

Formatos de preservación final

bargo, si la cantidad de bloques malos aumenta puede ser signo de que el disco ha comenzado a fallar. Existe el *software* que suministra una alarma del aumento en bloques malos, y también mide otras características físicas que pueden indicar problemas en los discos.

4.7 ¿SCSI o IDE?

- 4.7.1 Existen dos métodos de conectar HDD y otros dispositivos periféricos a computadoras personales: SCSI (interfase de sistema para computadora pequeña) y ATA/IDE (AT, adición electrónica de transporte integrada). Cada una tiene ventajas particulares en ciertas situaciones. SCSI permite velocidades de escritura más rápidas y facilita el acceso a un número mayor de transportes que los transportes IDE. Los discos SCSI pueden aceptar varios comandos al mismo tiempo sobre una corriente de datos (*bus*) SCSI y no sufre de solicitudes de interrupción como es el caso del ATA/IDE. Los transportes ATA/IDE son más baratos comparativamente. La velocidad de lectura es prácticamente la misma, y en un contexto de audio ninguna interfase limitará la operación de la estación de trabajo digital (DAW) más que la otra.
- 4.7.2 Los transportes SCSI de Fibre Channel son usados más frecuentemente en los sistemas de la industria de los negocios, mientras que los transportes ATA/IDE más baratos están diseñados para el mercado personal, y la calidad de los productos se refleja en los precios. Es posible diseñar y construir transportes ATA/IDE que sean tan robustos y confiables como los transportes FC SCSI, y en algunos casos existe poca diferencia entre los transportes actuales. En suma, algunos transportes de bajo costo contienen meramente un convertidor de SCSI a ATA para un transporte ATA/IDE, en cuyo caso ya no cumplen con los requisitos de un transporte SCSI. Sin embargo, el rango de los transportes ATA/IDE son, en su totalidad, menos confiables que los discos más robustos FC SCSI. Los discos FC SCSI tienen un puerto dual, permitiendo controladores redundantes, y los ATA/IDE solamente tienen un puerto, haciendo la redundancia un poco difícil. El ATA serial (SATA) se ha desarrollado y se espera la siguiente versión para aumentar significativamente la velocidad de datos, y de SATA versión 2 también se espera que introduzca interfases de puerto dual.
- 4.7.3 La interfase no es una indicación completamente consistente de la confiabilidad de un transporte dado y el comprador deberá estar atento a otros parámetros operativos, tanto es así que actualmente los transportes más confiables se asocian con la interfase FC SCSI. No obstante, los HDD no son permanentemente confiables, y todo dato de audio deberá respaldarse en una cinta conveniente (véase 6.3 Formatos y tipos de cinta de datos). (Más información en Anderson, Dykes y Riedel, 2003).

6.5 Sistemas manuales de almacenamiento masivo de datos digitales en pequeña escala

6.5.1 Introducción

- 6.5.1.1 En la segunda mitad de los años noventa, sólo las organizaciones grandes y con finanzas sanas —predominantemente con archivos sonoros de radio— pudieron digitalizar sus materiales en gran escala y almacenarlos en sistemas de almacenamiento masivo digital. Estos sistemas tienden a ser de almacenamiento dedicado a audio y audio-videos grandes y caros. En años más recientes, muchos archivos de sonido y grandes bibliotecas estatales han instalado e implementado sistemas de almacenamiento digital basados en computadoras más genéricas para sus colecciones. Las instituciones más pequeñas, que apenas se han iniciado en la digitalización, han tendido a almacenar la información digital sobre un formato digital final de soporte único, como el CD o el R-DAT. Con la rápida disminución de los precios en las cintas profesionales de respaldo

Formatos de preservación final

para computadora y HDD ahora es posible adoptar un almacenamiento digital más profesional que los relativamente menos confiables formatos digitales finales de soporte único.

- 6.5.1.2 Este capítulo describe un enfoque manual (no automatizado) del archivamiento digital en sistemas de almacenamiento en pequeña escala. Esta estrategia es mejor para colecciones de tamaño pequeño o mediano, en las que la automatización no es posible o apropiada, o que se encuentre en una etapa de proceso de automatización más grande. En este tema es importante consultar las secciones 6.2 Sistemas de almacenamiento digital masivo (DMSS); 6.3 Formatos y tipos de cinta de datos, y 6.4 Discos duros (HDD).

