

지능형 스토리지를 활용하는 기업의 머신 러닝 및 딥 러닝

요약

데이터, 인프라스트럭처의 발전, 머신 러닝 및 딥 러닝(ML/DL) 툴킷의 보편화로 AI(Artificial Intelligence) 솔루션이 엔터프라이즈 데이터 센터의 주류로 빠르게 자리 잡고 있다. AI는 자동차, 의료, 생명 과학, 금융, 기술, 소매 등 다양한 기업 분야의 데이터를 통찰력으로 전환한다. 보험 분야에서 예측 AI는 인수 시의 위험을 제거하고, 금융 분야에서는 실시간 딥 러닝이 부정 행위 발생 시 이를 인식하고, 데이터 센터 관리 분야에서는 패턴을 분석하여 장애 발생과 확장성 문제를 예측하는 등, 데이터는 이와 같은 업계에서 경쟁 우위를 차지할 수 있는 요소가 되었다.

인공 지능, 특히 딥 러닝은 데이터를 사용하는 컴퓨팅 엔진에 데이터를 제공하는 방식에 대한 새로운 요구를 불러일으킨다. 데이터 센터에 인공 지능을 배포하는 새로운 현실은 집적도, 처리량, 동시성 및 스케일 아웃 데이터 아키텍처 변경에 대한 요구를 변화시킨다. IT가 기업에 AI의 이점을 제공하기 위해서는 스토리지와 컴퓨팅을 결합하는 것에 대해 다르게 생각해야 한다.

이 백서에서는 기업의 딥 러닝과 인공 지능이 어떻게 데이터 센터 아키텍처에 새로운 워크플로와 당면 과제를 제기하는지 설명한다. 또한 스케일 아웃 컴퓨팅과 스토리지를 더 긴밀하게 통합하도록 특별히 설계된 인프라스트럭처 아키텍처를 통해 솔루션을 구축하는 방법을 다룬다.

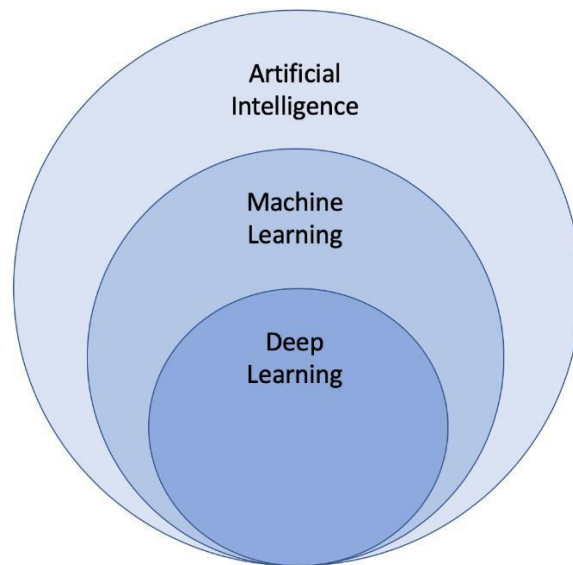
딥 러닝에서는 프로세서가 데이터를 기다리지 않고도 대량의 데이터를 공급받을 수 있어야 한다. Dell EMC Isilon Series와 같은 적합한 스토리지 기술에 컴퓨팅을 적절히 통합하면 데이터를 프로세서의 속도에 맞게 머신 러닝 파이프라인에 공급할 수 있다. 적절하게 균형 잡힌 시스템은 혁신을 가속화하고 IT 조직과 이 시스템을 사용하는 데이터 과학자 모두에게 유연성과 민첩한 대응 능력을 제공한다.

딥 러닝으로 인한 기업의 변화

많은 사람들이 유명 잡지인 The Economist의 석유 굴착 장치를 배경으로 한 표지에서 데이터가 “세계에서 가장 가치 있는 리소스”¹라고 선언한 것을 보았을 것이다.

ML(Machine Learning) 및 DL(Deep Learning)의 발전으로 인해 실제로 모든 조직의 데이터에 대한 중요성이 새롭게 부각되었으며, *데이터는 경쟁 차별화 요소가 되었다.*

그림 1: AI, ML, DL 간의 관계



출처: Moor Insights & Strategy

인공 지능은 컴퓨터가 일반적으로 인간의 지능과 연관되는 형태의 의사 결정을 내리거나 통찰력을 제공하는 일반 기술 분야를 의미한다. AI의 간단한 예로는 소매업체의 온라인 웹사이트에서 과거 쇼핑 기록과 현재 검색 중인 정보를 기반으로 제품을 추천하는 소매 추천 엔진이 있다.

¹ The Economist, The World's Most Valuable Resource is No Longer Oil, But Data. 2017년 5월 6일.

머신 러닝은 알고리즘이 데이터를 구문 분석하여 해당 데이터를 학습하고, 실제 환경에서 학습한 내용을 적용하여 의사 결정을 내리는 인공지능의 한 유형이다. 예를 들어, 이메일 스팸을 감지하는 머신 러닝 시스템은 매일 사용자가 스팸 폴더로 이동하는 수백만 개의 이메일의 샘플을 통해 학습할 수 있다.

