso independiente	4	xxxxx	$N + 1$	xxx	xxxx	xxxx
	5	xxxxx	$N + 1$	xxx	xxxx	xxxx
	6	xxxx	$N + 2$	xxxxx	xxxx	xxxx

comercialmente y no es probable que consigan aceptación industrial. En la tabla, las prestaciones de E/S se expresan tanto en términos de la capacidad de transferencia de datos, o capacidad para mover datos, como de la velocidad de petición de E/S, o capacidad de atender las peticiones de E/S, ya que estos niveles RAID operan inherentemente de forma distinta según sean estas dos métricas. El punto fuerte de cada nivel RAID se ha destacado sombreándolo. Las Figuras 6.8 a 6.9 muestran el uso de los siete esquemas RAID, que soportan una capacidad de datos para cuatro discos sin redundancia. En las figuras se destaca la organización de los datos del usuario y de los datos redundantes, y se indican los requisitos de almacenaje relativo de los distintos niveles. Nos referiremos a estas figuras a lo largo de la siguiente explicación.

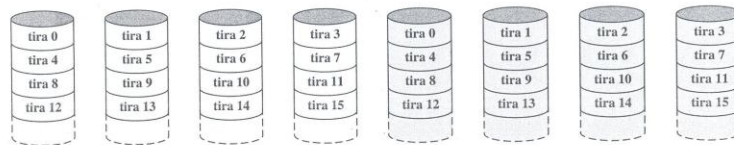
NIVEL 0 DE RAID

El nivel 0 de RAID no es un verdadero miembro de la familia RAID, porque no incluye redundancia para mejorar las prestaciones. Sin embargo, hay algunas aplicaciones, como algunas ejecuciones en supercomputadores, en los que las prestaciones y la capacidad son la preocupación primaria y un costo bajo es más importante que mejorar la seguridad.

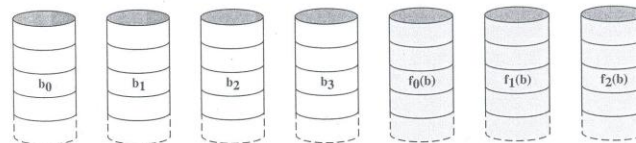
Para el RAID 0, los datos del usuario y del sistema están distribuidos a lo largo de todos los discos del conjunto. Esto tiene una notable ventaja frente al uso de un único y gran disco: si hay pendientes dos peticiones diferentes de E/S, para dos bloques de datos diferentes, entonces es muy



(a) RAID 0 (no redundante)

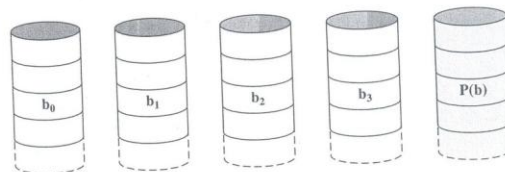


(b) RAID 1 (reflejado)

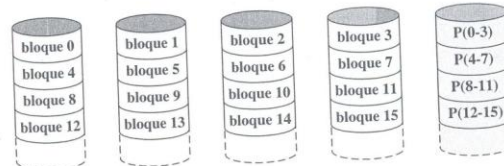


(c) RAID 2 (redundancia con código Hamming)

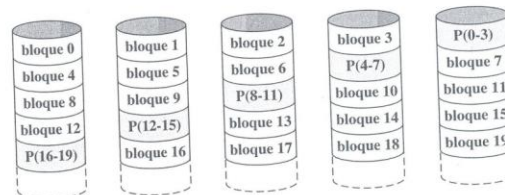
Figura 6.8. Niveles RAID 0 a 3.



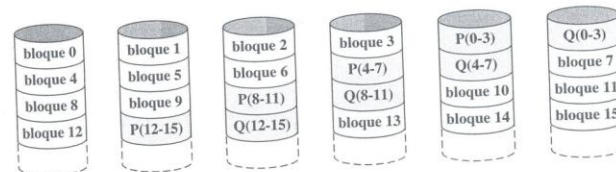
(a) RAID 3 (bit de paridad intercalada)



(b) RAID 4 (paridad en bloques)



(c) RAID 5 (paridad distribuida a nivel de bloque)



(d) RAID 6 (redundancia doble)

Figura 6.9. Niveles RAID del 3 al 6.

probable que los bloques pedidos estén en diferentes discos. Entonces, las dos peticiones se pueden emitir en paralelo, reduciendo el tiempo de cola de E/S.

Pero RAID 0, como todos los niveles RAID, va más lejos que una sencilla distribución de datos a través del conjunto de discos: los datos son *organizados en forma de tiras de datos* a través de los discos disponibles. Esto se entiende mejor considerando la Figura 6.10. Todos los datos del usuario y del sistema se ven como almacenados en un disco lógico. El disco se divide en tiras; estas tiras pueden ser bloques físicos, sectores o alguna otra unidad. Las tiras se proyectan cíclicamente, en

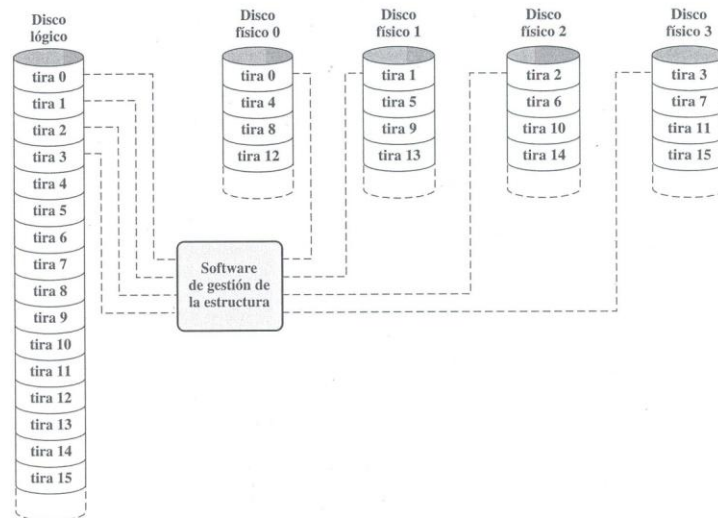


Figura 6.10. Mapa de datos para un conjunto RAID de nivel 0.

miembros consecutivos del conjunto. Un conjunto de tiras lógicamente consecutivas, que se proyectan exactamente sobre una misma tira en cada miembro del conjunto, se denomina franja. En un conjunto de n discos, las primeras n tiras lógicas (una franja) se almacenan físicamente en la primera tira de cada uno de los n discos, las segundas n tiras lógicas, se distribuyen en la segunda tira de cada disco, etc. La ventaja de esta disposición es que si una única petición de E/S implica a varias tiras lógicas contiguas, entonces las n tiras de esta petición se pueden gestionar en paralelo, reduciendo considerablemente el tiempo de transferencia de E/S.

