



© Foto: Worth1000.com

A evolução dos sistemas de Gestão de Acervos digitais

por Fabio Tsuzuki e José Olairson Valentim

Sistemas de gestão de acervos digitais são designados pela sigla DAM: *Digital Asset Management* ou MAM: *Media Asset Management*. O desenvolvimento destes sistemas iniciou-se há pelo menos 20 anos e agora o mercado já assimilou a sua necessidade. A sigla MAM tem sido usada na indústria de *broadcast* enquanto que a sigla DAM tem sido utilizada nas outras indústrias. Em essência designam sistemas que resolvem os mesmos problemas: problemas relacionados com a gestão de acervos digitais.

Em termos gerais são sistemas que se propõem gerenciar o conteúdo independente do seu formato. Então é possível gerenciar qualquer formato? Esta é a primeira dúvida. Qual o formato de arquivo que deve ser gerenciado?

Recomenda-se o uso de arquivos com maior resolução possível, pois assim é possível produzir qualquer outro formato de menor resolução. A decisão sobre qual formato gerenciar é bastante complexa, pois um sistema de gestão precisa contemplar as diferentes etapas do ciclo de vida do conteúdo: captura, edição, distribuição e arquivamento.

Para acervos de imagens fixas tais como fotos digitais, fotos digitalizadas e ainda documentos digitalizados o formato a ser gerenciado está relacionado com a finalidade (etapa de distribuição) destas imagens. Considerando imagens que serão utilizadas em internet, basta o formato JPEG RGB, para as imagens que precisam ser impressas será necessário gerenciar JPEG

CMYK. Caso a impressão seja em alta resolução para uso em painéis, cartazes ou mesmo *outdoor*, então o formato de mais alta resolução disponível são formatos que armazenam a imagem em um padrão RAW (imagem sem compressão). Um exemplo deste padrão são as imagens produzidas pelos equipamentos *Hasselblad* cuja extensão é FFF. Imagens com este padrão chegam a ter um tamanho de 50MB.

No universo de vídeos temos mais complexidade, pois o uso dos formatos está relacionado não apenas com a etapa de distribuição, mas com cada uma das etapas: captura, edição, distribuição e arquivamento. Muitos sistemas têm-se especializado em etapas específicas. Existem sistemas de MAM voltados para arquivamento – *Library Asset Management*, edição – *Production Asset Management* etc.

Esses sistemas especializados são focados em uma determinada etapa e assim tornam as decisões estratégicas mais simples. Alguns sistemas não gerenciam *assets*, mas apenas arquivos e assim são bastante limitados, pois um dado conteúdo pode ser representado por diferentes arquivos com características bem distintas. Tratando-se de vídeo podemos citar as principais características: resolução, *bitrate*, codec e *wrapper* (empacotamento do vídeo).

Um sistema que se propõe a apoiar todas as etapas do ciclo de vida do conteúdo deve contemplar o gerenciamento de vários formatos para um mesmo conteúdo. Trata-se de uma gestão multi formato. Adotar uma



solução que resolve uma etapa específica consiste em adiar a adoção de uma solução que gerencie todas as etapas. Em algum momento será necessário integrar todas as etapas e assim estabelecer a visão abrangente do ciclo de vida do conteúdo. Adotar um ambiente de exibição capaz de exibir e também editar qualquer formato ajuda, pois permite uma melhor compatibilidade entre cada uma das etapas. Decidir qual a solução para o ambiente de edição bem como qual a solução do ambiente de exibição depende de como é o ciclo de vida do conteúdo que está sendo elaborado. Vejamos alguns exemplos:

- Um canal que exibe muito material de estúdio: o formato a ser gerenciado deve ser o formato de captura e possivelmente o formato de exibição caso os sistemas de captura e exibição não tenham um formato em comum. Normalmente, este tipo de operação é projetado com esta compatibilidade mínima.
- Um canal que exibe muito material produzido: caso o formato editado seja compatível com o formato de exibição então é bastante natural gerenciar o formato de edição, mas se o formato não for compatível então é necessário realizar uma conversão compatibilizando com a exibição e este formato de exibição também deve ser gerenciado.