더 흥미롭고 활기를 띠고 있는 인공지능의 분야 중 하나인 딥 러닝은 머신 러닝의 하위 집합이다. 딥 러닝은 *신경망*이라고 하는 알고리즘을 사용하여 데이터를 접해가며 예측을 지속적으로 조정한다. 딥 러닝은 자율 주행 차량, 인간의 기분을 인식하는 감정 분석, 그 외 동적으로 의사 결정을 내리거나 권장 사항을 제시하기 위해 실제 데이터를 다루는 AI 기술 대부분의 핵심이다. 이 백서에서는 딥 러닝을 가장 직접적으로 다루지만, 어떤 유형의 데이터 집약적인 AI 시스템을 구현하든 아키텍처상으로 유사하다.

AI 활용 사례는 AI 기반 챗봇 및 음성 응답 시스템부터 고객 행동 예측, 최적화된 공급망 관리까지 많고 다양하다. 지능형 기업의 시대가 도래했으며 많은 이들이 머신 러닝 기술을 활용하여 업계에서 경쟁 우위를 최대한 확보하는 방법을 파악하는 당면 과제로 인해 어려움을 겪고 있다.

최근 IT 전문가 1,300명 이상을 대상으로 실시한 설문조사²에 따르면 응답자의 60% 이상이 IT 예산의 5% 이상을 인공지능에 지출할 계획인 조직에서 근무하고 있다. 응답자의 5분의 1은 IT 예산의 20% 이상을 AI에 지출할 계획인 조직에서 근무한다.

딥 러닝 기술을 사용하는 인공지능은 모든 비즈니스에 영향을 미치며, 이는 종종 예기치 않은 방식으로 이루어진다. 머신 러닝과 여기서 파생되어 더 특화된 딥 러닝이 현대 기업에 미치는 영향의 몇 가지 예는 다음과 같다.

- **M&E(Media and Entertainment)** 업계에서는 머신 러닝을 활용하여 다양한 작업에 인텔리전스를 제공한다. 감정 분석은 영화 및 TV 프로그램 상영에 대한 관객의 반응을 분류하는 데 사용된다. 또한 M&E 업계는 딥 러닝 알고리즘에 의해 학습된 이미지 인식을 활용하여 수많은 비디오 콘텐츠의 메타데이터를 자동으로 생성한다.

² O'Reilly Media, The State of Machine Learning Adoption in the Enterprise, 2019년.

- 다양한 업계의 최신 **제조** 조직은 인공 지능과 머신 러닝을 운영의 여러 측면에서 활용한다. 이미지 인식 시스템은 제조 라인의 제품을 분석하여 결함을 식별한다. 또한 머신 러닝 시스템은 공장 전체의 센서를 분석함으로써 해결되지 않으면 장애를 초래할 수 있는 패턴을 인식하고 식별하여 장애 예측 분석을 지원한다. 머신 러닝은 지능형 조달 및 물류에 최적화된 JIT(Just In Time) 운영을 유지하는 공급망 의사 결정을 추진하는 데도 사용된다.
- **자동차 및 운송** 업계는 딥 러닝을 사용함으로써 자동차에 대한 전 세계의 사고방식을 바꾸고 있다. 딥 러닝 기술은 자율 주행 차량을 제공하는 궁극적인 목표를 향한 경쟁에 활력을 불어넣고 있다. 그 과정에서 우리는 운전자 차량에서 일어나는 일을 인식할 수 있는 지능형 및 적응식 정속주행 시스템, 반자동 주행, 장애 예측 분석, 운전자 모니터링 등의 형태로 된 실제 애플리케이션을 경험하고 있다. 머신 러닝과 딥 러닝의 발전이 없었다면 이러한 일은 불가능했을 것이다.

AI 기술을 기업에 배포하는 것이 매우 파급력이 크지만, 대부분의 조직에서는 새로운 이야기일 뿐이다. AI 프로젝트를 시작하기 전에 공통 구성 요소를 간소화하여 살펴보는 것이 중요하다. 이미지 감지, 이미지 분류, 세분화, 자연 언어 처리 및/또는 예측 분석을 지원하는 거의 모든 딥 러닝 솔루션은 공통적인 핵심 기술을 활용한다. 이러한 기술은 기본적으로 지원되는 플랫폼에 기반하여 배포되며, 이러한 활용 사례를 구현하는 데 흔히 볼 수 있는 TensorFlow, PyTorch 및 Caffe2 등 일반적인 소프트웨어 패키지에 맞게 미세 조정된다.

대부분의 기존 IT 실무자는 다양한 활용 사례를 위해 AI 솔루션을 효율적으로 설계하고 구축하는 데 필요한 기술을 보유하고 있지 않다. 머신 러닝과 딥 러닝은 새로운 과제를 수반하며, 데이터에 대한 혁신적인 사고 방식을 요구하는 기술이다.

위에서 언급한 동일한 설문조사에 따르면 딥 러닝을 배포하는 방법에 대한 이해의 부족과 이러한 워크로드에 대비하지 않은 인프라스트럭처가 AI의 도입을 방해하는 주요 요소인 것으로 밝혀졌다.

