NVA-1173 NVIDIA DGX 시스템이 지원되는 NetApp AIPod

NetApp Solutions

NetApp
September 23, 2024

목차

NVA-1173 NVIDIA DGX 시스템이 지원되는 NetApp AIPod - 소개

POWERED BY

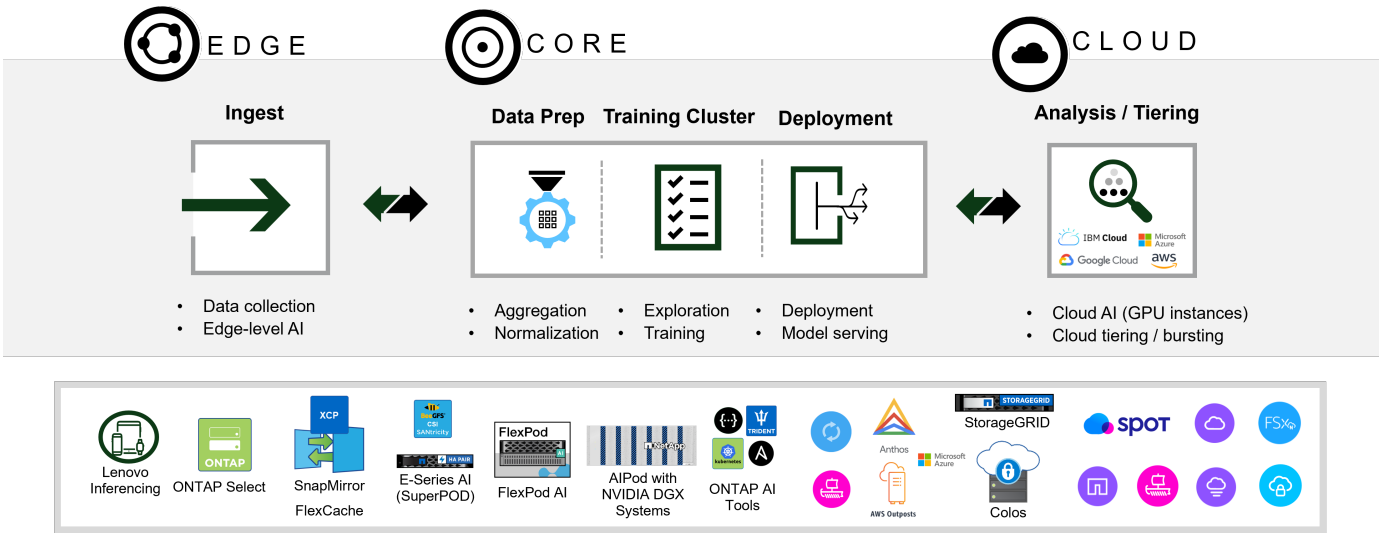
NVIDIA

NetApp 솔루션 엔지니어링

핵심 요약

NetApp 및 #8482, NVIDIA DGX 및 #8482, 시스템 및 NetApp 클라우드 연결 스토리지 시스템을 지원하는 AIPod는 설계 복잡성과 추측을 제거함으로써 머신 러닝(ML) 및 인공 지능(AI) 워크로드를 위한 인프라 구축을 간소화합니다. NVIDIA DGX BasePOD 및 #8482를 기반으로 구축하고 NVIDIA DGX 시스템이 탑재된 AIPod는 고객이 작은 규모로 시작한 후 중단 없이 성장하면서 에지에서 코어 및 클라우드까지 포괄하여 데이터를 지능적으로 관리할 수 있도록 NetApp AFF 스토리지 시스템을 추가합니다. NetApp AIPod는 NetApp AI 솔루션의 대규모 포트폴리오의 일부입니다. 아래 그림에 나와 있습니다.

_NetApp AI 솔루션 포트폴리오 _



이 문서에서는 AIPod 참조 아키텍처의 주요 구성 요소, 시스템 연결 및 구성 정보, 검증 테스트 결과 및 솔루션 사이징 지침을 설명합니다. 이 문서는 ML/DL 및 분석 워크로드를 위한 고성능 인프라를 구축하려는 NetApp 및 파트너 솔루션 엔지니어와 고객 전략적 의사 결정자를 위해 만들어졌습니다.

NVA-1173 NVIDIA DGX 시스템이 지원되는 NetApp AIPod - 소개

POWERED BY

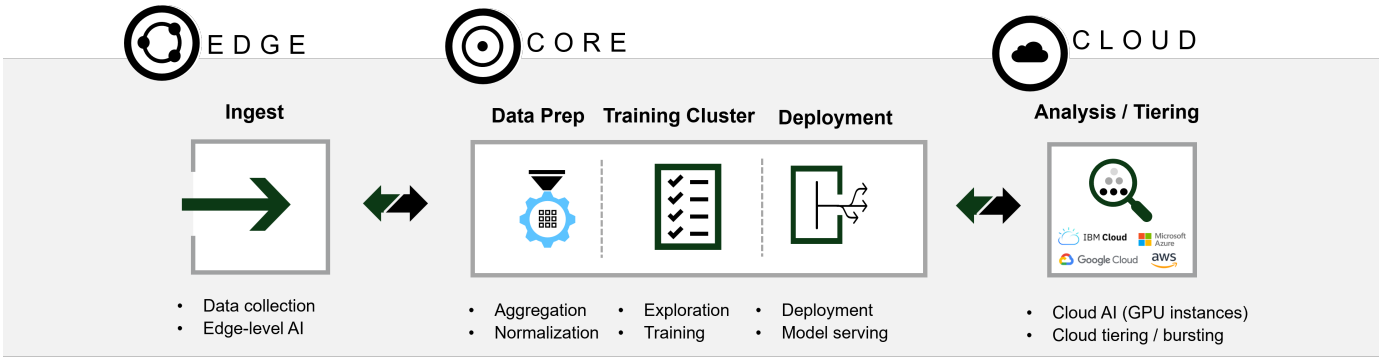
NVIDIA

NetApp 솔루션 엔지니어링

핵심 요약

NetApp 및 #8482, NVIDIA DGX 및 #8482, 시스템 및 NetApp 클라우드 연결 스토리지 시스템을 지원하는 AIPod는 설계 복잡성과 추측을 제거함으로써 머신 러닝(ML) 및 인공 지능(AI) 워크로드를 위한 인프라 구축을 간소화합니다. NVIDIA DGX BasePOD 및 #8482를 기반으로 구축하고 NVIDIA DGX 시스템이 탑재된 AIPod는 고객이 작은 규모로 시작한 후 중단 없이 성장하면서 에지에서 코어 및 클라우드까지 포괄하여 데이터를 지능적으로 관리할 수 있도록 NetApp AFF 스토리지 시스템을 추가합니다. NetApp AIPod는 NetApp AI 솔루션의 대규모 포트폴리오의 일부입니다. 아래 그림에 나와 있습니다.

_NetApp AI 솔루션 포트폴리오 _



이 문서에서는 AIPod 참조 아키텍처의 주요 구성 요소, 시스템 연결 및 구성 정보, 검증 테스트 결과 및 솔루션 사이징 지침을 설명합니다. 이 문서는 ML/DL 및 분석 워크로드를 위한 고성능 인프라를 구축하려는 NetApp 및 파트너 솔루션 엔지니어와 고객 전략적 의사 결정자를 위해 만들어졌습니다.

NVA-1173 NetApp AIPod 및 NVIDIA DGX 시스템 - 하드웨어 구성요소

이 섹션에서는 NVIDIA DGX 시스템이 지원되는 NetApp AIPod의 하드웨어 구성요소에 대해 중점적으로 다룹니다.

NetApp AFF 스토리지 시스템

NetApp AFF 최첨단 스토리지 시스템을 사용하면 IT 부서에서 업계 최고 수준의 성능, 탁월한 유연성, 클라우드 통합, 동급 최고의 데이터 관리 등을 통해 엔터프라이즈 스토리지 요구사항을 충족할 수 있습니다. 플래시 전용으로 설계된 AFF 시스템은 비즈니스 크리티컬 데이터를 더 빠르게 처리하고 관리, 보호할 수 있도록 지원합니다.

AFF A90 스토리지 시스템

NetApp ONTAP 데이터 관리 소프트웨어 기반의 NetApp AFF A90은 기본 데이터 보호, 선택적 랜섬웨어 방지 기능, 가장 중요한 비즈니스 워크로드를 지원하는 데 필요한 고성능 및 복원력을 제공합니다. 또한 미션 크리티컬 운영의 중단을 제거하고 성능 조정을 최소화하며 랜섬웨어 공격으로부터 데이터를 보호합니다. 다음과 같은 이점을 제공합니다. • 업계 최고의 성능 • 완벽한 데이터 보안 • 간소화된 무중단 업그레이드

_NetApp AFF A90 스토리지 시스템 _

업계 최고의 성능

AFF A90은 딥 러닝, AI, 고속 분석과 같은 차세대 워크로드와 Oracle, SAP HANA, Microsoft SQL Server, 가상 애플리케이션과 같은 기존 엔터프라이즈 데이터베이스를 손쉽게 관리합니다. 이 솔루션은 HA 쌍당 최대 2.4M IOPS와 지연 시간을 100µs의 낮은 속도로 비즈니스 크리티컬 애플리케이션을 최고 수준으로 실행하는 동시에, 이전 NetApp 모델에 비해 성능을 최대 50%까지 향상합니다. RDMA 기반 NFS, pNFS 및 Session Trunking을 사용하면 기존 데이터 센터 네트워킹 인프라를 사용하여 차세대 애플리케이션에 필요한 높은 수준의 네트워크 성능을 달성할 수 있습니다. 또한 고객은 SAN, NAS 및 오브젝트 스토리지에 대한 통합 멀티 프로토콜 지원을 통해 확장하고 성장할 수 있으며 온프레미스 또는 클라우드 데이터에 대한 통합 단일 ONTAP 데이터 관리 소프트웨어로 최대 유연성을 제공할 수 있습니다. 또한 Active IQ과 Cloud Insights에서 제공하는 AI 기반 예측 분석으로 시스템 상태를 최적화할 수 있습니다.

