

백서

데이터 중복 제거를 활용한 백업: 효율성 향상 및 IT 비용 절감

후원: EMC Corporation

Laura DuBois

2009년 5월

핵심 요약

데이터 중복 제거 기술은 분산된 사이트와 데이터 센터 모두에서 네트워크 대역폭, 백업 윈도우 및 스토리지 공간 요구량을 최적화하여 IT 환경의 경제성을 크게 향상시킨다. 기업들은 실제로 운영 환경에서 데이터 중복 제거를 활용하여 백업 효율성을 향상시키고 IT 비용을 절감할 수 있다. 본 백서에서는 백업 데이터에서 중복된 데이터를 제거할 수 있는 다양한 방법을 살펴보고 데이터 중복 제거 솔루션을 선택할 때 고려해야 할 사항을 간략하게 설명한다. 이 밖에도 데이터 중복 제거 기술이 포함된 EMC 백업 포트폴리오를 살펴보고 백업 효율성을 최적화하고 비용을 절감하기 위한 구체적인 활용 사례를 소개한다.

데이터 중복 제거의 도입 배경

기업들이 매년 2배 가까이 증가하는 스토리지 용량에 효과적으로 대처하기 위한 방안을 모색함에 따라 중간 규모 환경과 대규모 환경 모두에서 데이터 중복 제거에 대한 수요는 계속해서 증가하고 있다. 최근에는 새로운 애플리케이션 도입, 가상화 확대, 전자 문서 스토어 생성 및 문서 공유, Web 2.0 기술 사용, 디지털 기록 보존과 같은 요인까지 더해져 이러한 스토리지 증가세가 더욱더 가속화되고 있다. 한정된 IT 예산으로 인해, 기업들에게 자본 및 운영 비용을 절감하는 것이 시급한 과제로 대두되면서 스토리지 증가세에 제동을 걸어야 할 필요성은 더욱더 높아졌다. 물리적 측면에서 볼 때, 많은 데이터 센터 관리자들 또한 전력, 냉각, 설치 공간이 제약된 인프라스트럭처로 인해 어려움을 겪고 있다. 데이터 중복 제거 기술은 비용을 절감하여 스토리지 효율성을 향상시키는 데 도움을 주고 물리적 제약이 있는 데이터 센터의 문제점들을 완화한다.

데이터 중복 제거 기술은 관리, 백업 및 네트워크 비효율성과 관련된 과제도 해결해 준다. 데이터가 증가하면서 IT 담당자 수와 이들의 관리가 필요한 스토리지 용량 간의 불균형은 더욱 심화되고 있다. 데이터 중복 제거는 데이터 용량을 줄여 이 둘 간의 균형을 유지할 수 있게 해 준다. 마찬가지로 서버 처리 성능과 디스크 간의 격차가 갈수록 벌어짐에 따라 WAN을 통한 전체 운영 환경에서, 디스크 스토리지 서브시스템 내에서, 제한된 백업 윈도우 전반에 걸쳐 성능을 향상시킬 방법을 모색하는 기업이 늘고 있다. 데이터 중복 제거 기술은 로컬 또는 원격 네트워크 링크를 통해 전송되는 데이터의 양을 감소시켜 사용 가능한 물리적/가상화 인프라스트럭처를 최적화할 수 있을 뿐 아니라, 서비스 수준 응답 시간을 단축하고 갈수록 짧아지는 백업 윈도우를 충족할 수 있다. 데이터 중복 제거 기술은 또한 랜덤 액세스 미디어(디스크)를 사용하여 복구 시간, 데이터 보안 및 신뢰성을 향상시킨다.

최근에는 가상화에 따른 과제도 대두되고 있다. 기업들이 서버 통합 및 재해 복구를 지원하기 위해 가상 시스템 기술을 도입함에 따라 가상 시스템에서 중복 데이터를 처리하고 있는데, 이러한 데이터 또한 보호가 필요하다. 일반적으로 다양한 장애 시나리오에 대비하거나 이미지를 복구하려면 하나의 백업 솔루션과 백업 프로세스로 물리적 서버와 개별 파일을 관리해야 한다. 게스트 가상 시스템에 기존의 백업 에이전트를 구축하거나 VCB 프록시 백업을 사용하는 것과 같은 일반적인 방법으로는 백업할 가상 시스템 데이터 볼륨이나 네트워크 대역폭 요구 사항을 줄일 수 없다. 데이터 중복 제거는 백업 스토리지 용량 절감 효과가 뛰어나다. 또한 일부 데이터 중복 제거 기술은 백업할 데이터 양을 줄여 결과적으로 백업 속도를 향상시키고 네트워크에 미치는 영향을 최소화한다. 기업은 백업 소프트웨어와 데이터 중복 제거 기술을 함께 사용하여 가상 시스템 환경을 효율적이고 경제적으로 완벽하게 보호할 수 있다.

데이터 중복 제거의 이점

기업들은 이처럼 실질적인 과제들을 해결하기 위해 인프라스트럭처 스택의 여러 지점에 데이터 중복 제거 기술을 구축하고 있다. 데이터 중복 제거의 이점은 다음과 같다.

- ☒ **비용 절감.** 데이터 중복 제거는 데이터 센터의 전력, 냉각 및 공간 소비와 스토리지 용량, 네트워크 대역폭, IT 인력 등을 감소시켜 리소스 효율성을 향상시키고 비용을 절감한다.
- ☒ **탄소 배출량 감축.** 데이터 중복 제거는 스토리지의 전력, 냉각, 공간 소비를 줄여 결과적으로 탄소 배출량을 감축하고 환경 보호를 실현한다.
- ☒ **백업 및 복구 서비스 수준 향상.** 데이터 중복 제거 기술은 백업 성능을 획기적으로 향상시켜 제한된 백업 윈도우를 충족할 수 있게 해 준다. 또한 랜덤 액세스 디스크 스토리지를 활용하므로 순차 액세스(테이프) 방법보다 복구 성능이 훨씬 더 뛰어나다.
- ☒ **디스크 백업을 통해 테이프 백업과 동일하거나 더 높은 수준의 비용 절감 효과 달성.** 데이터 중복 제거 기술을 활용하면 더욱 폭넓은 애플리케이션을 대상으로 디스크 기반 백업을 수행할 수 있다. 테이프는 경제성과 아카이빙을 이유로 엔터프라이즈 데이터 센터에서 폭넓게 사용되어 왔다. 하지만 디스크 기반의 데이터 중복 제거 기술을 사용하면 GB당 비용이 줄어들어 디스크 비용이 테이프 비용 이하로 절감될 수 있다.

데이터 중복 제거 기술은 2배씩 증가하는 막대한 양의 데이터를 처리하고, 점점 더 단축되고 있는 백업 윈도우를 충족하며, 운영 및 재해 관련 장애가 발생했을 때 데이터를 보다 신속하게 복구하는 등 모든 규모의 기업에서 십 년 넘게 고민해 온 수많은 백업 과제를 해결해 준다.

표 1에는 현재 기업이 당면하고 있는 수많은 백업 관련 과제, 이와 관련해 데이터 중복 제거를 통해 실현할 수 있는 효과, 그리고 각각의 백업 관련 과제를 해결하는 데 가장 적합한 데이터 중복 제거 방식이 간략하게 정리되어 있다.

