

개요: 백업  
워크플로우에서 테이프를  
클라우드로 교체

비교 및 3가지 참조 아키텍처

2017년 6월



© 2017, Amazon Web Services, Inc. 또는 자회사. All rights reserved.

## 고지 사항

이 문서는 정보 제공 목적으로만 제공됩니다. 본 문서의 발행일 당시 AWS의 현재 제품 및 실행방법을 설명하며, 예고 없이 변경될 수 있습니다. 고객은 본 문서에 포함된 정보나 AWS 제품 또는 서비스의 사용을 독립적으로 평가할 책임이 있으며, 각 정보 및 제품은 명시적이든 묵시적이든 어떠한 종류의 보증 없이 "있는 그대로" 제공됩니다. 본 문서는 AWS, 그 계열사, 공급업체 또는 라이선스 제공자로부터 어떠한 보증, 표현, 계약 약속, 조건 또는 보증을 구성하지 않습니다. 고객에 대한 AWS의 책임 및 채무는 AWS 계약에 준거합니다. 본 문서는 AWS와 고객 간의 어떠한 계약도 구성하지 않으며 이를 변경하지도 않습니다.

# 목차

서론	5
테이프의 특성	5
백업	6
아카이브	7
비용 요인 분석	7
휴대성	8
안정성	9
성능	9
클라우드의 특징	10
비용 요인 분석	10
지속성과 가용성	11
상호 보완 서비스	11
성능	12
참조 아키텍처	14
백업 소프트웨어 통합을 통해 클라우드로 백업	14
하이브리드 클라우드 가상 테이프 라이브러리를 통해 클라우드로 백업	15
온프레미스 NFS 인터페이스를 통해 클라우드로 백업	18
결론	19
기여자	19
참고 문헌	20
주석	20

## 요약

IT 조직은 수십 년 동안 테이프 기반 백업 및 보관 프로세스를 대체할 방식을 강구해 왔습니다. 최근 들어, 클라우드 서비스가 테이프 및 테이프 자동화의 비용 및 복잡성을 줄일 수 있는 대안으로 떠오르고 있습니다. 설문 조사 결과, 현재 IT 전문가들은 클라우드 플랫폼의 최상위 사용 사례 중 하나로 백업을 꼽고 있습니다.<sup>1,2</sup> 본 문서에서는 기존의 마그네틱 테이프의 비용 및 성능과 클라우드 스토리지의 비용 및 성능을 백업 워크플로우 측면에서 비교하고 대조해 봅니다. 또한 간단하고 중단 기간이 최소화된 클라우드 백업 프로젝트에 대한 3가지 일반 참조 아키텍처를 소개하고 AWS가 왜 강력한 백업 스토리지 플랫폼인지를 알아봅니다.

## 서론

테이프 미디어, 자동화 시스템 및 백업 소프트웨어는 기존의 데이터 보호 전략에서 중요한 부분을 차지했습니다. 그러나 많은 조직들은 테이프 미디어를 구매하고 처리하고 보관하는 데 드는 비용 및 노력에 지쳐갔을 뿐 아니라 테이프 라이브러리 및 관련 소프트웨어 라이선스를 교체하는 비용도 만만치 않음을 느끼고 있었습니다. 백업 기능은 애플리케이션 및 해당 데이터를 보호하는 데 있어 매우 중요한 항목임에는 변함이 없었지만 테이프 관리 및 프로세스에 드는 노력 및 복잡성을 대체할 무언가가 필요했습니다. **AWS**는 기존의 테이프 백업, 온프레미스 테이프 라이브러리 및 오프사이트 물리적 보관 서비스를 개선하거나 이를 내구성, 대용량 확장성, 종량 과금제, 짧은 검색 시간 및 자동화된 관리 기능을 가진 클라우드로 대체할 수 있는 다양한 옵션을 제공합니다.

## 테이프의 특성

60여 년 동안 IT 업계는 마그네틱 테이프에 의존하여 데이터를 저장하고, 오프사이트로 이동하고 오랜 기간 보관해 왔습니다. 이렇게 오랜 기간 사랑을 받을 수 있었던 것은 하드 디스크 드라이브 미디어("디스크")에 비해 테이프 스토리지가 밀도 증가로 인해 기가바이트당 비용이 지속적으로 떨어졌으며, 테이프 시스템의 전력 및 냉각 비용이 낮을 뿐 아니라 수 년 간 사용되었던 테이프 관리 구조가 익숙했기 때문입니다.

일반적인 테이프 기반 백업 아키텍처에서는 프로덕션 애플리케이션에 사용되는 주요 데이터가 네트워크로 연결된 스토리지나 로컬 디스크 스토리지에 저장됩니다. 설정된 일정에 따라, 데이터가 애플리케이션 서버나 스토리지 어레이 스냅샷을 통해 백업 서버로 전송되며 마그네틱 테이프에 수집 및 기록됩니다. 이러한 테이프는 테이프 운영 직원에 의해 수동으로 관리되거나 로봇 공학의 형식으로 자동으로 관리되는 대용량의 온사이트 테이프 라이브러리에 저장됩니다. 일부 조직에서는 미션 크리티컬 애플리케이션 데이터를 WAN(Wide Area Network)을 통해 다른 디스크 어레이나 소형의 테이프 라이브러리 오프사이트에 복제하거나 또는 제3자를 고용하여 테이프를 오프사이트 스토리지 시설로 수동으로 옮기도록 합니다.