경쟁이 치열한 IT 조직에서 가장 중요한 과제는 이러한 격차를 해소하고, 유연하고 미래 지향적인 분석 아키텍처가 지원하는 딥 러닝을 배포하는 데 필요한 기술을 구축하는 것이다.

데이터 센터에서 딥 러닝 사용을 위한 아키텍처 설계

딥 러닝이 스토리지 아키텍처에 미치는 다차원적 영향을 이해하려면 일반적인 학습 워크플로에 대한 높은 수준의 이해가 필요하다. 학습 파이프라인의 각 단계는 기본 인프라스트럭처에 대해 서로 다른 사항을 요구한다. 이는 그림 2에서 볼 수 있다.

그림 2: 일반적인 머신 러닝 및 딥 러닝 파이프라인

	<u>INGEST</u>	<u>DATA PREP</u>	<u>REFINE</u>	<u>TRAIN</u>	<u>DEPLOY</u>	<u>RETENTION</u>
	IOT, Logs, Sensors, Users, Etc	CPU-intensive Servers	GPU-enabled Server & Workstations	High Performance GPU-based Servers	CPU or Inference Accelerated Edge, Client, or Server	Long-term Storage
Access Pattern	Sequential	Sequential or Random	Random	Random	Random	Sequential
Access Type	Write	Read & Write	Read	Read	Read	Write
Concurrency	Variable	Low	Moderate	High	Low	Low
Performance	High	High	High	High	Moderate	Low
Storage	Block, File, or Object	Block or File	File or Object	File or Object	Block or Memory	Block, File, or Object
Scale	MB-GB	MB-TB	TB-PB	TB-PB	KB-MB	TB-PB

출처: Moor Insights & Strategy

이러한 단계는 다음과 같이 요약된다.

- **데이터 수집** - 데이터는 엣지 디바이스, 로그 파일, 음성 또는 비디오 스트림, 고객 관계 관리 시스템과 같은 외부 소스(또는 여러 소스)에서 추출된다. 데이터는 추출된 후 저장된다. 스토리지 솔루션은 수신 데이터가 필요로 하는 성능 기준에 적합해야 한다.
- **데이터 준비** - 데이터가 교육을 위해 정리 및 변환된다. 이 중요한 단계를 통해 데이터의 일관성을 확보하고, 이상점을 식별하며, 데이터 세트가 교육 알고리즘에 최적화된다. 지도 학습과 같은 일부 머신 러닝 유형의 경우 이 단계에서 데이터에 레이블을 지정해야 한다.

- **데이터 검색 및 시각화** - 교육 알고리즘과 매개변수를 최적화하기 위해 데이터 과학자가 데이터 관련 작업을 수행한다. 이 작업은 매우 반복적인 과정이지만, 적은 양의 스토리지와 컴퓨팅만 필요하다.
- **모델 교육 및 개발** - 이 단계에서 대량의 작업이 발생한다. 정리된 데이터는 GPU 또는 기타 고성능 컴퓨팅 엔진 클러스터로 공급되며, 이러한 과정이 매우 오랜 기간 반복된다. 교육에는 동시성이 높은 랜덤 읽기 작업에 최적화되어 처리량이 높은 스토리지가 필요하다.
- **모델 배포 또는 운영 추론** - 교육 단계에서 생성된 모델은 실제 데이터를 기준으로 하여 배포된다. 이 단계의 특성은 배포되는 딥 러닝 유형에 따라 크게 달라진다. 예를 들어 이미지 인식은 스토리지 시스템과의 상호 작용이 거의 없는 스마트 카메라와 같은 클라이언트 디바이스에서 이루어질 수 있는 반면, 더 발전된 애플리케이션은 엔터프라이즈 데이터 센터 내에서 실행할 수 있다.
- **데이터 보존** - 모델 교육 또는 딥 러닝을 위한 반복 추론에 사용하는 데이터는 보관 또는 재사용 목적으로 유지된다. 데이터 보존은 매우 중요한 단계이다. 데이터를 아카이빙하면 모델을 다시 만들 수 있고, 미래에 대비한 통찰력을 위해 데이터를 마이닝할 수 있다.

이러한 단계는 딥 러닝을 지원하기 위한 인프라스트럭처를 구축할 때 고려해야 하는 일련의 설계 원칙을 토대로 한다.

- **성능 및 확장성**: 성능은 확장으로 인해 저하될 수 없다. 컴퓨팅, 스토리지 또는 네트워킹에 관계없이 각 구성 요소를 비례적이고 독립적으로 확장하여 시스템이 컴퓨팅, IO 및 네트워킹 병목 현상이 발생하지 않고 워크로드와 함께 원활하게 확장될 수 있도록 해야 한다.
- **유연성**: AI 시스템은 데이터를 중심으로 구축된다. 이러한 역동성에 내재된 현실은 AI 생태계가 발전함에 따라 소프트웨어, 분석 기술 및 활용 사례가 불가피하게 변화를 맞이하게 되지만, 조직의 데이터는 상대적으로 일정하게 유지된다는 것이다. 시스템은 비즈니스 요구의 변화에 맞추어 발전할 수 있는 유연성을 유지하면서도 오래 지속되는 데이터 스토리지를 지원해야 한다.