표 1

백업 관련 과제 및 데이터 중복 제거 효과

백업 관련 과제	데이터 중복 제거 효과	가장 적합한 데이터 중복 제거 방식
다운타임 비용을 최소화하기 위한 복구 시간 단축	데이터 중복 제거를 사용하면 보다 적은 비용으로 더 많은 백업 데이터를 디스크에 저장할 수 있다. 백업본을 테이프가 아닌 디스크에 보관하면 여러 애플리케이션의 복구 시간이 크게 단축된다.	소스 또는 타겟 측 데이터 중복 제거
백업 데이터의 복구 신뢰성 보장 문제	테이프 미디어에 의존한 백업 방식은 미디어 손상, 헤드 오염 등의 미디어 오류를 비롯해 사용 가능한 미디어 부족 또는 하드웨어 장애 등에 따른 여러 위험요소를 초래할 수 있다. 데이터 중복 제거는 데이터 보호 프로세스에서 디스크를 사용하기 때문에 이와 같은 장애가 발생할 가능성이 거의 없다.	소스 또는 타겟 측 데이터 중복 제거
글로벌 고객 요구에 부응하기 위한 24x7 업무 운영에 따른 백업 윈도우 단축	방대한 양의 중복된 데이터를 전송하는 기존의 백업 방식은 존재하지도 않거나 있더라도 빠른 백업 윈도우를 초과하기 일쑤이다. 데이터 중복 제거는 백업할 데이터 양을 감소시켜 결과적으로 사용할 수 있는 백업 윈도우에서 더 많은 데이터를 백업할 수 있게 해 준다.	소스 측 데이터 중복 제거
서버 가상화 확대 에 따른 사용 가능한 백업 리소스 감소로, 백업 시간 증가 및 백업 윈도우 부담 가중	소스 측 데이터 중복 제거를 사용하면 중복된 데이터의 공유 리소스 처리가 필요하지 않아 경합이 감소하고 가상 시스템의 백업 속도가 향상된다.	소스 측 데이터 중복 제거
급속한 데이터 증가 로 인해 사용할 수 있는 백업 윈도우에서 모든 데이터를 백업할 수 없음	기업에서는 보호해야 할 데이터의 양이 매년 평균 50% 씩 증가한다. 제한적인 야간 백업 윈도우와 기존의 백업 방식은 이러한 데이터 증가세를 따라잡지 못하는 실정이다. 그러나 데이터 중복 제거를 사용하면 데이터 증가 문제를 해결하고 지속적으로 증가하는 데이터 세트를 효율적으로 백업할 수 있다.	소스 측 데이터 중복 제거
테이프를 이용한 기존 오프사이트 복제본 보관 방식으로 인해 미디어 손실 또는 절도에 따른 데이터 위험 발생 가능	이동식 테이프 미디어를 오프사이트 보관용으로 사용하면 재해 발생 시 해당 물리적 미디어가 위험에 노출될 가능성이 있다. 데이터 중복 제거를 안전한 복제 프로세스와 함께 사용하면 전자 복제본을 오프사이트에 보관할 수 있어, 테이프 미디어를 수작업으로 처리하지 않아도 되고 데이터 보안이 향상된다.	소스 또는 타겟 측 데이터 중복 제거
용량 증가 및 백업 윈도우에 따른 백업 인프라스트럭처 비용 증가	대부분의 기업들은 테이프 인프라스트럭처를 늘리는 방법으로 데이터 증가 및 백업 윈도우 문제를 해결한다. 테이프 드라이브와 자동화를 사용하면 지금 당장의 성능 병목 현상을 해소하고 백업 시간을 단축할 수는 있어도 비용과 관리 부담은 상승하게 된다. 데이터 중복 제거를 사용하면 이처럼 데이터 및 백업 윈도우 변화에 맞게 테이프 인프라스트럭처에 지출하는 비용을 절감할 수 있다.	소스 또는 타겟 측 데이터 중복 제거

출처: 2009년 IDC

데이터 중복 제거: 의미, 위치, 시기 및 방법

데이터 중복 제거의 의미

IDC에서는 데이터 중복 제거를 ‘중복된 데이터를 하나의 공유 데이터 객체로 표준화하여 스토리지 용량 효율성을 높이는 기술’로 정의하고 있다. 즉, 데이터 중복 제거란 블록, 청크, 파일 등의 중복된 데이터 객체를 검색해서 제거하는 모든 알고리즘을 가리킨다. 중복된 데이터가 발견되면 더 이상 보존되지 않고, 대신 디스크에 이미 저장되어 있는 데이터 객체의 정확한 복제본을 스토리지 시스템이 참조하도록 “데이터 포인터”가 수정된다. 또한 데이터 중복 제거는 고유한 데이터만 처리하므로 동일한 데이터 객체의 복제본을 여러 개 보관하는 데 따른 비용이 절감된다.

파일이나 객체 레벨에서 중복을 제거하는 SIS(Single-Instance Storage)와 달리, 데이터 중복 제거는 하위 파일 중복 제거 프로세스와 연관이 있는 경우가 많다. 하위 파일 중복 제거에서는 파일을 검사하여 “청크” 단위로 분할한다. 그런 다음 이 작은 크기의 청크로 여러 시스템과 사이트에 중복된 데이터 콘텐츠가 있는지 확인한다. 데이터 중복 제거는 여러 파일이나 한 파일의 여러 조각이 아닌 하나의 객체가 차지하는 공간을 줄이는 압축과는 다르다. 그렇지만 중복 제거된 데이터를 압축하여 공간을 추가로 절약할 수는 있다.

데이터 중복 제거의 위치

백업 데이터 중복 제거는 소스나 타겟에서 수행할 수 있다. 소스 측 데이터 중복 제거는 백업 프로세스 동안 고유한 하위 파일 데이터만 네트워크를 통해 전송되도록 Exchange 또는 파일 서버 등의 클라이언트에서 백업 데이터의 크기를 줄이는 경우 등을 가리키며, 타겟 측 데이터 중복 제거는 백업 데이터가 네트워크를 통해 전송된 후 데이터 중복 제거 어플라이언스에 도달했을 때 크기를 줄이는 경우 등을 가리킨다. 소스 측 데이터 중복 제거는 네트워크 대역폭, 백업 윈도우 및 스토리지 공간의 절감 효과를 제공한다. 타겟 측 데이터 중복 제거는 스토리지 공간을 절약하고 기존 백업 소프트웨어를 활용하는 것이 가능하며 네트워크에 미치는 영향을 줄인다는 이점이 있지만, 이를 수행하려면 모든 위치에 하드웨어 어플라이언스가 구축되어 있어야 한다. 데이터 중복 제거 기술을 어느 위치에 구축하느냐에 따라 얻을 수 있는 이점과 구축 시간 및 비용이 달라지므로 기업들은 현재 당면한 백업 문제를 제대로 파악한 후 가장 적합한 데이터 중복 제거 방식을 선택해야 한다(표 1 참조).

소스 측 데이터 중복 제거

소스(클라이언트)에서 데이터 중복 제거를 수행하면 용량 최적화 이상의 효과를 얻을 수 있다. 즉, 전송되는 데이터 양을 크게 감소시켜 혼잡한 가상화/물리적 인프라스트럭처 및 LAN/WAN 링크의 워크로드를 줄일 수 있다. 또한 새로 추가되었거나 변경된 하위 파일 데이터 세그먼트만 전송되므로 이동되는 데이터 양이 크게 줄어 일일 전체 백업 속도가 획기적으로 향상된다. 소스 측 데이터 중복 제거를 수행하면 클라이언트 CPU에 대한 오버헤드는 최대 15%까지 증가하지만 기존 방법에 비해 훨씬 빠른 속도로 백업을 완료할 수 있다. 실제로 7일 동안 소스 측 데이터 중복 제거가 전체 환경에 미치는 영향은 기존 에이전트에 비해 훨씬 적다. 하지만 대규모의 데이터베이스 또는 일일 데이터 변경률이 높은 데이터베이스가 구축된 환경에서는 타겟 측 데이터 중복 제거 솔루션을 고려하는 것이 더 적합한 경우가 많다. 다행히도 대부분의 공급업체는 고객이 가장 적합한 솔루션을 선택할 수 있도록 도와 주는 데이터 진단 툴을 제공한다. 소스 측 데이터 중복 제거의 경우, 소규모 원격 사무소에서 로컬 하드웨어를 추가로 설치할 필요 없이 소프트웨어 백업 에이전트만 구축하여 수행할 수 있기 때문에 구축 유연성도 뛰어나다.

