

IBM SpectrumAI with NVIDIA® DGX™

AI 데이터 파이프라인을 위한 동력 엔진

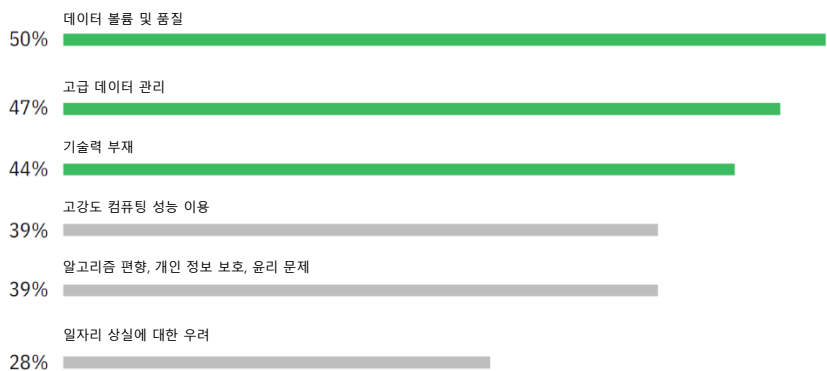
IBM SpectrumAI with NVIDIA DGX는 컴퓨팅/스토리지 통합 솔루션으로, 최첨단 시스템 및 소프트웨어의 혁신을 활용하여 AI 라이프사이클, 즉 데이터 준비부터 트레이닝, 추론까지의 전 과정을 지원합니다.

과제:

업계에서는 몇 년 내에 거의 모든 기술, 서비스, 과학 분야가 AI와의 접목을 통해 발전을 도모할 것으로 예측하고 있습니다. 이러한 예측을 입증하듯 고성능 GPU와 컨테이너의 형태로 제공되는 최적화된 ML/DL 프레임워크를 기반으로 하는 AI가 모든 산업 영역에서 빠르게 도입되고 있습니다. 혁신적인 컴퓨터 비전 및 개체 탐지, 사람/컴퓨터 상호 작용, 데이터 분류, 고급 패턴 탐지 기술이 현재 제공되고 있으며 그 사용 분야도 점차 확대되고 있습니다. 머신 러닝, 딥러닝, 신경망 트레이닝의 충실도는 가용 데이터의 품질 및 수량에 의해 좌우됩니다. 따라서, 강력한 서버로 데이터 인프라를 뒷받침하는 것이 관련될 수 있습니다. 방대한 공유 데이터에 빠른 속도로 액세스하는 것이 중요하기 때문입니다.

적합한 데이터에 액세스하지 못하면 데이터 과학 팀의 생산성이 크게 떨어질 수 있습니다. 널리 알려진 대로,¹ 데이터 과학자는 데이터를 준비하고 수집하는 데 대부분 시간을 보냅니다. 모든 프로젝트에는 트레이닝 및 테스트 데이터 세트가 필요하며, 이 데이터 세트는 모델 개발과 지속적인 검증에 쓰일 수 있도록 올바르게 구성되고 태그가 지정되어야 합니다. AI 데이터 파이프라인을 제대로 관리한다면 (데이터 거버넌스, 확장 가능한 메타데이터 태깅, 아카이빙 및 탄력적 공유 스토리지 등 포함), 이렇게 마련된 데이터 리포지토리에서 여러 모델을 개발하고 트레이닝할 수 있습니다.

그림 1: AI 워크로드 배포 시 해결 과제



IDC 설문 조사에서 AI 워크로드 배포 시 해결해야 할 3대 과제로 데이터 볼륨 및 품질, 고급 데이터 관리, 기술력 부재가 꼽혔습니다.

솔루션 설명:

데이터 과학의 생산성은 ML/DL 워크로드의 생성과 트레이닝에 쓰이는 인프라 및 소프트웨어에 의해 결정됩니다. 반드시 고성능 공유 스토리지와 최신 GPU 서버를 사용하여 최적화된 컨테이너 AI 워크로드를 실행해야 합니다. GPU 밀도가 증가하면 고밀도 고속 스토리지에 대한 요구도 커집니다. 플래시 스토리지 기술의 발전 덕분에 이러한 스토리지 시스템을 구축하고 최신 고속 네트워킹으로 연결하여 단일 랙 구성에서 최대 72개의 GPU를 가동하는 것이 가능해졌습니다. 더불어, 데이터 병목 지점을 없애고 시스템 성능을 최고 수준으로 유지하려면 전문적인 설계 및 구성이 필요합니다.

