Meeting Exa-Scale Archive Requirements In The Digital Age Of Media and Entertainment

Tom Coughlin
Coughlin Associates

The Storage Dilemma In The Media And Entertainment Industry

Digital technology is revolutionizing how the media and entertainment industry operates. Whether it is in broadcasting, motion pictures, music, publishing or video production, the media and entertainment industry is both reaping the benefits of, and struggling with the new challenges brought on by the digital revolution. The transition from analog to digital workflows in the media and entertainment industry has occurred rapidly and in a non-linear fashion, so much so that these digital advancements are creating other, unforeseen and unplanned for challenges and issues.

A great example of an important challenge with digital content is the issue of how to digitize older analog content and preserve all the new, ever larger, digital content that is being created at a blistering pace. The motion picture industry has a long standing commitment to archiving and preserving motion pictures for a minimum of 100 years and some industry stalwarts will settle for nothing less than "forever". With today's longer shooting times, higher frame rates, more bits per pixel, higher resolution and an increasing number of content capture devices used for a motion picture project combined with an increasing number of digital films produced annually, the size and number of individual digital files is getting larger and larger.

For example, the latest RED digital camera technology can generate over 85TB of digital video content during a 24-hour shoot with a single camera. Storing this "daily" content would require a minimum of twenty-one 4 TB hard disk drives or ten 8.5 TB digital tape cartridges. In another example, the production of the movie Avatar in 2009 resulted in the creation of over a petabyte (PB) of digital content, the vast majority of which has been retained and archived for possible re-use or for content preservation reasons. Within 10 years the total content created for a single high-end movie project could approach 1 Exabyte (1,000 PB), making the management and preservation of valuable media content a challenge that the industry will struggle with for many years to come.

Across the Media and Entertainment industry, re-use of digital content is integral to the monetization process. Therefore, protecting, archiving and providing access to digital content, at an affordable cost, requires continued improvements in content management, digital storage and industry best practices. The critical tensions between cost, access and long-term preservation present a significant challenge for the archiving of modern digital content.

As for the required digital storage hardware, in the last few years digital tape capacities have increased faster than hard disk drive (HDD) capacities. This has resulted in ever lowering costs for storing content on digital tape versus the cost of storing content on HDDs (even more so when operating costs are taken into account). Due to the significant overall cost advantages of digital tape, it continues to play an important role versus more expensive HDDs or flash memory in long-term storage of media assets.

Digital tape has proven to be an effective medium for both long-term media and entertainment content archiving as well as for various production workflow applications. In the world of digital tape technology, there are three supported formats in use today. Two of these digital tape formats are deemed "enterprise class", one from IBM and one from Oracle. The third format is the midrange Linear Tape Open (LTO) format. **Table 1** provides a quick comparison of these digital tape formats.

Format	Class	Raw Capacity Per Tape Cartridge	Hours Of 4K Content Per Tape Cartridge	Time To Transfer 1 Hour Of 4K Content
LTO-6	Midrange	2.5 TB	1.4 hrs.	3.0 hrs.
IBM TS1140	Enterprise	4.0 TB	2.3 hrs.	1.9 hrs.
Oracle StorageTek T10000D	Enterprise	8.5 TB	4.9 hrs.	1.9 hrs.

Table 1. Digital Tape Formats

Lowering the Costs of Digital Conversion and Archiving

In addition to the challenges of supporting digital workflows and archiving new content, there is a vast amount of analog content on film and videotape that is yet to be converted to modern digital formats. This "preservation" effort creates its own set of unique challenges including determining which content to preserve (not all existing content needs or deserves to be preserved), what format(s) to preserve it in, how to assign metadata and

how to keep the costs of conversion under control. As for the latter, new technology and services are making the conversion of the most common, older formats more affordable. And as time goes on, and as the capacity of digital storage devices and media increase, the total costs for storing a given amount of content is decreasing.