타겟 측 데이터 중복 제거

타겟에서 데이터 중복 제거를 수행하면 새롭게 추가된 고유한 하위 파일 데이터만 디스크에 저장되므로 백업 디스크 스토리지 용량을 최적화할 수 있다. 하지만 기존의 백업 소프트웨어를 통해 여전히 중복된 백업 데이터가 데이터 중복 제거 타겟에 전송되기 때문에 백업 윈도우를 단축하는 데는 도움이 되지 않는다. 타겟 측 데이터 중복 제거 방식을 사용할 때 중요하게 고려해야 할 요소는 백업 윈도우에 맞출 수 있는지 여부, 그리고 특정 워크로드에 대해 인라인(Inline) 또는 사후 처리(Post-process) 방식의 중복 제거가 보장되는지 여부이다. 인라인 또는 사후 처리 방식의 중복 제거에 대한 자세한 내용은 다음 섹션에서 참조할 수 있다.

그 밖에도 타겟 측 데이터 중복 제거 방식을 사용하려면 증가분 데이터 중복 제거 어플라이언스를 구입해야 하는데, 그러기 위해서는 예산을 추가로 확보해야 하고 어플라이언스를 구축한 후에도 다른 시스템과 마찬가지로 관리해야 한다. 어플라이언스의 용량이 부족하게 되면 데이터 중복 제거 어플라이언스를 추가로 구축해야 한다. 일부 솔루션의 경우에는 이러한 문제를 줄이기 위해 여러 어플라이언스의 글러스터링 기능을 제공한다. 타겟 측 데이터 중복 제거 방식은 대용량 데이터 관리를 위한 중앙 데이터 센터와 원격 사이트에서 모두 사용할 수 있다. 하지만 이를 위해서는 여러 지사의 데이터를 데이터 센터의 중앙 집중식 대용량 어플라이언스로 원격 복제할 수 있도록 데이터 중복 제거 어플라이언스를 각 지사에 구축해야 한다. 반면 소스 측 데이터 중복 제거는 이러한 원격 지사 하드웨어 투자가 필요하지 않다.

데이터 중복 제거의 시기

오늘날 데이터 중복 제거 프로세스의 실행 시기를 결정하는 방식은 대개 인라인과 사후 처리의 두 가지로 나뉜다. 일부 공급업체는 이와 다른, 하이브리드(Hybrid) 또는 어댑티브(Adaptive) 중복 제거라는 방식을 사용하기도 한다. 인라인 중복 제거는 중복된 데이터가 디스크에 작성되기 전에 제거하므로 디스크 스테이징 영역이 필요하지 않다. 반면에 사후 처리 중복 제거는 데이터가 디스크에 저장된 후에 데이터를 분석 및 제거하므로 데이터 중복 제거 프로세스를 시작할 수 있는 충분한 용량을 갖춘 스테이징 영역이 필요하다. 기업에서는 백업 속도와 디스크 용량을 고려해서 데이터 중복 제거 방식을 선택해야 한다.

인라인 프로세스는 비교적 용량 효율성이 높기 때문에 지연 없이 데이터 중복 제거 프로세스를 바로 시작할 수 있다. 백업 윈도우를 고려해야 하는 대용량 환경의 경우 사후 처리 중복 제거 방식은 백업 작업을 완료하는 데 우선 순위를 두지만, 이를 위해서는 훨씬 더 많은 초기 스토리지 용량이 필요하다. 이 두 가지 방식은 성능 및 용량 요구 사항 측면에서 장단점이 있다. 이 밖에 아직 개발 단계에 있는 데이터 중복 제거 방식으로 하이브리드 또는 어댑티브 중복 제거가 있다. 이 데이터 중복 제거 방식은 인라인 방식을 우선적으로 적용하다가 성능 임계값에 도달하면 자동으로 사후 처리 방식으로 전환함으로써 고객 환경의 현재 워크로드에 맞게 데이터 중복 제거 방식을 조정한다. 일부 주요 솔루션은 고객의 필요에 맞게 데이터 중복 제거 설정을 구성하여 즉시 또는 스케줄에 따라 데이터 중복 제거를 실행하거나, 데이터 세트 특성을 기준으로 데이터 중복 제거를 비활성화하도록 하는 정책 기반 중복 제거 기능을 제공한다. 예를 들어, 용량이 작은 데이터 세트와 비정형 데이터는 즉시 데이터 중복 제거 대상으로 설정하고, 용량이 큰 백업 작업은 사후 처리 방식으로 구성하는 한편, 리치 미디어나 암호화된 데이터의 백업에 대해서는 데이터 중복 제거를 해제할 수 있다. 정책 기반 중복 제거를 활용하면 고객 각자의 환경 조건에 맞게 데이터 중복 제거를 설정할 수 있어 유연성이 크게 향상된다.

데이터 중복 제거의 방법

데이터 중복 제거 프로세스가 실행되는 방법은 구축 방식에 따라 다르다. 해시 기반 중복 제거 방법은 파일 또는 백업 스트림을 고정 또는 가변 길이의 하위 파일 데이터 청크로 분할한다. 해시 값은 청크별로 계산된다. 이 프로세스에서 각각의 청크에 대해 계산된 고유한 번호는 인덱스에 저장된다. 파일이 업데이트되면 변경된 하위 파일 데이터만 저장되며, 변경 내용 때문에 완전히 새로운 파일을 저장할 필요는 없다. 해시 기반 구축은 청크 크기의 고정 또는 가변 여부에 따라 두 가지 방식으로 구분된다. 가변 길이 방식은 콘텐츠 유형에 따라 청크 크기를 탄력적으로 조정하여 파일 변경 내용 중 바이트 스트림에서 위치가 변경되었거나 오프셋된 중복된 데이터 청크를 지원할 수 있다. 반면에 고정 길이 방식은 위치가 변경되었거나 오프셋된 중복된 데이터를 인식하지 못해 이미 백업 저장소에 있는 청크도 다시 백업하므로 효율성이 떨어진다. 해시 기반 방식을 사용할 경우에는 성능 및 디스크 입출력과 관련한 문제가 발생할 수 있다. 해시 인덱스는 기본적으로 메모리에 유지되지만 해시 인덱스의 크기가 증가하면 메모리로는 공간이 부족해 디스크를 사용하게 되므로 조회 및 청크 검색에 디스크 입출력이 필요하게 된다. 공급업체에서는 이러한 실질적 기술 문제를 처리하기 위한 다양한 솔루션을 제공한다.

해시 기반 중복 제거를 대체하는 방식인 델타 기반 중복 제거는 기본 복제본과 다른 형식의 데이터를 저장 또는 전송한다. 여기서 기본 복제본이란 다른 버전의 데이터를 재생성하는 데 사용되는 완전한 데이터 복제본을 의미한다. 델타 기반 중복 제거는 블록 또는 바이트 레벨에서 수행할 수 있다. 해시 기반과 델타 기반 중 어떤 방식을 사용하는 데이터 중복 제거율은 성능과 반비례하게 된다. 쉽게 말해, 청크 크기가 클수록 데이터 중복 제거율은 떨어지고 청크 크기가 작을수록 인덱싱 오버헤드는 증가한다.