그러나 AI 프로젝트에서 진정한 확장성을 실현하려면 데이터 과학 팀과 IT 팀이 협업하여 데이터 수집부터 추론 및 아카이빙까지 포괄하는 AI 데이터 파이프라인의 전 범위에서 효율을 높여야 합니다. 통상적으로 다양한 데이터 소스가 사용되는데, 이를 최적화하려면 충분한 성능 및 유연성을 가진 공통 데이터 저장소에서 여러 소스의 데이터를 수집하고 분석 및 스크리핑 툴을 지원하며 사용 빈도가 낮은 데이터는 투명한 방식으로 차등화해야 합니다.

툴 선택에 제한을 두지 않으면서 데이터 이동을 불필요하게 만들면 팀의 생산성이 향상되고 비용이 절감되며 데이터 거버넌스 요구사항도 간소해집니다. 소프트웨어 정의 스토리지의 확장성 및 관리 기능을 통해 비즈니스 요구사항에 따라 데이터 흐름을 통합 모니터링하고 시스템을 스케일아웃 확장하거나 계층화할 수 있습니다. IBM Spectrum Scale을 사용하면, 기존 스토리지 어레이에서 흔히 겪는 데이터 액세스의 장애 요인과 데이터 복사의 오버헤드로부터 벗어날 수 있습니다.

¹ <https://www.information-management.com/slideshow/how-data-professionals-spend-their-time>

² 출처: Cognitive, ML, and AI Workloads Infrastructure Market Survey, IDC, 2018년 1월, n=405, 직원 수 1,000명 이상인 기업(미국), 직원 수 500명 이상인 기업(캐나다), IDC Technology Spotlight #US43977818 :

주요 특징

- 컨버지드 솔루션: 가장 까다로운 머신 러닝/딥러닝 프로젝트에서도 손쉽게 배포
- IBM Spectrum Scale v5: 소프트웨어 정의 스토리지로 AI 데이터 파이프라인을 통해 데이터 이동 간소화
- NVIDIA DGX-1 서버: AI 및 머신 러닝 전문 솔루션
- NVIDIA DGX 소프트웨어 스택: 최고의 GPU 트레이닝 성능을 제공하도록 최적화
- 검증된 데이터 성능: 100GB/s 이상의 처리량으로 단일 랙에서 DGX-1 서버를 9 대까지 지원
- 컨테이너화된 AI 워크로드를 위한 공유 데이터 서비스를 구현할 수 있는 기반
- 솔루션 전반에 대한 전문성을 갖춘 비즈니스 파트너가 제공하고 IBM 과 NVIDIA 가 뒷받침하는 형태의 간소화된 지원 모델

생산성, 유연성, 확장성

IBM SpectrumAI with NVIDIA DGX를 동력원으로 하는 AI 데이터 인프라에서 데이터 과학 서비스를 구현할 수 있습니다. IBM과 NVIDIA의 통합 솔루션인 IBM SpectrumAI 플랫폼은 데이터 준비, 트레이닝, 추론 등을 모두 포함한 데이터 과학의 라이프사이클 전 범위를 지원하도록 설계되었습니다. 하드웨어 및 소프트웨어 스택에서 소프트웨어 정의 스토리지와 최신 NVIDIA 혁신 기술을 활용하며 원하는 대로 구성 가능한 아키텍처인 IBM SpectrumAI는 확장된 데이터 파이프라인에도 잘 맞는 뛰어난 유연성뿐만 아니라 다수의 사용자 및 모델을 처리할 수 있는 우수한 성능을 제공합니다. 즉, IBM SpectrumAI는 데이터 과학자에게 필요한 성능, 데이터, 유연성을 모두 제공합니다.