Quando se passa para uma visão mais ampla percebe-se que não será possível ter um único formato sendo gerenciado. Como já foi citado veja a necessidade de ter um formato de captura, normalmente compatível com o ambiente de edição, e o formato de edição que nem sempre é compatível com o ambiente de exibição.

Formatos de arquivos

Em um ambiente *broadcast* a etapa de distribuição está intrinsecamente relacionada com a etapa de exi-

bição, e por esta razão não existe uma preocupação em gerenciar o formato de distribuição, pois ele é produzido ao longo do processo de exibição. Por outro lado, existem métodos alternativos de distribuição de vídeo, e a internet é um bom exemplo, são as novas mídias, e os formatos para distribuições desta que estão se somando aos formatos de cada uma das etapas já mencionadas.

Ter um formato compatível com todas as etapas simplifica bastante o processo de gestão e ainda permite uma maior agilidade em todos os fluxos de trabalho, pois não existe a necessidade de realizar conversões de formato. Caso a conversão seja apenas a troca do empacotamento do vídeo (*wrapper*) então é um processo simples, comparado com o processo de recodificar um vídeo que consiste na alteração do codec, da resolução ou mesmo do *bitrate* desse vídeo.

Outro aspecto relevante está relacionado com a evolução dos formatos ao longo do tempo. Mesmo que se adote um formato único e compatível com todas as etapas neste momento, depois de certo período (vários anos), esse formato poderá ser uma barreira operacional e então ele deverá ser abandonado e um novo formato adotado. Essa troca será inevitável, e a conclusão é bastante simples: os sistemas de MAM precisam ser capazes de gerenciar múltiplos formatos e se adaptarem a essa evolução.

O uso de multi-formatos é justificado para integrar diferentes etapas no tratamento do conteúdo ou para compatibilizar a operação com um novo formato que facilite a execução das atividades em cada etapa. Além de serem capazes de gerenciar múltiplos formatos os sistemas de MAM também precisam ser capazes de converter os formatos. Caso seja necessário re utilizar um conteúdo antigo, será necessário convertê-lo para o novo padrão adotado ou mesmo converter para compatibilizar o formato em cada etapa.

Abaixo está apresentado um quadro com informações estratégicas para apoiar a definição de um formato.

Matriz de Compatibilidade Codecs e Wrappers						
	MXF OP1a	MOV	DV	IMX	MPEG2	XDCAM
K2-SDA REC	S	S	S	S	S	N
K2-DAS PLAY	S	S	S	S	S	?
K2-SUMMIT REC	S	S	S	S	S	S
K2-SUMMIT PLAY	S	S	S	S	S	S
PIPELINE REC	S	S	S	S	N	N
PIPELINE PLAY	S	S	S	S	N	?
PREMIERE PRO	S	N	S	S	S	S

Matriz de compatibilidade de formato de vídeo Fonte: Autores



Fonte: Intel

Nas linhas estão descritos os equipamentos: K2 Standard e K2 Summit que são servidores de vídeo fabricados pela Grass Valley. Pipeline é um network encoder fabricado pela Telestream, e Adobe Premiere PRO é o software utilizado no ambiente de edição. Os K2 Summit estão no ambiente de exibição, enquanto todos os outros equipamentos estão no ambiente de captura. Os equipamentos de captura são utilizados tanto para capturar um sinal de estúdio, como para digitalizar um acervo legado de fitas de vídeo. Nas colunas estão apresentados diferentes formatos, variando *wrapper* e *codec*.

Através desta matriz é possível observar uma gama de formatos compatíveis entre todos os equipamentos.

