



MBC 콘텐츠 아카이브 시스템 고도화

글. 이용성 MBC 방송IT센터 유통인프라부

IT 기술이 급격하게 발달하면서 방송제작 및 유통환경도 급격하게 발전하고 있다. “라떼는 말이야”식 표현을 하자면 필자가 입사한 2006년만 해도 VCR 기반 녹화와 편집이 아주 일반적이었다. 파일 기반의 제작 환경은 소규모였던 것으로 기억한다. 하지만 NPS가 도입되면서 방송제작 환경은 급격하게 바뀌었다. 1:1 편집기 시절에는 상상도 할 수 없던 수많은 영상 소스를 다룰 수 있게 된 것이다. MBC에서는 이러한 방송환경의 변화에 능동적으로 대응하여 콘텐츠 아카이브 시스템(이하 DAMS, Digital Asset Management System)을 자체적으로 개발하여 15년 이상 지속적으로 고도화하며 안정적으로 운영하고 있다.

‘아카이브(Archive)’를 인터넷에서 찾아보면 역사적 가치 혹은 장기 보존의 가치를 지닌 기록이나 문서 등의 모음 또는 시설을 의미한다고 나온다. 과거의 아카이브가 VCR 테이프나 도서 등을 저장하는 장소를 의미했다면 오늘날의 아카이브는 전산화된 파일 저장소와 관련 장비(서버 등)의 묶음, 그리고 자료 조회를 위한 온라인 서비스를 의미하는 경우가 더 많다. VCR 기반 제작 환경이던 시절에 아카이브는 테이프 저장소를 의미했지만 IT 기술이 급격하게 발달하고 있는 오늘날에는 대용량 파일 저장 서비스로 인식하는 게 더 일반적이다. 본사 아카이브 시스템은 2005년에 처음 서비스를 시작하였다. 타사보다 이른 시기에 시작해서 영상은 SD 규격부터 아카이브를 시작하였다. 그 후 2008년부터 HD 아카이브를 본격적으로 시작하였고 2018년부터는 UHD 아카이브가 시작되었다.



그림 1. DAMS 개념도

해의 주요 방송사도 그렇지만 대부분의 아카이브 시스템은 운영상의 호환성 문제 발생을 최소화하기 위해 해상도별 특정 규격으로 통일한다. DAMS에서도 SD, HD의 경우 특정 코덱으로 통일했으나 제작 및 유통 형태의 다변화에 대응하고자 입고 규격을 다양화했다. UHD는 XAVC, PRORES 영상을 모두 저장할 수 있다. UHD 영상의 경우 아직 HD만큼 표준화된 포맷이 확정되지 않아서 CODEC이나 Frame Rate 등이 변화해도 문제가 없도록 하였다. 2019년부터는 원본급 영상에서 유튜브 등 클립규격의 H.264 영상까지 아카이브 범위를 확대했다. 이로써 유튜브 등 외부 동영상 플랫폼에서 많은 조회 수가 발생하는 ‘오분순삭’ 등의 각종 클립을 저장할 수 있게 되었고, 용량이 크지 않은 외부 영상의 경우에도 불필요한 Up-Converting이 필요 없게 되었다.