타협 없는 데이터 보안

AFF A90 시스템에는 NetApp 통합 및 애플리케이션 정합성을 보장하는 데이터 보호 소프트웨어의 전체 제품군이 포함되어 있습니다. 선점 및 공격 후 복구를 위한 내장 데이터 보호 및 최첨단 랜섬웨어 방지 솔루션을 제공합니다. 악성 파일은 디스크에 기록되는 것을 막아낼 수 있으며, 저장 문제를 쉽게 모니터링하여 통찰력을 얻을 수 있습니다.

단순화된 무중단 업그레이드

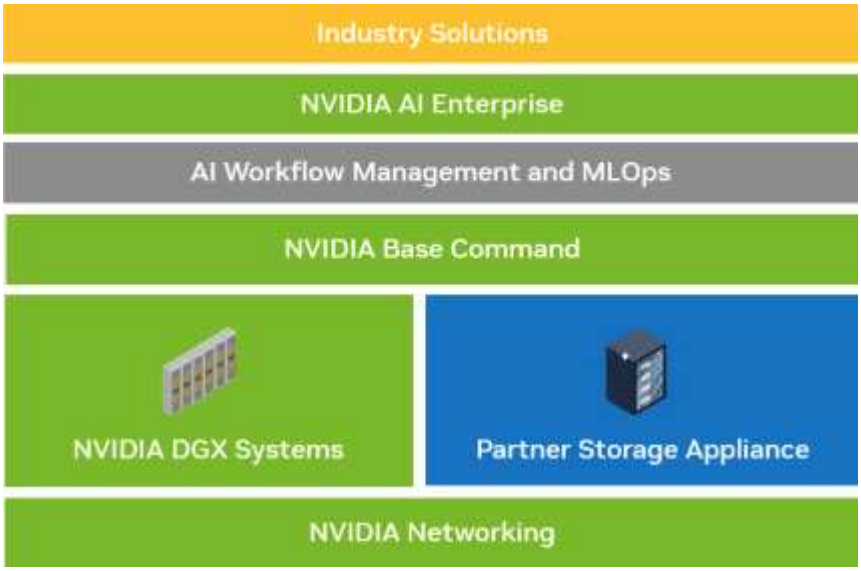
AFF A90은 기존 A800 고객을 위한 무중단 섀시 내 업그레이드로 사용할 수 있습니다. NetApp은 우수한 안정성, 가용성, 서비스 가능성 및 관리성(RASM) 기능을 통해 간단히 업데이트하고 미션 크리티컬 운영 중단을 제거합니다. 또한, NetApp ONTAP 소프트웨어는 모든 시스템 구성 요소에 대해 펌웨어 업데이트를 자동으로 적용하므로 운영 효율성을 높이고 IT 팀의 일상적인 활동을 간소화합니다.

대규모 구축의 경우 AFF A1K 시스템은 최고의 성능과 용량 옵션을 제공하는 한편, AFF A70 및 AFF C800와 같은 다른 NetApp 스토리지 시스템은 저렴한 비용으로 소규모 구축에 필요한 옵션을 제공합니다.

NVIDIA DGX 베이스POD

NVIDIA DGX BasePOD는 NVIDIA 하드웨어 및 소프트웨어 구성 요소, MLOps 솔루션, 타사 스토리지로 구성된 통합 솔루션입니다. 고객은 NVIDIA 제품과 검증된 파트너 솔루션으로 스케일아웃 시스템 설계의 모범 사례를 활용하여 AI 개발을 위한 효율적이고 관리하기 쉬운 플랫폼을 구축할 수 있습니다. 그림 1은 NVIDIA DGX BasePOD의 다양한 구성 요소를 보여줍니다.

NVIDIA DGX BasePOD 솔루션



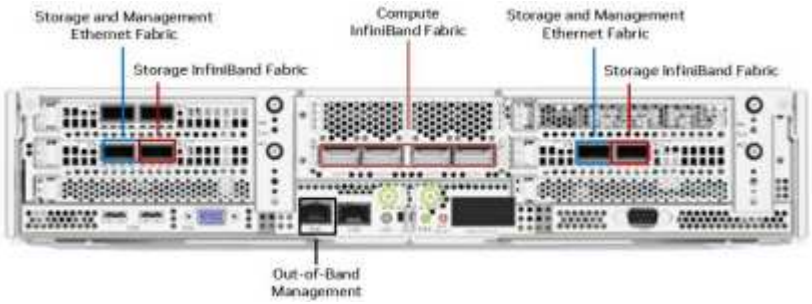
NVIDIA DGX H100 시스템

NVIDIA DGX H100 및 #8482; 시스템은 NVIDIA H100 Tensor 코어 GPU의 획기적인 성능을 통해 가속화된 AI 파워하우스입니다.

_NVIDIA DGX H100 시스템 _

DGX H100 시스템의 주요 사양은 다음과 같습니다. • 8개의 NVIDIA H100 GPU. GPU당 • 80GB GPU 메모리, 총 640GB • NVIDIA NVSwitch ™ 칩 4개 • PCIe 5.0을 지원하는 듀얼 56코어 인텔 ® 제온 ® 플래티넘 8480 프로세서. • 2TB의 DDR5 시스템 메모리. • 8개의 단일 포트 NVIDIA ConnectX 및 #174; -7(InfiniBand/이더넷) 어댑터 및 2개의 듀얼 포트 NVIDIA ConnectX-7(InfiniBand/이더넷) 어댑터를 제공하는 OSFP 포트 4개. • DGX OS용 1.92TB M.2 NVMe 드라이브 2개, 스토리지/캐시용 3.84TB U.2 NVMe 드라이브 8개 • 10.2kW 최대 출력. DGX H100 CPU 트레이의 후면 포트는 아래에 나와 있습니다. OSFP 포트 4개는 InfiniBand 컴퓨팅 패브릭에 8개의 ConnectX-7 어댑터를 제공합니다. 각 듀얼 포트 ConnectX-7 어댑터 쌍은 스토리지 및 관리 패브릭에 대한 병렬 경로를 제공합니다. 대역외 포트는 BMC 액세스에 사용됩니다.

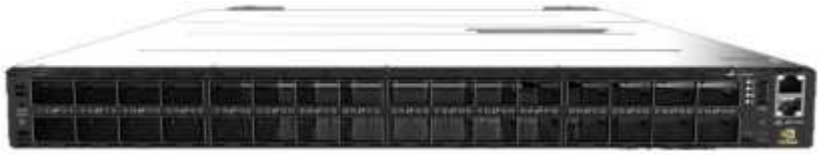
_NVIDIA DGX H100 후면 패널 _



NVIDIA 네트워킹

NVIDIA Quantum-2 QM9700 스위치

_NVIDIA Quantum-2 QM9700 InfiniBand 스위치 _



NVIDIA Quantum-2 QM9700 스위치는 400GB/s InfiniBand 연결 기능을 갖추고 있으며 NVIDIA Quantum-2 InfiniBand BasePOD 구성의 컴퓨팅 패브릭에 전력을 공급합니다. ConnectX-7 단일 포트 어댑터는 InfiniBand 컴퓨팅 패브릭에 사용됩니다. 각 NVIDIA DGX 시스템은 각 QM9700 스위치에 대한 이중 연결을 제공하므로 시스템 간에 대역폭이 높고 지연 시간이 짧은 경로를 여러 개 제공합니다.

NVIDIA Spectrum-3 SN4600 스위치

_NVIDIA Spectrum-3 SN4600 스위치 _

NVIDIA 스펙트럼 및 #8482; -3 SN4600 스위치는 총 128개의 포트(스위치당 64개)를 제공하여 DGX BasePOD의 대역 내 관리를 위한 이중화 연결을 제공합니다. NVIDIA SN4600 스위치는 1GbE 및 200GbE 사이의 속도를 제공할 수 있습니다. 이더넷을 통해 연결된 스토리지 어플라이언스의 경우 NVIDIA SN4600 스위치도 사용됩니다. NVIDIA DGX 이중 포트 ConnectX-7 어댑터의 포트는 대역 내 관리 및 스토리지 연결에 모두 사용됩니다.

NVIDIA Spectrum SN2201 스위치

_NVIDIA Spectrum SN2201 스위치 _

NVIDIA Spectrum SN2201 스위치는 대역 외 관리를 위한 연결을 제공하는 48개의 포트를 제공합니다. 대역 외 관리는 DGX BasePOD의 모든 구성 요소에 대한 통합 관리 연결을 제공합니다.

NVIDIA ConnectX-7 어댑터

_NVIDIA ConnectX-7 어댑터 _

NVIDIA ConnectX-7 어댑터는 25/50/100/200/400G의 처리량을 제공할 수 있습니다. NVIDIA DGX 시스템은 단일 포트 및 이중 포트 ConnectX-7 어댑터를 모두 사용하여 400GB/s InfiniBand 및 이더넷을 사용하는 DGX BasePOD 구축에 유연성을 제공합니다.

