front porch digital

Digital Video Archiving: The Evolving Reality of Any Content, Anywhere, Anytime

Rino Petricola

ABSTRACT

More and more the Information is driving the business. Access to any content, anytime, and anywhere will become the critical success factor for the broadcast industry to succeed. The convergence and the integration of Information Technologies and Broadcast Technologies such as servers, storage, automation systems, media asset management and video devices are allowing to deploy central repositories of content within the broadcast stations. Ultimately content will be accessible at a fingertip and distributed between the various departments of the broadcast stations and also between stations' affiliates. Ultimately content will be accessed anytime, anywhere.

After reviewing some key drivers of the broadcast market evolution, this paper analyses how the introduction of digital asset archive is fuelling the IT/Broadcast convergence and bringing increased automation, greater efficiency, global access and sharing of content.

INTRODUCTION

Almost ten years ago, the first video disk recorder was introduced in the broadcast industry and unleashed the tornado that is today the digital revolution, we mean the storage of video content as computer data instead of analogue or digital video tape recorder tapes.
What the next decade holds for us is an unprecedented explosion of digital information. As broadcasters and video producers throughout the world move to computer-based platforms, digital video files will proliferate and the demand for content storage and archive will drastically increase.

The rapid application of computing technology to broadcast television is demanding a focus on content, or "asset" management. And as computers become faster and network bandwidth wider, customers will expect more and more of their multimedia content to be available in a smaller amount of time and in different formats. In order for broadcasters to place more and more of their content (their video assets) at their users' fingertips, the processes of capturing, storing, preserving, converting, transcoding, moving and sharing this content will have to become simpler and less expensive.

Vendors must provide tools that support and provide higher availability and reliability while reducing overhead, maintenance, and administration costs. What we're talking about is providing much more content, across distributed networks, faster, and at a reasonable cost in such a way that the information becomes much more powerful for the user, and accessible anytime, anywhere.

In today's changing broadcast world, storage is the foundation upon which the broadcast facility is built. Cost effective storage management, archival and retrieval services are the framework built upon the storage foundation. While immediate-use clips are stored on

expensive hard disk storage, the archive system must be relied upon to automatically and transparently copy content from expensive on-line disk to less expensive near-line storage, and restore the content back online as needed.

EVOLUTION OF THE BROADCAST FACILITY

The need to automate

Over the last two decades, there has been a relentless drive to automate the daily on-air operations of broadcast television stations large and small.

One of the most revolutionary devices to enter the broadcast arena has been the disk based video storage device, commonly referred to as a video server.
Video servers are storage and playout devices that can hold many hours of video and audio and playout the programs or commercials on demand. A video server is controlled by an automation system which controls the ingest, storage, and playout of material according to an accurate schedule. The automation system controls many pieces of equipment throughout the broadcast facility and ensures that all material scheduled to go "on air" is available and is played out in accordance with the broadcasters schedules and contracts.

Advanced automated storage and playout systems employ a hierarchy of storage technologies to store multimedia files and associated information. In order to be highly automated and efficient these systems utilize archive management software similar to Hierarchical Storage Management (HSM) software currently used in the computer industry.

These automated storage and playout systems also provide large amounts of highly reliable digital data tape storage for protection and storage of their multimedia assets. They provide high bandwidth, high speed network connections to other servers, editing systems, and playout and production devices. Archive system management data can be exchanged with other systems via a robust Application Programmers Interface (API).

Many broadcast archives consist of not only various videotape formats, but also motion picture and photographic film. These media formats, particularly film, have a limited life span and can even be thought of as "perishable" items. Multimedia materials stored in the archives are essentially "assets". It is imperative to archive the material in the most reliable and highest quality format available.

Traditional methods

In order to view an archived piece of material in a broadcast facility you must somehow identify the tape or reel on which the material resides, manually remove from a shelf, and then transport it to a viewing suite which contains the appropriate format playback device such as a VTR or film projector. You must then watch the entire tape or reel to locate the material you are seeking. This is obviously a long, tedious process which ties up valuable equipment, people, and space.

In many cases (especially film), material may be stored at reduced temperature for a better preservation. Additional time is required to allow the media to return to normal temperature before it can be used.