For example, in 2010 a vendor specializing in digital conversion projects[1] provided a quote for converting a 10,000 hour mixed analog video tape archive into a digital archive, including media asset management (MAM) control for $958,569. By updating only the bulk ingest components, with the same SAMMA solos as used in the original quote for the digital conversion, the costs could be reduced in 2012 to $825,490. By using two, 4-channel Cinedeck MX recorders in place of six 1-channel SAMMA Solos this cost could be further reduced to $717,668. That is about a $240,000 cost reduction over a 2-year period (about 25% lower cost)! In 2013 this cost would probably decline by an additional 15%.

…new services and equipment are making the conversion of the most common, older formats less expensive…

These cost reductions are a consequence of using improved equipment that increases throughput and reduces the amount of manual media handling. This example shows how the cost reductions occurring in content conversion make preservation of historical content more affordable and thus more likely. As a consequence of lower cost conversion, the digitization of older content will swell the total amount of digital content in modern media archives.

Likewise, increases in the capacity of digital storage devices, especially HDDs and digital tape, are resulting in significant cost savings for archiving and preserving digital content. Current projections for the cost of storing 1 PB of content over a 20 year period beginning in 2013 (and assuming a hardware refresh every 5 years and regular capacity increases for HDDs and digital tape) show a cost of approximately $1.49M for HDD storage as compared to about $468,000 for tape storage (using Oracle's StorageTek T10000D tape technology). Thus we estimate a 3.2x savings for storing 1 PB of content for 20 years using StorageTek T10000D tape versus using legacy HDD technology.

..we estimate a savings of about 3.2x for storing 1 PB of content for 20 years using StorageTek T10000D digital tape

Tiered Storage Combines Flash, Disk, and Tape to Drive Archive Efficiency

The increase in the size of digital content over time will require archive storage systems that can scale to accommodate that growth. By scale, we mean that the storage system must not only be able to store content economically for a long period of time, but also

[1] Nicholas W. Lim, Chosun Group, Inc., www.chosungroup.com

support required access requirements, such as latency from a request to delivery of a required video file. While the answer to the former is modern digitaltape storage, the answer to the latter is a combination of digital tape storage AND disk or flash storage. *Tiered storage* refers to an archive storage environment in which digital content or media files can be stored on different types of storage media—such as high performance flash storage, capacity optimized disk storage, and highly economical digital tape storage—and moved automatically between those different platforms to optimize for cost, performance, access, and protection (see **Figure 1**).

Tiered storage is common in digital media archiving because it can deliver the following benefits:

- Fast access to digital content by staging active content on flash or disk storage before migrating to highly economical digital tape storage for long-term retention.
- Reduced storage infrastructure costs by balancing cost considerations with performance and scalability needs, leading to lower hardware and software acquisition costs as well as improved total cost of ownership (TCO) of the storage infrastructure for digital media archiving.
- Decreased operational cost by reducing manual intervention via automated, policy-based content placement and migration across the tiers of digital storage.
- Investment protection through a highly scalable architecture across both capacity (hours of digital content stored) and performance (access time and streaming rate of digital content) that supports multiple generations of backward compatibility.

For example, Oracle, has tiered storage solutions for digital media archiving that are deployed in large broadcast and media outlet operations. Oracle's tiered storage solutions scale both the capacity and performance of media archives, while minimizing the cost per minute or per