백업을 마그네틱 테이프에 기록하면 보관이 쉽고 복제가 간편하지만, 마그네틱 테이프에는 독점 형식 및 순차적인 읽기 동작과 같은 태생적 단점이 있어 백업 기간 및 복구 프로세스가 길어지게 되며 이에 따라 **IT** 팀이 목표 복구 시점(**RPO**) 및 목표 복구 시간(**RTO**)을 충족하기 힘듭니다. 또한, 마그네틱 테이프 백업은 자주 테스트하는 것이 현실적으로 어려우며, 테이프 자체는 특히 아주 깨끗한 무진 환경에 보관되지 않는 경우 백업 및 복구 중에 카트리지와 미디어 오류가 발생할 수 있습니다. 실제로, 테이프 미디어 단품 및 수익은 꾸준히 감소해 왔으며, **TDK, Imation** 및 **Tandberg Data**와 같은 공급업체는 디스크 및 클라우드 기반 대체품의 높아가는 압력으로 인해 시장을 통합하거나 시장에서 빠져나오고 있습니다.

이러한 단점으로 인해 테이프 기반 백업 솔루션은 복구 이벤트에 일반적으로 적합하지 않습니다. 현재 테이프 기반 백업 솔루션을 사용하고 있는 경우, 클라우드로 이전하면 물리적 미디어를 손수 다루는 복잡성 및 위험이 제거될 뿐 아니라 물리적 테이프 스토리지 미디어에의 백업과 관련된 총 비용을 줄일 수도 있습니다. 또한 클라우드에서는 애플리케이션 복구 지원이라는 추가 이점도 있습니다.

## 백업

본 문서에서는 "백업"을 시스템 가동 중지 및 악의적이거나 실수로 데이터가 손상되는 일이 발생할 때 이를 보호할 수 있도록 보조 위치에 데이터 및 애플리케이션 상태의 복사본을 저장하는 것으로 정의합니다. 백업에는 일반적으로 일 또는 주로 측정되는 사용 수명이 있으며 가끔 온사이트로 관리되기도 하지만 보통은 테이프 라이브러리에 관리됩니다. 대부분의 백업 소프트웨어 제품은 백업 작업의 위치 및 시간에 대한 메타데이터를 사용하여 백업 카탈로그를 유지 관리합니다. 제품 플레이와 같은 데이터 관리 역할은 보통 **IT** 관리자에게 매우 중요합니다. **D2D2T(Disk-to-disk-to-tape)** 솔루션은 디스크 용량이 늘어나는 반면, 백업 기간이 짧고 복원이 빠를 수 있습니다. 중복 제거 및 압축과 같은 데이터 감소 기술을 사용하면 가격이 비싼 디스크에서 사용량이 증가하는 문제를 해결하는 데 도움이 될 수 있습니다.

백업 복사본은 규정이나 내부 지침을 준수하기 위해 수년 또는 심지어 수십 년 동안 보관될 수 있습니다. 장기 데이터 보관 정책의 일부로 테이프 백업이 보관되는 경우 물리적 테이프의 내구성, 안정성 및 관련 시스템 비용은 **IT** 조직의 큰 해결 과제가 되기도 합니다.

## 아카이브

아카이브 애플리케이션에서도 테이프를 일반적인 스토리지 미디어로 활용할 수 있지만, 백업과 다르게 아카이브에서는 디지털 자산의 복사본만 규정 준수, 보존, 감사, 재사용, 분석 또는 상업적인 라이선스 가치로 인해 더 오래 저장하기도 합니다. 아카이브 프로세스는 업계나 규정 준수 요건 또는 조직의 모범 사례에 따라 구동되기도 합니다. 이러한 점 때문에 기가바이트당 비용이 가장 낮은 서비스로 데이터를 "가두는" 방식이 사용되는데, 이 경우 특별히 요청하는 경우에만 데이터를 사용할 수 있습니다. **RTO**는 일 수로 측정되기도 하며, 타사 공급업체가 특정 유형의 데이터 및 애플리케이션(예: 이메일)에 대해 특화된 서비스 및 솔루션을 제공하여 규정 준수나 기록 관리 전문가를 제공할 수도 있습니다.

데이터 관리 애플리케이션에서 클라우드 기술을 사용하여 백업을 아카이브로 원활하게 변환하고 아카이브 프로세스를 간소화할 수도 있지만, 본 백서에서는 주로 백업을 중점적으로 설명합니다. 기존 테이프 아카이브를 클라우드로 변환하고 이전하도록 도와주는 타사 솔루션(예: **Index Engines**)이 있더라도 본 백서에서는 이를 설명하지 않습니다.

## 비용 요인 분석

테이프의 가장 강력한 이점은 약간 낮은 미디어 비용으로, 일반적으로 데이터 보관에 드는 비용은 가장 낮고 밀도는 가장 높습니다. 테이프의 두 번째 주요 비용 이점은 낮은 전력 사용료로, 로봇 공학 시스템에 의해 테이프가 기록되고 읽히고 전송될 때만 전력이 사용됩니다. 그러나 이러한 요소만을 총 소유 비용(**TCO**) 분석에서 고려하면 누락된 부분이 보입니다. 추가 비용 요소는 다음과 같습니다.

- 테이프 라이브러리, 로봇 공학 및 확장 어레이에 드는 하드웨어 비용으로, 대규모 시설의 경우 수백억 달러에 달하며 5년에서 10년의 수명 주기 동안 분할 상환/감가상각됩니다.
- 대량 미디어 마이그레이션과 관련된 관리 및 미디어 비용으로, 데이터가 여러 세대의 테이프 미디어를 거치는 경우 장기 "비트 손상(**bit-rot**)"을 방지하기 위해 발생합니다. 미디어 갱신 및 마이그레이션은 보통 6~8년 간격으로 수행됩니다.

