

IBM Spectrum Storage for AI with NVIDIA DGX Systems

AI 워크로드를 위한 검증된 인프라 솔루션



목차

서론.....	3
IBM Spectrum Storage for AI with NVIDIA DGX Systems의 구성요소 - 컨버지드 솔루션	4
NVIDIA DGX System 제품군.....	4
NVIDIA DGX-1 System.....	4
NVIDIA DGX-2 System.....	5
Mellanox InfiniBand Network	5
IBM Spectrum Scale 기반의 스토리지 시스템.....	6
IBM Elastic Storage System 3000.....	6
IBM Spectrum Scale 및 IBM Spectrum Scale RAID	7
GPU에 발맞춰 확장	8
IBM Spectrum Storage for AI with NVIDIA DGX Systems - DGX-1을 포함한 레퍼런스 아키텍처	12
IBM Spectrum Storage for AI with NVIDIA DGX Systems - DGX-2를 포함한 레퍼런스 아키텍처	13
스토리지 및 네트워크 성능 테스트	14
DGX-1 System 사용 시 시스템 처리량 결과.....	15
DGX-2 System 사용 시 시스템 처리량 결과.....	18
트레이닝 결과 - 단일 DGX-1 시스템	20
트레이닝 결과 - 멀티 DGX-1 시스템	21
추론(Inference) 결과 - 멀티 DGX-1 시스템.....	22
트레이닝 결과 - 멀티 DGX-2 시스템	23
추론(Inference) 결과 - 멀티 DGX-2 시스템.....	24
엔터프라이즈 데이터 파이프라인.....	25
결론.....	28
추가 자료.....	28
감사의 말.....	28

서론

인공지능(AI)과 딥러닝(DL)은 의료부터 자율주행차, 농업에 이르기까지 다양한 산업 분야에서 빠른 혁신을 위해 동력을 공급하는 엔진입니다. IBM은 2020년까지 전 세계의 디지털 데이터 용량이 44 제타바이트를 초과할 것으로 예측합니다^[1]. 의사결정과 실행을 하는 데 있어서 데이터의 가치를 인식한 조직들은 지금 그리고 앞으로 생성될 방대한 양의 새로운 데이터를 신속하게 수집하고 정확하게 해석하며 그러한 데이터로부터 빠르게 핵심적인 데이터 인사이트를 제공할 수 있는 딥러닝(DL) 시스템으로 관심을 돌리고 있습니다.

기업들은 AI 연구와 혁신에 대한 투자를 늘리고 있으며 지난 10년간 관련 특허는 30% 이상, 학술 논문은 13% 증가했습니다^[2]. 논란의 여지는 있지만, 미래에도 살아남아 번창할 수 있는 기업은 데이터를 기반으로 한 기업뿐일 것입니다.

확장 가능한 고성능 딥러닝(DL) 시스템은 조직 내에서 딥러닝(DL)에 대한 증가하는 수요를 충족하면서 데이터 수집, 데이터 트레이닝 및 검증 그리고 추론과 분류를 훌륭하게 수행할 수 있어야 합니다. 신경망 규모를 늘리고 모델 트레이닝 단계에 사용되는 데이터의 양과 질을 향상하면 딥러닝(DL) 알고리즘의 정확성을 향상할 수 있습니다. 그러나 이런 방법으로 정확성을 향상할 경우, 필연적으로 컴퓨팅 관련 복잡성과 딥러닝(DL) 컴퓨팅, 스토리지, 네트워크 리소스에 대한 수요가 증가합니다.

NVIDIA는 대규모 프로세서 코어 수와 병렬 아키텍처를 통해 현대적인 GPU의 탁월한 성능을 활용하며 AI 컴퓨팅의 혁명을 이끌어왔습니다. 이 병렬 아키텍처는 딥러닝(DL)에 핵심인 대규모 병렬 처리에 특히 적합하며 기존 CPU 기반 아키텍처의 한계를 뛰어넘습니다. IBM은 이미지, 오디오, 비디오, 텍스트 또는 시계열 데이터 등 다양한 데이터 유형을 대규모로 병렬 처리하도록 지원하는 입증된 확장성과 함께, Latency가 낮고 경제적인 업계 최고 수준의 고성능 올플래시 스토리지와 소프트웨어 정의 클러스터링을 제공합니다.

NVIDIA와 함께 IBM은 딥러닝(DL) 트레이닝 및 추론의 가속화를 위해 플래시에서 GPU까지 End-to-End 병렬 처리를 제공하고 개별적으로 확장 가능하면서도 통합된 형태의 컴퓨팅 및 스토리지 솔루션을 탄생시켰습니다. 본 문서는 IBM Spectrum Storage for AI with NVIDIA® DGX™ Systems가 AI를 간소화하고 가속화하는 방식에 대해 알아보려고 하는 기업 경영진, 솔루션 아키텍트 및 기타 독자를 위해 작성되었습니다. 이 확장 가능한 인프라 솔루션은 NVIDIA DGX-1™ 시스템 및 NVIDIA DGX-2™ 시스템과 IBM® Spectrum Scale™ 파일 스토리지 소프트웨어를 통합한 것이며, 이 IBM® Spectrum Scale™ 파일 스토리지 소프트웨어는 새로운 IBM Elastic Storage System(ESS 3000)을 포함한 IBM Elastic Storage Server(ESS) 스토리지 시스템 제품군을 지원합니다.

또한 본 문서에서는 synthetic 워크로드와 ImageNet 데이터를 모두 사용하여 DGX-1 시스템을 1대에서 9대로, 또는 DGX-2 시스템을 1대에서 3대로 확장함에 따라 어떻게 이 솔루션이 선형적인 스토리지 성능을 구현하는지에 대해서도 설명합니다.

¹ <https://www.ibm.com/blogs/systems/ibm-and-nvidia-further-collaboration-to-advance-open-source-gpu-acceleration/>

² <https://web.luxresearchinc.com/hubfs/18%20for%202018/Lux%20Research%2018%20for%202018.pdf>

IBM Spectrum Storage for AI with NVIDIA DGX Systems의 구성요소 - 컨버지드 솔루션

NVIDIA DGX System 제품군

슈퍼컴퓨터로 구성된 NVIDIA DGX 시스템 제품군은 AI 애플리케이션 용도로 특별히 제작된 완전 통합형 하드웨어 및 소프트웨어 솔루션입니다. NVIDIA DGX POD™는 레퍼런스 아키텍처에 포함된, 딥러닝(DL)을 위한 데이터센터 설계 베스트 프랙티스를 구현한 것입니다. 본 문서에서는 DGX-1 시스템과 DGX-2 시스템 모두를 다룹니다.

DGX-1 시스템과 DGX-2 시스템은 GPU로 가속 성능을 극대화하기 위해 NVIDIA가 설계한 최신 딥러닝(DL) 컨테이너의 최적화된 버전과 함께, 운영 체제(DGX 시스템 OS)와 NVIDIA GPU Cloud(NGC) 딥러닝 스택을 포함하는 DGX 시스템 소프트웨어에 의해 구동됩니다. 이 소프트웨어 스택은 단일 DGX 시스템(또는 다수의 DGX 시스템)에서 빠른 개발 및 배포, 그리고 DGX 시스템 플랫폼에서 애플리케이션에 대한 다중 GPU, 다중 시스템 스케일업을 용이하게 해주므로 시간을 절약하고 개발자의 수고를 덜 수 있습니다. DGX POD와 함께 제공되는 소프트웨어 구성요소로는 이 문서에서 설명하는 아키텍처의 관리에 활용할 수 있는 클러스터 관리, 라이브러리, 프레임워크, 워크로드 스케줄링 등이 있습니다. [NVIDIA GPU Cloud\(NGC\)를 방문하여](#) 자세한 내용을 알아보고 시작하세요.

NVIDIA DGX-1 System



그림 1: NVIDIA DGX-1 System

DGX-1 시스템(그림 1)에는 여덟 개의 NVIDIA Tesla™ V100 Tensor Core GPU가 통합되어 있으며, 이 GPU들이 NVIDIA NVLink™ 기술을 사용하는 하이브리드 큐브 메시 토폴로지를 구성합니다. 이 패브릭은 GPU를 더 많이 사용할 때 비선형적 성능을 유발하는 기존 아키텍처의 PCIe 기반 인터커넥트 병목 현상을 제거하는 동시에, 확장 가능한 다중 GPU 트레이닝을 지원하는 낮은 Latency의 고대역폭 GPU 간 통신을 제공합니다. DGX-1 시스템에는 네 개의 Mellanox VPI 카드가 포함되어 있어 다중 노드 클러스터링을 위한 EDR InfiniBand 또는 100GbE 네트워크 포트를 지원하고 고속 RDMA 기능을 통해 스토리지-DGX 시스템 간에 탁월한 데이터 속도를 제공합니다.

NVIDIA DGX-2 System

DGX-2 시스템(그림 2)에는 16개의 Tesla V100 Tensor Core GPU가 통합되어 있으며, 이 GPU들이 NVLink 기술을 사용하는 NVSwitch™ 패브릭 x 16GPU 인터커넥트 패브릭을 구성합니다. 이 낮은 Latency의 고대역폭 GPU-GPU 패브릭은 병목 현상과 중간의 GPU 홉을 없앴으로써 2petaFLOPS의 딥러닝(DL) 컴퓨팅 용량을 제공하며, 기존 아키텍처에서 볼 수 있는 PCIe 기반 인터커넥트의 병목 현상을 제거하면서 확장 가능한 다중 GPU 트레이닝을 지원합니다.

DGX-2 시스템에는 여덟 개의 Mellanox VPI 카드가 포함되어 있어, 다중 노드 클러스터링을 위한 EDR InfiniBand 또는 100GbE 네트워크 포트를 지원하고 고속 RDMA 기능을 통해 탁월한 스토리지-DGX 시스템 간 데이터 속도를 제공합니다.



그림 2: NVIDIA DGX-2 System

Mellanox InfiniBand Network

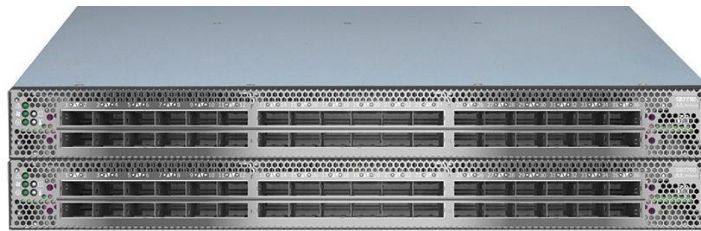


그림 3: Mellanox SB7700 시리즈

데이터 중심 조직이 직면한 현재의 과제는 기하급수적으로 증가하는 데이터를 짧은 시간 안에 수집하고 분석해야 하며, 실행 가능한 인사이트와 대응 방안을 실시간으로 확보해야 하는 경우가 많은 보다 복잡한 모델과 시뮬레이션을 지원해야 한다는 것입니다. 이와 같은 현재의 새로운 데이터 관련 과제의 요구사항을 충족하려면 효율적인 방식으로 데이터센터를 확장해야 합니다.

이 레퍼런스 아키텍처의 경우, IBM Spectrum Scale 기반의 NVMe 스토리지는 Mellanox EDR InfiniBand 네트워크를 통해 DGX-1 또는 DGX-2 시스템에 연결되어 단일 DGX 시스템 이상의 GPU 워크로드 및 데이터 세트 확장성을 가장 효율적으로 제공하는 동시에 DGX 시스템 사이에 노드 간 통신을 제공합니다.

가장 빠른 데이터 속도, 가장 낮은 Latency, 지능적 오프로드 기능을 제공하는 InfiniBand는 세계 최고 수준의 HPC 및 AI 슈퍼컴퓨터를 연결하기 위한 최상의 선택으로 리더십을 유지하고 있습니다. Mellanox는 RDMA, GPU 간 통신을 가속하는 GPUDirect® 기술, 머신러닝 알고리즘을 가속하는 SHARP™ 기술 등 최첨단 네트워크 컴퓨팅 기능을 제공하여, AI 워크로드에 사용되는 고성능 GPU 클러스터를 위한 완전한 End-to-End 인터커넥트 솔루션을 제공합니다.