- **엔터프라이즈 데이터 관리:** 딥 러닝에 활용되는 데이터는 일반적인 용도가 아님에도 불구하고 엔터프라이즈 데이터이며, 이와 같이 관리되어야 한다. 보안, 데이터 보호, 규정 준수 및 기타 기존 데이터 관리 관련 우려 사항은 딥 러닝 데이터에도 적용된다. 이러한 환경에 구축된 스토리지 솔루션은 엔터프라이즈 데이터 관리를 위해 기존 정책 및 절차와 긴밀하게 통합되어야 한다.

머신 러닝과 딥 러닝에 대한 논의는 자연스럽게 컴퓨팅에 중점을 두지만, 이러한 솔루션이 데이터에 대한 새로운 사고방식을 강요한다는 것은 분명하다. 딥 러닝은 데이터를 관리하고, 분석하고, 저장하는 방법에 대한 다른 사고방식을 필요로 한다.

딥 러닝 환경의 데이터

딥 러닝 워크플로의 데이터 특성은 다음과 같이 대부분의 다른 IT 애플리케이션과 다르다.

- **데이터는 주로 정형화되지 않은 형태로,** 이미지, 오디오, 자유 형식 텍스트나 시계열 데이터 스트림으로 구성된다. 딥 러닝 환경을 위한 스토리지 아키텍처는 비정형 데이터에 최적화되어야 한다. 또한 스토리지는 SMB, NFS, HDFS, S3 및 HTTP와 같은 여러 데이터 액세스 프로토콜을 지원하여 운영 유연성을 최대한 높여야 한다.
- 비디오 및 엣지 센서와 함께 **데이터 규모**가 급격히 증가하고 있으며, 단기간에 이루어지는 분석으로 생성되는 수 테라바이트의 고해상도 콘텐츠가 포함되는 경우 특히 더 그렇다. 추후 분석 또는 재교육을 위해 이 데이터를 유지하면 페타바이트 규모의 스토리지가 필요할 수 있다. 딥 러닝을 통해 신뢰할 수 있는 통찰력을 도출하려면 분석할 데이터에 대한 심층적인 기간별 레코드가 필요하다. 이 환경의 스토리지 솔루션은 중단 없이 간단하게 스케일 아웃할 수 있어야 한다.

- **데이터 사용량**은 학습 파이프라인 각 단계의 다양한 요구에 따라 크게 달라진다. 예를 들어 데이터 정리 또는 레이블 지정은 교육 또는 실시간 추론을 위해 동일한 데이터를 클러스터에 공급하는 프로세스와는 매우 다른 성능을 요구한다. 이 파이프라인의 한쪽 끝은 기존 로컬 스토리지, DAS 또는 중형급(mid-tier) 스토리지로 충족할 수 있다. 파이프라인의 다른 쪽 끝은 최신 처리 기술에 대응하는 처리량 및 엔터프라이즈 기능을 필요로 한다.
- **데이터는 도처에서 추출된다.** 딥 러닝 애플리케이션에는 매우 다양한 데이터 소스가 있다. 분석 또는 모델 생성을 위한 데이터는 엣지, 클라우드 네이티브 애플리케이션, 음성 서비스 및 서버 로그 집계 애플리케이션에서 추출될 수 있다. 다양한 소스에서 데이터를 수집하도록 스토리지를 설계해야 한다.
- **데이터 모델의 수명주기:** AI에는 기업의 전체 AI 워크플로에 맞게 조정된 스토리지 아키텍처에서 데이터를 유지하기 위해 고성능에서 딥 아카이브 스토리지를 포괄하는 일관된 관리 툴 세트가 필요하다. 마찬가지로, 기존 데이터를 새로운 AI 기능을 위한 입력 데이터로 전환하려면 IT 조직이 기존 스토리지에 새로운 솔루션을 배포할 수 있는 데이터 관리 툴이 필요하다.

이러한 높은 수준의 특성은 딥 러닝을 위한 데이터 관리 솔루션을 선택할 때 실제 고려 사항이 된다. 이 데이터는 여전히 “엔터프라이즈 데이터”이며 하드웨어 및 소프트웨어 장애 및 보안 침해로부터 보호되어야 하고, 효율적으로 관리되어야 한다는 점을 염두에 두는 것이 중요하다.

조직에서 배포하는 딥 러닝 유형은 이러한 워크플로를 지원하는 스토리지 아키텍처에도 영향을 미친다. 예를 들어 미디어 및 엔터테인먼트, 제조 및 자동차 업계와 같은 업계에서 많이 사용되는 이미지 인식은 CNN(Convolutional Neural Network) 및 DNN(Deep Neural Network) 애플리케이션을 기반으로 한다.

