

Generative AI를 위한 Dell Technologies의 개방형 이더넷 솔루션

IT 인프라스트럭처의 새로운 영역 개척

'AI 패브릭(GPU 간 연결을 위한 백엔드 스위칭)은 12억 달러(2022년)에서 152억 달러(2027년)로 5년 CAGR이 65% 증가할 것으로 예상됩니다.

AI 패브릭에 대해 이더넷은 32%의 매출 점유율과 37%의 포트 배송률을 달성할 것으로 예상됩니다(2027년).'

Dell'Oro Research¹

GPU 성능은 네트워크 성능에 크게 의존합니다. 컴퓨팅 노드와 스토리지 간의 지속적인 통신이 필요한 대규모 서버 클러스터에서 실행되는 AI 워크로드의 경우, 병목 현상을 방지하기 위해 강력한 네트워킹이 필요합니다. 워크로드에 대한 네트워킹 성능이 부족하면 GPU가 유휴 상태가 되고 교육 및 추론 시간이 늘어나서 데이터 처리 속도가 느려지고 통찰력을 확보하기까지 오래 걸립니다.

소개: Generative AI의 네트워크 수요

GenAI(Generative AI) 솔루션이 계속 발전하고 데이터 처리 및 컴퓨팅 요구 사항이 끝없이 늘어나면서 IT 인프라스트럭처는 이러한 환경의 엄청난 요구 사항을 지원할 수 있는 방법을 찾는 데 어려움을 겪고 있습니다. 이러한 모델, 특히 LLM(Large Language Model)에는 더 많은 인프라스트럭처뿐만 아니라 GPU 클러스터 전반에 걸친 대규모 연결 요구 사항을 관리하기 위해 신중하게 설계된 시스템도 필요합니다. 기존 네트워크 솔루션은 빠르게 병목 현상을 일으키고 있으며 GenAI 이니셔티브의 실행과 성공을 위협하고 있습니다. AI 패브릭에는 낮은 레이턴시, 무손실 성능, 최대 대역폭이 필요합니다. 대규모 데이터 및 애플리케이션 처리 요구 사항으로 인해 프런트엔드 및 백엔드 패브릭에 대한 요구 사항이 늘어났습니다.

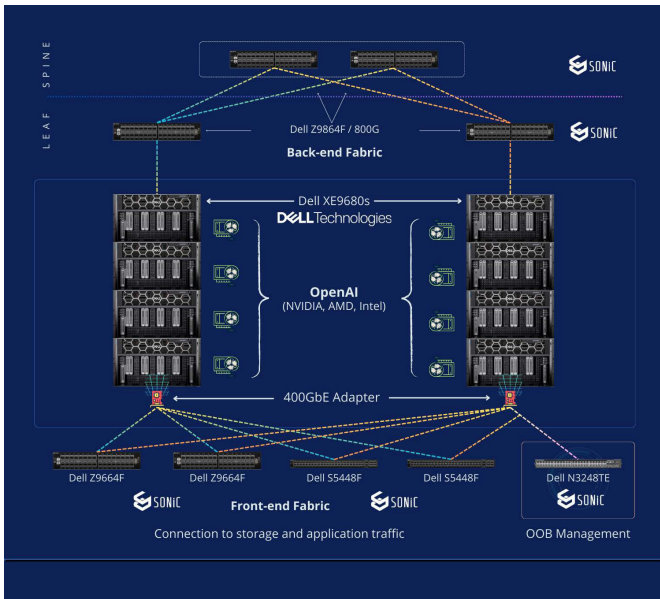
GenAI 인프라스트럭처 구축의 당면 과제와 요구 사항

GenAI 기술을 구축하는 과정에는 새로운 아키텍처와 관련된 기술적 복잡성부터 이러한 구축을 관리할 수 있는 숙련된 전문가의 부족에 이르기까지 다양한 당면 과제가 따릅니다. InfiniBand와 같은 독점 기술에 의존하는 솔루션은 복잡성을 가중시키고, 리소스 가용성을 제한하며, 기존 모니터링 또는 오케스트레이션 플랫폼과의 통합을 복잡하게 만듭니다. 또한 독점 솔루션으로 인해 발생하는 높은 비용, 긴 평가 시간, 공급업체 종속 현상은 특히 공급망 불확실성 시대에 상당한 장애 요소가 됩니다. 이러한 당면 과제는 GenAI 워크로드의 고유한 요구 사항을 해결할 수 있는 개방적이고 유연하며 강력한 GenAI 인프라스트럭처 솔루션이 절실히 필요하다는 사실을 잘 보여줍니다.