IBM Spectrum Scale 기반의 스토리지 시스템

IBM과 NVIDIA는 IBM Spectrum Storage for AI의 NVMe(Non-Volatile Memory Express) 플래시 스토리지와 DGX-1 시스템 및 DGX-2 시스템을 함께 테스트하는 두 번의 벤치마크를 최근 실시했습니다. 각 벤치마크에는 IBM Elastic Storage Server(ESS) 스토리지 제품군 및 NVMe 올플래시 ESS 3000의 기반이 되는 선도적인 소프트웨어 정의 파일 스토리지인 IBM Spectrum Scale이 포함되었습니다. ESS는 사전 통합된 스토리지 하드웨어 제품군입니다. IBM Spectrum Storage for AI with NVIDIA DGX Systems 컨버지드 솔루션은 모든 유형의 ESS 올플래시 모델 또는 ESS 3000을 사용하여 구성이 가능합니다(그림 4).

IBM Elastic Storage System 3000



그림 4: IBM Elastic Storage System 3000

IBM Elastic Storage System 3000은 플래시 및 End-to-End NVMe의 성능과 IBM Spectrum Scale의 풍부한 기능이 결합된 강력한 2U 사이즈의 스토리지 시스템으로, 100Gb/s, IB 등 다양한 초고속 네트워크 연결 옵션을 제공합니다.

또한 ESS 3000은 DGX-2 시스템 테스트를 포함하여, 삼성 NVMe 드라이브와 같은 업계 표준 NVMe 플래시 드라이브 옵션으로도 테스트를 실시했습니다. 이러한 드라이브 옵션을 활용하여 NVMe 기반 IBM Spectrum Scale은 ESS 3000 유닛당 약 40GB/s의 대역폭, 세 개의 ESS 3000 유닛으로 구성된 클러스터에 대해 약 120GB/s의 지속 읽기(sustained read), 100마이크로초의 Latency를 구현하여 올플래시 성능과 확장성을 기준으로 시장을 선도하는 제품임을 입증하였습니다.

데이터 기반 멀티클라우드 스토리지 용량을 제공하는 ESS 3000은 IBM Spectrum Scale storage™의 소프트웨어 정의 기능과 긴밀하게 통합되어 있으므로 AI 데이터 파이프라인에 쉽게 연결할 수 있습니다.

ESS 3000은 다양한 연결 옵션과 기타 산업 표준 NVMe 플래시 드라이브 옵션을 지원합니다. 또한 우수한 설계를 바탕으로 모든 범위의 AI 워크로드와 비즈니스 사례를 지원합니다.

IBM Spectrum Scale 및 IBM Spectrum Scale RAID

IBM Spectrum Scale은 업계를 선도하는 고성능 병렬 파일 시스템 소프트웨어입니다. 이 제품은 Summit과 Sierra라는 세계 1, 2위의 가장 빠른 슈퍼컴퓨터와 그 밖의 다른 슈퍼컴퓨터에도 활용되고 있습니다. “Summit은 300억개 파일과 300억개 디렉토리의 용량을 제공하며 초당 260만 회 이상의 I/O 파일 작업 속도로 파일을 만들 수 있습니다. 이러한 속도를 활용하면 미국 의회도서관의 모든 책을 10초 안에 열 수 있습니다.” (<https://www.ibm.com/blogs/systems/fastest-storage-fastest-system-summit>)

Spectrum Scale이 제공하는 핵심 기능은 각 데이터 소스가 NFS, SMB, Object 또는 POSIX 인터페이스를 사용하여 데이터를 저장소에 추가할 수 있도록 단일 네임스페이스(또는 데이터 영역)를 제공하는 것입니다. 이 단일 데이터 영역을 통해 데이터 준비 톨은 불필요한 데이터의 이동이나 재배치 없이 기존 데이터에 액세스할 수 있습니다. 추론 애플리케이션과 마찬가지로 AI 트레이닝도 기존 데이터에 액세스할 수 있으며 모든 경우 복제가 필요하지 않고 산업 표준 인터페이스가 사용됩니다.

Spectrum Scale의 또 다른 주요 장점은 데이터가 하드 디스크 드라이브(HDD), 테이프, 클라우드 등 가장 비용 효율적인 스토리지로 자동으로 티어링(Tiering)된다는 점입니다. IBM은 전체 AI 파이프라인에 대한 지원을 제공합니다. 자세한 내용은 본 문서의 엔터프라이즈 데이터 파이프라인 섹션을 참조하세요.

IBM Spectrum Scale RAID는 IBM Spectrum Scale 내에서 스토리지 Erasure Code 기술을 소프트웨어로 구현한 것으로 정교한 데이터 배치 및 Error-Correction 알고리즘을 제공하여 높은 수준의 스토리지 성능, 가용성, 신뢰성을 제공합니다. 공유 파일 시스템은 IBM Spectrum Scale RAID로 정의된 NSD(Network Shared Disk) 상에서 생성됩니다. 이 파일 시스템은 구성 환경 내의 모든 컴퓨팅 노드가 동시에 액세스할 수 있으므로 AI와 같은 현대적인 스케일아웃 애플리케이션의 용량 및 성능 요구사항을 효율적으로 충족합니다.

IBM Spectrum Storage for AI의 일부로 제공되는 IBM Spectrum Scale RAID는 GPU 프로세서를 많이 사용하는 AI 워크로드를 위해, 높은 수준의 데이터 보호와 지원기능을 DGX POD와 같은 멀티 시스템 GPU 클러스터의 최고 수준의 성능 환경에서도 제공합니다.

GPU에 발맞춰 확장

IBM Spectrum Storage for AI with DGX Systems는 용량 및 성능 면에서 최고의 확장성을 제공합니다. 워크로드 수요에 따라 ESS 3000과 DGX-1 또는 DGX-2 시스템을 하나 또는 여러 대 사용하여 구축을 시작할 수 있으며, 어떤 경우에도 비용 효율적인 솔루션을 제공합니다.

그림 5는 DGX-1 시스템 및 DGX-2 시스템을 위한 DGX POD 랙에서 가능한 ESS 3000의 구성을 보여줍니다. 각 구성은 성능을 위해 최적화할 수 있으며 기존 프로덕션 워크플로우의 확장이나 딥러닝(DL) 애플리케이션의 추가를 위한 확장이 쉽게 가능합니다. 랙 공간과 전력 요구사항을 따져봤을 때, ESS 3000은 2U~6U의 크기와 낮은 전력 요구량을 제공하여 최대의 유연성을 달성할 수 있습니다.



그림 5: 스케일아웃 구성의 예

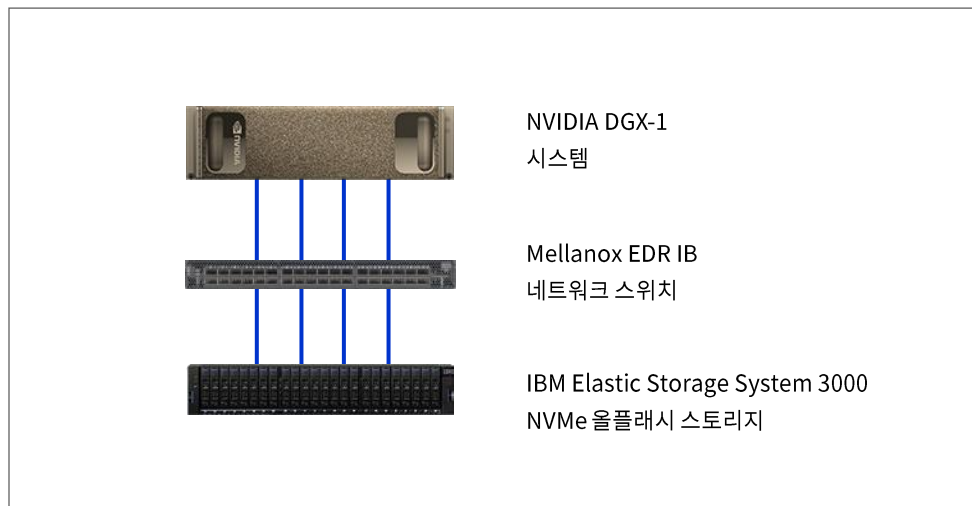


그림 6: DGX-1 시스템 1대 + ESS 3000 1대

그림 6은 1대의 DGX-1 시스템과 1대의 ESS 3000으로 이루어진 구성을 보여줍니다.

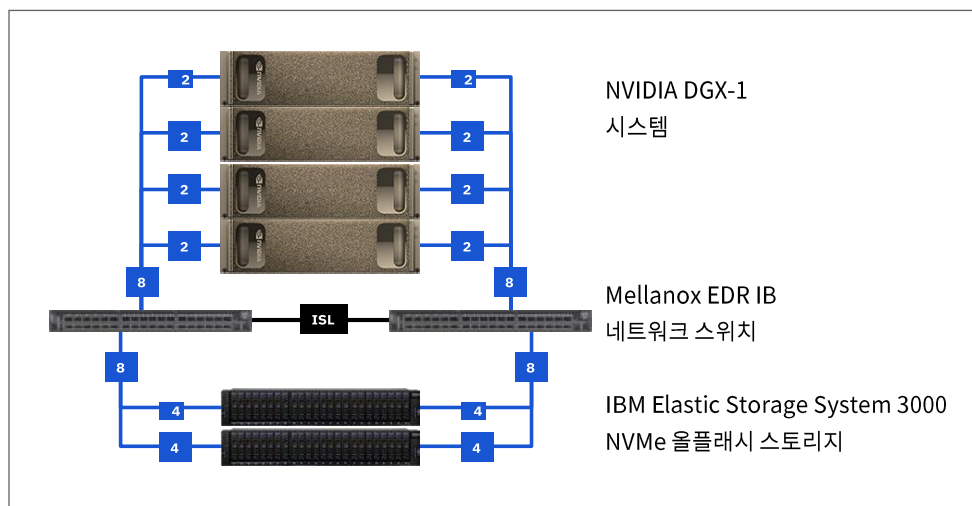


그림 7: 4대의 DGX-1 시스템과 2대의 ESS 3000 유닛으로 구성된 IBM Spectrum Storage for AI with NVIDIA

그림 7은 2대의 ESS 3000 유닛과 4대의 DGX-1 시스템으로 구성된 예를 보여주고 있습니다. 그림에 제시된 대로 2개의 스위치를 사용하거나, 원하는 경우 1개의 스위치를 사용하는 다른 아키텍처 옵션도 가능합니다. 플랫(flat) 네트워크 토폴로지가 제공하는 수준을 넘어 기존의 IO 인프라 성능을 스케일업하거나 신규 인프라를 구성하는 경우, 팟 트리와 같은 다른 네트워크 토폴로지도 고려해야 합니다.

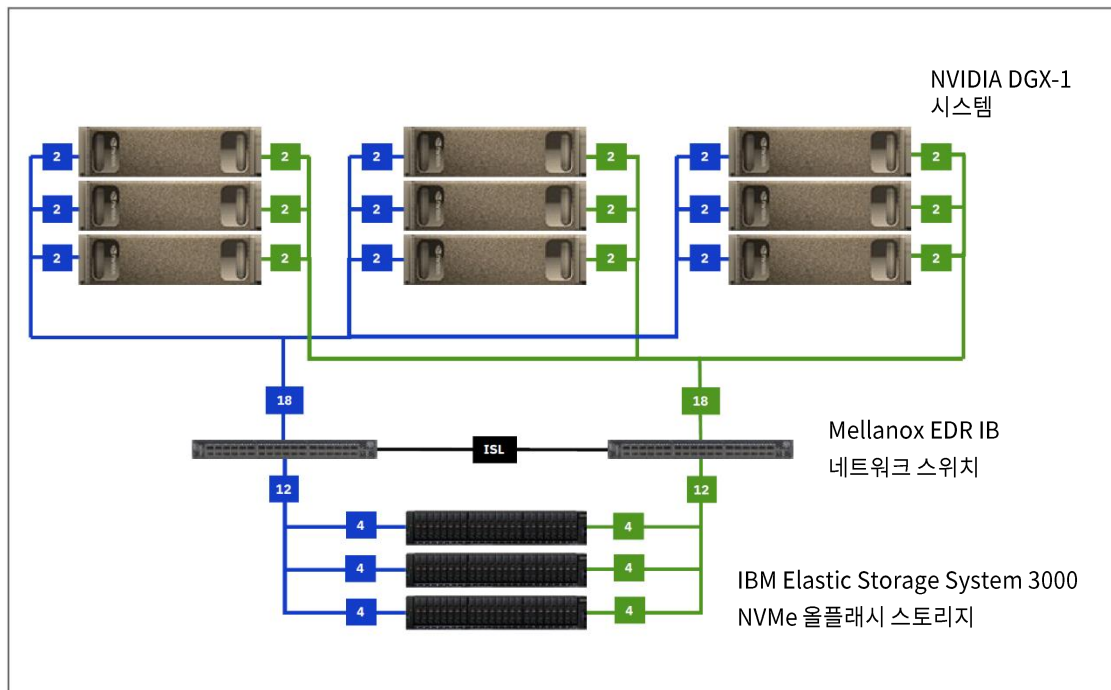


그림 8: 9대의 DGX-1 시스템 + 3대의 ESS 3000 유닛

그림 8은 3대의 ESS 3000 유닛을 사용하는 전체 DGX POD 구성을 보여줍니다. 이것이 IBM Elastic Storage System 3000 with DGX-1 Systems의 벤치마크에서 사용된 레퍼런스 아키텍처이기도 하지만, 본 문서에 제시된 DGX-1 시스템 모델의 초당 이미지 개수를 결과로 얻는 데에는 2대의 ESS 3000 유닛이면 충분했습니다.