En la Figura 6.10 se indica como el software de gestión de un conjunto proyecta el espacio del disco físico sobre el disco lógico. Este software se puede ejecutar tanto en el subsistema de disco como en un computador anfitrión.

RAID 0 para alta capacidad de transferencia de datos. Las prestaciones de cualquiera de los niveles RAID dependen críticamente de los patrones de petición del sistema anfitrión y de la distribución de los datos. Estas emisiones pueden ser más claramente direccionadas en RAID 0, donde el impacto de la redundancia no interfiere con el análisis. Primero, consideremos el uso de RAID 0 para lograr una velocidad de transferencia de datos alta. Se deben cumplir dos requisitos para que las aplicaciones tengan una velocidad de transferencia alta. Primero, debe existir una capacidad de transferencia alta en todo el camino entre la memoria del anfitrión y las unidades de disco individuales. Esto incluye controladores de buses internos, buses de E/S del anfitrión, adaptadores de E/S, y buses de memoria del anfitrión.

El segundo requisito es que la aplicación debe hacer peticiones de E/S que se distribuyan eficientemente sobre el conjunto de discos. Esta condición se satisface si la petición típica es de una gran cantidad de datos lógicamente contiguos, comparados con el tamaño de una cinta. En este caso, una única petición de E/S implica la transferencia paralela de datos desde varios discos, aumentando la velocidad efectiva de transferencia, en comparación con la de un único disco.

RAID 0 para altas frecuencias de petición de E/S. En los entornos orientados a transacciones, el usuario se suele preocupar más del tiempo de respuesta que de la velocidad de transferencia. Para una petición individual de E/S de una pequeña cantidad de datos, el tiempo de E/S está dominado por el movimiento de las cabezas del disco (tiempo de búsqueda) y el movimiento del disco (latencia rotacional).

En un entorno de transacción, puede haber cientos de peticiones de E/S por segundo. Un conjunto de discos puede proporcionar velocidades altas de ejecución de E/S, balanceando la carga de E/S a través de los distintos discos. El balanceo de la carga efectiva, se consigue solamente si hay varias peticiones de E/S pendientes. Esto, por turnos, implica que hay varias aplicaciones independientes o una única aplicación orientada a transacción que es capaz de generar varias peticiones de E/S asíncronas. Las prestaciones también se verán influidas por el tamaño de la franja. Si la franja es relativamente grande, de forma que una única petición de E/S solo implique un único acceso a disco, entonces las peticiones de E/S que están esperando pueden ser tratadas en paralelo, reduciendo el tiempo en cola para cada petición.

NIVEL 1 DE RAID

RAID 1 se diferencia de los niveles 2 al 6 en cómo se consigue la redundancia. En estos otros esquemas RAID, se usan algunas formas de cálculo de paridad para introducir redundancia; en RAID 1, la redundancia se logra con el sencillo recurso de duplicar todos los datos. Según muestra la Figura 6.8b, se hace una distribución de datos, como en el RAID 0. Pero en este caso, cada franja lógica se proyecta en dos discos físicos separados, de forma que cada disco del conjunto tiene un disco espejo que contiene los mismos datos. RAID 1 también se puede implementar sin franja de datos, pero es menos común.

En la organización RAID 1 hay una serie de aspectos positivos:

1. Una petición de lectura puede ser servida por cualquiera de los discos que contienen los datos pedidos, cualquiera de ellos implica un tiempo de búsqueda mínimo más la latencia rotacional.
2. Una petición de escritura requiere que las dos tiras correspondientes se actualicen, y esto se puede hacer en paralelo. Entonces, el resultado de la escritura viene determinado por la menos rápida de las dos escrituras (es decir, la que conlleva el mayor tiempo de búsqueda más la latencia rotacional). Sin embargo, en RAID 1 no hay «penalización en la escritura». Los niveles RAID del 2 al 6 implican el uso de bits de paridad. Por tanto, cuando se actualiza una única tira, el software de gestión del conjunto debe calcular y actualizar primero los bits de paridad así como actualizar la tira en cuestión.
3. La recuperación tras un fallo es sencilla. Cuando una unidad falla, se puede acceder a los datos desde la segunda unidad.

La principal desventaja es el coste; requiere el doble del espacio de disco del disco lógico que puede soportar. Debido a esto, una configuración RAID 1 posiblemente está limitada a unidades que almacenan el software del sistema y los datos, y otros ficheros altamente críticos. En estos casos, RAID proporciona una copia de seguridad en tiempo real de todos los datos, de forma que en caso de fallo de disco, todos los datos críticos están inmediatamente disponibles.

En un entorno orientado a transacciones, RAID 1 puede conseguir altas velocidades de petición de E/S si la mayor parte de las peticiones son lecturas. En esta situación, las prestaciones de RAID 1 son próximas al doble de las de RAID 0. Sin embargo, si una parte importante de las peticiones de E/S son peticiones de escritura, entonces la ganancia en prestaciones sobre RAID 0 puede no ser significativa. RAID 1 puede también proporcionar una mejora en las prestaciones de RAID 0 en aplicaciones de transferencia intensiva de datos con un alto porcentaje de lecturas. Se produce una mejora si la aplicación puede dividir cada petición de lectura de forma que ambos miembros del disco participen.