- 테이프 드라이브 및 네트워크 패브릭에 대한 비용으로, 백업 기간 요건을 충족할 수 있도록 적절한 처리량을 제공하기 위해 드는 비용. 새로운 테이프 형식을 지원하거나 성능을 개선하기 위해 새로운 드라이브가 제공됩니다.
- 지원 계약 비용, 일반적으로 하드웨어 구매에 사용한 자본의 **20%** 정도에 해당.
- 테이프 백업 관리자의 인건비, 오프라인 원본에서 테이프를 가져오는 시간 및 노력 포함.
- 서비스의 오프사이트 볼트에 대한 계약(예: **Iron Mountain**).
- 하드웨어 및 관리 비용, 미디어나 기계적 오류를 예방하기 위해 여러 온사이트 및 오프사이트 데이터 복사본을 보관하는 관습에 따라 몇 배가 되기도 함.
- 총 비용은 데이터 오버헤드 및 낮은 미디어 활용률 등의 요소로 인해 증가될 수도 있습니다. 압축을 사용하면 이러한 요소가 완화될 수 있지만, 대부분의 최신 백업 애플리케이션 워크플로우에서는 데이터에 대해 중복 제거, 압축 및/또는 암호화를 수행하므로 완화되는 부분이 전혀 없을 수도 있습니다. 미디어 파일(예: 비디오, 오디오 또는 이미지) 아카이브 워크플로우에는 파일 형식이나 코덱에 일부 압축 기술이 이미 사용되어 있을 수도 있으므로 원시 압축 효과가 나타나지 않을 수 있습니다.

요약하면, 테이프 미디어의 다소 낮은 비용은 운영 비용(OpEx) 및 비미디어 자본 지출(CapEx) 요소를 완화함에 따라 세부 TCO 분석에서는 상쇄될 수도 있습니다.

## 현대성

테이프 카트리지는 물리적으로 처리, 전송 및 저장하기 쉽도록 설계되었으며 매우 좁은 범위의 업계 표준을 준수합니다. 고용량 하드 디스크 드라이브(HDD) 스토리지 미디어는 이동 중의 충격 및 진동에 견디도록 설계되지 않은 반면, 엔터프라이즈 솔리드 스테이트 드라이브(SSD)는 HDD보다 복원력이 뛰어납니다. GB당 SSD 가격은 현재 엔터프라이즈 HDD와는 경쟁할 수 없는 수준이며, IDC에 의하면 예측 가능한 미래에도 이는 동일할 것으로 예상됩니다.<sup>3</sup>



## 안정성

테이프 기반 스토리지는 장기 보관을 위해 엔지니어링된 것이지만 실제로는 다양한 안정성 위험을 가지고 있습니다. 따라서 여러 개의 테이프 사본(온사이트 및 오프사이트 모두)을 생성해야 하는데, 그렇게 하면 저렴한 비용이라는 테이프의 이점이 없어집니다. 카트리지, 라이브러리 로봇 공학 및 테이프 드라이브는 기계적인 특성을 가지고 있습니다. 오류가 발생하면 데이터가 손실되거나 가용성이 떨어지게 됩니다. 테이프의 데이터 무결성은 시간이 흐르면서 예측할 수 없게 무력화되어 주기적인 대량 미디어 마이그레이션 프로그램이 필요합니다. 테이프를 마운트하고 읽는 행위는 안전하지 않습니다. 산화물 코팅이 없어지고 얼룩이 테이프 카트리지에 적용되며, 선형 디바이스로부터 임의의 데이터를 가져올 때 "슈샤인(Shoeshine)"과 같은 특정 동작이 발생합니다. 빈번한 액세스용으로 설계된 테이프 시스템에는 고유한 설계 요소가 포함되지만, 가격이 비싸기 때문에 비용 관련 이점이 없어집니다.

## 성능

잘 프로비저닝된 테이프 시스템은 높은 읽기 및 쓰기 처리량을 제공할 수 있습니다. 그러나 성능을 예측할 수 없는데, 특히 로드를 받는 경우 더욱 그러합니다. 대용량 데이터세트를 가져오는 작업은 데이터 양, 테이프 드라이브 사양, 함께 멀티플렉스된 테이프 드라이브 수 및 로봇 공학 테이프 가져오기 속도에 따라 몇 시간에서 며칠이 걸릴 수 있습니다. 웨어하우스 기반 오프사이트 아카이브는 데이터 센터로부터의 지리적 분리 덕분에 안전성을 제공하지만, 데이터 테이프를 복구하려면 가져오기 속도를 높여주는 "러시" 서비스에 대해 비용을 지불하지 않는 한 며칠이 걸릴 수 있습니다.

디스크 기반 시스템은 일반적으로 최근에 보호된 데이터를 유지하여 테이프와 관련된 높은 가져오기 지연 시간을 완화하기 위해 배포됩니다. 파일 시스템 및 가상 테이프 라이브러리(VTL) 솔루션은 모두 기존의 테이프 관리 소프트웨어 및 하드웨어를 마그네틱 디스크 속도로 보완하므로 많이 사용됩니다. 이러한 설계는 디스크에 저장된 데이터의 성능을 개선하지만, 스테이징 디스크를 통과하여 물리적 테이프로 완전히 마이그레이션된 데이터에는 해당되지 않습니다.

## 클라우드의 특징

클라우드 공급자는 집계된 고객 요구를 활용하여, 고도로 중복되고 효율적인 데이터 센터, 가용 영역(AWS의 경우) 및 지역적 중복성 등 비용 및 안정성 면에서 규모의 경제를 제공하는 인프라에 투자합니다. 이러한 규모의 경제 덕분에 애플리케이션에 대한 검색, 분석 및 클라우드 기반 재해 복구와 같이 저장된 데이터를 활용하는 무료 부가 가치 서비스를 배포할 수 있습니다. 규모가 커질수록, 비용 절감 효과가 고객에게 전달되고, 고객은 더 많이 사용하여 수익을 높이게 되고 선순환 투자가 발생하게 됩니다. 클라우드의 장기 스토리지 플랫폼은 다른 클라우드 스토리지 제품과 함께 결합되어 저비용 스토리지 계층을 제공하며 매월 GB당 0.004달러(문서 작성 시점 기준)의 저렴한 비용으로 수 밀리초에서 수 시간까지의 데이터 검색 시간을 제공합니다.