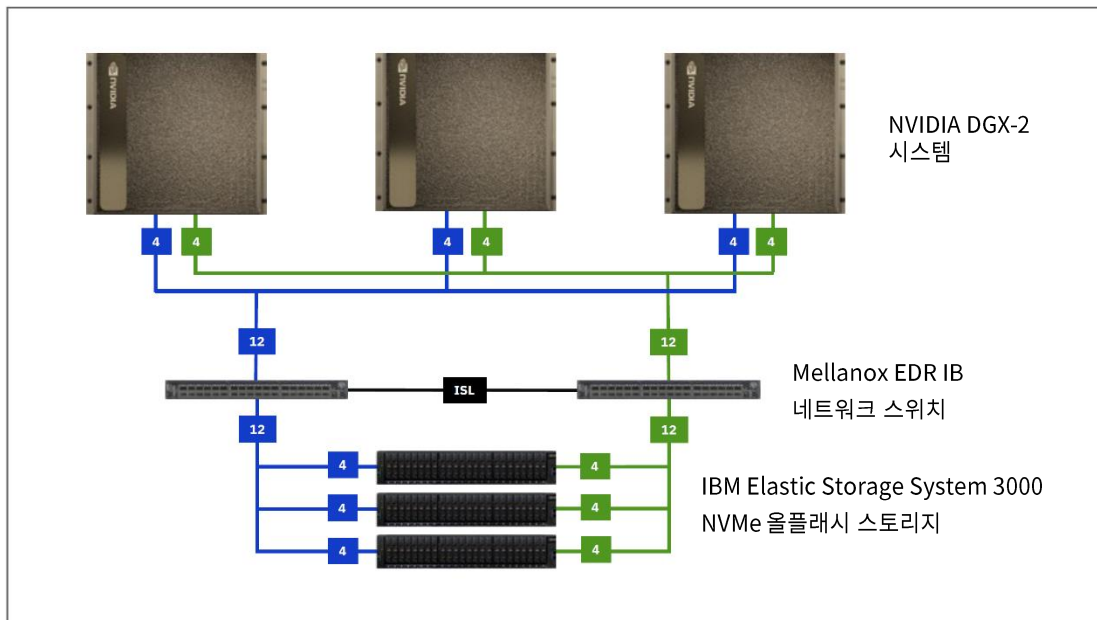


그림 9: 3대의 DGX-2 시스템 + 3대의 ESS 3000 유닛

그림 9는 3대의 DGX-2 시스템과 3대의 ESS 3000 유닛으로 된 구성을 보여줍니다. 이 구성이 IBM Elastic Storage System 3000 with NVIDIA DGX-2 Systems의 벤치마크에서 사용된 레퍼런스 아키텍처이기도 하지만, 본 문서에 제시된 DGX-2 시스템 모델의 초당 이미지 개수를 결과로 얻는 데에는 2대의 ESS 3000 유닛이면 충분했습니다.

IBM Spectrum Storage for AI with NVIDIA DGX Systems - DGX-1을 포함한 레퍼런스 아키텍처

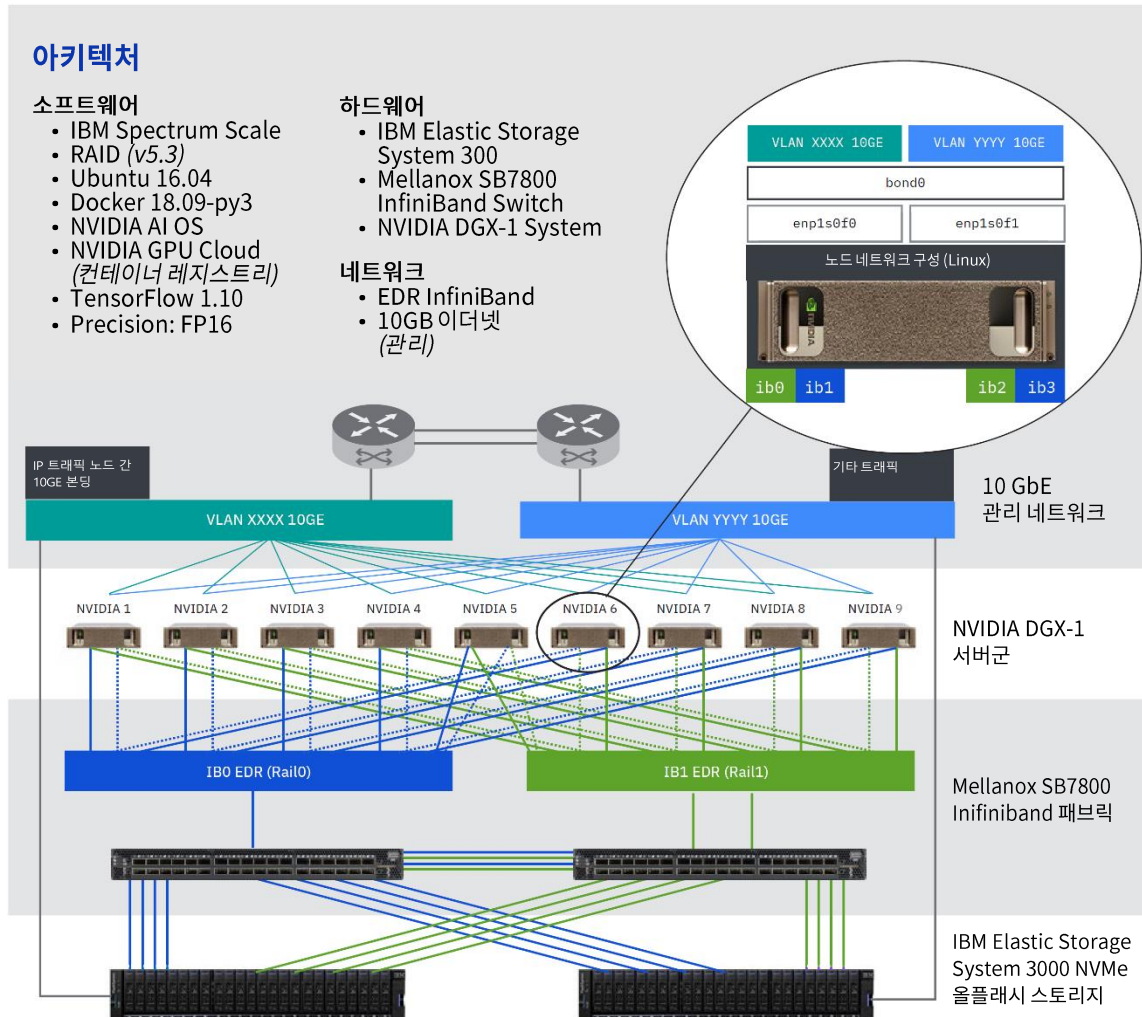


그림 10: DGX-1 시스템 솔루션 레퍼런스 아키텍처 다이어그램

그림 10은 DGX-1 시스템 테스트를 위해 배포된 IBM Spectrum Storage for AI with NVIDIA DGX Systems 솔루션 환경의 End-to-End 디버닝(DL) 레퍼런스 아키텍처를 보여줍니다. 이러한 첨단 DGX-1 시스템 테스트 환경에서 IBM Spectrum Scale RAID v5는 ESS 3000의 베이스 Linux OS상에 설치됩니다. IBM Spectrum Scale RAID 소프트웨어는 일반적으로 IBM Elastic Storage Server(ESS) 배포 환경을 위한 Spectrum Scale 소프트웨어 스택의 일부로 제공됩니다. 구성된 바와 같이 각 ESS 3000 유닛은 Spectrum Scale 클러스터 내에서 완전히 이중화된 NSD 시스템 한 쌍을 제공합니다. ESS 3000은 두 개의 Mellanox SB7800 EDR 스위치로의 여덟 개의 링크를 통해 EDR InfiniBand로 연결됩니다. 또한, DGX-1 시스템은 네 개의 링크를 통해 두 개의 InfiniBand 스위치에 연결됩니다. ESS 3000 및 DGX-1 시스템은 고속 EDR InfiniBand 패브릭 외에도 10GbE 관리 네트워크를 통해 이더넷 스위치에 연결됩니다.

IBM Spectrum Storage for AI with NVIDIA DGX Systems - DGX-2를 포함한 레퍼런스 아키텍처

아키텍처

소프트웨어

- IBM Spectrum Scale RAID (v5.3)
- Ubuntu 18.04
- Docker 19.04-py3
- NVIDIA AI OS
- NVIDIA GPU Cloud (컨테이너 레지스트리)
- TensorFlow 1.13.1
- Precision: FP16

하드웨어

- IBM Elastic Storage System 3000
- Mellanox SB7800 InfiniBand Switch
- NVIDIA DGX-2 System

네트워크

- EDR InfiniBand
- 10GB 이더넷 (관리)

그림 11은 DGX-2 시스템 테스트를 위해 배포된 IBM Spectrum Storage for AI with NVIDIA DGX Systems 솔루션 환경의 End-to-End 딥러닝(DL) 레퍼런스 아키텍처를 보여줍니다. 이 DGX-2 시스템 테스트 환경에서는 DGX-1 시스템에서 별도로 테스트된 것과 유사하게 IBM Spectrum Scale RAID v5가 ESS 3000의 베이스 Linux OS상에 설치됩니다. [IBM Spectrum Scale RAID](#) 소프트웨어는 일반적으로 [IBM Elastic Storage System\(ESS\)](#) 배포 환경을 위한 Spectrum Scale 소프트웨어 스택의 일부로 제공됩니다.

구성된 바와 같이 각 ESS 3000 유닛은 Spectrum Scale 클러스터 내에서 완전히 이중화된 NSD 시스템 한 쌍을 제공합니다. 각 ESS 3000 유닛은 두 개의 [Mellanox SB7800](#) 패브릭 인터커넥트 스위치로의 여덟 개의 링크를 통해 EDR InfiniBand로 연결됩니다. 또한, DGX-2 시스템은 여덟 개의 링크를 통해 EDR InfiniBand 스위치에 연결됩니다. ESS 3000 및 DGX-2 시스템은 고속 EDR InfiniBand 패브릭 외에도 10GbE 관리 네트워크를 통해 이더넷 스위치에 연결됩니다.