데이터 중복 제거율은 데이터 중복 제거 엔진이 특정 백업 애플리케이션, **Microsoft Exchange** 데이터 같은 특정 데이터 형식을 인식할 수 있는지 여부에도 영향을 받는다. 데이터 형식을 인식할 수 있으려면 애플리케이션별 메타데이터가 스트림에 삽입되는 위치를 알아야 한다. 그래야 데이터 중복 제거 엔진이 본래 애플리케이션의 데이터 형식에 맞게 청크 크기를 조정하여 데이터 중복 제거의 효과를 향상시킬 수 있다.

데이터 중복 제거 기술 평가 시 고려할 사항

현재, 데이터 중복 제거 기능을 갖춘 다양한 유형의 제품이 시판되고 있다. 백업 애플리케이션에서 어플라이언스, 가상 테이프 라이브러리, WAN 최적화 솔루션, 운영 디스크 스토리지 서브시스템 등에 이르는 모든 제품이 다양한 형태의 데이터 중복 제거 기능을 제공한다. 기업에서는 데이터 중복 제거 제품을 선택하기에 앞서 해당 제품을 통해 애플리케이션 또는 데이터 유형별로 어떤 문제를 해결할 수 있는지 꼼꼼히 따져봐야 한다. 데이터 중복 제거 방식에 따라 지원되는 용량, 성능 및 네트워크 효율성도 제각기 다르기 때문이다.

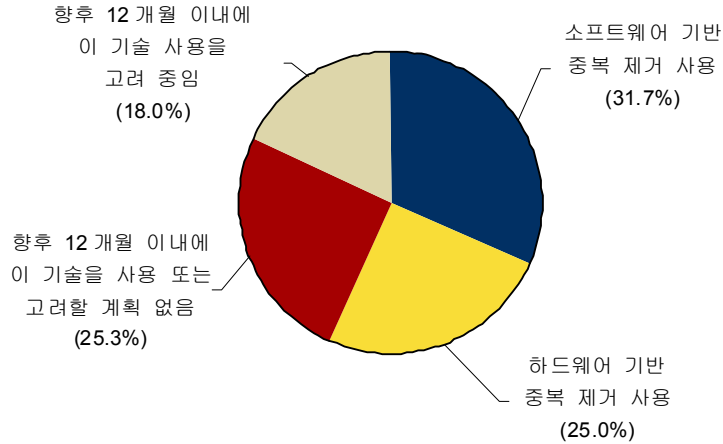
1. **데이터 중복 제거율.** 데이터 유형, 데이터 변경률, 보존 기간, 가변 및 고정 길이 세그먼트, 백업 정책, 파일 형식 인식 등의 다양한 요인에 따라 실현할 수 있는 데이터 중복 제거율에도 차이가 있다. IDC가 앞서 언급한 요인을 바탕으로 실제 환경에서 전체 백엔드 디스크 스토리지 데이터 중복 제거율을 조사한 결과에서는 그 값이 8:1에서 22:1 사이인 것으로 나타났다. 소스 측 중복 제거 솔루션은 기존의 일일 전체 백업 방식에 비해 일일 네트워크 대역폭 요구량을 큰 폭으로 절감할 수 있다. 몇몇 주요 공급업체에서는 최고 500:1 수준까지 줄일 수 있다고 주장하기도 한다. 그러나 모든 성능 지표가 그렇듯이 이러한 수치는 해당 환경에 따라 달라진다. 그러므로 기업에서는 데이터 중복 제거 솔루션이 제공하는 처리량, 확장성, 성능을

명확하게 파악하고 실제 운영 환경에서 자체 데이터 세트로 해당 데이터 중복 제거 솔루션을 직접 테스트해 볼 필요가 있다.

2. **압축, 암호화 및 멀티플렉싱의 역할.** 데이터 인코딩을 통해 저장 용량을 줄이는 압축은 데이터 중복 제거를 보완하는 기술로 활용할 수 있다. 압축은 단일 객체에 최적화된 기술로 단일 객체의 저장 공간을 줄이는 데 반해, 데이터 중복 제거는 여러 객체를 대상으로 수행된다. 하지만 압축을 이미 중복 제거된 데이터에 적용하면 공간을 추가로 절약할 수 있다. 반대로 데이터 중복 제거를 이미 압축되었거나 암호화된 파일에 적용할 경우에는 공간 절약 효과를 거의 얻을 수 없다. 소프트웨어 계층에서 암호화 또는 압축을 사용하면서 타겟 기반 중복 제거를 활용하는 기업에서는 데이터 중복 제거 효과를 얻기 위해 이러한 기능을 비활성화하거나 하드웨어 기반 압축을 활용해야 할 수 있다. 또한 여러 클라이언트의 데이터를 하나의 스트림으로 인더리빙하여 테이프 드라이브에 전송하는 현재의 백업 멀티플렉싱 방식에 대해서도 고려해야 한다. 하지만 이 프로세스를 사용하면 이미 존재하는 데이터 세그먼트를 감지하기가 어렵다. 데이터 중복 제거의 효과를 실현하려면 멀티플렉싱 기능을 사용하지 않아야 한다.
3. **가상 시스템의 데이터 중복 제거.** 운영 환경에서 가상 시스템을 사용함에 따라 가상 시스템, 물리적 호스트 및 파일을 보호하고 복구해야 할 필요성이 높아졌다. 가상 시스템 백업 옵션으로는 각 게스트의 에이전트, VCB 프록시 서버 백업 또는 서비스 콘솔의 에이전트 등을 들 수 있다. 기존의 백업 솔루션은 대량의 중복된 데이터를 전송하고 백업을 실행하는 데 많은 CPU 리소스를 소모하여 결과적으로 백업 성능을 저하시키고 서버 통합 수준을 떨어뜨리기 때문에 가상 시스템을 효율적으로 백업하기 위한 솔루션으로는 적합하지 않다. 그러나 데이터 중복 제거 기술을 사용하면 이러한 한계를 극복할 수 있다. 소스 측 중복 제거를 사용하면 중복된 데이터가 기본 물리적 인프라스트럭처를 통해 전송되지 않으므로 일일 전체 백업이 빠르고 효율적으로 수행된다. 또한 데이터 중복 제거는 VMDK 전체에서 전사적으로 실행되기 때문에 시스템 전체에서 중복된 데이터를 백업할 필요가 없다. 이외에도 VMDK 파일이 업데이트될 때 오프셋된다는 점을 고려해야 한다. 가변 길이 중복 제거를 사용할 경우에만 이 오프셋을 인식하고 VMDK 내에서 고유한 변경 내용을 찾아서 전송할 수 있다.
4. **원격 사이트의 데이터 중복 제거.** 데이터 센터의 경우와 마찬가지로 원격 사이트에서도 로컬 복구 작업과 재해(원격) 복구 작업을 수행해야 한다. 하지만 원격 사이트는 그 특성상 몇 가지 해결해야 할 과제를 안고 있다. 원격 사이트는 일반적으로 WAN 대역폭이 제한되어 있고 전담 IT 인력이 없으며, 해당 지사 수와 지역 또는 주 데이터 센터 수 간에 균형이 맞지 않는 경우가 많다. 한편 데이터 중복 제거는 WAN을 통해 이동하는 데이터 양을 최소화하며, 전사적 중복 제거는 원격 사이트와 데이터 센터 전반에 걸쳐 중복된 데이터를 제거한다. 원격 지사의 IT 인력이 부족한 대부분의 기업에서는 여러 곳에 분산된 지사에서 스토리지 하드웨어 설치 규모를 최소화할 방법을 찾고 있다. 소스 측 중복 제거는 소프트웨어를 통해 구축할 수 있으므로 이들 기업에게 훌륭한 해결책이 되고 있다. 최근, IDC는 원격 지사의 데이터 중복 제거 사용과 구축 유형에 대한 조사를 실시했다(그림 1 참조).