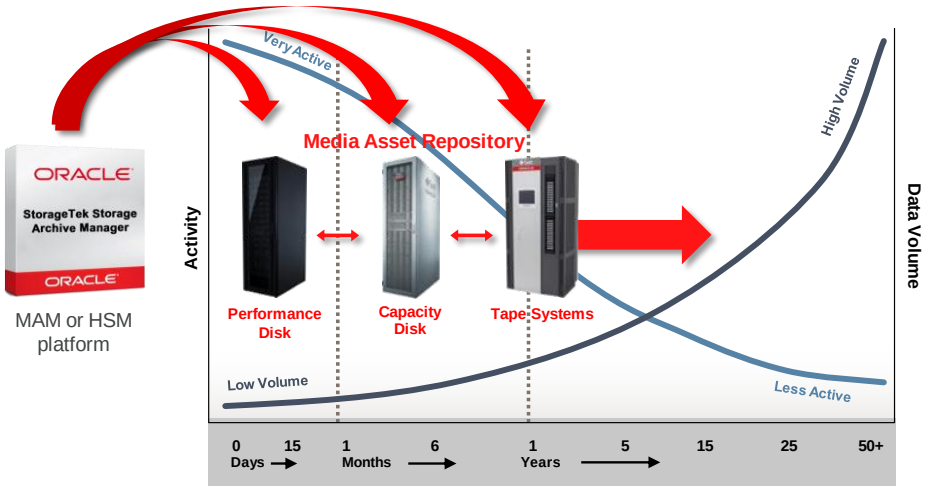

Figure 1 - Tiered storage can increase efficiency, boost performance, and streamline media archives.

For example, Oracle, a leading provider of both disk and tape storage solutions—and the #1 enterprise database, middleware, and application vendor—has tiered storage solutions for digital media archiving that are deployed in large broadcast and media outlet operations. Oracle's tiered storage solutions scale both the capacity and performance of media archives, while automating data movement and minimizing the cost per minute or per hour of digital content stored.

Integrating The Archive Into A Digital Workflow

Digital tape is used in many ways in modern digital media workflows: content capture, content transport, and content libraries as well as in genuine, long-term media asset archives. Today, many of these processes involve considerable automation while others still require some manual intervention.

For example, during movie and large video projects it is common to bond and insure the production to protect the investors. This bonding requires the production company to create more than one copy of captured content either at the time of initial capture or immediately thereafter. To fulfill this requirement many production companies use devices in the field to which captured content can be downloaded, stored and then transferred to digital tape for aggregation, backup and transport. The digital tape then serves as the copy to meet bonding and insurance requirements and provides a mechanism for transport and ingest into the greater digital workflow process and, if required, the long-term archive.

In large, more sophisticated, digital productions, workflows become complex and may require additional layers of integration to support numerous production stages, various transcoding requirements and multiple points of data protection. Increasingly, content libraries or "working archives" in modern workflows are becoming large, actively managed content repositories. Consequently, whether large or small, content workflows will want to use digital storage with "drag and drop" capabilities that are standards-based, self-describing and operating system aware. These capabilities provide a simple and efficient means of data interchange between the "working archive" and the various stages of the digital workflow.

The Age of LTFS

A continuous stream of innovation is making digital tape storage an increasingly critical tool for the media and entertainment sector. Of particular importance, Linear Tape File System (LTFS) is a new development in digital tape storage that is undergoing rapid adoption in the motion picture and television broadcast industries. As a relatively new

technology, LTFS is currently moving through the standardization process. The co-chairs of the SNIA working group chartered with developing the specification standard are IBM and Oracle. LTFS is currently supported by midrange LTO and the enterprise-class StorageTek T10000 digital tape drives and IBM TS1140 digital tape drives.

LTFS creates tape cartridge partitions. Partition 0 can contain indexing metadata, including directory tree structures that make accessing data on a digital tape similar to finding files on a HDD. The directory structure in Partition 0 allows swift access to the desired data on a mounted tape. The LTFS specification is open source, reducing the barriers to implementation and enabling users to use digital tape storage from multiple vendors as part of a rapidly accessible archive.

StorageTek LTFS, Library Edition extends LTFS capabilities across an entire automated tape library. This allows aggregation of content metadata from all the tapes in the library. LTFS Library Edition allows highly effective integration of a digital tape library into a hierarchical storage system to provide rapid access to archived content.

LTFS is extremely useful for media and entertainment applications. Increased data mobility using open file formats makes sharing across workflows much simpler, more intuitive and easier. In addition, LTFS brings the familiar drag-and-drop procedure to files on digital tape making a tape's directory information viewable, accessed and managed without special applications, such as costly third party backup software. Files are displayed similar to how they appear on disk or flash storage and can be dragged and dropped from disk to tape (or vice-versa), so the use of digital tape in workflows becomes more transparent.

