Achieving A Faster, More Scalable Media Active Archive

Coughlin Associates

Introduction

Active archives are a critical part of the media workflow. Once a resting place for assets that might be needed "someday," the archive is now a profit center as content is needed for on-going monetization through reuse and redistribution. This has led to frequent use of Active Archives that act like a content library, rather than traditional passive archives. Furthermore, with media files increasing in number and size, storage capacities are increasing as well. This double impact — more frequent archive access and increasing capacity requirements — has created a compelling demand for new archive solutions.

Increasing Capacity Demands

Exponential capacity growth results from the transition from analog video content to digital content, which has lowered the overall cost of video production, from capture to production and delivery. This has encouraged video producers to produce more content and to use the rapid development of digital video technology to provide a more immersive video experience to their customers. Two simultaneous changes — more data from each media source, and more sources employed in each production — now combine to increase the capacity required per hour of finished product.

14X More Capacity per Hour of Raw Content

Increased data from each source is the obvious byproduct of greater resolution: 4K is here now, and soon 8K content will become common in many video workflows. Note that 8K content, whether UHD-2 or Digital Cinema format, produces significantly larger files than 4K or 2K (HD).

Figure 1: Capacity Required Per Media Source — New formats will require up to 70TB per hour of raw content, a 14X increase over today's demand.

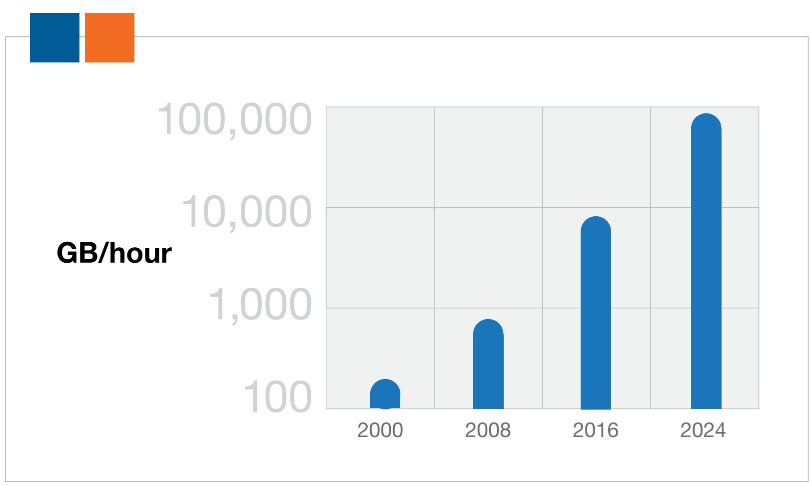


Figure 1 shows how the size of raw captured professional content could grow with time.

Beyond 4K and 8K content, video resolutions as high as 16K X 8K pixels are sought by some content producers to create experiences like those in an IMAX film or to provide an immersive 360-degree composite video for virtual reality.

Video frame rates have also increased to avoid motion artifacts that can occur with higher resolution. 120 frames per second are likely to become common for 8K resolution, and slo-motion camera with thousands of frames per second are often used for special effects. Dynamic range and color depth are also increasing to create richer more life-like content. Over the next seven years, this translates to an expected 14X increase in capacity required per hour of raw content.

2X More Hours of Content Per Hour of Completed Work

Compounding this capacity growth, producers now employ more media sources. 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 up to 30 video cameras, with up to 4K resolution. The video from the multiple cameras are combined together to create what they call a “free viewpoint” video where a viewer can look at scenes and people in the field of view from any perspective.

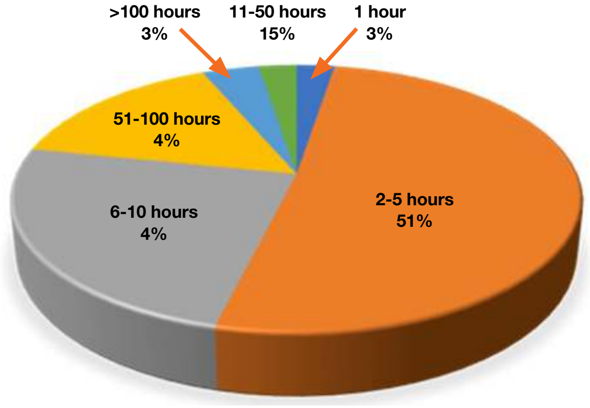


Figure 2: Content Shot Per Hour of Completed Work — Each hour of completed work now requires more raw content than in the past, a trend that is expected to continue.