철저한 테스트와 튜닝을 거쳐 제공되는 IBM SpectrumAI는 컴퓨팅, 스토리지, 네트워킹이 통합되었지만 개별 확장이 가능한 컨버지드 솔루션으로 NVIDIA DGX-1, IBM Spectrum Scale 올플래시 스토리지, 그리고 Mellanox 네트워킹으로 구성되어 있습니다. 또한, IBM SpectrumAI with NVIDIA DGX는 소규모로 출발하여 필요에 따라 확장할 수 있습니다. 즉 DGX-1 서버 3대와 단일 NVMe 스토리지 시스템으로 시작할 수 있습니다. NVMe 시스템 3대와 DGX-1 서버 9대로 이루어진 풀 랙 구성은 Tesla V100 Tensor Core GPU 72개까지 확장하여 120GB/s의 데이터 처리량을 지원할 수 있습니다. IBM Spectrum Scale은 소프트웨어 정의 스토리지이므로 복수의 랙 또는 다른 스토리지 유형까지 포괄하면서 자동으로 데이터를 계층화하여 보호하고 엑사바이트 단위의 데이터를 아카이빙할 수 있습니다.

NVIDIA DGX-1은 풀 스택 AI 컴퓨팅 솔루션이며, 세계 최고의 데이터 센터 액셀러레이터인 NVIDIA Tesla V100 Tensor Core GPU 8개를 사용합니다. 각 DGX-1은 1petaFLOPS의 혼합 정밀도 트레이닝(mixed precision training) 성능을 제공하며 즉시 사용 가능하도록 설계되었습니다. 이를 위해 NVIDIA CUDA, 딥러닝 라이브러리, Docker 컨테이너 관리 기능 등이 포함된 최적화된 소프트웨어 스택을 활용하여 빨리 시작할 수 있습니다. DGX 컨테이너 레지스트리는 NVIDIA 최적화 AI 및 데이터 과학 컨테이너로 구성된 확장 카탈로그를 제공합니다. 데이터 과학자는 NVIDIA DGX를 통해 신속하게 머신 러닝 및 딥러닝 모델을 개발하고, 손쉽게 반복하고 실험하며, 어떤 규모든 트레이닝 및 추론을 위해 배포할 수 있습니다.

IBM Spectrum Scale v5는 요구가 까다로운 최신 워크로드도 지원하도록 설계된 소프트웨어 정의 병렬 파일 시스템입니다. 40,000개 이상의 NVIDIA GPU를 실행하는, 규모 및 성능 면에서 세계 최고 수준의 슈퍼컴퓨터도 지원할 수 있는 확장성을 갖추었으며³ 최신 올플래시 스토리지에 최적화되어 AI 워크로드의 처리량 및 복합 크기 파일 요구사항을 지원할 수 있습니다. NFS, SMB, Object, HDFS 등 여러 프로토콜을 지원하므로 유연하게 사용할 수 있습니다. 고급 데이터 관리 기능이 있어 스토리지 또는 클라우드의 전 범위에서 스크립트, 사용 패턴, 정책에 따라 자동으로, 단 사용자에게는 투명한 방식으로 데이터를 계층화하여 관리합니다.

IBM Spectrum Discover는 더 효과적인 데이터 분류, 추적, 거버넌스를 지원하는 최신 메타데이터 관리 소프트웨어로서 IBM Spectrum Scale 및 IBM Cloud Object Storage를 대상으로 비정형 스토리지에 대한 데이터 인사이트를 제공합니다. 수십억 개의 파일 및 오브젝트에 대한 메타데이터를 신속하게 수집, 통합, 인덱싱하여 충실한 메타데이터 계층을 제공하므로, 데이터 과학자, 스토리지 관리자, 데이터 실무자는 이를 활용하여 방대한 비정형 데이터를 효율적으로 관리, 분류하고 인사이트를 얻을 수 있습니다.

IBM SpectrumAI 네트워크 연결 기능은 Mellanox에서 제공합니다. Infiniband 옵션을 선택하면 안정적인 100Gbps 네트워크에서 RDMA와 같은 첨단 기술을 활용할 수 있습니다. 적절한 사이징(sizing) 및 구성으로 스토리지 처리량을 확보하고 NVIDIA DGX-1 서버 사이에 노드 간 통신을 구현할 수 있습니다.