## 비용 요인 분석

퍼블릭 클라우드 스토리지는 매우 다른 소비 및 비용 모델을 제공하는데, 모든 비용이 사용량을 기준으로 매월 "사용한 만큼"만 청구됩니다. 온프레미스 테이프 솔루션의 방대한 비용을 클라우드 기반 스토리지의 비용 및 민첩성 이점과 비교해 보는 경우, 오늘날의 클라우드 스토리지는 테이프와 동일한 TCO를 제공할 수 있다는 점을 고려하십시오.

- 테이프 기반 백업 및 아카이브와 다르게, 클라우드 스토리지에는 선행 자본 투자가 필요하지 않습니다. 원하는 용량을 사용할 수 있으며 데이터를 저장하면서 곧바로 크기를 조정할 수 있습니다.
- 미디어 및 시스템 노후화와 데이터 마이그레이션과 관련된 비용이 들지 않습니다.
- 테이프 시스템의 크기 조정, 복구 및 구성과 관련된 관리 비용이 들지 않습니다.
- 복원을 위해 오프라인 미디어를 가져오거나 새로운 미디어를 라이브러리에 채우는 데 있어 관리 비용이 들지 않습니다.
- 지원 계약 비용이 추가로 들지 않습니다.
- 내구성 및 가용성 비용이 추가로 들지 않습니다(모두 서비스에 포함되어 있음).

- 클라우드 공급자가 효율성 절감("상용 곡선")을 달성함에 따라 시간이 지날수록 일반적으로 가격이 떨어지는 이점을 얻을 수 있습니다.

## 지속성과 가용성

클라우드 플랫폼은 구성 요소 및 시설의 오류에 대비해 매우 높은 안정성을 제공하도록 내구성 서비스 수준 계약(SLA)인 99.999999999%를 만족하는 Amazon Simple Storage Service(S3) 및 Amazon Glacier와 같은 스토리지 솔루션이 적용 및 설계되었습니다. 인프라 오류로 인한 데이터 손실 가능성이 매우 낮으며, 온프레미스 스토리지 솔루션으로 쉽게 달성할 수 있는 수준보다 일반적으로 낮습니다. 테이프 형식 및 백업 소프트웨어 애플리케이션이 변경될 수 있으며 노후화된 형식 및 주변 장치와의 역방향 호환성을 유지해야 하는 기존 솔루션과 달리 기본 스토리지 아키텍처 및 미디어 선택 사항이 사용자에게 표시되지 않습니다.

또한, 클라우드 플랫폼은 매우 가용성이 높습니다. 업계 최고의 클라우드 스토리지 서비스는 미디어, 시스템, 네트워크 및 완전한 데이터 센터의 오류도 견딜 수 있도록 설계된 복수 사이트 지역을 기반으로 합니다. 전 세계 지역 간에 데이터가 복제될 수도 있으므로 안정성 및 글로벌 가용성 이점이 더욱 증가합니다. 클라우드의 내구성 및 가용성 수준을 가진 온프레미스 테이프나 디스크 시스템을 사용하여 강력한 솔루션을 구성하는 것은 특히 오늘날처럼 일반적으로 자본 지출(CapEx)이 제한되어 있는 IT 예산 속에서 불가능에 가깝습니다.

## 상호 보완 서비스

클라우드 스토리지 플랫폼은 데이터를 자동으로 정렬 및 이동하여 가성비 목표를 달성하도록 도와주는 관리 도구 등의 부가 가치 도구를 사용하여 데이터를 처리할 수 있습니다. 이러한 관리 도구는 다른 클라우드 기반 애플리케이션 내에서 보안, 감사 추적, 액세스 제어 및 동작과 통합 및 구동될 수 있습니다. 건강 관리 또는 금융 서비스와 같은 업계의 규정 준수 요건을 만족할 수 있도록 WORM(한 번 쓰고 여러 번 읽기) 정책을 적용할 수 있습니다.

다음은 테이프 교체 프로젝트 동안 적용되는 클라우드 기반 데이터 수명 주기 관리 도구의 몇 가지 예입니다.

- RTO 요건이 낮은 새로운 백업을 고성능 객체 스토리지 계층에 30일 동안 기록하고, 90일 동안 비용이 좀 더 저렴한 계층으로 이동한 다음, 비용이 가장 저렴한 아카이브 계층으로 이동하여 규정 준수를 위해 잠근 다음 7년 후 삭제할 수 있습니다.
- 특정 타임스탬프 범위 내에서 새롭게 가져온 백업은 아카이브 계층으로 바로 이동하고 선택적으로 잠글 수 있습니다.
- 새롭게 가져온 백업에는 메타데이터 태그를 지정하여(작업 ID, 날짜, 물리적 위치 등) 태그별로 정책을 적용할 수 있도록 합니다. 이를 통해 점점 늘어나는 클라우드 테이프 아카이브 내에서 추가 인덱싱 및 관리 범위를 제공할 수 있습니다.

일부의 경우에는 백업 데이터세트에 **LOB(Line of Business)** 사용자가 규정 준수나 실용적인 비즈니스 가치를 위해 분석하고자 하는 유용한 기록 데이터가 포함되어 있습니다. 그러나 테이프 기반 시스템의 경우, 대량 분석을 위해 데이터에 액세스하는 것이 어렵거나 불가능할 수 있습니다. 그러나 클라우드 스토리지에서는 필요 시 데이터로부터 가치를 순조롭게 추출할 수 있는 컴퓨팅 서비스를 즉시 사용할 수 있습니다. 이러한 점 때문에 테이프의 데이터를 클라우드로 이전하는 경우 원래 데이터 및 파일 형식을 보존하기 위해 독점적 백업 형식을 피하는 것이 좋습니다. 따라서 클라우드로의 이전은 "의미 없는 비트를 스마트한 비트로 전환"하는 것이 될 수 있으며, 내부 이해관계자에게 빅 데이터 가치를 추출할 수 있습니다.

