ORACLE

DIVA

Digital Video Archiving

The Evolving Reality of Any Content, Anywhere, Anytime

ORACLE WHITE PAPER | UPDATED MAY 2015

ORACLE

Executive Summary

More and more, information drives the business. Access to any content, anytime and anywhere, will become the critical success factor for the broadcast industry to succeed. The convergence and integration of information technologies and broadcast technologies, such as servers, storage, automation systems, media asset management, and video devices, are enabling companies to deploy central repositories of content within the broadcast stations. Ultimately, content will be fingertip accessible and distributed between the various departments of the broadcast stations and their affiliates—and content will be accessed anytime, anywhere.

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

Effects of the Digital Revolution

Almost 10 years ago, the first consumer video disk recorder was introduced in the broadcast industry and unleashed the digital revolution. At that time, video content began to be stored as computer data—instead of on analog or digital videotape recorder (VTR) tapes. The next decade will bring 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 archiving will drastically increase.

The rapid application of computing technology to broadcast television is demanding a focus on content, or asset, management. As computers become faster and network bandwidth becomes wider, customers will expect more and more of their multimedia—that is, video and audio—content to be available in a smaller amount of time and in different formats. For broadcasters to place more of their video and audio assets at 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 administrative costs. Together, vendors and broadcasters must provide much more content—across distributed networks, faster, and at a reasonable cost. The information must be delivered in such a way that it becomes much more powerful for users and accessible anytime, anywhere.

In today's changing broadcast world, storage is the foundation upon which the broadcast facility is built, and cost-effective storage management, archival, and retrieval services are the framework of that foundation. Whereas immediate-use clips are stored on expensive hard disk storage, the archive system must be relied on to automatically and transparently copy content from expensive online disk to less expensive near-line storage, and then to 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 content to be played out on demand. A video server is controlled by an automation system, which controls the ingest, storage, and play out of programs and commercials, 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 both available and 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. 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 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 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 lifespan 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

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 it from a shelf, and transport it to a viewing suite that 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 using valuable equipment, people, and space.

In many cases (including that of film), material can be stored at reduced temperature for better preservation. Additional time is required to enable the media to return to normal temperature before it can be used.

Introduction of the 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 ideal for the application of computerized and robotic systems.** Archival storage is highly repetitive in nature, so computers can track material and robots can store and move cartridges to digital ingest and playout devices. Human intervention (and therefore human error) is minimized.
- **Because 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.
- **Metadata makes it easy to find files.** Metadata is information that describes a specific multimedia data 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*, helps manage volume.** To work efficiently and rapidly in a busy broadcast environment, video servers and archives are often networked so that data can 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 allow automation benefits to extend to other areas.** Video servers and data tape archives do not operate independently. They must interface to many other devices, such as the plant automation 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 enables sharing of content from multiple sources.** Real-time transcoding during archive or restore operations enables various video devices to share and play out 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 of the following sources:

- **Videotape.** Videotape can be delivered via a VTR connected to an input of the video server—usually via the plant routing switcher.
- **Satellite video feeds.** 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.** Data tape can be brought to the video server 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, optical carrier 3, or another data network technology.

Capture, digitization, and compression of the material is performed at ingest. Once converted to a digital format 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, and each type has its own requirements.

Online

Online storage is the mission-critical storage for play-to-air material. It holds several hours of spots and programs that are scheduled to be played out during the day. It is constantly checked and updated by the automation system, which is usually the only application allowed to access the online storage. Online storage typically contains a minimum of six hours of material and can reach enough capacity to store material for up to two or three days of programming.

Near-line

Archival storage in a data tape library is much less expensive than storing the same amount of information on disk-based 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 that requires near-instantaneous access to any piece at any time. The trade-off 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.

Cataloging

A particular thorn in the paw of multimedia archivists is the cataloging 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 capability of the database to return desired matches is greatly reduced because all the scenes would be indexed as *good shot*.

Information kept to describe or accompany a piece of multimedia material is called **metadata**. Metadata often contains important information, such as copyright data, contractual usage information, and file origination information. 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. An archival storage system must ensure that metadata is linked to, moved with, and stored with the multimedia data file.

Emerging technologies, such as voice recognition, scene recognition, and scene change recognition, will enable the cataloging process to become more automated with time.

Browsing

Today‘s production environment is hectic and rushed. Production staff is busy searching through new and archived media files to edit together new material for broadcast. Compounding matters is the fact that other staff members often need nearly simultaneous access to the same material. In the past, this was handled by making *dubs*, or copies, 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 enables many users to simultaneously search for and manipulate archived data files.

