Archiving in Media and Entertainment Market:

Strategies and Best Practices

DigitalFilm Tree

About DigitalFilm Tree

DigitalFilm Tree is a research and development based production, post, creative, and consulting environment. Since 1998, DigitalFilm Tree has played a definitive role in bringing change to the post-production industry by pioneering innovative tools and workflows. Clients include television, motion picture, independent, and first time filmmakers alike. Consulting clients are as diverse as today's media landscape. Services include production, post, consulting, workflow design, and creative. For more information: www.digitalfilmtree.com

Summary

This whitepaper focuses on the current growth in the media and entertainment market within the digital domain. It overviews the quantity of media generated, what the client is doing to store that media, and what potential steps the client is taking for archiving. The white paper provides thought leadership around the strategies behind these topics, with some technical discussion.

SECTION I

Introduction

The past twenty years have seen a huge growth in the area of media and entertainment. Starting with the availability of "digitizing" analog imagery, and later the introduction of new digital acquisition formats, the world has seen a phenomenal growth in the sources and the output of media and entertainment.

The average consumer has new and incredible choices in the age of the "digital lifestyle." No longer is a consumer required to drive to a store and buy physical media for music and movies. As of January 2009, Apple's iTunes Store has sold over 6 billion downloadable songs since it opened in 2003. Netflix on average mails 2 Million movies per day out to its 10.6 million subscribers.

In the pre-iTunes days, those kinds of numbers would be the envy of every music and film chain outlet across the globe.

At the same time, the amount of content created in the professional sector of the media and entertainment space has significantly accelerated. The last two decades have seen the emergence of affordable digital capture devices, non-linear editing systems, special effects software and computer hardware that has allowed for a true democratic vision of content creation to come true.

The digital storage marketplace has responded to this opportunity and the storage density of devices has increased dramatically, while at the same time the prices have plunged.

The one remaining frontier that still remains to be conquered is the one of archiving.

In the age of analog, archiving was a straight forward process. The medium at hand, film or tape, after being used would be saved in vaults where temperature and humidity would be carefully controlled. Every few years the materials would be retrieved and the archives "refreshed" by cleaning and restoration.

Moving into the new world of digital media and entertainment, the archiving process has not been solidified and still poses questions and opportunities. There are numerous options, but much remains unclear to most professionals seeking to make informed choices about their archiving strategies.

This whitepaper will present a brief overview of some of the strategies and ideas in the digital archiving solutions marketplace.

The analog age

In the old age of analog media and entertainment, film was the highest resolution format. It was also the final archiving format for materials that originated on film. Film negatives were archived in temperature and humidity controlled vaults for the future. Next in line was the video revolution. Videotapes became a dominant form of media capture and storage. At the end of a TV show that originated on tape, the final edited master and the raw footage was stored in controlled environment tape vaults. Anytime access was needed to these archives, it was simply a matter of retrieving the tape and playing it back in the VTR.

Next came the digital revolution and the rules changed significantly. Digital media, due to its inherent nature needs to be captured and manipulated onto disk drives. Despite steep declines in disk drive prices, drive space is always in short supply. Along with the world going digital, the rules of archiving changed as well.

Current state of media and entertainment market

There are huge and sweeping changes occurring within the media and entertainment market. The ability of an average desktop computer to generate digital content has grown dramatically. All areas of media and entertainment: acquisition, editing, conversion and distribution have gone digital and are expected to have explosive growth rates in the coming years. According to T.M Coughlin's, 2007, Entertainment Content Creation Digital Storage Report, the annual capacity for creation of professional moving image content will grow from 976 Petabytes in 2008 to over 2,441 Petabytes by 2012. (A Petabyte is equal to 1024 Terabytes.)

A major area of growth for the media and entertainment market is the Internet. Studies estimate that online content will grow 600% by the year 2012. There is a movement towards set-top hardware appliances that allow consumers to view movies and TV content via their computer, through websites like Hulu. User generated content services, such as YouTube, have exploded and continue to grow exponentially.

All of this growth creates an insatiable need for media-savvy storage solutions. Even with the quantum leaps in storage density and an inverse drop in price, there exists a huge need for reliable and affordable storage. At the same time archiving of digital content has languished in terms of technology and strategies. Most archiving technologies that have existed are seen as cumbersome, even by the facilities employing them on a daily basis.

The following are some of the key areas that have experienced this recent growth:

Film

According to the Motion Picture Association of America (MPAA), Hollywood produced 503 films in 2008. Approximately 20-30 of those major motion pictures were shot digitally, using HD cameras such as Panavision's Genesis. This number is expected to increase. With the advent of new affordable cameras such as the Red One, the independent market has also taken to creating "digitally originated" films. 16mm and Super 16mm, which previously used to be the standard format for independents, have fallen out of style overnight. The ease of using a Red-style digital camera, whose files can then be viewed, edited and output from any laptop, is too compelling to resist. None of the costs associated with 16mm film processing and prints are incurred in the digital production. Currently, the Red One has sold over 5,000 units. That is a phenomenal number for a camera that claims to provide 4K output, which requires massive 4K data to be stored and even archived somehow.

Even today films that are shot on celluloid film technology are then "scanned" to create Digital Intermediates (DI). These DI's are often edited and finished in the digital domain. Later these digital files are recorded back to celluloid via laser recorders for final theatrical distribution. The Academy of Motion Picture Arts and Sciences, expects that celluloid film will be replaced by an all-digital process, but the timing of that replacement is unknown.

As more and more film theatres convert to digital projection, the need for celluloid distribution copies will also diminish. According to 2008 statistics by the MPAA, there are a total of 8,614 digital screens worldwide, of which 64% or over 5,400 are located within the U.S. Just within the last year there was a 33% increase in the number of digital screens. And this number will continue to grow. As the need for digital release prints increases, there is a growing need for digital storage and transportation within the theatrical domain. Instead of transporting film-cans to theatres, disk drives are now physically transported. Archiving of these digital release prints is also a need that can be expected to grow in parallel.