6.5.2 Descripción del sistema

- 6.5.2.1 Esta sección describe ampliamente dos situaciones en las cuales los sistemas de almacenamiento en baja escala pueden implementarse: únicamente un operador digitalizando sobre un solo dispositivo de almacenamiento, y el caso en el que más de un operador necesite entrar en el dispositivo de almacenamiento. Existen diferencias en la complejidad de cada sistema, pero también muchas similitudes fundamentales. Cualquier sistema presupone cumplir con todos los otros componentes mencionados en los *Lineamientos*, incluyendo convertidores apropiados de análogo a digital, tarjetas de sonido adecuadas, estaciones de trabajo digitales (DAW) y dispositivos apropiados de reproducción.

6.5.2.2 Sistema de almacenamiento de uno o dos operadores

- 6.5.2.2.1 Esta configuración solamente es posible para instituciones con un operador en el proceso de digitalización. Un requisito para el éxito de este enfoque es un plan bien estructurado de digitalización y la asignación de un dispositivo de disco de audio exclusivo, de manera que el trabajo se lleve a cabo continuamente y sin interrupciones. Esto asegurará que el HDD en la DAW se llene y vacíe continuamente siempre que se alcance el tope para llenar de datos el medio. Se recomienda el uso de dos dispositivos de disco separados, uno para el sistema operativo y el otro para los datos de audio. Esto facilitará el mantenimiento y optimización de los dispositivos de disco y asegurará que el respaldo de datos y el acceso se logren eficientemente.

- 6.5.2.2.2 Conforme aumenta la capacidad (normalmente 200 GB) de las cintas de datos como la LTO (véase la sección 6.3.2 Cinta de datos, formatos y tipos), una cantidad considerable de datos de audio pueden almacenarse en un dispositivo de disco antes de que sean respaldados en cinta. Este procedimiento es peligroso debido a que el HDD en la estación de trabajo digital (DAW) puede fallar en cualquier momento y la labor de muchos días –probablemente de varias semanas– puede perderse. Es, por lo tanto, recomendable que un segundo dispositivo de disco de audio sea usado como almacenamiento momentáneo antes de que los datos de audio se envíen al medio final, aumentando considerablemente su seguridad con esta medida.

- 6.5.2.2.3 Si dos estaciones de trabajo con sus operadores se encuentran realizando la tarea de digitalización, será necesario suministrar acceso a uno o varios dispositivos de disco compartidos. Es posible compartir tales recursos definiendo una de las computadoras como servidor y configurándola de forma que administre los dispositivos de disco, implementando además la capacidad de compartir los recursos con un cable. Tal enfoque es relativamente sencillo de implementar y permite compartir los recursos entre dos operadores, pero aun así se necesitan algunos acuerdos para evitar conflictos.

- 6.5.2.2.4 La organización lógica de los datos y el procedimiento estricto de denominación son una necesidad de los sistemas manuales de almacenamiento en pequeña escala. Sin un *software*

Formatos de preservación final

de administración jerárquica del almacenamiento (HSM) y bases de datos integrados bien desarrolladas, es obligado organizar y estructurar los datos de manera que la información permanezca accesible y localizable. El sistema y el procedimiento deben tomar en cuenta la posibilidad de saber qué proyectos particulares se interrumpirán y cuáles no serán completados en su orden natural o designado. La solución más común es aumentar el espacio de almacenamiento, de modo que el proyecto pueda ser copiado en cinta de datos a tiempo y ordenadamente.

5.2.3 Almacenamiento de datos en HDD

5.2.3.1. Aunque más costosa, la solución ideal para el problema de la seguridad de los datos y HDD es conectar los HDD a un conjunto de HDD configurados en un arreglo RAID (véase la sección 6.4.3.2). El nivel 1 de RAID es un poco mayor que dos dispositivos de disco en redundancia espejo; por esto es necesario mantener dos copias de los datos en diferentes dispositivos de disco por separado; si un disco falla, el otro se encuentra disponible. Niveles superiores de RAID (del 2 al 5) implementan sistemas cada vez más complejos de redundancia de datos y verificación de paridad que aseguran que la integridad de los datos se mantenga. Los niveles más altos de RAID logran el mismo nivel de seguridad que el nivel 1, o espejo, pero con un espacio significativamente menor de espacio de almacenamiento. El RAID 5, por ejemplo, tiene una pérdida de 25% de almacenamiento comparado con el 50% del RAID 1.

5.2.3.2. Sin embargo, los arreglos de discos son relativamente caros y no particularmente flexibles. Una alternativa es un administrador de volumen basado en *software* el cual se desempeña con funciones similares en *software*. Un administrador de volumen puede soportar también el respaldo y los procedimientos similares de administración de datos tratando a los múltiples dispositivos de disco como un solo volumen. Existen en el mercado varios desarrolladores de *software* de administración de volumen.