INTRODUCTION OF DIGITAL ASSET ARCHIVE

Benefits

Digital multimedia archiving technologies provide the following advantages:

- Multimedia material is stored digitally as computer data. Digital data is not prone to generational loss due to copying. Each copy is "loss-less" and is exactly identical to the master material.
- Archival storage is highly repetitive in nature and is ideal for the application of computerized and robotic systems. By using computers to track material and robots to store and move cartridges to digital ingest and playout devices, human intervention (and therefore human error) is not necessary.
- Since the material is stored digitally, no quality check is required. The produced copy will be identical to the master. This reduces the time and resources required to retrieve an archived media file.
- Association of metadata: metadata is information pertaining to a multimedia data file that is used to describe the file. It is commonly stored in a computer database and is used to locate and identify the data files.
- Networking for data, also called essence, and metadata: in order to work efficiently and rapidly in a busy broadcast environment, video servers and archives are often networked so that data may be interchanged rapidly and without any quality loss. Managing the network bandwidth of a video server network is a difficult task, especially when several mission-critical transfers are taking place simultaneously.
- Interfaces to other systems: video servers and data tape archives do not operate independently. They must interface to many other devices such as the plant automations system, video and data routers, and news editing systems. Automation systems typically control the video server which, in turn, controls the data tape archive.
- Interoperability with different video devices: real-time transcoding during archive or restore operations allows various video devices to share and playout content from different sources.

Ingest

Ingest is an application or a function provided by the video server. Material is brought into the video server from any one of the following four sources:

- Video tape via a video tape recorder connected to an input of the video server (usually via the plant routing switcher).
- Satellite delivery: satellite video feeds are recorded directly into an input of the video server via the plant routing switcher. The plant automation system typically controls the recording of the material.
- Data tape from archival storage or from a similar archive system within the same facility or from outside the facility.

- Wide Area data Network: the material is sent from a remote location via ATM, Frame Relay, OC3, or other data network technology.

Capture, digitization and compression of the material is performed at ingest. Once converted to digital and compressed, the material will stay in this format throughout its stay in the video server and the data tape library. When the material is played out to air it is converted back to baseband video and audio by the CODECs in the video server and sent to the transmitter.

Storage

Within a broadcast facility there are different types of multimedia storage, each with its own requirements.

Online:
Online storage is the mission-critical storage for play-to-air material. It holds several hours of spots and programs which are scheduled for playout during the day. It is constantly checked and updated by the automation system. The automation system is usually the only application allowed to access the online storage. Online storage typically contains a minimum of 6 hours of material and can reach enough capacity to store material for the next two or three days programs.

Near-line:
Archival storage in a data tape library is much less expensive than storing the same amount of information on disk-base storage. For this reason it makes economic sense to store the bulk of your material in a data tape library and use your hard disk storage for material which requires near instantaneous access to any piece at any time. The trade off here is obvious: the less expensive storage medium (data tape) requires more time to access the material in storage. Near-line automated tape libraries are the primary medium for archival storage. They are also ideal for short-term storage.

Cataloguing

A particular thorn in the paw of multimedia archivists is the cataloguing of information about a particular piece of material at the time it is ingested into the library. It is vital to the success of the indexing and searching system that definitive, specific terms are used to describe the piece of material. For instance, if all operators used a term such as "good shot" to describe a scene they wanted to index, then the ability of the database to return desired matches is greatly reduced as all the scenes would be indexed as "good shots".

Emerging technologies such as voice recognition, scene recognition and scene change recognition will allow the cataloguing process to become more automated with time. Information kept to describe or accompany a piece of multimedia material is called metadata. An archival storage system must ensure that metadata is stored linked to, and moved with, the multimedia data file.

The term metadata refers to information pertaining to or describing a multimedia file. It often contains important information such as copyright data, contractual usage information, and information about the file's origination. Metadata must be tied to the data file by index keys within the media management database. If the metadata becomes separated from its data file, neither will be useful and it will be impossible to locate a data file.

Browsing

Today's production environment is hectic and rushed. Production people are busy searching through new and archived media files to edit together new material for broadcast. Compounding matters is the fact that these people often need nearly simultaneous access to the same material. In the past, this was handled by making dubs of the target material and giving everyone a tape copy to use. Tomorrow's broadcast facility must include a media or asset management system that allows many users to simultaneously search for and manipulate archived data files.

By combining the catalogue metadata with the ability to see and hear the contents of the multimedia file, a user can quickly preview or "browse" the data file and its pertaining metadata. It also means that a user can quickly search for a particular shot or scene using metadata "keywords" contained in the database record. Once a user has located a desired piece of material, he or she can transfer the material to a production server for further editing or processing.