NIVEL 2 DE RAID

Los niveles 2 y 3 de RAID usan una técnica de acceso paralelo. En un conjunto de acceso paralelo, todos los discos miembro participan en la ejecución de cada petición de E/S. Típicamente, el giro de cada unidad individual está sincronizado de forma que cada cabeza de disco está en la misma posición en cada disco en un instante dado.

Como en los otros esquemas RAID, se usa la descomposición de datos en tiras. En el caso de RAID 2 y 3, las tiras son muy pequeñas, a menudo tan pequeñas como un único byte o palabra. Con RAID 2, el código de corrección de errores se calcula a partir de los bits de cada disco, y los bits del código se almacenan en las correspondientes posiciones de bit en varios discos de paridad. Normalmente, se usa el código Hamming, que permite corregir errores en un bit y detectar errores en dos bits.

Aunque RAID 2 requiere menos discos que RAID 1, es todavía bastante caro. El número de discos redundantes es proporcional al logaritmo del número de discos de datos. En una sola lectura, se accede a todos los discos simultáneamente. El controlador del conjunto proporciona los datos pedidos y el código de corrección de errores asociado. Si hay un error en un solo bit, el controlador lo puede reconocer y corregir instantáneamente, con lo que el tiempo de acceso a lectura no se ralentiza. En una escritura sencilla, la operación de escritura debe acceder a todos los discos de datos y de paridad.

RAID 2 debería ser solamente una elección efectiva en un entorno en el que haya muchos errores de disco. Si hay una alta seguridad en los discos individuales y en las unidades de disco, RAID 2 es excesivo y no se implementa.

NIVEL 3 DE RAID

RAID 3 se organiza de manera similar a RAID 2. La diferencia es que RAID 3 requiere solo un disco redundante, sin importar lo grande que sea el conjunto de discos. RAID 3 utiliza un acceso paralelo, con datos distribuidos en pequeñas tiras. En vez de un código de corrección de errores, se calcula un sencillo bit de paridad para el conjunto de bits individuales en la misma posición en todos los discos de datos.

Redundancia. En el caso de un fallo en una unidad, se accede a la unidad de paridad y se reconstruyen los datos desde el resto de los dispositivos. Una vez que se sustituye la unidad que ha fallado, los datos que faltan se restauran en la nueva unidad y se reanuda la operación.

La reconstrucción de los datos es bastante sencilla. Consideremos un conjunto de cinco discos de los que de X0 a X3 contienen datos y X4 es el disco de paridad. La paridad para el i -ésimo bit se calcula de la siguiente forma:

$$X4(i) = X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$$

donde \oplus es la función exclusive-OR.

Supongamos que la unidad X1 ha fallado. Si sumamos $X4(i) \oplus X1(i)$ a ambos miembros de la ecuación, tenemos que:

$$X1(i) = X4(i) \oplus X3(i) \oplus X2(i) \oplus X0(i)$$

Por lo tanto, se puede regenerar el contenido de cualquier tira de datos en X1 a partir del contenido de las correspondientes tiras del resto de los discos del conjunto. Este principio es válido para los niveles 3 a 6 de RAID.

Caso de que un disco falle, todos los datos estarán todavía disponibles en lo que se denomina modo reducido. En este modo, para lecturas, los datos que faltan se recuperan «al vuelo» con la operación *exclusive-or*. Cuando se escriben datos en un conjunto RAID 3 reducido, se debe mantener la consistencia de la paridad para regeneraciones posteriores. Volviendo al funcionamiento global, se requiere que el disco que ha fallado se reemplace y se regenere todo su contenido en el nuevo disco.

Prestaciones. Puesto que los datos se dividen en tiras muy pequeñas, RAID 3 puede conseguir velocidades de transferencia de datos muy altas. Cualquier petición de E/S implicará una transferencia de datos paralela desde todos los discos de datos. Para grandes transferencias, la mejora de prestaciones es especialmente notable. Por otra parte, solo se puede ejecutar a la vez una petición de E/S. Por tanto, en un entorno orientado a transacciones, el rendimiento sufre.

NIVEL 4 DE RAID

Los niveles 4 al 6 de RAID usan una técnica de acceso independiente. En un conjunto de acceso independiente, cada disco opera independientemente, de forma que peticiones de E/S separadas se atienden en paralelo. Debido a esto, son más adecuados los conjuntos de acceso independiente para aplicaciones que requieren velocidades de petición de E/S altas, y son menos adecuados para aplicaciones que requieren velocidades altas de transferencia de datos.

Como en otros esquemas RAID, se usan tiras de datos. En el caso de RAID 4 a 6, las tiras son relativamente grandes. Con RAID 4, se calcula una tira de paridad bit a bit a partir de las correspondientes tiras de cada disco de datos, y los bits de paridad se almacenan en la correspondiente tira del disco de paridad.

RAID 4 lleva consigo una penalización en la escritura cuando se realiza una petición de escritura de E/S pequeña. Cada vez que se realiza una escritura, el software de gestión del conjunto debe actualizar no solo los datos del usuario, sino también los bits de paridad correspondientes. Consideremos un conjunto de cinco unidades en las que de X0 a X3 contienen datos y X4 es el disco de paridad. Supongamos que se realiza una escritura que implica solo una tira del disco X1.

(Continúa)

Inicialmente, para cada bit i , tenemos la siguiente relación:

$$X4(i) = X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$$

Después de la actualización, indicamos con prima los bits que han sido alterados:

$$\begin{aligned} X4'(i) &= X3(i) \oplus X2(i) \oplus X1'(i) \oplus X0(i) \\ &= X3(i) \oplus X2(i) \oplus X1'(i) \oplus X0(i) \oplus X1(i) \oplus X1(i) \\ &= X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i) \oplus X1(i) \oplus X1'(i) \\ &= X4(i) \oplus X1(i) \oplus X1'(i) \end{aligned}$$

El conjunto anterior de ecuaciones se ha obtenido de la siguiente forma. La primera línea muestra que un cambio en $X1$ también afecta a la paridad del disco $X4$. En la segunda línea, se han añadido los términos $[\oplus X1(i) \oplus X1(i)]$. Como la XOR de cualquier número consigo mismo es 0, no afecta a la ecuación. Sin embargo, esto se usa por conveniencia, para crear una tercera línea reordenando. Finalmente, la Ecuación (11.1) se usa para reemplazar los cuatro primeros términos por $X4(i)$.