The first of the non-linear editing software emerged in the early 1980's and today applications such as Apple's Final Cut Pro have made editing and handling possible of all types of digital media; consumer formats, HD, film and even 2K and 4K files.

With the entire film sector looking to change to digital, the issue of archiving has become paramount. No longer can the editing or post facility simply stock away their film negatives or the edit masters. More often than not what exists at the very end of a long digital process within these facilities is a digital file. There is the need to archive this final print digitally and with the assurance that the materials will be retrievable and intact in the future.

As a point of reference, a detailed report issued by the Academy of Motion Picture Arts and Sciences titled, "The Digital Dilemma" looks at the issue of archiving for the film industry and is a must read for anyone involved in digital film content creation and management. The report overviews the growth of digital content, specifically in the film industry and evaluates the "digital dilemma" of archiving for the future. From a film industry perspective Hollywood faces an archiving challenge and according to the report the key solution has yet to materialize.

Television

For the television industry, converting to High Definition (HD) as required by the Federal Communications Commission (FCC), as well as simple economics, has created a vast incentive to switch to digital content creation. More and more TV shows are being shot digitally and working in the digital domain for the small screen has proven more viable and economical. The new channels of distribution, such as online locations for TV shows, have also driven this migration to digital media. Numerous TV channels have created their own online distribution and marketing outlets for consumers and continue to look for additional revenue generating opportunities to grow their business.

As an example, NBC-Universal CEO, Jeff Zucker, recently said that NBC represented over 40% of the market share on the video side of Apple's iTunes Store. According to available statistics and market data, by October of 2008, Apple had sold over 200 million TV episodes via the Store.

For Television, archiving has always consisted of saving videotapes in a tape vault with controlled temperature and humidity. This equation is challenged in the age of digital content.

Digital Media and Imaging

Numerous other developments in the imaging area have also created a huge pool for digital media content. Photography was one of the earliest technologies to go digital and this has resulted in a massive growth of content in still imaging.

To use just one example, Snapfish is the world's number one online photo service. The company has over 70 million members in more than 20 countries and houses over 5 billion unique photos. Within the year Snapfish expects to have over 10 Petabytes of storage.

The mobile phone market has also impacted the media and entertainment markets by creating yet another viable channel for which content has to be created from the original media. Yet again, for obvious reasons this digital content needs to find storage and archiving space.

One key benefit of digitally acquired content is that repurposing it is much easier and more affordable. It is far more efficient and cost effective to create digital cinema "prints", internet-streaming copies, Blu-ray DVDs and mobile versions of a digitally originated film, than a celluloid one.

As more and more content is created, it needs to find storage. And with the value placed on these assets, it needs to find a proper archiving technology and strategy.

All of these recent trends and statistics point to an entirely digital future. Whereas digital acquisition, storage, processing and creation have taken quantum leaps in the recent past, archiving remains a challenge, as we'll see in the next few sections.

SECTION II

Media and Entertainment Industry growth in digital content

The Media and Entertainment Industry is one of the largest sectors of the US economy. According to a report by Deliotte Touche Tohmatsu, about $950 billion was spent by M&E companies in 2006. According to the same report this number was expected to grow by 38% over the next five years to a total of $1.3 trillion, despite a negative downturn in the economy.

No question that a large factor in the huge growth predicted has been the emergence of digital media in the M&E space.

A few trends in the media and entertainment industry have contributed to the growth of digital content:

Changing concepts of saving content

It is easy to miss the fact that when talking from a content creator's point of view, the challenge of archiving is it involves much more than just the final product. To put it into perspective, for a final program or feature film of 2 hours, the raw media involved could be hundreds of hours. This equation is especially steep depending on the nature of the final product. For example, the ratio of raw footage versus final product for a reality show or nature program is much higher than in the case of a scripted, narrative show. Of course, if the narrative, scripted show is heavy on special effects, or is an animated program, the equation goes up again.

Historically, with analog film, not all takes were "printed". Many of the takes never even made it to the lab for processing. In the age of digital there are never any discarded takes. Everything is available. That in itself creates a need for archiving and preservation in a cost effective manner.

Producers are increasingly choosing to save all raw content for their shows. The reasons are obvious. The countless channels of distribution and outlets, from the Internet to mobile, allow for endless repurposing of content. Ancillary footage, extras, outtakes and the growing trend of "Director's Cuts" have all led to a new sense of posterity for all material created in the making of a final program.

From an archiving perspective, this trend creates a large opportunity to bring additional revenue vehicles and a challenge with the magnitude of information that needs to be stored and managed.

Growth of high resolutions (HD, Red, 4K etc)

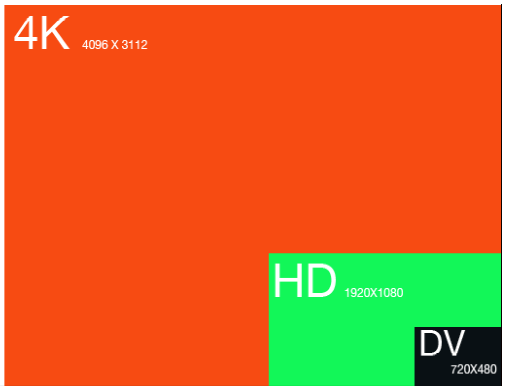
One of the causes of a recent surge in digital content has been the emergence of high-resolution formats. For example, a standard HD frame (1920 X 1080 pixels) is over six-times the size of a Standard Definition video (720 X 486 Pixels.) Hence the data size is proportionally larger.