Typical browsing architectures rely on a low resolution copy, or "mirror", of the original material which resides on a disk-based Browse Server and is connected to the media management database. The Browse Server contains highly compressed "mirror" copies of all the files on the media server. To minimize the cost of hard-disk storage, the content of the browse server is highly compressed. Since the purpose of the browse server is to allow fast and easy identification of scenes and material, image quality degradation and size reduction as a result of high compression rates is assumed to be admissible.

Users of the browse server will have the ability to fast-forward, stop, rewind, jog and shuttle, and mark material just as they would if using linear or non-linear editing systems. Once a scene has been identified, marks are placed in the database and the original high resolution material can be transferred to a production server for further processing.

ARCHIVAL STORAGE

Automated tape library

The cost of a mass storage system is determined by the costs of the program medium per hour of storage and by the expenditures for the necessary storage room (cm3 per hour of storage). Based on its extremely low thickness, tape medium exceeds any other type of storage medium regarding the volume density of stored information. Magnetic tape is the medium of choice. Optical tape may prove beneficial over time, but is still too new to be practical and cost-efficient. Most broadcasters have relied on a storage model similar to a traditional lending library. The key to finding any particular book is the library's catalogue. For decades, lending libraries used index card files. Today, we have replaced card catalogues with computers and databases. Instead of using index cards, we use records in a database. The task then becomes one of managing the database. Video and audio (or multimedia) storage is moving away from being stored on a tape on a shelf to being managed as digital data files within a computer system. Reduced personnel, greater reliability, better asset control, and content re-purposing are only a few of the competitive advantages broadcasters obtain by going to a digital file based broadcast facility.

Originally, disk-based servers were employed as cache devices for videotape cart machines. Now, disk-based servers are the central core of the broadcast facility. Because of the high cost of hard disk storage, it is not economical to store all of a broadcast station's material on a

disk-based server. A near-line automated tape library is a cost-effective and reliable mass storage device which easily replaces an aging videotape cart machine and effectively expands the storage capacity of a video server.

Automated tape libraries provide a key piece to the storage solution puzzle. Tape libraries, because they store multimedia files in a data format compatible with the video server and not as video and audio, can be considered as an extension of the video server. In a tape library, multimedia (that is, video and audio) material is stored as files. These data files are created by the video server and are comprised of data that has passed through the video and audio codecs and has been compressed and combined into a single stream. Because the files are stored as video server data files, generational loss normally incurred as files are moved and copied is eliminated.

The archive manager

An archive manager is a middleware software application which bridges the gap between disk-based storage and the automated tape library. Its purpose is to provide an interface, or more accurately an abstraction layer, between the video server and the automated tape library's tape drives and robotics. Not only does it provide much needed connectivity, it also provides media management in the form of a database to keep track of the contents of the data tapes and the bins inside the library.

It is very important that the middleware archive manager application be compatible with many automation systems as well as video servers. Most archive managers are compatible with automation systems from Encoda, Harris/Louth, Omnibus, Probel, and others. They must support video servers or edit systems from Avid, Omneon, Pinnacle, Quantel, SeaChange, Thomson/Grass Valley Group, and others. Additionally, they must provide support for asset managers and content servers such as ArkeMedia, Blue Order, Konan, Omnibus, and others.

The archive manager software must do more than just provide connectivity between video servers and automated tape libraries. It must provide a complete management environment for archived multimedia. In the past, video and audio were previously recorded onto tapes and stored on shelves. These tapes then had to be physically retrieved from the shelf by a librarian in order to be placed into a video tape recorder for playback. Now, with video servers and automated tape libraries, all multimedia can be accessed electronically via networked systems. Not only does this simplify operations, it makes all the multimedia assets within the facility more accessible and, therefore, more valuable.

When video is copied from the video server to the automated tape library, it is said to have been "archived". Once archived, multimedia material is treated as a file and is stored by the archive manager software. Archival functions are initiated by one of the following applications:

- Applications running on the video server such as channel record, time delay, or video tape recorder emulator application.
- Automation system. The automation system runs external to the video server.
- A Media Asset Manager (MAM) supervising content distribution.

By providing an abstraction layer for the applications, the archive manager middleware relieves automation and MAM systems from storage tasks, which would otherwise be tedious and complex, to a few simple commands necessary to control the archive software. The archive manager middleware manages the tapes drives, the library robotics, and the media database, performs the necessary complex logic, and issues the multitude of commands to the

various devices in order to perform commands which, to the video server appear simple such as "archive", "restore" and "delete".
When it needs to store a piece of material in the automated tape library, an application such as an automation system sends the material to be archived to the archive manager middleware. The archive manager performs all the necessary steps required to copy the material to a data tape within the library. It then notifies the sending application of the status of the operation. Subsequently, when the application needs to use the material again, it knows that the material has been stored in the archive. It sends a request to the archive manager to retrieve the video by the file ID which was assigned to it during the archive operation. The archive manager middleware takes care of the underlying storage and retrieval operations.