Os componentes lógicos da infraestrutura

A infraestrutura para implantação de um projeto de MAM depende de vários fatores. Um sistema de MAM é composto por componentes físicos e por componentes lógicos. Os seguintes componentes lógicos essenciais são:

- Banco de dados para armazenamento de informações gerenciais e metadados;
- Aplicação que apresenta todas as informações de forma organizada para facilitar os trabalhos de gerenciamento;
- Serviço de *streaming* da respectiva baixa resolução de cada arquivo de vídeo gerenciado
- Serviço de transcodificação para elaboração da respectiva baixa resolução de cada item gerenciado
- Serviço para movimentação de arquivos

O serviço de transcodificação também é fundamental para compatibilizar os formatos de arquivos entre os diferentes ambientes que foram integrados ao sistema, bem como compatibilizar os formatos antigos com os formatos padrões que estão em uso.

Estes, bem como os serviços de movimentação podem ser designados por agentes do sistema de gerenciamento.

to. Já foram observados sistemas com agentes próprios, sistemas com agentes de terceiros e ainda sistemas híbridos com um misto de agentes próprios e de terceiros.

As informações gerenciais armazenadas em banco de dados são relativas à segurança: quais usuários têm acesso ao acervo gerenciado? Quais operações cada usuário pode executar sobre o acervo gerenciado? As informações gerenciais também são opções de como os arquivos estão armazenados e como cada usuário pode interagir com os mesmos. Aplicações mais modernas têm uma modelagem da infraestrutura relacionando cada item da infraestrutura física. Essa modelagem serve para endereçar o acervo gerenciado e assim ter a sua exata localização. Essas aplicações são capazes de gerenciar um item gerenciado como sendo um conjunto de arquivos. Esses arquivos podem ser cópias do original bem como transcodificações desse original. Todo o conjunto de arquivos é gerenciado como um único ativo digital.

O ativo digital é composto por esse conjunto de arquivos, somado aos seus respectivos metadados. Conforme já foi discutido, o original é o arquivo no formato gerenciado, a princípio com maior resolução possível. Além destes componentes lógicos descritos na forma de serviços, o sistema também pode estar integrado com outros serviços tais como:

- Análise para controle de qualidade
- Análise para extração inteligente de informações
- Extração de quadros chaves (*keyframes*);
- Transcrição de voz para texto;
- Identificação da pessoa que está falando;
- Identificação do tipo de áudio: barulho, música, silêncio, conversa etc.
- Identificação de rosto;
- Análise de imagem com identificação de objetos e descrição de cena;
- OCR

Todos estes serviços têm a finalidade de enriquecer a descrição dos arquivos que estão sendo gerenciados. Este enriquecimento consiste em agregar mais metadados ao respectivo ativo facilitando as pesquisas. Os sistemas de MAM precisam ter um “*engine*” de pesquisa integrado ao mesmo. Estes podem ser nativos do banco de dados ou *engines* mais sofisticados que integram um dicionário controlado.

Existem outros baseados em indexação de arquivos binários, são *engines* mais elaborados e dependendo do acervo gerenciado ele pode ser aplicado. Um exemplo interessante é o *engine* que compara imagens. Ele recebe uma foto e então pesquisa as fotos semelhantes que constam no acervo. Outro exemplo interessante consiste em ter uma amostra da voz de uma pessoa gravada e então pesquisar as locuções da pessoa que fala nessa amostra.