그림 1

원격 지사 데이터 보호를 위한 데이터 중복 제거 기술 사용 현황



응답자 수= 300

출처: 2009년 IDC의 Remote Branch Special Study

- 5. 운영/재해 복구 데이터 센터의 데이터 중복 제거.** 데이터 센터 환경에서는 데이터 볼륨이 급속히 증가하지만 LAN 접속 구성과 재해 복구 사이트로의 신속한 접속과 같은 이점이 있다. 여전히 많은 대규모 데이터 센터에서는 중요한 일부 애플리케이션에 대한 백업 윈도우를 충족하면서 일정 수준의 백업 성능을 유지하기 위해 애쓰고 있다. 그렇기 때문에 해당 애플리케이션 및 환경에 따라 소스 및 타겟 중복 제거를 포함한 정책 기반 중복 제거 방식이 필요할 수 있다. 데이터 센터 내에서 네트워크 대역폭을 최적화하는 작업은 재해 복구 사이트로 데이터를 원격 복제하는 작업에 비해 우선 순위가 낮을 수 있다. 하지만 백업 윈도우는 계속해서 줄어들고 있는 추세이므로 앞으로 네트워크 대역폭은 심각한 문제로 대두될 가능성이 크다.
- 6. 전사적 데이터 중복 제거.** 전사적 데이터 중복 제거는 다양하게 정의할 수 있다. 전사적 데이터 중복 제거의 범위를 사이트 전체로 보는 견해도 있고, 단일 스토리지 프레임 이내로 보는 견해도 있다. 그러나, 전사적 데이터 중복 제거의 이점을 최대한 누리기 위해서는 그 범위를 사이트와 프레임 전체로 보는 것이 바람직하다. 타겟 측 중복 제거 기술을 사용할 경우 다수의 원격 사무소에서 단일 프레임 및 그 복제 페어를 대상으로 전사적 데이터 중복 제거를 수행할 수 있다. 그렇지만 용량 또는 성능 임계값에 도달하면 새로운 어플라이언스를 설치해야 하는데, 이 경우 독립 실행형 데이터 중복 제거 어플라이언스도 도입해야 한다. 소스 측 중복 제거 기술을 사용할 경우에도 다수의 원격 사무소 데이터 센터 서버 및 그 복제 페어를 대상으로 전사적 데이터 중복 제거를 수행할 수 있다. 전사적 데이터 중복 제거는 공급 업체와 그 접근 방식에 따라 의미가 달라질 수 있는 용어임을 명심해야 한다.

7. **데이터 중복 제거 및 복제.** 복제는 데이터 중복 제거 기술이 사실상 다시 한번 승부를 벌일 분야이다. 데이터 중복 제거는 업계의 선두 공급업체나 이에 도전하는 공급업체를 불문하고 공히 효과적이라는 사실이 검증되었으며, 데이터 중복 제거 기술을 평가한 사용자의 반응도 매우 긍정적이었다. 데이터 중복 제거 기술은 중앙 사이트와 원격 사이트를 포함한 모든 엔터프라이즈 환경에 구축되어 높은 효율성 향상 및 인프라스트럭처 비용 절감 효과를 제공하고 있다. 데이터 센터의 테이프를 보다 전략적 차원에서 사용하여(예: 규정 준수와 같은 활용 사례) 원격 사이트의 테이프 사용을 최소화하려는 기업이 늘어나면서 원격 복제의 역할은 점점 중요해지고 있다. 복제에 대한 사용자들의 요구 수준도 다음과 같이 점차 고도화하고 있는 추세이다.

- **전체 볼륨이 아닌 중복 제거된 데이터 세트만 복제하는 데이터 중복 제거 지원 복제.** 일부 공급업체는 데이터 중복 제거를 지원하는 제품과 함께 복제 서비스를 제공한다. 단, 복제 기능이 데이터 중복 제거를 지원하는지는 각 솔루션 별로 확인해야 한다.
- **전체 복제 또는 디렉토리/테이프 레벨에서의 선택적 복제.** 활용 사례에 따라 시스템 전체 복제를 지원하거나 복제할 공유 저장소 또는 테이프를 유연하게 결정할 수 있는 기능이 필요할 수도 있다.
- **복제 모니터링, 성능 튜닝 및 문제 해결.** 데이터 중복 제거에도 불구하고 대부분의 대기업에서는 여전히 많은 양의 데이터를 복제해야 한다. 데이터 복제는 스케줄에 따라 또는 비동기식으로 수행되는 복제 프로세스를 통해 관리되며 복제 프로세스와 대역폭 사용량에 대한 모니터링이 수행된다. 튜닝 및 문제 해결 툴은 복제 프로세스가 허용된 복제 윈도우를 벗어나지 않도록 지원한다.
- **링크 지연을 고려한 스케줄에 따른 복제/실시간 복제.** 일부 링크/사이트에는 실시간 복제가 필요한 반면, 또 다른 링크/사이트는 스케줄을 지정하여 복제를 수행하는 것으로도 충분할 수 있다. 원격 지사마다 특성이 다르고 링크 지연이 낮을 수도 있으며, 두 데이터 센터 간 링크에서도 지연 양상이 매번 다르게 나타날 수 있기 때문이다.

8. **시드 서비스 및 마이그레이션.** 데이터 중복 제거는 스토리지 공간을 절약하고 중복된 데이터의 전송을 감소시키는 데는 효과적이지만 기본 복제본 또는 최초 백업본이 설정되어 있어야 한다. **Edge-to-Core** 데이터 중복 제거를 수행하려면 대역폭이 제한된 링크를 통해 이 기본 복제본을 어떻게 생성할 것인지 검토가 필요하다. 대부분의 공급업체는 배치된 여러 대의 시스템에서 데이터 중복 제거를 지원하는 대량 복제 프로세스를 사용하거나, 데이터 중복 제거 노드로 사용할 수 있는 시스템에 최종 전체 백업본 테이프에 저장된 데이터를 로컬로 복구하는 방법을 사용하여 기본 복제본을 신속하게 생성할 수 있는 일종의 시드 서비스를 제공한다. 스토리지 교체 주기가 3년~5년 정도인 경우에는 마이그레이션 방식과 마이그레이션이 기존 환경에 미치는 영향도 고려해야 한다.

9. **공급업체 선택.** 공급업체마다 데이터 중복 제거 방식과 관련하여 각양각색의 주장을 펼치고 있다. IDC 조사에 따르면 시중에 발표된 모든 데이터 중복 제거 제품의 실제 성능이 광고만큼 충실한 것은 아니다. 따라서 데이터 중복 제거 기능을 갖춘 특정 제품의 판매 기간, 해당 제품을 운영 환경에 사용하고 있는 고객의 수, 실제 환경에서 제품이 제공하는 기술의 성숙도를 고려해야 한다. 또한 제품의 확장성에 대한 철저한 조사를 위해 애플리케이션 및 시스템 지원 매트릭스를 요청할 필요가 있다. POC(Proof-of-Concept)를 수행하지 않을 경우 성능과 확장성 면에서 예기치 못한 위험을 겪을 수도 있다.