CNN은 매우 반복적인 여러 단계를 통해 이미지를 분류하고 인식하는 방법을 학습하는 신경망의 한 유형이다. 교육 및 인식 과정에 있는 CNN의 데이터 액세스 패턴에는 기본 스토리지 어레이에 대한 수많은 작은 블록 읽기 액세스에 맞게 조정된 스토리지 아키텍처가 필요하다.

실례를 들어보면, Dell EMC와 NVIDIA가 수행한 벤치마킹에서는 Dell EMC Isilon F800 스토리지 시스템을 NVIDIA Tesla V100 GPU 여러 개로 구성된 NVIDIA DGX-1 서버에 결합했다. 각 GPU는 5,000개 이상의 병렬 스레드를 실행했으며, 이는 GPU당 평균 703개의 파일을 동시에 읽는 것과 같다³. 딥 러닝 시스템과 결합된 스토리지 시스템이 처리 요소가 데이터를 기다리는 것을 지연시키지 않으면서도 규모에 맞는 탁월한 동시성으로 데이터를 제공할 수 있어야 한다는 점이 중요하다.

이것은 하나의 예에 불과하다. 다른 딥 러닝 시스템은 저마다의 요구 사항이 있다. 예를 들어 금융 부정 행위 탐지를 위한 실시간 패턴 인식을 제공하는 지능형 시스템에는 초고성능 블록 스토리지가 필요할 수 있다. 이러한 제약이 있는 애플리케이션은 Dell EMC PowerMax Series와 같이 처리량이 높고 레이턴시가 짧은 블록 스토리지 어레이를 통해 더 원활한 지원을 받을 수 있다.

블록 크기, 파일 I/O 패턴 및 규모에 대해서도 마찬가지로 고려해야 할 사항이 존재한다. 중요한 점은 머신 러닝 및 딥 러닝을 위한 데이터를 제공하는 것이 다른 엔터프라이즈 워크로드와는 매우 다르다는 것이다. 딥 러닝을 위한 데이터를 관리하려면 단일 네임스페이스를 통한 계층화 및 일관된 툴 세트를 통한 간편한 관리를 사용하여 높은 동시성과 다차원적인 성능을 지원하도록 설계된 솔루션을 구축해야 한다.

Dell EMC: 딥 러닝을 위한 스토리지 제공

AI의 힘은 효율적이면서 성능 기준에 맞는 데이터 제공을 통해서만 발휘될 수 있고, 머신 러닝 및 딥 러닝 애플리케이션을 위한 스토리지 솔루션을 설계할 때에는 학습 파이프라인 단계마다 성능, 확장성 및 동시성에 대한 요구 사항이 다르므로, 이에 대해 여러 요소를 고려해야만 한다.

³백서: 딥 러닝을 위한 Dell EMC Isilon 및 NVIDIA DGX-1 서버, <https://www.dell.com/ko-kr/collaterals/unauth/white-papers/products/storage/Dell EMC Isilon and NVIDIA DGX 1 servers for deep learning.pdf>

이와 동시에 딥 러닝 워크로드의 단계별 요구 사항을 모두 충족하기 위해 원활하게 계층화되고 확장되는 스토리지 아키텍처를 구축하는 것이 합리적이다.

Dell EMC Isilon 제품군은 엔터프라이즈 딥 러닝의 전체 수명주기를 지원하는 스토리지 기능을 제공할 수 있는 견고한 기반을 공급한다. 교육에서 학습, 배포, 그리고 궁극적으로 장기 아카이빙 요구에 이르는 워크플로를 따른다.

Dell EMC Isilon OneFS

모든 스토리지 시스템의 힘은 기본 운영 체제 소프트웨어에 있다. Dell EMC Isilon OneFS 운영 체제는 Dell EMC Isilon 스케일 아웃 NAS 스토리지 솔루션의 기반이 되는 인텔리전스를 제공한다.

OneFS의 강력한 기능은 모든 인공 지능 워크플로의 핵심에서 데이터 스토리지를 최적화하고 간소화한다. 이 소프트웨어는 원활한 계층화를 제공하는 한편, 단일 네임스페이스를 공급하고, 데이터 배치를 관리하고, 감지된 트래픽 패턴을 기반으로 각 어레이의 성능을 최적화 및 튜닝하고, 운영 중단 없는 비례적인 스토리지 확장을 지원한다. Dell EMC Isilon OneFS 운영 체제에서 이러한 각 기능을 제공한다.

스토리지 관리의 단순성으로 인해 데이터 과학자는 기본 스토리지 인프라스트럭처의 세부 사항에 대해 걱정할 필요 없이 머신 러닝 프로세스 관리에 집중할 수 있다. 또한 이러한 단순성으로 IT 관리자는 머신 러닝 및 딥 러닝의 요구를 포괄하는 유연하고 효율적인 스토리지 솔루션의 적절한 조합을 구축할 수 있다.