LTFS, Library Edition takes the LTFS capability to the tape library level. StorageTek LTFS, Library Edition software manages the file index of every tape cartridge in an enterprise tape library. LTFS LE further broadens the appeal of digital tape in media and entertainment, enabling customers to take advantage of digital tape's low cost per terabyte for backup, content archiving and preservation projects. .

There are new developments for the creation of media content archives in the cloud using digital tape storage. With the introduction of the LTFS file system, many storage vendors think it will be easier to store content in the cloud using digital tape. The self-describing file system characteristics of LTFS combined with a RESTful API for digital tape storage will enable Archive-as-a-Service providers to more easily and efficiently take advantage of the economics of digital tape storage as the infrastructure foundation for their public or private cloud offerings.

LTFS and LTFS LE have received a lot of interest as innovations that make digital tape technology a key component for low cost, highly reliable media content archiving. However, it can also be used for data transfers within a workflow and as a third tier of storage in a mixed media (disk and digital tape), tiered architecture for content libraries. With LTFS, digital tape can be used cross-platform and in various parts of the workflow. Digital tape storage can replace pro-video tape in a modernized workflow and it is less

expensive than disk-based storage both per GB and in terms of operating expenses. For example, Fox Studios uses digital tape for archiving all their high definition content to reduce cost and increase the density of their digital archive

Long-Term Data Retention and Integrity

As higher definition formats take over the industry, the size of digital files is increasing rapidly. Consequently, storing and transferring that content is becoming more difficult. Storage devices with faster data rates and new, low latency, high bandwidth transports are required. For this reason, workflow architects and digital archivists must move to new (and faster) storage technologies over time. If the capabilities of the digital infrastructure supporting the workflow (and its integrated archive) don't keep up with the increasing demands of new digital content, a breaking point will occur. Thus, with the continual growth in the size of content data files, existing digital workflow and archiving systems can quickly become obsolete.

The StorageTek T10000D tape drive has backward read compatibility for three generations, to the T10000A. This gives the end user an extended period of time to migrate from prior generations to the latest generation tape. This extended backward read compatibility be very attractive to an archivist whose budgets may vary from year to year.

In the world of digital storage, tape storage provides, perhaps, the best protection against this obsolescence issue. New generations of digital tape storage, with significant improvements in both capacity and speed are introduced approximately every 2-3 years. This cycle of innovation in digital tape technology provides a strong hedge against the transition to higher definition formats by the media and entertainment industry. Furthermore, the tape storage industry is very cognizant of the need for each new generation of its technology to be able to read the content stored on previous generation digital tape cartridges. This "backward read" compatibility enables workflow architects and archivists to take advantage of the latest advancements in digital tape storage while still being able to access older, archived content for re-use and incremental monetization.

Another important characteristic of digital tape compared to other digital storage technologies is its inherent "remove-ability". Because files are stored on a removable tape cartridge, content on digital tape can easily be placed in long term "vaulting" facilities. Though not necessarily on par with legacy film-based content, under proper storage and handling conditions, up to 30 years of media shelf life can be expected from digital tape.

Many casual observers may not recognize that digital tape is likely the most durable and reliable digital storage technology available today. For example, whereas disk storage systems typically need to be refreshed every 3-5 years, digital tape storage refresh cycles can be extended to 8-10 years. This allows older generations of digital tape storage technology to reside side-by-side with new generations of the technology in large automated digital tape systems. The 8-10 year useful life of tape drive technology along with the multi-generational support of automated tape systems provide customers long-term protection of initial investments in digital tape technology.

While LTO tape drives offer this data integrity checking, there is a large price to pay in terms of network and server bandwidth. LTO tape drives require that all of the preservation data is recalled from tape, and then sent across the network to a server that performs the data integrity checking. With Oracle's StorageTek Library Media Validation feature, data integrity checking is performed in the tape library within the tape drive, eliminating the need to consume valuable network and server bandwidth.

