

솔루션 브리프

AI를 위한 에버그린//원

SLA 기반 서비스형 스토리지로 AI의 예측 불가능성 제거

거의 모든 조직에서 인공지능(AI)이 현실의 문제를 해결하고 조직에 경쟁 우위를 제공하는 혁신적인 힘을 가지고 있다는 사실을 인식하고 있습니다. 하지만 AI 애플리케이션과 사용 사례가 기하급수적으로 증가하고 GPU 가속 컴퓨팅이 확산되면서, 이러한 데이터 집약적인 워크로드를 지원할 수 있는 스토리지 환경의 규모를 조정하기가 어려워졌습니다. 스토리지를 소규모로 구축하면 고가의 컴퓨팅, 데이터 과학, AI 개발 투자에 성능 병목 현상이 발생하고, 오버프로비저닝은 비용 비효율을 초래합니다. 뿐만 아니라 많은 AI 프로젝트가 소규모로 시작해 시간이 지남에 따라 규모가 커지기 때문에 부담이 가중됩니다. 따라서 스토리지 규모를 예측하기 훨씬 더 어려워지며, 3~5년 교체 주기가 시작될 때 추측하는 것만으로는 충분하지 않습니다. 클라우드를 활용하는 것도 하나의 옵션이지만, AI 프로젝트가 커짐에 따라 비용은 급증하게 됩니다. 또한, 차별화된 AI 모델을 학습시키는 데 필요한 중요한 데이터의 대부분은 데이터 프라이버시, 보안, 컴플라이언스 및 성능 요구 사항으로 인해 직접 관리해야 하는 온프레미스 스토리지 환경에 있는 경우가 많습니다.



인증

Nvidia OVX 서버, DGX BasePOD 및 DGX SuperPOD* 인증으로 AI 성능이 입증되었습니다.



에너지 효율

GPU와 가속기에 에너지 소비를 집중하세요. 퓨어스토리지 제품은 에너지 사용량과 탄소 배출량을 최대 85%까지 줄여줍니다.



독보적인 성능

성능에 한계가 없습니다. AI용 에버그린//원은 모든 AI 워크로드에 대해 최대 수 TB/s 이상의 성능을 보장합니다.

AI용 에버그린//원 소개

AI 인프라 구축을 위해 특별히 설계된 퓨어스토리지 AI용 에버그린//원(Evergreen//One)은 더 이상 규모를 예측할 필요가 없게 합니다. AI용 에버그린//원은 예측 불가능한 AI의 특성을 완벽하게 지원합니다. 실험적이고 반복적인 개발 프로세스로 인해 AI의 데이터, 성능, 확장성 요구 사항은 매우 다양하고 빠르게 변화하며 진화할 수 있습니다. GPU 및 가속기의 최대 대역폭 요구 사항을 기반으로 성능을 보장하므로 스토리지가 병목현상을 일으킬 걱정이 없습니다. AI용 에버그린//원은 Nvidia OVX 서버, DGX BasePOD, 심지어 DGX SuperPOD¹ 클러스터의 가장 극단적인 성능 요구 사항도 충족할 수 있습니다. 이는 서비스수준협약(SLA) 기반 서비스형 스토리지(StaaS)를 통해 퓨어스토리지가 지킬 약속입니다.

AI의 불확실성 속에서 가장 중요한 예측 가능성

다양한 AI 워크로드에는 각기 다른 스토리지 요구 사항이 있습니다. 데이터 가변성, 모델 복잡성, 데이터 품질, 훈련 역학, 진화하는 성능 요구 사항, 추가적인 외부 요인으로 인해 이러한 워크로드를 지원하는 데 적합한 스토리지 풋프린트를 조정하기가 더 어려워집니다.

AI용 에버그린//원은 워크로드 강도나 GPU 수에 관계없이 일관되고 예측 가능한 스토리지 성능을 보장하는 SLA를 제공하여 중요한 AI 투자가 제대로 활용되도록 보장합니다. 스토리지 비용은 워크플로우에 관계없이 필요한 최대 대역폭을 기반으로 하는 간단한 요금제로 미리 확인할 수 있습니다. 추가 용량은 클드 스토리지와 유사한 경쟁력 있는 비용으로 청구됩니다. 소중한 아티팩트, 체크포인트, 로그를 삭제할 필요가 없습니다. 예측 가능한 월별 지출과 SLA의 확실성을 통해 현재 필요한 스토리지 용량을 확보하고 절감된 비용을 향후 보다 전략적인 AI 지출에 투자할 수 있습니다.