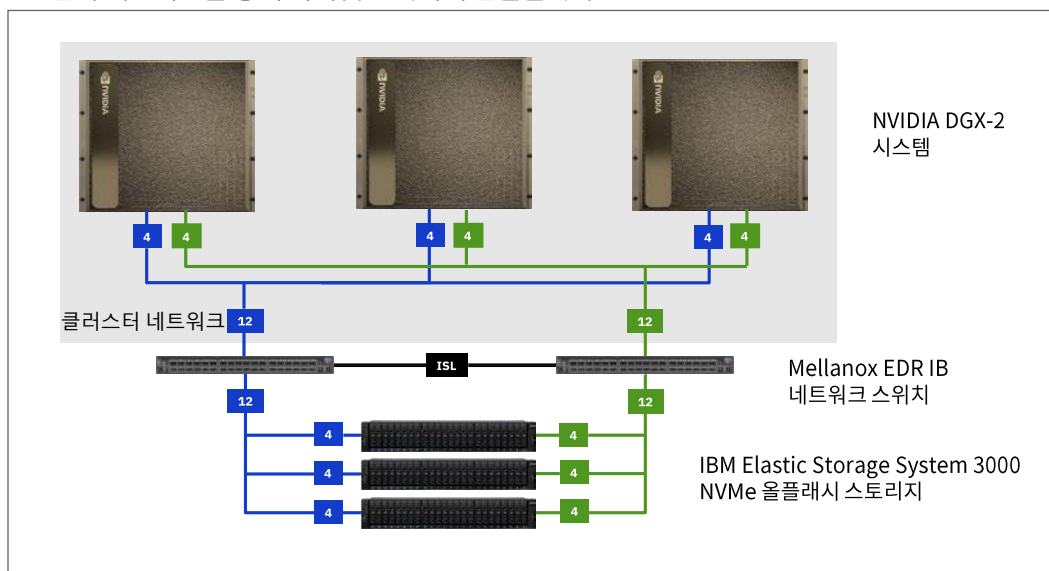


그림 11: DGX-2 시스템 솔루션 레퍼런스 아키텍처

스토리지 및 네트워크 성능 테스트

DGX-1 시스템 및 DGX-2 시스템 벤치마크 모두 딥러닝(DL) 개발 팀이 인프라를 효과적으로 계획하는 데 도움이 되도록 시스템 성능 역량을 제공하는 부분에 주로 중점을 두었습니다. 예를 들면, DGX-1 시스템 벤치마크 테스트는 IBM Elastic Storage System 3000을 사용하여 한 대의 DGX-1 시스템부터 전체 DGX POD까지 스케일업할 때 딥러닝(DL) 모델 트레이닝 성능과 더불어 딥러닝(DL) 모델 추론 성능 및 선형성 그리고 전체 처리 능력을 주로 살펴보았습니다.

마찬가지로 DGX-2 시스템 벤치마크 테스트는 ESS 3000을 사용하여 한 대의 DGX-2 시스템부터 3대의 DGX-2 시스템까지 스케일업할 때 딥러닝(DL) 모델 트레이닝 성능과 더불어 딥러닝(DL) 모델 추론 성능 및 선형성 그리고 전체 처리 능력을 주로 살펴보았습니다. 두 벤치마크의 경우 모두 IOR 및 fio와 같은 합성(Synthetic) 처리량 테스트 애플리케이션과, ResNet-50, ResNet-152, Inception-v3, 그리고 ILSVRC2012(ImageNet Large Scale Visual Recognition Challenge 2012) 데이터세트를 포함하는 다른 네트워크 등 여러 모델을 사용하는 딥러닝(DL) 프레임워크 TensorFlow로 성능을 테스트했습니다.

벤치마크에서는 NVIDIA NGC 컨테이너를 활용합니다. 각 AI 컨테이너에는 사전 통합된 GPU 가속화 소프트웨어 세트인 NGC Software Stack이 포함되어 있습니다. 이 스택은 선택된 애플리케이션 또는 프레임워크, NVIDIA CUDA Toolkit, NVIDIA 딥러닝(DL) 라이브러리 및 Linux OS를 포함하며 이들은 모두 추가 설정 없이 즉시 함께 작동하도록 테스트 및 조정을 거칩니다.

NVIDIA Deep Learning SDK는 개발자에게 Caffe2, Cognitive 툴킷, MXNet, PyTorch, TensorFlow 등의 딥러닝(DL) 프레임워크 개발을 위한 강력한 툴과 라이브러리를 제공합니다. 이 프레임워크들은 고성능 다중 GPU 가속화 트레이닝을 제공하기 위해 cuDNN 및 NCCL과 같은 GPU 가속 라이브러리를 사용합니다. 개발자, 연구자, 데이터 사이언티스트는 NVIDIA GPU에 맞게 성능을 조정하고 테스트를 거친 NVIDIA 최적화 딥러닝(DL) 프레임워크에 쉽게 액세스할 수 있습니다. 따라서 패키지와 종속성을 관리하거나 소스에서 딥러닝(DL) 프레임워크를 직접 빌드할 필요가 없습니다.

TensorFlow 벤치마크는 GitHub에서 TensorFlow가 제공하는 스크립트를 사용하여 수행했습니다. https://github.com/tensorflow/benchmarks/tree/master/scripts/tf_cnn_benchmarks를 참조하세요.

DGX-1 System 사용 시 시스템 처리량 결과

1~9대의 DGX-1 시스템과 1~3대의 ESS 3000 유닛을 사용할 경우 전체 시스템 처리량 성능은 각 벤치마크에서 모든 DGX-1 시스템 GPU를 포화 상태로 유지할 만큼 원활하게 증가하는 것으로 나타났습니다. 테스트 결과 ESS 3000 솔루션은 유닛이 한 대 ~ 여러 대 사용된 경우 약 40GB/s의 읽기 성능으로부터 이에 비례하여 선형적으로 확장되면서, 데이터 인프라의 잠재적 처리량을 극대화하는 것으로 나타났습니다.

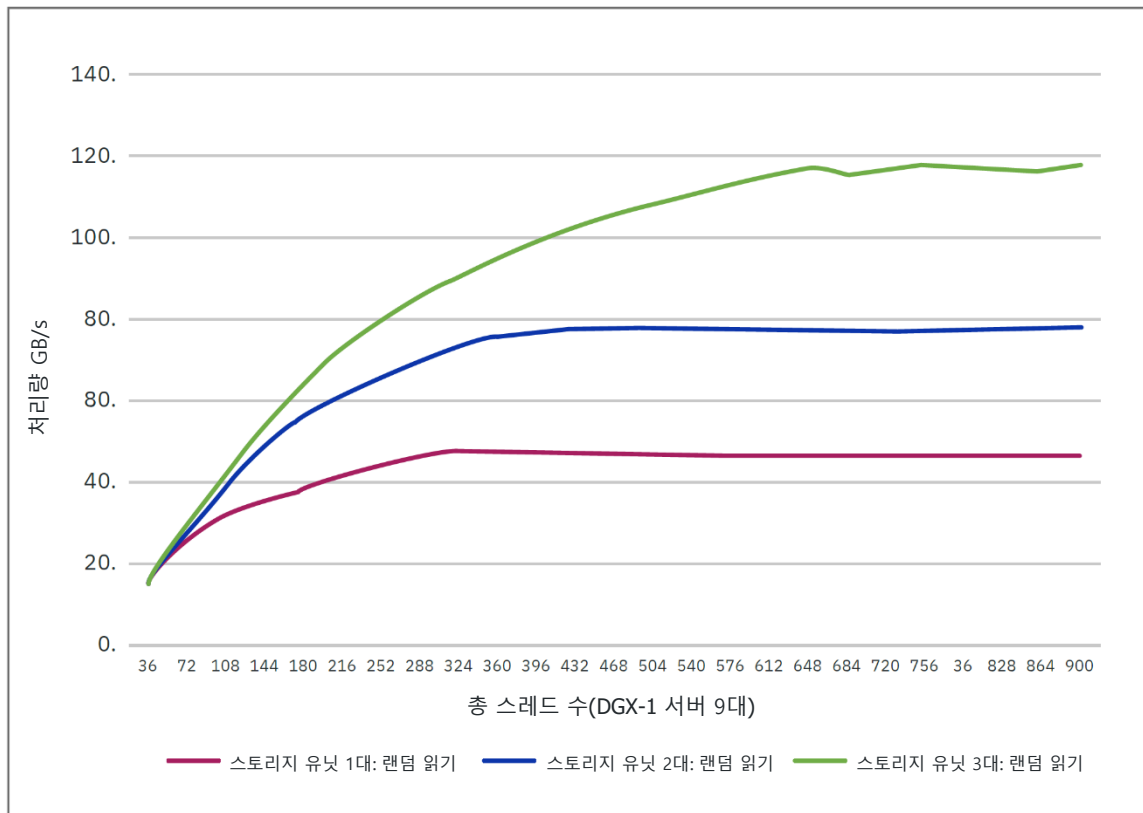


그림 12: 처리량의 선형 확장성 (fio 벤치마크, DGX-1 시스템 사용)

그림 13은 ESS 3000이 DGX 시스템에 효율적으로 데이터를 공급하여 딥러닝(DL) 워크로드에 대해 평균적으로 거의 100%의 전체 GPU 사용율을 달성하는 한편, 단일 DGX-1 시스템의 전체 대역폭 요청량이 Spectrum Scale NVMe 어플라이언스의 처리 성능에 부담이 되는 수준에는 미치지 못한다는 점을 보여줍니다. 두 개의 차트를 바탕으로 스토리지 처리 능력과 요구사항을 접목하면, 테스트한 단일 ESS 3000은 본 문서에서 제시한 모델 트레이닝 워크로드를 처리할 수 있는 것으로 나타났습니다.

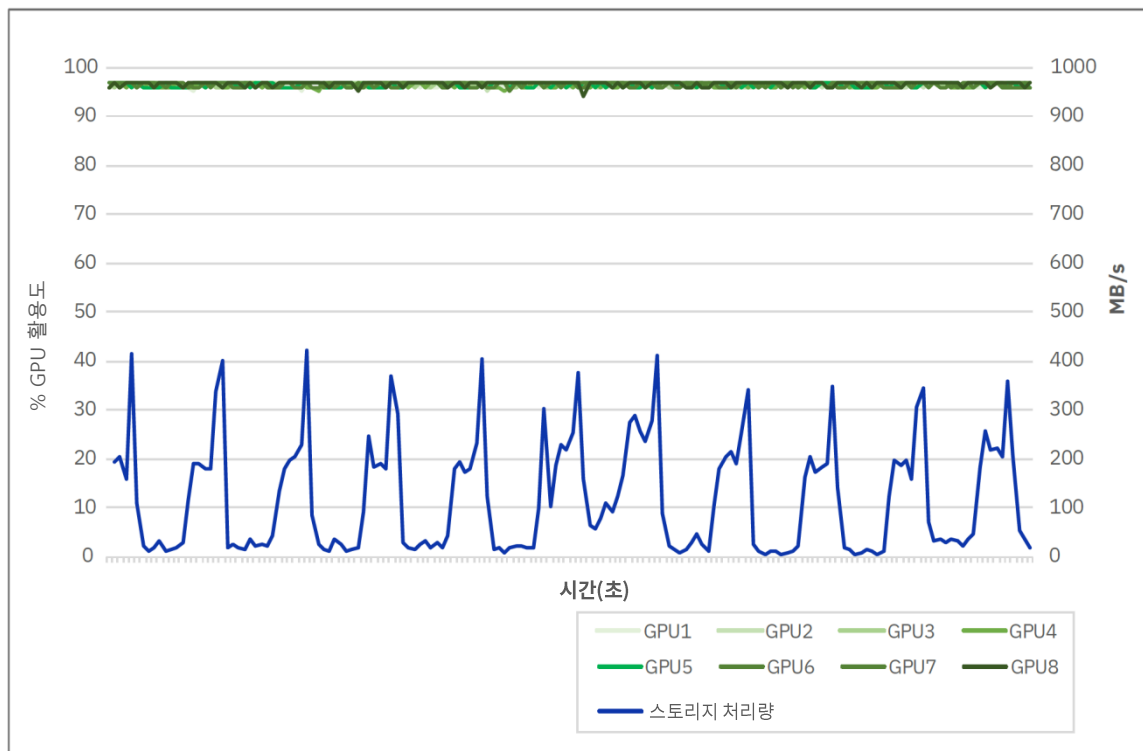


그림 13: 단일 DGX-1 시스템의 GPU 활용도 및 IO 대역폭 비교(ResNet-50)

ESS 3000 스토리지 솔루션의 유연성을 증명하게 위해 순차 및 랜덤 IO 액세스 패턴을 비교하는 추가 테스트를 실행했습니다(그림 14). 순차 읽기 성능을 랜덤 읽기 성능을 비교한 결과 프리페치(prefetch)의 이점이 어느 정도 나타났으나, 이는 작업 스레드 수가 증가할수록 약해지는 경향을 보였습니다. ESS 3000은 IO 유형에 무관하게 강력한 처리 능력을 보여주었습니다.