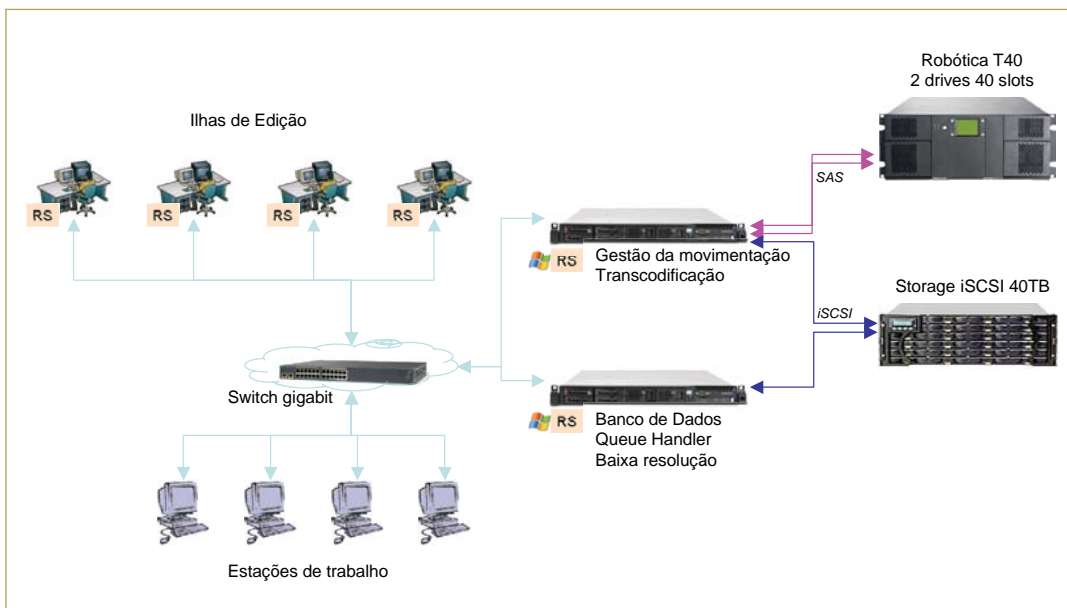


Gráfico 1: Diagrama com arquitetura para sistema com pequenas cargas operacionais
Fonte: Autores

Esses engines que manipulam e indexam os arquivos binários de áudio, foto, vídeo são realidades na indústria de espionagem e é apenas uma questão de tempo para amadurecer um engine que seja atraente para ser integrado com sistemas de gestão de acervos digitais.

Os componentes físicos da infraestrutura

A infraestrutura física é composta de servidores e áreas para armazenamento, classificadas conforme o período de retenção. Tipicamente existem duas áreas, uma área de armazenamento temporária e outra área de armazenamento permanente. Definir o tamanho necessário de cada área depende da carga operacional. Por exemplo, se são produzidas cerca de 12TB de conteúdo anualmente e o conteúdo de até 3 anos de operação

é bastante reutilizado então é interessante que esses três anos de conteúdo estejam armazenados na área temporária, resultando em uma capacidade de 36TB de armazenamento. A área de armazenamento permanente precisa ter uma capacidade de crescimento ilimitada, pois o acervo apenas cresce a cada ano. Após 10 anos de operação o acervo deve estar com 120TB de conteúdo.

Os dispositivos com capacidade de crescimento ilimitado correspondem a sistemas robotizados para realização de arquivamento em fitas magnéticas.

A seguir mostramos como alguns diagramas (Gráfico 1) estão apresentados para descrever a arquitetura dos componentes físicos que compõe uma solução de MAM.

