EPFL Build Disk-based Video Archive for Legendary Montreux Jazz Festival

Leveraging the Amplidata Object Storage technology to build a live video archive

Data archiving has been the topic of many debates in the storage industry. Think about the many papers that have been written about compliance regulations for data archives or the numerous panels that discussed best practices for data migration. New topics are gaining popularity, however. While data archiving used to be considered as a dreadful, money consuming process, companies are starting to see the value of their data. Data archives need to become accessible resources, which opens a whole range of new discussions: tape vs. disk archives, latency vs. cost etc.

One organization that definitely understands the value of their "archive" is **Montreux Sounds**, which preserve the **Montreux Jazz Festival** archives. More than 5,000 hours of concerts have been recorded since the inception of the Montreux Jazz Festival in 1967 by the visionary Claude Nobs. Over 4,000 groups and artists have followed each other there to create a page in the international heritage of music. Having adopted leading edge technology in material for audio-visual recording at a very early stage, the Montreux Festival has accumulated a wealth of visual and audio recordings that is unique in its genre. This represents by far the largest collection of concerts recorded live to be brought together in one place. The greatest names in jazz appear on them and a great number of improvised jam sessions are extremely rare because of the combination of talents playing in them.

Since its beginnings in 1967, the Montreux Jazz Festival has systematically recorded the concerts, using a wide range of television formats. With the advent of new audio-visual technologies, the Festival began recording in high definition in 1991, and in 2010 the Festival experimented with 3D live recording. The result of this is an immense library of concerts in various formats and on various media. The physical deterioration and technological obsolescence of the audio-visual media – of which there are no backup copies – has prompted Montreux Sounds to find a solution to manage these media in the long term.

It is the musical and technological restoration of this inheritance that is at the origin of the creation of the **Montreux Sounds Digital Project**. In that perspective, Montreux Sounds set up a collaboration with the Swiss Federal Institute of Technology of Lausanne (EPFL) to "digitize" and valorize the over 5000 hours of video archive of memorable concerts at the past editions of the Festival. It is a cultural and educational project to safeguard this heritage for future generations and to emphasize its value in scientific and artistic projects. It is continuously financed by companies or private entities wishing to associate themselves to this unique initiative.

Step 1: securing the footage on tape, redundantly

The first phase of this mega-project was to copy all existing footage to digital tape. But before this long and meticulous process could be started, all the original recordings had to be identified and a reference database of their contents had to be created. In the past, tapes were not tagged with the terminology we use today. Some tapes even lacked any form of detectable information. The Festival captures a lot of information during

concerts. That metadata was crucial to the project as it contains valuable information about what happened during each event, as well as timing data that can be used to retrieve specific songs. Some tapes only contained hand-written information, which was also digitised.

After creating the inventory, the actual digitization process began by transferring the old tapes to a huge, redundant LTO tape library (two copies uncompressed, two copies compressed). This part of the project will be completed in two years. Today, about 20% of the footage has been processed. The digital tape archive will be a huge improvement compared to the hardly maintained collection of – often raw – footage. The library is progressively available internally at EPFL and at Monteux Sounds for projects such as the release of new concert dvd's, rockumentaries etc. Fortunately the Montreux Jazz organization recognized there is more potential for the archive.

Building a digital archive of live concert recordings is more than just copying the data to digital tape. Montreux Sounds and EPFL have designed an error-proof process that facilitates archiving approved assets for an indefinite time. One of the main challenges in this process is for the data transfer to keep up with the encoding. Each month, about 200 hours of video and 200 hours of audio are encoded. All concert videos are manually sampled to control parameters like chroma or white levels. For these and other post-processing tasks such as bit checking or video and audio control, 40TB of high-end post-processing storage is available. When manipulating large video files in batches, disks of high capacity and high throughput are crucial. Also, to ensure no data can ever be lost, the two copies must never be on the same location, which adds to the complexity of the process.