For a few reasons within the film and TV industry 4K digital acquisition is considered the most desirable and ideal digital acquisition format. There is some technical and scientific basis for this, but from a layman's perspective, certain format sizes have always been dominant in the industry and 4K is twice as much as a 2K frame. A 2K frame is slightly larger than High Definition size. There are times where 6K scans are utilized for Vista Vision and other large format films, but it is agreed that 6K is simply too cumbersome.

A standard digital 4K frame is 4096 X 2048 pixels. (A 4K scan of a 35mm celluloid frame for a Digital Intermediate is often 4096 X 3112 pixels). Compared to HD's 1920 X 1080, or DV's 720 X 480, a 4K frame is a much higher data rate. 4K data is also commonly stored at a 4:4:4 sampling, creating much more data than a 4:2:2 sampling found in many formats.

The terms 4:4:4 and 4:2:2 are used as brightness versus color sampling ratios. Where as in a 4:2:2 file the color (which is represented by the last two digits) is sampled at half the rate as the brightness. At a 4:4:4 sampling rate all channels are sampled equally, discarding no color information. All other things being equal 4:4:4 is much more desirable.

New cameras like Sony's F35, the Genesis from Panavision, and even the Red One allow high resolution digital captures, approaching 4K resolution. The 4K data rate is quite high. One second of a 24fps 4K file will need approximately 1.2 GB of storage! Accordingly, archiving of 4K data has much higher needs than any other formats common in the last few years.



The relative frame size of 4K is seen here compared to an HD and DV frame.

Changing trends in industry

Various other trends in the industry have also increased the amount of content and media, which require storage and archiving. Clearly, the rise of high-resolution formats is one reason. Another is the changing nature of the industry itself. For example, the rise of Reality Shows has created a very large increase in the amount of raw footage being acquired for each show.

Paul Raudonis, director of post production at Bunim-Murray, one of the largest creators of reality TV shows, said in a recent bNet article by Claudia Kienzle that for one of their reality shows they shoot with multiple cameras for 12-15 hours a day, 7 days a week, generating 30-50 hours of footage. This amount later gets edited to 5 one-hour weekly shows. He also stated that the show has a starting raw to edit ratio of 200:1 and often requires 50 editors working in two shifts to come to its final shape.

That amount of raw media would be simply unheard of in the days before reality TV. More often than not all of the raw materials are saved in archives on their original tape formats.

In many cases such large amounts of media is ingested into disk storage at low resolution, offline storage. But in many cases facilities have moved away from offline due to the availability of very high-density storage systems at much more affordable prices than a few years ago. Another reason for moving away from offline is also the recapture issues encountered during online archiving, as well as the time needed for it. The availability of new compression engines such as the Apple Pro Res and Avid's DxNs has also blurred the line between offline and online. These high quality compression schemes allow editors to output in online form what is essentially an offline data-rate ingest.

In many cases, even when there's no particular need for it, shows are choosing to capture at RAW uncompressed quality. The difference in quality between raw and compressed (for instance at 2:1) is mostly negligible when it comes to broadcast viewing, due to all of the compression done during the broadcast path. However, the storage requirements for raw can be many times more. Most TV outlets have no need for a raw capture for their final broadcast air masters, but it is done anyway, in the interest of "future-proofing" or to simply start at the highest quality possible. This has also contributed to another significant factor in the increase in digital content, which later has to find a place to be saved and archived.

Given the incredible amount of content being created today, it is quite interesting to note that Hollywood still has not settled on a viable archiving solution.

In 2007, the Science and Technology Council of the Academy of Motion Picture Arts and Sciences issued an 80 page report entitled "The Digital Dilemma: Strategic Issues in archiving and accessing digital motion picture materials." This report is a comprehensive and in-depth review of the archiving universe from the Academy's vantage point.

Essentially, the report says, that the Hollywood industry has not found an archiving solution that meets their criteria or long term needs.

The report further says that technologies such as data-tape are widely being used, but most large-scale archive facilities that desperately need a solution are taking a wait and see approach. In the meantime Hollywood is moving fast and furious into the digital arena.

SECTION III

Various approaches to saving content

Saving and archiving information has a long history that dates back to ancient times. In this context however, we are concerned with recent trends within the media and entertainment industry. Saving content is about as critical as producing it. Consider the assets for a major motion picture and the financial investment they represent to the company. The copyright value of much of the materials being generated in the US film market serves to illustrate the point. In 2008, MPAA films in the US grossed $9.79 Billion domestically and $28.1 Billion worldwide. With such a high value attached to these digital assets, saving and archiving them is a top priority.

In the media and entertainment industry there are a few loosely defined approaches to saving content:

Saving all

This is considered the safest choice of all. Saving all has become quite dominant in an industry that can spin off DVDs, specials, web and mobile content from the same, original source footage. The key benefit here is that no decision maker or committee is needed to decide what to save or what to discard. Everything is saved. This may be practical in the case of a feature film, but not in the case of, say, C-SPAN. This approach of course is expensive, both in terms of time and money.

Saving limited portion of data

The drawback in this approach is simply the decision making process. Who gets to decide and how the decision is made. Of course the advantage here is the savings of space and cost involved. However the cost may be offset by the time and schedule it takes to make these decisions. In some cases there is always the fear of discarding something that is then needed later. In general, facilities tend to discard only the items that can later be recreated. These could be final renders of assets, which if needed in the future can be re-rendered from the saved project files.

Saving essentials

When saving essentials, for example in the case of a documentary, which originated on digital videotape, the producers may decide to delete all captured media from the disk. Of course, the original tape is saved in an archive for recapture later. The "essentials" in this case may be the project file for the non-linear editing system, which can be used later for recapture of material from tape.

Various formats for archiving in the M&E space

Hard disks

Hard disk drives, or "magnetic hard drives", have shown a phenomenal growth in storage capacity in the last two decades. Their prices have fallen proportionally. Early hard drives tended to be as small as 5 MB at a cost of hundreds, or even thousands of dollars. Today a 1 TB hard drive can be purchased for as little as $120.