Para calcular la nueva paridad, el software de gestión del conjunto debe leer la antigua tira del usuario y la antigua tira de paridad. Entonces, se pueden actualizar estas dos tiras con nuevos datos y calcular la nueva paridad. Por tanto, cada escritura de una tira implica dos lecturas y dos escrituras.

En el caso de una escritura de E/S de mayor tamaño que implique tiras en todas las unidades de disco, la paridad se puede obtener fácilmente con un cálculo usando solamente los nuevos bits de datos. Por tanto, la unidad de paridad puede ser actualizada en paralelo con las unidades de datos, y no habrá lecturas o escrituras extra.

En cualquier caso, cada operación de escritura implica al disco de paridad, que por consiguiente se convertirá en un cuello de botella.

NIVEL 5 DE RAID

RAID 5 está organizado de manera similar a RAID 4. La diferencia es que RAID 5 distribuye las tiras de paridad a lo largo de todos los discos. Una distribución típica es un esquema cíclico, como se muestra en la Figura 6.9c. Para un conjunto de n discos, la tira de paridad está en diferentes discos para las primeras n tiras, y este patrón se repite.

La distribución de las tiras de paridad a lo largo de todas las unidades evita el potencial cuello de botella de E/S encontrado en RAID 4.

NIVEL 6 DE RAID

El nivel 6 de RAID se introdujo en un artículo de los investigadores de Berkeley [KATZ89]. En el esquema del nivel 6 de RAID, se hacen dos cálculos de paridad distintos, que se almacenan en bloques separados en distintos discos. Por tanto, un conjunto RAID 6 cuyos datos requieran N discos consta de $N + 2$ discos.

La figura 6.9d ilustra este esquema. P y Q son dos algoritmos de comprobación de datos distintos. Uno de los dos calcula la *exclusive-OR* usada en los niveles de 4 y 5 de RAID. Pero el otro es

un algoritmo de comprobación de datos independiente. Esto hace posible la regeneración de los datos incluso si dos de los discos que contienen los datos de los usuarios fallan.

La ventaja del RAID 6 es que proporciona una disponibilidad de los datos extremadamente alta. Tendrían que fallar tres discos en el intervalo MTTR (tiempo medio de reparación) para no poder disponer de los datos. Por otra parte, RAID 6 incurre en una penalización de escritura ya que cada escritura afecta a dos bloques de paridad.

La Tabla 6.4 es un resumen comparativo de los siete niveles.

Tabla 6.4. Comparación de RAID.

Nivel	Ventajas	Inconvenientes	Aplicaciones
0	Las prestaciones de E/S se mejoran mucho repartiendo la carga de E/S entre varios canales unidades. No hay cálculo de paridad de cabecera. Diseño muy sencillo. Fácil de implementar.	El fallo de una sola unidad afectará a todos los datos de una estructura, perdiéndose.	Producción y edición de video. Edición de imágenes. Aplicaciones de pruebas de imprenta. Cualquier aplicación que requiera ancho de banda grande.
1	Una redundancia del cien por cien de los datos implica que no sea necesaria la reconstrucción en caso de fallo de disco, solo una copia del disco a reemplazar. Bajo ciertas circunstancias RAID 1 puede soportar varios fallos de unidades. El diseño del subsistema de almacenamiento RAID más sencillo.	La mayor sobrecarga de todos los tipos RAID (100%) ineficiente.	Contabilidad. Nóminas. Finanzas. Cualquier aplicación que requiera una disponibilidad muy alta.
2	Son posibles velocidades de transferencia de datos extremadamente altas. Cuanto mayor es la velocidad de transferencia requerida, mejor es la relación entre discos de datos y discos ECC. Diseño del controlador relativamente sencillo en comparación con los de los niveles 3, 4 y 5.	Relación muy alta entre discos ECC y discos de datos con tamaños de palabra pequeños (ineficiente). Coste del nivel de entrada muy alto (requisitos de velocidades de transferencia muy altas para justificarlo).	No existen implementaciones comerciales / no es comercialmente viable.
3	Velocidad de transferencia de datos de lectura muy alta. Velocidad de transferencia de datos de escritura muy alta. Un fallo de disco tiene un impacto insignificante en el rendimiento. Una baja relación entre discos ECC (paridad) y discos de datos implica una alta eficiencia.	Velocidad de transacción igual que la de una única unidad de disco como mucho (si la velocidad de giro está sincronizada). El diseño del controlador es bastante complejo.	Producción de video y secuencias en vivo. Edición de imagen. Edición de video. Aplicaciones de prueba de imprenta. Cualquier aplicación que requiera un alto rendimiento.

(Continúa)

Tabla 6.4. Comparación de RAID (continuación).

Nivel	Ventajas	Inconvenientes	Aplicaciones
4	Velocidad de transacción de datos de lectura muy alta. Una baja relación entre discos ECC (paridad) y discos de datos implica una alta eficiencia.	Diseño del controlador bastante complejo. Peor velocidad de transacción de escritura y velocidad de transferencia total de escritura. Reconstrucción de datos difícil e ineficiente en caso de fallo de disco.	No existe implementación comercial / no es comercialmente viable.
5	La mayor velocidad de transacción de datos. Una baja relación entre discos ECC (paridad) y discos de datos implica una alta eficiencia. Buena velocidad de transferencia en su conjunto.	Diseño del controlador más complejo. Es difícil la reconstrucción en caso de fallo de disco (comparado con el nivel 1 de RAID).	Servidores de ficheros y aplicaciones. Servidores de bases de datos. Servidores de páginas web, correo electrónico y noticias. Servidores Intranet. Nivel RAID más versátil.
6	Proporciona una tolerancia a fallos extremadamente alta y puede soportar varios fallos de unidades simultáneos.	Diseño del controlador más complejo. La sobrecarga del controlador para calcular las direcciones de paridad es extremadamente alta.	Solución perfecta para aplicaciones con objetivos críticos.