NVA-1173 NVIDIA DGX 시스템 및 NetApp AIPod - 소프트웨어 구성 요소

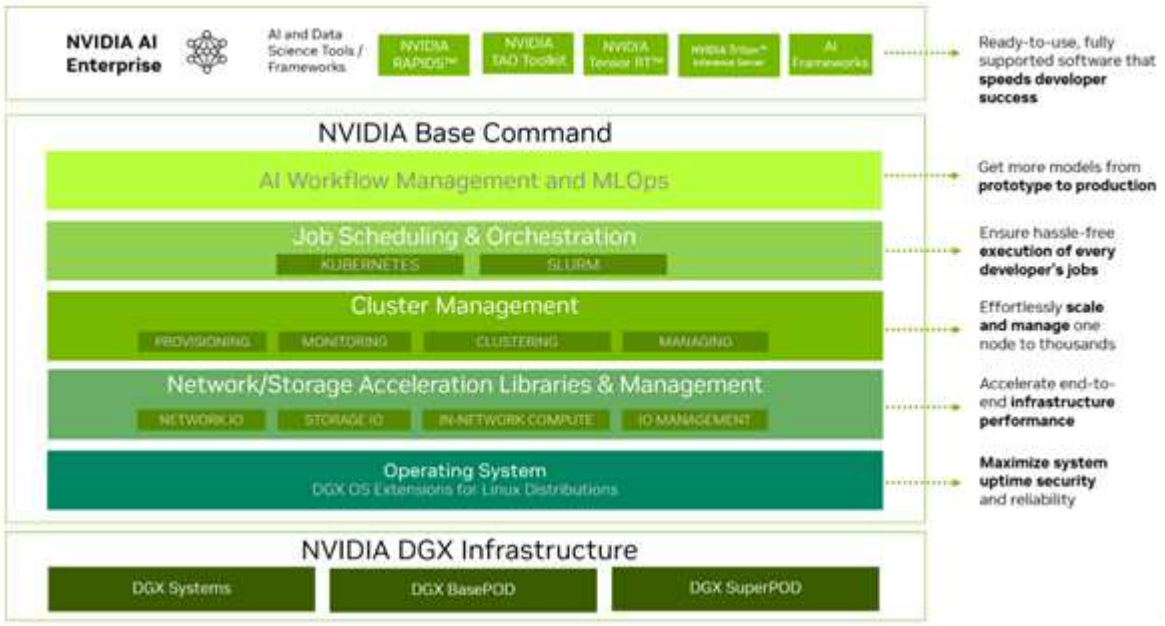
이 섹션에서는 NVIDIA DGX 시스템이 지원되는 NetApp AIPod의 소프트웨어 구성요소에 대해 중점적으로 다룹니다.

NVIDIA 소프트웨어

NVIDIA Base 명령

NVIDIA Base Command & #8482; 는 모든 DGX BasePOD를 지원하므로 조직이 최고의 NVIDIA 소프트웨어 혁신을 활용할 수 있습니다. 기업은 엔터프라이즈급 오케스트레이션 및 클러스터 관리, 컴퓨팅 속도를 높이는 라이브러리, 스토리지 및 네트워크 인프라, AI 워크로드에 최적화된 운영 체제(OS)를 포함하는 검증된 플랫폼을 통해 투자의 잠재력을 모두 활용할 수 있습니다.

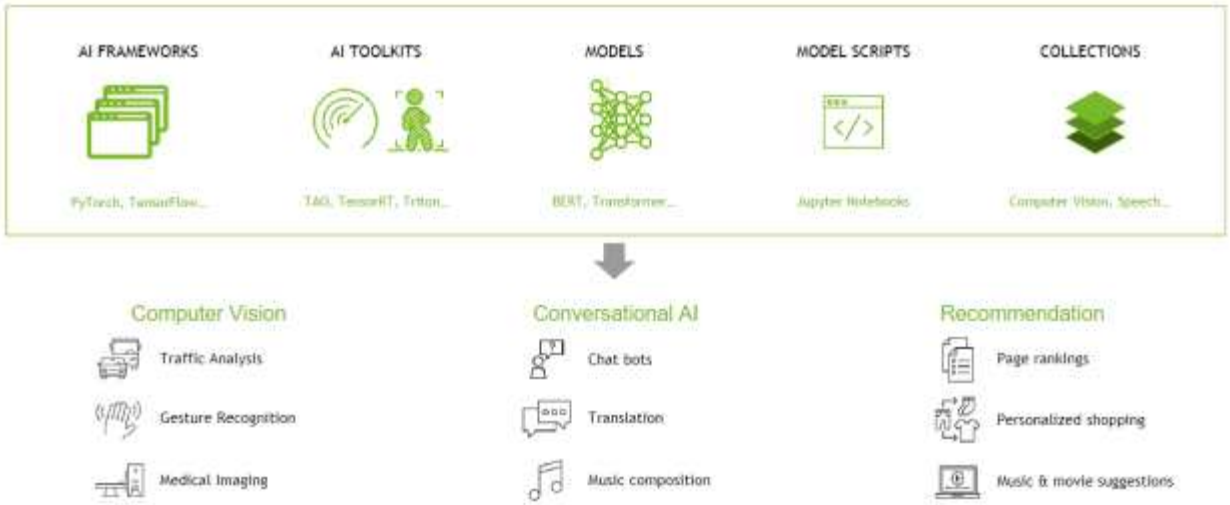
_NVIDIA BaseCommand 솔루션 _



NGC(NVIDIA GPU Cloud)

NVIDIA NGC ™ 는 다양한 수준의 AI 전문 지식을 갖춘 데이터 과학자, 개발자 및 연구자의 요구를 충족하는 소프트웨어를 제공합니다. NGC에서 호스팅되는 소프트웨어는 종합적인 CVE(Common Vulnerability and Exposure), 암호화 및 개인 키 집합을 검사합니다. 테스트를 거쳐 여러 GPU로 확장하고, 대부분의 경우 다중 노드로 확장하여 사용자가 DGX 시스템에 대한 투자를 극대화할 수 있도록 설계되었습니다.

_NVIDIA GPU 클라우드 _



NVIDIA AI 엔터프라이즈

NVIDIA AI Enterprise는 모든 기업에 세대 AI를 제공하여 NVIDIA DGX 플랫폼에서 실행하도록 최적화된 세대 AI 기반 모델에 가장 빠르고 효율적인 런타임을 제공하는 엔드 투 엔드 소프트웨어 플랫폼입니다. 운영 수준의 보안, 안정성 및 관리 효율성을 통해 생성 가능한 AI 솔루션 개발을 간소화합니다. NVIDIA AI Enterprise는 엔터프라이즈 개발자가 사전 훈련된 모델, 최적화된 프레임워크, 마이크로서비스, 가속 라이브러리 및 엔터프라이즈 지원에 액세스할 수 있도록 DGX BasePOD에 포함되어 있습니다.

NetApp 소프트웨어

NetApp ONTAP를 참조하십시오

NetApp의 최신 세대 스토리지 관리 소프트웨어인 ONTAP 9는 기업이 인프라를 현대화하고 클라우드 지원 데이터 센터로 전환할 수 있도록 지원합니다. ONTAP는 업계 최고 수준의 데이터 관리 기능을 활용하여 데이터가 상주하는 위치와 상관없이 단일 툴셋으로 데이터를 관리하고 보호할 수 있습니다. 필요에 따라 에지, 코어, 클라우드 등 어느 위치로도 데이터를 자유롭게 이동할 수 있습니다. ONTAP 9에는 데이터 관리를 단순화하고, 중요 데이터를 더 빨리 처리하고, 보호하며, 하이브리드 클라우드 아키텍처 전체에서 차세대 인프라 기능을 지원하는 다양한 기능이 포함되어 있습니다.

데이터 가속화 및 보호

ONTAP는 탁월한 수준의 성능과 데이터 보호를 제공하며 다음과 같은 방법으로 이러한 기능을 확장합니다.

- 성능 및 짧은 지연 시간: ONTAP는 NFS over RDMA, pNFS(Parallel NFS) 및 NFS 세션 트렁킹을 사용하는 NVIDIA GPUDirect Storage(GDS)를 지원하는 등 가장 짧은 지연 시간으로 가장 높은 처리량을 제공합니다.
- 데이터 보호: ONTAP는 내장 데이터 보호 기능과 모든 플랫폼에서 공통된 관리를 통해 업계에서 가장 강력한 안티-랜섬웨어 보장을 제공합니다.
- NVE(NetApp 볼륨 암호화). ONTAP는 온보드 및 외부 키 관리를 모두 지원하는 기본 볼륨 레벨 암호화를 제공합니다.
- 스토리지 멀티 테넌시 및 다단계 인증. ONTAP를 사용하면 인프라 리소스를 최고 수준의 보안으로 공유할 수 있습니다.

데이터 관리를 단순화하십시오

데이터 관리는 AI 애플리케이션에 적합한 리소스를 사용하고 AI/ML 데이터 세트를 교육할 수 있도록 엔터프라이즈 IT 운영 및 데이터 과학자에게 매우 중요합니다. NetApp 기술에 대한 다음 추가 정보는 이 검증의 범위에 포함되지 않지만, 배포에 따라 달라질 수 있습니다.

ONTAP 데이터 관리 소프트웨어에는 운영을 간소화 및 단순화하고 총 운영 비용을 절감하는 다음과 같은 기능이 있습니다.