- **통합 Data Lake** – 분석 워크플로 전반에 걸친 데이터를 한 곳으로 통합하여 데이터 분석 파이프라인을 간소화한다.
- **멀티 프로토콜 지원** - “한 번의 저장으로 여러 번 사용”하는 방식을 지원하는 데이터 분석을 통해 민첩한 대응 능력을 높인다.
- **엔터프라이즈 데이터 거버넌스** - 기본 제공되는 복원력 및 보안 기능으로 데이터를 보호한다.
- **원활한 계층화** – 동일한 클러스터의 올플래시, 하이브리드 및 아카이브 노드 간의 스토리지 계층화를 통해 경제적인 페타바이트 단위 확장 및 대규모 데이터 세트에 대한 액세스가 가능하다.

- **지능형 캐싱** - 데이터를 사용하는 워크로드를 기반으로 스토리지 시스템의 캐싱 특성을 동적으로 튜닝하는 기능을 제공한다. Isilon OneFS 캐싱은 딥 러닝 워크플로에서 중요한 성능 특성인 동시 읽기 성능을 대상으로 한다.
- **비례적 확장성** - Isilon 시스템이 딥 러닝 구현의 특징인 동시 병렬 워크로드를 제공하면서 일관된 성능을 유지할 수 있도록 한다.
- **별도의 설정이 필요 없는 DevOps 및 As-a-Service 지원** - 기업이 개발, 테스트 및 운영 데이터 환경을 구축하거나 동일한 Isilon 클러스터 내의 여러 액세스 존을 통해 테넌트가 명확히 분리된 여러 운영 데이터 환경을 제공할 수 있다.

이 소프트웨어는 Dell EMC Isilon Series에 내재된 전반적인 환경과 인텔리전스를 관리한다. 단순한 관리 용이성, 어레이의 견고한 성능 및 확장성 특성의 결합은 Isilon을 딥 러닝 워크로드를 위한 매력적인 플랫폼으로 만든다.

Dell EMC Isilon: 머신 러닝 및 딥 러닝을 위해 설계된 플랫폼

Dell EMC Isilon 스토리지 제품군의 최상위 계층은 Isilon F800 올 플래시 스케일 아웃 NAS이다. Dell⁴에 따르면 F800은 업계 최고 수준의 성능과 용량을 제공한다. F800은 단일 4U 새시에서 초당 15GB의 집계 처리량으로 최대 250,000 IOPS의 성능을 발휘하며, 전체 252노드 클러스터에서 최대 15.75M IOPS 및 초당 945GB의 성능을 발휘한다.

용량을 살펴보면 Isilon F800은 단일 네임스페이스에서 수십 테라바이트 규모의 스토리지부터 수십 페타바이트까지 중단 없이 확장할 수 있다. Isilon은 최대 85%의 스토리지 효율성과 데이터 스토리지 용량 요구 사항을 최대 3:1 비율로 줄일 수 있는 중복 제거 및 압축 기술을 제공하여 솔루션의 유효 용량을 증가시킨다.

Isilon F800은 딥 러닝 컴퓨팅 노드의 잘 공급된 상태를 유지한다. 60개의 고성능 SSD와 8개의 40Gbps 이더넷 연결을 갖춘 이 기계는 딥 러닝에 필요한 높은 수준의 동시성으로 일관된 성능을 제공하도록 설계되었다. 단순히 일관된 성능을 제공하는 것 외에도 Isilon F800은 Isilon 하이브리드 및 Isilon 아카이브 노드로 계층화하여 페타바이트 확장성을 쉽게 관리할 수 있다.

⁴ Dell EMC Isilon F800 사양: <https://www.dellemc.com/ko-kr/collaterals/unauth/data-sheets/products/storage/h15963-ss-isilon-all-flash.pdf>

공동 개발을 통해 PowerEdge C4140, DSS 8440 및 NVIDIA DGX-1과 같은 NVIDIA Tesla V100 GPU 가속 서버와 Isilon F800의 기능이 결합된 Dell EMC 레퍼런스 아키텍처보다 성능이 더 뛰어난 것으로 입증된 솔루션은 없다. 이러한 솔루션에 대한 벤치마크의 경우, ResNet-50 벤치마크에서 GPU 사용률 97%를 기준으로 하여 초당 선형 이미지 성능이 8~72개 GPU로, 최대 72개 GPU를 달성하는 성능을 보여주었다⁵.

이러한 벤치마크 수치는 현재 사용 가능한 최고 성능의 딥 러닝 컴퓨터 중 하나에서 프로세서가 병목 현상을 일으키지만, Dell EMC Isilon F800은 데이터가 완전히 공급되는 상태를 유지한다는 것을 입증한다.

Dell EMC PowerMax: 고성능 블록 스토리지

데이터 수집, 데이터 준비 및 운영 추론 도중 실시간 대응 속도를 위해 매우 낮은 레이턴시를 요구하는 AI 워크플로의 단계와 특정 머신 러닝 및 딥 러닝 알고리즘이 있다.

블록 스토리지 솔루션 중 Dell EMC PowerMax Series는 현재 사용 가능한 최고 성능의 스토리지 아키텍처 중 하나로, 이러한 시나리오를 효과적으로 지원하도록 설계되어 있다. PowerMax는 포괄적인 NVMe에 구축되어 170만~1천만 IOPS(각각 PowerMax 2000 및 PowerMax 8000)에서 300ms 미만의 레이턴시를 제공하며 브릭당 최대 13TB의 용량을 제공한다.⁶

Dell은 오늘날 기업에 배포되는 가장 까다로운 실시간 AI 워크로드를 지원할 수 있는 PowerMax를 제공한다.