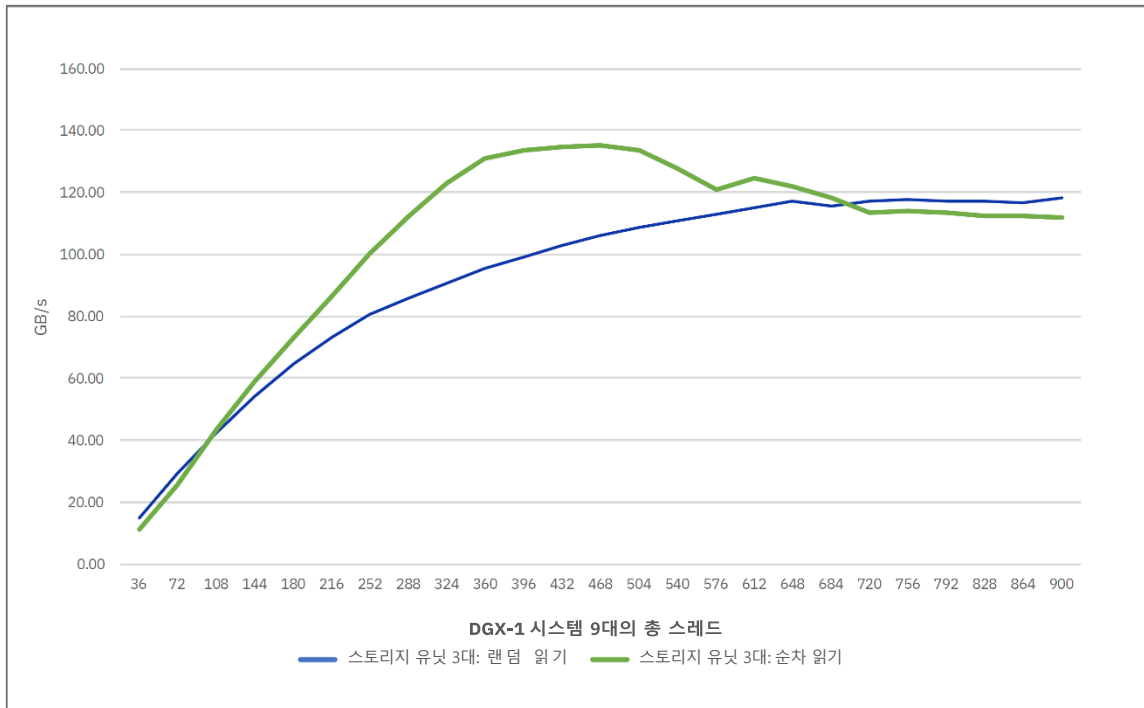


그림 14: 순차 vs 랜덤 읽기 처리량 비교 (fio 벤치마크, DGX-1 시스템 사용)

전반적으로, DGX-1 시스템을 사용하는 ESS 3000 처리량 결과는 ESS 3000이 1~9대의 DGX-1 시스템이 GPU를 최대한 활용하게 하면서도 우수한 성능을 보여줌을 입증합니다.

이러한 성능 덕분에 개발 팀은 필요할 때 컴퓨팅 리소스를 추가할 수 있으며, 워크로드 수요가 증가하더라도 스토리지 성능이 그것을 수용 가능하다는 확신을 가질 수 있습니다. 그리고 필요한 경우 ESS 3000 유닛을 업무중단 없이 AI 스토리지 클러스터에 원활하게 추가할 수 있습니다.

DGX-2 System 사용 시 시스템 처리량 결과

그림 16을 보면 단일 스토리지 유닛이 최대 처리량을 달성하도록 하려면 3대의 DGX-2 시스템이 필요합니다.

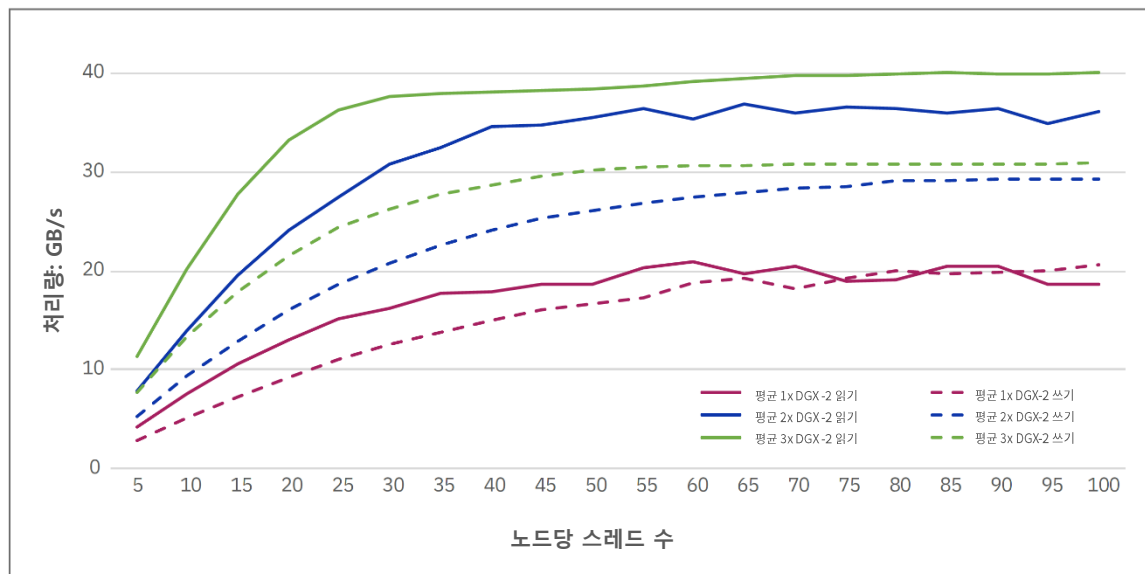


그림 16: 처리량 벤치마크 (단일 ESS 3000, 멀티 DGX-2 시스템)

이 차트는 1, 2, 3대의 ESS 3000 유닛을 사용한 경우 3대의 DGX-2 시스템의 처리 능력 합계를 보여줍니다.

이전에 DGX-1 시스템으로 테스트했을 때와 마찬가지로, 2대의 ESS 3000 유닛은 최대 80GB/s의 랜덤 읽기 파일시스템 대역폭을 제공할 수 있고 3대의 ESS 3000 유닛은 최대 120GB/s의 처리량을 제공할 수 있으므로, 3대의 DGX-2 시스템의 최대 시스템 대역폭 합계인 60GB/s에 도달하려면 ESS 3000 스토리지 유닛은 2대면 충분합니다. DGX-1 시스템으로 테스트한 것과 유사하게 1, 2, 3대의 ESS 3000 NVMe 스토리지의 선형적 성능확장에 대한 추가 테스트는 불가능했는데, 이는 테스트에 사용할 수 있는 DGX-2 시스템의 수가 제한되었기 때문입니다.

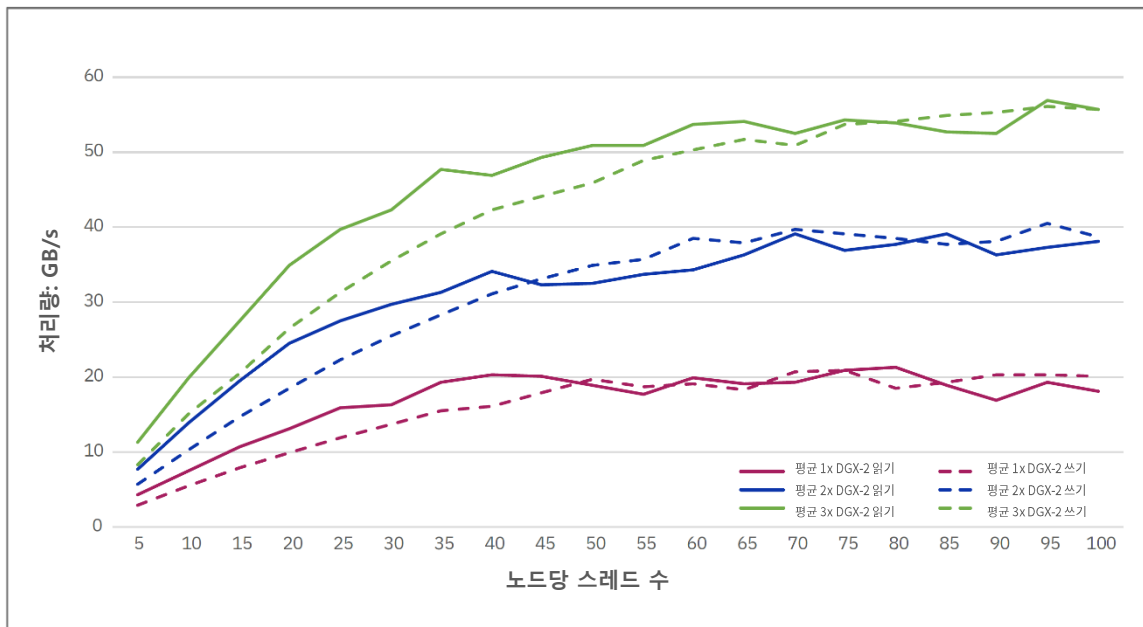


그림 17: 처리량 벤치마크 (3대의 ESS 3000, 3대의 DGX-2 시스템)

그 다음, DGX-2 시스템의 GPU를 활용하는 모델 트레이닝 및 추론 테스트를 살펴보았습니다. 그 결과, DGX-1 시스템 테스트 결과와 유사하게 NVMe 플래시 스토리지가 DGX-2 시스템에 효율적으로 데이터를 공급하여 딥러닝(DL) 워크로드에 대해 평균적으로 거의 100%에서 전체 GPU 활용도를 달성하는 한편 단일 DGX-2 시스템의 전체 대역폭 요청량이 ESS 3000 유닛의 처리 성능에 부담이 되는 수준에는 미치지 못한다는 것을 알 수 있었습니다.

또한, DGX-2 시스템을 사용하는 ESS 3000에 대한 처리량 결과는 다시 한번 ESS 3000이 1~3대의 DGX-2 시스템이 GPU를 충분히 활용하게 하면서도 우수한 성능을 보여줌을 입증하였습니다. 이러한 성능 덕분에 개발 팀은 필요할 때 컴퓨팅 리소스를 추가할 수 있으며, 워크로드 수요가 증가하더라도 스토리지 시스템 성능이 그것을 수용 가능하다는 확신을 가질 수 있습니다. 그리고 필요한 경우 ESS 3000 유닛을 업무 중단없이 AI 스토리지 클러스터에 원활하게 추가할 수 있습니다.

트레이닝 결과 - 단일 DGX-1 시스템

그림 18은 GPU당 별도의 컨테이너를 사용하는 단일 DGX-1 시스템에서 다양한 수의 GPU를 사용하는 AlexNet, ResNet-50, ResNet-152, Inception-v3, LeNet, Inception-v4 및 GoogLeNet 모델의 트레이닝 처리량을 초당 이미지 수로 보여주고, IBM Spectrum Scale 파일 시스템과 로컬 RAM 디스크의 트레이닝 실행을 비교하고 있습니다. 여기서 볼 수 있듯이 ESS 3000은 Spectrum Scale 클러스터에서 DGX-1 시스템 GPU에 효과적으로 데이터를 공급하여 DGX-1 시스템을 데이터로 완전히 포화된 상태로 유지함으로써 모든 모델에서 최대의 트레이닝 역량을 달성할 수 있습니다.

또한 GPU를 사용하는 딥러닝(DL) 모델에서 단일 DGX-1 시스템에 대한 모델 트레이닝 결과는 Spectrum Scale 파일 시스템 성능이 로컬 RAM 디스크 성능과 비교할 때 전반적으로 페널티가 최소한의 수준, 또는 전혀 발생하지 않음을 보여줍니다.