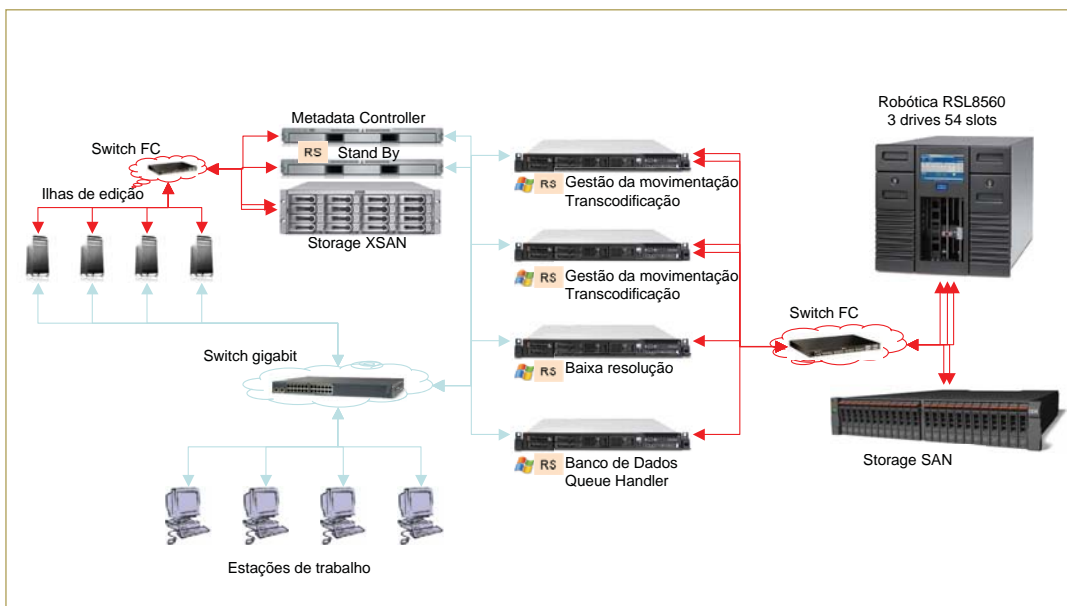


Gráfico 2: Diagrama com arquitetura para sistema com grandes cargas operacionais
Fonte: Autores



O dimensionamento da infraestrutura depende da carga operacional. Para pequenas cargas é possível usar storages NAS com tecnologia Ethernet, storages iSCSI e para grandes cargas é interessante usar storages SAN com tecnologia fiber channel. Naturalmente que para cargas excepcionais será interessante a melhor tecnologia disponível como, por exemplo: tecnologia Infini-Band.

Definição de hierarquias para arquivamento

Conforme pode ser observado nos diagramas da seção anterior, é possível segmentar a arquitetura em três camadas. Na primeira constam os equipamentos para operação propriamente dita, no ambiente de broadcast, são os servidores de captura, as ilhas de edição, os servidores de exibição. Na segunda camada constam os storages para armazenamento temporário e na terceira e última camada consta o sistema robotizado para arquivamento permanente. Desta forma é possível representar a arquitetura simplificada conforme a figura abaixo (Gráfico 3).



Gráfico 3: Esquema representando a infraestrutura de um sistema de MAM Fonte: Autores

Mídias para arquivamento

Dentre as mídias disponíveis, podemos citar a tradicional família de fitas LTO, que se encontra na 6ª

geração com capacidade de armazenamento de 2.5TB. LTO é acrônimo para Linear Tape-Open (LTO), trata-se de uma tecnologia para armazenamento de dados em meio magnético. Foi firmado como padrão aberto por um consórcio de empresas: Seagate, Hewlett-Packard e IBM para competir com os padrões proprietários, DLT, vigentes no início dos anos 2000. A versão lançada inicialmente podia armazenar 100 GB de dados em um único cartucho. A versão mais recente foi lançada em 2012 e pode armazenar até 2.5 TB (Quadro 1).

O tamanho físico do cartucho é padrão e não varia a cada nova geração, ele é sempre o mesmo e é possível observar que conforme a capacidade de armazenamento aumenta, existe um aumento na quantidade de trilhas, um alongamento da fita e diminuição de sua espessura para manter as dimensões físicas do cartucho inalteradas.

Também existem cartuchos para armazenamento ótico desenvolvidos pela Sony. O anúncio deste produto foi feito na NAB de 2012 e agora deve ser lançado a segunda geração, desenvolvido pela Sony em parceria com a Panasonic. A segunda geração tem capacidade de arquivamento de até 3.6TB, velocidade de leitura de 2Gbps e velocidade de gravação de 1Gbps.