- 스냅샷 및 클론을 통해 ML/DL 워크플로에 대한 협업, 병렬 실험 및 향상된 데이터 거버넌스를 지원할 수 있습니다.
- SnapMirror를 사용하면 하이브리드 클라우드 및 다중 사이트 환경에서 데이터를 원활하게 이동할 수 있으며 필요한 시간과 장소에 데이터를 제공할 수 있습니다.
- 인라인 데이터 컴팩션 및 확대된 중복제거: 데이터 컴팩션은 스토리지 블록 내부의 낭비되는 공간을 줄이고, 중복제거는 실제 용량을 상당히 늘려줍니다. 이는 로컬에 저장된 데이터와 클라우드로 계층화된 데이터에 적용됩니다.
- 최소, 최대 및 적응형 서비스 품질(AQoS): 세부적인 서비스 품질(QoS) 제어로 고도의 공유 환경에서 중요 애플리케이션의 성능 수준을 유지할 수 있습니다.
- NetApp FlexGroup을 사용하면 스토리지 클러스터의 모든 노드에 데이터를 분산하여 매우 큰 데이터 세트에 필요한 대용량 및 높은 성능을 제공할 수 있습니다.
- NetApp FabricPool를 참조하십시오. AWS(Amazon Web Services), Azure, NetApp StorageGRID 스토리지 솔루션을 포함한 퍼블릭 클라우드 및 프라이빗 클라우드 스토리지에 콜드 데이터를 자동으로 계층화합니다. FabricPool에 대한 자세한 내용은 를 참조하십시오 "TR-4598: FabricPool 모범 사례".
- NetApp FlexCache 를 참조하십시오. 파일 배포를 간소화하고 WAN 지연 시간을 줄이며 WAN 대역폭 비용을 낮추는 원격 볼륨 캐싱 기능을 제공합니다. FlexCache를 사용하면 여러 사이트에 분산된 제품 개발을 지원하고, 원격지에서 기업 데이터 세트에 더 빠르게 액세스할 수 있습니다.

미래 지향형 인프라

ONTAP은 다음과 같은 기능을 통해 끊임없이 변화하는 까다로운 비즈니스 요구사항을 충족할 수 있도록 지원합니다.

- 원활한 확장 및 무중단 운영: ONTAP는 기존 컨트롤러 및 스케일아웃 클러스터에 온라인으로 용량을 추가할 수 있도록 지원합니다. 고객은 고비용이 따르는 데이터 마이그레이션이나 운영 중단 없이 NVMe 및 32Gb FC와 같은 최신 기술로 업그레이드할 수 있습니다.
- 클라우드 연결: ONTAP은 클라우드에 가장 많이 연결된 스토리지 관리 소프트웨어로, 모든 퍼블릭 클라우드에서 ONTAP Select(소프트웨어 정의 스토리지) 및 NetApp Cloud Volumes Service(클라우드 네이티브 인스턴스)에 대한 옵션을 제공합니다.
- 새로운 애플리케이션과 통합: ONTAP은 기존 엔터프라이즈 앱을 지원하는 인프라와 동일한 인프라를 사용하여 자율주행 차량, 스마트 시티, Industry 4.0과 같은 차세대 플랫폼 및 애플리케이션을 위한 엔터프라이즈급 데이터 서비스를 제공합니다.

NetApp DataOps 툴킷

NetApp DataOps 툴킷은 고성능 스케일아웃 NetApp 스토리지가 지원하는 개발/교육 작업 공간 및 추론 서버의 관리를 단순화하는 Python 기반 툴입니다. DataOps 툴킷은 독립 실행형 유틸리티로 작동할 수 있으며, NetApp Astra Trident를 활용하여 스토리지 운영을 자동화하는 Kubernetes 환경에서 더 효과적입니다. 주요 기능은 다음과 같습니다.

- 고성능 스케일아웃 NetApp 스토리지를 기반으로 하는 새로운 고용량 JupyterLab 작업 공간을 빠르게 프로비저닝합니다.

- 엔터프라이즈급 NetApp 스토리지를 통해 지원되는 새로운 NVIDIA Triton Inference Server 인스턴스를 빠르게 프로비저닝합니다.
- 실험이나 신속한 반복을 지원하기 위해 고용량 JupyterLab 작업 공간의 거의 즉각적인 클론 복제
- 백업 및/또는 추적 기능/기준선 설정을 위한 대용량 JupyterLab 작업 공간의 거의 즉각적인 스냅샷
- 대용량 고성능 데이터 볼륨의 거의 즉각적인 프로비저닝, 복제, 스냅샷

NetApp Astra Trident

Astra Trident는 Anthos를 비롯한 컨테이너 및 Kubernetes 배포를 위한 완전히 지원되는 오픈 소스 스토리지 오케스트레이터입니다. Trident는 NetApp ONTAP를 비롯한 전체 NetApp 스토리지 포트폴리오와 연동되며 NFS, NVMe/TCP, iSCSI 연결도 지원합니다. Trident는 최종 사용자가 스토리지 관리자의 개입 없이 NetApp 스토리지 시스템에서 스토리지를 프로비저닝 및 관리할 수 있도록 하여 DevOps 워크플로우를 가속합니다.

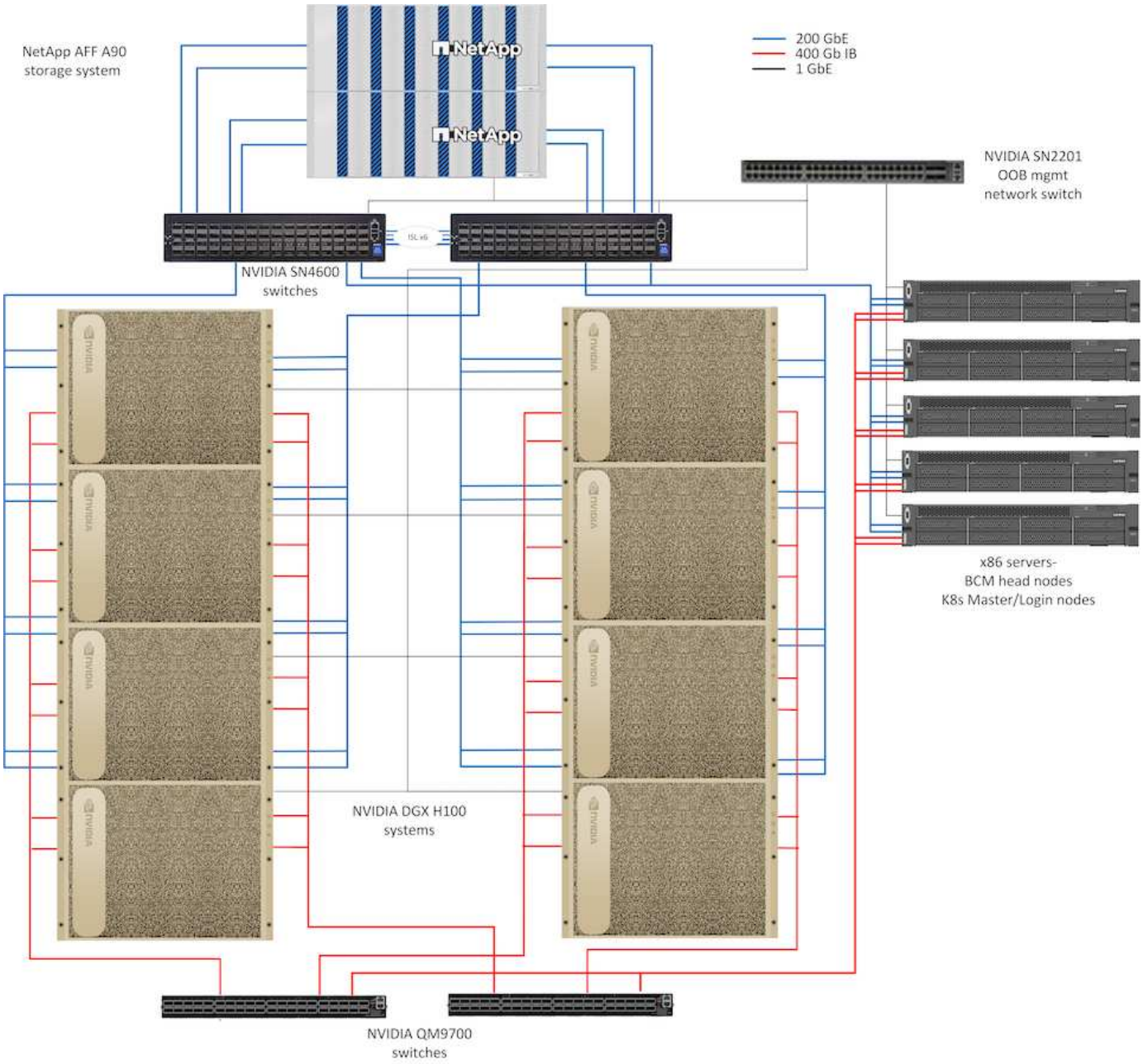
NVA-1173 NetApp AIPod 및 NVIDIA DGX H100 시스템 - 솔루션 아키텍처

이 섹션에서는 NVIDIA DGX 시스템이 지원되는 NetApp AIPod의 아키텍처에 중점을 둡니다.

DGX 시스템이 장착된 NetApp AIPod

이 참조 아키텍처는 컴퓨팅 노드 간의 400GB/s InfiniBand(IB) 연결을 통해 컴퓨팅 클러스터 인터커넥트 및 스토리지 액세스를 위해 별도의 패브릭을 활용합니다. 아래의 그림은 DGX H100 시스템이 장착된 NetApp AIPod의 전반적인 솔루션 토폴로지를 보여줍니다.

_NetApp AIPOD 솔루션 토폴로지 _



네트워크 설계

이 구성에서 컴퓨팅 클러스터 패브릭은 고가용성을 위해 함께 연결된 한 쌍의 QM9700 400GB/s IB 스위치를 사용합니다. 각 DGX H100 시스템은 8개의 연결을 사용하여 스위치에 연결되며, 짝수 번호 포트는 하나의 스위치에 연결되고 홀수 번호 포트는 다른 스위치에 연결됩니다.