³ <https://www.ibm.com/blogs/systems/fastest-storage-fastest-system-summit/>

AI 데이터 파이프라인에 최강의 성능 제공

IBM SpectrumAI with NVIDIA DGX는 기업의 AI 전략을 뒷받침하는 기초가 됩니다. IBM 및 NVIDIA의 전문 채널 파트너가 즉시 배포 가능한 형태로 공급하므로, 데이터 과학자들은 곧 바로 뛰어난 생산성을 발휘하고 미래에도 대비할 수 있습니다. NVIDIA의 전문 기술과 지원을 근간으로 하므로, 데이터 과학자들이 스토리지 성능에 대한 자신감을 갖고 최신 NVIDIA 최적화 GPU 액셀러레이터 툴을 활용할 수 있습니다.

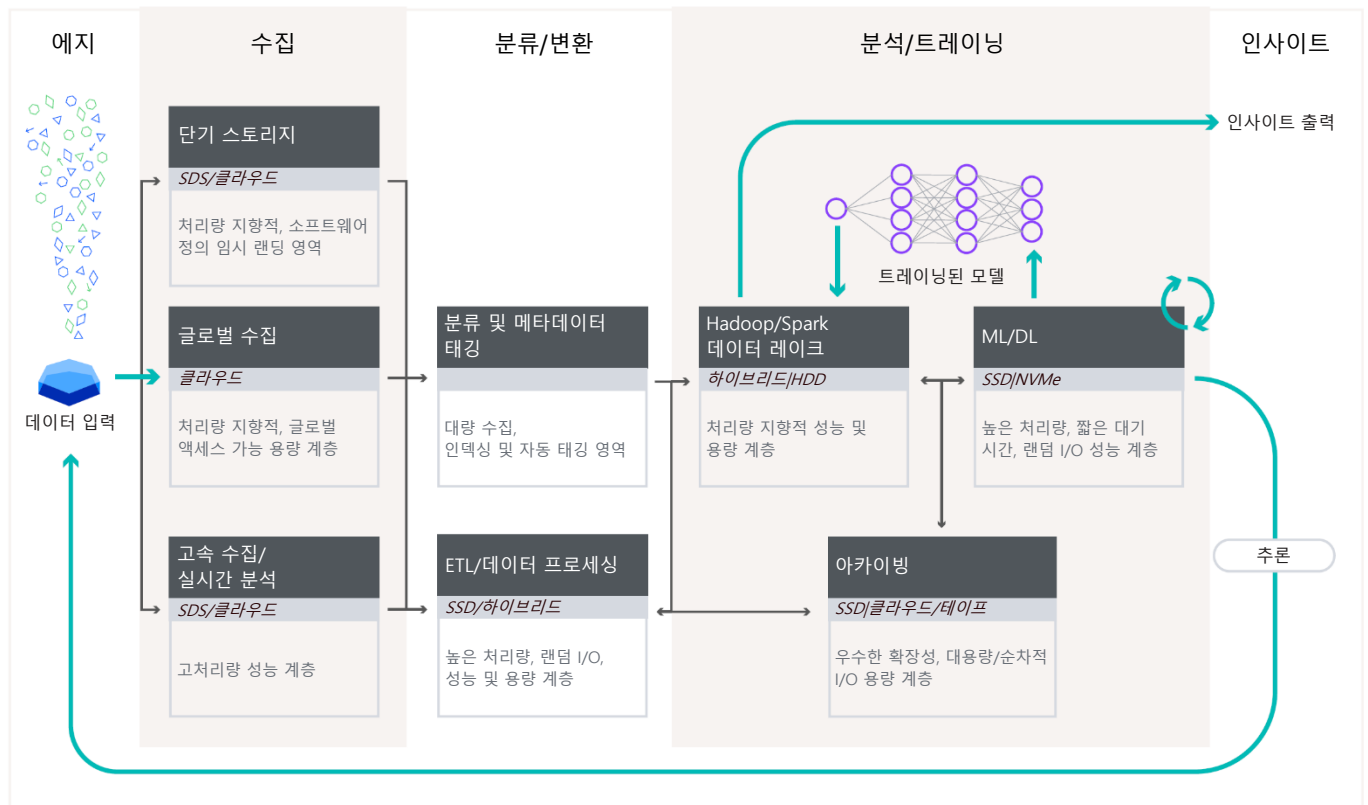
데이터 및 비즈니스 요구사항이 증가하면, IBM Storage에서 제공하는 여러 옵션과 혁신적인 기술을 통해 모든 AI 데이터 파이프라인의 성능, 경제성, 그리고 데이터 거버넌스 요구사항을 해결할 수 있습니다. IBM Spectrum Scale은 AWS 퍼블릭 클라우드를 포함하여 다양한 미디어를 포괄하는 스토리지 서비스를 제공합니다. 데이터와 메타데이터를 IBM Cloud Object Storage 및 테이프와 공유하면서 지리적으로 분산된 스토리지 또는 로컬 스토리지를 탄력적으로 제공할 수 있습니다.