In addition to using more cameras for a video event, media and entertainment professionals are also shooting more hours with each camera. In a 2017 survey of media and entertainment professionals we found that 46% of the survey participants said that they captured 6 hours or more of original content for an hour of completed work . The full survey result is shown in the pie chart below. A continued push towards more immersive experiences will further drive production, resulting in a doubling in hours of raw content per hour of completed work over the next seven years.

28X More Capacity Required per Hour of Finished Product

Together, the combined impact of more resolution and more cameras is expected to create an exponential increase in capacity requirements during the coming years for many content creators. The table below shows how this could progress.

	2017	2024	Increase
Capacity per hour shots	5 TB/hr	70 TB/hr	**14X**
Hours shot per hour of completed work	5	10	**2X**
Capacity per hour of completed work	25 TB/hr	700 TB/hr	**28X**

For content owners, this creates a compelling challenge with big implications: either significantly increase storage budgets, or move data from primary storage to static archives at a faster rate than in the past. Both of these changes would negatively impact production cost and complexity.

Furthermore, these issues will only become greater over time. For instance, by 2024, for a 2-hour completed project with 5 hours shot per hour of completed work, the total capacity shot is 700 TB. With 100 hours shot per hour of completed work, the total capacity shot is about 35 PB.

ACTIVE ARCHIVES ADDRESS MULTIPLE CHALLENGES

A potentially more attractive option for storing frequently used content is to increase the use of active archives.

Traditional archiving of media and entertainment content uses removable media such as tape and optical discs that can be put on a shelf or in an automated library system until needed. While this is an adequate means of storing media, especially for a long time, retrieving that media takes time that is not always available in time-pressured workflows.

Format Proliferation Leads to More Frequent Archive Access

Indeed, the media access time challenge is more pressing than ever with the increasing demand to re-use and access content in video archives.

Video content is now frequently transcoded into multiple formats for various distribution channels (for instance, traditional cable or satellite transmission, OTT Internet distribution, YouTube and other online video sites, Video on Demand and digital cinema), and an active archive often serves as a content library.

Active Archive Option 1: Local Disk Arrays

Hard disks are increasingly used for long-term storage, a trend driven by the need to access content frequently and rapidly. This is because access latency is far lower using hard drive arrays (on the order of milliseconds) than for tape or optical discs in a library (seconds and even minutes), and because of the declining cost of hard disk drive storage (particularly for the large hard disk drives used for archives). For instance, waiting to load a new file can take up to 5,000 times longer if the data is on tape versus disk. Slower file access means slower workflows, and unhappy clients.

The challenge to using a disk-based archive is cost. Capacity increase requirements far outstrip the traditional disk system cost decline. Traditional large disk storage systems, capable of scaling to petabyte-plus sizes, are expensive to buy and manage, leaving the media owner with a budget challenge.

Active Archive Option 2: Cloud Storage

Cloud storage solutions promise low cost and scalable capacity. Indeed, some video production facilities are moving workflows there. However, public clouds do not suit all use cases for several reasons:

- **Security:** Many media and entertainment companies are not willing to put their assets into public cloud storage for fear of losing control or piracy.
- **Performance:** The large size of video content archives would be difficult to access easily if those assets were off-site in a public cloud and away from the professionals that are using the content.

ACTIVE ARCHIVE, IN BRIEF

Archives that allow ready access to their contents are referred to as "active archives."

These are typically highly-scalable, disk-based systems where media access times are measured in milliseconds. This compares with access times of many seconds or minutes for traditional tape libraries.

Active archives therefore accelerate the workflow, make it easier and faster to re-use/reformat media, and increase the value of existing assets.

- **Access Costs:** Moving data to the cloud is inexpensive, but retrieving it may cost you. Bi-directional workflows where data goes both ways may not be cost effective with content in a public cloud.

Active Archive Option 3: On-Premises Object Storage

A third option — object storage systems— is rapidly gaining attention as a viable option for local active video archives.

These are disk-based systems that store media files, but are designed in a way that makes them more scalable and more cost-effective than conventional disk systems. They provide the immediate access of disk, but are built on a low-cost modular design.

The result is an affordable starting point and petabyte scalability — at about 70% less cost than traditional high-capacity disk.