## 성능

클라우드 스토리지는 수 밀리초에서 수 시간까지 다양한 가성비 옵션으로 제공되므로 맞는 옵션을 선택하여 워크플로우의 필요를 충족할 수 있습니다. 복원 작업을 위한 대량 데이터 가져오기와 같은 성능이 중요한 워크플로우의 경우, 기본 스토리지 성능보다 WAN 연결(예: 인터넷 연결)이 RTO에 더 많은 영향을 줄 수 있습니다. 따라서, 테이프를 클라우드로 전환할 때는 네트워크 연결 요건뿐만 아니라 최근 데이터의 온프레미스 캐싱 여부도 꼼꼼히 고려해야 합니다.

가장 저렴한 클라우드 스토리지 계층은 가장 낮은 읽기 성능을 나타내기도 하며, 거의 액세스하지 않는 장기 아카이브 사용 사례를 위해 만들어진 것입니다.

**Amazon Glacier** 객체 스토리지와 같은 아카이브 플랫폼은 데이터를 수 분 내에 전달하는 잠재적인 "고속" 가져오기와 함께 데이터를 수 시간 내에 전달하는 표준 가져오기 시간을 제공할 수 있습니다. 이러한 점은 "항상 전송된다고 보장할 수 없음" SLA를 기반으로 하므로 성능 보장 옵션이 선불로 제공될 수도 있습니다. 미디어나 지리 공간 애플리케이션과 같은 사용 사례의 경우 페타바이트를 하루 안에 가져오는 작업이 "대량" 가져오기 옵션이나 오프라인 솔루션을 통해서 수행될 수 있습니다.

테이프 교체 시 한 가지 문제는 백업 데이터를 클라우드로 이전하는 것입니다. 공급업체는 다양한 온프레미스 네트워크 전송 방법(예: **AWS Direct Connect** 또는 **AWS Storage Gateway**), 물리적 전송 방법(예: **AWS Snowball** 및 **AWS Snowmobile** 디바이스) 또는 기존 백업 소프트웨어 공급업체와의 통합(백업이 클라우드로 직접 기록됨)을 제공할 수 있습니다.

테이프를 클라우드로 이전하는 프로젝트를 계획할 때는 다음 질문을 고려하여 규모, 규정 준수, **RPO** 및 **RTO**에 대한 프로젝트 요구 사항을 평가해야 합니다.

1. 클라우드로 이전 백업을 저장해야 하니까 아니면 새로운 데이터만 클라우드로 백업되도록 시작한 후 기존 데이터가 수명이 종료되면서 점차적으로 온프레미스 테이프 사용이 줄어들도록 할 수 있습니까?
2. 대용량의 초기 데이터세트를 클라우드로 이전해야 하는 경우, 어떠한 방식을 사용하는 것이 좋습니까?
3. 테이프에서 클라우드로 데이터를 이전할 때 인덱스 지정, 트랜스코딩 또는 리패키지 등 먼저 수행해야 하는 처리 옵션이 있습니까?
4. 나의 백업 공급업체에서 클라우드로의 이전을 기본적으로 지원합니까? 먼저 새로운 기능에 대한 라이선스를 구입하거나 주요 버전 업데이트를 수행해야 합니까? 나의 백업 공급업체가 클라우드 가상 머신(VM)으로의 데이터 복원을 직접 지원합니까? 나의 백업 소프트웨어 공급업체(및 관련 백업 카탈로그)에 대한 내 계획이 내 결정에 영향을 미칩니까?
5. 나의 백업 소프트웨어가 클라우드 기반 인터페이스(**Amazon S3** 프로토콜 등)를 사용하여 데이터를 클라우드로 직접 기록할 수 있습니까? 나의 백업 기간이 이를 지원합니까?

6. 내 백업 소프트웨어가 네트워크 파일 시스템(NFS)과 같은 기존 스토리지 프로토콜을 사용하도록 해야 합니까?
7. 내 조직이 클라우드 기반 VTL로 기존 테이프 기반 워크플로우를 유지 관리할 가치가 있습니까?
8. 지속적인 백업 활동을 방해하지 않고 대규모 변경 사항을 진행할 수 있을 만큼의 테이프 드라이브 용량을 갖추고 있습니까?
9. 기본적으로 클라우드의 데이터를 액세스하고자 합니까? 파일 형식 및 카탈로그 시스템의 영향이 어떻습니까?
10. 클라우드의 글로벌 위치를 활용하여 광범위하게 분산된(예: ROBO) 백업 워크플로우를 간소화할 수 있습니까?
11. 분석, 수명 관리 또는 규정 준수 기능과 같은 클라우드 기능 중 어느 것이 필수 요건입니까?
12. 작업의 규모 및 복잡성 면에서, 마이그레이션을 달성한 외부 통합업체나 전문가 사례를 살펴봐야 합니까?

전체 테이프 아카이브를 읽고 클라우드에 다시 기록하는 것은 만만치 않은 작업이 될 수 있습니다. 테이프 처리, 소프트웨어, 장비 및 이러한 테이프를 소싱, 읽기, 표준화 및 마이그레이션하는 절차와 관련하여 전문 지식을 보유한 타사 파트너의 사용을 고려할 수 있습니다.

## 참조 아키텍처

테이프 백업을 AWS 클라우드로의 백업으로 교체한 다음 아키텍처 예제를 고려해 보십시오.

