



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 366 591**

51 Int. Cl.:
G06F 3/06 (2006.01)
H04N 5/76 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07732004 .2**
96 Fecha de presentación : **12.03.2007**
97 Número de publicación de la solicitud: **1999554**
97 Fecha de publicación de la solicitud: **10.12.2008**

54 Título: **Sistema de almacenamiento de disco duro.**

30 Prioridad: **28.03.2006 US 786364 P**

45 Fecha de publicación de la mención BOPI:
21.10.2011

45 Fecha de la publicación del folleto de la patente:
21.10.2011

73 Titular/es: **VERACITY UK LIMITED**
6 Barns Street
Ayr KA7 1XA, GB

72 Inventor/es: **Mcleod, Alastair, Campbell y**
Sprucefield, Chris

74 Agente: **Elzaburu Márquez, Alberto**

ES 2 366 591 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Sistema de almacenamiento en disco duro.

5 El presente invento se refiere al campo de sistemas de almacenamiento de datos y en particular a un método y aparato perfeccionados para almacenamiento de datos en unidades de disco duro (HDD) que emplean técnicas de grabación de datos secuencial para aplicaciones de grabación secuencial.

Las unidades de disco duro modernas son generalmente muy fiables para aplicaciones de almacenamiento de datos típicas. En los últimos años recientes, las aplicaciones para el almacenamiento intensivo de datos tales como sistemas [DVR] de grabación de video digitales han aprovechado la velocidad y capacidad de las HDD para almacenamiento de video, audio y otros datos generados sobre una base estrictamente secuencial.

10 Por ejemplo, la solicitud de Patente Europea EP 1.178.390 describe el aumento de la fiabilidad para un funcionamiento continuo para un dispositivo de almacenamiento de información que tiene una función de grabación de información de entrada de modo continuo.

15 Estas aplicaciones intensivas de datos solicitan a las HDD hasta su límite de tal modo que la ocurrencia de eventos de fallo del disco han resultado más comunes, si no evitables, en tanto en cuanto de hecho, es generalmente aceptado por los expertos en la técnica que la HDD es el punto más débil en la fiabilidad de cualquier sistema mayor de almacenamiento de datos basado en el disco.

20 A fin de mitigar el problema de eventos de fallo del disco cuando el volumen de datos generado por los procesos de grabación es grande, y/o el período de almacenamiento requerido (antes de la sobre-escritura, si la hay) es prolongado, entonces se emplean típicamente grupos o disposiciones de HDD para almacenar los datos. Cuando se emplean cantidades de HDD, la probabilidad de que ocurra un fallo en la unidad es aumentada en proporción directa al número de unidades presentes.

Como los datos grabados pueden ser de importancia crítica, varios esquemas de grupos de disco redundantes o tolerantes a los fallos desarrollados para la industria de IT general (por ejemplo RAID1, RAID5, etc.) son explotados para requisitos de almacenamiento mayores con el fin de mejorar la fiabilidad completa del sistema total.

25 Aunque el diseño actual de sistemas de HDD puede ser muy sofisticado, el concepto básico es muy simple. Uno o más platos giratorios apilados de superficies de grabación (material magnético) son accedidos por un número igual de cabezas de lectura/escritura apiladas en un brazo móvil que puede girar a través del espacio o intervalo completo del área de grabación (muy parecido al brazo de reproducción de un tocadiscos pasado de moda). La posición del brazo móvil con relación a un plato giratorio asociado es controlada por un accionador de brazo (imán y bobina).

30 Debido a las densidades de datos y velocidades implicadas, la exactitud de posicionamiento requerida para las cabezas de grabación es extrema. Por ejemplo, con el fin de que un brazo móvil alcance rápidamente el punto de lectura/escritura, el accionador de brazo está diseñado para acelerar y decelerar rápidamente el brazo. Aunque el brazo y el conjunto de cabeza de lectura/escritura están diseñados para tener una masa tan baja como sea posible, las rápidas aceleraciones implicadas requieren una energía significativa, y esta energía es disipada generalmente en forma de calor dentro del conjunto accionador. Es por esta razón por la que una HDD que está continuamente leyendo/escribiendo (con el brazo saltando constantemente hacia atrás y hacia delante del directorio o directorios a las áreas de datos) funcionará mucho más caliente que una HDD que está funcionando sin ninguna actividad de lectura/escritura.

40 Por las razones antes esbozadas la electrónica de interfaz de sistemas de HDD conocida en la técnica está diseñada típicamente para ser muy robusta y por tanto en la práctica rara vez fallan. Sin embargo, las HDD a menudo fallan cuando la superficie física de los discos resulta dañada o contaminada. Esto puede dar como resultado una cabeza que toca la superficie del plato debido a un choque o vibración, o cuando la exactitud de posicionamiento del brazo de lectura/escritura se degrada debido a un choque, vibración o desgaste general. Es sabido que temperaturas elevadas también acortan la vida de una HDD y que el movimiento continuo del brazo también actúa para desgastar el accionador. Por lo tanto, la exactitud de posicionamiento del brazo disminuye eventualmente en una magnitud en la que la HDD resulta más propensa a los errores. Esto se ha encontrado lo más frecuente durante el ciclo de escritura.

45 La vibración, el sobrecalentamiento y el desgaste y desgarro son las causas principales de fallos eventuales en la HDD. Así, cuanto más elevados son los niveles de vibración (en funcionamiento) y más elevadas las temperaturas de funcionamiento, más corto es el tiempo medio entre fallos [MTBF] para cualquier HDD dada. Estos factores son reconocidos por la mayoría de los fabricantes de HDD quienes proporcionan por tanto gráficos que muestran el MTBF esperado cuando varía con la temperatura de funcionamiento. Típicamente esto muestra un aumento exponencial en las tasas de fallo de unidad con un aumento de la temperatura de funcionamiento, especialmente para HDD de coste inferior.

55 También es apreciado por los expertos en la técnica que la mayoría de los fallos de HDD (es decir, pérdidas irre recuperables de datos) ocurren durante los procesos de escritura, ya que los datos de escritura implican cambiar

activamente el estado magnético de la superficie de grabación, tanto en la propia escritura de datos como en la actualización del directorio de archivo de la HDD. El directorio de archivo es dónde es conservado un índice a las posiciones de todas las partes de todos los archivos en el disco. Como este directorio es el que es escrito más a menudo, este es el área de la HDD más susceptible de fallar. Desgraciadamente un error serio en el directorio de archivos puede causar una pérdida catastrófica de datos ya que este es el índice a las posiciones físicas de los archivos de datos en toda la HDD. Típicamente, todos los archivos en el área de datos de la HDD están fragmentados, y por tanto la recuperación estos archivos puede ser imposible, o muy difícil, al menos sin el uso de servicios especializados de recuperación de datos de HDD. Tales servicios son tanto caros como muy inconvenientes, especialmente cuando los datos pueden ser legalmente, comercialmente, oficialmente o sensibles de otro modo, o cuando el acceso a los datos es requerido tan rápidamente como sea posible.

Por comparación, es apreciado por los expertos en la técnica que simplemente leer datos desde una HDD es un procedimiento relativamente pasivo ya que la superficie de grabación es dejada en un estado inalterado (aunque un choque de cabeza, es decir, la cabeza que toca el plato debido al choque externo o vibración, puede aun causar una pérdida de datos).

Las soluciones convencionales para proporcionar una elevada integridad de datos con sistemas de HDD implica el uso de agrupaciones de HDD (tales como una Agrupación Redundante de Discos Baratos [RAID]) con alguna forma de redundancia, por ejemplo, obtener una copia especular, esquemas de paridad, etc. Por ejemplo, los sistemas de almacenamiento DVR a nivel de empresa emplean típicamente sistemas RAID 5 en los que una agrupación de n HDD proporciona una agrupación efectiva de $n-1$ HDD, ya que los datos de paridad de cada HDD se extienden a través de todas las demás unidades en un diseño definido. En el caso de un fallo, el controlador RAID 5 regenera los datos en la HDD que ha fallado sobre una HDD nueva a partir de los datos de paridad en todas las demás HDD supervivientes.

El problema con este esquema particular es que la reconstrucción de los datos perdidos en la nueva HDD es procesador y HDD intensiva, así como ser algo complejo. En aplicaciones de grabación en tiempo real, es además complicado por la necesidad de soportar la escritura de nuevos datos (a través de TODAS las HDD, incluyendo la nueva) al tiempo que se reconstruyen simultáneamente los datos perdidos.

Para HDD grandes, (es decir, aquellas del orden de unos cientos de GB) el proceso de reconstrucción puede requerir muchas horas. Durante todo este período la agrupación de unidades está bajo una sollicitación de lectura/escritura incrementada considerablemente, especialmente si las corrientes de datos entrantes (para grabación) son de un ancho de banda elevado. El problema fundamental aquí es que si sucede que falla otra HDD durante este proceso (o en realidad antes de que la unidad que ha fallado sea reemplazada) entonces todos los datos a través de la agrupación entera se perderán. Este caso es denominado como una pérdida de datos catastrófica. Como resultado, los sistemas RAID 5 son particularmente inadecuados para aplicaciones intensivas de datos (tal como se ha encontrado en sistemas DVR) debido a que las corrientes de datos de ancho de banda elevado constantes generadas por tales aplicaciones (por ejemplo, video, audio, datos de instrumentación a tiempo real, etc.). Se ha observado por los autores que a pesar de este hecho, los sistemas RAID 5 están a menudo especificados para tales aplicaciones, más probablemente debido a una falta de cualquier otra solución tolerante con los fallos de coste eficiente.

Una opción alternativa conocida en la técnica es la opción de agrupación de unidades copiadas especularmente (RAID 1). Esta es más atractiva desde un punto de vista de integridad de datos, pero por definición, es el doble de cara. En los sistemas RAID 1 los mismos datos son escritos para dos HDD (o dos agrupaciones) simultáneamente. La reconstrucción de una HDD que ha fallado es relativamente sencilla ya que la HDD superviviente de un par coincidente es copiada simplemente a una nueva HDD que reemplaza a la HDD que ha fallado. Sin embargo, esto pone de nuevo a la HDD superviviente bajo mucha más tensión ya que tiene que leer todos los datos a través de la nueva HDD y al mismo tiempo continuar grabando nuevos datos.

Un problema general de todos los sistemas de agrupación de discos conocidos en la técnica es que por razones de coste, espacio y accesibilidad, las HDD son normalmente empaquetadas muy juntas y montadas frontalmente para una retirada fácil. Una agrupación de discos de estante de 19" típica puede contener hasta 15 unidades montadas verticalmente y apiladas de forma horizontal. Este esquema es muy común, y aunque hay muchas variaciones de montajes verticales y horizontales, el resultado general es el mismo. Todas las unidades están funcionando constantemente (y normalmente los datos son escritos a través de las unidades de forma irregular) y así la disipación de calor en una HDD servirá para aumentar la temperatura de la HDD inmediatamente circundante. Así, in la práctica, una agrupación de 15 unidades montadas estrechamente puede hacerse muy caliente en realidad. La refrigeración forzada de tal agrupación es entonces empleada. Tal refrigeración requiere muchos ventiladores (y una fuente de aire entrante enfriado y así a menudo acondicionamiento de aire caro) y es a menudo insuficiente debido al flujo de aire restringido alrededor de las HDD empaquetadas. Como se ha hecho notar previamente, hay normalmente un aumento exponencial en las tasas de fallo cuando la temperatura de funcionamiento de las HDD aumenta.

Otro problema inherente a las agrupaciones de HDD conocidas es que la vibración de una HDD (debido al motor de accionamiento y al rápido movimiento de los brazos de lectura/escritura) será transmitida a las otras unidades,

5 aumentando así el desgaste y la rotura, reduciendo así el MTBF de cada HDD individual. Este efecto es compuesto cuando todas las unidades están escribiendo al mismo tiempo, como es típico con la mayoría de los sistemas de archivado intensivo de datos y es esencial con la mayoría de los esquemas RAID, incluyendo RAID 5.

5 Además de los problemas de temperatura y vibración, las agrupaciones de HDD exhiben un consumo de potencia elevado. Como un ejemplo, una agrupación de 15 discos, que disipa 12W por disco, con una PSU de 400W funcionando a una eficiencia del 75% (y que disipa por ello 75W por sí misma, ventiladores de refrigeración que gastan 25W, y electrónica de control (usualmente una placa madre de un PC y un procesador muy rápido para control de RAID complejo) que gasta 100W, totaliza 380W. Esto representa un gasto de potencia significativo con un coste de funcionamiento elevado, especialmente cuando tales agrupaciones son usadas 24 horas al día, 365 días al año y los costes del acondicionamiento de aire requerido para enfriar salas de equipo son tenidos en cuenta.