5.2.4 Sistemas de almacenamiento con operadores múltiples

5.2.4.1 Para cualquier número de conexiones mayor a dos se deben implementar los sistemas en red de almacenamiento y respaldo. Tales sistemas en red permiten el acceso a múltiples usuarios de acuerdo con las reglas establecidas por el sistema de administración de datos. Las redes de pequeña escala son relativamente comunes, más o menos conocidas, ofrecen implementación fácil y tienen precios accesibles. Una cantidad adecuada de almacenamiento se puede lograr con un dispositivo de almacenamiento en red adicionado de nivel empresa (NAS). Comparados con dispositivos más caros, la mayoría de los dispositivos NAS de bajo costo tienen un ancho de banda reducido, con lo que se obtienen tiempos de acceso más lentos o menor disponibilidad de accesos simultáneos. Esto no deberá presentar mayor problema para colecciones menores, ya que la necesidad de accesos simultáneos es baja, especialmente si para ello se usan medios convencionales como el CD, o derivados de los MP3 que se obtienen de las copias maestras de preservación.

5.2.4.2 Un sistema típico de almacenamiento en red en pequeña escala puede estar compuesto de una computadora de escritorio como servidor conectada a un dispositivo NAS. El NAS podría tener la capacidad de montar varios discos duros en un arreglo RAID. En promedio, un NAS barato puede contener entre 0.5 a 3 TB de almacenamiento en disco (note que para el RAID hay menos almacenamiento que el indicado para el disco individualmente). Las estaciones digitales de audio (DAW) acceden al NAS vía un conmutador Ethernet o dispositivo similar que, si se configura adecuadamente, tiene el efecto de separar el almacenamiento de la red de área local (LAN) y mejora la seguridad del sistema de almacenamiento. Los HDD podrían respaldarse en cinta de datos en un formato confiable para su almacenamiento a largo plazo.

Formatos de preservación final

6.5.2.5 Respaldo en cinta

6.5.2.5.1 Aunque no hay un solo componente de sistema digital que pueda considerarse confiable, esta cualidad se logra a través de múltiples copias extra en cada etapa. El componente final y más importante en la cadena de almacenamiento es la cinta de datos. En el pasado reciente LTO ha ganado popularidad para este propósito (véase la sección 6.3.9 Selección y monitoreo del medio de cinta de datos); sin embargo, otros formatos de datos en cinta pueden ser los apropiados en cada caso particular.

6.5.2.5.2 Todos los datos almacenados en disco deberán ser duplicados sobre una cinta de datos adecuada. Un mínimo de dos conjuntos de cintas de datos deben producirse y almacenarse en lugares distintos. Como es frecuente que el segundo conjunto de copias se emplee para la restauración de datos, muchos archivos hacen tres conjuntos de copias: dos guardados cerca del sistema para tenerlos a la mano, y un tercero, lejos, en prevención de algún desastre. Se ha hecho una costumbre que los conjuntos de cintas de datos separados se hagan con diferentes productos, de los cuales una cantidad considerable del mismo lote se adquieren al mismo tiempo. Esto facilita el control de calidad y las medidas de rescate, una vez que un lote de un producto dado falla. El *software* apropiado de administración del volumen ayudará en el proceso de respaldo y recuperación, especialmente si el sistema incorpora un número de dispositivos de almacenamiento.

6.5.3 Riesgos

6.5.3.1 Los sistemas de almacenamiento automatizado pueden configurarse para copiar y renovar constantemente los datos, desechando las cintas de datos que se han convertido en no confiables. Los sistemas de almacenamiento masivo digital en gran escala están diseñados profesionalmente y funcionan apoyados en organizaciones consolidadas, que pueden sostener y garantizar todas las medidas necesarias para la seguridad de los datos. Con los sistemas de respaldo y recuperación manual no deben subestimarse los peligros de pérdidas de datos asociados con los sistemas exclusivamente diseñados y administrados para digitalización manual y semiautomatizada. La seguridad de que los datos archivados de audio permanezcan válidos y accesibles cae dentro de un ámbito individual y requiere que la persona responsable verifique regularmente las cintas de datos. Esta situación específica se agrava porque la mayoría de las instituciones culturales y de investigación no disponen de grandes recursos.