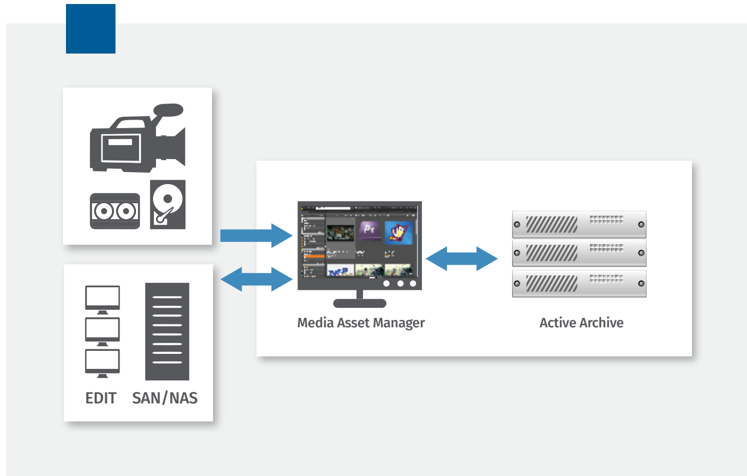
Key attributes of object storage include:

- **High-speed:** Architecture combines disk plus flash storage for optimal cost/performance.
- **Low-cost platform:** Built on conventional servers for lowest possible hardware cost.
- **Modular:** Software combines multiple storage "nodes" into a single group that act like one big, fast device.

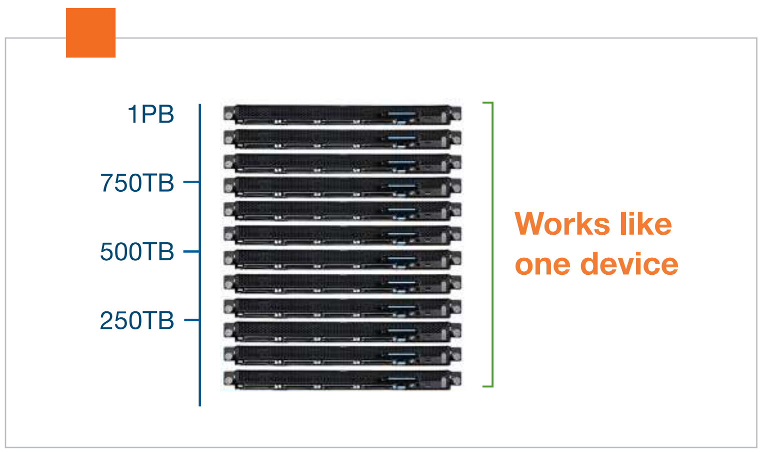
These attributes set object storage apart from conventional SAN and NAS, and make it a compelling option for media storage.

Same Underlying Technology as Netflix, Amazon Prime Video and YouTube

Object storage technology is the storage architecture of the cloud. Netflix, Comcast, Amazon Prime Video, and YouTube all use object storage as their primary storage to solve the same challenges: the need for scalability and performance that exceeds what's possible otherwise. They all concluded that object storage was the only viable solution. Now, enterprise object storage system vendors bring that same



Active Archive Workflow
An active archive provides a quickly-accessible media store for the media asset manager.



Capacity Scaling
Object storage "nodes" combine to form a cluster that acts as a single system to simplify management.

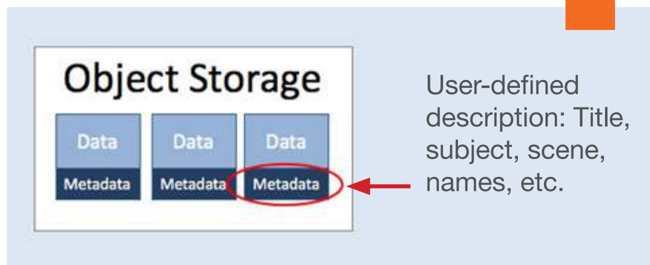
technology to the data center.