A critical requirement of many digital tape storage systems used for long-term data preservation is the capability to periodically perform data integrity checking. This capability provides reassurance that the integrity of the archived content has not degraded over time and it provides visibility into issues that facilitate preventative action before an issue becomes a major problem. For example, with Oracle's Library Media Validation feature, data integrity checking is performed by tape drives within the automated tape library, thus eliminating the need to consume valuable network and server bandwidth. Oracle offers multiple options for initiating Library Media Validation. One option is to set policies within Oracle's StorageTek Tape Analytics tool. In addition to continuously monitoring the health of every tape drive and every cartridge in the tape storage environment, it also automates the process of data integrity checking.

For these reasons (and more), automated digital tape library systems are the perfect storage target for hosting fast growing, large scale, modern media repositories.

Traditional Archives Are Becoming Managed Content Repositories

The media and entertainment industry supports an increasing number of use cases and workloads requiring a diverse array of workflow architectures to support them. For purely archival purposes, content can be off-line, as when digital tapes (or legacy film canisters) are placed on a shelf in a controlled environment, or near-line on a networked tape library or disk storage array where they can be accessed more rapidly. As mentioned earlier, near-line content repositories are more often being designed with multiple tiers of digital storage including a disk tier and a tape tier, and perhaps even flash memory, to provide optimal trade-offs between performance and cost. In many of these digital

workflows it is most cost effective to employ these "tiered storage" architectures so as to create effective, fully integrated media content repositories.

By strict definition, media archives do not require high performance, "on-line" access. However, modern content archives, whether directly integrated into the workflow or not, are becoming more active in order to support asset re-use and the desire to further monetize existing content. Content re-use and higher degrees of content monetization have become important factors in the increasingly competitive media and entertainment sector. Let's look at three examples of how content re-use and monetization are driving the need for modern managed content repositories.

Sonuma

Sonuma, the Society for the Digitalization and Commercialization of Audiovisual Archives, was created in 2009 by the French Community of Belgium (located in the Walloon region), and the Radio Télévision Belge Francophone (RTBF) (the national television broadcasting company). As a subsidiary of RTBF, with the goal of building an infrastructure dedicated to the preservation, digitization, and commercialization of RTBF's audiovisual archives, Sonuma became the owner of all RTBF's audiovisual archives. This meant that Sonuma needed to implement a scalable storage solution that could store a vast and steadily growing archive of digitized media content.

Sonuma was looking for a robust storage solution that would enable it to store its digital archive that includes 120,000 hours of content from a variety of media, as well as support future digital archive growth. The solution also needed to support the organization's goal of preserving, managing, and commercializing this content. With Oracle's StorageTek SL3000 modular library system, Sonuma has digitized thousands of hours of content while providing scalable storage space for future additions to the library. Furthermore, with the help of Oracle partner Netia, the organization created a Web portal to facilitate journalist access to this content.

Using these solutions the society has realized a highly automated workflow, from receipt of digital content through delivery to specific journalists and the general public. This has simplified the task of preserving, managing and accessing content within its rapidly growing digital library and provides a storage solution through which large groups of journalists in many locations can easily and quickly search low and high resolution content.

USC Shoah Foundation

The USC Shoah Foundation Institute was established by filmmaker and USC trustee Steven Spielberg. Its purpose is to overcome prejudice, intolerance and bigotry and the suffering they cause, through the educational use of more than 100,000 hours of visual

testimonies. The Institute uses Oracle's StorageTek T10000C tape drives to support the cataloguing and preservation of visual testimonies from Holocaust survivors and witnesses. The Institute's holographic preservation project began in 2008 with Oracle's StorageTek SL8500 Modular Library System and StorageTek T10000B tape drives. The Oracle storage solution, which also includes Oracle's Sun x86 servers with Oracle Solaris and Oracle disk arrays, digitizes about 80 TB of content per month.