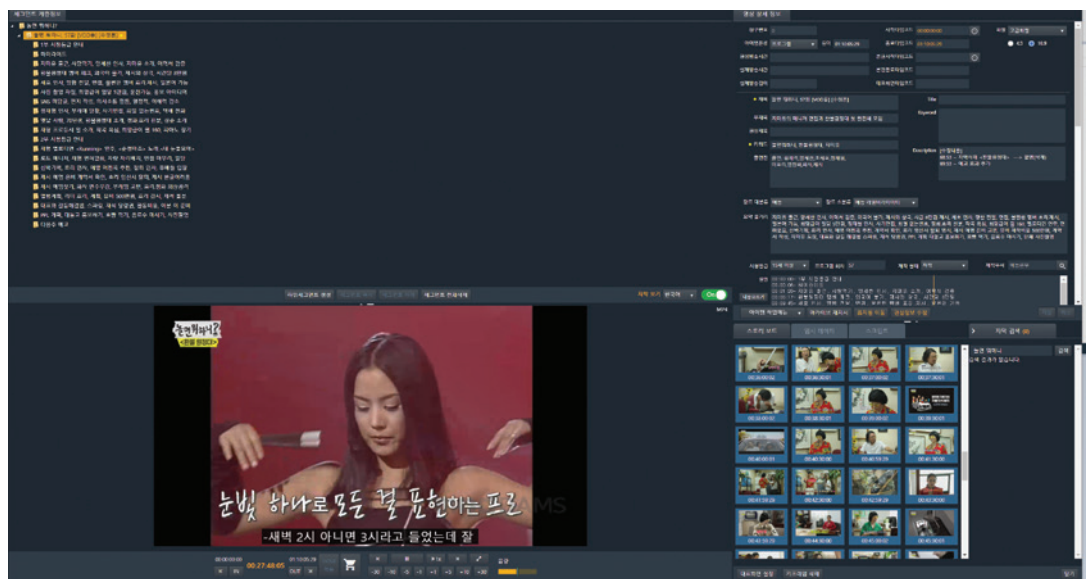


그림 2. 방송본 아카이브 예시

아카이브 시스템으로 입고된 영상 데이터는 LTO(Linear Tape Open)라는 미디어에 저장된다. 첨단 IT 시대에 테이프 미디어를 쓰는 것을 이해하지 못하는 사람이 많다. 하지만 LTO는 요즘 표현으로 가성비가 좋은 저장 미디어이다. HDD 기반 스토리지의 경우 속도가 빠르지만 장기간 저장이 불가능하다. 기본적으로 HDD의 권장 수명이 5년이

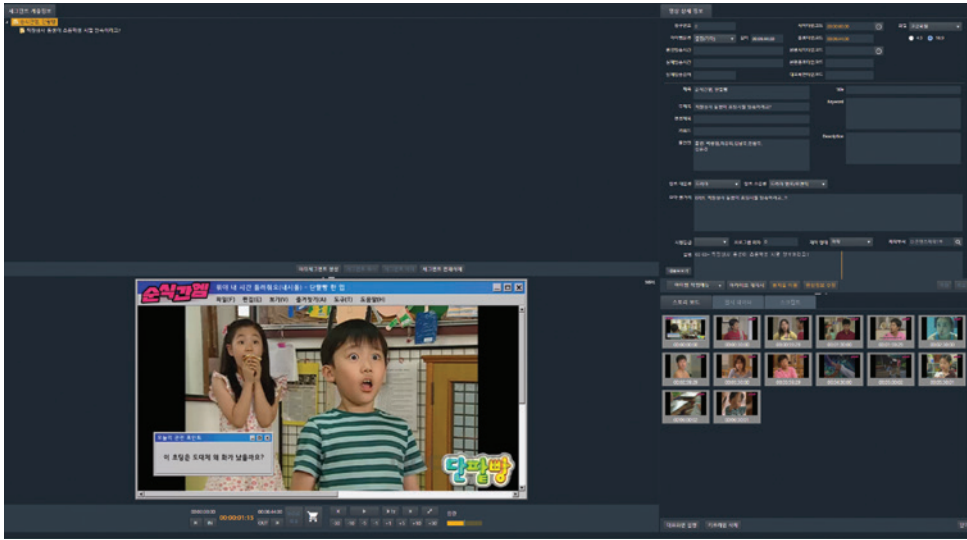


그림 3. 스펀오프 콘텐츠 아카이브 예시

며, 대부분의 스토리지가 납품 후 7년 이내에 EOS(End of Service, 제조사가 고장난 부품 교체 등 기술지원을 중단하는 상태)를 선언하기 때문에 장기간 사용이 불가능하다. 스토리지를 교체하는 경우 기존에 저장했던 대용량 데이터를 모두 이전해야 하는데 이 과정에서 일부 데이터가 유실될 수 있는 위험이 있다.