GenAI 네트워킹에 대한 Dell Technologies의 접근 방식

이러한 당면 과제에 대응하기 위해 Dell Technologies는 GenAI 인프라스트럭처의 복잡한 요구 사항을 충족하도록 설계된 포괄적인 개방형 이더넷 기반 솔루션을 개척해 왔습니다. Dell Technologies는 AI, 모델링 및 HPC(High Performance Computing) 환경에서 쌓은 방대한 경험을 활용하여 프런트엔드 및 백엔드 요구 사항을 모두 해결하는 솔루션 제품군을 제공합니다. Dell PowerEdge XE 서버와 같은 가속화에 최적화된 모듈형 컴퓨팅 시스템부터 PowerScale과 같은 AI 중심의 스토리지 솔루션에 이르기까지, Dell Technologies는 성공적인 GenAI 구축을 위한 필수 구성 요소를 제공합니다. 이러한 접근 방식의 핵심은 첨단 네트워크 실리콘으로 구동되는 차세대 이더넷 패브릭을 구축하는 것입니다. GenAI 애플리케이션에 필수적인 800GbE의 비차단 네트워크 성능을 제공하는 Dell PowerSwitch Z9864-ON을 통해 고객은 고급 라우팅, ROCEv2, 향상된 해싱 및 우선순위 흐름 제어 등 Dell의 Enterprise SONiC Distribution by Dell Technologies에서 제공하는 높은 대역폭 스위칭과 새로운 기능을 사용하여 레이턴시가 낮고 처리량이 높은 AI 클러스터를 구축함으로써 패브릭 성능을 높이고 혼잡도 모니터링을 개선할 수 있습니다.

¹ Dell'Oro Brief: Advanced Research Report on AI Networks for AI Workloads



GenAI 패브릭 아키텍처 예시

최첨단 실리콘을 활용하는 Dell PowerSwitch Z Series는 수천 개의 노드를 지원할 수 있는 고성능 확장형 네트워크를 위한 백본을 제공하여 GenAI 애플리케이션에 내재된 연결 문제를 해결합니다.

Dell Technologies를 통해 GenAI 구축 가속화

GenAI의 급증으로 인해 IT 인프라스트럭처에는 많은 당면 과제가 발생했으며, 혁신적이고 유연한 네트워크를 위한 새로운 접근 방식이 요구됩니다. Dell Technologies는 GenAI 구축의 즉각적인 요구 사항을 충족할 뿐 아니라 향후 발전을 위한 기반을 마련하는 개방형 이더넷 기반 솔루션으로 이러한 문제에 대해 해답을 제시합니다.

AI 하드웨어 솔루션에 대한 불확실성 해소를 지원하기 위해 Dell Technologies는 랩 검증을 거쳐 AI 워크로드에 최적화된 것으로 입증된 레퍼런스 아키텍처를 제공합니다. 이러한 Validated Designs에는 아키텍처 개념, 전체 솔루션 개요, 성능 그리고 설계 시 고려된 워크로드에 대한 솔루션의 기능을 입증하는 기타 랩 검증이 포함됩니다. 더 빠르고 심층적인 통찰력을 더 쉽게 제공할 수 있는 검증된 솔루션을 통해 AI 지원 가능 단계에서 AI에 대해 검증된 단계로 전환해 보세요.

Dell Technologies를 선택하는 조직은 전문 지식과 포괄적인 솔루션 스택을 갖췄으며 GenAI 이니셔티브의 성공을 보장하기 위해 헌신적으로 노력하는 파트너를 확보하는 것입니다. Dell Technologies와 함께 협력함으로써 기업은 GenAI 아키텍처의 복잡성을 해결할 장비를 갖춰 프로젝트를 실행할 뿐만 아니라 성공을 위한 입지도 확보할 수 있습니다.

Dell Technologies와 함께 GenAI 환경 구축 및 가치 실현 시간을 단축하여 위험과 운영 복잡성을 줄일 수 있습니다. 개방적이고 유연하며 지속 가능한 네트워크 솔루션을 통해 GenAI 이니셔티브를 검증된 방식으로 전환하고 혁신과 효율성의 새로운 시대로 나아가는 방법을 알아보세요.

GenAI 네트워킹에서 찾아볼 수 있는 Dell Technologies의 혁신 기술

Dell Technologies는 GenAI 네트워킹 혁신의 최전선에 있으며, 엣지에서 코어 그리고 클라우드에 이르기까지 현재와 미래의 GenAI 환경 요구 사항을 충족하는 솔루션을 제공합니다. Dell Technologies는 상용 실리콘과 오픈 소스 기반 소프트웨어를 활용하는 확장 가능한 개방형 솔루션에 초점을 맞춰 최고 수준의 성능과 유연성을 보장합니다.

UEC(Ultra Ethernet Consortium)에 적극적으로 참여하고 기여할 뿐만 아니라 SONiC과 같은 상용 오픈 소스 네트워크 운영 체제를 사용함으로써 Dell Technologies는 이더넷 분야에서 개방형 표준과 협력적인 개발을 위해 노력하고 있음을 보여줍니다. 이렇게 노력하는 이유는 차세대 AI 환경을 지원하는 데 이더넷이 계속해서 핵심적인 역할을 하도록 하기 위해서입니다.



Dell Networking에 대한 자세한 정보



Dell Technologies 전문가에게 문의



ESG의 전문 분석 기관 브리프 보기



IDC의 전문 분석 기관 브리프 보기