스토리지 시스템 액세스, 대역 내 관리 및 클라이언트 액세스의 경우 한 쌍의 SN4600 이더넷 스위치가 사용됩니다. 스위치는 스위치 간 링크로 연결되고 여러 VLAN으로 구성되어 다양한 트래픽 유형을 격리합니다. 기본 L3 라우팅은 특정 VLAN 간에 활성화되어 고가용성을 위해 동일한 스위치에서 클라이언트와 스토리지 인터페이스 간의 다중 경로를 지원할 수 있습니다. 대규모 구축의 경우 필요에 따라 스파인 스위치 및 리프(leaf)를 위한 추가 스위치 쌍을 추가하여 이더넷 네트워크를 리프-스파인 구성으로 확장할 수 있습니다.

컴퓨팅 상호 연결 및 고속 이더넷 네트워크 외에도 모든 물리적 장치가 대역 외 관리를 위해 하나 이상의 SN2201 이더넷 스위치에 연결됩니다. "배포 세부 정보"네트워크 구성에 대한 자세한 내용은 페이지를 참조하십시오.

DGX H100 시스템의 스토리지 액세스 개요

각 DGX H100 시스템은 관리 및 스토리지 트래픽을 위해 2개의 이중 포트 ConnectX-7 어댑터를 사용하여 프로비저닝됩니다. 이 솔루션의 경우 각 카드의 두 포트가 같은 스위치에 연결됩니다. 그런 다음, 각 카드에서 하나의 포트를 각 스위치에 연결하고 있는 LACP MLAG 결합으로 구성하고, 대역 내 관리, 클라이언트 액세스 및 사용자 레벨 스토리지 액세스를 위한 VLAN이 이 결합에서 호스팅됩니다.

각 카드의 다른 포트는 AFF A90 스토리지 시스템에 대한 연결에 사용되며 워크로드 요구사항에 따라 여러 구성으로 사용할 수 있습니다. NVIDIA Magnum IO GPUDirect 스토리지를 지원하기 위해 NFS over RDMA를 사용하는 구성의 경우 포트가 별도의 VLAN에서 IP 주소와 함께 개별적으로 사용됩니다. RDMA가 필요하지 않은 배포의 경우 LACP 본딩을 사용하여 스토리지 인터페이스를 구성하여 고가용성 및 추가 대역폭을 제공할 수도 있습니다. RDMA 사용 여부에 관계없이 클라이언트는 NFS v4.1 pNFS 및 세션 트렁킹을 사용하여 스토리지 시스템을 마운트하여 클러스터의 모든 스토리지 노드에 대한 병렬 액세스를 지원할 수 있습니다. "배포 세부 정보"클라이언트 구성에 대한 자세한 내용은 페이지를 참조하십시오.

DGX H100 시스템 연결에 대한 자세한 내용은 를 "NVIDIA BasePOD 설명서"참조하십시오.

스토리지 시스템 설계

각 AFF A90 스토리지 시스템은 각 컨트롤러의 200GbE 포트 6개를 사용하여 연결됩니다. 각 컨트롤러의 포트 4개는 DGX 시스템에서 워크로드 데이터에 액세스하는 데 사용되며, 각 컨트롤러의 포트 2개는 클러스터 관리 아티팩트 및 사용자 홈 디렉토리에 대한 관리 플레인 서버의 액세스를 지원하기 위해 LACP 인터페이스 그룹으로 구성됩니다. 스토리지 시스템에서 수행하는 데이터 액세스는 모두 NFS를 통해 제공되며, AI 워크로드 액세스 전용 SVM(Storage Virtual Machine) 및 클러스터 관리 전용 SVM은 별도로 제공됩니다.

"배포 세부 정보"스토리지 시스템 구성에 대한 자세한 내용은 페이지를 참조하십시오.

관리 플레인 서버

이 레퍼런스 아키텍처에는 관리 플레인을 위한 5개의 CPU 기반 서버도 포함되어 있습니다. 이러한 시스템 중 2개가 클러스터 구축 및 관리를 위한 NVIDIA Base Command Manager의 헤드 노드로 사용됩니다. 다른 3개의 시스템은 작업 예약을 위해 Slurm을 활용하는 구축에 Kubernetes 마스터 노드 또는 로그인 노드와 같은 추가 클러스터 서비스를 제공하는 데 사용됩니다. Kubernetes를 활용하는 구축에서는 NetApp Astra Trident CSI 드라이버를 활용하여 AFF A900 스토리지 시스템의 관리 및 AI 워크로드를 위한 영구 스토리지를 통한 자동 프로비저닝 및 데이터 서비스를 제공할 수 있습니다.

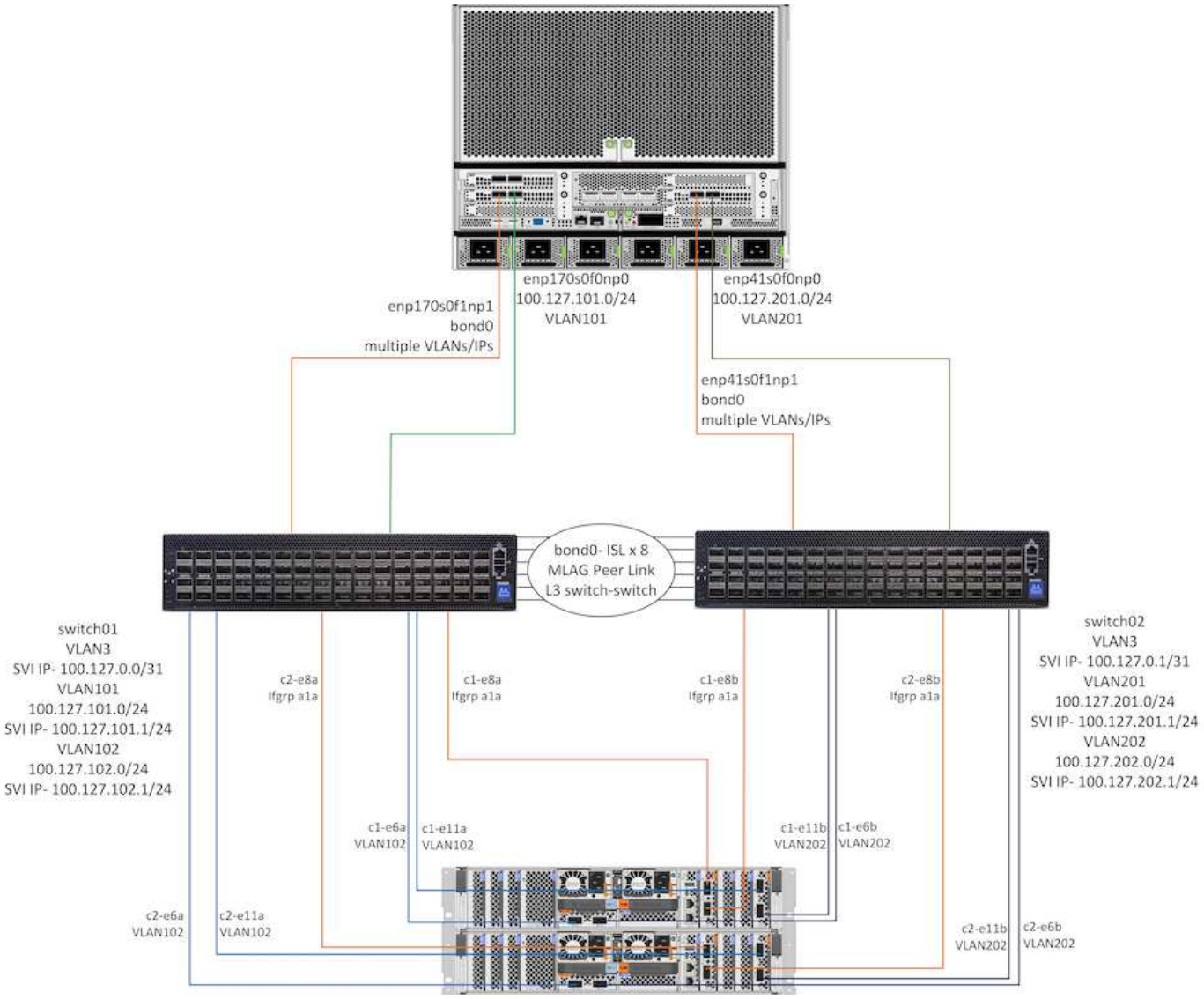
각 서버는 IB 스위치와 이더넷 스위치 모두에 물리적으로 연결되어 클러스터 배포 및 관리를 지원하며, 앞서 설명한 대로 클러스터 관리 아티팩트를 저장하기 위해 관리 SVM을 통해 스토리지 시스템에 NFS 마운트로 구성됩니다.

NVA-1173 NVIDIA DGX 시스템 및 NetApp AIPod - 구축 세부 정보

이 섹션에서는 이 솔루션을 검증하는 동안 사용되는 배포 세부 정보를 설명합니다. 사용된 IP 주소는 예이며 배포 환경에 따라 수정해야 합니다. 이 구성을 구현하는 데 사용되는 특정 명령에 대한 자세한 내용은 해당 제품 설명서를 참조하십시오.

아래 다이어그램은 1개의 DGX H100 시스템 및 1개의 AFF A90 컨트롤러 HA 쌍에 대한 상세 네트워크 및 연결 정보를 보여줍니다. 다음 섹션의 배포 지침은 이 다이어그램의 세부 정보를 기반으로 합니다.

_NetApp AIPOD 네트워크 구성 _



다음 표는 최대 16개의 DGX 시스템 및 2개의 AFF A90 HA 쌍에 대한 케이블 연결 할당의 예를 보여줍니다.