이에 비해 LTO의 경우 평균 수명이 30년 이라서 장기간 보존이 가능하다. 그리고 대량생산이 쉬운 간단한 형태라 가격도 저렴하다. 게다가 차세대로 발전할수록 같은 미디어 크기에 용량이 2배씩 증가하여 테이프 라이브러리를 장기간 사용할 수 있다. 보안 문제에서도 LTO가 상당히 유리하다. 특정 서버에 랜섬웨어가 유입되는 경우 해당 서버에 연결된 모든 스토리지에 저장된 파일은 암호화되어 사용할 수 없게 된다. 온라인 스토리지의 경우 한 번에 수많은 파일을 접속할 수 있어서 연결된 서버가 감염되면 대부분의 데이터가 한 번에 암호화될 수 있다. 하지만 LTO의 경우 저장 미디어가 오프라인 형태로 저장되기 때문에 랜섬웨어로 데이터가 파괴되기 어려운 구조라고 볼 수 있다. 그리고 LTO 테이프는 사람이 직접 마운트해서 사용하는 방식이 아니라 테이프 라이브러리 안에 보관되어 고속의 로봇이 마운트 작업을 수행하기 때문에 비교적 빠른 속도로 온라인화 할 수 있다.



그림 4. 다량의 LTO가 저장된 테이프 라이브러리 내부

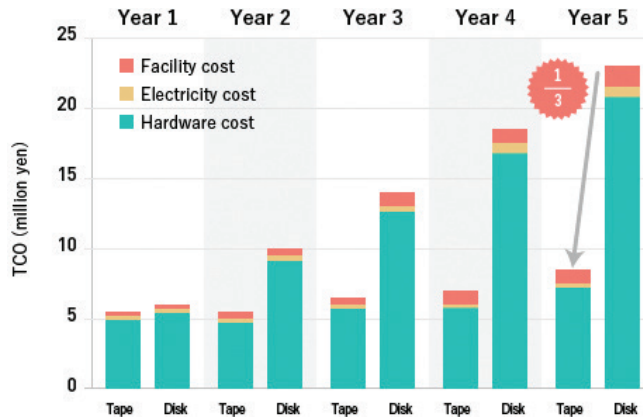


그림 5. 저장 매체별 비용 비교

LTFS

Linear Tape File System

그림 6. LTFS 로고

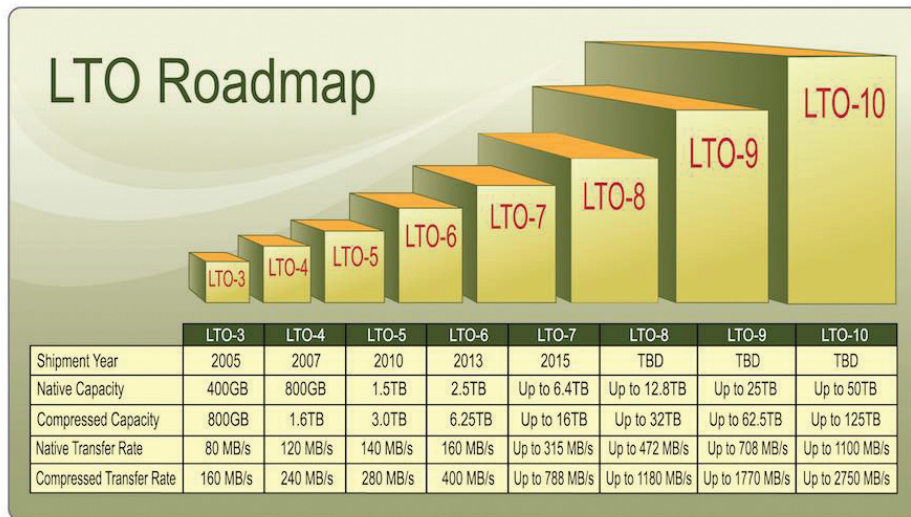


그림 7. LTO 세대별 용량

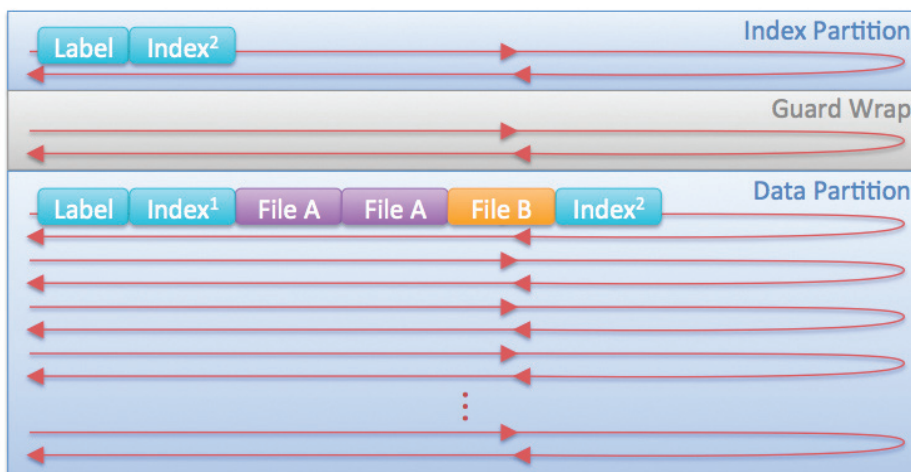


그림 8. LTFS 파일 저장 구조

DAMS의 경우 아카이브의 역사가 길어서 LTO2 시절부터 개발을 시작하였고 LTO3부터 본격적인 저장을 시작하였다. 이후 LTO4, LTO5가 도입되었고, 현재는 LTO7까지 업그레이드되었다. 기존 LTO의 경우 본사에서 자체적으로 개발한 규격으로 저장했지만 2018년에 도입된 LTO7부터는 글로벌 표준 기술인 LTFS 기술을 저장하였다.