6.3. MEMORIA ÓPTICA

En 1983, se introdujo uno de los productos de consumo de más éxito de todos los tiempos: el disco compacto (CD, *Compact Disk*) digital de audio. El CD es un disco no borrrable que puede almacenar más de sesenta minutos de información de audio en una cara. El gran éxito comercial del CD posibilitó el desarrollo de la tecnología de discos de memoria óptica de bajo coste, que revolucionó el almacenamiento de datos en un computador. Se han introducido una gran variedad de discos ópticos (Tabla 6.5). Vamos a ver cada uno de ellos brevemente.

DISCOS COMPACTOS

CD-ROM. Tanto el CD de audio como el CD-ROM (*compact disk read-only memory*, memoria de disco compacto de solo-lectura) comparten una tecnología similar. La principal diferencia es que los lectores de CD-ROM son más robustos y tienen dispositivos de corrección de errores para asegurar que los datos se transfieren correctamente del disco al computador. Ambos tipos de disco se hacen también de la misma forma. El disco se forma a partir de una resina, como un policarbonato.

Tabla 6.5. Discos ópticos.

CD	Disco compacto. Un disco no borrable que almacena información de audio digitalizada. El sistema estándar usa discos de doce cm y puede grabar más de sesenta minutos de tiempo de ejecución ininterrumpido.
CD-ROM	Disco compacto de memoria de solo-lectura. Un disco no borrable usado como memoria de datos de un computador. El sistema estándar usa discos de doce cm y puede guardar más de 650 MB.
DVD	Disco versátil digital. Una tecnología para producir representación de información de vídeo digitalizada y comprimida, así como grandes cantidades de otros datos digitales. Se usan en formatos de ocho y doce cm de diámetro, con una capacidad con doble cara de hasta 17 GB. El DVD básico es de solo-lectura (DVD-ROM).
DVD-R	DVD grabable. Es similar al DVD-ROM. El usuario puede escribir en el disco solo una vez. Solo se utilizan discos de una cara.
DVD-RW	DVD grabable. Es similar al DVD-ROM. El usuario puede borrar y reescribir el disco varias veces. Solo se utilizan discos de una cara.

La información grabada digitalmente (ya sea música o datos del computador) se graba como una serie de hoyos microscópicos en la superficie reflectante. Esto se hace, primero de todo, con un láser de alta intensidad y enfocado con precisión, para crear el disco patrón. El patrón se usa, sin embargo para hacer una matriz para estampar copias en policarbonato. La superficie con los hoyos se cubre con una superficie altamente reflectante como aluminio u oro. Esta superficie brillante se protege contra el polvo y los arañazos con una última capa de laca transparente. Finalmente, se puede imprimir una etiqueta sobre la laca.

La información del CD o CD-ROM se recupera con un láser de baja potencia situado en un lector o unidad de disco óptico. El láser pasa a través de la capa protectora transparente mientras un motor hace girar el disco sobre el láser (Figura 6.11). La intensidad de la luz reflejada cambia si se encuentra un hoyo. En concreto, si el haz de láser cae sobre un hoyo, que de alguna manera es una superficie rugosa, la luz se dispersa y una luz de baja intensidad llega a la fuente. Las áreas entre hoyos se llaman *valles*. Un valle es una superficie lisa, que refleja con mayor intensidad. El cambio entre hoyos y valles es detectado por un fotosensor y convertido en una señal digital. El sensor barre la superficie a intervalos regulares. El principio o fin de un hoyo representa un 1; cuando no hay cambios en la altura entre intervalos, se graba un 0.

Recordemos que en un disco magnético, la información se graba en pistas concéntricas. Con el sistema de velocidad angular constante (CAV), el número de bits por pista es constante. Se puede conseguir un incremento en la densidad con la grabación de varias zonas, en la que la superficie se divide en una serie de zonas, de forma que las zonas lejanas al centro contienen más bits que las zonas cercanas al mismo. Aunque esta técnica incrementa la capacidad, no está todavía optimizada.

Para conseguir mayor capacidad, los CD y CD-ROM no se organizan en pistas concéntricas. En su lugar, el disco contiene una única pista en espiral, que comienza en el centro y se extiende hacia el

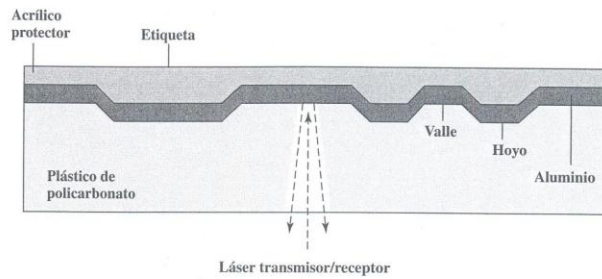


Figura 6.11. Funcionamiento de un CD.

borde del disco. Los sectores cercanos al filo del disco tienen la misma longitud que los cercanos al centro. Por tanto, la información está empaquetada uniformemente a lo largo del disco en segmentos del mismo tamaño y son escaneados a la misma velocidad rotando el disco a velocidad variable. Los hoyos son leídos por un láser a **velocidad lineal constante (CLV)**. El disco rota más despacio en los accesos cercanos al filo externo que en los cercanos al centro. Por tanto, la capacidad de una pista y el retardo rotacional es mayor cercano al centro. La capacidad de un CD-ROM es de unos 650 MB.

Los datos de un CD-ROM se organizan en una secuencia de bloques. En la Figura 6.12 se muestra un formato típico de un bloque. Este consta de los siguientes campos:

- **Sincronización:** el campo de sincronización identifica el principio de un bloque. Consta de un byte de 0s, 10 bytes de 1s, y un byte de 0s.
- **Cabecera:** la cabecera contiene la dirección del bloque y el byte de modo. El modo 0 especifica un campo de datos en blanco; el modo 1 especifica el uso de un código de corrección de errores y 2048 bytes de datos; el modo 2 especifica 2336 bytes de datos del usuario sin código de corrección de errores.
- **Datos:** datos del usuario.
- **Auxiliar:** datos del usuario adicionales, en modo 2. En modo 1, es un código de corrección de errores de 288 bytes.