⁵ Dell EMC 백서. 딥 러닝을 위한 Dell EMC Isilon 및 NVIDIA DGX-1 서버. 2018년 11월.
https://www.dell.com/ko-kr/collaterals/unauth/white-papers/products/storage/Dell EMC Isilon and NVIDIA DGX_1_servers_for_deep_learning.pdf

⁶ Dell EMC PowerMax 스펙 시트: <https://korea.emc.com/collateral/data-sheet/h16739-powermax-2000-8000-ss.pdf>

Dell EMC: 전체 스택 딥 러닝

스토리지와 컴퓨팅은 딥 러닝 환경에서 서로 얽혀 있다. 딥 러닝을 위해 효과적으로 설계된 인프라스트럭처는 데이터 관리와 관련된 모든 복잡성을 통해 균형, 상호 운용성, 성능 및 유연성을 제공한다. 구현 전반에 걸쳐 높은 수준의 유사성에도 불구하고 하나의 올바른 방법이란 없다. 모든 배포와 각 환경은 조금씩 다르다.

머신 러닝 및 딥 러닝 워크로드를 배포하는 데에는 다양한 옵션이 있다. 단계마다 다른 데이터 액세스뿐만 아니라 다른 컴퓨팅 솔루션도 필요하다. AI 실무자는 베어 메탈 서버, 가상 머신 또는 Docker와 유사한 컨테이너에서도 워크로드를 실행하도록 선택할 수 있다.

Dell EMC는 단순히 개별 요소를 딥 러닝 인프라스트럭처에 제공하는 것 외에도 IT 실무자가 신속하게 배포할 수 있는 솔루션을 구현하기 위해 노력하고 있다. Dell EMC는 문제를 해결하기 위한 요소를 결합한 Ready Solutions 및 RA(Reference Architecture)를 통해 아키텍처 관련 의사 결정을 간소화하고 배포 시간을 단축한다. Dell EMC는 기업이 데이터 분석 및 AI 솔루션을 확장하고 특정 워크로드 요구 사항에 맞게 조정하는 데 도움이 되는 솔루션 구성 지침을 제공한다.

Ready Solutions와 RA는 적절한 규모의 Dell PowerEdge 서버와 Dell EMC 네트워크 스위치, Isilon 스토리지 및 솔루션에 최적화된 소프트웨어 스택을 결합한다. Ready Solutions는 AI 이니셔티브를 가속화하도록 최적화되고 검증된 하드웨어 및 소프트웨어 스택으로서 새로운 솔루션을 설계하는 데 소요되는 시간을 6~12개월이나 단축시킨다. Dell EMC Ready Solutions for AI에 추가적인 성능과 이점을 제공하는 것은 Dell Technologies 컨설팅, 지원, 자금 조달 및 구축 서비스다. 이러한 서비스는 모두 함께 작동하여 원활한 솔루션 구축을 보장한다.

레퍼런스 아키텍처는 Dell의 고객 및 솔루션 파트너를 대상으로 테스트와 검증을 거친 스택이다. Ready Solutions는 Dell에서 직접 주문할 수 있지만, RA는 IT 실무자가 Dell Technologies의 검증된 제품을 기반으로 동급 최고 수준의 솔루션을 구축할 수 있도록 지원하는 것을 목표로 한다.

표 1: 사용 가능한 Ready Solutions 및 레퍼런스 아키텍처 예시

유형	솔루션	핵심 요소	주요 파트너
Ready Solutions for AI	인텔 기반 딥 러닝	Isilon H600 PowerEdge R740xd PowerEdge C6420	인텔
	NVIDIA 기반 딥 러닝	Isilon F800 PowerEdge R740xd PowerEdge C4140	NVIDIA
	Hadoop 기반 머신 러닝	Isilon H500/H600 PowerEdge R640	Hortonworks
AI에 대한 레퍼런스 아키텍처	딥 러닝을 위한 Dell EMC Isilon 및 NVIDIA DGX-1	Isilon F800 NVIDIA DGX-1	NVIDIA
	딥 러닝을 위한 Dell EMC Isilon 및 PowerEdge C4140	Isilon F800 PowerEdge C4140	NVIDIA
	딥 러닝을 위한 Dell EMC Isilon 및 DSS 8440	Isilon F800 DSS 8440	NVIDIA
	알고리즘 거래를 위한 Dell EMC Isilon 및 PowerEdge R940	Isilon F800 PowerEdge R940	인텔

출처: Moor Insights & Strategy

결론

데이터는 많은 조직에서 가장 전략적이고 차별화되는 자산이 되었다. AI 기술은 데이터를 해석하고 활용하는 방식에 혁신을 일으키고 있다. 기업은 이러한 현실을 지원하기 위해 지식을 쌓고 인프라스트럭처를 구축하는 데 많은 투자를 하고 있다.