LTFS 기술은 자체 규격이나 DIVA 등과 다르게 파일에 관한 메타데이터를 LTO 미디어 자체에 저장해서 시스템을 통한 접속을 하지 않아도 파일에 접근이 가능하다. 직관적으로 표현하자면 LTO 마운트 시 윈도우 탐색기에서 파일 정보를 직접 확인할 수 있고 온라인 스토리지에 직접 복사할 수 있다. 기존 LTO의 경우 어떤 LTO에 어떤 파일이 저장되었는지 알려주는 메타데이터가 삭제된 경우 복구하기 어렵지만 LTFS의 경우 복구가 가능하다. 무엇보다도 오픈된 기술이라서 시스템 운영 시 경비 절감이 가능

하다. 고가의 솔루션의 경우 도입가의 22%를 매년 요구하고, 용량 증가에 따른 라이선스 추가 구매가 필요하지만 관련 비용지출이 불필요하다. 하지만 무료 기술이기 때문에 도와주는 사람이 없어서 문제 발생 시 엔지니어 스스로 찾아서 해결해야 한다.

텍스트나 이미지 데이터는 파일 크기가 작아서 스토리지의 여러 디스크에 분산되어 저장되는게 일반적이다. 그래서 RPM이 높은 SAS Disk 또는 SSD를 사용한다. 이에 비해 영상 미디어 파일은 용량이 몇십~몇백 배 크다. 그래서 일반적인 데이터와는 다른 취급 방법이 필요하다. 온라인 스토리지의 경우 일반적으로 Random Access를 얼마나 빨리 할 수 있는지 IOPS(Input/Output Operations Per Second)로 평가하는 것이 일반적이다. 하지만 아카이브 시스템의 경우 대용량 파일이기 때문에 디스크에 Sequential에 가까운 형태로 저장해서 Random Access에 대한 이슈가 적은 편이다. 그래서 Read/Write 성능(Throughput)으로 측정하는 방식이 더 일반적이다. 아카이브 시스템에 온라인 스토리지를 구성하는 경우 RPM이 높은 SAS를 사용하지 않고 대용량 구성이 가능한 NL-SAS로 구성해도 성능에 큰 문제가 없다. 대부분의 데이터 사용 케이스를 보면 과거의 데이터보다는 최근의 데이터를 사용하는 경우가 많다. 아카이브 시스템 사용에서도 비슷한 패턴으로 사용된다. DAMS의 경우 최근 몇 달의 콘텐츠를 저장할 대용량의 온라인 스토리지를 구성하여 처리 속도가 늦은 LTO 사용을 최소화하였다.