미래 지향적인 기업이라면 엔드 투 엔드 AI 데이터 파이프라인을

검토하고 구축함으로써 중단 없는 생산력, 비용 절감 및 간소화된 데이터 거버넌스의 이점을 누릴 것입니다. 데이터 과학 팀은 데이터 복사본 수를 최소화하면서 신속하게 인사이트를 발굴할 수 있습니다. 동시에 인프라 팀은 간소화된 관리, 확장성, 향상된 TCO의 이점을 누릴 수 있습니다.

IBM SpectrumAI with NVIDIA DGX는 새로운 AI/ML/DL 애플리케이션을 개발하고 제공할 수 있는 데이터 과학 플랫폼으로서 데이터 준비부터 트레이닝, 추론, 시각화까지 모두 지원합니다. 엔터프라이즈 환경에서 검증된 이 솔루션은 고속, 플러그인/파워업 배포 방식으로 획기적인 성능을 제공합니다. IBM SpectrumAI with NVIDIA DGX는 AI 컴퓨팅 및 고성능 스토리지/네트워킹 분야의 최신 혁신 기술과도 통합하여 개발 라이프사이클의 전 범위에서 다양한 프로젝트를 지원하면서 가장 빠르고 효율적으로 인사이트를 발굴할 경로를 제공합니다. 이를 뒷받침하는 간소화된 지원 모델을 통해 AI 워크로드를 중단 없이 실행할 수 있습니다.

AI 데이터 파이프라인과 스토리지 요구사항



IBM SpectrumAI의 기술적 특징

- DGX-1 서버 - AI 및 머신 러닝 전용 솔루션
- NVIDIA DGX 소프트웨어 스택 - 최고의 GPU 액셀러레이터 트레이닝 성능을 제공하도록 최적화
- IBM Spectrum Scale v5 - 최고의 소프트웨어 정의 파일 스토리지. AI 워크로드 용도로 설계되었으며, 향상된 소형 파일, 메타데이터, 랜덤 IO 성능 제공
- NVMe 올플래시 스토리지 - 대기 시간 최소화, 전력 효율 및 데이터 밀도 향상
IBM Spectrum Scale 분산 데이터 보호 기능으로 2U 빌딩 블록마다 300TB 이상의 용량을, 단일 랙에서 최대 120GB/s의 데이터 처리량을 제공합니다.
- 막힘없는 데이터 파이프라인 연결 - 복수의 랙, 기타 IBM SpectrumAI 구성, 워크스테이션을 포괄하는 연결을 통해 데이터 과학자들에게 IBM SpectrumAI 데이터
- IBM Spectrum Discover와의 통합 - IBM Spectrum Scale과 IBM Cloud Object Storage의 전 범위에서 확장 가능한 데이터 거버넌스 및 메타데이터 태깅 지원



© Copyright IBM Corporation 2018
IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of
America

IBM, IBM 로고 및 ibm.com 은 미국 또는
기타 국가에서 사용되는 International
Business Machines Corporation 의 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 IBM 이 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 및 DGX-1은 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 영입하고 있는 모든
국가에서 이를 사용할 수 있다는 것을
의미하지는 않습니다.