Step 2: creating an "Active" Archive

The Montreux Jazz Festival archives are a unique treasure, and without doubt one of the greatest musical documents of the last 45 years. Storing this treasure on digital tape only would make it cumbersome for people to access the data. As a matter of fact, the archive would hardly be accessible at all. Therefore, as a second phase of the project, EPFL is building a storage infrastructure that will enable the Festival to make the library more easily available for future developments.

Very few companies had experience with building active archives when the requirements for the Montreux Jazz Festival active archive were defined. For EPFL, the following were of key importance:

- Long life cycle ("a hundred year archive")
- Low energy consumption
- Immediate access to the data; low latency
- High security, protect access to media by unauthorized users
- High availability (built-in redundancy to prevent loss of data)
- Cost-efficient
- "Fast" and easy replication mechanisms (in case of loss of data)

The selection of requirements made the technology choice all but impossible. Especially the combination of low latency (rules out tape) with low cost and high availability was a challenge.

For the first phase of the project, LTO tapes were a logic choice: availability and cost efficiency could not be higher. Latency and management, however, were far from acceptable. The teams decided to wait for disk archive systems to become more affordable and expected that increased efforts to build green datacenters would lead to interesting options soon. Several EPFL labs were already involved in storage research projects so a solution had to present itself soon.

The solution would be Amplidata's AmpliStor, which offered all the benefits of disk at the cost of tape. AmpliStor's main benefit compared to tape was low latency and low management. The teams of Montreux Jazz and EPFL agreed that the innovative platform would enable them to make their dream come true. For the EPFL research labs, the project with Amplidata would serve as a reference platform to assess the relevance of a disk-based infrastructure to "run" an archive. Ultimately this would feed EPFL research labs with specific data and challenges for further developments.

The unique benefits of Amplidata's technology that convinced EPFL to choose the AmpliStor platform:

- Amplidata reduces the power usage with 90% compared to other disk systems. Power consumption is currently less than 4W/TB and future features such as powering down nodes or disks will bring the cost even closer to that of tape
- The platform dramatically reduces the latency to read data from the archive
- The Bitspread technology in AmpliStor remarkably reduces required disk capacity and gives a much higher availability than traditional RAID systems

Step 3: sustaining the archive on the long run

Archiving is a continuous process. Maintaining such an archive can only be done efficiently when it is well planned and well designed. The two main challenges are scalability, providing for growth, and hardware redundancy.

Scalability

The Montreux Jazz Festival will continue to record content that needs to be preserved. Also, each year, the quality of the recordings improves, which impacts the storage that is needed to store the footage. HD video needs five times the capacity of standard definition. 3-D video will further double that figure and there are already emerging video capturing technologies which will multiply capacity needs by four, eight and even sixteen in the coming years (based on current codecs).

Hardware redundancy

It is very important for the project to build an archive that can evolve as new technologies become available, so the system won't become obsolete in 10 years. Otherwise the whole effort would have to be started from scratch again. Tape systems have a lifetime of more than 30 years, for disk that is a lot less. Therefore it is important

to build a system that does not depend on the underlying hardware, with software that can evolve as well.

When equipment is made "End-of-Life", this means it will become increasingly difficult to have spare parts and support, as is the situation of analogue tapes today. We cannot afford to have such a problem ten years from now, as the archive will be multitudes larger than today's library.

By combining tape and disk, EPFL will be able to compare the weaknesses and strengths of both systems and design an optimal video archive infrastructure. Over the next five years, EPFL will address the sustainability problem and carefully evaluate the following challenges to preserve the archive on the long run:

- Grow the archive with new content
- Manage archiving technology shifts
- Operate the archive
- Secure the integrity of the archive
- Minimize operation costs
- Minimize costs of technology evolution and smooth this process

Architecture

The Montreux Sounds technicians have evaluated several disk based archiving systems to make most of the Montreux Jazz Festival content accessible online for display in Montreux Jazz Cafe's and for other purposes. Amplidata has been selected as the disk storage technology for this archive because of its high reliability, sufficient streaming throughput for content of all encoding qualities and extreme scalability which makes it a reliable and future proof foundation for this project. Also, a strong interest in "green" data center was key in the decision making process.