10 Otro problema aún de los sistemas de la técnica anterior son los efectos de los cortes del suministro eléctrico y variaciones de potencia. Las variaciones de la red de suministro o principal pueden afectar a la fiabilidad de la HDD. Las puntas de la red principal particularmente pueden causar sobretensiones en el lado de salida de corriente continua de la unidad de alimentación de corriente [PSU] del controlador y pueden, o bien dañar la electrónica de la HDD o bien provocar errores de datos durante los procesos de escritura. Además, reducciones de tensión (tensiones bajas), aumentos (sobretensiones) y pérdidas momentáneas (caídas) de potencia pueden tener todos efectos impredecibles en la electrónica del sistema. La experiencia con la instalación de miles de sistemas de almacenamiento en discos sobre muchos tipos diferentes de lugares industriales y comerciales muestra que los problemas de la red principal pueden no tener un efecto inmediato sobre la fiabilidad de la HDD, pero con el tiempo los efectos de puntas de tensión, de reducciones y de aumentos reducirá la vida de la HDD y eventualmente producirá fallos en la HDD. Las HDD son más sensibles a problemas de alimentación de corriente cuando están escribiendo datos ya que cualesquiera variaciones pueden corromper los datos, o provocar errores de control, pero pueden no afectar permanentemente en el sistema de HDD. Así, la exposición de las HDD a problemas de alimentación de corriente es exacerbada en aplicaciones tales como grabaciones de video o audio (o de hecho grabaciones de datos continuas de cualquier tipo) ya que implican la escritura continua de datos, y así las HDD están siempre escribiendo.

15 Como resultado de los efectos de interrupciones del suministro y variaciones de potencia, la mayor parte del equipamiento de IT (incluyendo las agrupaciones de discos a nivel de empresa) tiende a emplear unidades de suministro de corriente ininterrumpido [UPS], que añaden grandes costes al sistema total. En la práctica se ha encontrado que deben emplearse unidades UPS de línea interactiva verdaderas ya que estas condicionan y filtran la alimentación de corriente mientras que unidades más baratas no lo hacen.

20 Los fabricantes de HDD son conscientes de muchos de los problemas esbozados con anterioridad encontrados en distintas aplicaciones. Como resultado ofrecen una gama de productos de HDD, con diferentes características operativas y de rendimiento. Naturalmente, cuanto más robusta es la HDD (por ejemplo, diseños de ciclo de trabajo más elevados, tolerancia a la vibración o temperatura de funcionamiento elevada), más elevado es el precio. En el extremo superior de la gama, los fabricantes de HDD producen las así llamadas HDD de prestaciones de empresa, que son razonablemente robustas contra modos de fallo comunes. Sin embargo, tales unidades son muy caras y son por lo tanto usadas solamente para el denominado almacenamiento en línea crítico (dónde las velocidades de escritura de datos y el rendimiento elevado son esenciales). Estas unidades también tienden a consumir más energía y un funcionamiento más caliente. Incluso así, estas unidades de rendimiento elevado aun exhiben un MTBF disminuido con temperaturas de funcionamiento y vibraciones aumentadas. Además, para aplicaciones de almacenamiento en masa (tales como grabaciones de video/audio o con propósitos de archivo de IT y de copia de seguridad), estos discos son tan caros que una solución que usa muchas HDD de empresa es a menudo totalmente impráctica.

25 En resumen, las agrupaciones de HDD de la técnica anterior sufren de un número de problemas que incluyen temperatura, vibración, refrigeración, consumo de potencia, exposición constante a variaciones de suministro de corriente y coste. Otros medios distintos de usar refrigeración forzada o discos más caros, las soluciones de la técnica anterior a estos problemas emplean generalmente técnicas diseñadas únicamente para luchar con el fallo de la HDD, en vez de acceder a las razones fundamentales de por qué fallan las HDD en primer lugar. Por lo tanto emplean una aproximación técnica que lucha con los síntomas pero no logra proporcionar una cura.

30 Es por tanto un objeto del presente invento proporcionar un sistema de HDD más barato, más fiable y más conveniente para el almacenamiento de datos que emplea una o más técnicas de grabación de datos secuencial. Tal grabación secuencial puede ser a través del empleo secuencial de agrupaciones de HDD de bajo coste y/o el empleo secuencial de sectores, o grupos de sectores, situados dentro de las propias HDD individuales.

35 RESUMEN DEL INVENTO

De acuerdo con un primer aspecto del presente invento se ha proporcionado un método de almacenamiento de datos en un sistema de una unidad de disco duro como se ha definido por la reivindicación 1.

De acuerdo con un segundo aspecto del presente invento se ha proporcionado un método de escritura de datos en

una unidad de disco duro como se ha definido por la reivindicación 9.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Los aspectos y ventajas del presente invento resultarán evidentes con la lectura de la descripción detallada siguiente y con referencia a los dibujos siguientes en los que:

5 La fig. 1 presenta (a) una vista frontal; y (b) una representación esquemática, de un sistema de almacenamiento de datos HDD de acuerdo con un aspecto del presente invento;

La fig. 2 presenta (a) una vista superior de una única Unidad de Disco Duro de plato; y (b) una representación esquemática de una cuarta Unidad de Disco Duro de plato, siendo ambas adecuadas para su uso dentro del sistema de almacenamiento de datos de HDD de la fig. 1;

10 La fig. 3 presenta una vista frontal de un sistema de almacenamiento de datos HDD de agrupación bidimensional de acuerdo con una realización del presente invento;

La fig. 4 presenta una representación esquemática de cinco etapas de un sistema de almacenamiento de datos de HDD de rendimiento de escritura copiada especularmente de acuerdo con una realización del presente invento. En particular las etapas mostradas corresponden a los datos que son escritos a:

- 15
- (a) un primer par de HDD;
 - (b) un segundo par de superposición de HDD;
 - (c) un tercer par de superposición de HDD;
 - (d) un par de superposición final de HDD; y
 - (e) un par de HDD que comprende la última HDD y la primera HDD en la agrupación de discos de datos; y

20 La fig. 5 presenta una representación esquemática de un plato de disco de datos que emplea un sistema de archivado secuencial (SFS) de acuerdo con un aspecto del presente invento.

DESCRIPCIÓN DETALLADA

25 Con el fin de ayudar a la comprensión de diferentes aspectos del presente invento, las figs. 1(a) y (b) presentan una vista frontal y una representación esquemática, respectivamente, de un sistema 1 de almacenamiento de datos de HDD.

30 El sistema 1 de almacenamiento de datos de HDD comprende una agrupación de quince Unidades de Disco Duro (HDD) 2 cuyo funcionamiento está controlado por un sistema 3 controlador de agrupación de HDD. El sistema controlador 3 comprende una interfaz 4 de datos (tal como Ethernet, SCSI, USB, Canal de Fibra u otra interfaz rápida), una placa controladora 5 que incluye un controlador 6 de CPU y un controlador 7 de HDD, y una alimentación de corriente 8. El sistema controlador 3 comprende además tres indicadores 9 de estado de LED asociados con cada HDD 2. Durante el funcionamiento los indicadores 9 de estado que corresponden a cada HDD 2 en la agrupación, son accedidos y controlados por el controlador 6 de CPU. Los indicadores 9 proporcionan información en cuanto al estado de la HDD asociada 2 por ejemplo en cuanto a cuando la HDD 2 es, por ejemplo, Seleccionada, Activa, está Leyendo, Escribiendo o ha Fallido.

35 Otro detalle de las HDD 2 es presentado en la fig. 2. En el ejemplo mostrado en la fig. 2(a), la HDD 2 comprende un conjunto accionador 10 empleado para controlar la posición relativa entre una cabeza 11 de lectura/escritura, situada en el extremo distal de un brazo 12 de lectura/escritura, y una superficie de datos de un plato asociado 13. Como es práctica normal en la técnica el plato 13 dentro de la HDD 2 está montado sobre un eje de giro 14.

40 Como es apreciado por los expertos en la técnica, cada plato 13 de disco duro tiene (potencialmente) dos superficies de grabación. Las HDD alternativas 2 que comprenden múltiples platos 13 de discos y combinaciones de la cabeza 11 de lectura/escritura también son conocidas en la técnica. A modo de ejemplo solamente, la fig. 2(b) presenta una HDD 2 que comprende cuatro platos 13 montados en un eje de giro común 14 con ocho cabezas 11 de lectura/escritura asociadas la totalidad de las cuales están controladas por el conjunto accionador 10 a través de los brazos 12 de lectura/escritura asociados.

45 La superficie de datos de grabación de un plato 13 está dividida en una serie de cilindros concéntricos 15 espaciados de forma precisa. Cada uno de estos cilindros 15 actúa como una pista de grabación y cada pista está además dividida en sectores 16 de igual tamaño (así hay más sectores en las pistas exteriores que en las pistas interiores). En casi todos los platos 13 de discos duros modernos, cada sector puede grabar 512 bytes de información. Los autores hacen notar, sin embargo, que hay una propuesta actual dentro de la técnica para
50 aumentar la capacidad estándar de sectores de disco a 4.096 bytes por sector. Sin embargo, como la electrónica dentro de la HDD 2 tiene cuidado de acceder a todos los sectores dentro de una HDD 2 a través del firmware

(soporte lógico inalterable) específico del fabricante, el sistema de la fig. 2(b) puede simplemente ser considerado como una extensión de la HDD 2 de plato único de la fig. 2(a). Cualquiera que sea el sistema de archivado de disco (DFS) empleado entonces proporciona los medios requeridos para mantener la pista de todos los archivos grabados y su posición en el plato 13.

- 5 Dentro del presente sistema, puede usarse cualquier combinación de HDD 2 y capacidades de datos. Estos son sistemas RAID distintos en los que las HDD 2 han de ser coincidentes estrechamente si no idénticas en términos de rendimiento y capacidad.

10 Se apreciará que dentro del presente sistema si han de ser usadas HDD de capacidad variable con uno de los modos de escritura copiada especularmente, entonces la disposición óptima de la HDD tiene el fin de aumentar la capacidad, ya que esta proporcionará la máxima capacidad de almacenamiento efectiva. Además, también se apreciará que cualquier tipo de plato estándar 13, y métodos de lectura/escritura conocidos, pueden ser empleados dentro de cada HDD 2.

15 Un primer modo de funcionamiento del sistema de almacenamiento de datos 1 será descrito ahora con referencia a la fig. 1(b). En primer lugar los datos son grabados en un elemento de almacenamiento de datos que comprende la primera HDD 2, hasta que el HDD 2 está llena, por lo que los siguientes datos son grabados en el siguiente elemento de almacenamiento de datos, en particular la segunda HDD 2. Este proceso continúa a través de la tercera y cuarta HDD 2 hasta que la agrupación de HDD está llena. En este punto los siguientes datos son grabados nuevamente en la primera HDD 2, de modo que sobrescriba los datos más antiguos en la secuencia. Este proceso puede efectivamente continuar sobre una base de bucle sin fin. En la realización descrita actualmente, la duración del tiempo de los datos grabados disponibles es ajustada por la tasa media de datos generados (bits/seg) y la capacidad total de la agrupación de HDD.

20 Es preferible que el controlador 6 de la CPU proporcione un ajuste del Límite de Retención de Archivo [FRL], que permite un límite definido de usuario en el período de tiempo de almacenamiento de datos (por ejemplo, 30 días). Este límite de tiempo es siempre menor que la capacidad potencial total de la agrupación de HDD, a una tasa de datos media dada. En esta realización, los datos más antiguos comienzan a ser sobrescritos tan pronto como el FRL es alcanzado, dejando ningún espacio en blanco de capacidad de HDD restante. Esta puede ser una característica importante para seguir las leyes, regulaciones locales y las guías de buena práctica para ciertas aplicaciones en algunos países.

30 Hay muchas ventajas para la grabación de datos basados en el tiempo en esta forma estrictamente secuencial. En primer lugar el controlador 6 de CPU actúa para desconectar totalmente (es decir, retirar la corriente) de todas las HDD 2 que no están siendo escritas, o leídos. Así en circunstancias normales sólo una HDD2 es conectada en cualquier instante. El controlador 6 de CPU vigila la tasa media de uso de la capacidad de la HDD 2 cuando los datos son grabados y conecta la siguiente HDD 2 en la secuencia en tiempo suficiente de tal modo que la HDD 2 es hecha girar y está lista para aceptar datos antes de que la HDD 2 previa sea llenada. El período de conexión avanzado también permite que el controlador 6 de CPU compruebe la siguiente HDD 2 en la secuencia, encuentre el primer sector disponible en la HDD y lea la capacidad de HDD 2. Empleando unidades modernas y un buen sistema de archivado de disco, el proceso de conexión avanzado sólo requiere unos pocos segundos, incluyendo la comprobación de HDD.