### 백업 소프트웨어 통합을 통해 클라우드로 백업

Amazon S3 및 Amazon Glacier는 기본적으로 API 기반이기 때문에 인터넷에서도 사용할 수 있습니다. 백업 소프트웨어 공급업체가 아래 그림과 같이 자신의 애플리케이션을 AWS 스토리지 솔루션에 직접 통합할 수 있는 이유도 여기에 있습니다.



그림 1: Amazon S3 또는 Amazon Glacier에 연결되는 백업 커넥터

이 시나리오에서는 테이프에 백업을 저장하기 위해 일반적으로 사용되어 온 기존 온프레미스 백업 및 아카이브 소프트웨어가 AWS와 직접적으로 작동합니다. 백업 소프트웨어는 백업 대상으로 AWS 스토리지 계층에 기록하고 데이터 및 백업 카탈로그에 대한 완전한 가시성을 유지 관리합니다. 이는 작업에 방해 받지 않고 테이프에서 클라우드로 전환할 수 있으며 테이프에 수행되고 테이프 라이브러리나 오프사이트 아카이브에 저장된 백업 작업의 지속성을 유지할 수 있음을 의미합니다. 시간이 지나면서 이러한 백업 작업, 테이프 및 테이프 하드웨어는 서서히 만료될 수 있습니다. AWS와 직접적으로 작동하는 백업 소프트웨어를 사용하는 것은 클라우드 사용을 시작하는 간단하고 효율적인 방법이지만, 기능 통합 세부 사양 및 소프트웨어 버전을 백업 소프트웨어 공급업체와 확인하는 것이 좋습니다. 또한 WAN 성능이 백업 기간 요건을 충족하는지 확인하는 것이 좋습니다.

## 하이브리드 클라우드 가상 테이프 라이브러리를 통해 클라우드로 백업

수십 년 동안 테이프 백업을 수행해 온 많은 IT 조직에서는 테이프 구성과 관련하여 유효성이 검증된 프로세스를 개발해 왔습니다. 백업을 위해 테이프 미디어를 클라우드 스토리지로 교체할 수 있는 두 번째 비중단적 방법은 동일한 프로세스를 사용하는 가상 테이프 라이브러리 및 온프레미스 환경과 AWS 스토리지를 연결하는 브릿지를 사용하는 것입니다. 이러한 설계에서는 기존 온프레미스 테이프 백업 소프트웨어를 사용하여 로컬 디스크와 함께 온프레미스

가상 테이프 라이브러리에 기록함으로써, 최근에 백업된 데이터에 대한 지연 시간을 낮추고 테이프 미디어 대신 클라우드 스토리지 계층을 사용함으로써 보관의 내구성을 높입니다. 로봇 공학이 적용된 테이프 라이브러리, 테이프 슬롯 및 테이프와 같은 제품을 백업하는 것으로 나타나므로, 소프트웨어 및 프로세스가 항상 동일한 방식으로 작동합니다(물리적 테이프 접촉, 라벨 부착, 이동, 저장 및 가져오기 등의 수동 작업 제외). 실제 백업은 로컬 디스크에 수행된 다음 클라우드로 저장됩니다.

**AWS Storage Gateway** 서비스는 업계 표준 iSCSI 기반 VTL로 작동하도록 구성될 수 있으며, 온프레미스 환경과 프로덕션 애플리케이션을 **Amazon S3** 및 **Amazon Glacier**에 연결합니다. 이러한 **AWS Storage Gateway** 구성은 가상 테이프 드라이브 및 테이프와 함께 가상 미디어 체인으로 기존 백업 애플리케이션에 표시됩니다. 기존 백업 애플리케이션 및 워크플로우는 **AWS Storage Gateway**의 온프레미스 가상 테이프 라이브러리에 저장된 가상 테이프 모음에 데이터를 기록합니다. 라이브러리의 가상 테이프는 **Amazon S3**에 비동기적으로 백업됩니다. 가상 테이프에 포함되어 있는 데이터에 빠르게 또는 빈번하게 액세스할 필요가 더 이상 없는 경우 백업 애플리케이션이 가상 테이프 라이브러리의 해당 데이터를 **Amazon Glacier**에 보관하도록 하면 스토리지 비용을 더욱 줄일 수 있습니다. **AWS Storage Gateway**는 **Veritas NetBackup** 및 **Backup Exec**, **EMC Networker**, **Arcserve**, **Veeam**, **Dell Netvault**, **HPE Data Protector**, **Microsoft Data Protection Manager** 및 업계 표준 iSCSI 호환 VTL과 직접 통신하는 다른 제품을 포함한 업계를 선도하는 다양한 백업 애플리케이션과 호환됩니다.

VTL 또는 "테이프 게이트웨이"로 구성된 **AWS Storage Gateway**에는 비용이 많이 드는 테이프 자동화 선수 자본 지출이 없으며 다년 간의 유지 관리 계약 약정을 맺지 않아도 되고 미디어 비용이 지속적으로 발생하지도 않습니다. 테이프에 기록한 데이터 양에 대해서만 요금을 지불하며 비즈니스 규모가 커지면 서비스도 확장됩니다. 이제는 스토리지 미디어를 오프사이트 시설로 배송하고 테이프 미디어를 직접 다루지 않아도 되며, 백업 및 아카이브에 있어 **AWS** 클라우드 스토리지 플랫폼의 우수한 설계 및 내구성의 이점을 누릴 수 있습니다.

다음은 VTL로 배포된 일반적인 **AWS Storage Gateway** 토폴로지를 보여주는 다이어그램입니다.