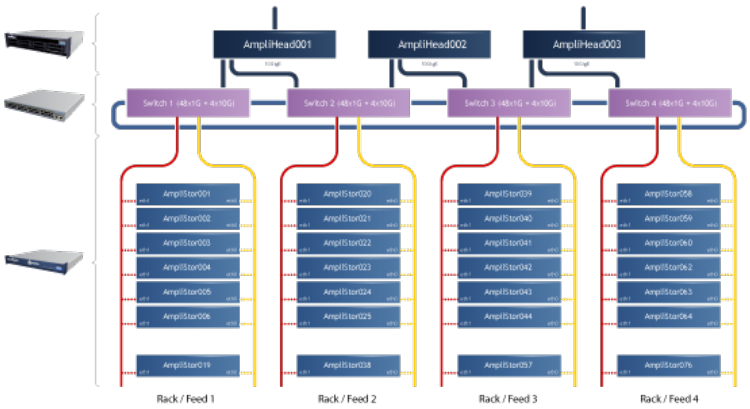
EPFL has several use cases that represent different challenges and opportunities to evaluate disk-based solutions:

- Static archive: how do disks behave on the long run and what is the energy required to run the system
- Large file repository: systems need to support files of up to 1TB
- Large file data transfer
- High speed streaming: peak throughput put above 1Gbps

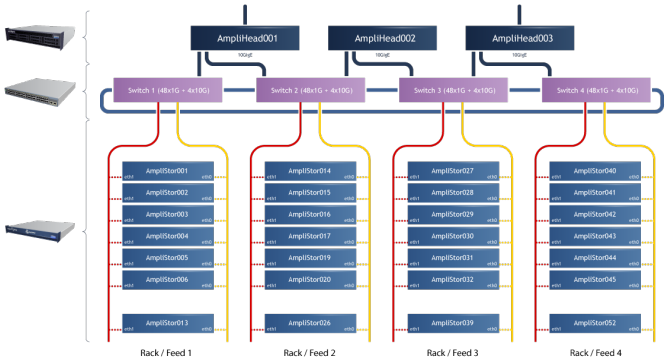
EPFL acquired and installed a total of 128 Amplidata 8TB[1] storage nodes, totaling 1PB, and 6 storage controllers. The controllers, which provide access to the storage pool, each have 2 10GB ethernet ports; the storage nodes have regular 1GB ethernet. The hardware is set up in two datacenters:

[1] The 8TB storage nodes have since been discontinued and were replaced with the larger 20TB nodes (AmpliStor AS20). As the infrastructure grows, the EPFL infrastructure will seamlessly scale by adding new AS20 nodes.

At the EPFL data center, 76 nodes will provide storage for further EPFL research. EPFL will use the rich heritage as a base for launching its MetaMedia Center, a center of competence for innovation and technology transfer to the media/IT industry and the development of research in this domain.



The plan is to eventually maintain a disk-only archive. EPFL will investigate what the potential is of the active archive and what the limits are.



52 nodes are installed in the Montreux Jazz data center, where the archive will be used for internal projects that require easier access to the archive than what the tape library provides, for example post processing. For such activities tape would be just too cumbersome.

Future projects

One project that will particularly benefit from the Active Archive is the plan to make most concerts available through the Montreux Jazz Café outlets.

The availability of these archives should result in new research options and opportunities for studying the domain of musicology, as well as research into sound and video. There will be a progressive and partial opening of the archives to the public, with some recordings being visible at branches of the Montreux Jazz Café (already open in Geneva and Sydney; opening soon in Zurich, London and Paris).