40 En uso normal, el proceso de grabación significa que solo una HDD 2 está activa al mismo tiempo. Sin embargo, una solicitud de lectura puede ser recibida por el sistema 1 de almacenamiento de datos en cualquier momento, para un segmento de datos desde cualquier lugar dentro del período de tiempo actualmente almacenado en la agrupación. Para acomodar tales solicitudes el controlador 6 de CPU mantiene un índice de ejecución de los tiempos inicial y final de los datos almacenados dentro de cada HDD 2 dentro de la agrupación. El índice de ejecución puede estar almacenado en una memoria no volátil en forma de una memoria flash o incluso otra HDD 2 que es accedida directamente por el controlador 6 de CPU. Alternativamente, el índice de ejecución puede ser almacenado en una RAM del controlador 6 de CPU.

Aunque el almacenamiento del índice de ejecución en una HDD 2 separada es una opción de trabajo no es particularmente deseable ya que esta HDD 2 necesitaría estar conectada en todo momento.

50 El concepto del índice de ejecución puede extenderse además de tal modo que pueda replicar la función completa de un sistema de gestión de base de datos (DBMS), empleado para mantener una pista de todos los archivos de datos en el sistema 1 de almacenamiento de datos. Como será apreciado por los expertos en la técnica tal índice es mejor almacenado en una RAM por velocidad o para los procesos de lectura, modificación y escritura. Sin embargo, una pérdida de corriente significaría una pérdida completa del índice de la base de datos. Mientras el simple índice de ejecución puede ser reconstruido de forma relativamente rápida a la recuperación de corriente, la reconstrucción de un índice DMNS completo consumiría significativamente más tiempo sin las copias de seguridad periódicas.

55 Para acomodar estas distintas opciones es por ello preferible que el índice de ejecución sea almacenado inicialmente en RAM y a continuación copiado a una memoria flash volátil interna a intervalos periódicos.

Independientemente de la forma de almacenamiento exacta del índice de ejecución el proceso de recuperación de

datos permanece igual. Cuando una solicitud de lectura es recibida, el controlador 6 de CPU simplemente conecta la HDD 2 correspondiente y lee los datos relevantes. Los sistemas de HDD 2 modernos giran muy rápidamente y así tienen un retraso de sólo unos pocos segundos después de la conexión antes de responder con los datos requeridos. Normalmente la HDD 2 que contiene los datos desde la que se están leyendo los datos se desconectará automáticamente después de un período de interrupción definible por el usuario (por ejemplo 10 minutos).

Con los sistemas 1 de almacenamiento de datos antes descritos los datos nunca son borrados excepto por sobrescritura con nuevos datos. Esto es una parte esencial del esquema, especialmente, para el almacenamiento de datos de evidencia potenciales, de tal modo que una vez que los datos están grabados, no pueden ser modificados en ningún instante, ni pueden ser borrados hasta que expiren y sean sobrescritos. Así el sistema 1 de almacenamiento de datos descrito proporciona el rendimiento total de escritura de datos en línea en el ancho de banda máximo de los subsistemas de HDD, mientras proporciona simultáneamente el rendimiento de acceso de datos cerca de línea cuando se requiere la repetición de los datos registrados basados en el tiempo.

El sistema 1 de almacenamiento de datos también proporciona un número de ventajas adicionales sobre el sistema descrito en la técnica anterior. En primer lugar el sistema 1 de almacenamiento de datos usa mucha menos energía que una agrupación de HDD convencional. Por ejemplo, en una agrupación de 15 unidades, poco más de 1/15 parte de la potencia será usada sobre el promedio comparado con una agrupación convencional. De modo similar, en una agrupación de 64 discos, sólo se usaría 1/64 parte de la potencia. Dicho de otro modo, el hecho de que sólo una HDD 2 esté funcionando en un instante significa que cada HDD 2 estará conectada y en uso durante sólo una fracción de tiempo comparado con el funcionamiento normal, prolongando así dramáticamente la vida de la HDD 2. Por ejemplo, en la agrupación de 15 unidades descrita, cada unidad estaría desconectada durante 14/15 partes (93%) del tiempo. Otro modo de considerar esto es que cada HDD 2 solo estaría en uso durante 26 días en un año, mientras que en los sistemas de almacenamiento de datos convencionales cada unidad funcionaría durante 365 días al año.

Como sólo una HDD 2 está generalmente en funcionamiento en cualquier instante hay una disipación de calor mínima desde otras HDD 2 y la refrigeración forzada es hecha redundante, o al menos es reducida considerablemente. Las HDD 2 también se han encontrado que funcionan a una temperatura muy inferior, aumentando así el MTBF y la fiabilidad.

Otra ventaja del hecho de que sólo una HDD 2 esté funcionando a la vez, es que hay una vibración mínima transmitida desde la otra HDD 2 y así el desgaste y rotura inherentes de la unidad son reducidos considerablemente, dando como resultado de nuevo un MTBF y una fiabilidad total aumentados.

Se hará resaltar que las grabaciones en cada HDD 2 cubren un período de tiempo contiguo completo (por ejemplo, 10:03:27 en 12 de Enero de 2006 hasta 12:32:13 en 14 de Enero de 2006). El controlador 6 de CPU puede por ello ser empleado para presentar este período de tiempo ya que cada HDD 2 es seleccionado a su vez por el controlador 7 de HDD. Alternativamente, el período de tiempo puede ser presentado en una pantalla de ordenador conectado a una red que ejecuta un software de vigilancia. Así, la HDD 2 física real en la que han de ser encontrados los datos correspondientes a cualquier período de tiempo particular, puede ser identificado modo extremadamente rápido y conveniente. Esta es una característica extremadamente útil para sistemas de datos probatorios (tales como grabaciones de vigilancia), particularmente cuando la HDD 2 correspondiente al instante de un evento particular o incidente puede ser identificado rápidamente y retirado físicamente para conservarlo de modo seguro (por ejemplo por la policía u otras autoridades). Tal situación fácil de datos físicamente en una agrupación no es posible con métodos tradicionales, especialmente configuraciones de RAID. Obsérvese que en un sistema RAID 5, la agrupación entera, incluyendo el bastidor y el controlador debe ser tomada para ser de algún uso práctico ya que cada HDD 2 contiene sólo información inútil parcial y efectivamente sin que estén presentes todas las demás HDD 2.

Una consecuencia significativa y útil de los sistemas 1 de almacenamiento de datos descritos es que el período de tiempo almacenado puede ser hecho efectivamente infinito retirando la HDD 2 grabada en cualquier instante antes de que sea sobrescrita, y reemplazándola con nuevas, HDD 2 en blanco. Así pueden crearse archivos de datos muy largos retirando y reemplazando las HDD 2 en una base continua. Además, como el período de tiempo básico de la agrupación es probablemente del orden de semanas (es decir, 30 días), entonces la HDD 2 puede ser reemplazada a conveniencia, y a intervalos irregulares de cualquier tiempo menor de 30 días en el ejemplo.

Dentro de los sistemas 1 de almacenamiento de datos descritos actualmente, el fallo de la HDD 2 es mucho menos probable, debido al aumento en las vidas del MTBF y de la HDD 2. Si una HDD 2 fallara, sin embargo, los datos son automáticamente redirigidos a la siguiente HDD 2 en la secuencia de HDD 2 y la HDD 2 que ha fallado resaltada con el indicador 9 de estado correspondiente. La notificación del fallo de HDD 2 puede ser por medio de sonido audible, presentación del controlador de CPU o presentación en cualquier aplicación de gestión remota que se ejecuta en un ordenador conectado a una red. La pérdida de datos está por lo tanto limitada a una HDD 2, aunque se ha reconocido que cualquier pérdida de datos, sin embargo menos probable en el sistema actual, es aún una desventaja significativa.

Una consecuencia ventajosa de los sistemas 1 de almacenamiento de datos es que cualquier HDD 2 que no está activo o ha fallado puede ser retirada y reemplazada en los sistemas 1 de almacenamiento de datos en cualquier

instante. Además, cualquier HDD 2 que esté activa (es decir, que está escribiendo), puede ser retirada por un salto forzado a la siguiente HDD 2 en la secuencia a través de la placa controladora 5 o la interfaz de software, o simplemente retirando la unidad y provocando un “fallo” efectivo.

5 Varias realizaciones alternativas a los sistemas 1 de almacenamiento de datos antes descritos serán descritas en detalle a continuación.

Sistema de Almacenamiento de Datos de Agrupación de Dos y Tres Dimensiones

La fig. 3 presenta una vista frontal de un sistema 17 de almacenamiento de datos de agrupación de dos dimensiones. Como puede verse el sistema 17 de almacenamiento de datos de agrupación de dos dimensiones es una extensión del sistema 1 de almacenamiento de datos descrito previamente en el que ahora se emplea una agrupación de dos dimensiones de HDD 2 dispuestas en secuencia. El sistema 17 de almacenamiento de datos de agrupación de dos dimensiones proporciona una disposición conveniente que permite el uso de múltiples unidades apiladas de agrupaciones esclavas 1a, 1b y 1c conectadas a una placa controladora maestra 5. Con tales esquemas, literalmente cientos de HDD 2 pueden ser accedidas, estando sólo siempre activa una HDD 2 en cualquier instante. Para tales instalaciones de gran capacidad, los ahorros de potencia, los costes de refrigeración y de acondicionamiento de aire resultantes pueden ser considerables.

Será fácilmente evidente que los principios de los dos sistemas 17 de almacenamiento de datos de agrupación de dos dimensiones pueden ser extendidos simplemente de modo que formen un sistema de almacenamiento de datos de agrupación tridimensional que funciona de una manera sustancialmente idéntica. Sin embargo, debería observarse que no hay requisitos para que la siguiente HDD 2 en la secuencia sea situada físicamente a continuación de la HDD 2 previa. El requisito esencial de todos los sistemas es que existe una relación secuencial para el orden en el que se accede a las HDD 2 individuales que es mantenido durante el funcionamiento del sistema.

Una funcionalidad incrementada para los sistemas de almacenamiento de datos de agrupación descritos tiene lugar a través del empleo de dos o más placas controladoras maestras 5. Cada placa controladora maestra 5 está conectada a la fuente de datos originaria (tal como Ethernet o SCSI) en tal disposición que las placas maestras secundarias 5 actúan efectivamente como esclavas, excepto porque también retienen una copia del índice de período de tiempo del disco. Así, si fallara una placa controladora maestra 5, una placa maestra secundaria 5 podría captar inmediatamente las tareas del controlador y el sistema de agrupación podría ponerse a trabajar. Debido a la arquitectura de tal disposición, las HDD 2 asociadas con la placa maestra 5 que ha fallado pueden ser retiradas y puestos en agrupaciones de reemplazamiento (o incluso en agrupaciones de HDD de una unidad esclava) que son a continuación incorporadas de nuevo al sistema en lugar de las asociadas con la unidad defectuosa. Todo esto puede ser realizado sobre una base de intercambio caliente.

Sistema de Almacenamiento de Datos de Escritura Copiada Especularmente

Como se ha descrito en detalle antes, casi todos los fallos de HDD 2 ocurren durante el proceso de escritura y cualquier pérdida de datos sin embargo pequeña, es aún una desventaja significativa para cualquier sistema de almacenamiento de datos. Con la ayuda de la fig. 4, un sistema de almacenamiento de datos que proporciona prestaciones de escritura y protección copiadas especularmente sin tener que comprometer dos veces la capacidad de almacenamiento del sistema (como se requiere por ejemplo con una sistema RAID 1), se ha descrito a continuación.

El sistema 18 de almacenamiento de datos de escritura copiada especularmente es puesto en práctica como sigue. En vez de escribir a un elemento de almacenamiento de datos que comprende una única HDD 2, los mismos datos son escritos simultáneamente a un elemento de almacenamiento de datos que comprende un par de HDD 2. En el ejemplo descrito actualmente, los datos son inicialmente escritos a la primera y a la segunda HDD 2 (denominadas como “Escritura 1 y 2”), véase la fig. 4(a).

Cuando estas HDD resultan llenas, la tercera HDD 2 es conectada de forma que esté lista para su uso. La escritura de datos cambia entonces desde la primera y la segundo HDD 2 (“Escritura 1 y 2”), al siguiente elemento de almacenamiento de datos, en particular la segundo y la tercera HDD 2 (“Escritura 2 y 3”), véase la fig. 4(b). La primera HDD2 es a continuación desconectada. Como resultado, los datos que acaban de ser escritos a la segundo HDD 2 son a continuación sobrescritos inmediatamente con nuevos datos. Como una copia de estos datos es almacenada de forma segura en la primera HDD 2, que ha sido desconectada, es muy improbable que falle.

La escritura de datos continúa en este diseño, escribiendo a elementos de almacenamiento de datos subsiguiente que comprenden pares de HDD 2 por ejemplo “Escritura 3 y 4” véase la fig. 4(c), “Escritura 4 y 5”, etc. Como el emparejamiento de escritura avanza, la HDD 2 completa posterior es desconectada. Esta escritura a pares solapados de HDD 2 continúa hasta que se ha alcanzado la HDD, véase la fig. 4(d), en cuyo punto los pares de secuencia de escritura serán “Escritura (n-1) y n”, a continuación “Escritura n y 1”, véase la fig. 4(e) (es decir, la última y primera HDD 2) y a continuación vuelve a través de la secuencia de “Escritura 1 y 2”, “Escritura 2 y 3”, etc.