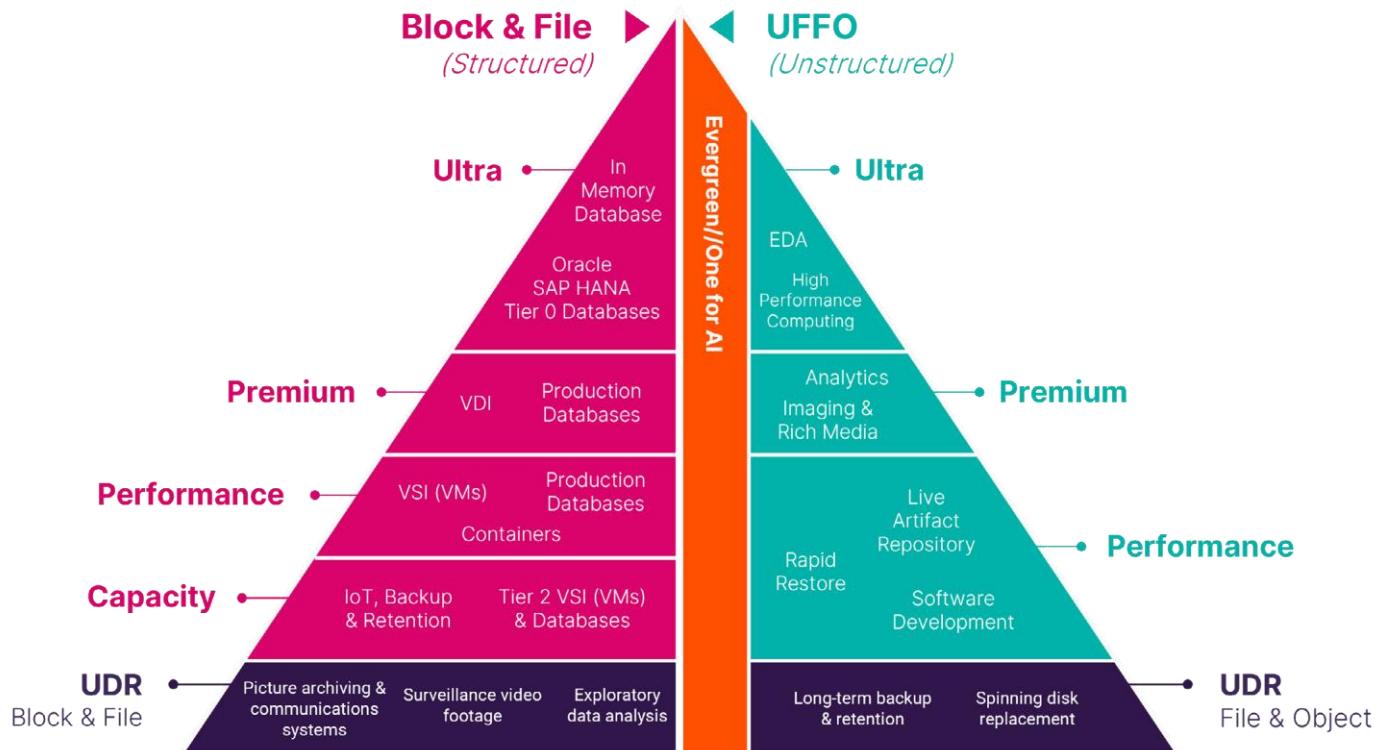


그림 1 에버그린//원 서비스 카탈로그

운영 중단 없이 지속적인 혁신 제공

AI용 에버그린//원은 운영 중단 없이 지속적인 혁신을 제공함으로써 퓨어스토리지의 에버그린 아키텍처 원칙에 충실히 합니다. AI 워크로드의 변화와 성장에 따라, AI용 에버그린//원은 스토리지 인프라를 간편하게 업그레이드할 수 있도록 지원합니다. 에버그린 아키텍처는 비용이 많이 드는 전면적인 업그레이드나 데이터 마이그레이션 없이도 최신 기술 도입, 용량 확장, 성능 향상, 기능 업데이트를 제공하여 AI 워크플로우가 중단되지 않도록 보장합니다.

AI용 에버그린//원 구독에는 AI 기반 데이터 서비스 플랫폼인 퓨어1(Pure1)이 포함되어 있습니다. 인공지능 엔진인 퓨어1 메타(Pure1 Meta)를 사용한 지속적인 모니터링과 선제적인 문제 해결을 통해 애플리케이션과 인프라 또는 서비스 요구 사항을 정확하게 예측하고 미래를 내다볼 수 있습니다. 퓨어1은 예측 인텔리전스, 폴스택 분석, 글로벌 엔드-투-엔드 가시성, 지속 가능성 및 데이터 보호를 위한 평가를 통해 에버그린//원 구독을 모니터링 및 확장할 수 있는 단일 공간입니다.

스토리지 풋프린트 관리가 아닌 AI 혁신에 집중하기

AI용 에버그린//원은 스토리지 인프라의 수명주기 전반에 걸쳐 포괄적인 관리 및 지원을 제공함으로써 관리 부담을 덜어주며, 조직이 리소스와 전문성을 AI 이니셔티브의 발전에 집중할 수 있도록 지원합니다. 퓨어스토리지는 초기 구축 및 구성부터 지속적인 모니터링, 관리 및 최적화에 이르기까지 모든 측면을 관리합니다. 훈련, 투닝, 추론, 딥 러닝, 검색 증강 생성(RAG) 워크플로우 도입 등, 서비스 구독을 통해 AI용 에버그린//원은 투자를 극대화하는 보장된 성능 수준으로 모든 예측 또는 생성형 AI 워크로드를 지원할 수 있습니다.

참고 자료

- [에버그린//원](#)에 대해 자세히 알아보세요.
- 에버그린//원 [ESG 경제성 검증](#) 분석 보고서를 확인해 보세요.
- 에버그린//원 [비용 계산기](#)를 사용해 보세요.
- 퓨어스토리지의 [AI 솔루션](#)에 대해 자세히 확인해 보세요.
- [에이리\(AIRI\) AI 인프라](#)에 대해 자세히 알아보세요.

1 Nvidia DGX SuperPOD 인증은 2024년 3분기에 예상됩니다. 퓨어스토리지는 이러한 인증을 취득하기 위해 최선을 다하고 있지만, 인증에 대한 모든 미래 지향적인 언급은 현재의 예상을 기반으로 하며, 약속이나 보장이 아님을 양해해 주시기 바랍니다.