By combining the catalog metadata with the capability to see and hear the contents of the multimedia file, a production employee can quickly preview or browse the data file and its pertaining metadata. It also means that the employee can quickly search for a particular shot or scene using metadata keywords contained in the database record. Once a desired piece of material is located, the employee 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 that resides on a disk-based browse server and that is connected to the media management database. The browse server contains mirror copies of all files on the media server. To minimize the cost of hard disk storage, the contents of the browse server is highly compressed. Because the purpose of the browse server is to enable fast and easy identification of scenes and material, image quality degradation and size reduction as a result of high compression rates are acceptable.

Users of the browse server will be able to fast-forward, stop, rewind, jog and shuttle, and mark material just as they would if they were using linear or nonlinear 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 (cm^3 per hour of storage).

Based on its extremely low thickness, the 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 might prove beneficial over time, but it 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 catalog, and for decades, lending libraries used index card files. Today, card catalogs have been replaced with computers and databases. Instead of using index cards, libraries use database records. 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, and toward being managed as digital data files within a computer system. Reduced personnel, greater reliability, better asset control, and content repurposing 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 that 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 material is stored as files. These data files are created by the video server and comprise 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, the generational loss normally incurred as files are moved and copied is eliminated.

Archive Manager

An **archive manager** is a middleware software application that 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. It not only provides much-needed connectivity but also 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 archive manager be compatible with many automation systems and video servers. Most archive managers are compatible with automation systems from vendors such as Encoda, Louth (acquired by Harris), Omnibus, and Pro-Bel. The archive manager must also support multiple video servers or edit systems from companies that include Avid, Omneon, Pinnacle, Quantel, SeaChange, and Grass Valley Group (acquired by Thomson). Additionally, they must provide support for asset managers and content servers such as ArkeMedia (acquired by Encoda), Blue Order (acquired by Avid), Konan, and Omnibus.

The archive manager 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 so that they could be placed into a VTR for playback. Now, with video servers and automated tape libraries, all multimedia can be accessed electronically via networked systems. This not only simplifies operations but also 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 be archived. Once archived, multimedia material is treated as a file and is stored by the archive manager. Archival functions are initiated by one of the following applications:

- Applications running on the video server, such as channel record, time delay, or VTR emulator
- An automation system that 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 relieves automation and MAM systems of storage tasks, which would be tedious and complex for them to execute. With a few simple commands, the archive manager can control and manage tape drives, library robotics, and the media database. It performs the necessary complex logic and issues the multitude of commands to the various devices, to perform commands that appear simple to the video server, such as the `archive`, `restore`, and `delete` commands.

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. The archive manager performs all the 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 takes care of the underlying storage and retrieval operations.

Managing 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 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 selects a tape on which to store the file and mounts the tape in an available tape drive; then it 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 and mounts the appropriate tape in an available tape drive; then it moves the file data from the tape to the video server where the application can use it.

Because the archive manager specializes in archive management, it can offer the following features that 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 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.
- **Defragmentation.** When objects are deleted from tape, tape fragmentation inevitably occurs. The archive manager can then operate defragmentation manually or automatically as a low-priority process.
- **Grouping.** The same kinds of content (such as news clips) can be grouped together on a separate set of physical media (for example, tapes).

Managing Content

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

- **Transcoding.** Software transcoding enables 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 50 Mb/sec, retrieving a two-hour movie from tape can take a minimum of 40 minutes even with the fastest tape. It is preferable to only restore a short section of the whole clip (for example, five minutes). The clip can be viewed in a much lower resolution for the selection of mark-in and mark-out points expressed in time code. 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 intervals with a specified frequency or according to special operations, such as archive.
- **Multiple restore.** Multiple restore enables a single content to restore to multiple destinations simultaneously. Usually this feature is used for mirrored transmission servers so that the automation can restore a clip to two servers in a single command.

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 of a relational database. This database contains information needed by the archive manager to move video material between servers and archives, and 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 in 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. 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 or middleware package that provides a level of control over the library that would be tedious and cumbersome for either the video server application or the automation system 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, and cost-effective storage management, archival, and retrieval services are the framework of that foundation. As technology advances and disk capacity become greater at a lower cost, so too will data tape capacity and tape transport data rates increase at a 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. Integrating digital storage with legacy video applications and devices is allowing any content to be accessible anytime, anywhere in the broadcast facility.

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

blogs.oracle.com/oracle
facebook.com/oracle
twitter.com/oracle
oracle.com

Hardware and Software, Engineered to Work Together

Digital Video Archiving: The Evolving Reality of Any Content, Anywhere, Anytime
Updated May 2015

Oracle is committed to developing practices and products that help protect the environment