CLOUDIAN®

WGBH Builds a Hybrid Cloud Active Archive Around Cloudian HyperStore

The Challenge

If you watch PBS, you've probably seen the WGBH logo at the end of many programs; their productions include Nova, Frontline, Masterpiece, Antiques Roadshow, American Experience, and the children's series Arthur. Past productions include Julia Child's the French Chef, the New Yankee Workshop and the children's programs ZOOM! and Curious George. And WGBH is creating new programming all the time, using data-intensive formats like 4K video.

50 years of content creation results in an enormous archive, most of which was stored in a tape library or on external hard drives, all residing in a large vault at the WGBH studios in Boston.

The station's many production teams shared a common archive, which strained the IT team's ability to keep pace with the growing rate of ingest. An ongoing transition to 4K and 8K media created ballooning capacity demand, while the growing rate at which media was created added to the sheer volume of material to be archived.

Media retrieval time was a challenge as well. When producers needed access to archived material they submitted a request to the archive team, which then located the media. With a 24 hour SLA for retrieval requests — which could stretch to 72 hours if the request came on a Friday — production teams on tight deadlines sometimes had to find alternative sources.

Data protection was also problematic. Only a portion of the media could be moved to the offsite DR location, which represented a tremendous liability. "If the greater Boston area was in some way compromised we would have zero ability to restore that archive," said Shane Miner, WGBH Senior Director of Technical Services.

The Decision

In the search for a new solution, the WGBH IT team first determined that a cloud-only archive solution would not meet their needs, largely due to workflow requirements. Downloading a large file from the cloud would consume time, in effect adding another step to the workflow. Cloud access charges would also add cost for this busy archive.

The WGBH team decided instead that a hybrid cloud approach would best meet their needs. A hybrid cloud environment combines on-premises storage and public cloud storage: on-prem for the working copy, plus public cloud for a disaster recovery copy. This would combine the immediacy of an on-prem system with simplicity of cloud-based DR.

This left open the question of what to use for on-premises storage. WGBH considered storage area networks (SAN) and network attached storage (NAS), but eliminated them from consideration because of cost, scalability, and their inability to handle capacity WGBH expected to store without suffering performance issues. They also lacked the ability to tag files with the detailed metadata WGBH desired to facilitate search.

Eventually, WGBH's research led them to object storage — specifically, HyperStore from Cloudian — which offered what WGBH needed: fast access to data, limitless

INDUSTRY
Media & Entertainment

CHALLENGES
- Workflow efficiency: difficult to keep pace with accelerating archive ingest rates
- Scalability: capacity to accommodate growing data volume
- Time-consuming search: difficult to locate and retrieve media
- Labor-intensive DR: high workload to create and move additional tape copies
- Integration with Sony Ci

SOLUTION
- On-prem object storage system: 3PB cluster in 12U of rack height
- Rich metadata tagging with Elastic Search: allows rapid search based on descriptive tags
- Automated DR: policy-based replication to Amazon Glacier
- Integrates with Sony Ci via S3 API

Stacey Decker, CTO

"We've done a lot of work to determine what works the best for us. The Cloudian solution rose to the top very quickly."

Stacey Decker
CTO, WGBH

VIEW VIDEO »

capacity, modular and easily-managed growth, high density, metadata tagging to facilitate search, and low cost. The initial deployment consisted of a three petabyte cluster, housed in three 4U-high appliances. Consuming just 21" of rack height, this cluster consumes less than 1/10 the space of the equivalent tapes and library facilities.

Benefits of HyperStore

The benefits of HyperStore's capacity and performance were immediately apparent. The producers of Frontline saw production on a new episode brought to a halt when they ran out of capacity on their editing platform. To resume work, it was essential to offload some capacity immediately. With the newly-installed Cloudian HyperStore in place, the team could free up editing space in minutes, a job that previously would have taken hours or days to complete.

Capacity challenges are a thing of the past with the easily scaled HyperStore system, which can scale as the amount of data in the library expands. HyperStore clusters are implemented by deploying nodes into a peer-to-peer architecture. As physical nodes are added, all resources are automatically aggregated into a common pool of storage and CPU resources across the cluster.

WGBH employs HyperStore's policy-based management for automatic replication to Amazon Glacier, replicating data from the archive as it's stored. "With Cloudian, DR became automatic," said Miner. "We store data to the archive and it's automatically replicated to the cloud. That's a lot simpler and more reliable than managing tapes."

Metadata for Fast Media Search

Unlike file storage, object storage allows extended metadata to be added to each file. "The real power of object storage is that we can quickly find content by searching the metadata stored with the object," Miner said. "This value will grow over time as we use new AI tools to enrich metadata with detailed scene descriptions and audio transcriptions that will further enhance the value of our content."

Because metadata can be defined by users, and updated and enriched over time, it can provide a far greater depth of detail about the content, allowing production staff to quickly locate relevant content through a Google-like search.

HyperStore's Impact

The WGBH team ultimately defined a workflow that incorporated HyperStore for the on-site archive, Amazon S3 for hybrid cloud capacity expansion, Amazon Glacier for hybrid cloud disaster recovery, and Sony Ci to facilitate production workflows.

When a program episode is complete, WGBH keeps a copy of the raw material and the final show in the Cloudian-based archive, and duplicates the media to Amazon Glacier for data protection. Completed shows are also stored in the Sony Ci production workflow software for use by sales teams and for other production prep work. All content is tagged extensively. WGBH also sells stock footage, but in the past, the station was unable to fully capitalize on this source of potential revenue because of the time required to search. Today, stock footage can be sold as soon as the content is tagged and searchable.

"With Cloudian, DR became automatic. We store data to the archive and it's automatically replicated to the cloud. That's a lot simpler and more reliable than managing tapes."

Shane Miner
Senior Director of Technical Services
WGBH