The Institute upgrades its tape technology and migrates its data on tape every three years to ensure the preservation of these testimonials on the best possible technology. With the StorageTek T10000C drives maximum storage capacity under a single point of control is 55 PB and the operating costs are 35% less than with the prior generation tape technology. Using the StorageTek T10000C tape drives in conjunction with Oracle's StorageTek Storage Archive Manager software and Oracle Solaris, the Institute is able to checksum its files any time they are moved on the network to ensure data integrity.

Thanks to the additional storage capacity the Institute has begun to digitize and preserve testimonies from the Armenian and Rwandan genocides.

T3Media

T3Media offers cloud-based storage, access, and licensing for enterprise-scale video libraries. Its' technology platform and services enable media owners to generate new value from their content while managing cost and complexity. Through www.t3licensing.com and its global sales force, the company licenses sports, news, and creative footage to producers in advertising, entertainment, publishing, and emerging media.

T3Media uploads tens-of-thousands-of hours of programming content and more than 100,000 media assets per year, it was looking to deploy a cloud-based, digital storage environment that was cost-effective and highly accessible. In addition, it needed to ensure its storage environment could scale to meet clients' future needs. Because many T3Media clients are in the media and entertainment industry—with digital assets that have more than a 30-year life span—the company also needed to ensure short- and long-term integrity for the content it is trusted to store.

After reviewing many vendors, T3Media deployed a storage environment built on Oracle's StorageTek SL8500 modular library system, StorageTek T10000 tape drives, Sun Oracle disk storage, and Sun Storage Archive Manager. The primary storage site alone contains more than five petabytes of T3Media's customer content. With the new storage environment, T3Media acquired the ability to add storage incrementally, on a per-customer basis, and efficiently save content on tape for pennies on the dollar, compared to "disk only" storage.

T3Media also benefited from exceptional performance. With Oracle's storage solutions, the company realized a three- to-five-times performance improvement, compared to its legacy "disk only" storage environment.

It's a Wrap: The Resurgence of Tape

As video resolution and frame rates increase; stereoscopic and even "free viewpoint" projects multiply and as the total hours of captured digital content increase the storage capacity needed for production and archiving becomes staggering. A calculation shows that 16K x 8K pixel resolution, 24 bits/pixel color, 300 fps raw video content could require 115 GB/s data rates and generate 414 TB/hour of content. If this were full stereoscopic capture then these requirements would double. If 4 cameras were used to create content for a "free viewpoint" presentation, the raw data would be 1.66 PB for an hour of content. Within a few years' time movie production stored content for a single project could approach 1 exabyte (1,000 PB).

The need to store this amount of content not only for long-term preservation but also for incremental monetization drives the requirement for cost effective archival storage solutions. Tape has demonstrated its archival storage value over the years as a reliable, long-life medium, superior to disk for this use case. Recent innovations such as Oracle's StorageTek T10000D tape drive, which delivered world record capacity and throughput, and LTFS, Library Edition, a leap ahead in usability, have made tape an even more attractive option for the media and entertainment industry. These innovations have led to the resurgence of tape while continued evolution of tape technology promises to keep it as the medium of choice for the future exabyte scale archives.

Appendix

Feature movie resolutions are increasing in general and in particular for the very high end of the market. 2K resolution movies are very common in some parts of the world, but as production costs go down and as the number of cinemas (and televisions) capable of using 4K content increases, 4K resolution video is becoming more common. All high-end digital video cameras now support 4K or even higher resolution. For some sequences involving fast motion or low light 6K and even 8K resolutions are sometimes used for the creation of video content. For example, the Peter Jackson movie, *The Hobbit*, released in late 2012 was filmed at 48 frames per second (fps), in 3D, using RED RAW format with RED 5K Epic cameras. James Cameron's follow-on Avatar film is being shot at 60 fps. Note that most professional movie production is commonly done at 24 fps. Film frame rates as high as 120 fps are supported in current high-end video cameras and to create the ultimate in slow motion special effects much higher frame rates (1,000 and even 3,000 fps) are appearing in some specialized cameras.