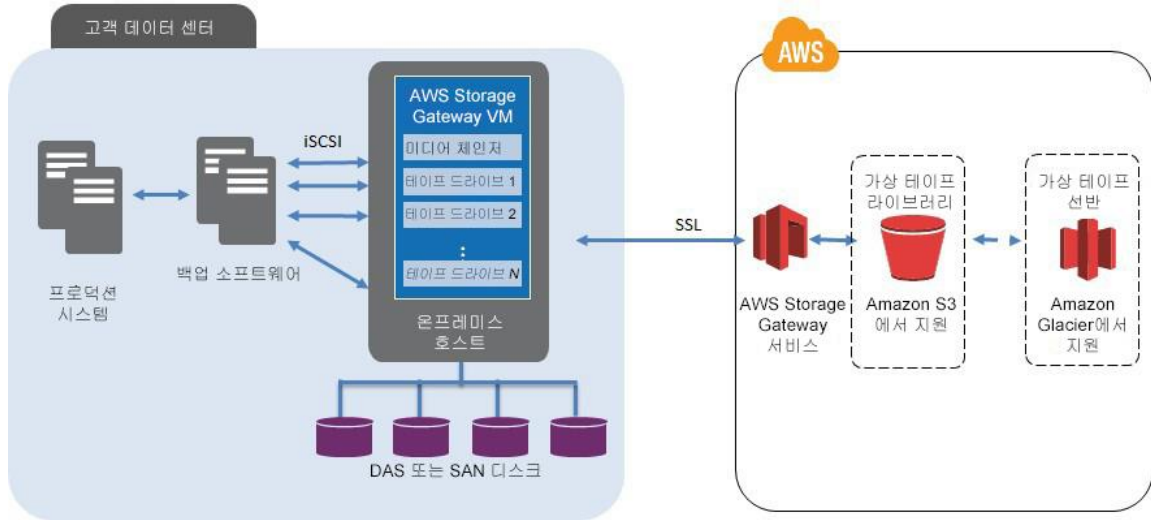


그림 2: 업계 표준 iSCSI VTL로 배포된 AWS Storage Gateway

이 토폴로지는 Amazon Cloud 스토리지에도 백업되어 있는 데이터에 낮은 지연 시간으로 온프레미스 액세스를 가능케 해줍니다. 데이터가 로컬 캐시에 있는 경우, 로컬 디스크 및 네트워크 속도로 복구가 진행되며, 로컬 VTL에 있는 복원해야 하는 데이터 양에 따라 복구 시간이 결정됩니다. 이에 따라 오프사이트 테이프 스토리지나 VTL 사이트별 복제와 관련한 복잡성이 사라집니다. 복구 소프트웨어가 백업을 암호화할 수도 있고 아니면 AWS Storage Gateway와 함께 AWS Key Management Service를 사용할 수도 있습니다.

AWS로부터 AWS Storage Gateway 어플라이언스를 VM으로 다운로드하고 이를 로컬 하이퍼바이저에 설치하면 됩니다. VM 구성은 VMware ESXi 또는 Microsoft Hyper-V에 배포할 수 있습니다. 온프레미스 게이트웨이를 배포하는 경우, 게이트웨이 VM을 다운로드하여 배포한 후 게이트웨이를 활성화합니다. 게이트웨이는 클라우드의 Amazon Elastic Compute Cloud(EC2) 인스턴스에 배포할 수도 있는데, 이 방식은 애플리케이션 워크로드를 Amazon EC2로 이전한 후에도 테이프 백업 구성에 일관된 조직 프로세스를 유지하고자 하는 기업에 유용합니다.

## 온프레미스 NFS 인터페이스를 통해 클라우드로 백업

AWS Storage Gateway는 "파일 게이트웨이"로도 구성할 수 있습니다. 파일 게이트웨이는 온프레미스 NFS 백업 스토리지 대상으로 작동하고 백업 파일을 AWS 클라우드로 저장하여 내구성이 뛰어난 보관을 가능하게 합니다. 비공개로 호스팅되는 VMware 환경에 배포된 파일 게이트웨이는 백업 대상으로 사용되어 지정된 AWS 리전에서 호스팅되는 Amazon S3 버킷과 프라이빗 데이터 센터의 NFS 호환 백업 시스템 간에 성능이 최적화된 연결로서 작동합니다. 파일 게이트웨이는 로컬에 연결된 스토리지를 사용하여 읽기/쓰기 캐시를 제공함으로써 파일 게이트웨이와 동일한 LAN에 있는 백업 서버(NFS 클라이언트로 작동)의 지연 시간을 줄입니다.

그런 다음 게이트웨이 서비스에서 이러한 백업 파일을 지원되는 AWS 리전에 전송하여 Amazon S3 객체로 저장되도록 합니다. 이것이 NFS를 사용할 수 있는 백업 제품용 하이브리드 클라우드 옵션입니다. 복원해야 할 필요성이 사라지면서 데이터는 저렴하고 내구성이 높은 클라우드 스토리지에 저장되고 계층화되어 스토리지 비용이 점차 감소할 수 있습니다. 그림 3에서는 백업이 1년 동안 유지된다고 가정한 예제 아키텍처를 보여줍니다. 30일이 지나면 복원의 필요성이 적어지며 60일이 지나면 거의 없게 됩니다.