Product	Introduced	Capacity (GB)	Write Type
ODC300R	2013	300	Write-Once
ODC300RE	2013	300	Rewritable
ODC600R	2013	600	Write-Once
ODC600RE	2013	600	Rewritable
ODC1200RE	2013	1200	Rewritable
ODC1500R	2013	1500	Write-Once

Quadro 2: Quadro apresentando a capacidade dos discos óticos Fonte: Autores

Além das tecnologias citadas, existe a tecnologia RDX que também pode ser utilizada para arquivamento digital. Trata-se de uma tecnologia baseada em discos magnéticos desenvolvido pela ProStor System em 2004, como uma alternativa para armazenamento em

Propriedades	Geração									
	LTO-1	LTO-2	LTO-3	LTO-4	LTO-5	LTO-6	LTO-7	LTO-8	LTO-9	LTO-10
Data de lançamento	2000	2003	2005	2007	2010	2012	TBA	TBA	TBA	TBA
Capacidade nativa de armazenamento	100 GB	200 GB	400 GB	800 GB	1.5 TB	2.5 TB	6.4 TB	12.8 TB	26 TB	48 TB
Taxa de transferência máxima (MB/s)	20	40	80	120	140	160	315	472	708	1100
Compressão	ALDC	ALDC	ALDC	ALDC	ALDC	LTO-DC				
Capacidade WORM	Não	Não	Sim	Sim	Sim	Sim				
Capacidade de encriptação	Não	Não	Não	Sim	Sim	Sim				
Particionamento	Não	Não	Não	Não	Não	Não				
Espessura da fita	8,9 µm	8,9 µm	8,0 µm	6,6 µm	6,4 µm	6,1 µm				
Comprimento da fita	609 m	609 m	680 m	820 m	846 m	846 m				
Trilhas	384	512	704	896	1280	2176				
Elementos de escrita	8	8	16	16	16	16				
Voltas por banda	12	16	11	14	20	34				
Densidade linear (bits/mm)	4880	7398	9638	13250	15142	15143				
Codificação	RLL	PRML	PRML	PRML						
Memória integrada	4 KB	4 KB	4 KB	8 KB	8 KB	16 KB				

Quadro 1: Quadro apresentando a evolução das fitas da família LTO Fonte: Autores



cartucho	capacidade unitária		custo unitário	retenção anos	qtd cartuchos 1000hs		custo 1000hs
LTO-6	2.5 TB	US\$	41.00	30	9 cartuchos	US\$	369.00
RDX	2.0 TB	US\$	381.71	30	11 cartuchos	US\$	4198.81
ODA	1.5 TB	US\$	146.00	50	15 cartuchos	US\$	2190.00

Quadro 3 apresentando a capacidade das mídias para arquivamento e o custo para arquivar 1000hs de conteúdo @50Mbps (21,97TB) Fonte: Autores

fitas magnéticas. Os cartuchos RDX têm 2,5 polegadas, são a prova de choque, contra quedas de até 1 metro de altura, e oferecem uma capacidade de arquivamento para até 30 anos (coercitividade magnética). Existem cartuchos com diferentes capacidades chegando a até 2TB de capacidade.

O quadro apresenta o cartucho de maior capacidade para cada tecnologia estudada. Na última coluna o quadro apresenta o custo estimado para armazenar 1.000hs de vídeo em cada tecnologia. Para LTO-6 são necessários 9 cartuchos, para RDX são necessários 11 cartuchos e para Optical Disc Archive são necessários 15 cartuchos.

O custo total é menor para a tecnologia LTO. O custo é um parâmetro importante, mas o tempo de retenção também é bastante relevante. Uma retenção menor significa necessidade de migrações de dados mais frequentes. Antes de atingir 30 anos, conforme indicado nas tecnologias LTO e RDX, é necessário fazer uma cópia para outro cartucho. Esse procedimento aumenta o custo operacional.