The key limitation with hard drives as an archive medium is that their life expectancy is expected to be around 5-7 years. And more importantly hard drive mechanisms are designed to be powered up and not remain dormant on an archive shelf. As recently as 5 years ago, this lack of long-term reliability made hard disks unsuitable for archiving. However, newer emerging disk-based archiving schemes have shown great promise, as will be discussed later in this paper.

Data Tape

Also known as LTO (Linear Open-Tape) or DLT (Digital Linear Tape) standards, the data tape archive is the most widely used and accepted archive format in the media and entertainment space. LTO, a jointly developed open standard by HP, IBM and Quantum is by far the most widely used. The latest iteration of the LTO4 standard allows for over 800 GB per cartridge and has a transfer rate of 120 MB/sec.

The lifetime for an LTO tape is assumed to be around 10 years. However, the LTO development has often been backwards compatible in its technology and could extend that lifetime slightly.

As much as the manufacturers would like to present the LTO as an "open" format, incompatibilities exist even between the three joint developers of the LTO standard: HP, IBM and Quantum. For example, on the LTO1 standard, tapes written on the IBM cannot be read on the HP and the tapes written on the HP can only be read at half speed on the IBM.

Videotape

Videotapes have proven to be a surprisingly viable and robust form of archive format in the media and entertainment industries. The HDCAM-SR format from Sony and the D5 format from Panasonic have become the dominant form of videotape in the high-end film and post industries.

Under the proper conditions the lifetime of videotape is around 10-15 years.

Of course, when it comes to restoration, videotape is the simplest solution. The "restore" (or recapture) happens in real time, a rarity otherwise in any restore operation.

Optical Disk Storage

Optical CD-ROM disks have been available since the early 1980's. An average CD-ROM stores 700 MB of data. Later, DVDs came into use and allowed for up to 9 GB of storage.

Today is the age of Blu-ray discs, which can hold up to 50 GB. Future specs for Blu-ray allow for over 400 GB of storage per disc. The cost for Blu-ray storage is about $1/GB. There have been very few comprehensive studies done on the lifetimes of optical discs but it is generally assumed to be around 15-20 years, when stored in ideal conditions. Evidence exists however that optical disks can degrade in some cases within 5-7 years.

The future potential for optical disks is by no means exhausted. With cutting edge research much work is going into multi-layering and using lasers of various wavelengths to pack more and more information into the optical disks. For instance, in April of 2009, The New York Times reported that engineers from GE have created holographic optical disks, able to hold 500 GB of data, or over 10 Blu-ray disks worth of materials.

Online, near-line and offline storage

Within the archive domain online, near-line and offline are loose definitions that can sometimes overlap.

Online storage

Data is immediately accessible. In most cases the online storage would be the main disk of the host computer.

Near-line storage

A term for data that is stored in an intermediate disk storage. It is still easily accessible, but the performance requirements are not as fast as online. An example of near-line storage is a host of users connected to a central JBOD (Just a Bunch of Disks). The JBOD is not required to have the top performance of the online storage disks, since it is only used for storage and retrieval of data and not used for direct off-disk playback in most cases.

Off-line storage

Data is backed up on tape and put away in an archive. Data is available, but much slower than online or near-line.

In general, online and near-line storage tends to be disk based. Whereas offline storage is data-tape based.

Evaluating archive choices

There are numerous factors to be considered when designing an archive strategy. Some of these are:

Cost per gigabyte

A key evaluation for any archive system is its cost per gigabyte of storage. Many leading studies and manufacturers' "fact" sheets often misrepresent this evaluation. For example, the disk-archive storage proponents will often list the price of just the storage drive itself (as opposed to a fully working disk system with controller cost, etc). The tape-archive proponents, when listing the price of a tape drive and cartridge, will often fail to mention that virtually no data-tape library will be using just a single drive! The cost per gigabyte rises steeply when one also considers the personnel needed for tape-drive archive management.

Power usage

Power usage is often a hidden cost for a facility that may not be so obvious up front. This cost not only includes the power that the archive consumes, but also the power required for cooling the electronics. In a large archive environment, the cost of properly designed air-conditioning can often exceed the cost of power consumed by the unit itself.

Transfer Speeds

Manufacturers clearly state the transfer speeds of their systems. And in many cases they do distinguish the faster compressed rates from the slower uncompressed rates. However, when storing imagery (and this is most often the case with the media and entertainment industry) the slower uncompressed rates should be used for estimation. The faster, compressed speeds are not suitable for imagery archives.

At other times real world transfer rates simply don't match advertised numbers. For example a USGS (U.S Geological Survey) study found that a real world transfer rate for an LT04 was 108.46 MB/sec (90.4% of its advertised uncompressed transfer rate figure of 120 MB/sec).

Access speed

Another key criterion for any archiving strategy is the latency time between request and access of data. In many facilities this latency time may need to be within seconds, while in some others a latency time of hours or even days might be considered acceptable. In many ways this single issue can be the biggest deciding factor between a data tape-based archive versus a disk-based archive. In any tape-based scheme the access speed tends to be within hours while disk-based systems provide instant access.

Lifetime of media

When designing an archive the lifetime of media must be carefully considered. Various studies have been done to determine the real-world lifetimes of media and these are often significantly lower than the industry's determinations for their own media formats. As part of the overall cost estimation, one must consider the cost of migrating the archive to a newer set of media within the safety of the media's real-world lifetime estimation period.

Lifetime of technology

Of course, there's always the concern that certain technology itself may get obsolete. For the moment the LTO-data tape family has shown good promise into the future. Future iterations of the LTO spec are generally compatible and the format has enjoyed longstanding support from manufacturers.