10. **데이터 중복 제거 활용 사례.** 데이터 중복 제거는 스토리지 인프라스트럭처의 가치를 한 단계 높이는 기술이다. 현재, 대량의 중복 데이터가 존재하는 백업 분야에서는 데이터 중복 제거 기술이 이미 광범위하게 구축되어 있다. 똑같은 데이터를 매주 백업하게 되면 서버, 네트워크 및 스토리지 리소스가 불필요하게 소모된다. 일부 기업들은 **NAS(Network Attached Storage)** 방식을 통해 운영 스토리지 환경에서 기존 데이터 중복 제거 기술을 검토하거나 테스트하기 시작했는데, 이러한 환경을 구축하려면 지연 및 응답 시간 문제를 피할 수 있는 향상된 성능이 필요하다. 오늘날 데이터 중복 제거 기술은 가상 시스템, 원격 사무소/지사 및 데이터 센터 환경 백업에 적합한 기술로 자리잡았다.

데이터 중복 제거 기술을 지원하는 EMC 솔루션 포트폴리오

EMC는 IT 비용을 절감하고 백업 효율성을 향상시키는 데 도움이 되는 다양한 데이터 중복 제거 지원 제품을 제공한다. 데이터 중복 제거 백업 솔루션으로는 소스 측 중복 제거 방식의 **EMC Avamar**, 타겟 측 중복 제거 방식의 **EMC Disk Library**, 그리고 소스 측/타겟 측 또는 양쪽에 구축할 수 있는 **EMC NetWorker**가 있다. 또한 본 백서에서는 다루지 않지만 EMC는 **EMC Celerra NAS** 시스템을 통해 운영 스토리지 및 백업 데이터를 위한 데이터 중복 제거 솔루션을, **Centera** 제품군을 통해 디스크 아카이브 데이터 중복 제거 솔루션을 제공하고 있다.

EMC Avamar

EMC Avamar 백업 및 복구 솔루션은 통합된 데이터 중복 제거 기술을 통해 소스에서 중복된 데이터를 파악함으로써 LAN/WAN을 통해 전송되는 백업 데이터의 양을 최소화한다. Avamar 솔루션을 활용하면 VMware 환경, 원격 사무소 및 데이터 센터 LAN/NAS 서버에서 백업되는 데이터의 양을 줄이고 매일 신속하게 전체 백업을 수행할 수 있다. 또한 Avamar는 시간 경과에 따라 사이트 및 서버 전반에서 전사적으로 백업 데이터의 중복을 제거한다. 기존 복구 방법을 활용하는 제품과 달리 Avamar는 한 번에 신속하게 데이터를 복구할 수 있기 때문에 전체 백업본 및 이후 증가분 백업본에서 데이터를 복구해야 하는 번거로움 없이 원하는 시점으로 복구할 수 있다. 이와 같은 Avamar의 기능은 기존 백업 애플리케이션에서 근본적으로 탈피한 것이라 할 수 있다.

Avamar 에이전트는 새로 추가되거나 데이터가 변경된 파일을 추적한다. Avamar 에이전트는 새로 추가되거나 변경된 데이터를 찾기 위해 전체 파일 시스템 트리를 검색하기 전에 해당 파일의 로컬 캐시를 먼저 확인한다. 새로 추가되거나 변경된 파일이 확인되면 해당 파일을 하위 파일 레벨의 가변 길이 데이터 세그먼트로 나눈 다음 각 세그먼트에 해시 값(고유 ID)을 지정한다. 그 다음으로 이 에이전트는 Avamar 서버와 통신을 수행하여 해시가 고유한지, 아니면 이미 존재하는지 확인한다. 새로운 데이터 세그먼트로 확인되면 일일 전체 백업을 수행하는 동안 LAN/WAN을 통해 해당 데이터 세그먼트를 전송한다.

이와 같은 백업 프로세스는 기존 백업 에이전트보다 호스트의 CPU 활용도를 증가시킨다. 그렇지만 완전히 새로운 데이터 세그먼트만 백업하는 효율적인 방식을 지원하므로, Avamar를 사용할 경우 전체 백업을 수행한 다음 다시 증가분 백업을 수행하는 기존 방식보다 훨씬 빠르게 백업이 완료된다. 예를 들어, 일반적으로 10시간이 소요되던 증가분 백업에 Avamar를 활용하는 경우 1시간 정도면 완료할 수 있기 때문에 월요일부터 금요일까지 증가분을 백업할 경우 주당 백업 소요 시간이 50시간에서 5시간으로 크게 줄어든다. 또한 Avamar는 기존 전체 백업보다 훨씬 향상된 속도로 매일 신속하게 전체 백업을 수행할 수 있다.

Avamar 백업 및 복구 솔루션은 소스 측에서의 전사적 데이터 중복 제거를 제공하므로 다음과 같은 환경의 기업에게 적합하다.

- ☑ **가상 시스템**을 구축하고 물리적 서버, 가상 서버, 개별 객체들을 복구할 새로운 보호 전략을 평가하려는 기업
- ☑ **원격 지사** 백업을 향상시켜 신속하게 매일 전체 백업 수행, 중앙 집중식 관리, 신뢰성 향상, 안전한 복제, 혼잡한 WAN 링크의 백업 트래픽 감소 등의 이점을 얻으려는 기업
- ☑ 로컬 **NAS 및 파일 서버** 환경의 백업을 위해 데이터, 백업 윈도우 및 네트워크 트래픽의 증가를 억제하려는 기업

EMC Avamar는 네 가지 구성으로 구축할 수 있다.

- ☑ **Avamar 소프트웨어.** 규모가 작은 원격 사무소의 경우 로컬 하드웨어를 추가로 구축할 필요 없이 보호할 시스템(클라이언트)에 Avamar 소프트웨어 에이전트를 구축하면 된다.
- ☑ **타사 서버에 Avamar 구축.** Avamar 소프트웨어를 구입해서 내장형 디스크 스토리지가 있는 다양한 업계 표준 서버에 구축할 수 있다.
- ☑ **Avamar Data Store.** 확장이 가능한 일체형 솔루션이며 Avamar 소프트웨어가 EMC 하드웨어에 미리 설치 및 구성되어 있어 주문, 구축 및 서비스가 간편하다.
- ☑ **Avamar Virtual Edition for VMware.** 업계 최초의 이 구성을 사용하면 기존 ESX Server에 Avamar 서버를 가상 어플라이언스로 구축할 수 있으므로 연결된 리소스 및 디스크 스토리지를 효과적으로 활용할 수 있다.

Avamar는 시중의 다른 소스 측 중복 제거 제품과는 차이가 있다. 일례로 Avamar 데이터 중복 제거는 하위 파일 레벨의 가변 길이 데이터 세그먼트를 사용하므로 고정 길이의 세그먼트를 사용하는 솔루션에 비해 효율성과 성능이 뛰어나다. Avamar는 그리드 아키텍처를 사용하여 성능과 용량을 확장하는데, 이 경우 노드가 증가할 때마다 전체 시스템의 CPU, 메모리, 입출력 및 스토리지가 증가한다.