No caso da família LTO ainda tem a questão dos drives, são os dispositivos que fazem gravação e leitura

das fitas LTO. Os drives acompanham a evolução das gerações e assim existem drives para todas as gerações. Os LTO têm compatibilidade de gravação com a geração de referência e a geração anterior, e compatibilidade de leitura com até duas gerações anteriores. Por exemplo: um drive LTO-6 pode gravar cartuchos LTO-6 e LTO-5 e pode ler cartuchos LTO-6, LTO-5 e LTO-4. Por esta razão trocar a geração de cartuchos LTO não é um trabalho muito simples, após 3 gerações será necessário migrar todo o conteúdo.

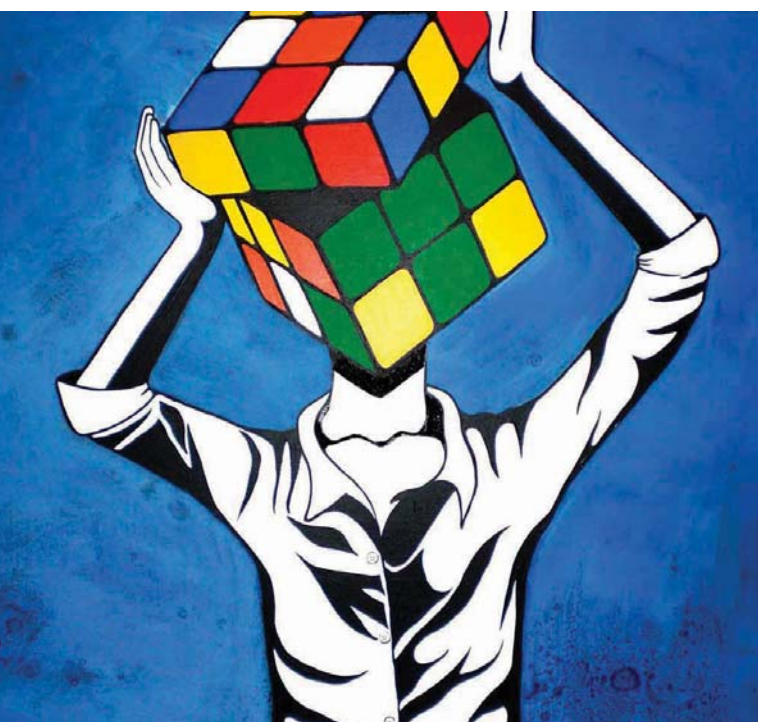
Juntamente com o lançamento da geração LTO-5 foi apresentado o padrão LTFS, desenvolvido pela IBM, para compatibilizar o acesso aos dados gravados nas fitas independentemente do software que fez a formatação das mesmas. LTFS é um acrônimo de *Linear Tape File System* que estabelece um padrão para gravação de arquivos em fitas LTO. Fitas gravadas neste padrão são capazes de expor os arquivos para o sistema operacional sem a necessidade de complexos sistemas de software.

Uma característica importante na tecnologia LTO consiste na incorporação de algoritmos de compressão e descompressão (sem perda de dados) por parte dos drives nos processos de gravação e leitura de dados das fitas. Apesar de aparentemente aumentar o tempo de processamento devido à aplicação desses algoritmos, tempo necessário para que o algoritmo possa executar o seu trabalho, observa-se que o tempo para gravar e ler um arquivo da fita diminui com essa compactação, pois menos dados são efetivamente tratados. Considerando arquivos de vídeo com codificação MPEG-2 (*iframe only*) é observado uma compactação média de 17%, é simples entender que os arquivos devem ser gravados 17% mais rapidamente do que se não estivesse compactado.

Cada tecnologia tem vantagens e desvantagens para gestão de grandes acervos, o uso misto de todas as tecnologias é bastante interessante para aproveitar o melhor de cada uma. Considerando acervos de tamanho médio é interessante explorar ao máximo uma determinada tecnologia.