Este sistema 18 de almacenamiento de datos de escritura copiada especularmente proporciona redundancia copiada especularmente durante la parte de escritura crítica, y es tan eficiente como los sistemas RAID 5 en

proporcionar una capacidad total $n-1$. Este esquema también mantiene todas las ventajas de los sistemas 1 de almacenamiento de datos descritos previamente, con sólo reducciones menores experimentadas en las ventajas de uso de potencia, aislamiento de temperatura y de vibración.

5 Empleado el sistema 18 de almacenamiento de datos de escritura copiada especularmente, se ha proporcionado un sistema de almacenamiento de datos que puede enfrentarse con una HDD 2 que ha fallado sin ninguna pérdida de datos. Considérese por ejemplo un fallo la cuarta HDD 2 cuando el par que se está escribiendo es "Escritura 3 y 4". Inmediatamente cuando un fallo de la HDD 2 es detectado en la cuarta HDD 2, la escritura continúa normal dentro de la tercera HDD 2, mientras la quinta y la sexta HDD 2 ("Escritura 5 y 6") están conectadas. Después de unos pocos segundos estas unidades están listas para aceptar datos, y toda la escritura es cambiada a ellas. La
10 tercera y la cuarta HDD 2 son a continuación desconectadas, preservando los datos dentro de la tercera HDD 2 (la HDD buena) y permitiendo que la cuarta HDD 2 (la HDD mala) sea reemplazada a conveniencia. El indicador 9 de estado de "Fallo" asociado con la cuarta HDD 2 se iluminaría a continuación para indicar que la HDD 2 había fallado y así que se requería reemplazarla.

15 Debe observarse que no hay urgencia en reemplazar la HDD 2 mala, como sería el caso dentro de un sistema RAID en el que un segundo fallo de HDD 2 significaría una pérdida catastrófica de todos los datos hasta el instante en el que una nueva HDD 2 ha sido prevista y se ha reconstruido completamente la HDD 2 que ha fallado. Dentro del sistema 18 de almacenamiento de datos de escritura copiada especularmente, cuando la HDD 2 mala es reemplazada, no hay necesidad de hacer nada con ella hasta que el proceso de escritura forma un bucle alrededor y se requiere que los datos sean grabados dentro de la HDD 2. Esto sólo ocurrirá después de que el período de
20 almacenamiento de la agrupación de datos completa haya expirado. Otro punto significativo a destacar es que la HDD 2 superviviente de un par que ha fallado sólo es expuesto a unos pocos segundos de escritura no copiada especularmente, haciendo así insignificantes las posibilidades de una pérdida total de datos.

25 El sistema de almacenamiento de datos de escritura copiada especularmente descrito puede extenderse además de tal modo que el primer elemento de almacenamiento de datos comprende m HDD 2 dispuestas en secuencia. El siguiente elemento de almacenamiento de datos superpuesto comprendería entonces también m HDD 2 dispuestas en secuencia, con un desplazamiento menor de m introducido entre la numeración secuencial de las HDD 2 del primer y segundo elementos de almacenamiento de datos.

30 Por ejemplo, con m ajustado a tres y el desplazamiento secuencial entre elementos de almacenamiento de datos ajustado para que sea dos es producido un sistema de escritura de datos copiados especularmente triple que proporciona un almacenamiento de datos copiados especularmente doble. Escribiendo los mismos datos en las tres HDD 2, y a continuación saltando en secuencia dos HDD 2 antes de que el siguiente proceso de escritura tenga lugar, sólo una de las tres HDD 2 del elemento de almacenamiento de datos es siempre sobrescrita, dejando así dos copias idénticas detrás. Este método de proporcionar pares copiados especularmente almacenados encuentra
35 aplicación particular dentro de sistemas que son extremadamente críticos para los datos. Además, esta técnica tiene una ventaja sobre los sistemas de almacenamiento copiados especularmente normales (por ejemplo sistemas RAID 1) porque si una HDD 2 falla aún se tiene un par copiado especularmente almacenado completo de HDD 2. Un sistema RAID 1 le dejaría solo con una sola copia, que debe ser entonces duplicada de modo que se reconstruya el par de HDD 2, poniendo así la única y sola HDD 2 original superviviente bajo tensión adicional y riesgo de fallo cuando sus datos contenidos están siendo leídos.

40 Además, empleando la técnica de triple escritura, y doble almacenamiento permite la retirada de una HDD2 de cada elemento de almacenamiento de datos con propósitos de almacenamiento y transferencia permanentes.

Sistema de Archivado de Disco Duro

45 Como se ha descrito antes, los sistemas de almacenamiento de datos 1, 17 o 18 pueden emplear métodos de lectura/escritura estándar conocidos para los expertos en la técnica. La electrónica integrada y el firmware dentro de las HDD 2 convencionales permiten generalmente el acceso de todos los sectores en la HDD 2 por número de sector. Los Sistemas de Archivado de Discos (DFS) convencionales agrupan entonces típicamente un número de sectores en grupos o "cluster" y la HDD 2 es formateada de acuerdo con estos tamaños de grupo. Por ejemplo, un tamaño de grupo típico para un DFS moderno sería de 16 sectores que totalizan 4.096 bytes o 4 kB. Un archivo mayor de 4 kB usará tantos grupos como se requiera para el archivo, y un archivo menor de 4 kB aún usará hasta 4
50 kB ya que este es el menor bloque accesible en el sistema de archivado. Los expertos en la técnica serán conscientes de que el tamaño del sector de disco estándar, actualmente 512 bytes, es probable que aumente a 4.096 bytes en el futuro. Por consiguiente, los DFS tradicionales también adoptarán en el futuro diferentes tamaños de grupo.

55 Los DFS conocidos en la técnica emplean distintos esquemas de directorio para mantener la pista de todos los archivos grabados y sus posiciones. El directorio es conservado usualmente al comienzo de un plato de datos ya que esta posición proporciona tiempos de lectura más rápidos. Un DFS dividirá archivos mayores a través de muchos grupos, y con el fin de usar el espacio de la HDD de forma más eficiente, estos grupos pueden ser dispersados a través de los platos. Esto es denominado fragmentación, y la grabación, modificación y borrado de muchos miles de archivos provoca eventualmente una fragmentación severa y reduce el rendimiento de la HDD 2 ya

que la HDD 2 se llena con muchos archivos fragmentados. Se ha observado, sin embargo, que algunos esquemas de DFS modernos mitigan el efecto de fragmentación en cierta medida. Los métodos de DFS convencionales también hacen una estructura de directorio grande y compleja ya que todas las posiciones del grupo de todos los archivos y los fragmentos de archivo son grabados. Incluso una HDD 2 vacía, formateada con un DFS estándar, puede usar algo entre un 1,8% y un 4% de la capacidad de grabación, solo para la estructura de directorio, y esto puede aumentar además en un uso prolongado si muchos archivos pequeños son escritos a la HDD 2, dependiendo del tipo de DFS.

Obsérvese que con la mayor parte de los esquemas DFS, si un archivo es borrado, es simplemente borrado del directorio, no borrado realmente del área de datos de archivo. Esto hace tiempos de borrado de archivo muy rápidos.

Un sistema de archivado de disco duro alternativo, y preferido, será descrito en detalle a continuación, con referencia a la fig. 2(a) y a la fig. 5. La fig. 5 presenta una representación esquemática de un plato 13 de disco de datos que incorpora este sistema de archivado de disco duro. El sistema de archivado de disco duro explota nuevamente el principio de almacenamiento de datos secuencial de datos generados sobre una base de tiempo estrictamente secuencial. El sistema de archivado en el disco es denominado en lo sucesivo el Sistema de Archivado Secuencial (SFS).

A partir de la fig. 5 puede verse que el SFS divide los sectores en un número de elementos de almacenamiento de datos situado en secuencia denominado Bloques de Asignación AB 19. Cada AB 19 consiste de 129 sectores 16 y en particular un sector de encabezamiento 20 y 128 sectores 21 de datos. Así cada AB 19 comprende 129 x 512 bytes o 64,5 kB de tamaño. Puede verse que los AB 19 son mucho mayores que los grupos de disco típicos. Normalmente esto sería ineficiente para un sistema de almacenamiento en el que los tamaños de archivo que han de ser grabados varían dramáticamente y son almacenados muchos archivos muy pequeños. Sin embargo, tamaños grandes de AB son ideales para grabar datos secuenciales en que las aplicaciones de grabación generan típicamente grandes archivos.

El sector de Encabezamiento 20 del AB 19, es clave para el funcionamiento del sistema SFS ya que contiene al menos el tiempo de inicio y la fecha para los datos almacenados dentro del AB 19. En este contexto se definirá un "archivo" por la aplicación de generación de datos externos y abarcará generalmente múltiples AB 19. Opcionalmente el sector de Encabezamiento 20 también puede almacenar la información siguiente, en particular: número de canal, posición del AB anterior para este canal, posición del AB siguiente para este canal, marcadores de inicio de archivo y final de archivo, y opcionalmente también datos cifrados para este AB 19.

En funcionamiento, los SFS leen y escriben datos a los AB 19 de una forma estrictamente secuencial en una base de tiempo. Esta tasa de datos puede sin embargo variar dramáticamente. Puede también comenzar y parar impredeciblemente y han de ser grabados probablemente muchos canales diferentes de forma simultánea (por ejemplo, canales de video, pistas de audio, corrientes de datos, etc.). Así, el SFS es requerido para gestionar correctamente todos estos eventos y así incluye métodos para conseguir una gestión total de datos de múltiples canales irregulares y variables y así el control completo de los procesos de escritura/lectura del disco.

Para conseguir los requisitos anteriores el SFS emplea el controlador 6 de CPU para mantener una memoria tampón de 64 kB para cada canal desde el que se ha de grabar información. Tan pronto como cada memoria tampón está completa el controlador 6 de CPU selecciona a continuación el primer AB 19 disponible y escribe el bloque entero a este AB 19, y cualesquiera AB 19 posicionados en secuencia de modo adicional requeridos. La HDD 2 entera es llenada en secuencia de este modo, asegurando así la utilización completa de la capacidad de la HDD y la corriente secuencial completa de los datos. Una vez que la HDD 2 ha sido llenada hasta su capacidad (y cualesquiera otras HDD en la agrupación también han sido llenadas), el controlador 6 de CPU forma un bucle de nuevo de tal modo que cualquier dato subsiguiente sobrescribe los datos almacenados originalmente dentro del primer AB 19.

Debería observarse, que si los canales de datos separados están en uso, los canales individuales pueden no estar necesariamente en secuencia estricta, pero esto no tiene que ver con el rendimiento del sistema ya que el sector de Encabezamiento 20 contiene información acerca de los AB 19 de canal anterior y siguiente.

Otro punto a observar es que como el SFS emplea un sector de Encabezamiento 20 para cada 128 sectores de datos significa que solo es usado el 0,78% de la capacidad de la HDD 2 para información de encabezamiento de archivo, comparado con más del 5% para un sistema de directorio convencional.

El SFS puede funcionar de esta simple manera debido a que los datos nunca son borrados o modificados, sólo completamente sobrescritos con nuevos datos (al final del período de almacenamiento). El sistema de control SFS no contiene realmente el comando u orden de Borrado en absoluto. Así el SFS trata cada HDD 2 como si fuera un cinta muy grande (larga). Como resultado el SFS proporciona medios para formatear en marcha, permitiendo así que discos nuevos completamente sin formatear (o incluso formateados previamente con algún otro sistema de archivado) sean usados sin ningún procedimiento de formateado que consume tiempo.

Varias características muy útiles y ventajosas surgen como un resultado de esta aproximación, incluyendo:

- 1) Optimización del rendimiento de escritura (máxima velocidad de disco);

- 2) Reproducción rápida de secuencias continuas;
- 3) Movimiento mínimo del brazo de lectura/escritura, así:
 - o Menor desgaste del accionador del brazo;
 - o Menor temperatura de funcionamiento del disco (el accionador consume menos potencia);
 - 5 o Menor vibración (el brazo y el accionador son efectivos aun la mayor parte del tiempo);
- 4) Una estructura de datos más robusta debido a la ausencia de un directorio;
- 5) Recuperación de datos simple; y
- 6) Búsqueda fácil por fecha y hora, a pesar de que no hay directorio.

Optimización del rendimiento de escritura

- 10 Como el brazo de lectura/escritura 12 no tiene que moverse constantemente hacia atrás y hacia delante para actualizar un directorio, puede conseguirse el máximo rendimiento sostenible del sistema de disco. Esto es ideal para usar dentro de sistemas dentro de los cuales son generados y requeridos para ser grabados grandes cantidades de datos de una manera robusta y fiable.