Three Reasons Why Object Storage is Compelling for Media

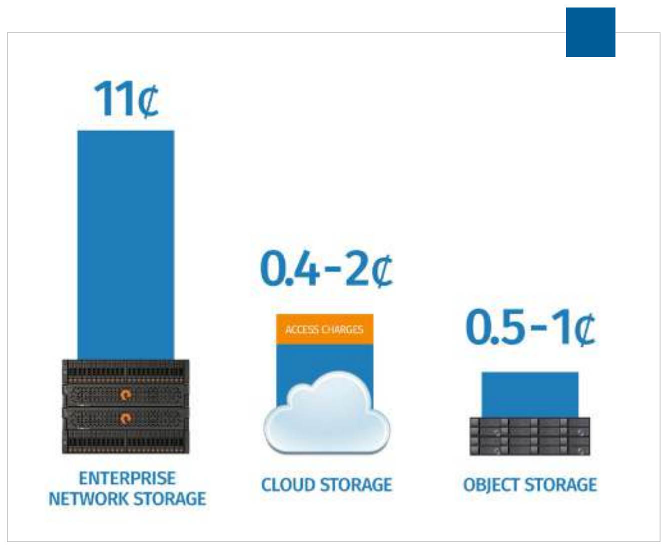
Compared with traditional enterprise network storage (NAS and SAN systems), object storage has three important differences:

1. **Limitless Capacity:** With traditional network storage, the structure is usually arranged in hierarchies that have practical limits, typically in the range of 1 petabyte of capacity per system. In the age of 4K and 8K formats, that limit can be exceeded with just a few hundred hours of media. Object storage has no hierarchy and therefore eliminates these limits. The system can grow to whatever size is needed, and can accomplish this growth in cost-effective increments and with little or no downtime.
2. **Self-describing Media:** Object storage integrates rich metadata to tag (or “label”) assets. Tagging assets with complete descriptions, such as title, scene, subject, performers, or whatever else will be helpful in future searches makes it easier to access that data in the future. Wherever the asset is in the object storage, media can be found using a simple Google-like search.
3. **70% Less Cost:** Because object storage is built on non-proprietary industry-standard servers, it costs much less, often as much as 70% less. Most enterprise network storage systems employ complex architectures to achieve high reliability and enterprise functionality. Object storage uses conventional server technology.

These attributes—capacity, embedded metadata, and lowest possible cost—explain why object storage accounts for the majority of storage used by the major public cloud providers.



Embedded Metadata Enables Search
User-defined metadata, embedded with the object, acts as a “label”. Large media libraries can then be searched using Google-like tools.



Cost Comparison
Object storage offers on-premises storage at costs comparable to public cloud storage.

SUMMARY

Four factors now create an urgent need for high-capacity storage innovation:

- **Exponentially increasing capacities:** On-site storage capacity growth rates now exceed 50% per year. To put that in perspective, if you manage 100TB of capacity today, in five years that will be 800TB.
- **Time pressured workflows:** Distribution outlets will continue to grow, leading to a relentless push for content re-use and monetization. Instantaneous access to assets will be essential to keep up.
- **Public cloud limitations:** Many workflows will continue to demand on-premises storage to achieve required data latencies. Reliance on the public cloud as a cure-all is not realistic.
- **Traditional network storage is cost prohibitive:** Conventional SAN and NAS storage are usually too costly to address archive requirements. Scalability becomes a limitation as well when multi-PB storage is needed.

Cloud providers faced identical capacity and cost challenges, and uniformly moved to object storage technologies as a solution. Commercial options now make on-premises object storage systems both affordable and achievable in most studio environments. Production houses should investigate these new systems to prepare for the next wave of media capacity growth.

CASE STUDY

Object Storage Helps Major Weekend Comedy Show Archive Over 40 Years of Media

Active archives are an increasingly critical part of the media workflow. Today, the need for highly scalable and quickly accessible archives is greater than ever as higher resolution media demands more capacity and media re-use requirements multiply in multi-format, time-pressured workflows.

A well-known weekend comedy show knew these challenges first-hand. They needed a new archive solution, having outgrown the capabilities of their tape archive. Their post-production staff required speedy access to content from 40 years of programming — over 800 episodes, millions of digital assets, and petabytes of data. And they required a less labor-intensive process that would eliminate tape handling.

After thorough analysis, the studio deployed two object storage systems from Cloudian to support the production active archive and DR site.

The challenge of achieving flexible and searchable archive

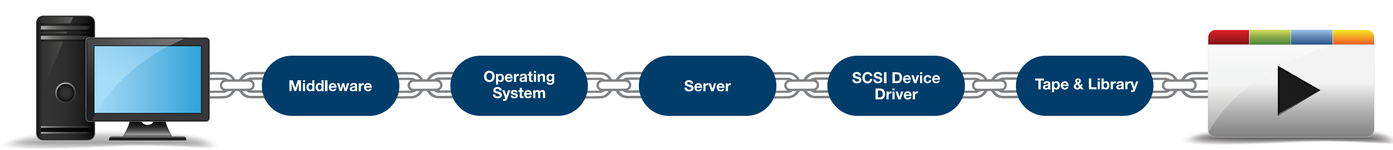
The show's previous archive solution employed tape libraries plus an offsite warehouse for long-term storage. Archiving was controlled via a media asset manager which maintained the asset database. The show's older episodes were originally recorded to analog tape, which had since been moved to digital tape.