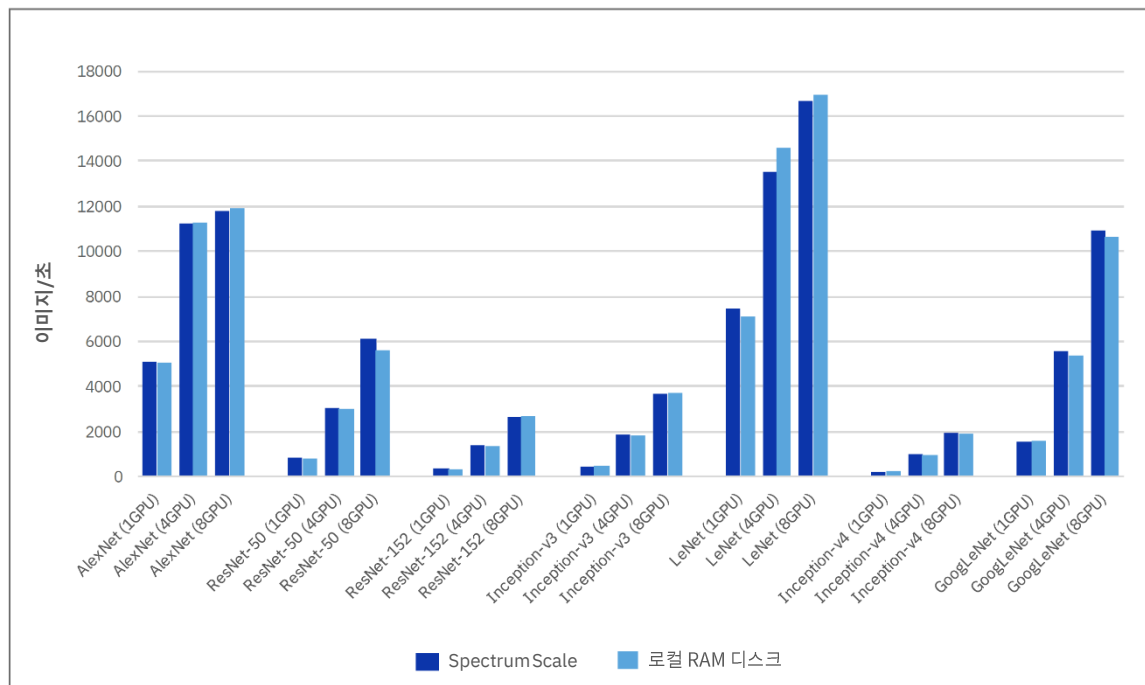


그림 18: 모델-GPU 성능 (단일 DGX-1 시스템, Spectrum Scale 파일 시스템과 RAM 디스크 성능 비교)

그림에서 볼 수 있듯이, GPU 개수가 증가함에 따라 일부 모델은 선형적으로 스케일업되지만, 나머지 모델은 ESS 3000 스토리지를 사용하건, 로컬 RAM 디스크를 사용하건 상관없이 일관되게 비선형적 스케일업 패턴을 보입니다. 이것이 뜻하는 바는 특정 케이스에서는, 확장성이 로컬/공유 스토리지에 상관없이 스토리지 IO에 의해 제한되는 것이 아니고, 컴퓨팅 인프라 자체 내의 딥러닝(DL) 모델 확장성 패턴에 의해 제한된다는 것입니다.

트레이닝 결과 - 멀티 DGX-1 시스템

개별 컨테이너를 사용하는 멀티 DGX-1 시스템의 경우, ESS 3000은 DGX-1 시스템 1대부터 9대(GPU 총 72개)까지 동시 실행환경에서 모든 DGX-1 시스템 GPU에 대해 최대 포화 상태까지 선형적으로 스케일업하는 것으로 나타났으며, 그림 19에 제시된 모델에 대해 ImageNet 데이터셋을 사용할 때 초당 이미지 개수를 집계한 성능 결과는 아래와 같습니다.

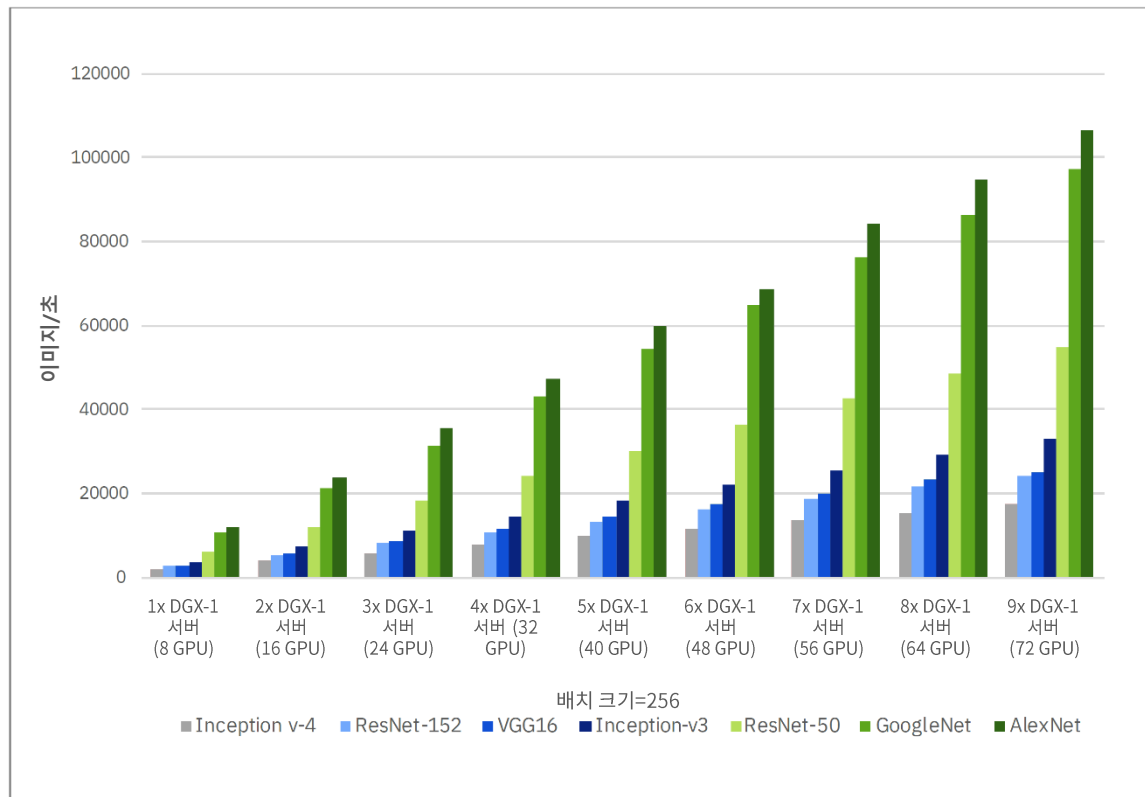


그림 19: 트레이닝 속도 (TensorFlow 모델, 멀티 DGX-1 시스템 환경)

멀티 DGX-1 시스템을 사용하는 경우 이미지 처리 속도(그림 19)는 DGX-1 시스템마다 8개의 GPU를 사용하는 Inception-v3, ResNet-50, GoogLeNet, AlexNet 모델을 적용한 ESS 3000에서 트레이닝 애플리케이션 성능의 확장성을 보여줍니다. ESS 3000 솔루션은 8개의 GPU를 사용하는 DGX-1 시스템을 1대부터 9대까지 추가하고 테스트된 72개 GPU 모두가 일관되게 GPU 완전 포화 상태를 보일 때까지 선형적으로 스케일업하는 것으로 나타났습니다.

추론(Inference) 결과 - 멀티 DGX-1 시스템

개별 컨테이너를 사용하는 멀티 DGX-1 시스템의 경우, ESS 3000은 DGX-1 시스템 1대부터 9대(GPU 총 72개)까지 동시 실행환경에서 모든 DGX-1 시스템 GPU에 대해 최대 포화 상태까지 선형적으로 스케일업하는 것으로 나타났으며, 그림 20에 제시된 모델에 대해 ImageNet 데이터셋을 사용할 때 초당 이미지 개수로 집계한 추론 성능 결과는 아래와 같습니다.

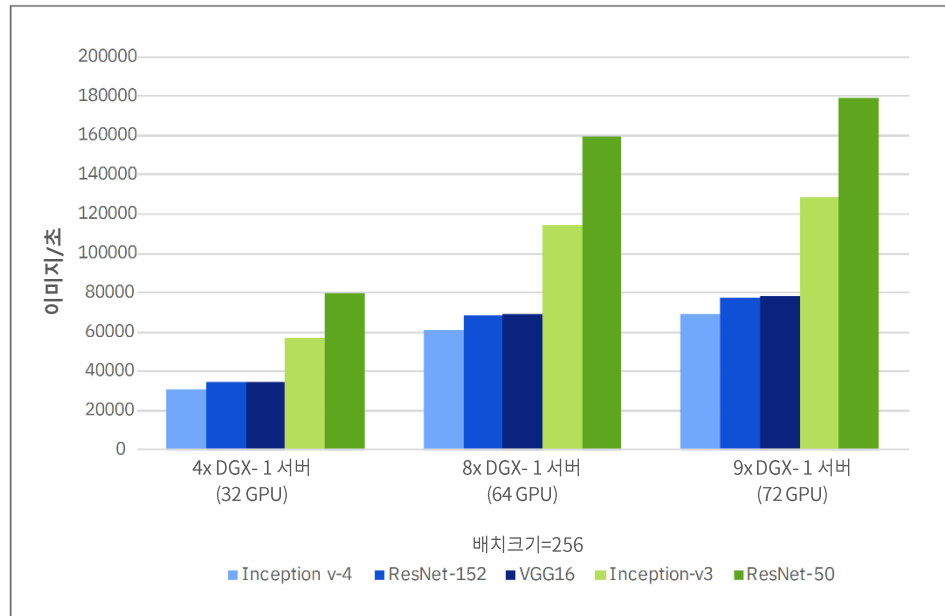


그림 20: 추론 속도 (멀티 DGX-1 시스템, TensorFlow 모델)

테스트된 바와 같이 추론 이미지 처리 속도는 해당 TensorFlow 모델의 트레이닝 속도의 1.5배에서 거의 4배 사이였습니다. ESS 3000 솔루션을 사용하는 DGX-1 시스템을 통해 데이터 사이언티스트는 필요에 따라 단일 DGX-1 시스템에서 트레이닝 및 추론이 혼합된 모드를 실행하여 GPU 한두 개를 추론 전용으로 할당하고 DGX-1 시스템의 나머지 GPU를 트레이닝 작업에 할당할 수 있습니다.

트레이닝 결과 - 멀티 DGX-2 시스템

GPU마다 개별 컨테이너를 사용하는 멀티 DGX-2 시스템의 경우, ESS 3000은 DGX-2 시스템 1대부터 3대(GPU 총 48개)까지 동시 실행 환경에서 모든 DGX-2 시스템 GPU에 대해 최대 포화 상태까지 선형적으로 스케일업하는 것으로 나타났습니다. ResNet-50, ResNet-152, Inception-v3, Inception-v4, VGG-16 model 모델에 대한 트레이닝 처리량을 초당 이미지 개수로 집계한 성능 결과는 아래와 같으며, Synthetic 환경(인공 테스트 데이터는 GPU 내에서 생성, CPU 나 데이터 이동은 수반되지 않음), IBM Spectrum Scale 파일 시스템 환경 및 로컬 RAM 디스크 환경 간 트레이닝 실행을 비교합니다.

그림 21에서 볼 수 있듯이, ESS 3000은 Spectrum Scale 클러스터에서 DGX-2 시스템 GPU에 효과적으로 데이터를 공급하여 DGX-2 시스템을 데이터로 완전히 포화된 상태로 유지함으로써 모든 모델에서 최대의 트레이닝 처리 능력을 달성할 수 있습니다.

또한 GPU를 사용하는 딥러닝(DL) 모델에서 DGX-2 시스템에 대한 모델 트레이닝 결과는 Spectrum Scale 파일 시스템 성능이 로컬 RAM 디스크 성능과 비교할 때 전반적으로 페널티가 최소한의 수준, 또는 전혀 발생하지 않음을 보여줍니다.