Reproducción Rápida

- 15 Las aplicaciones de grabación de datos secuencial son la mayor parte probablemente para ser leídas (es decir reproducidas) de modo secuencial. Con el SFS, los datos están dispuestos dentro de la HDD 2 de la misma manera estrictamente secuencial en la que han sido generados (es decir, no están fragmentados y en un orden de bloque aleatorio como con un Sistema de Archivado de Disco convencional). La disposición de archivo SFS es idealmente adecuada para el requisito de recuperación de datos más común, siendo optimizada la velocidad de lectura, ya que
- 20 la cabeza de lectura/escritura 11 no necesita buscar constantemente hacia atrás y hacia delante para reconstruir un archivo completo, o coger el siguiente archivo – simplemente permanece en posición en un cilindro 15 que lee los datos dentro de los AB 19 y que sigue lentamente a través de los platos 13 cuando se mueve desde el cilindro actual al siguiente.

- 25 Incluso las lecturas inversas (reproducción inversa, que es común con aplicaciones de video) son rápidas, ya que los AB 19 son leídos simplemente en orden inverso ya que giran por debajo la cabeza de lectura/escritura 11, siguiendo de nuevo lentamente con la cabeza 11 (hacia atrás esta vez) cuando se mueve desde un cilindro 15 al anterior.

Mínimo Movimiento del brazo de lectura/escritura

- 30 Una consecuencia obvia del SFS es que el brazo de lectura/escritura 12 no tiene que moverse constantemente hacia atrás y hacia delante entre los AB 19 y un directorio. De hecho, como se ha visto, el brazo 12 sólo necesita moverse muy lentamente a conveniencia cuando llena (o lee) todos los AB 19 contenidos dentro de un cilindro 15 y salta al siguiente. Esto reduce dramáticamente la energía requerida para accionar el brazo 12, dando como resultado una disipación de calor inferior y por ello menores temperaturas de funcionamiento.

- 35 Esta situación ventajosa es además mejorada por el desgaste inferior resultante en el conjunto accionador 10 y la reducción extrema de vibraciones (que de otra manera serían normalmente causadas por las rápidas aceleraciones y deceleraciones del brazo 12).

Estructura de Datos Robusta

- 40 La estructura de datos SFS es muy robusta como una consecuencia directa de su simplicidad y ordenación secuencial. Si las áreas de las superficies de los platos 13 de datos resultan contaminadas o dañadas en algún modo (por polvo, o un golpe en la cabeza, o simplemente errores de escritura a través de inexactitudes de la cabeza provocadas por vibración o golpe), los datos restantes son completamente utilizables. En un sistema convencional, se perdería un archivo entero si incluso una parte del archivo se deposita sobre un sector 16 que tiene un área dañada dentro de él. Más importante, el SFS no es sensible a los errores de directorio (que pueden provocar una pérdida completa de datos) ya que no hay requerimiento para un directorio. Así incluso con un disco dañado de forma severa (suponiendo que la cabeza de lectura/escritura está aun operativa), todos los datos que pueden aún
- 45 ser leídos permanecen útiles.

Recuperación de Datos Simple

- 50 Si falla una HDD 2, incluso catastróficamente, y se requieren los servicios de recuperación de datos especializados, entonces el SFS también ofrece ciertas ventajas adicionales. Como los AB 19 de sectores de datos están efectivamente autocontenidos, entonces cualquier AB 19 no dañado será aun capaz de ser leído. Por ello, una secuencia de datos útil (si está incompleta) es aun capaz de ser reformada a partir de un plato 13 de datos dañado

de forma severa.

Búsqueda fácil por fecha y hora, a pesar de no haber directorio

Emplear una placa controladora maestra 5 permite durante el período de tiempo que cada HDD 2 sea rastreada. Además, como los datos son almacenados en una secuencia estricta en cada HDD 2, es una razón bastante trivial encontrar un instante particular dentro de ese período de tiempo, incluso aunque no haya ningún directorio.

Dados los tiempos de comienzo y final para un plato 13 de datos particular, el sistema puede hacer una estimación razonable para la localización del punto de comienzo para cualquier secuencia de datos requerida a través de una interpolación en línea recta. El sello de tiempo y el período de tiempo dentro del bloque en el punto estimado es a continuación leído desde el sector de encabezamiento 20 y a continuación otra, estimación más exacta, hecha y el tiempo de lectura desde ese sector de encabezamiento 20 y así sucesivamente en un proceso iterativo hasta que los datos requeridos son encontrados.

Debería observarse que el método de localización de la secuencia de datos anterior se mantiene verdadero incluso si los datos almacenados no son generados de una manera estrictamente continua pero son de hecho almacenados de una forma continua escalonada. Por ejemplo, algunas grabaciones pueden no ser verdaderamente continuas en el tiempo, pero podrían comenzar y parar de acuerdo con algunos eventos externos tales como disparadores de alarma, o los datos pueden variar de volumen tal como con la tasa de marco o calidad de imagen en aplicaciones de vigilancia por video. Tales variaciones o discontinuidades de tasa de datos darán como resultado simplemente un ligero aumento en el tiempo de búsqueda (número de intentos de búsqueda) requerido para acceder al punto inicial para la secuencia de datos requerida.

Como el brazo de lectura/escritura 12 es capaz de moverse muy rápidamente, incluso un proceso de búsqueda que requiere muchos pasos iterativos será completado en unos milisegundos. Esto no es un problema para la recuperación de datos dentro de los sistemas descritos actualmente 1, 17 o 18, y sus aplicaciones pretendidas. Además, dada la arquitectura total del sistema, los datos requeridos para reproducción pueden de hecho estar situados en un disco de datos desconectado, requiriendo así aproximadamente 5 o 6 segundos para hacer girar el disco de datos listo para interrogación. Obsérvese sin embargo, que una vez que el punto inicial de la secuencia de datos requerida es encontrado, los datos pueden ser hechos circular para la reproducción a la máxima velocidad de lectura de la HDD 2 ya que la propia reproducción será secuencial (incluso si es en sentido inverso).

Índice de Pistas

Se ha encontrado que puede ser beneficioso incorporar un índice de pistas 22 al comienzo (o final) de un elemento de almacenamiento de datos de HDD. El índice de pistas 22 es una lista de los tiempos de inicio (y fechas) de cada archivo grabado, el nombre del archivo y el número de canal. El propósito de este índice de pistas 22 es proporcionar un modo más rápido para localizar un archivo particular por la fecha, la hora y el número de canal.

El índice de pistas 22 es secuencial en sí mismo en tiempo, y lo bastante pequeño para que el índice entero pueda ser leído en la memoria del controlador de CPU de forma extremadamente rápida, así la búsqueda del índice de pistas durante un instante particular y el canal es un proceso trivial y rápido. El índice de pistas 22 apunta a un número AB 19 para el instante de inicio de un archivo particular, y así el brazo de lectura/escritura puede ser movido directamente a esa posición y el archivo apropiado leído. En el sistema descrito actualmente cada entrada del índice de pistas 22 es de 32 bytes de largo, de tal modo que habrá 16 entradas por sector. Como resultado el índice de pistas 22 ocupa sólo alrededor del 0,0001% del espacio disponible en una HDD 2.

Por motivos de claridad debería resaltarse que el índice de pistas 22 no es el mismo que un directorio, como es empleado típicamente dentro de los sistemas de la técnica anterior. Existen varias características del índice de pistas 22 que lo distinguen de tales directorios, incluyendo el hecho de que el índice de pistas 22 es:

- Muy compacto, aumentando así la capacidad efectiva del plato 13 de datos (que usa menos del 0,0001% del espacio del disco, comparado con el 4-5% usado por un sistema de directorio convencional);
- Actualizado infrecuentemente, por ejemplo una vez cada pocos minutos, o alternativamente una vez cada vez que un número predeterminado de archivos ha sido escrito;
- No esencial para el proceso de localizar secuencias de datos; y
- Muy fácil y rápido de reconstruir, si es necesario.

El índice de pistas 22 es simplemente empleado como una guía para encontrar cualquier secuencia de datos particular de acuerdo con un sello de tiempo ya que esto mejora el tiempo de búsqueda, especialmente en el caso de que sea solicitada la reproducción a partir de muchos puntos diferentes dentro de un corto espacio de tiempo.

A partir de la descripción anterior puede verse que el SFS reduce además cualquier temperatura residual o vibración emitidas que permanecen dentro de los sistemas de almacenamiento de datos descritos previamente 1, 17 o 18, al

tiempo que mitiga el aumento de consumo de potencia resultante del empleo de dos HDD 2 una vez que está dentro del sistema 18 de datos de escritura copiada especularmente.

5 Explotando el SFS, la fiabilidad de los sistemas 1, 17 o 18 de almacenamiento de datos puede ser aun mejorada y la situación óptima conseguida para velocidades de escritura de datos. Este método es ideal para la grabación de video digital, la grabación de audio digital, la grabación de instrumentación, copia de seguridad de datos, archivado y aplicaciones similares en las que grandes volúmenes de datos son grabados en secuencia y típicamente sólo leído raras veces.

Cifrado de Disco de Datos

10 Como los datos almacenados en la HDD 2 son a menudo de naturaleza sensible o probatoria, un sistema de cifrado puede ser añadido a cualquiera de los procesos de lectura o escritura antes descritos. El proceso de cifrado de disco de datos es empleado a un nivel bajo y fundamental, haciéndolo así independiente de cualquier cifrado que pueda o no ser realizado a un nivel más alto (por ejemplo un DVR). Distintos esquemas de cifrado pueden ser así puestos en práctica para proporcionar seguridad a los datos, permitiendo solo que personas autorizadas lean los datos por ejemplo, o más importante, usar una clave de cifrado generada en máquina para proporcionar un método de codificar de modo seguro los datos de tal modo que los datos no podrían ser modificados o falsificados y vueltos a grabar posteriormente. El último esquema proporcionaría un método seguro de identificar donde y cuando fueron escritos los datos de una manera que podría ser verificada por el fabricante de los sistemas 1, 17 y 18 de almacenamiento de datos.

20 Los anteriores sistemas de almacenamiento de datos por ello proporcionan métodos y aparatos más baratos, más fiables, más eficientes en potencia y más convenientes para el almacenamiento de datos que los descritos dentro de la técnica anterior. Los sistemas de almacenamiento de datos descritos emplean una o más técnicas de grabación de datos secuencial. Tal grabación secuencial es conseguida mediante el empleo secuencial de agrupaciones de HDD 2 de bajo coste y/o el empleo secuencial de sectores individuales situados dentro de las propias HDD 2.

25 La descripción anterior del invento ha sido presentada con propósitos de ilustración y descripción y no pretende ser exhaustiva o limitar el invento a la forma precisa descrita. Las realizaciones descritas fueron elegidas y descritas a fin de explicar mejor los principios del invento y su aplicación práctica para permitir con ello a otros expertos en la técnica utilizar mejor el invento en distintas realizaciones y con distintas modificaciones ya que son adecuadas al uso particular considerado. Por ello, pueden ser incorporadas otras modificaciones o perfeccionamientos sin salir del marco del invento según ha sido definido por las reivindicaciones adjuntas.