To ensure integrity, it was periodically necessary to read the tapes and re-write them to new tape media. An ongoing challenge was the regular format transitions that occurred when tape technology progressed from one tape generation to the next. As the post production supervisor put it, "we had systems under glass, a sort of technology time capsule that let us access older media types. If anything broke on those old systems, we had real challenges getting to those assets."

"We had systems under glass, a sort of technology time capsule that let us access older media types. If anything broke on those old systems, we had real challenges getting to those assets."

A risky chain of dependencies

Compatibility is a lurking problem: to retrieve media, it is essential that all parts work together. This creates a "chain of dependencies" which must remain connected for continued access. For example, consider the process of reading media from tape. A tape can only be read by a tape device from specific generations. Those tape devices require a specific driver, which only runs on certain software, which in turn runs on a specific operating system. Finally, that software maintains a database which is essential to accessing all the assets.



EDITING BAY **CHAIN OF DEPENDENCIES** **MEDIA CLIP**

Maintaining full interoperability is straightforward only in the short run. Over time, through multiple release cycles and changing vendor strategies, some parts of the chain will inevitably break: something won't work. As the manager put it, "over the course of my career, I've seen every part of the chain get changed. Ensuring access through all of those transitions has itself been a full-time job."

"Over the course of my career, I've seen every part of the chain get changed. Ensuring access through all of those transitions has itself been a full-time job."

Tape's limitations as an archive

Beyond the interoperability challenges, tape was unreliable. The tape libraries, tape readers and tapes themselves could all fail at inopportune moments. The manager recalled, "many times I've had my hand in the library trying to fix a jam while the robot is whizzing around in there. When we're on a deadline, we do what it takes, but sometimes it's perilous."

Tape also presents logistical challenges. Tapes had to be moved among sites, and when it was necessary to retrieve them, even street traffic could be a factor in meeting a deadline.

To use an asset, you must first find it

Another limitation of tape is the ability to find assets. The search capabilities were only as good as the underlying media asset manager. When searching for specific clips, the producers were limited to the MAM capabilities and the indexing decisions that had been made years ago. "It seemed crazy that in the era of Google we would be limited by primitive search of our most valuable resource, but we were. Finding media could take hours if not days," the manager added.

"It seemed crazy that in the era of Google we would be limited by primitive search of our most valuable resource, but we were. Finding media could take hours if not days."

Objectives for the next-gen archive solution

From these challenges, the studio's engineering department compiled objectives for their new active archive.

The top three objectives were:

1. Break the chain of dependences
Re-think the archival strategy and implement a solution that would ensure long term, risk-free access to media.

2. Rapid search
With millions of assets in the database, a more scalable search solution was essential, one that could change as search tools and search requirements evolved.

3. Planning for growth
In the entertainment business, capacity requirements grow exponentially over time. Even if the number of assets only grows linearly, capacity demands will still be driven skyward by increased format resolution. 4K and 8K will inevitably be followed by other new formats, so it's essential to plan for the unknown.

The answer: object storage

After lengthy consideration, the program's engineering department concluded that object storage was the only means to achieve their objectives. This storage type would address their objectives with a disk-based system which overcomes the scalability and cost issues found on other enterprise storage.

Breaking the chain of dependencies

One of the show's objectives was to remove the risks inherent in media access: the driver, hardware, and software limitations all had to work perfectly to ensure access. Object storage works differently than conventional storage in several ways that break the chain of dependencies.

- Freedom from drivers: There are no proprietary hardware or drivers – it's all HTTP. Object storage is the only storage type invented in the internet-era, so it's the only one to capitalize on internet connectivity tools to ensure universal access.
- Portability: Objects can be moved from one storage environment to another. Move between vendors or to the cloud. Cloudian even has functionality built in that can make this automatic, if you choose.

"In previous environments, such as tape, we were stuck with whatever organizing scheme we started with. Now we can change things and even go back and re-tag media if needed."