많은 콘텐츠를 아카이브 해도 찾을 수 없다면 아무 소용이 없다. DAMS에서는 빠른 검색이 가능하고 유동 등에 사용이 유리하도록 장면 단위로 메타데이터를 입력할 수 있다. 예를 들자면 드라마는 Scene 단위, 뉴스는 꼭지 단위, 예능은 코너 단위, 음악 방송의 경우 곡 단위로 메타데이터가 입력된다. 각 장면은 우측의 이미지처럼 계층화된 입력이 가능하다. 이렇게 입력된 메타데이터는 DB에 입력한 상태로는 인덱싱의 한계로 빠른 검색이 불가능하다. 그래서 별도의 검색엔진 도입이 필수적이다. DAMS에서도 제목, 출연진 등 주요 데이터 외에도 장면 메타데이터까지 검색엔진에 색인하여 빠른 검색이 가능하도록 하였다. 2020년에는 자막 아카이브 기능을 추가하여 자막 기반 Speech 검색이 가능하도록 하였다. 자막 검색의 경우 오픈소스 기술인 Elasticsearch 기술이 도입되었다.

(‘월간 방송과기술’ 2020년 5월호 참조)

Segment 계층정보 (Segment = 구간)

Segment 계층 설명

- L1 : 상위 계층
- L2 : L1의 하위계층
- L3 : L2의 하위계층

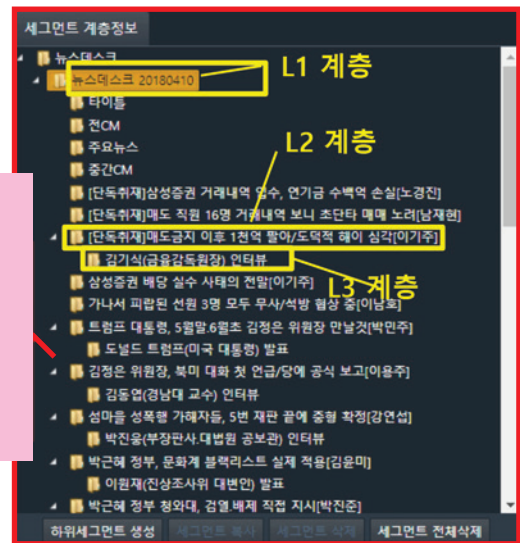


그림 9. 장면 메타데이터 UI

이렇게 DAMS의 인프라 및 S/W가 발전하면서 아카이브 범위도 계속 증가하고 있다. 초창기의 아카이브는 TV, 라디오 방송프로그램 저장에 대부분이었다. 그 후 뉴스NPS가 도입되면서 취재, 제보영상이 추가되었고 다큐멘터리 중요 소재, 과거 영상자료 등으로 확대되었다. 예능 제작에서 과거 영상 활용이 증가하면서 클린본 아카이브가 시작되었다. 과거 영상을 사용하는 경우 기존의 자막을 흐리게 처리해야 했지만 클린본이 아카이브되면서 보다 깔끔한 영상을 사용할 수 있게 되었다. 2019년에는 콘텐츠 저장 범위가 Mashup으로 확대되었다. 유튜브 등의 클립 유통이 늘어나면서 클립에 대한 체계적인 저장의 필요성이 생겨서 오분순삭, 옛능 등 다양한 클립들이 아카이브되고 있다.