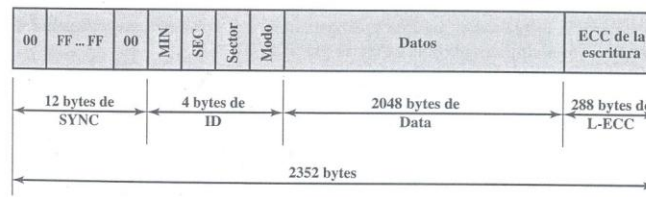


Figura 6.12. Formato de los bloques de un CD-ROM.

Usando CLV, el acceso aleatorio se hace más difícil. Localizar una dirección específica implica mover la cabeza al área general, ajustar la velocidad de rotación y leer la dirección, y hacer pequeños ajustes para encontrar y acceder al sector específico.

Los CD-ROM son apropiados para la distribución de grandes cantidades de datos a un gran número de usuarios. Debido al gasto del proceso inicial de escritura, no es adecuado para aplicaciones individuales. Comparado con los discos magnéticos tradicionales, el CD-ROM tiene dos ventajas:

- El disco óptico junto con la información almacenada en él, se puede replicar en grandes cantidades de forma barata (a diferencia de los discos magnéticos). Las bases de datos en un disco magnético se reproducen copiando uno a uno, usando dos unidades de disco.
- El disco óptico es extraíble, permitiendo usar el mismo disco como memoria de archivo. La mayoría de los discos magnéticos no son extraíbles. La información que contiene tiene que copiarse en una cinta antes de que se pueda usar la unidad de disco/disco para almacenar nueva información.

Las desventajas del CD-ROM son:

- Es de solo lectura y no se puede actualizar.
- El tiempo de acceso es mayor que el de una unidad de disco magnético, tanto como medio segundo.

CD grabable. Para adaptarse a aplicaciones en las que solo se necesitan unas pocas copias de un conjunto de datos, se han desarrollado los CD de una escritura y varias lecturas, conocido como CD grabable (CD-R, *CD-recordable*). Para hacer un CD-R, se prepara un disco de forma que se pueda escribir una vez con un haz láser de intensidad modesta. De esta forma, con algún controlador de disco especial, más caro que para CD-ROM, el cliente puede escribir una vez, además de leer el disco.

El material del CD-R es similar pero no idéntico al de un CD o CD-ROM. En los CD y CD-ROM, la información se graba haciendo pequeños agujeros en la superficie del material, de forma que cambie su reflectividad. En un CD-R, el medio incluye una capa de tinte. El tinte se utiliza para cambiar la reflectividad y se activa con un láser de alta intensidad. El disco resultante se puede leer en una unidad de CD-R o CD-ROM.

El disco óptico CD-R resulta atractivo como almacén de documentos y ficheros. Proporciona una copia permanente para gran cantidad de datos.

CD regrabable. El disco óptico CD-RW se puede escribir y reescribir como un disco magnético. A pesar de las numerosas técnicas que se han probado, la única puramente óptica que ha conseguido ser atractiva se denomina **cambio de fase**. El disco de cambio de fase utiliza un material que presenta dos tipos de reflexión, significativamente diferentes, en dos estados diferentes. Hay un estado amorfo, en el que las moléculas presentan una orientación aleatoria y que refleja mal la luz; y un estado cristalino, que presenta una superficie lisa que refleja bien la luz. Un haz de láser puede cambiar el material de una fase a otra. La principal desventaja del cambio de fase de los discos ópticos es que el material finalmente y de forma permanente pierde sus propiedades. Los materiales actuales se pueden borrar entre 500 000 y un millón de veces.

Los CD-RW tienen la ventaja obvia sobre los CD-ROM y CD-R que se pueden regrabar y por tanto usarse como verdaderos almacenes secundarios. Por tanto, compiten con los discos

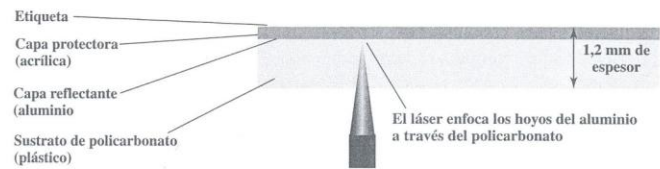
magnéticos. Una ventaja clave de los discos ópticos es que las tolerancias de los parámetros de construcción en los discos ópticos es mucho más severa que para los discos magnéticos de gran capacidad. Por tanto, tienen una mayor fiabilidad y vida.

DISCO DIGITAL VERSÁTIL

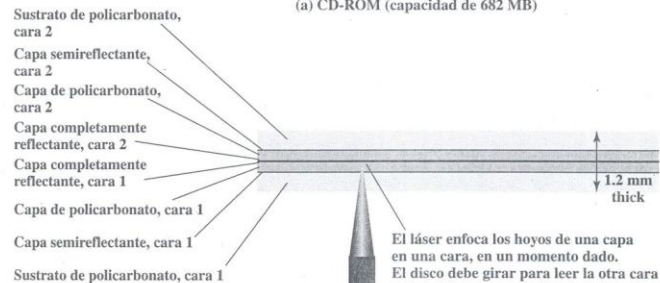
Con la gran capacidad de almacenamiento del disco digital versátil (DVD, *Digital Video Disk*), la industria de la electrónica ha encontrado por fin un sustituto razonable de las cintas VHS de vídeo analógicas. El DVD sustituirá a las cintas de vídeo usadas en los reproductores de vídeo (VCR) y, lo que es más importante para este texto, sustituirá al CD-ROM en los PC y servidores. El DVD lleva al vídeo a la edad digital. Proporciona películas con una calidad de imagen impresionante, y se puede acceder a ellos aleatoriamente como en los CD de audio, que pueden también leer los DVD. En un disco se puede grabar un gran volumen de datos, en la actualidad siete veces más que en un CD-ROM. Con esta gran capacidad de almacenamiento y alta calidad de los DVD, los juegos para PC serán más reales y el software educativo incorporará más vídeo. Como consecuencia de estos desarrollos habrá un nuevo pico en el tráfico en Internet e intranets corporativas, ya que este material se incorporará a los sitios web.