- Hardware independence: Object storage is built on industry-standard servers, so hardware can be refreshed at minimal expense.
- MAM independence: Object storage leverages tags that are stored with the media. Locate media via standard search tools, independent of the MAM database. A database can always be rebuilt, if needed, using those tags.

For this comedy show, these capabilities changed their storage expectations. The manager stated, “Now we can always get to our data without having to worry about vendors and what their agenda might be.”

Plus, the scheme for organizing and finding assets can evolve over time. “In previous environments, such as tape, we were stuck with whatever organizing scheme we started with. Now we can change things and even go back and re-tag media if needed,” recalled the manager.

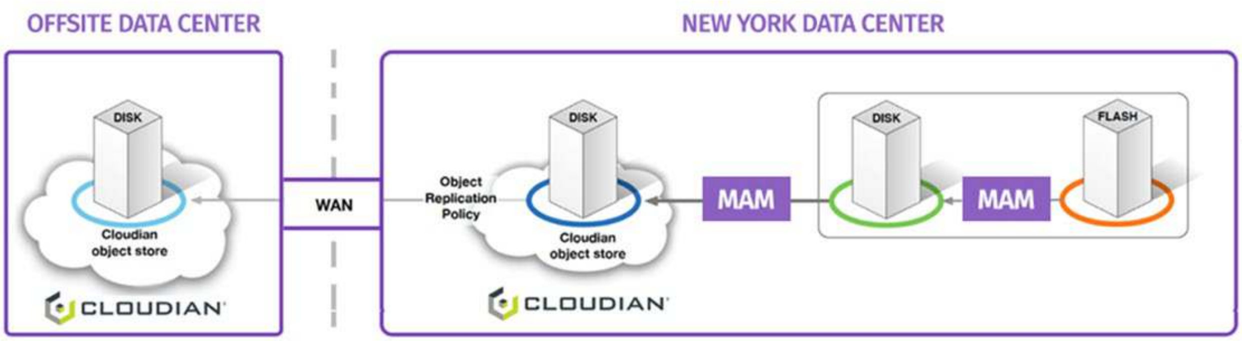
Simple scalability

Object storage simplified the ever-present task of adding capacity. The technology is built on “nodes”, like storage bricks, that can be added to the cluster as needed. They integrate themselves, with new capacity simply added to the common pool.

For this show, that allows them to start with the capacity they need now, keeping costs down. They can add new capacity at any time, without even a service call, so they always have the flexibility to grow.

A new storage environment

Currently, the show employs an all-flash SAN for primary storage and Cloudian object storage as the active archive. A second Cloudian cluster at an offsite location holds the disaster recovery copy. Replication is managed by Cloudian’s built-in data management features.



“This is where we’ve been trying to go since I ingested that first tape and watched storage space disappear,” the manager said. “My job is to make sure our assets are stored, safe, and accessible. We’re finally there with an answer that will hold up over time. We’ve had our last ever data migration, and that feels good.”

“My job is to make sure our assets are stored, safe, and accessible. We’re finally there with an answer that will hold up over time. We’ve had our last ever data migration, and that feels good.”

About Cloudian Object Storage

Cloudian simplifies storage management with a limitlessly scalable platform that consolidates massive data sets to a single, easily managed on-premises environment. Available as appliances or as software-defined-storage, Cloudian HyperStore can scale from just three nodes to hundreds, allowing systems to be right-sized for any application or organizational need. Cloudian dramatically reduces enterprise storage costs with up to 95% less management overhead, 30% less power/space/cooling, and a highly robust design that ensures maximum productivity with up to 14 nines data durability. Cloudian use cases including media and entertainment, video surveillance, data protection, bioinformatics, IoT, and more.

About the Author

Thomas M. Coughlin, President, Coughlin Associates is a widely respected storage analyst and consultant. He has over 35 years in the data storage industry. 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, published by Newnes Press. Tom publishes the Digital Storage Technology Newsletter, the Digital Storage in Media and Entertainment Report, and other reports.

Tom is active with SNIA, SMPTE, IEEE, and other professional organizations. He is Education Chair for the SNIA Solid State Storage Initiative. He is Chair of Future Directions for the IEEE Consumer Electronics Society as well as past Director for IEEE Region 6. He is a long-standing member of the CE Society BoG and was Vice President of Operations for three years. Tom is the founder and organizer of the Annual Storage Visions Conference as well as the Creative Storage Conference. He has been the general chairman of the annual Flash Memory Summit.

For more information on Tom visit www.tomcoughlin.com.

Coughlin Associates