Avamar 그리드는 RAIN(Redundant Arrays of Independent Nodes) 구성을 사용하여 그리드 전반에서 내결함성 및 고가용성을 기본적으로 제공하고, 이중화를 구현한다. Avamar는 Avamar 노드 전체로 내부 인덱스를 배포하여 신뢰성, 로드 밸런싱 및 확장성을 구현한다. 또한 Avamar는 백업 데이터가 완벽하게 복구 가능한 상태인지 매일 자동으로 확인하고, Avamar 서버는 매일 두 차례 자체 검사를 실행하여 서버 무결성을 점검한다. 마지막으로 Avamar는 Exchange, SQL, Oracle, DB2, SharePoint, Lotus Notes, NDMP를 포함하여 다양한 애플리케이션과 클라이언트를 지원한다.

Avamar는 다양한 방법으로 물리적 시스템과 가상 시스템을 보호한다. VMware 가상 시스템 환경에서 지원되는 Avamar 백업 옵션은 다음과 같다.

- ☒ **게스트 OS의 Avamar 에이전트.** 각 게스트 OS에 포함된 Avamar 에이전트는 기존 에이전트 백업 방식에 비해 보다 효율적인 백업 방식을 제공한다. 최소 구성의 Avamar 에이전트는 게스트에서 백업 데이터의 양을 줄여 네트워크 요구 사항과 공유 CPU, NIC, 디스크 및 메모리 리소스에 대한 경합을 최소화한다. 새로 추가되었거나 고유한 하위 파일 데이터만 백업되므로 Avamar는 매일 신속하게 전체 백업을 완료할 수 있다.
- ☒ **VCB 백업을 위한 Avamar.** VCB 프록시 서버에서 실행되는 Avamar 에이전트는 고유한 데이터만 백업함으로써 게스트 시스템의 처리 부담을 덜어 준다. 데이터 중복 제거는 VMDK 파일 내부 및 VMDK 파일 간에 이루어지며 VCB 파일/이미지 레벨 백업을 지원한다. Avamar는 효율적인 복제 기능을 기반으로 WAN을 통해 VMDK 파일을 신속하게 전송하여 재해 복구 목표가 달성되도록 돕는다.
- ☒ **ESX 콘솔의 Avamar 에이전트.** ESX 콘솔의 Avamar 에이전트는 VMDK 파일 내부 및 VMDK 파일 간에 데이터 중복 제거를 수행할 수 있다. 이 방법은 VMware VCB 또는 공유 스토리지에 의지하지 않고 이미지 레벨 백업 및 복구 옵션을 제공하지만, 파일 레벨 복구는 제공하지 않는다.

EMC Disk Library

EMC DL(Disk Library) 제품군은 DL1500, DL3000, DL4000 시리즈 시스템에서 정책 기반 데이터 중복 제거를 제공한다. EMC DL1500/3000은 데이터 중복 제거 기능이 포함된 LAN 기반 B2D(Back-to-Disk) 솔루션을 제공한다. DL1500은 성능 향상, 온사이트 보존 기간 연장, 복제 비용 절감을 원하는 중소, 성장, 중견 기업을 위해 개발된 제품이다. DL1500은 기본 가용 용량이 4TB로 최대 36TB까지 확장할 수 있으며 Immediate 데이터 중복 제거 옵션 사용 시에는 0.72TB/hr, Scheduled 데이터 중복 제거 옵션 사용 시에는 최대 0.84TB/hr의 백업 수집 속도를 유지한다.

DL3000은 기본 가용 용량이 8TB로 최대 148TB까지 확장할 수 있으며 Immediate 데이터 중복 제거 옵션 사용 시 1.44TB/hr의 백업 수집 속도를 유지한다. DL1500과 DL3000에는 정책 기반 데이터 중복 제거 기능이 포함되어 있다. DL4000은 DL1500 및 DL3000 모델과 달리 신규 또는 기존에 설치된 DL4000 VTL(Virtual Tape Library) 시스템의 하드웨어 추가 옵션을 통해 데이터 중복 제거 기능을 제공한다. DL4000 시스템을 구축하면 B2D 용량 요구 사항과 데이터 센터 간 복제에 필요한 네트워크 트래픽을 줄일 수 있다.

EMC DL 시리즈의 데이터 중복 제거는 데이터 센터, 대용량 스토리지 및 변경이 잦은 데이터베이스 환경에서 B2D를 사용하고자 할 때 적합하다. DL 시리즈의 데이터 중복 제거 기능을 사용하면 다음과 같은 이점을 얻을 수 있다.

- ☒ 기존 EMC VTL 환경에서 백업이 필요한 **대용량 데이터**의 증가 억제
- ☒ 새로운 **B2D 전략** 구축으로 스토리지 비용을 최소화하면서 복구 성능/신뢰성 향상
- ☒ 기존 **EMC DL** 환경에서 디스크 기반 백업 및 데이터 중복 제거 모두 지원

☒ 재해 복구 데이터 센터로 **전자적 방식의 백업 볼팅(Vaulting)**을 활용하여 물리적 테이프 사용 최소화

☒ 현재의 백업 작업에 거의 영향을 미치지 않고 **백업용 테이프를 디스크로 교체**

DL 시리즈 시스템은 타겟 측 중복 제거 방법을 사용한다. 즉, 전체 DL 제품군 전반에서 데이터 중복 제거 기능을 동일하게 수행하여 타겟에서 블록 레벨의 가변 길이 해시 기반 중복 제거 기능을 제공한다. DL 시리즈 데이터 중복 제거는 “애플리케이션 감지 필터”를 사용하여 데이터 스트림 형식을 감지하고 애플리케이션별 메타데이터가 스트림에 삽입되는 위치를 파악한다. 필터가 메타데이터를 마커로 표시하여 구분하기 때문에 데이터 중복 제거 효과가 한결 높다.

DL 시스템은 고객이 직접 구성할 수 있는 정책 기반 데이터 중복 제거 프로세스를 사용한다. 데이터 중복 제거를 즉시 수행하거나 스케줄에 따라 수행하도록 설정할 수도 있고, 야에 비활성화할 수도 있다. 정책은 파일 공유 또는 가상 라이브러리 레벨에서 설정할 수 있다. “Immediate” 또는 인라인 중복 제거는 크기가 작은 데이터 세트와 비정형 데이터에 적합하다. “Scheduled” 모드로 설정할 경우 백업이 수행되는 동안 데이터 중복 제거 프로세스를 지연시키고 백업이 완료되면 데이터 중복 제거를 시작한다. 용량이 큰 데이터 세트에 적합한 이 방법은, EMC 자체 조사 결과에 따르면 Immediate 모드보다 150%~200% 신속하게 백업을 완료한다.

데이터 중복 제거 효과를 크게 기대할 수 없는 데이터 유형의 경우 기능을 비활성화할 수 있다. 재해 복구를 목적으로 데이터 중복 제거 지원 원격 복제를 활용하는 경우, 시스템/애플리케이션/디렉토리 레벨 또는 가상 테이프 카트리지로 복제 작업을 구성할 수 있다. DL 시리즈의 데이터 중복 제거 인덱스는 분기로 클러스터링되고 유사한 객체는 버킷으로 그룹화되므로, 인덱스 조회 효율은 높아지고 디스크 입출력은 최소화된다. 하드웨어 압축은 스토리지 최적화 수준을 한층 더 높여 준다.