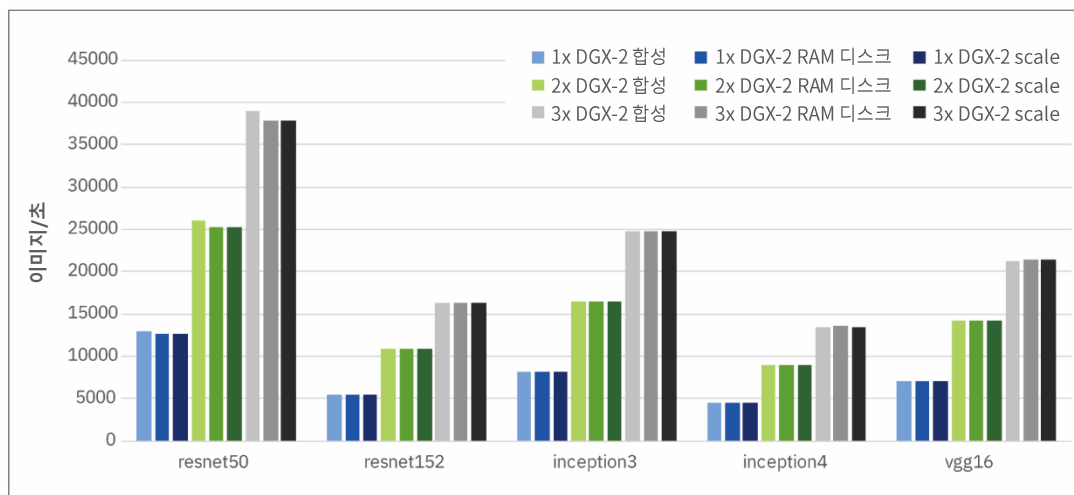


그림 21: 모델-GPU 성능 (Synthetic/Spectrum Scale 파일 시스템/RAM 디스크 성능 비교, 멀티 DGX-2 시스템)

DGX-1 시스템과 관련하여 앞에서 살펴보았듯이, DGX-2 시스템의 경우에도 시스템/GPU 개수가 증가함에 따라 일부 모델은 거의 선형적으로 스케일업되지만, 나머지 모델은 ESS 3000 스토리지 환경이건, Synthetic 환경이건, 로컬 RAM 디스크 환경이건 상관없이 일관되게 비선형적 스케일업 패턴을 보입니다. 이것이 뜻하는 바는 특정 케이스에서는, 확장성이 로컬/공유 스토리지에 상관없이 스토리지 IO에 의해 제한되는 것이 아니고 컴퓨팅 인프라 자체 내의 딥러닝(DL) 모델 확장성 패턴에 의해 제한된다는 것입니다.

전반적으로, ESS 3000 솔루션은 DGX-2 시스템 1대(16 GPU)부터 3대(48 GPU)까지 일관되게 GPU 완전 포화 상태까지 거의 선형적으로 스케일업하는 것으로 나타났습니다.

추론(Inference) 결과 - 멀티 DGX-2 시스템

개별 컨테이너를 사용하는 멀티 DGX-2 시스템의 경우, ESS 3000은 DGX-2 시스템 1대부터 3대(GPU 총 48개)까지 동시 실행환경에서 모든 DGX-2 시스템 GPU에 대해 최대 포화 상태까지 거의 선형적으로 스케일업하는 것으로 나타났으며, 그림 22에 제시된 모델에 대해 ImageNet 데이터셋을 사용할 때 초당 이미지 개수로 집계한 추론 성능 결과는 아래와 같습니다.

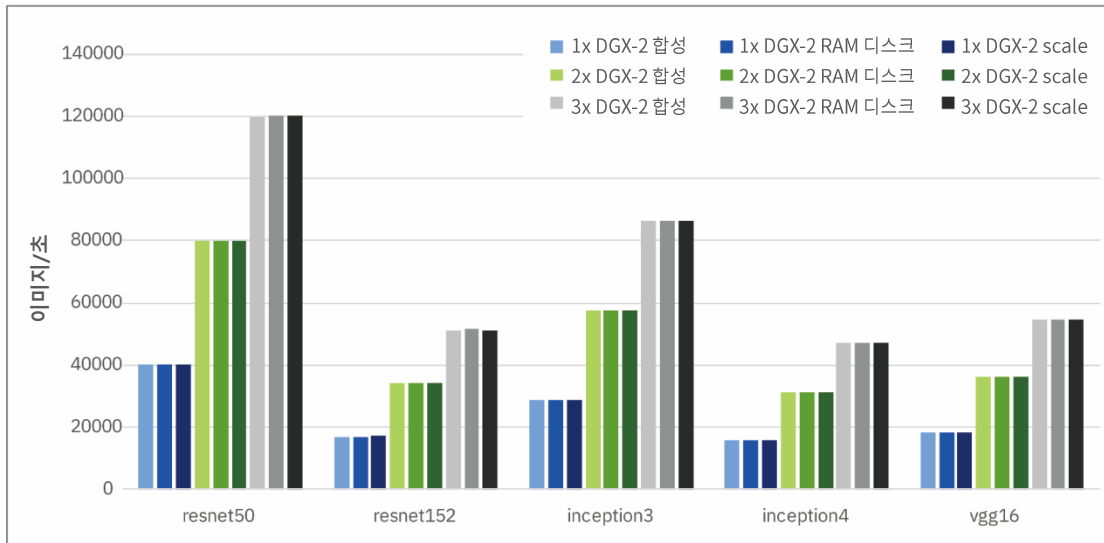


그림 22: 추론 속도 (멀티 DGX-2 시스템, TensorFlow 모델, Synthetic/RAM 디스크/Spectrum Scale 성능 비교)

DGX-1 시스템과 유사하게 DGX-2 시스템의 추론 이미지 처리 속도는 Synthetic, RAM 디스크 또는 Spectrum Scale 파일 시스템 스토리지 유형 중 무엇을 사용하든 해당 TensorFlow 모델의 트레이닝 속도의 1.5배에서 거의 4배 사이였습니다. 또한, 테스트 결과 DGX-1 시스템과 유사하게, ESS 3000 솔루션을 사용하는 DGX-2 시스템을 통해 데이터 사이언티스트는 필요에 따라 단일 DGX-2 시스템에서 트레이닝 및 추론이 혼합된 모드를 실행하여 원하는 개수의 GPU를 추론 전용으로 할당하고 DGX-2 시스템의 나머지 GPU를 트레이닝 작업에 할당할 수 있음을 확인할 수 있었습니다.

엔터프라이즈 데이터 파이프라인

데이터 사이언스 팀의 생산성은 최신 개발 프레임워크, 풍부한 컴퓨팅 파워, 데이터 액세스 가능성을 즉시 활용할 수 있는지 여부에 따라 좌우됩니다. 성능이 중요하기는 하지만 유일한 고려사항은 아닙니다. AI 개발 일정 중에는 대부분의 시간이 데이터 준비와 수집 작업에 소모됩니다. 프로젝트마다 데이터는 각기 다른 소스로부터 추출하고 적절하게 구조화해야 모델 트레이닝에 사용할 수 있습니다. 모델이 개발된 후에는 추적을 위해 데이터를 보관해야 합니다. 데이터의 가치는 다수의 사용자, 시스템, 모델에 의해 다양하게 사용되면서 증대됩니다. 데이터 사이언티스트의 생산성은 전반적인 데이터 파이프라인의 효과와 AI 워크로드 실행에 사용되는 인프라의 성능에 따라 좌우됩니다. 또한, 기반 스토리지 및 네트워크 기술은 이 두 워크플로우 측면 모두에서 중대한 역할을 수행합니다.

AI의 프로토타입 제작 단계에서 프로덕션 환경 배포 단계로 넘어갈 때 해결해야 할 첫 번째 과제는 기존 엔터프라이즈 데이터 파이프라인에 AI를 포함시키는 작업 또는 기존 데이터 저장소를 활용할 수 있는 데이터 파이프라인을 구축하는 일입니다. 파이프라인의 각 단계에 대한 스토리지 요구사항을 포함하여, 일반적인 엔터프라이즈 데이터 파이프라인이 그림 23에 제시되어 있습니다.

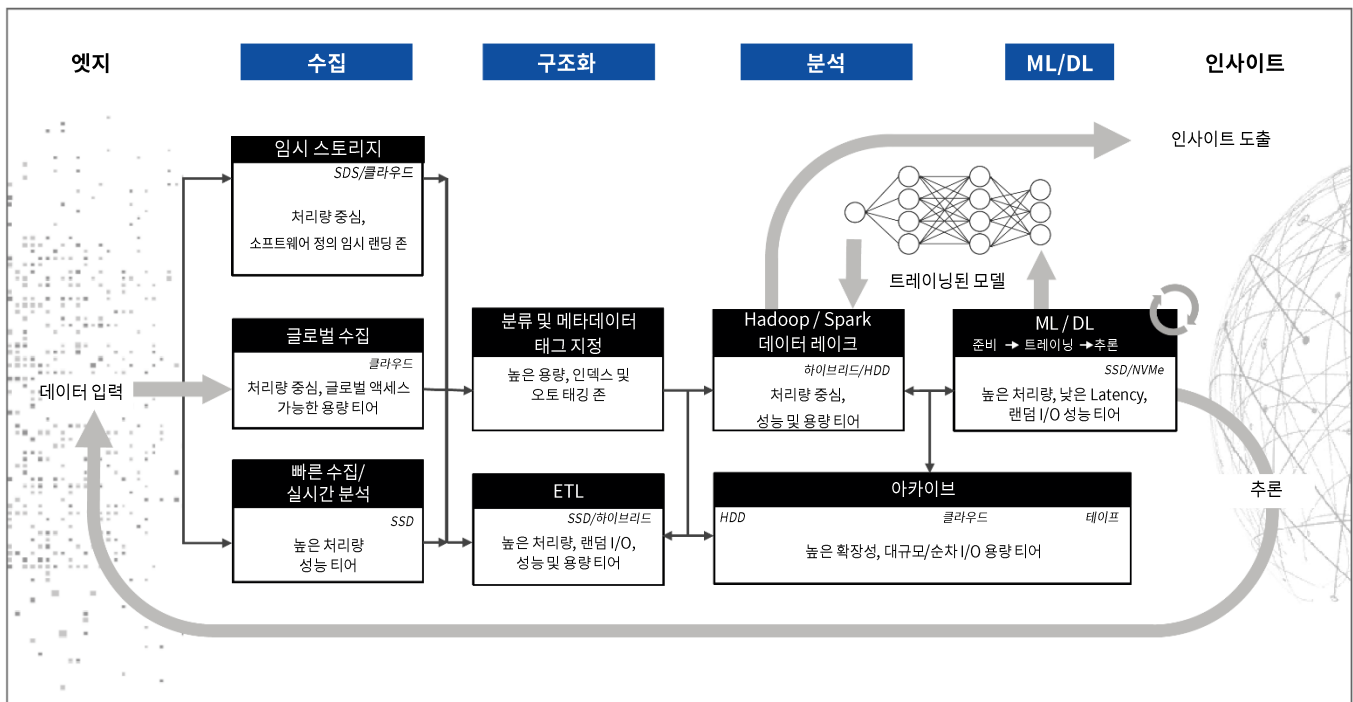


그림 23: 스토리지 요구사항을 포함한 엔터프라이즈 데이터 파이프라인

IBM은 고객이 각 단계에 적합한 성능과 비용 특성으로 엔터프라이즈 데이터 파이프라인을 구축할 수 있도록 포괄적인 소프트웨어 정의 스토리지 포트폴리오(그림 24)를 제공합니다. 이러한 포트폴리오에는 글로벌 수집 및 지리적으로 분산된 저장소를 위한 IBM Cloud Object, 다양한 Elastic Storage Server 모델, 고성능 파일 스토리지와 확장 가능한 공유 데이터 레이크를 위한 IBM Spectrum Scale을 지원하는 Elastic Storage System 3000이 포함됩니다. 또한, IBM Spectrum Archive는 저렴한 아카이브 구현을 위해 테이프에 저장된 데이터에 대한 직접적인 파일 액세스를 지원합니다.