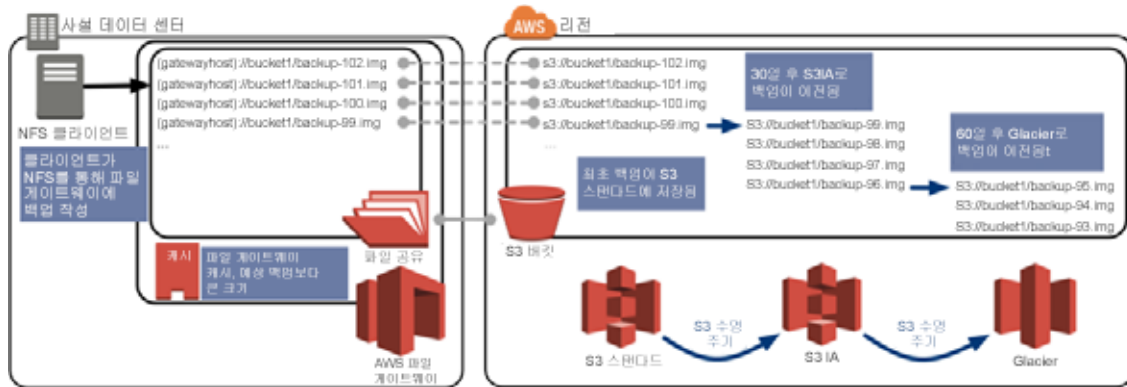


그림 3: Amazon S3 Standard에 파일을 객체로 저장하고 Amazon S3 Standard - IA 및 Amazon Glacier로 전환하는 파일 게이트웨이

이 솔루션에서는 Amazon S3 Standard가 처음 30일 동안 초기 대상입니다. 백업 소프트웨어 또는 스크립트가 백업을 NFS 공유에 기록합니다. 전체 솔루션에서 볼 때 Amazon S3 Standard - Infrequent Access(IA) 및 Amazon Glacier의 낮은 스토리지 비용과 낮은 수명 주기 이전 비용을 포함하여 더 적은 횟수의 이전이 필요한 대용량 파일의 비용 최적화 효과가 더 큼니다.



또 다른 30일이 지나면 백업이 Amazon Glacier로 이전됩니다. 처음 생성 후 1년이 지나 삭제될 때까지 여기에 보관됩니다.

이러한 유형의 솔루션에서 파일 게이트웨이용 캐시 크기를 조정하는 경우 백업 프로세스 자체를 이해하는 것이 중요합니다. 캐시 크기는 완전한 전체 백업을 포함할 수 있을 만큼 충분히 커야 캐시로부터의 직접 복원이 가능하며 속도도 WAN 링크를 통하는 것보다 훨씬 빠르게 복원할 수 있습니다.

백업 솔루션에서 기존 백업을 읽은 다음 진행 중인 백업을 기록하여 백업 파일을 통합하는 소프트웨어를 사용하는 경우, 캐시 크기도 고려하는 것이 중요합니다. 이러한 유형의 작업 도중에는 로컬 캐시에서 읽어야 작업 비용이 감소하고 진행 중인 백업 작업의 전체 성능이 올라가기 때문입니다. NFS 기반 대상으로 AWS Storage Gateway를 사용하여 클라우드에 백업하는 방법에 대한 자세한 내용은 백서 [File Gateway for Hybrid Architectures: Overview and Best Overview and Best Practices](#)를 참조하십시오.<sup>4</sup>

## 결론

테이프 미디어의 비용은 기가바이트당 순수 비용만을 본다면 확실히 낮습니다. 그러나 인건비, 관리 및 초기 구매 비용, 유지 관리, 갱신 및 테이프 미디어 라이선스 비용과 테이프 자동화 시스템 등의 테이프 기반 백업 솔루션의 추가 비용도 다시 평가해 보아야 합니다. 테이프 백업을 클라우드 백업으로 바꾸는 작업은 조직의 데이터를 클라우드로 이전함으로써 추가로 이점을 얻을 수 있는 간단한 작업이며, 다른 작업보다 우선되어야 합니다.

## 기여자

이 문서의 작성에 도움을 준 개인 및 조직은 다음과 같습니다.

- Peter Baltatzidis, 제품 관리자, AWS Storage Gateway
- Peter Levett, 솔루션 아키텍트, Amazon Web Services
- Dylan Locsin, 제품 마케팅 관리자, Amazon Web Services
- Mike Davis, 비즈니스 개발, Amazon Web Services

## 참고 문헌

자세한 내용은 다음을 참조하십시오.

- [간단한 프로젝트 및 구현 가이드](#) 테이프 백업을 클라우드 스토리지로 바꾸는 단계 설명<sup>5</sup>
- 백서: [File Gateway for Hybrid Architectures Overview & Best Practices](#)<sup>6</sup>
- [VTL과 함께 Backup Exec를 사용하는 방법에 대한 Veritas 블로그](#)<sup>7</sup>
- [S3 Connector 또는 VTL과 함께 NetBackup을 사용하는 방법에 대한 Veritas 블로그](#)<sup>8</sup>

## 주석

- <sup>1</sup> IDC, Public Cloud Storage Spending Becomes Strategic and Business-Driven - 유럽에서 실시된 스토리지 관리자 설문 조사, 2015, #EMEA40869715
- <sup>2</sup> IDC, Storage User Demand Study, 2014 - Fall Edition: Growing Demand for Private Storage Cloud, #259067
- <sup>3</sup> "Looking at HDD and SSD Markets Through the Lens of the 'Laws of Economics,'" Nov 2016, IDC, Jeff Janukowicz, John Rydning
- <sup>4</sup> <https://do.awsstatic.com/whitepapers/Storage/aws-storage-gateway-file-gateway-for-hybrid-architectures.pdf>
- <sup>5</sup> <https://aws.amazon.com/getting-started/projects/replace-tape-with-cloud/>
- <sup>6</sup> <https://do.awsstatic.com/whitepapers/Storage/aws-storage-gateway-file-gateway-for-hybrid-architectures.pdf>
- <sup>7</sup> <https://vox.veritas.com/t5/Backup-Exec/Backup-Exec-and-AWS-Storage-Gateway/ba-p/828669>
- <sup>8</sup> <https://vox.veritas.com/t5/Netting-Out-NetBackup/Use-Amazon-Web-Services-to-Store-Your-Backups/ba-p/826330>