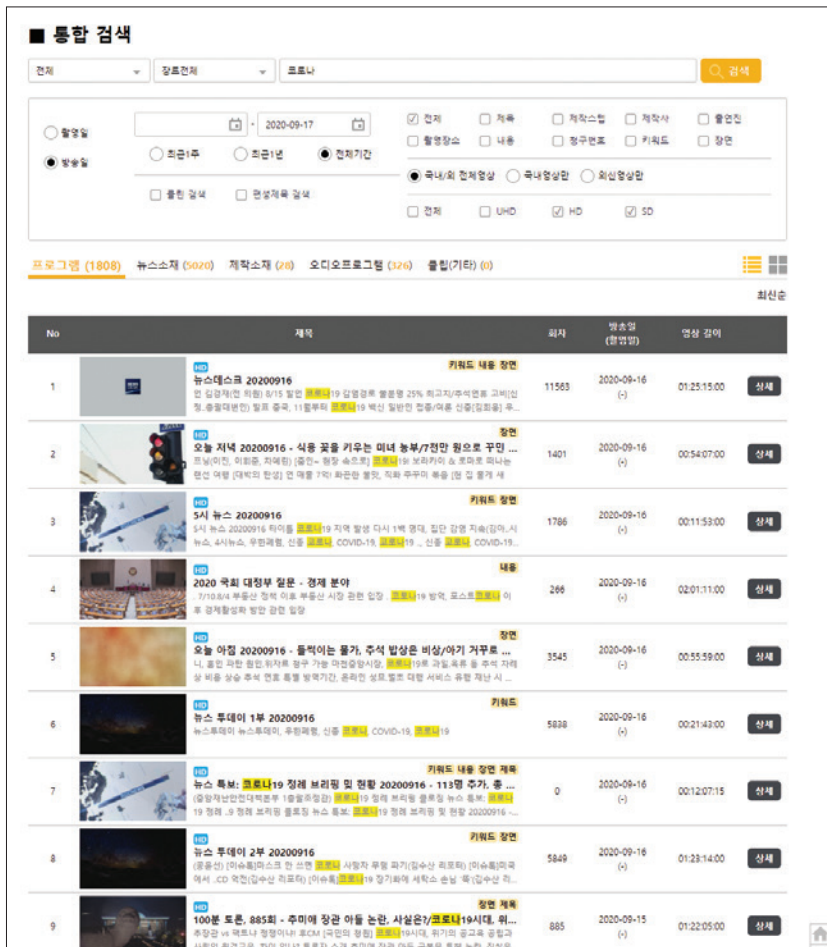


그림 10. DAMS 검색 화면



그림 11. 자막 검색 결과

2020년에도 DAMS는 계속 고도화되고 있다. 작년부터 Tapeless 제작이 일반화되어 VCR 테이프 저장이 중단되었다. 이로 인한 콘텐츠 유실 방지를 위해 외부 클라우드 시스템과 연동되는 콘텐츠 백업 워크플로우를 테스트하고 있다. 그리고 LTO의 용량이 증가하면서 저장 매체 오염으로 인한 파일 손상의 위험성이 증가하고 있다. 이를 방지하기 위해 쓰기 작업이 완료된 테이프의 무결성을 검증하는 프로세스를 추가하고 있다.

아카이브 범위가 확대되면서 아카이브 영상 사용량도 점차 증가하고 있다. 뉴스, 예능, 시사교양 등 주요 프로그램 제작에서 뉴미디어 유통용 영상 제작으로 확대되고 있다. 몇 년 전 본사에 새로운 뉴스NPS가 도입되면서 대용량의 Nearline Storage가 도입되었다. 대용량의 영상 저장이 가능하여 아카이브의 사용량이 감소하였으나 예능, 클립 제작용으로 사용량이 급증하면서 전체적인 사용량은 오히려 증가하였다.

정유 공장을 보면 RFCC(Residue Fluid Catalytic Cracking Unit)라는 공정이 있다. 돈이 안 되는 중질유를 촉매 등으로 분해하여 돈이 되는 경질유로 변환하는 과정이다. 아카이브 시스템 활용도 비슷한 맥락이라고 본다. 기존의 아카이브는 방송의 기록을 남겨 놓은게 주목적이라서 유지하는데 많은 돈이 드는 업무라고 생각하는게 일반적이었다. 하지만 각종 콘텐츠 제작에 아카이브 영상 활용이 증가하면서 제작에서 점점 더 많은 비중을 차지하고 있다. DAMS도 이렇게 변화하는 환경에 대응하여 좀 더 고품질의 콘텐츠 제작에 도움이 되기 위해 지속적으로 고도화할 예정이다.

자료 출처

- 아카이브 의미 / ko.wikipedia.org/wiki/%EC%95%84%EC%B9%B4%EC%9D%B4%EB%B8%8C
- 그림 5. 저장 매체별 비용 비교 / www.tape-storage.net/en/storage_comparison/article_01
- 그림 6. LTO 로고 / www.lto.org/2016/03/lto-the-inside-scoop
- 그림 7. LTO 세대별 용량 / insidehpc.com/2015/09/spectra-logic-to-offer-lto-7-technology
- 그림 8. LTO 파일 저장 구조 / www.smaller-systems.com/blog/2011/06/how-does-lto-work