It's clear that tape is not an option for streaming of media. The only alternative is a disk-based archive, but at a reasonable cost. The challenge is that the Montreux Jazz Library will soon pass the 1PB threshold and will continue to increase. Traditional disk storage systems by average require 200% overhead for availability purposes (RAID/Replication). For a 1PB archive, that would mean 3PB of raw storage. Amplidata

provides a much higher availability with only 60% overhead, which is even lower than the 100% tape overhead. As the Amplidata appliances are also optimized for low power consumption, Montreux Jazz will be able to offer the service with an affordable power cost (90% less power than tape).

Conclusion

Alexandre Delidais: "This partnership is a fantastic opportunity to evaluate and innovate a real platform with such various and antagonist constraints. The high demand of video and the appetite of the Montreux Jazz organization to always use latest technologies for capturing concerts, will make this platform an anticipating case of mainstream needs 5 years from now."

Trivia: Claude Nobs, founder of a legendary festival

On a trip to New York, Claude Nobs met Roberta Flack and invited her to the Rose d'Or de Montreux. Later, Aretha Franklin made her first visit to Europe thanks to him. At the age of 31, he organized the first Montreux Jazz Festival featuring artists such as Charles Lloyd, Keith Jarrett, Ron McLure and Jack DeJohnette. This new Festival was an immediate success, and gained a reputation far beyond Switzerland. Claude Nobs quickly transformed his Festival into an international gathering place for lovers of jazz. During the 1990s, Claude Nobs shared the directorship of the Festival with Quincy Jones, and made Miles Davis an honorary host. The Festival continued to diversify and was no longer exclusively devoted to jazz.

All video recordings of the Montreux Jazz Festival since the first edition in 1967 are being archived on (analog and digital) tapes in the chalet of Claude Nobs. End 2010, Claude Nobs and his company Montreux Sounds, which manages the archive, agreed with EPFL to start a project to digitize and preserve all this unique and exclusive content.

http://en.wikipedia.org/wiki/Claude_Nobs
http://www.montreuxsounds.com/?a=ad&v=288
http://www.tsr.ch/video/info/journal-19h30/#id=2777509;nav=info/journal-19h30/?q=EPFL

About EPFL

EPFL is one of the two Swiss Federal Institutes of Technology. With the status of a national school since 1969, the young engineering school has grown in many dimensions, to the extent of becoming one of the most famous European institutions of science and technology. Like its sister institution in Zurich, ETHZ, it has three core missions: training, research and technology transfer. Associated with several specialised research institutes, the two Ecoles Polytechniques (Institutes of Technology) form the EPF domain, which is directly dependent on the Federal Department of Home Affairs.

With over 350 laboratories and research groups on campus, EPFL is one of Europe's most innovative and productive scientific institutions. Ranked top 3 in Europe and top 20 worldwide in many scientific rankings, EPFL has attracted the best researchers in their fields. The School's unique structure fosters trans-disciplinary research and promotes partnerships with other institutions. It continuously combines fundamental research and engineering.

Disk vs. Tape comparison

Tape's advantages as a medium for long-term archiving are very clear: very low media cost, low power cost (zero for offline tapes), extended archive life and removability. For these reasons, tape still is the incumbent technology for long-term data archiving of large-scale data, including the critical digital assets managed in media and entertainment.

It is clear that tape does have challenges that must be carefully considered. Among them are access performance, accessibility, reliability, data integrity assurance, technology compatibility, scalability and management cost. Archives maintained on tape are typically offline, with high retrieval latencies, and hence are impractical and inconvenient for use as an "active archive". The management of large tape libraries also increases the overall cost of ownership beyond what is promised by the media by itself.

Guaranteeing data integrity is another key challenge for electronic archives, and maintaining data integrity over time requires trusting the tape media. The problem of tape media refresh and migration will continue to become worse as capacities increase to hundreds of Petabytes and beyond. The problem of managing, migrating and refreshing tapes at this scale becomes analogous to painting the Golden Gate Bridge from end-to-end while the bridge is growing in length – it's by definition a never ending problem! We already know that current digital tape readers will be at the end of their life in several years. Transferring to a new tape technology again, is time consuming and costly.