30

REIVINDICACIONES

1. Un método de almacenar datos en un sistema (18) de unidad de disco duro, cuyo método comprende las operaciones de:
 - 5 1) seleccionar un primer elemento de almacenamiento de datos que comprende m unidades de disco duro (2) dispuestas en secuencia, en el que m es dos o más;
 - 2) escribir en secuencia los datos simultáneamente en cada una de las m unidades de disco duro (2) del primer elemento de almacenamiento de datos hasta que el primer elemento de almacenamiento de datos es llenado hasta su capacidad;
 - 10 3) seleccionar un elemento de almacenamiento de datos siguiente de m unidades de disco duro (2) dispuestas en secuencia;
 - 4) escribir en secuencia los datos simultáneamente en cada una de las m unidades de disco duro (2) del siguiente elemento de almacenamiento de datos hasta que el siguiente elemento de almacenamiento de datos es llenado hasta su capacidad;
 - 15 5) repetir las operaciones 3 y 4 hasta que un número predeterminado de elementos de almacenamiento de datos ha sido llenado hasta su capacidad;

caracterizado porque el siguiente elemento de almacenamiento de datos se solapa al elemento de almacenamiento de datos anterior en menos de m unidades de disco duro (2) del elemento de almacenamiento de datos anterior.
2. Un método según la reivindicación 1, cuyo el método comprende además repetir las operaciones 1 a 5 una vez que el número predeterminado de elementos de almacenamiento de datos ha sido llenado hasta su capacidad.
- 20 3. Un método según la reivindicación 1 o la reivindicación 2, en el que los elementos de almacenamiento de datos consisten de dos unidades de disco duro (2) dispuestas en secuencia.
4. Un método según la reivindicación 3, en el que el siguiente elemento de almacenamiento de datos se solapa al elemento de almacenamiento de datos anterior por una de las unidades de disco duro (2) del elemento de almacenamiento de datos anterior.
- 25 5. Un método según la reivindicación 4, cuyo método comprende además seleccionar un siguiente elemento de almacenamiento de datos en respuesta al fallo de una unidad de disco duro (2) del elemento de almacenamiento de datos anterior, dicho siguiente elemento de almacenamiento de datos desplazado del elemento de almacenamiento de datos anterior por m unidades de disco duro.
6. Un método según cualquiera de las reivindicaciones precedentes, en el que la operación de grabar los datos dentro de los elementos de almacenamiento de datos comprende además la operación de cifrar los datos.
- 30 7. Un método según cualquiera de las reivindicaciones precedentes, en el que la operación de escribir datos en secuencia en una unidad de disco duro (2) comprende las operaciones de:
 - A) seleccionar un primer bloque de asignación (19) de una agrupación de bloques de asignación dispuestos en secuencia situados dentro de la unidad de disco duro (2);
 - 35 B) grabar datos dentro del primer bloque de asignación (19) hasta que el primer bloque de asignación es llenado hasta su capacidad;
 - C) seleccionar un bloque de asignación (19) siguiente de la agrupación de bloques de asignación dispuestos en secuencia;
 - 40 D) grabar datos dentro del siguiente bloque de asignación (19) hasta que el siguiente bloque de asignación en secuencia es llenado hasta su capacidad; y
 - E) repetir las operaciones C y D hasta que un número predeterminado de los bloques de asignación (19) dispuestos en secuencia de la agrupación ha sido llenado hasta su capacidad.
8. Un método según la reivindicación 7, en el que el método comprende además repetir las operaciones A a E una vez que el número predeterminado de los bloques de asignación dispuestos en secuencia de la agrupación ha sido llenado hasta su capacidad.
- 45 9. Un método según la reivindicación 7 o la reivindicación 8, en el que la operación de grabar datos dentro de un bloque de asignación (19) comprende además la operación de grabar tiempos de inicio y final de los datos almacenados dentro del bloque de asignación.

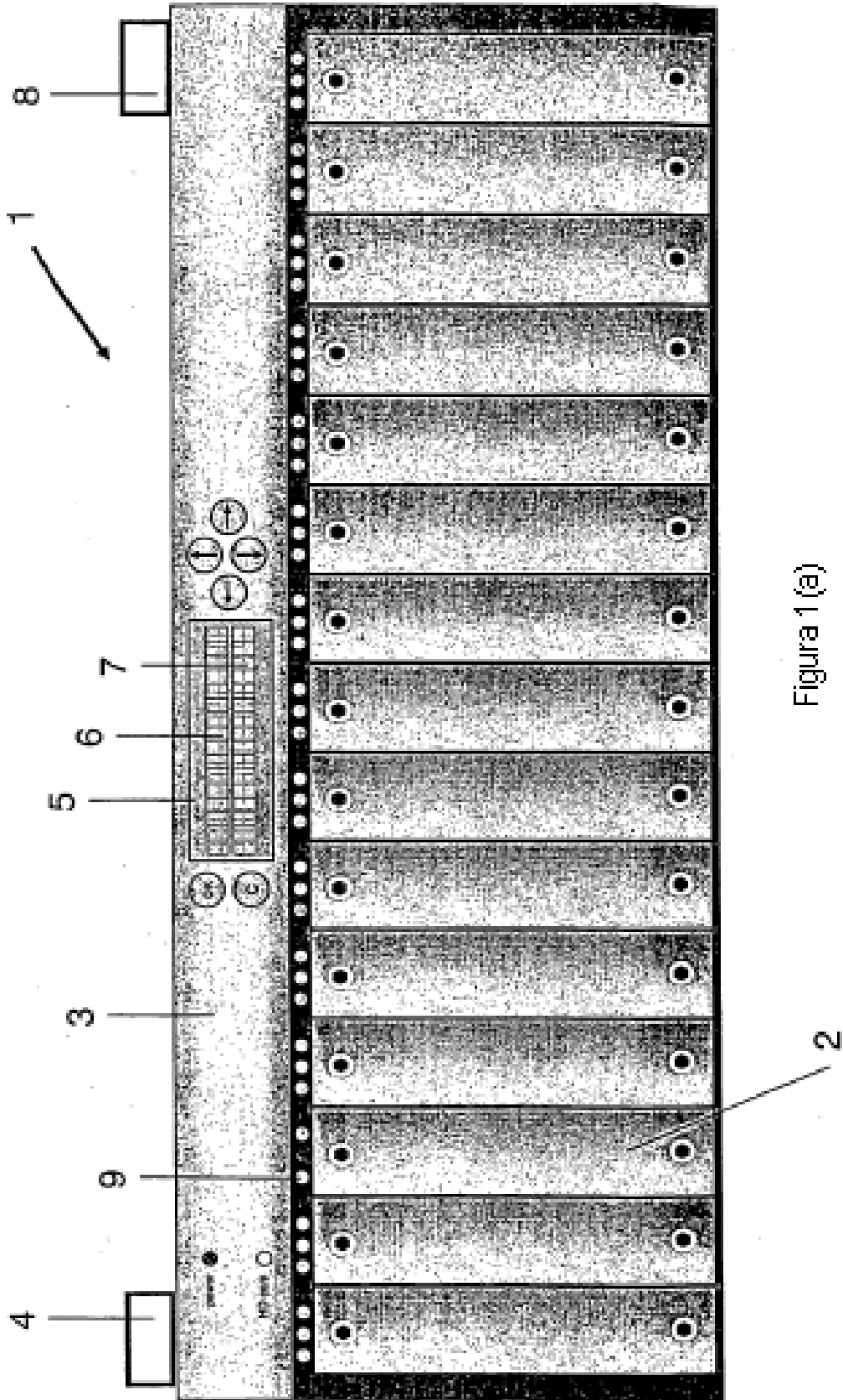
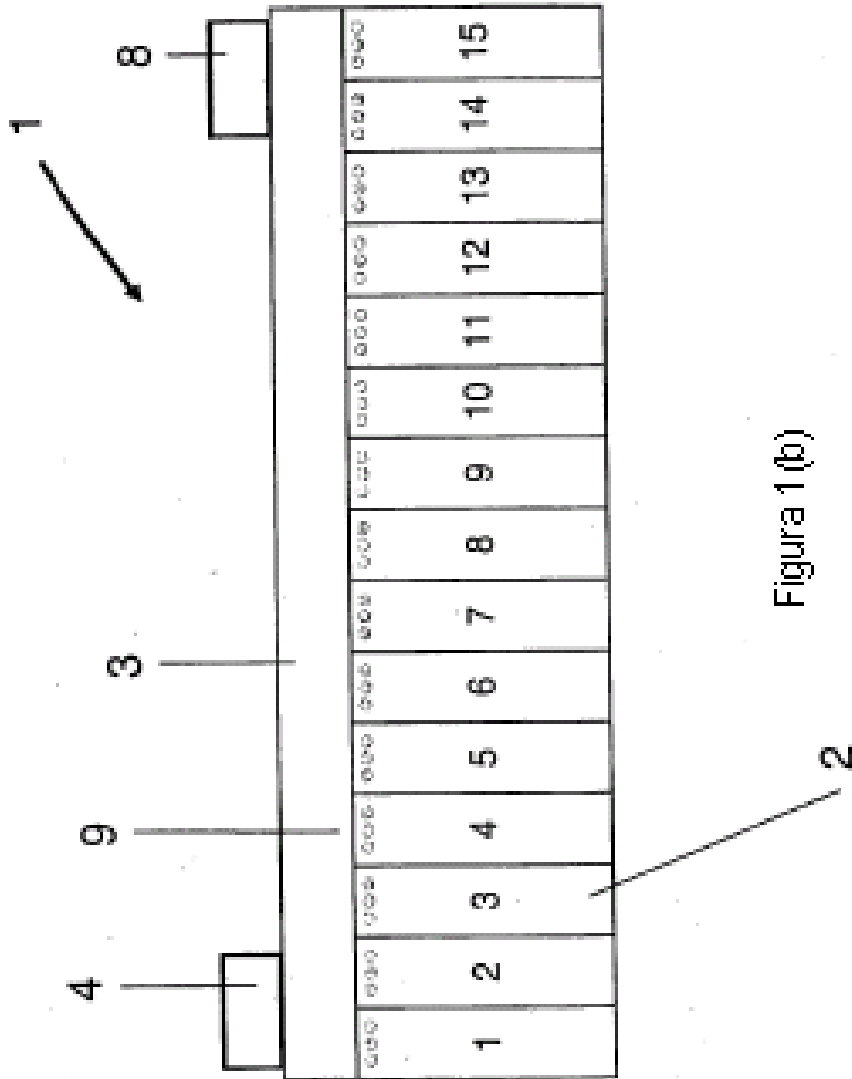


Figura 1 (a)



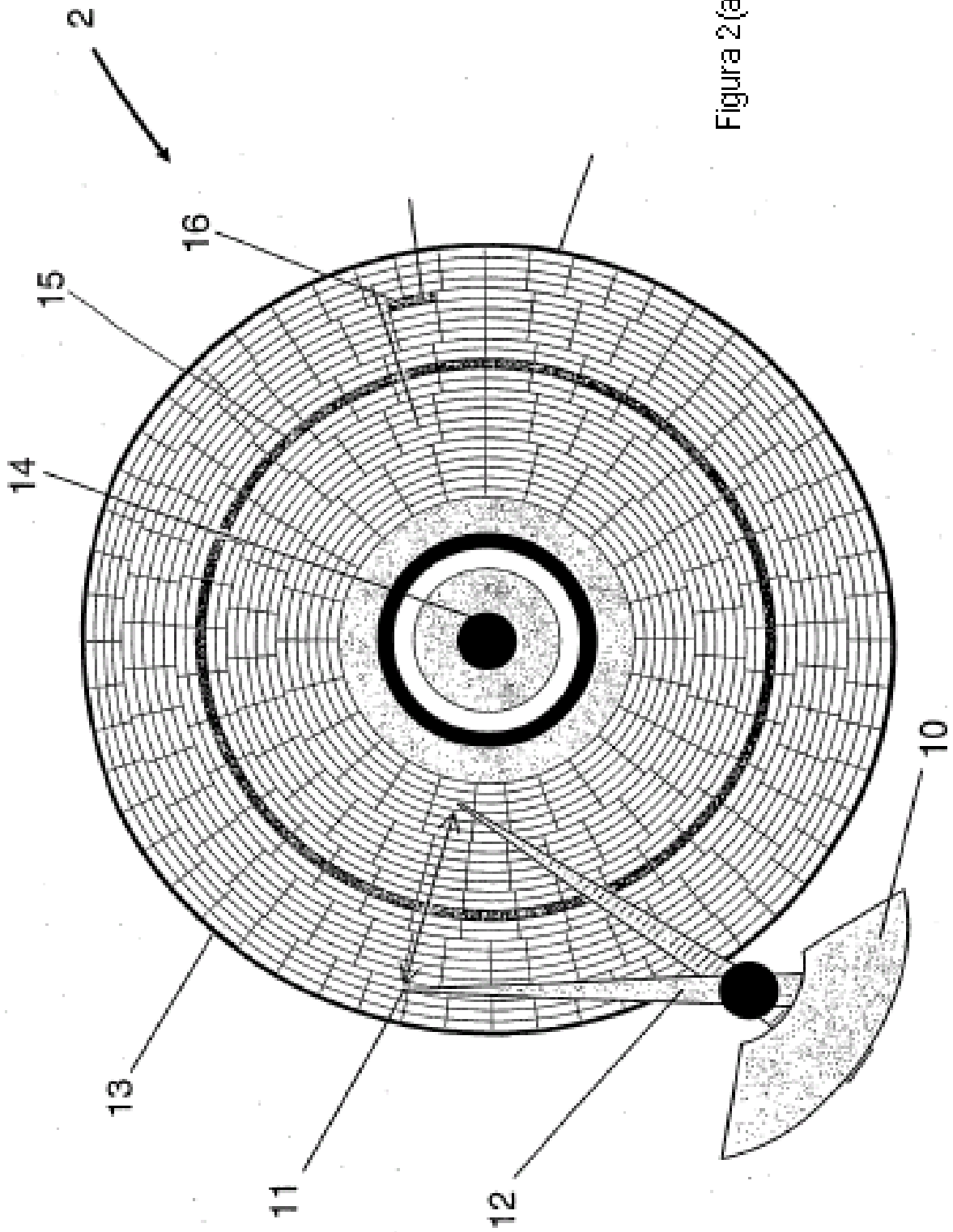


Figura 2 (a)