Pros and cons of various archiving methods

All archive systems have a few advantages and some disadvantage:

Data Tape Based Systems

Data tape, or LTO, is considered the safest choice overall for long-term archiving. One main issue that often creates a hurdle is the transfer rate of the tape system. Other disadvantages of tape-based systems are around the issue of data integrity and the wear and tear on the physical tape itself. Tape systems are typically subjected to high degrees of physical wear and tear during the relocation of the tape during the write and read processes. This is the primary cause of data damage.

Disk-Based Systems

The emergence of near-line disk-based systems has provided an alternative to the long-term tape archive systems. The cost factor, which was the main hurdle for these systems has been managed to compete successfully with tape-based systems. RAID-based disk systems are more geared for data protection than a tape-based system. The fast seek times on disk are also an advantage. For any facility evaluating these systems, their biggest challenge will be to factor the future cost of growth they may experience.

Deciding on archive methods

Perhaps the most useful task a facility can implement during the design of its digital archive is to carefully study and categorize its assets. The key division in the categories must be based on the level of access that the media requires.

For example, if the media needs to be rarely accessed, then perhaps it is ideally suited for an LTO-based archive library, or "off-line" storage.

If on the other hand the assets need to be available on a semi-regular basis, then they are ideally suited for a near-line type, or disk-based archive, environment.

Any media and assets that are currently needed on a daily basis, of course, need to be available "online" or on the main working drives of the systems.

Beyond the basic categories, depending on the workflow of the facility, there needs to be in place an automated archive strategy. This archive strategy should be designed to automatically move items from one mode or archive to another. For instance, it can be mandated that any materials that have been stored "near-line" may be backed up to LTO tape for "off-line" storage after a certain period of inactivity. In this manner space can be freed up on the near-line archives for more materials.

SECTION IV

Some hurdles in archiving

Numerous hurdles exist for clients to navigate through as they walk the thicket of digital archiving solutions. Just a few are listed below:

Slower than ideal transfer speeds

Often times in an archiving situation it is the slow transfer speed that holds up work and lowers productivity, more so than the density of the storage media.

For example, at the LT04 transfer rate of 120 MB/sec it will take just over 7 hours to backup or restore a 3 TB High Definition feature!

It is important to design an archive solution with this limitation in mind. Solutions exist by creation of parallel schemes, but add to the cost significantly. For example, a two-drive LT04 system would cut that time of 7 hours in half, and so on.

Error detection concerns

Data tape drive systems reduce data loss through CRC (Cyclic Redundancy Check), which allows data to be read after a read error. As critical as a CRC is during a write cycle, it is more important during a read pass, because of the assumption that media will degrade over time.

Many archivists also shy away from helical scan based tape archive systems, due to their unreliability. Tape wear is also a concern. A single read/write on the LT04 can be 56 or more passes, due to the various tracks and channels being used.

Lifetimes of media is never as advertised

Various real world studies have shown that the lifetime of media as advertised by manufacturers rarely holds up in the real world. For example, various MTBF (Mean Time Between Failures) issued by manufacturers for hard disks have fallen short of their real world performance. Any archive strategy should plan conservatively around the expected lifetime of the media.

Lifetimes of archive technology are an unknown

Another valid concern that arises is obsolescence of the archive technology itself. For example, there is no guarantee that a DVD drive able to read today's archived DVD may be found 10 years from now. Just consider having to look for a floppy drive today to read a floppy disk created in 1990.

Hence, any archive strategy will assume an entire technology refresh rate of 5-7 years to prevent being outdated.

One of the most common "migration" strategies for example is for facilities to keep copying old LTO tape based archives to newer LTO standard tapes, as they emerge. For example, in 2003 the USGS (U.S Geological Survey) migrated 50,000 tapes to the newer LTO2 tapes over a period of 5.5 months.

Also as a precaution method it is ideal to use an open source file format (such a tarball) for archiving in case the archive software is no longer around in 5-7 years.

"Digital Origination" masters are nothing more than videotapes or LTOs (for now)

At the moment the archives for digitally originated materials consist of either HDCAM-SR tapes, as mentioned above, or in many cases on LTO data archives. When designing an LTO-based archive it is imperative that allowances be made for future migration. Almost all major LTO-based archives assume that within 7-10 years the data tapes would have to be migrated to a newer LTO tape format.

Bit for bit confirmation: A key for any viable digital film archive

A key requirement for any image archive solution is the robustness of its CRC (Cyclic Redundancy Check) error detection and correction. CRC ensures that even if the data archive tape suffers physical damage, the mechanism will detect the error and correct it via the extra bits encoded on the tape for just that very purpose.

It needs to be emphasized that for any imaging archive, this is an absolute and firm requirement. Missing bits in an image are just that. They will never be recreated, unless there has been a valid CRC in place within the archive system.

For example: The total capacity of an LTO4 tape is 850 GB of user data. 50 GB of this total amount is reserved to cover any media defects or degradation that occurs as the tapes age.

In addition to the user data the LTO format requires that two levels of Error Correction Codes (ECC) be generated over the user data allowing any lost data to be rebuilt. This ECC overhead, coupled with additional CRC error prediction and data recovery fields, results in the LTO4 format being 71.6% efficient.

In an ideal situation two exact copies would be made on the LTO4 archive, and stored separately. This allows for redundancy and for any missing data to be recreated.

Of course, it is ideal that the data archive system has an extremely low BER (Bit Error Rate) to begin with, so that very little correction or recovery is needed. For example, HP's Ultrium LTO-3 drives have an uncorrected BER of 10^{17}. Essentially that means there is a 1 in 10^{17} chance the drive will encounter an error it is unable to fix. In other words for every 100,000,000,000,000,000 bits transmitted there will be one bit in error.

"Refreshing" digital archives

The concept of refreshing archives comes from traditional film storage. Periodically the archived film negatives are retrieved, unwound, cleaned and rewound back for storage. This process extends the lifetime of the film in archive but it is cost intensive from labor and materials point of view.