IBM Spectrum Discover는 이 포트폴리오에 새롭게 추가된 제품입니다. Spectrum Discover는 온프레미스와 클라우드에서 엑사바이트 규모의 비정형 스토리지에 대한 데이터 인사이트를 제공하는 메타데이터 관리 소프트웨어입니다. Spectrum Discover는 NFS 및 S3 인터페이스를 통해 IBM Cloud Object Storage, IBM Spectrum Scale 및 서드파티 스토리지에 쉽게 연결되어 수십억 개의 파일과 오브젝트에 대한 메타데이터를 빠르게 수집 및 통합하고 인덱싱합니다. IBM Spectrum Discover는 전체 데이터 파이프라인 중 데이터 분류 단계에서 가장 중요한 역할을 수행하지만 파이프라인을 따라 거버넌스 요구사항을 지원하고 스토리지 최적화를 가능하게 하는 기능도 제공합니다.

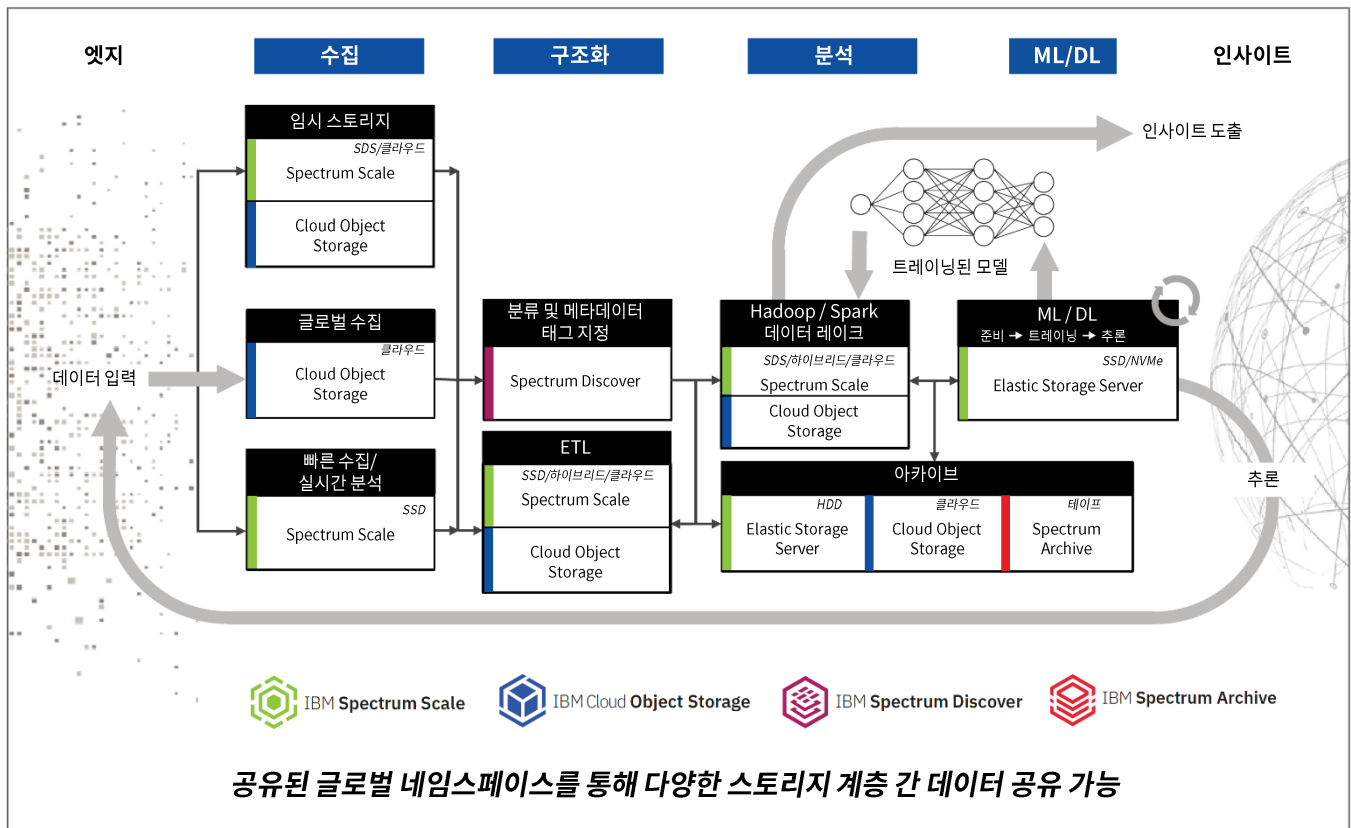


그림 24: AI 데이터 파이프라인을 위한 IBM 솔루션

스토리지 기술을 선택할 때 이처럼 포괄적인 관점에서 파이프라인을 보는 고객이라면 데이터 거버넌스 개선이라는 이점을 누릴 수 있습니다. 데이터 사이언스 팀은 최소한의 데이터 복제본을 사용하여 인사이트 획득 시간을 단축함으로써 이점을 누릴 수 있습니다. 인프라 팀은 관리 간소화, 확장성 및 TCO 개선이라는 이점을 얻을 수 있습니다.

아래의 그림 25에서와 같이 이 파이프라인을 확대하여 ML/DL 단계를 중점적으로 살펴보면 이 단계는 반복적 프로세스임을 알 수 있습니다. 트레이닝된 신경망 모델을 개발한 후에는 모델을 계속 최신 상태로 유지하고 정확성을 향상하기 위해 지속적으로 테스트하고 재트레이닝(retraining)해야 합니다. 프로젝트가 첫 번째 테스트 시스템을 넘어 더욱 확장됨에 따라, 이러한 ML/DL 시스템이 지속적으로 발전하여 궁극적으로 비즈니스 의사결정을 내리는 데 필요한 인사이트를 제공할 수 있으려면 적절한 스토리지 및 네트워킹 인프라를 갖추어야 합니다.

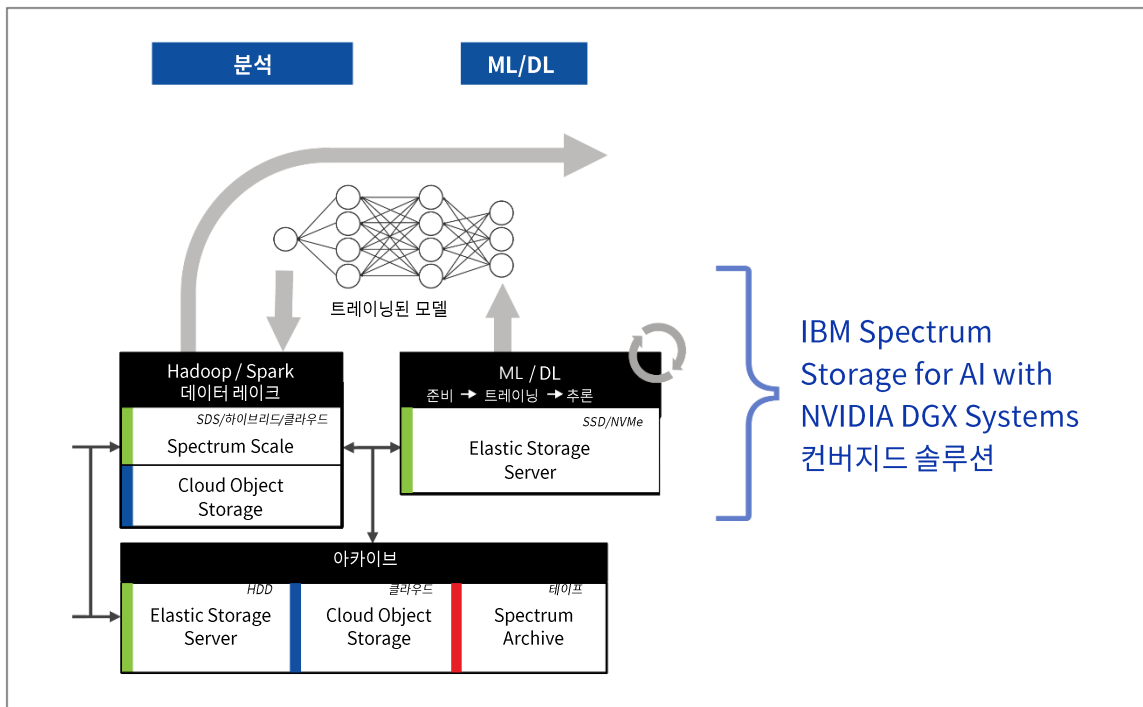


그림 25: ML/DL 단계를 위한 컨버지드 솔루션, IBM Spectrum Storage for AI with NVIDIA DGX System(s)

IBM Spectrum Storage for AI with NVIDIA DGX Systems 컨버지드 솔루션은 이러한 ML/DL sub-stage (데이터 준비, 모델 트레이닝, 추론)를 계획하고 구현하는 데 필요한 시간과 복잡성을 최소화하는 것을 목표로 설계되었습니다.

결론

테스트에서 입증된 대로, Mellanox EDR InfiniBand 패브릭 기반의 ESS 3000 아키텍처를 사용하는 IBM Spectrum Storage for AI는 딥러닝(DL) 워크로드 트레이닝 및 추론을 위해 DGX-1 및 DGX-2 시스템을 활용할 때 최고의 성능을 제공하며, 다수의 DGX-1 또는 DGX-2 시스템과 구동될 경우 높은 대역폭과 낮은 Latency로 GPU의 활용도를 극대화하여 고성능 병렬 처리를 수행합니다.

IBM Spectrum Storage for AI 및 NGC Software Stack과 통합된 NVIDIA DGX POD 시스템은 IBM Spectrum 소프트웨어 정의 스토리지 솔루션 제품군과 함께, 워크로드 통합, 데이터 준비 및 관리, 그리고 프로세스 자동화 기능을 제공하여 End-to-End AI 데이터 파이프라인 개발을 효율화하고 기존 인프라로의 통합을 쉽게 구현할 수 있도록 해줍니다.

추가 자료

Introduction Guide to the IBM Elastic Storage Server

<http://www.redbooks.ibm.com/redpapers/pdfs/redp5253.pdf>

NVIDIA DGX SuperPOD Delivers World Record Supercomputing to Any Enterprise

<https://devblogs.nvidia.com/dgx-superpod-world-record-supercomputing-enterprise/>

NVIDIA DGX-2 POD: From Concept to World-Record Setting Supercomputer in Three Weeks

<https://www.nvidia.com/en-us/data-center/resources/nvidia-dgx-superpod-reference-architecture>

감사의 말

IBM Spectrum Scale 레퍼런스 아키텍처에 여러모로 기여해 주신 NVIDIA의 파트너, Jacci Cenci, Satinder Nijjar, Darrin Johnson, Robert Sohigian, Tony Paikeday께 감사의 말을 전합니다. 또한, Arrow Electronics 팀에게도 감사드립니다.

IBM 팀원인 Dave McDonnell, Brian Porter, Constantine Arnold, Udayasuryan A Kodoly의 지원과 지도가 없었다면 이 연구를 완료하지 못했을 것입니다. 이 연구에 큰 도움이 된 인사이트와 전문 지식을 제공해주신 이 모든 분께 진심으로 감사드립니다.



© Copyright IBM Corporation 2019
IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America

IBM, IBM 로고 및 ibm.com은 미국 또는 기타 국가에서 사용되는 International Business Machines Corporation의 상표 또는 등록상표입니다. 이와 함께 기타 IBM 상표가 기재된 용어가 상표 기호(® 또는 ™)와 함께 이 정보에 처음 표시된 경우, 이와 같은 기호는 이 정보를 발행할 때 미국에서 IBM이 소유한 등록상표 또는 일반 법적 상표입니다. 또한 이러한 상표는 기타 국가에서 등록상표 또는 일반 법적 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보"(ibm.com/legal/copytrade.shtml)에 있습니다.

NVIDIA, NVIDIA 로고, DGX, DGX-1 및 DGX-2는 미국 및/또는 기타 관할권에서 사용되는 NVIDIA, Inc. 또는 그 계열사의 등록상표 또는 상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

기타 회사, 제품 및 서비스 이름은 해당 회사의 상표 또는 서비스표입니다.

본 문서에서 IBM의 제품 또는 서비스를 언급하는 것이 IBM이 영업하고 있는 모든 국가에서 이를 사용할 수 있다는 것을 의미하지는 않습니다.



재활용하세요