6.5.3.2 Aun cuando el diseño de tales sistemas parece requerir un alto número de copias extra, se debe tener en mente que los soportes y componentes digitales pueden fallar en cualquier momento, sin ninguna indicación previa. Por lo tanto, es imperativo tener en cualquier etapa del proceso de digitalización y en el subsiguiente almacenamiento al menos dos copias del archivo lineal. Cualquier falla producirá una pérdida de pequeñas o grandes cantidades de datos, pero si se ha seguido la estrategia adecuada, la falla no será irrecuperable gracias a las copias extra disponibles. Dado que el proceso de transferencia consume mucho tiempo –sin mencionar la pérdida de materiales más viejos– cualquier esfuerzo deberá hacerse para no tener que volver a digitalizar materiales por inconsistencia en el diseño de seguridad o por ambigüedad en el flujo de trabajo.

6.5.4 Complejidad del sistema

6.5.4.1 Una vez completados e instalados los sistemas de almacenamiento de datos son relativamente sencillos de operar y mantener. Sin embargo, durante las etapas iniciales de implementación y en cualquier problema siguiente o de modernización, se recomienda mucho la asesoría de soporte IT especializado.

Formatos de preservación final

6.5.5 Costos y crecimiento

- 6.5.1 La etapa final de un sistema de almacenamiento masivo debería ser eventualmente la introducción de una caja robotizada para manejar automáticamente el creciente número de cintas de datos, combinado con un *software* de administración jerárquica del almacenamiento (HSM). Pero en este punto debe advertirse que el *software* HSM es proporcionalmente todavía muy caro. Puede esperarse, sin embargo, que con la creciente demanda a nivel mundial de almacenamiento masivo en pequeña escala en el rango bajo de TB, el *software* para la administración de esos objetos digitales llegue a ser más barato.
- 6.5.2 El costo de instalación de un sistema de almacenamiento de datos en pequeña escala puede parecer relativamente alto si se le compara a comprar un quemador de CD individual; pero si la comparación toma en cuenta cada bit almacenado en unas cuantas horas de audio, la diferencia relativa se reduce notablemente cuando se costean todos los requerimientos de un archivo. En resumen, una instalación de almacenamiento de datos bien administrada en su conjunto es un sistema muy confiable y permitirá la futura transferencia de datos de audio hacia la siguiente solución de almacenamiento cuando esto inevitablemente ocurra.

6.6 Discos ópticos: CD/DVD para grabar

6.6.1 Introducción

- 6.6.1.1 El CD para grabar (CD-R) y el DVD para grabar (DVD-R) se han convertido en una parte integral en la grabación y distribución de muchos tipos de materiales de audio y audiovisual. La fácil disponibilidad y bajos costos de preparación los han hecho atractivos para colecciones pequeñas y no tan pequeñas. Sin embargo, cuando se trata de contar con un servicio confiable para una administración exitosa de archivos digitales de audio en CD-R o DVD-R, se requiere una considerable inversión en conocimiento y tecnología que va más allá de la requerida para producir un disco. También es importante enfatizar que los discos ópticos de cualquier tipo para grabar no representan una solución permanente. El decaimiento del soporte y la obsolescencia del formato requerirán que el contenido del disco sea migrado a otro formato en el mediano plazo. Como ya se ha dicho, existen otras soluciones que pueden ser más efectivas en el largo plazo.
- 6.6.1.2 Teniendo en mente estas restricciones es posible usar los discos ópticos para grabar como soportes confiables en el corto plazo, teniendo cuidado en tomar en cuenta puntualmente las siguientes recomendaciones.

6.6.2 Formatos de grabación en CD-R y DVD-R

- 6.6.2.1 Hay dos maneras diferentes para codificar el audio y el video en CD o DVD para grabar: ya sea como "corriente" de datos o como un archivo de datos. En la primera, el sonido es grabado en el formato CD-DA pudiéndose realizar en un reproductor de CD ordinario, o bien codificarlo en formato MPEG para DVD, el cual no se reproducirá necesariamente en todos los reproductores de DVD. Las grabadoras individuales grabarán solamente esos formatos, pero el equipo basado en computadoras puede producir opcionalmente discos en esos estándares domésticos. El uso de los formatos restringe severamente la posibilidad de su acceso en línea y eligiendo esta opción puede crear un problema de migración cuando vuelva a ser necesario cambiar de soporte.
- 6.6.2.2 La alternativa —grabar un archivo empleando un sistema de edición de audio basado en computadora y escribiendo el archivo a CD-R o DVD-R— es un enfoque más confiable. Grabando archivos en