A similar strategy needs to be applied to digital archives. If a facility has chosen an LTO-based data tape archive for its assets, for example, then periodically the tapes need to be inspected and refreshed.

The LTO-tapes should also be "migrated" to newer tapes depending on the assumed lifetime. And as a further long-term strategy, the entire tape library should be migrated to a newer LTO format as it emerges.

Refreshing digital archives in this manner prevents the general deterioration of the physical tape as well as keeps the entire technology from being outdated. It is important to keep the "refreshing" costs in mind when planning an archive for the future.

The assumed lifetime of videotape, for instance, stored in an ideal environment, is 10 years, or more in some cases. It is quite economical in many scenarios to store HDCAM-SR, or D5 masters as the final "archives". Of course, the tapes must be "refreshed" by winding and rewinding, every few months. And a plan of action put in place to "migrate" the materials to a new tape every 5-7 years.

Current archiving choices for the media and entertainment industry

Currently, the choices for archiving in the media and entertainment industry break down into two groups.

For any long term, off-line archiving the best choice right now is the LTO technology. Given the overall advantages of LTO, data tape archiving is still considered the most reliable. Future migrations of the LTO data tapes as a refreshing strategy should be part of every LTO-based library.

The second choice, which has become quite popular in the media and entertainment space, is the disk-based near-line archiving.

Advantages of disk-based near-line archives

A careful study of the tape-based archiving system comes up with a single shortcoming; data integrity. Even with all the fancy algorithms and gyrations, it is almost impossible to fully confirm the viability of all bits laid down to tape. At the same time, the failure of any one tape cartridge during restore can spell doom for the entire operation causing delays in schedules and incurring additional cost.

With a disk-based RAID system, data integrity is assured. However, a large-scale disk-based near-line system has its own big disadvantage: cost. Data tape wins hands down over disk–based systems when it comes to cost. However, cost is only one part of the equation to consider and should be evaluated with other factors.

In an ideal world, an archive solution would present all the advantages of a disk-based system, while maintaining the price point advantage of LTO.

That is precisely what the new generation of near-line archive systems offer. For example, the new HP X9000 Storage System is an ideal case in point. It offers all the advantages of a disk-based system, while matching the price point of a LTO archive solution.

It is important to bear in mind that near-line archive solutions such as HP's X9000 Storage System are much more than just a few cheap RAID disks snapped together. The disk system is optimized for archiving, and the software manager is especially designed to ensure data integrity. In addition, these systems are vastly scalable and their price point is well designed to match that of a data tape archive.

SECTION V

Current strategies for archives

As mentioned earlier, the media and entertainment industry is experiencing unprecedented growth. The industry's reliance on archiving solutions is growing proportionally. Currently, given the fast changing pace of technology, customers are adopting a few strategies:

Develop smart strategies

Facilities are developing protocols and smart strategies for reducing the amount of material that is slated for archiving.

If a post facility relies heavily on videotapes for example, they may choose to delete all of the digital media that was acquired through capture from those videotapes.

The logic here is that since all digital media captured from the tape is time coded, and those time code references are stored within the project file, which was used for capture, there is no real need to archive all of the media. Simply saving the videotapes in a traditional videotape archive and the project file on a data tape archive is enough.

The entire project may be brought online in the future by restoring the project file from the data tape and recapturing from the videotapes.

Keep archives as they are

For now, many facilities have simply chosen to stay with traditional archiving methods, for example saving videotapes, if they contain the original footage. In many cases the Sony HDCAM-SR format has become the de facto archive method. This is especially the case if the program to be saved was digitally originated.

And in the case of film, facilities are saving the negatives, or separation masters, as has been the tradition in film archiving for many, many years.

Decide on the best method, for now

Another approach that some facilities have chosen is to start an archive strategy based on what seems to be the best choice for them.

In most cases the facilities may distinguish between a long term archive and short term one. In a long term archive, the facility may not expect to access the data within the next 5-7 years. And in the short term, the materials may be needed on a daily or a weekly basis.

Many facilities have implemented a long-term program of saving materials on LTO data tapes. Bear in mind that such a strategy without a proper "refreshing" plan for the future may be far more risky than even a traditional videotape archive.

In most cases the facilities may distinguish between a long term archive and short term one. In a long term archive, the facility may not expect to access the data within the next 5-7 years. And in the short term, the materials may be needed on a daily or a weekly basis.

Many facilities have implemented a long-term program of saving materials on LTO data tapes. Bear in mind that such a strategy without a proper "refreshing" plan for the future may be far more risky than even a traditional videotape archive.

At the same time the facility may choose a short-term archive solution that is disk-based.

When deciding on any archive strategy a balancing act needs to play out between cost, time and longevity. How much does it cost? How much time will it consume? And how long will it last? There will always need to be a tradeoff in that equation where the cost may be low but the longevity may be less. In other cases the cost may be high but it may require a lot less time and perhaps even allow for longevity.

Conclusion

The world of media and entertainment has become too diverse for a single solution. Just as workflows have quickly multiplied, so have archiving solutions. Each workflow may have its own archive solution. And even each phase of the workflow may have its own archive strategy.

Short term and long-term recommendations

In the short term, perhaps the best archive strategy is one that takes into account the original raw materials, and a proper consideration of the lifetime of the format chosen. For example, with the increase in digital-originated masters, many facilities are choosing to store their materials on digital videotape formats such as HDCAM-SR, taking into account the limitations and the lifetime of the tape format.

Any materials that may not be needed for sometime are almost always best backed up to LTO tape. Ideally, two copies must be created, separated by location. More than one copy of the LTO allows a better chance of restoration if there are bit confirmation issues during restore.