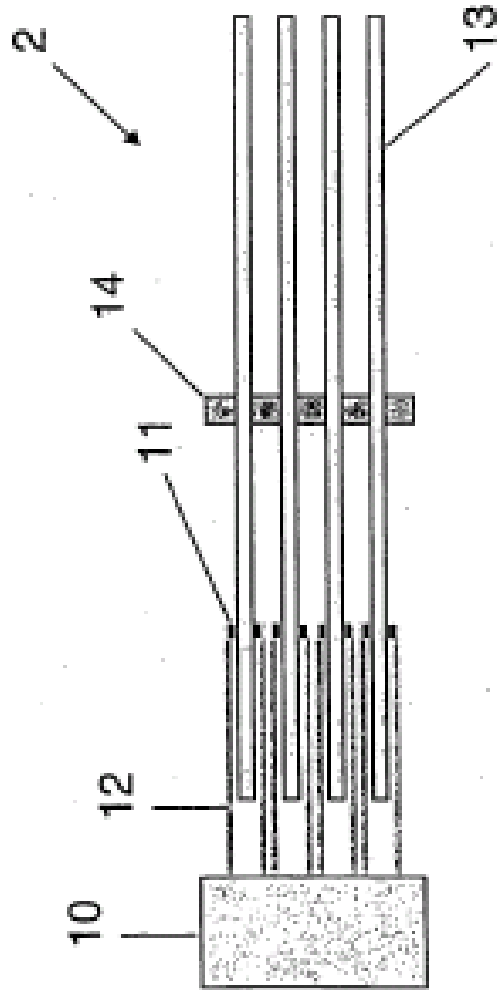


Figura 2(b)

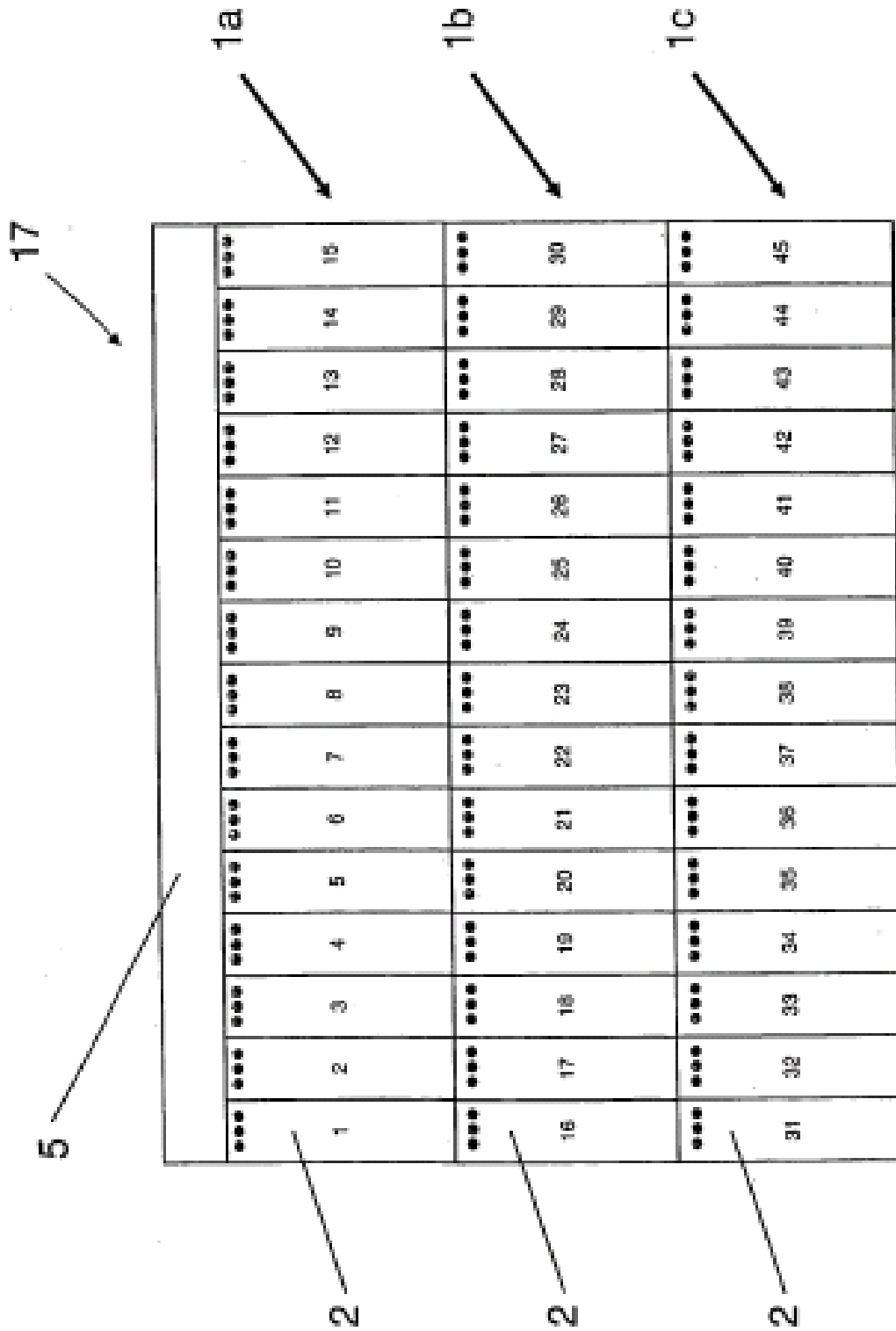


Figura 3

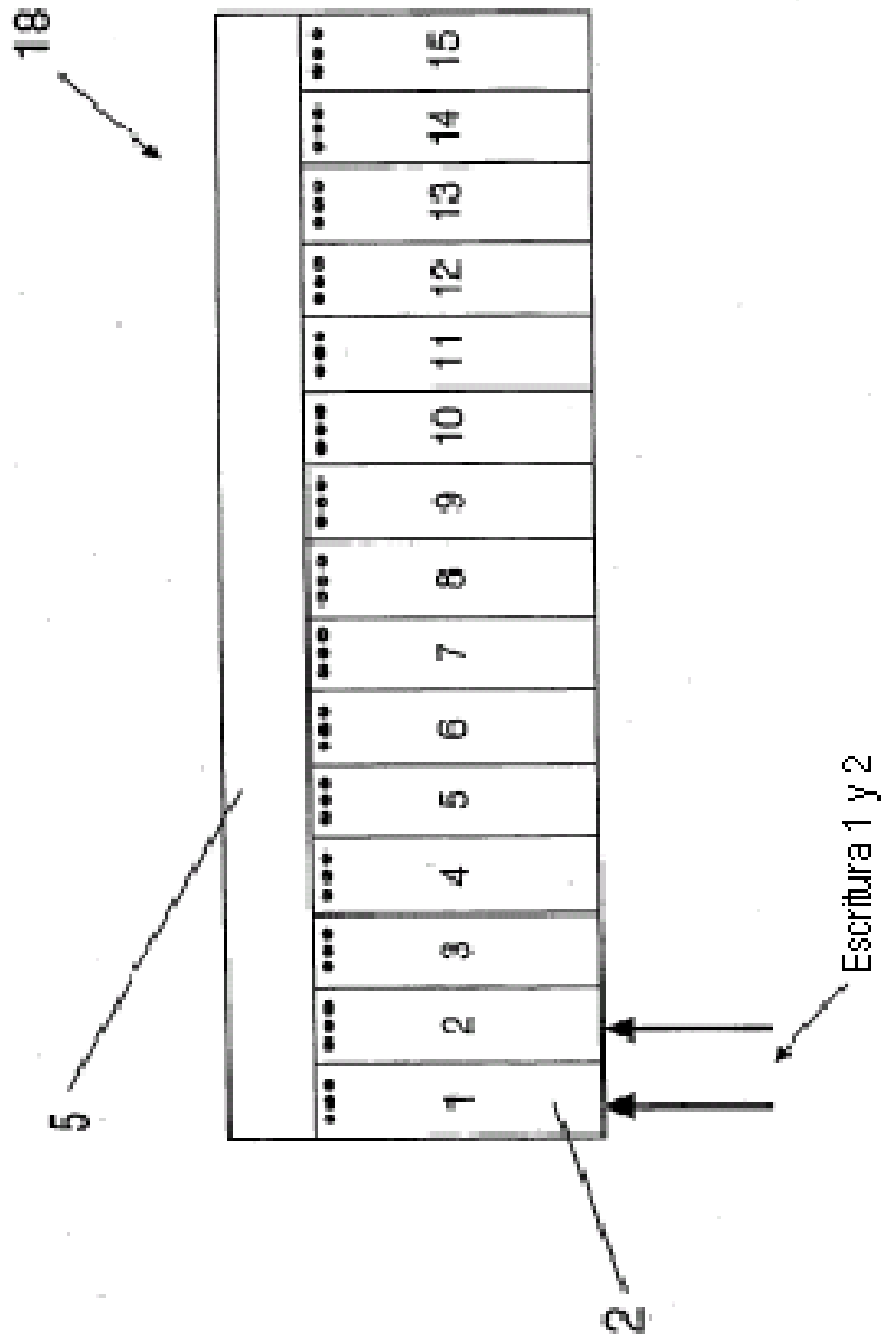


Figura 4(a)

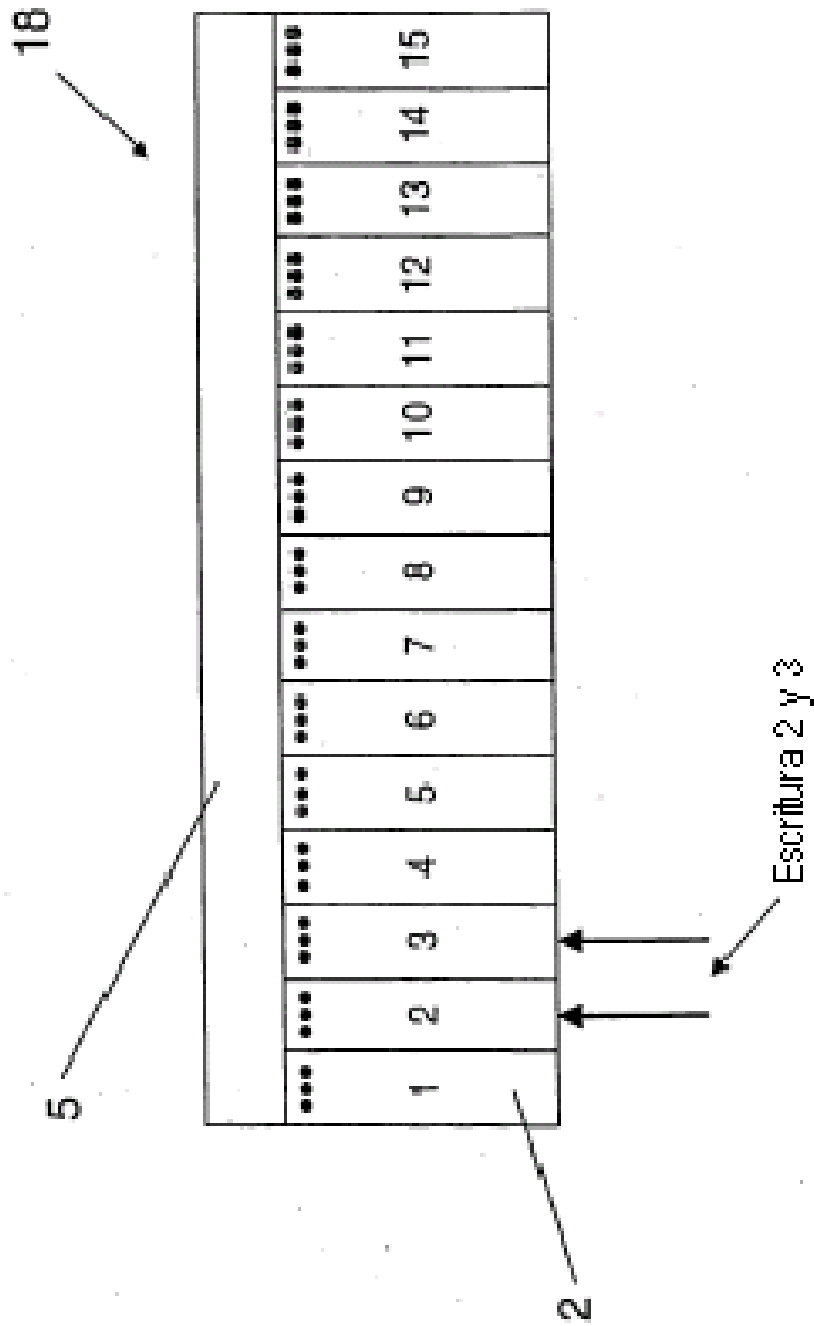


Figura 4(b)