EMC NetWorker

EMC NetWorker는 중앙 집중식으로 백업 및 복구 작업을 관리하는 엔터프라이즈 백업 애플리케이션이다. NetWorker는 물리적 환경과 가상 환경에서 B2D, 복제, CDP(무중단 데이터 보호), 데이터 중복 제거 등의 다양한 데이터 보호 옵션을 지원하는 공통 플랫폼을 제공한다. 광범위한 지원 기능을 제공하는 NetWorker는 대규모 데이터 센터에서 원격 사무소에 이르는 모든 환경에서 관리를 간소화하려는 고객에게 이상적인 백업 소프트웨어이다. NetWorker의 핵심 애플리케이션은 EMC Avamar 데이터 중복 제거 기술과 통합되어 소스 측 중복 제거를 제공하는 동시에, 지원 범위 내에서 EMC DL 시리즈 등의 타겟 측 중복 제거 솔루션과도 연동할 수 있다.

NetWorker 데이터 중복 제거를 사용하면 다음과 같은 이점을 얻을 수 있다.

☒ 기존 NetWorker 환경에서 대용량 데이터의 증가 억제

☒ 새로운 B2D 전략 구축으로 아카이빙 또는 장기간 보존을 위해 물리적 테이프를 사용하는 환경의 복구 성능 향상

- ☒ 기존 EMC DL 환경에서 디스크 기반 백업 및 데이터 중복 제거 모두 지원
- ☒ 혼합 요구 사항 충족 — 소스 측 중복 제거와 타겟 측 중복 제거 모두 지원 가능
- ☒ 다양한 데이터 보호 전략을 하나의 애플리케이션에 통합하여 비용 절감 및 복잡성 해소

NetWorker 애플리케이션의 데이터 중복 제거 방식은 기존 백업 애플리케이션과 데이터 중복 제거를 통합했다는 점에서 높은 경쟁력을 갖추었다고 할 수 있다. NetWorker 클라이언트 소프트웨어는 단일 에이전트에서 데이터 중복 제거 백업과 기존 백업을 모두 지원한다. 또한 완벽하게 통합된 소스 중복 제거 기능은 구축 및 유지 관리 부담을 최소화한다. NetWorker 콘솔로 기존 백업과 데이터 중복 제거 백업을 모두 관리하고 모니터링할 수 있으므로, NetWorker 고객은 클라이언트 측 비용을 추가로 부담하지 않고 데이터 중복 제거의 이점을 활용할 수 있다.

다른 솔루션과 달리 NetWorker는 소프트웨어 SKU나 비용 추가 없이도 데이터 중복 제거 통합을 지원한다. NetWorker 고객은 Avamar나 EMC DL 시리즈 백엔드 솔루션과 같은 적절한 데이터 중복 제거 엔진을 백업 환경에 추가할 수 있다. NetWorker의 데이터 중복 제거는 물리적 테이프를 지원하므로, 테이프를 계속 사용해야 하는 사용자의 요구를 단일 애플리케이션으로 충족할 수 있다는 이점이 있다. 백업 애플리케이션의 데이터 중복 제거 기술은 암호화 및 압축의 순서 조정과 용량 할당을 정확하게 수행할 수 있다는 이점도 제공한다. NetWorker를 통해 기업은 현재의 백업 환경에 영향을 미치지 않고 강력한 데이터 중복 제거 기술을 활용할 수 있다.

당면 과제: 적합한 방식 선택

지금까지 살펴본 바와 같이 각 데이터 중복 제거 기술 및 접근 방식이 제공하는 장점은 활용 사례별로 많은 차이가 있다. 따라서 각 환경에서 최고의 효율성을 제공하는 EMC 제품이 무엇인지 손쉽게 파악할 수 있는 방법이 필요하다. 표 2에는 적합한 EMC 데이터 중복 제거 지원 제품을 선택할 때 고려해야 할 사항이 나와 있다.

EMC는 데이터 중복 제거 기능을 갖춘 다양한 제품을 보유하고 있다. 데이터 중복 제거는 독립 실행형 제품이 아니라 하나의 기능 또는 기술이므로 EMC는 고객사 운영 환경의 당면 과제를 감안하여 데이터 중복 제거 기능을 가장 적절하게 활용할 수 있는 부분에 대한 고객 교육을 더 강화해야 한다. 문서화된 사례 연구, 기준 및 성능 테스트 벤치마크 자료를 이러한 교육과 함께 제공한다면 고객은 더욱 큰 신뢰를 바탕으로 EMC 제품의 데이터 중복 제거 기술을 활용하게 될 것이다.

표 2

EMC 데이터 중복 제거 지원 제품 선택 방법

	EMC NetWorker	EMC Disk Library	EMC Avamar
백업을 위한 데이터 중복 제거 솔루션	<ul style="list-style-type: none"> 소스 측 인라인 중복 제거 	<ul style="list-style-type: none"> 타겟 측 정책을 기반으로 데이터 중복 제거를 Immediate/Scheduled 로 설정하거나 비활성화하도록 구성 가능 	<ul style="list-style-type: none"> 소스 측 인라인 중복 제거
적합한 환경	<ul style="list-style-type: none"> NetWorker 환경 물리적 테이프를 지원해야 하는 경우 대규모 이기종 환경 	<ul style="list-style-type: none"> 고속 백업 및 복구가 필요한 경우 복제를 통해 오프사이트 백업을 수행하는 환경 운영에 영향을 미치지 않고 현재의 백업 환경을 지원해야 하는 경우 데이터 센터 및 원격 사이트 지원이 필요한 경우 	<ul style="list-style-type: none"> 가상화 환경 원격 지사 LAN/NAS 서버
구축 옵션	<ul style="list-style-type: none"> 단일 NetWorker 에이전트 규모가 작은 원격 사무소에 적합한 에이전트 데이터 중복 제거 노드로 사용할 수 있는 시스템의 경우: <ul style="list-style-type: none"> Avamar Data Store — 포괄적인 통합 솔루션 (하드웨어 및 소프트웨어 포함) 타사 서버 — 자체 Avamar 서버 구축 Avamar Virtual Edition — 기존 ESX Server 및 디스크를 활용하는 가상 어플라이언스 	<ul style="list-style-type: none"> 어플라이언스 하드웨어 	<ul style="list-style-type: none"> 에이전트만 지원 - 규모가 작은 원격 사무소에 적합 Avamar Data Store — 포괄적인 통합 솔루션(하드웨어 및 소프트웨어 포함) 타사 서버 — 자체 Avamar 서버 구축 Avamar Virtual Edition — 기존 ESX Server 및 디스크를 활용하는 가상 어플라이언스

출처: IDC, 2009년

결론

데이터 중복 제거는 백업 효율성을 향상시키고 IT 비용을 절감할 수 있게 해 주는 기술이다. 급증하는 백업 데이터와 관련된 수많은 비용 및 운영상의 당면 과제를 해결하고자 많은 기업이 다양한 데이터 중복 제거 지원 솔루션을 구축하고 있다. 이러한 당면 과제를 해결하기 위해 다양한 스토리지 솔루션이 반드시 갖추어야 할 핵심 기능은 데이터 중복 제거라는 것이 IDC의 견해이다. EMC는 다양한 환경과 시나리오를 대상으로 광범위한 솔루션을 제공함으로써 백업과 관련된 오랜 문제를 해결하고 향후 5년간 고객의 기술적 요구를 충족할 수 있는 공급업체이다.

저작권 고지 사항

IDC Information and Data의 대외 발행 — 광고, 보도 자료 또는 홍보 자료에 사용되는 모든 IDC 정보는 IDC의 담당 부사장 또는 지사장의 사전 서면 승인을 받은 후 사용해야 합니다. 제안서 초안은 반드시 관련 요청서와 함께 제출해야 합니다. IDC는 이유를 막론하고 외부 사용에 대한 승인을 거부할 권리를 보유합니다.

Copyright 2009 IDC. 서면 허가를 받지 않은 모든 복제 행위는 엄격히 금지됩니다.