In the long term, a facility needs to organize and understand the nature of its assets. Any long-term archive solution, if implemented, must take into account the limitations imposed by the lifetime of the archive format and the archive technology itself. As a long-term archive strategy, the cost and time involved for "refreshing" an archive must be taken into account.

SECTION VI

Best practices for Digital Archiving

When evaluating the list of IT vendors who support the media and entertainment industry, it is imperative to look at the whole picture of their capabilities and their experience. As technology evolves, are they visionaries and continue to contribute innovation to help build the longevity of technology for the digital era?

One of our main criteria for this search was to consider the entirety of offerings from the vendor. Archiving schemes do not exist in a vacuum. It is key that they seamlessly integrate within the overall infrastructure of the facility. It is also important that the vendor be a proven party within the storage solutions market, a key link to archiving.

We found that HP is without precedent in the industry for the full scope of its offerings within the media and entertainment space. Uniquely, HP provides a complete solution- including storage, servers, software and service- from a single vendor.

HP is one of the three key companies critical in the formation of the LTO standards and it has maintained its commitment and kept its promises for future development in that area. HP offers an extensive array of LTO drives and management software.

The latest offering from HP in the archive space is the X9000 Storage System. The X9000 is a disk-based, near-line archive solution that is extremely scalable and matches the price point of LTO-tape based archive solutions.

HP's breadth of portfolio in the area of storage and archiving allows for a complete and all encompassing solution for even the largest of clients. DreamWorks Animation Studios has used HP's portfolio of products to its benefit to meet the needs of a demanding production environment.

The following case study overviews DreamWorks' use of HP's latest offerings in the archive arena:

Best Practices in Action: DreamWorks Animation

Introduction

DreamWorks studios started out in 1994, with collaboration between Steven Spielberg, Jeffrey Katzenberg and David Geffen. In 2005 the studio was sold to Viacom, the owner of Paramount pictures. DreamWorks Animation (DWA) is a spin off from the original studios and is currently producing feature length animation films, distributed by Paramount Pictures. Some of the biggest successes from DWA are the Shrek series of films.

DWA's recent hit Kung Fu Panda, grossed over $626 Million worldwide, making it one of the most successful films of 2008, and the most successful DWA original film ever.

In a recent announcement DWA stated the sequel for Kung Fu Panda is slated for release in 2011 and will be produced in stereoscopic 3D technology.

Recent growth for DWA

DWA is experiencing an unprecedented growth in the amount of media it generates. One of the reasons for this growth is the increase in complexity in their animated features. According to Derek Chan, the Head of Digital Operations at DWA, the original Shrek, when finished at DWA occupied 6 TB of space. DWA's recent Monsters vs. Aliens, on the other hand, took up over 90 TB of disk space.

Reasons for growth at DWA include:

DWA's entry into 3D

Recently DWA has gone through a change where they have decided to create all of their animated films in stereoscopic 3D. This decision has created an almost doubling of the media generated with two images have to be created, one for each eye.

Increase in overall output

DWA has also announced an increase in their output. The studio is moving from creating two movies a year to creating five movies every two years. This increase in output has created a larger demand for storage, and hence archiving.

Anticipation of 4K resolution

Chan suggests that the studio is looking ahead to 4K resolution becoming standard across the board. Again, this is expected to create a quantum jump in the amount of media generated, and hence the storage and archiving requirements.

Strategies for saving content

As a general strategy for archiving at DWA, they classify media into three general categories. There is a standard policy for treating each of these three different categories:

Final Images

The set of final images that are the ones sent to film-out, digital release print, or DVD etc are all saved without reservation. The standard policy is to save all of these comprehensively. These final images are kept "online" and accessible. The main reason for keeping this media instantly accessible is the fact that many of DWA's films are sequel based and when working on the current sequel designers and animators often need to refer to the original film for design and other references.

It is important to point out that in the past these final images were archived to data-tape based archives. But DWA discovered over the years that the cost of archiving them to disk-based, online and near-line systems was worth it due to the instant access these systems provided. Explains Chan, "People who work in the entertainment industry, they are very used to more, in-the-moment, real-time decision making."

Source files

DWA describes their "source files" as materials that are not necessarily seen on the final output, but are files that are needed to create the final images. Source files may include models and textures, etc. The source files for older releases that are not expected to have sequels may be archived to data-tape. But source files for films that have sequels in the making are always archived to disk-based online systems, where instant access is key.

Transient or "re-creatables"

These are files that are generated in the process of making final output, but are not the source material. They are intermediate files effectively used for the creation of the final output. And DWA could regenerate, or re-render them, if they need to in the process of creating the final image again. These re-creatable files are not saved.

Formats for Archiving

At DWA a mix of formats are used for archiving purposes. For older projects, which are not expected to have follow-up sequels, the films are backed up to data-tape archives. Examples of such films from DWA include Flushed Away and Over the Hedge. But for all other "in the works" projects, which may be expected to have sequels, the materials are kept more and more on disk-based, near-line archives. Chan feels that DWA is moving increasingly towards the newer disk-based archive solutions due to their low-cost and the instant access they provide.

Challenges in archiving for DWA

DWA, like any other major production facility in the media and entertainment space faces its very own set of challenges when it comes to archiving data:

Locating materials

Perhaps the biggest hurdle DWA faces in archiving is the ability to locate and access their data. Over the years DWA has discovered that in a data-tape archive an inordinate amount of time goes into locating a particular piece of media and then restoring it for access. To a greater extent, DWA has moved towards disk-based systems for archiving mainly due to their instant access. Chan describes the difference of time between tape and disk archive access as being between "days versus sub 10 seconds."

Access Speed

DWA's artists and animators require instant access to their archived materials. "Having data in archives is not really useful if you can't get access to it," explains Chan. DWA found that the access and speed times are quite debilitating when it came to tape-based archives.