La mayor capacidad del DVD se debe a tres diferencias respecto al CD (Figura 6.13):

1. Los bits se empaquetan más juntos en un DVD. El espacio entre las vueltas de una espiral en un CD es de $1,6 \mu\text{m}$ y la distancia mínima entre hoyos a lo largo de la espiral es de $0,834 \mu\text{m}$.



(a) CD-ROM (capacidad de 682 MB)



(b) DVD-ROM, doble superficie-doble cara (capacidad 17 GB)

Figura 6.13. CD-ROM y DVD-ROM.

s de cons-
an capaci-

, la indus-
nológicas.
e es más
a la edad
ellos ale-
se puede
esta gran
y el soft-
jevo pico
tios web.

spiral en
,834 μm .

El DVD utiliza un láser con una longitud de onda menor y consigue un espaciado entre vueltas de $0,74 \mu\text{m}$ y una distancia mínima entre hoyos de $0,4 \mu\text{m}$.

El resultado de estas dos mejoras supone un incremento de capacidad en un factor de siete, de alrededor de 4,7 GB.

2. El DVD utiliza una segunda capa de hoyos y valles sobre la primera capa. Un DVD de doble capa tiene una capa semirreflectante sobre la capa reflectante, y, ajustando el enfoque, el láser de la unidad de DVD puede leer cada capa por separado. Esta técnica casi dobla la capacidad del disco, hasta 8,5 GB. La baja reflectividad de la segunda capa limita su capacidad de almacenamiento por lo que no se consigue doblar la capacidad.
3. El DVD-ROM puede tener dos superficies, mientras que en un CD los datos se graban solo en una superficie. Esto da una capacidad total de más de 17 GB.

Como con los CD, los DVD tienen versiones grabables y de solo-lectura (Tabla 6.5).

6.4. CINTA MAGNÉTICA

Los sistemas de cinta usan las mismas técnicas de lectura y grabación que los discos. El medio es una cinta de poliéster flexible (parecido al usado en ropa) cubierta por un material magnetizable. La cubierta puede consistir en partículas de un metal puro en concreto un revestimiento o película de metal plateado vaporizado. La cinta y la unidad de cinta son análogas a las cintas de grabación domésticas. Los anchos de las cintas pueden variar entre 0,38 cm (0,15 pulgadas) y 1,27 cm (0,5 pulgadas). Una cinta ubicada en un carrete abierto tienen que enrollarse en otro carrete ubicado en un segundo cabezal. Hoy día, prácticamente todas las cintas vienen cerradas en cartuchos.

Los datos en la cinta, se estructuran en una serie de pistas paralelas longitudinales. Los primeros sistemas de cintas usaban nueve pistas. Esto hace posible almacenar datos de un byte en un instante dado, con un bit de paridad adicional, en la novena pista. Los nuevos sistemas de cintas usan 18 o 36 pistas, correspondiendo a una palabra o doble palabra digital. La grabación de datos de esta forma se denomina **grabación paralela**. Los sistemas más modernos utilizan en su lugar **grabación serie**, en la que los datos se disponen como una secuencia de bits a lo largo de cada pista, como se hace en los discos magnéticos. Como con el disco, los datos se leen y escriben en bloques contiguos, llamados *registros físicos* de cinta. Los bloques en la cinta están separados por bandas vacías llamadas *bandas interregistros*. Como en el disco, la cinta se formatea para facilitar la localización de los registros físicos.

La técnica típica utilizada en la grabación de cintas en serie se denomina **grabación en serpentina**. En esta técnica, cuando se graban los datos, el primer conjunto de bits se graba a lo largo de toda la cinta. Cuando se alcanza el fin, las cabezas se posicionan para grabar una nueva pista y la cinta se graba de nuevo a todo lo largo, esta vez en dirección contraria. Este proceso continua, hacia atrás y hacia delante, hasta que la cinta se llena (Figura 6.14a). Para aumentar la velocidad, la cabeza de lectura-escritura es capaz de leer y escribir una serie de pistas adyacentes simultáneamente (usualmente entre dos y ocho pistas). Los datos se graban en serie a lo largo de las pistas individuales, pero los bloques se almacenan en pistas adyacentes, como se sugiere en la Figura 6.14b. La Tabla 6.6 muestra los parámetros de un sistema, conocido como cinta DLT.

Una unidad de cinta es un dispositivo de *acceso secuencial*. Si la cabeza de la cinta se posiciona en el registro 1, entonces para leer el registro N , es necesario leer los registros físicos del 1 al $N-1$.

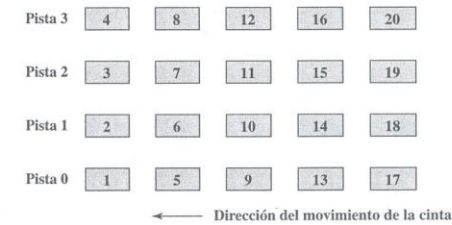


Figura 6.14. Características de una cinta magnética típica.

Tabla 6.6. Unidades de cintas DLT.

	DLT 4000	DLT 8000	SDLT 600
Capacidad (GB)	20	40	300
Velocidad de transferencia (MB/s)	1.5	6.0	36.0
Densidad de bits (Kb/cm)	32.3	38.6	92
Densidad de pistas (p/cm)	101	164	587
Longitud media (m)	549	549	597
Ancho medio (cm)	1.27	1.27	1.27
Número de pistas	128	208	448
Número de pistas de lectura/escritura simultáneas	2	4	8

uno a uno. Si la cabeza está actualmente situada más allá del registro deseado, es necesario rebobinar la cinta una cierta distancia y empezar a leer hacia delante. A diferencia del disco, la cinta está en movimiento solamente durante las operaciones de lectura o escritura.