스위치 및 포트	장치	장치 포트입니다
스위치 1 포트 1-16입니다	DGX-H100-01 ~ -16	enp170s0f0np0, 슬롯1 포트 1
스위치 1 포트 17-32	DGX-H100-01 ~ -16	enp170s0f1np1, 슬롯1 포트 2
스위치 1 포트 33-36	AFF-A90-01 ~ -04	포트 e6a
스위치 1 포트 37-40	AFF-A90-01 ~ -04	포트 e11a
스위치 1 포트 41-44	AFF-A90-01 ~ -04	포트 e8a
스위치 1 포트 57-64	ISL에서 switch2로	포트 57-64
스위치 2 포트 1-16입니다	DGX-H100-01 ~ -16	enp41s0f0np0, 슬롯 2 포트 1
스위치 2 포트 17-32	DGX-H100-01 ~ -16	enp41s0f1np1, 슬롯 2 포트 2
스위치 2 포트 33-36	AFF-A90-01 ~ -04	포트 e6b
스위치 2 포트 37-40	AFF-A90-01 ~ -04	포트 e11b

스위치 및 포트	장치	장치 포트입니다
스위치 2 포트 41-44	AFF-A90-01 ~ -04	포트 e8b
스위치 2 포트 57-64	ISL에서 switch1로	포트 57-64

다음 표에는 이 검증에 사용된 다양한 구성 요소의 소프트웨어 버전이 나와 있습니다.

장치	소프트웨어 버전
NVIDIA SN4600 스위치	Cumulus Linux v5.9.1입니다
NVIDIA DGX 시스템	DGX OS v6.2.1(Ubuntu 22.04 LTS)
Mellanox OFED	24.01
NetApp AFF A90를 참조하십시오	NetApp ONTAP 9.14.1

스토리지 네트워크 구성

이 섹션에서는 이더넷 스토리지 네트워크 구성에 대한 주요 세부 정보를 간략하게 설명합니다. InfiniBand 컴퓨팅 네트워크 구성에 대한 자세한 내용은 을 "NVIDIA BasePOD 설명서"참조하십시오. 스위치 구성에 대한 자세한 내용은 을 "NVIDIA Cumulus Linux 설명서"참조하십시오.

SN4600 스위치를 구성하는 데 사용되는 기본 단계는 다음과 같습니다. 이 프로세스에서는 케이블 연결 및 기본 스위치 설정(관리 IP 주소, 라이센스 등)이 완료되었다고 가정합니다.

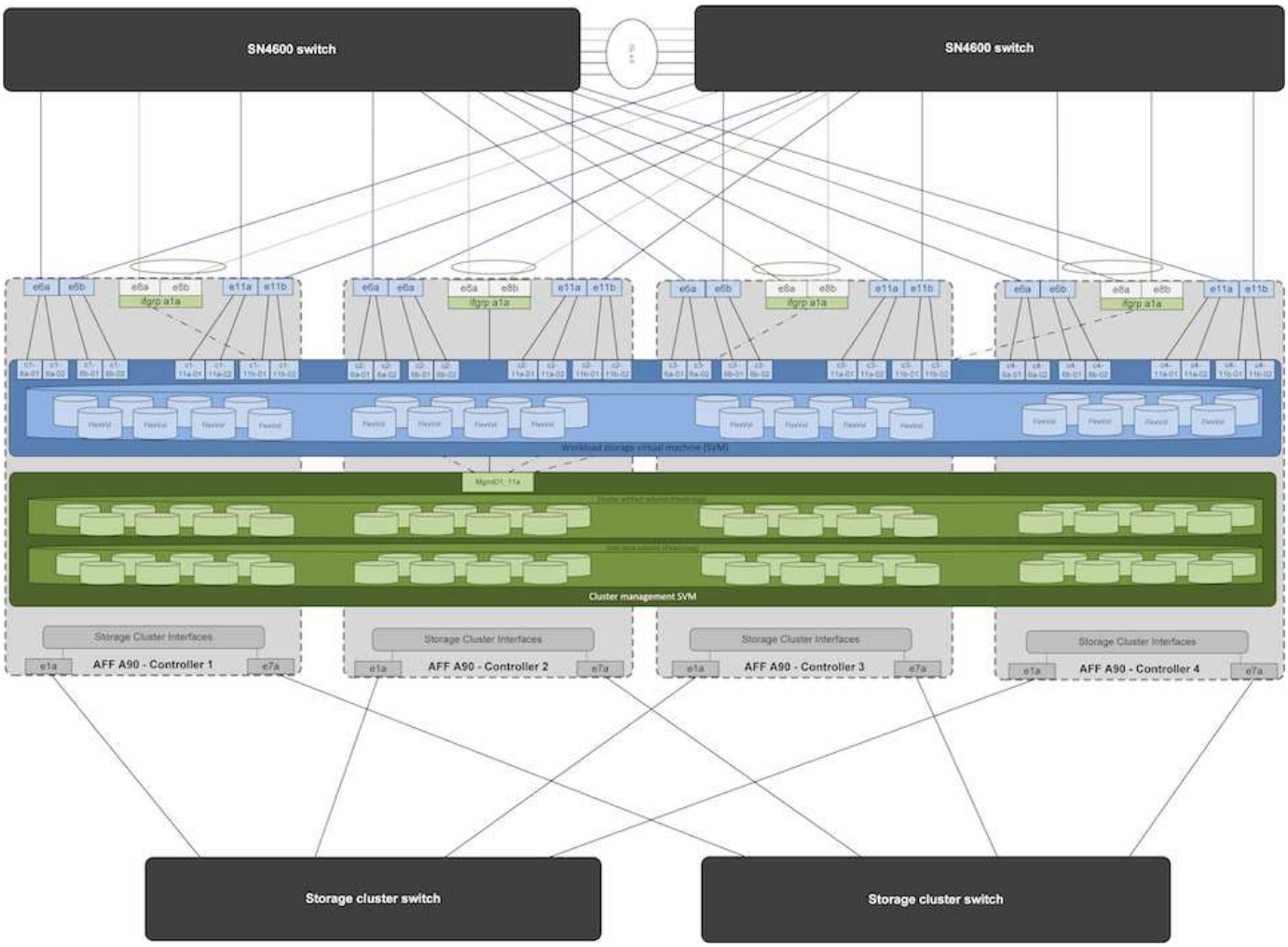
1. MLAG(Multi-Link Aggregation) 및 페일오버 트래픽을 사용하도록 스위치 간 ISL 본드를 구성합니다
   - 이 검증에서는 8개의 링크를 사용하여 테스트 중인 스토리지 구성에 충분한 대역폭을 제공했습니다
   - MLAG 활성화에 대한 자세한 지침은 Cumulus Linux 설명서를 참조하십시오.
2. 두 스위치의 각 클라이언트 포트 및 스토리지 포트 쌍에 대해 LACP MLAG를 구성합니다
   - DGX-H100-01의 각 스위치의 포트 swp17(enp170s0f1np1 및 enp41s0f1np1), DGX-H100-02의 포트 swp18 등(bond1-16)
   - AFF-A90-01(e8a 및 e8b)용 각 스위치의 포트 swp41, AFF-A90-02용 포트 swp42 등(bond17-20)
   - NV set interface bondX bond 멤버 swpX
   - NV set interface bondx bond MLAG id X
3. 기본 브리지 도메인에 모든 포트 및 MLAG 결합을 추가합니다
   - NV set int swp1-16,33-40 브리지 도메인 br_default
   - NV set int bond1-20 브리지 도메인 br_default
4. 각 스위치에서 RoCE를 활성화합니다
   - NV가 RoCE 모드 무손실을 설정합니다
5. 클라이언트 포트용 VLAN-2, 스토리지 포트용 2개, 관리용 1개, 스위치에 대한 L3 스위치용 1개를 구성합니다
   - 스위치 1 -
     - 클라이언트 NIC 장애 발생 시 L3 스위치에 대한 VLAN 3
     - 각 DGX 시스템의 스토리지 포트 1용 VLAN 101(enp170s0f0np0, 슬롯 1 포트 1)

        - 각 AFF A90 스토리지 컨트롤러의 포트 e6a 및 e11a용 VLAN 102
        - 각 DGX 시스템 및 스토리지 컨트롤러에 대한 MLAG 인터페이스를 사용하는 관리용 VLAN 301
    - 스위치 2 -
        - 클라이언트 NIC 장애 발생 시 L3 스위치에 대한 VLAN 3
        - 각 DGX 시스템의 스토리지 포트 2용 VLAN 201(enp41s0f0np0, slot2 포트 1)
        - 각 AFF A90 스토리지 컨트롤러의 포트 e6b 및 e11b에 대한 VLAN 202
        - 각 DGX 시스템 및 스토리지 컨트롤러에 대한 MLAG 인터페이스를 사용하는 관리용 VLAN 301
6. 각 VLAN에 물리적 포트(예: 클라이언트 VLAN의 클라이언트 포트 및 스토리지 VLAN의 스토리지 포트)를 적절하게 할당합니다
    - NV set int <swpX> bridge domain br_default access <vlan id>
    - MLAG 포트는 필요에 따라 연결된 인터페이스에서 여러 VLAN을 사용할 수 있도록 트렁크 포트로 유지되어야 합니다.
7. 각 VLAN에 SVI(스위치 가상 인터페이스)를 구성하여 게이트웨이 역할을 하고 L3 라우팅을 활성화합니다
    - 스위치 1 -
        - NV int VLAN3 IP address 100.127.0.0/31 설정
        - NV set int vlan101 IP address 100.127.101.1/24
        - NV set int vlan102 IP address 100.127.102.1 / 24
    - 스위치 2 -
        - NV int VLAN3 IP address 100.127.0.1/31 설정
        - NV set int vlan201 IP address 100.127.201.1 / 24
        - NV set int vlan202 IP address 100.127.202.1 / 24
8. 정적 라우트를 생성합니다
    - 동일한 스위치에 있는 서브넷에 대한 정적 경로가 자동으로 생성됩니다
    - 클라이언트 링크 장애 시 스위치에 대한 라우팅 전환을 위해 추가적인 정적 라우트가 필요합니다
        - 스위치 1 -
            - NV는 VRF 기본 라우터 정적 1000.127.128.0/17을 1000.127.0.1로 설정합니다
        - 스위치 2 -
            - NV는 VRF 기본 라우터 정적 100.127.0.0/17 을 100.127.0.0을 통해 설정합니다

스토리지 시스템 구성

이 섹션에서는 이 솔루션의 A90 스토리지 시스템 구성에 대한 주요 세부 정보를 설명합니다. ONTAP 시스템 구성에 대한 자세한 내용은 [ONTAP documentation] 을 참조하십시오. 아래 다이어그램은 스토리지 시스템의 논리적 구성을 보여 줍니다.