Managing the tapes in the library

Inside the automated data tape library, the data files are stored on magnetic data tape cartridges. The size and capacity of each tape cartridge depends on the tape format being used. The archive manager middleware controls the robotics, the tape drives, and the tape media in the library. When an application needs to archive a multimedia file, the archive manager middleware selects a tape on which to store the file, mounts the tape in an available tape drive, and then moves the file data from the video server and writes it to the tape. When an application needs to retrieve a file from the archive, it sends a restore command to the archive manager. The archive manager determines the physical location of the file in the library, mounts the appropriate tape in an available tape drive, then moves the file data from the tape to the video server where the application can then use it.

Because the archive manager specializes in archive management, it can offer a number of features which would be difficult to implement within a video application:

- Spanning: a video file can be made to span across multiple tapes with no intervention by the application. If a video file being sent to it is longer than the current tape, the archive manager software will automatically retrieve another tape to finish the storage process. Upon retrieval, the application does not have to know that multiple tapes were used for storage. The application simply requests the video by its file ID and the archive manager retrieves the tapes needed to restore the file.

- De-fragmentation: when objects are deleted from tape, tape fragmentation inevitably occurs. The archive manager can then operate de-fragmentation or manually or automatically as a low priority process.

- Grouping: same kinds of content (e.g. news clips) can be grouped together on a separate set of physical media (e.g. tapes).

Managing the content

The archive manager can also specialize in content management to offer more content related features:

- Transcoding: software transcoding allows during archive or restore operations to transcode from one format to another in real time. This opens the door to multiple applications. For example, a video file archived in broadcast quality can be automatically restored into a browse quality copy or a streaming format.

- Video partial restore: with formats such as MPEG-2 I-frame only at 50Mbit/s, retrieving a two hours movie from tape can take a minimum of 40 minutes even with the fastest tape

drives available today. In many cases, especially in News or Post-Production environments, it is preferable to only restore a short section of the whole clip (e.g. 5 minutes). The clip can be viewed in a much lower resolution for the selection of mark-in and mark-out points expressed in timecode. Then the archive manager can retrieve the corresponding clip section based on the selected time codes (IN and OUT).

- Migration policies: storage plans can be configured and used to manage the lifecycle of content, according to storage rules and policies. Execution of the storage plans can be configured at will: at regular interval with a specified frequency or according to special operations, like archive.
- Multiple restore: multiple restore enables to restore a single content to multiple destinations simultaneously. Usually such feature is used for mirrored transmission servers, so that the automation can restore a clip to two servers in a single command..

The media database

The video and audio files are not the only type of data that must be stored in the archive. The archive manager also stores a metadata file associated with each multimedia file in the archive in a relational database. This database contains information needed by the archive manager to be able to move video material between servers and archives. It also contains information that describes the content of the multimedia file. This information is used by the video applications to locate and identify the material.

The metadata stored in the archive manager's database is used both by the archive manager and by the applications the archive.

Through standard automation systems interfaces, the material needed for recording and playback is determined. If a particular video file is not on the desired playback server, the automation system initiates a request to the archive manager software. The archive manager then finds the video file and routes it to the designated server.

Video servers have revolutionized the broadcast facility. Tape archival libraries provide affordable mass storage. They are, more accurately, extensions of video servers and can provide nearly unlimited storage for just pennies per megabyte. The video server and the tape library are bound together by a software/middleware package which provides a level of control over the library which would be tedious and cumbersome for either the video server application or the automation system to have to assume. Both the video server and the data tape library are managed by the station automation system.

CONCLUSION

Storage is the foundation upon which the broadcast facility is built. Cost-effective storage management, archival and retrieval services are the framework built upon the storage foundation. As technology advances and disk capacity becomes greater at a lower cost, so too will data tape capacity and tape transport data rates increase at lower cost. For the foreseeable future the architecture will remain the same: video server caches with data tape library archival storage. Capacities will go up and prices will go down. Interface issues will continue to be of primary interest. Both disk and data tape storage will continue to provide a solid foundation for many years to come and its integration with the legacy video applications and devices is allowing any content to be accessible anytime, anywhere in the broadcast facility.