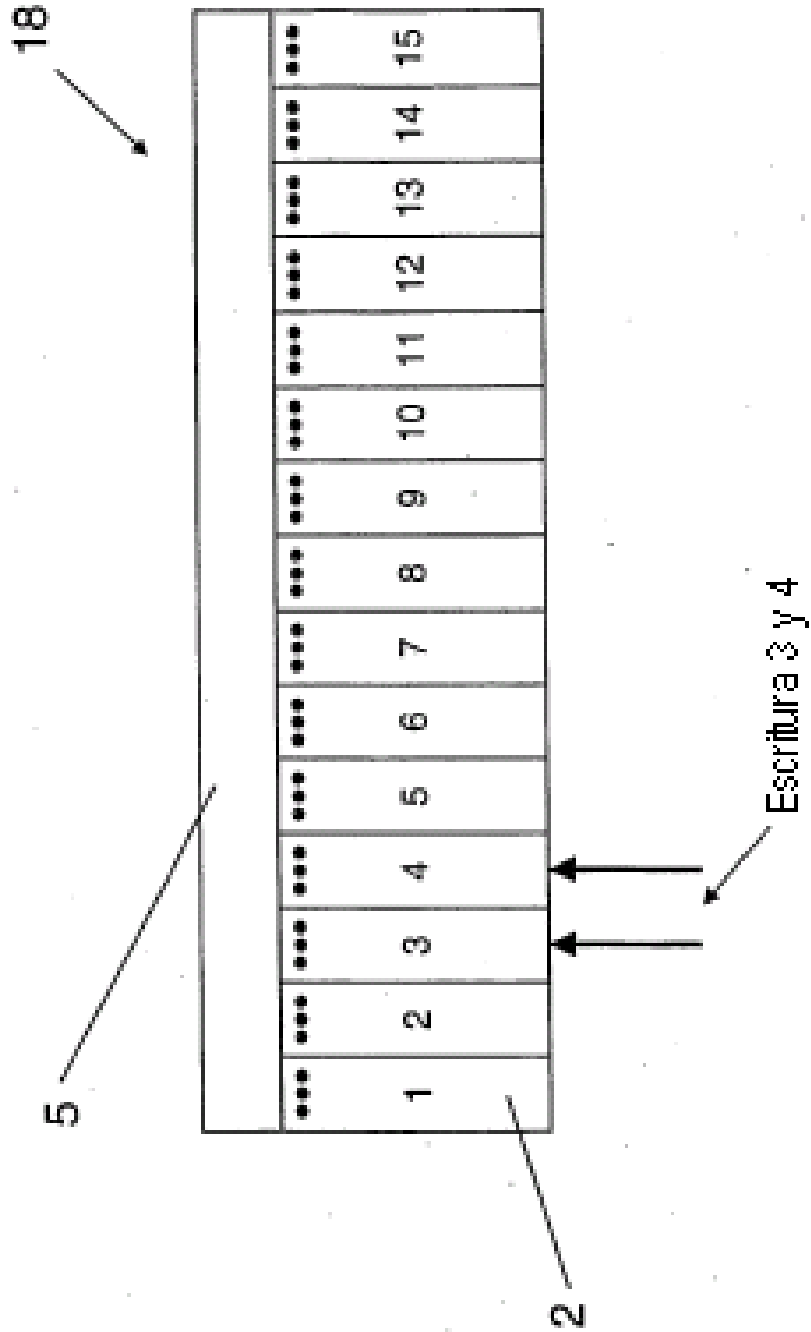


Figura 4(c)

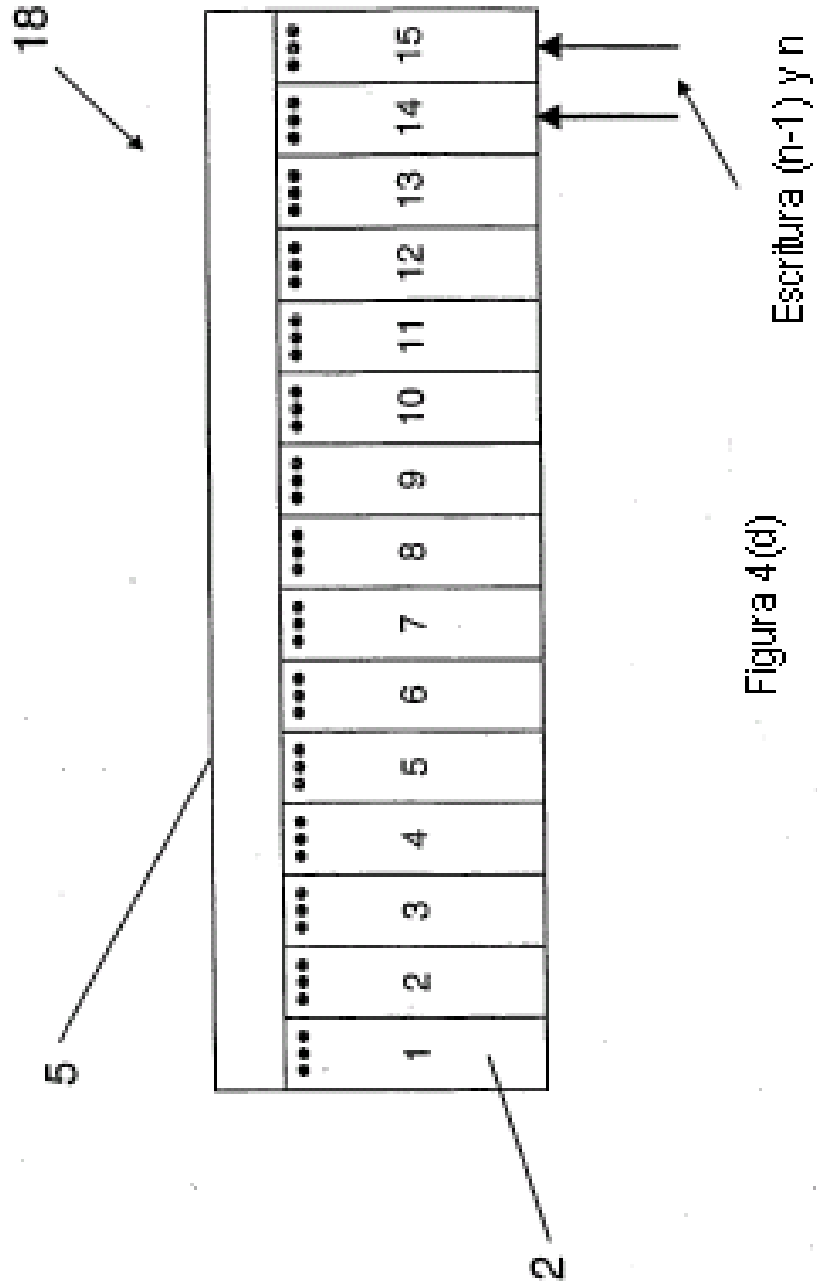
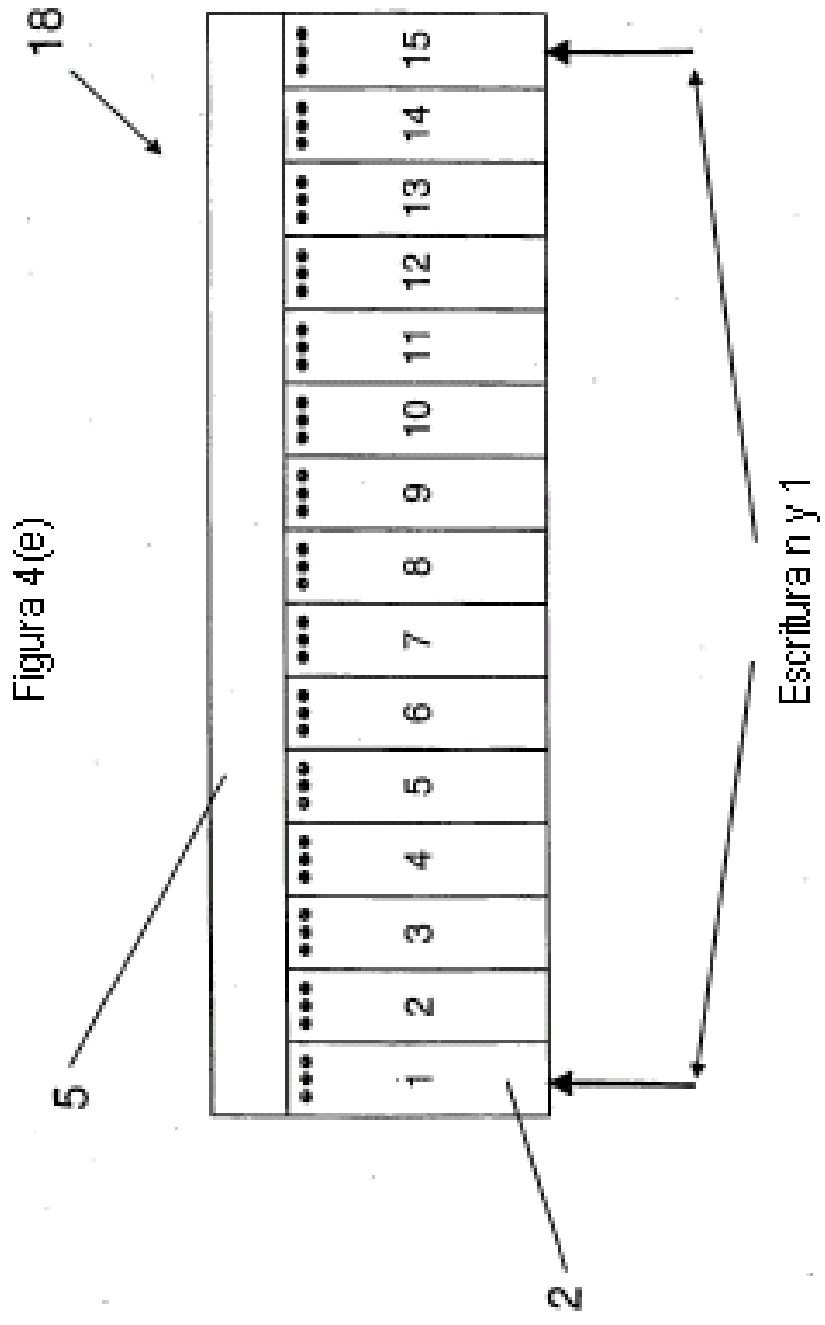


Figura 4(d)



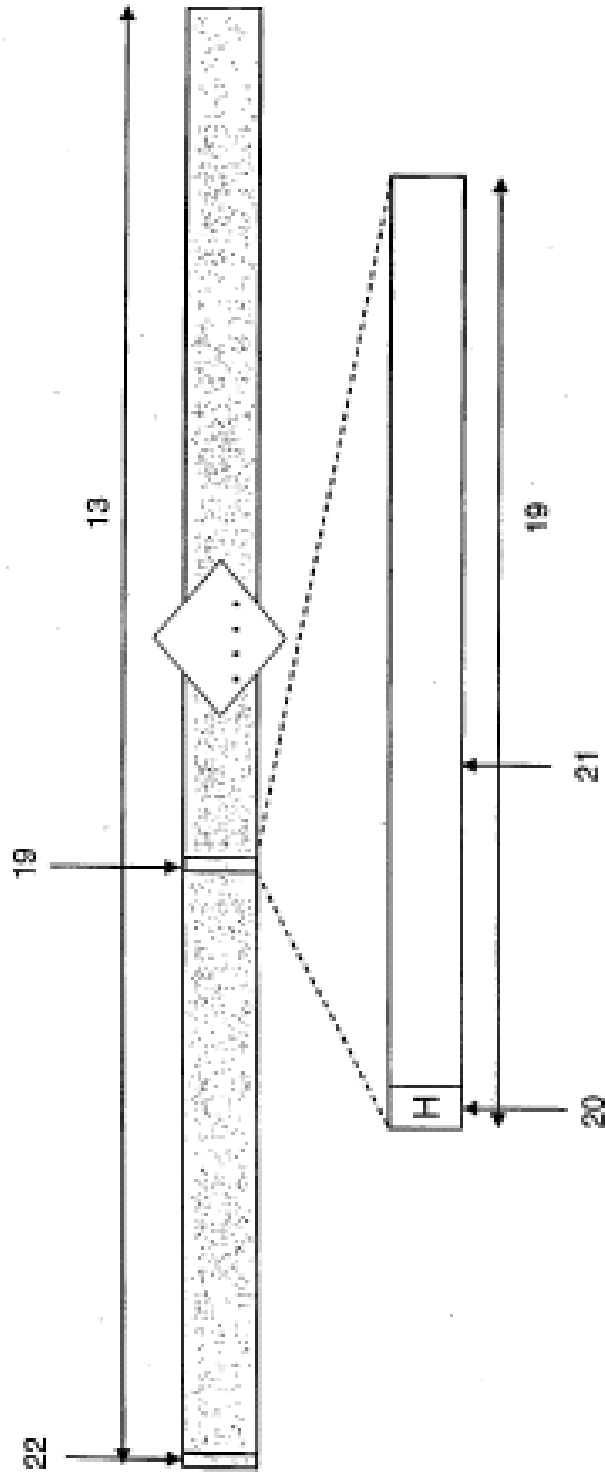


Figura 5