이와 동시에 인공 지능은 머신 러닝이든 딥 러닝이든 IT 조직이 데이터 및 스토리지 아키텍처에 대해 기존의 엔터프라이즈 워크로드를 지원하는 것과는 다른 사고방식을 갖추도록 요구한다. 데이터의 특성이 다르고, 분석의 복잡성이 다르고, 해당 데이터에 대한 소비자의 요구가 다르다. 가속 컴퓨팅 노드에 계속해서 데이터를 공급하는 능력이 가장 중요하다. Dell EMC Isilon을 기반으로 하는 AI 솔루션은 이러한 요구 사항을 정확하게 충족하도록 설계되었다.

딥 러닝 솔루션을 구축하려면 신중하게 생각해야 하며 이러한 새로운 환경의 요구를 이해하는 기술 공급업체와 협력하여 해당 환경을 이용하는 IT 실무자의 고충을 덜어주는 데 필요한 폭넓은 대상 솔루션을 제공해야 한다.

Dell EMC는 이러한 파트너의 좋은 예다. 딥 러닝은 데이터를 최우선으로 하며, Dell EMC는 데이터 센터, 프라이빗 클라우드 및 퍼블릭 클라우드 및 엣지 네트워크의 데이터 관리 분야에서 세계적인 선두 주자 중 하나이다. 폭넓은 Dell EMC AI 포트폴리오는 이에 대한 차별화된 포지셔닝으로 고객의 요구를 충족하는 최상의 환경을 설계할 수 있도록 지원한다. Dell EMC는 고객 데이터를 관리하고 보호하는 광범위한 스토리지 포트폴리오와 AI를 통한 성공에 최적화된 서비스 및 솔루션을 보유하고 있다.

Dell EMC 전용 웹사이트에서 자세히 알아보기:

<https://www.dell EMC.com/ko-kr/solutions/artificial-intelligence/index.htm>

본 백서에 대한 중요 정보

기고자

Moor Insights & Strategy 수석 분석가 [Steve McDowell](#)

게시자

Moor Insights & Strategy 설립자, 사장 겸 선임 분석가 [Patrick Moorhead](#)

문의

당사로 [연락](#)해 주시면 본 보고서와 관련한 문의에 Moor Insights & Strategy가 신속히 답해 드리겠습니다.

인용

공인된 언론과 전문 분석 기관에서 본 백서를 인용할 수 있지만, 문맥상 적절한 경우에 한해 작성자 이름, 작성자 직함 및 “Moor Insights & Strategy”를 명시해 인용해야 합니다. 언론 및 전문 분석 기관 이외의 개인 또는 기관에서 인용하려는 경우 Moor Insights & Strategy의 사전 서면 동의를 받아야 합니다.

라이선싱

모든 참고 자료를 포함한 본 문서의 소유권은 Moor Insights & Strategy에 있습니다. Moor Insights & Strategy의 사전 서면 동의 없이 어떤 형태로든 본 발행물을 복제, 배포 또는 공유할 수 없습니다.

고지 사항

본 백서는 Dell의 의뢰로 작성되었습니다. Moor Insights & Strategy는 본 백서에 언급된 여러 첨단 기술 기업에 연구, 분석, 조언 및 컨설팅을 제공합니다. Moor Insights & Strategy의 직원 어느 누구도 본 문서에서 언급한 기업에 지분 관계가 없습니다.

법적 고지 사항

본 문서에 제시된 정보는 정보 제공용이며 기술적으로 부정확하거나 누락된 내용이 있거나 식자 오류가 있을 수 있습니다. Moor Insights & Strategy는 그러한 정보의 정확성, 완전성 또는 적합성에 대해 모든 보증을 부인하며 그러한 정보의 오류, 누락 또는 부적합성에 대해 책임을 지지 않습니다. 이 문서는 Moor Insights & Strategy의 의견으로 구성되어 있으며 사실의 진술로 해석되어서는 안 됩니다. 본 문서에 언급된 의견은 예고 없이 변경될 수 있습니다.

Moor Insights & Strategy는 미래에 발생할 일에 대한 정확한 예측이 아니라 지향성 지표로서 예측 정보 및 향후 전망에 대한 진술을 제시합니다. 당사의 예측 및 향후 전망에 대한 진술은 미래에 대한 현재 당사의 판단에 근거한 것으로, 위험 요소와 불확실성으로 인해 실제 결과와 상당한 차이가 있을 수 있습니다. 본 문서의 발행 당시 당사의 견해만 반영된 이 같은 예측 및 향후 전망에 대한 진술에 지나치게 의존하지 않도록 주의해야 합니다. 당사는 새로운 정보 또는 미래에 발생할 일을 반영하여 이 같은 예측 및 향후 전망에 관한 진술을 수정하거나 그러한 수정 결과를 공개적으로 발표할 의무가 없습니다.

©2019 Moor Insights & Strategy. 기업 및 제품 이름은 참고용으로만 제공되며, 각각 해당 소유자의 상표일 수 있습니다.