_NetApp A90 스토리지 클러스터 논리적 구성 _



스토리지 시스템을 구성하는 데 사용되는 기본 단계는 다음과 같습니다. 이 프로세스에서는 기본 스토리지 클러스터 설치가 완료된 것으로 가정합니다.

1. 각 컨트롤러에서 사용 가능한 모든 파티션에서 1개의 스페어를 제외한 1개의 애그리게이트를 구성합니다
   - Aggregate create-node <node>-aggregate <node>_data01-diskcount <47>
2. 각 컨트롤러에서 ifgrp를 구성합니다
   - NET 포트 ifgrp create-node <node>-ifgrp A1A-mode multimode_LACP-Distr-Function 포트
   - NET 포트 ifgrp add-port-node <node>-ifgrp <ifgrp>-ports <node>:e8a, <node>:e8b
3. 각 컨트롤러의 ifgrp에서 관리 VLAN 포트를 구성합니다
   - NET 포트 vlan create-node AFF-a90-01-port A1A-vlan-id 31
   - NET 포트 vlan create-node AFF-a90-02-port A1A-vlan-id 31
   - NET 포트 vlan create-node AFF-a90-03-port A1A-vlan-id 31
   - NET 포트 vlan create-node AFF-a90-04-port A1A-vlan-id 31
4. 브로드캐스트 도메인을 생성합니다
   - broadcast-domain create-broadcast-domain vlan21-mtu 9000-ports AFF AFF-a90-01:e6a, AFF AFF AFF-a90-01:e11a, AFF AFF-a90-02:e6a, AFF-a90-02:e6a
   - broadcast-domain create-broadcast-domain vlan22-mtu 9000-ports aaaaaaaa90 AFF-01:e6b, AFF AFF

AFF-a90-01:e11b, AFF AFF-a90-02:e6b, AFF-a90-03:e6b: e6b: e6b: e6b: e6b: e6b: e6b

- Broadcast-domain create-broadcast-domain vlan31-mtu 9000-ports AFF-a90-01:A1A-31, AFF-a90-02:A1A-31, AFF-a90-03:A1A-31, AFF-a90-04:A1A-31

5. 관리 SVM * 을 생성합니다
6. 관리 SVM 구성
   - LIF를 생성합니다
     - net int create-vserver basepod-mgmt-lif vlan31-01-home-node AFF-a90-01-home-port a1A-31-address 192.168.31.X-넷마스크 255.255.0
   - FlexGroup 볼륨 생성 -
     - vol create-vserver basepod-mgmt-volume home-size 10T-auto-provision-as FlexGroup-jection-path/home
     - vol 생성 - vserver basepod-mgmt-volume cm-size 10T-auto-provision-as FlexGroup-jection-path/cm
   - 엑스포트 정책을 생성합니다
     - export-policy rule create-vserver basepod-mgmt-policy default-client-match 192.168.31.0/24-rorule sys-rwrule sys-superuser sys
7. 데이터 SVM * 을 생성합니다
8. 데이터 SVM 구성
   - RDMA 지원을 위해 SVM을 구성합니다
     - vserver modify -vserver basePOD -data-RDMA가 활성화되었습니다
   - LIF 생성
     - net int create-vserver basepod-data-lif C1-6a-lif1-home-node AFF-a90-01-home-port e6a-address 100.127.102.101-netmask 255.255.0
     - net int create-vserver basepod-data-lif C1-6a-lif2-home-node AFF-a90-01-home-port e6a-address 100.127.102.102-netmask 255.255.0
     - net int create-vserver basepod-data-lif C1-6b-lif1-home-node AFF-a90-01-home-port e6b-address 100.127.202.101-netmask 255.255.0
     - net int create-vserver basepod-data-lif C1-6b-lif2-home-node AFF-a90-01-home-port e6b-address 100.127.202.102-netmask 255.255.0
     - net int create-vserver basepod-data-lif C1-11a-lif1-home-node AFF-a90-01-home-port e11a-address 100.127.102.103-netmask 255.255.0
     - net int create-vserver basepod-data-lif C1-11a-lif2-home-node AFF-a90-01-home-port e11a-address 100.127.102.104-netmask 255.255.0
     - net int create-vserver basepod-data-lif C1-11b-lif1-home-node AFF-a90-01-home-port e11b-address 100.127.202.103-netmask 255.255.0
     - net int create-vserver basepod-data-lif C1-11b-lif2-home-node AFF-a90-01-home-port e11b-address 100.127.202.104-netmask 255.255.0
     - net int create-vserver basepod-data-lif c2-6a-lif1-home-node AFF-a90-02-home-port e6a-address 100.127.102.105-netmask 255.255.0
     - net int create-vserver basepod-data-lif c2-6a-lif2-home-node AFF-a90-02-home-port e6a-address 100.127.102.106-netmask 255.255.0
     - net int create-vserver basepod-data-lif c2-6b-lif1-home-node AFF-a90-02-home-port e6b-address

100.127.202.105-netmask 255.255.0

- net int create-vserver basepod-data-lif c2-6b-lif2-home-node AFF-a90-02-home-port e6b-address 100.127.202.106-netmask 255.255.0
- net int create-vserver basepod-data-lif c2-11a-lif1-home-node AFF-a90-02-home-port e11a-address 100.127.102.107-netmask 255.255.0
- net int create-vserver basepod-data-lif c2-11a-lif2-home-node AFF-a90-02-home-port e11a-address 100.127.102.108-netmask 255.255.0
- net int create-vserver basepod-data-lif c2-11b-lif1-home-node AFF-a90-02-home-port e11b-address 100.127.202.107-netmask 255.255.0
- net int create-vserver basepod-data-lif c2-11b-lif2-home-node AFF-a90-02-home-port e11b-address 100.127.202.108-netmask 255.255.0

9. RDMA 액세스에 대해 LIF를 구성합니다
   - ONTAP 9.15.1을 통한 배포의 경우 물리적 정보에 대한 RoCE QoS 구성에는 ONTAP CLI에서 사용할 수 없는 운영 체제 수준 명령이 필요합니다. RoCE 지원을 위한 포트 구성에 대한 지원을 받으려면 NetApp 지원에 문의하십시오. RDMA 기반 NFS는 문제 없이 작동합니다
   - ONTAP 9.16.1부터 엔드 투 엔드 RoCE 지원을 위한 적절한 설정으로 물리적 인터페이스가 자동으로 구성됩니다.
   - net int modify -vserver basepod-data-lif * -rdma-protocols RoCE
10. 데이터 SVM에서 NFS 매개 변수를 구성합니다
    - nfs modify -vserver basepod -data-v4.1 enabled -v4.1-pNFS enabled -v4.1-trunking enabled -tcp-max -transfer-size 262144
11. FlexGroup 볼륨 생성 -
    - vol create -vserver basePOD -데이터 볼륨 데이터 -크기 100T -자동 프로비저닝 -FlexGroup-접합 경로 /데이터로
12. 엑스포트 정책을 생성합니다
    - export-policy rule create-vserver basepod-data-policy default-client-match 100.127.101.0/24-rorule sys-rwrule sys-superuser sys
    - export-policy rule create-vserver basepod-data-policy default-client-match 100.127.201.0/24-rorule sys-rwrule sys-superuser sys
13. 루트 생성
    - Route add-vserver basepod_data-destination 100.127.0.0/17 - 게이트웨이 100.127.102.1 메트릭 20
    - Route add-vserver basepod_data-destination 100.127.0.0/17 - 게이트웨이 100.127.202.1 메트릭 30
    - route add-vserver basepod_data-destination 100.127.128.0/17-gateway 100.127.202.1 메트릭 20
    - Route add-vserver basepod_data-destination 100.127.128.0/17-gateway 100.127.102.1 메트릭 30

RoCE 스토리지 액세스를 위한 **DGX H100** 구성

이 섹션에서는 DGX H100 시스템 구성을 위한 주요 세부 정보를 설명합니다. 이러한 구성 항목 중 다수는 DGX 시스템에 배포된 운영 체제 이미지에 포함되어 있거나 부팅 시 Base Command Manager에 의해 구현될 수 있습니다. 이러한 이미지는 참조를 위해 여기에 나열되어 있습니다. BCM에서 노드 및 소프트웨어 이미지를 구성하는 방법에 대한 자세한 내용은 을 참조하십시오"BCM 설명서".