Some directors want digital video resolutions as high as 16K x 8K pixels to create resolutions as high as IMAX film. In addition, KDDI in Japan and the Free-viewpoint Immersive Networked Experience (FINE) project in Europe have been shooting concerts and sports events with simultaneous input from four to 30 video cameras, with up to 4K resolution.

The images from the multiple cameras are combined to create what is being called a "free viewpoint" video where a viewer can look at scenes and people in the field of view from any perspective. Of course the raw video footage from the multiple cameras requires enormous amounts of digital storage and bandwidth to move it around.

Still image resolution increases often precede moving image developments. Gigapixel image projects created from many smaller resolution images have become more common. Arrays of still and even moving image cameras that work together have been demonstrated at the National Association of Broadcasters (NAB) conference and other industry events. Duke University demonstrated its Aware-2 for still and moving image capture[2].

The Duke team installed 96 small cameras, each with a 14 megapixel sensor, on the outside of a small sphere about the size of a football. A computer connected to the sphere stitches the images together to create a composite whole. 50 gigapixel color images should be possible with this technology (note that 8K x 4K video is about 33 megapixels). Although it takes about 18 seconds to shoot a single frame today, as computer processor speeds increase and the speed of the cameras improves, gigapixel video images could be possible in professional video production by the next decade.

Figure 2 shows the hours of content shot for an hour of completed work[1]. In 2013 about 68% of the survey participants said that they captured 6 hours or more of original content for an hour of completed work. 7% said they captured more than 51 hours of video for an hour of final content. Based upon prior Coughlin Associates surveys it appears the number of hours captured for an hour of content is increasing.

[2] M. Grotticelli, New Gigapixel Camera Can Zoom In After An Image is Captured, Beyond the Headlines, Broadcast Engineering, June 28, 2012

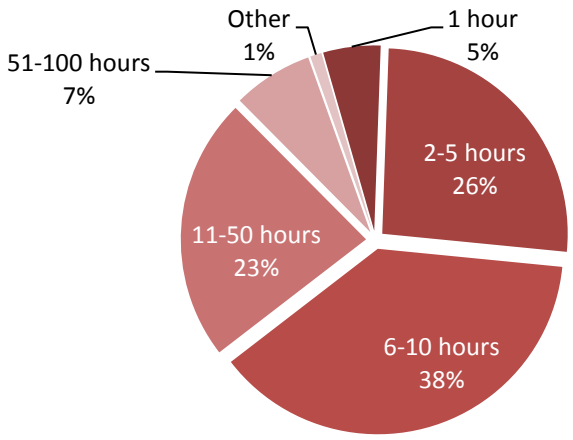

Figure 2. Content Shot for an Hour of Completed Work

About the Author

Tom Coughlin, President, Coughlin Associates is a widely respected storage analyst and consultant. He has over 30 years in the data storage industry with multiple engineering and management positions at high profile companies.

Dr. Coughlin has many publications and six patents to his credit. Tom is also the author of Digital Storage in Consumer Electronics: The Essential Guide, which was published by Newnes Press. Coughlin Associates provides market and technology analysis (including reports on several digital storage technologies and applications and a newsletter) as well as Data Storage Technical Consulting services. Tom publishes the *Digital Storage Technology Newsletter, the Media and Entertainment Storage Report*, *and the Capital Equipment and Technology for the Hard Disk Drive Industry Report.* **He also publishes a blog on digital storage at Forbes.com.**

Tom is active with SMPTE, SNIA, the IEEE Magnetics Society, IEEE CE Society, and other professional organizations. He is in charge of Future Directions for the IEEE Consumer Electronics Society as well as Director Elect for IEEE Region 6. Tom is the founder and organizer of the Annual Storage Visions Conference (www.storagevisions.com), a partner to the International Consumer Electronics Show, as well as the Creative Storage Conference (www.creativestorage.org). He is the general chairman of the annual Flash Memory Summit. He is a senior member of the IEEE, leader in the Gerson Lehrman Group Councils of Advisors and a member of the Consultants Network of Silicon Valley (CNSV). For more information on Tom Coughlin and his publications go to www.tomcoughlin.com.