En contraste con las cintas, a la unidad de disco se le llama dispositivo de *acceso directo*. Una unidad de disco no necesita leer todos los sectores de un disco secuencialmente para llegar al sector deseado.

Solo debe esperar a los sectores dentro de una pista y puede hacer accesos sucesivos a cualquier pista.

Las cintas magnéticas fueron el primer tipo de memorias secundarias. Se usan todavía ampliamente como los miembros de la jerarquía de memoria de menor coste y de menor velocidad.

6.5. LECTURAS Y SITIOS WEB RECOMENDADOS

[MEE96a] ofrece un buen resumen de la tecnología de grabación subyacente de los discos y cintas. [MEE96b] se centra en las técnicas de almacenamiento de datos en discos y cintas.

[COME00] es un artículo corto pero instructivo sobre las tendencias actuales en tecnologías de almacenamiento en discos magnéticos.

Un excelente estudio sobre la tecnología RAID, escrito por los inventores del concepto RAID, está en [CHEN94]. Un buen artículo resumen es [FRIE96]. Una buena comparación de las arquitecturas RAID se encuentra en [CHEN96].

[MARC90] da una excelente visión del campo de las memorias ópticas. Un buen examen de las tecnologías subyacentes de grabación y de lectura es [MANS97].

[ROSC03] proporciona una visión comprensiva de todos los tipos de memorias externas, con una modesta cantidad de detalles técnicos de cada uno. [KHUR01] es otra buena revisión.

ANDE03 ANDERSON, D.: «You Don't Know Jack About Disks». *ACM Queue*, junio, 2003.

CHEN94 CHEN, P.; LEE, E.; GIBSON, G.; KATZ, R. y PATTERSON, D.: «RAID: High: High-Performance, Reliable Secondary Storage». *ACM Computing Surveys*, junio, 1994.

CHEN96 CHEN, S. y TOWSLEY, D.: «A Performance Evaluation of RAID Architectures». *IEEE Transactions on Computers*, octubre, 1996.

COME00 COMERFONRD, R.: «Magnetic Storage: The Medium that Wouldn't Die». *IEEE Spectrum*, diciembre, 2000.

FRIE96 FRIEDMAN, M.: «RAID Keeps Going and Going and ...». *IEEE Spectrum*, abril, 1996.

KHUR01 KHURSHUDOV, A.: *The Essential Guide to Computer Data Storage*. Upper Saddle River, NJ. Prentice Hall, 2001.

MANS97 MANSURIPUR, M. y SINCERBOX, G.: «Principles and Techniques of Optical Data Storage». *Proceedings of the IEEE*, noviembre, 1997.

MAR90 MARCHANT, A.: *Optical Recording*. Reading, MA. Addison-Wesley, 1990.

MEE96a MEE, C. y DANIEL, E. eds.: *Magnetic Recording Technology*. New York. McGraw-Hill, 1996.

MEE96b MEE, C. y DANIEL, E. eds.: *Magnetic Storage Handbook*. New York. McGraw-Hill, 1996.

ROSC03 ROSCH, W.: *Winn L. Rosch Hardware Bible*. Indianapolis, IN. Que Publishing, 2003.

600

rebobi-
nta está

to. Una
el sector



SITIOS WEB RECOMENDADOS

- **Asociación de tecnología óptica de almacenamiento:** buena fuente de información sobre tecnologías de almacenamiento óptico y vendedores, más una extensa lista de enlaces importantes.
- **Cintas DLT:** buena colección de información técnica y enlaces a vendedores.

6.6. PALABRAS CLAVE, PREGUNTAS DE REPASO Y PROBLEMAS

PALABRAS CLAVE

banda	disco magnético	pista
cabeza	disco no extraíble	Plato
CD	disquete	RAID
CD-R	DVD	retardo rotacional
CD-ROM	DVD-R	Sector
CD-RW	DVD-ROM	sustrato
cilindro	DVD-RW	tiempo de acceso
cinta magnética	grabación en serpentina	tiempo de búsqueda
datos divididos	Grabación en varias zonas	tiempo de transferencia
disco de cabeza extraíble	hoyo	valle
disco de cabeza fija	magnetorresistivo	velocidad angular constante (CAV)
disco extraíble	memoria óptica	velocidad lineal constante (CLV)

PREGUNTAS DE REPASO

- 6.1. ¿Cuáles son las ventajas de usar un sustrato de cristal en un disco magnético?
- 6.2. ¿Cómo se escriben los datos en un disco magnético?
- 6.3. ¿Cómo se leen los datos en un disco magnético?
- 6.4. Explicar la diferencia entre un sistema de grabación CAV y de varias zonas.
- 6.5. Definir los términos *pista*, *cilindro* y *sector*.
- 6.6. ¿Cuál es el tamaño típico de un sector en un disco?
- 6.7. Definir los términos *tiempo de búsqueda*, *retardo rotacional*, *tiempo de acceso* y *tiempo de transferencia*.
- 6.8. ¿Qué características comunes comparten todos los niveles RAID?
- 6.9. Definir brevemente los siete niveles RAID.
- 6.10. Explicar el término *datos divididos*.
- 6.11. ¿Cómo se consigue redundancia en un sistema RAID?
- 6.12. En el contexto de RAID, ¿cuál es la diferencia entre acceso paralelo y acceso independiente?
- 6.13. ¿Cuál es la diferencia entre CAV y CLV?
- 6.14. ¿En qué se diferencia un CD de un DVD en lo que respecta a la capacidad de este último?
- 6.15. Explicar la grabación en serpentina.

PROBLEMAS

- 6.1. Considérese un disco con N pistas numeradas desde 0 hasta $(N - 1)$ y suponer que los sectores requeridos están distribuidos aleatoria y uniformemente a lo largo del disco. Calcular el número medio de pistas atravesadas en una búsqueda.