Outdated software

DWA has faced the challenge that sometimes the software used to create certain images may be outdated. "If I pulled out an asset from the year 2000, can I run it through to a shot? I don't think we have the software to do that," explains Chan. He further explained that the outdated software not just includes the software used to create the final images, but often times the operating system that it can perform on as well. DWA has faced this challenge and is looking into "virtualization" where archives are stored with their accompanying software and even the operating system for a secure access in the future.

Outdated technology

Many times the platform or the hardware that was used to create the materials has gone out of date as well. With the speed of processor and technological development, this cycle can be as fast as 5-6 years. Again DWA is looking into various archiving systems based on "virtualization" that allow for future-proof access.

Power requirements

As DWA has moved more and more towards the disk-based near-line solutions it has faced the challenges of increased power usage. All disk-based systems do have the need to stay powered up, as the disks need to be kept "alive" and spinning. Data-tapes do not have such a high power requirement. However, DWA has decided that this increased cost in power demand is worth it due to the speed and access allowed by disk-based systems.

Strategies for planning against obsolesce

Refreshing archives

DWA refreshes its archives periodically to prevent obsolesce. The tape-based archives are migrated to newer formats every 5-6 years and the disk-based archives are migrated to newer systems about every 4-5 years. Chan further explains that often times the motivation for this refresh cycle comes from "disruptive changes." These disruptive changes may be a new archiving standard set by the Library of Congress or perhaps newer formats that have been accepted as standards.

Using Virtualization

DWA is looking ahead to creating 'virtual containers' for future proofing their archives. The idea behind these containers is that they are programs that emulate an entire system- from hardware all the way to the operating system to the application needed to create a certain asset. That way the assets are stored away with the application and the operating system needed to run it. In this way the future revisions of the operating system and software do not render the asset useless and it can be recreated many years into the future.

Tape vs. Disk: Choosing HP's X9000 Storage System

Over the last few years, for DWA, a natural move occurred from data-tape based archives to disk-based online and near-line solutions. The primary motivation for this shift was the attraction of archives that are instantly accessible. DWA's current archive is based on HP's disk-based X9000 Storage System near-line archive solution.

There were other points of comparison as well, which have helped cause this shift:

Labor Costs

DWA felt that the labor cost of a tape library was more than they'd like to take on for their archives. In many ways a tape library requires more personnel than a disk storage system. In the case of DWA, they have just three people managing all of the disk-based archives. A tape library with its tape cartridges and robots requires many more people to run and maintain.

Floor Space

DWA like any other large facility is always looking to maximize their floor space. In a growing company with many features on its plate it is critical for DWA to be able to expand and utilize its floor space efficiently. In term of archiving, this means using as little space as possible for as much data as can be archived. With this important criterion DWA saw HP's X9000 to be ideal in the density they packed into their size and area.

Flexibility

HP's Disk-based near-line solution allows for greater flexibility and expandability into the future. Chan explains, "I don't need to buy it all today; I also know that I have the capacity and the capability to grow more when I need it. We know that in our solution we can go up in terms of density by adding more of these capacity blocks."

Speed

When evaluating the disk-based HP X9000 system DWA was also impressed with the speed and access they provided. Tape-based archives due to their inherent nature cannot provide this type of access and speed."

Solution Strategy at DWA

As DWA has grown it has quickly recognized the value of the new disk-based near-line archive solutions. In contrast to tape-based archives these disk-based systems are instantaneous when it comes to providing access to archives. In the past, price has been the main hurdle for disk-based archival systems to overcome. The new X9000 system from HP is an example of the latest generation of near-line archive systems that addresses this issue successfully. HP's X9000, for example, is competitive with the cost of a tape-based system. At the same time it allows for huge growth by adding additional modules.

DWA's Chan says, "If I were to do it from scratch today, I would go with disk-based system." He credits HP's engineering team for having created a viable and ideal solution for DWA's archiving needs with the new X9000 system. He says, "We feel that the X9000 is a pretty good step into the network-attached storage world, and HP has plans that continue out as far as the eye can see in terms of new features and functionalities. We feel confident that this platform will have longevity. We'll have the features that we need."

Summary

The last two decades have seen an enormous growth in the amount of digital content. Storage needs have risen proportionally. This trend shows no end of slowing down. As the world's content is going digital, the concept of archiving has taken center stage. What to save and how to save it have become paramount questions in the area of media and entertainment markets. This paper has discussed various strategies and approaches to saving content that are being utilized in the media and entertainment industry While at the same time offering an overview of various technologies that are currently popular for archiving. Numerous hurdles remain, but some new and emerging technical approaches look promising and are being adopted, which perhaps points to the future of archiving.

References

The following studies and documents were referenced throughout this whitepaper for their facts and ideas.

• The Digital Dilemma
2007. The Science and Technology Council of the Academy of Motion Picture Arts and Sciences

• Digital Storage in the Post Production Market
July 2008, Coughlin Associates, Inc

• Archiving Movies in a Digital World
White Paper Jan 10, 2007. Sun Microsystems

• Digital Archiving Solutions for the Media and Entertainment Industry
White Paper February, 2008. Sun Microsystems

• Reliability modeling for long term digital preservation
Constantopoulos, Doerr, Petraki. Institute of Computer Science, Athens University

• Media and Entertainment: 2009 Industry Outlook
2009 Deloitte Touche Tohmatsu

• Archive Media Trade Study
2008, USGS (United States Geological Survey) Report ID: TSSC-00-02-4.2-0001

• Real-world data transfer results
19 Feb 2009, Jim Fancher- Deluxe Digital Media. HPA Tech Retreat 2009

• Storage: The year in review and looking ahead to 2009
2008, Rob Korbin-CTO, Integrated Media Technologies, Inc. HPA Tech Retreat 2008

• The future direction of optical data storage
2006, Richard G. Zech, Ph.D. Media Tech 2006 Long Beach.