1. 추가 패키지를 설치합니다

- 이피툴입니다
- python3-PIP

2. Python 패키지를 설치합니다
   - 파라미코
   - 매트릭스 플로lib
3. 패키지 설치 후 dpkg을 다시 구성하십시오
   - dpkg --configure-A를 참조하십시오
4. MOFED를 설치합니다
5. 성능 튜닝을 위한 MST 값을 설정합니다
   - mstconfig -y -d <aa:00.0,29:00.0> set advanced_pci_settings = 1 NUM_OF_VFS = 0 MAX_ACC_OUT_READ = 44
6. 설정을 수정한 후 어댑터를 재설정합니다
   - mlxfwreset -d <aa:00.0,29:00.0>-y 재설정
7. PCI 장치에서 MaxReadReq를 설정합니다
   - setpci -s <aa:00.0,29:00.0> 68.W = 5957
8. RX 및 TX 링 버퍼 크기를 설정합니다
   - ethtool-G <enp170s0f0np0,enp41s0f0np0> Rx 8192 TX 8192
9. mlnx_qos를 사용하여 PFC 및 DSCP를 설정합니다
   - mlnx_qos-i <enp170s0f0np0,enp41s0f0np0>--pfc 0,0,0,1,0,0,0—trust=dscp—cable_len=3
10. 네트워크 포트에서 RoCE 트래픽에 대해 ToS를 설정합니다
    - echo 106>/sys/class/InfiniBand/<mlx5_7,mlx5_1>/tc/1/traffic_class
11. 각 스토리지 NIC를 적절한 서브넷에 있는 IP 주소로 구성합니다
    - 스토리지 NIC 1의 경우 100.127.101.0/24
    - 스토리지 NIC 2의 경우 100.127.201.0/24
12. LACP 결합을 위한 대역 내 네트워크 포트 구성(enp170s0f1np1, enp41s0f1np1)
13. 각 스토리지 서브넷에 대한 운영 및 보조 경로에 대한 정적 경로를 구성합니다
    - Route add – net 100.127.0.0/17 GW 100.127.101.1 metric 20
    - Route add – net 100.127.0.0/17 GW 100.127.201.1 metric 30
    - Route add – net 100.127.128.0 / 17 GW 100.127.201.1 metric 20
    - Route add – net 100.127.128.0 / 17 GW 100.127.101.1 metric 30
14. 마운트/홈 볼륨
    - mount-o vers=3, nconnect=16, rsize=262144, wsize=262144 192.168.31.X:/home/home
15. 마운트/데이터 볼륨
    - 데이터 볼륨을 마운트할 때 다음과 같은 마운트 옵션이 사용됨
      - Servers = 4.1#에서는 여러 스토리지 노드에 대한 병렬 액세스를 위해 pNFS를 사용합니다

- PROTO=RDMA#은 전송 프로토콜을 기본 TCP 대신 RDMA로 설정합니다
- MAX_CONNECT = 16#은(는) NFS 세션 트렁킹을 활성화하여 스토리지 포트 대역폭을 집계합니다
- Write=eager#은 버퍼링된 쓰기의 쓰기 성능을 향상시킵니다
- rsize = 262144, wsize = 262144# 입출력 전송 크기를 256K로 설정합니다

NVA-1173 NetApp AIPod 및 NVIDIA DGX 시스템 - 솔루션 검증 및 사이징 지침

이 섹션에서는 NVIDIA DGX 시스템 기반 NetApp AIPod에 대한 솔루션 검증 및 사이징 지침에 대해 중점적으로 설명합니다.

솔루션 검증

이 솔루션의 스토리지 구성은 오픈 소스 툴 FIO를 사용하는 일련의 가상 워크로드를 사용하여 검증되었습니다. 이러한 테스트에는 딥 러닝 교육 작업을 수행하는 DGX 시스템에서 생성된 스토리지 워크로드를 시뮬레이션하기 위한 읽기 및 쓰기 I/O 패턴이 포함됩니다. 스토리지 구성은 FIO 워크로드를 동시에 실행하는 2소켓 CPU 서버 클러스터를 사용하여 DGX 시스템 클러스터를 시뮬레이션하여 검증되었습니다. 각 클라이언트는 앞서 설명한 것과 동일한 네트워크 구성으로 구성되었으며 다음 세부 정보가 추가되었습니다.

이 검증에는 다음과 같은 마운트 옵션이 사용되었습니다.

vers = 4.1	여러 스토리지 노드에 대한 병렬 액세스에서 pNFS를 사용합니다
PROTO = RDMA	전송 프로토콜을 기본 TCP 대신 RDMA로 설정합니다
포트 = 20049	RDMA NFS 서비스에 대한 올바른 포트를 지정합니다
최대_연결 = 16	스토리지 포트 대역폭을 집계하기 위해 NFS 세션 트렁킹을 활성화합니다
쓰기 = 열망	버퍼링된 쓰기의 쓰기 성능을 개선합니다
rsize = 262144, wsize = 262144	입출력 전송 크기를 256K로 설정합니다

또한 클라이언트가 NFS max_session_slot 값 1024로 구성되었습니다. 이 솔루션은 NFS over RDMA를 사용하여 테스트되었으므로 스토리지 네트워크 포트는 액티브/패시브 결합으로 구성되었습니다. 이 검증에 사용된 연결 매개 변수는 다음과 같습니다.

모드 = active-backup	본드를 액티브/패시브 모드로 설정합니다
운영 = <interface name>	모든 클라이언트의 기본 인터페이스가 스위치 전체에 분산되었습니다
MII-MONITOR-INTERVAL = 100	100ms의 모니터링 간격을 지정합니다
장애 조치 - Mac - 정책 = 활성	활성 링크의 MAC 주소가 본드의 MAC 주소임을 지정합니다. 이는 연결된 인터페이스에서 RDMA가 올바르게 작동하는 데 필요합니다.

스토리지 시스템은 각 HA 쌍에 24개의 1.9TB NVMe 디스크 드라이브로 구성된 NS224 디스크 쉘프 2개가 장착된 A900 HA 쌍(컨트롤러 4개)으로 설명한 대로 구성되었습니다. 아키텍처 섹션에서 설명한 것처럼, 모든 컨트롤러의 스토리지 용량은 FlexGroup 볼륨을 통해 결합되었으며 모든 클라이언트의 데이터가 클러스터의 모든 컨트롤러에 분산되었습니다.

스토리지 시스템의 사이징 지침

NetApp은 DGX BasePOD 인증을 성공적으로 완료했으며, 테스트 결과 A90 HA 쌍 2개는 16개의 DGX H100 시스템 클러스터를 손쉽게 지원할 수 있습니다. 스토리지 성능 요구사항이 더 높은 대규모 구축의 경우 단일 클러스터에서 최대 12개의 HA 쌍(24개 노드)까지 AFF 시스템을 NetApp ONTAP 클러스터에 추가할 수 있습니다. 이 솔루션에 설명된 FlexGroup 기술을 사용하여 24노드 클러스터는 단일 네임스페이스에서 40PB 이상의 처리량과 최대 300GBps 처리량을 제공할 수 있습니다. AFF A400, A250 및 C800 같은 다른 NetApp 스토리지 시스템은 낮은 비용으로 소규모 구축을 위한 낮은 성능 및/또는 더 높은 용량 옵션을 제공합니다. ONTAP 9에서 혼합 모델 클러스터가 지원되므로, 고객은 초기 설치 공간을 작게 시작한 후 용량 및 성능 요구사항이 증가함에 따라 클러스터에 규모가 더 큰 스토리지 시스템을 더 추가할 수 있습니다. 아래 표는 각 AFF 모델에서 지원되는 A100 및 H100 GPU의 수를 대략적으로 보여줍니다.

_NetApp 스토리지 시스템 사이징 지침 _

		Throughput[2]	Raw capacity *(typical / max)*	Connectivity	# NVIDIA A100 GPUs supported[3]	# NVIDIA H100 GPUs supported[4]
NetApp® AFF A900	**1 HA pair[1]**	28GB/s	182TB / 14.7PB	**100 GbE**	1 - 64	1-32
	12 HA pairs	336GB/s	2.1PB / **176.4PB**		**768**	**384**
AFF A800	**1 HA pair**	25GB/s	368TB / 3.6PB	**100 GbE**	1 - 64	1-32
	12 HA pairs	300GB/s	4.4PB / **43.2PB**		**768**	**384**
AFF C800	**1 HA pair**	21GB/s	368TB / 3.6PB	**100 GbE**	1-48	1-24
	12 HA pairs	252GB/s	4.4PB / **43.2PB**		**576**	**288**
AFF A400	**1 HA pair**	11GB/s	182TB / 14.7PB	**40/100 GbE**	1 - 32	1-16
	12 HA pairs	132GB/s	2.1PB / **176.4PB**		**384**	**192**
AFF C400	**1 HA pair**	8GB/s	182TB / 14.7PB	**40/100 GbE**	1 - 16	1-8
	12 HA pairs	128GB/s	2.1PB / **176.4PB**		**192**	**96**
AFF A250	**1 HA pair**	7.4GB/s	91.2TB / 4.4PB	**25 GbE 40/100GbE**	1 - 16	1-8
	4 HA pairs	29.6GB/s	364.8TB / 17.6PB		64	32
AFF C250	**1 HA pair**	5 GB/s	91.2TB / 4.4PB	**25 GbE 40/100GbE**	1-8	1-4
	4 HA pairs	20 GB/s	364.8TB / 17.6PB		32	8

1 – 1 AFF = 1 HA pair = 2 Nodes. 12 HA pairs = 24 nodes
2 – 100% sequential read
3 – Based on workload testing in NVA-1153
4 – Based on BasePOD validation test results