Outras mídias

Já foi observado uso de outras mídias para formação de acervos digitais. Dentre essas mídias existem: CDs, DVDs, Blue Ray, discos externos, etc. Essas outras mí-



Fonte: Banco de Dados



cartucho	capacidade unitária		custo unitário	qtd cartuchos 1000hs		custo 1000hs
CD	700 MB	US\$	0.20	32.911 cartuchos	US\$	6582.20
DVD	8,7GB	US\$	1.50	2.586 cartuchos	US\$	3879.00
BLU-RAY	25 GB	US\$	4.00	900 cartuchos	US\$	3600.00

Quadro 4 apresentando a capacidade das outras mídias para arquivamento e o custo para arquivar 1000hs de conteúdo @50Mbps (21,97TB) Fonte: Autores

dias são apropriadas para uso doméstico e pessoal. No entanto, para grandes acervos esse uso pode ser desastroso. Por exemplo, o uso de HDs externos não oferece um armazenamento 100% seguro, pois basta um choque mecânico para que o HD fique inutilizado e seu conteúdo inacessível.

É sempre importante realizar cópias de segurança, mesmo para as mídias apresentadas no capítulo anterior que são as mais robustas, mas também não oferecem 100% de garantia. É importante realizar cópias de segurança, o *backup*.

Apenas para estudo é apresentado um novo quadro com as outras mídias (Quadro 4).

O custo para arquivamento de 1.000hs de vídeo é relativamente superior àquele observado com as mídias mais robustas. Só este fato já deveria desestimular o uso destas outras mídias. Outro aspecto importante é que um vídeo com duração de 1hora deve ocupar cerca de 20GB e assim não cabe completamente em uma única mídia. A quantidade de mídias é bastante superior, o que representa um espaço físico maior para acomodar estas mídias.

Outro aspecto importante é justamente a durabilidade, vida útil da mídia. No caso dos CDs, foi apresentado o preço de um CD padrão com alumínio. Um CD adequado para arquivamento digital é o dourado e seu custo unitário é superior à US\$ 1,50 demonstraram ser um arquivamento bastante dispendioso. Muitos estudos feitos com CDs e DVDs demonstrando não serem adequados para formação de acervos digitais devido a problemas na constituição física: camada de plástico protetora fina, aplicação de canetas e etiquetas decalques que

corroem o material das mídias dentre outros fatores. Os discos Blu-ray têm um processo de fabricação que lhes prove uma maior robustez, a camada de proteção do blu-ray é feito em resina acrílica e assim mais rígida que a proteção de plástico dos CDs e DVDs.

Ainda neste tópico de outras mídias convém comentar sobre o uso de discos externos para arquivamento digital. É bastante comum observar empresas formando acervos digitais baseados em discos externos. Essa prática não é recomendável, pois os HDs externos não são dispositivos robustos para formação de acervos digitais. Uma queda pode danificar permanentemente o dispositivo e impossibilitar a recuperação do conteúdo armazenado. Produzir o mesmo conteúdo novamente é inviável e se possível, um trabalho que pode ser evitado usando dispositivos apropriados. Justamente para atender a este perfil de uso que os dispositivos RDX foram desenvolvidos, pois são efetivamente discos externos mais robustos. ■

Continuará...



Fabio de Sales Guerra Tsuzuki é engenheiro eletrônico e mestre em engenharia pela Escola Politécnica da Universidade de São Paulo, doutor em engenharia pela Universidade de Tóquio. Sócio fundador da Media Portal Soluções Ltda. Contato: fabio.tsuzuki@mediaportal.com.br



José Olairson Valentim é engenheiro da computação graduado pela Universidade Unisantana de São Paulo, com experiência em emissoras de Televisão nas áreas de Treinamento, Projeto e Implementação de sistemas de TV. Especialista em Media Asset Management. Membro da SET desde 1999, tendo participado também dos grupos de trabalho do Fórum Brasileiro de TV Digital. Gerente de Projetos da Videodata, na Divisão de Arquitetura Digital. Contato: olairson@gmail.com



Fonte: LG Consulting