



SEAGATE

백서

최적화된 아키텍처

빅 데이터 및 AI/ML 작업 부하용
스토리지 아키텍처 설계

목차

- 03 서론
- 04 데이터 아키텍처에서 ML 및 AI의 중요성
- 05 증가하는 복잡한 데이터 관리 요구 사항 해결
- 06 우선 순위 변화에 대처하기 위한 스토리지 아키텍처 정의
- 08 AI/ML 워크플로의 다양한 단계를 위한 여러 데이터 흐름 필수 요소
- 11 보다 원활한 AI/ML을 간소화할 미래의 패브릭
- 13 결론



서론

번영 중인 기업 데이터 경제는 모든 것을 포착하고, 이를 효율적으로 관리하며, 잠재력을 활용하는 세 가지 필수 요소를 기반으로 합니다.

인공 지능(AI), 머신 러닝(ML), 대규모 데이터를 생성 및 사용하는 기타 빅 데이터 애플리케이션과 시스템이 급격하게 증가하고 있습니다. 최근 설문 조사에 따르면, 기업 중 절반 이상은 현재 ML을 사용 중이며 대부분의 기업들이 몇 년 이내로 사용할 예정인 것으로 나타났습니다. 빅 데이터 및 AI/ML 작업 부하를 위해서는 방대한 양의 정형 및 비정형 데이터를 처리하고 분석하기 위한 기능이 필요하지만, 하이브리드 클라우드 및 멀티 클라우드 전략으로 인해 온프레미스, 오프프레미스 및 엣지 등 이러한 데이터가 여러 위치에서 생성됨에 따라 기업은 데이터 관리, 현재 및 미래의 스토리지 용량, 효율성과 관련된 수 많은 문제를 재고하여 이러한 데이터 소스 및 애플리케이션을 관리할 수 있는 아키텍처상의 이점을 정의해야 합니다.

데이터 전략을 혁신하기 위해서는 올바른 데이터 아키텍처가 핵심입니다. 이상적인 데이터 아키텍처는 무한대의 규모와 효율성에 최적화되므로, 필요한 데이터 및 필요성을 인식하지 못한 데이터 모두를 캡처할 수 있습니다. 이러한 아키텍처는 데이터의 원활한 이동에 최적화된 아키텍처이므로, 장애물 없이 적절한 시간에 올바른 위치로 이동하는 것이 가능하여 조직이 최상의 가치를 창출하는 곳에서 최대한의 잠재력을 활용할 수 있습니다.

그러나 리소스 부족과 현재 기술 및 솔루션의 제한으로 인해 이를 달성하는 것은 어렵습니다. 이 백서에서는 빅 데이터 및 AI/ML 작업 부하에 적합한 확장성과 함께 비용 대비 효율적인 IT 인프라를 원하는 기업을 위한 하이브리드 클라우드 및 스토리지 아키텍처에 대해 알아봅니다.



데이터 아키텍처에서 AI 및 ML의 중요성

사실상 AI와 ML은 최신 기업 운영의 모든 측면을 바꾸고 있습니다. 자연어 처리(NLP) 및 이미지 인식은 소비자와 기업 모두가 머신과 상호 작용하는 방식에 영향을 줍니다. 새로운 스마트 시스템은 AI 지원 솔루션을 활용하여 제조를 혁신하고 있습니다. 자율 주행 자동차는 운송과 배송의 근본적인 변화를 약속합니다. 또한, 많은 기업들은 이러한 새롭고 혁신적인 시스템을 한 두 개가 아니라 여러 개 동시에 실험하고 있습니다.

Deloitte의 엔터프라이즈에서의 AI 현황 리포트¹에 따르면, 응답자 중 67%는 현재 ML을 사용 중이며 97%는 내년 중에 ML을 도입할 계획이라고 응답했습니다. IDC의 전망²에 따르면, 2025년까지 410억 개의 IoT 장치가 매년 79제타바이트(ZB) 규모의 데이터를 생성할 것이며, 최근 Seagate Rethink Data 보고서에서 IDC는 코어 데이터 센터 및 엣지에서 생성되는 데이터 중 44%가 분석, 인공지능 및 딥 러닝으로부터 생성될 것으로 전망했습니다³. 확인하지 못한 데이터 증가로 인한 문제는 늘 있어 왔지만, IT 혁신 기업인 451 Research에 따르면 2020년의 한 해 동안에도 데이터가 약 25% 증가할 것으로 예상됩니다.⁴

AI/ML은 이전에는 대규모 데이터 집합 속에 숨겨져 있거나 추출할 수 없었던 가치를 데이터 과학자와 분석가가 확보할 수 있는 길을 열어주고 있습니다. 그러나 IT 아키텍트에게는 이는 바로 비용, 거버넌스 및 규정 준수를 최적화하는 동시에 대규모로 데이터 가용성, 무결성, 기밀성 및 내구성을 유지할 수 있는 방법을 고려해야 한다는 새로운 당면 과제도 나타나고 있습니다. 여러 위치에 걸친 데이터 확산, 복제 및 분산은 이러한 진화 중인 환경과 관련된 일부 당면 과제에 불과합니다.

이러한 공통적인 문제를 완화하고 급증하는 데이터 물결의 강력한 힘을 보다 효율적으로 활용하려면, IT 아키텍트는 계층화된 스토리지 방식, 진화 중인 새로운 패브릭, 오픈 소스 오케스트레이션 도구 등의 여러 요소를 반드시 고려하여 구조적 이점을 공략해야 합니다.

갈수록 복잡해지는 데이터 해결 관리 요구 사항

10년도 더 전에 등장한 클라우드 배달 모델은 특정 작업 부하에 적합한 클라우드 또는 데이터 센터 리소스 및 관련 복잡성뿐만 아니라 광범위한 공용 및 개인 플랫폼에 대한 수문을 열었습니다. 어떠한 솔루션도 획일적이지 않으며, 단일 기업 내에서도 작업 부하별로 다른 경우가 많습니다.

기업은 어떤 온프레미스 및 개인 클라우드 시스템과 기능이 공용 클라우드 스토리지 서비스를 보완할 수 있는지, 그리고 필요에 따라 비클라우드 비즈니스 애플리케이션에 적응할 수 있는지 결정해야 합니다. 비즈니스 운영 및 고객 대면 서비스 모두에서 AI/ML이 부상함에 따라 기업은 데이터 센터와 스토리지 결정에서 미래 지향적인 접근법을 취해야 할 것입니다. 데이터 센터 및 스토리지 아키텍처는 하이브리드와 계층형 스토리지 구현의 활용이 증가하고 세분화 및 컴포저블 기능의 중요성이 높아짐에 따라 최근 10년 동안 크게 발전했습니다.

최근 대부분의 기업은 확장성, 데이터 가용성, 성능, 보안 및 비용의 최적화를 위해 공용 플랫폼과 개인 플랫폼을 혼합하는 경향을 보이고 있습니다. 451 Research의 2020년도 Cloud Confidence Report⁵에 따르면, 절반 이상의 기업(57%)이 통합 방식으로 온프레미스 시스템과 오프프레미스 클라우드/호스팅형 리소스를 통합하는 하이브리드 IT 환경으로 이동하고 있습니다.

451 Research 리포트에서 살펴본 현재와 미래 기업의 공통 인프라 분석은 다음과 같습니다.

- **49%**의 응답자는 현재 온프레미스에서 비클라우드 인프라를 사용하고 있으며, **25%**는 향후에 이를 사용할 계획이다.
- **36%**는 현재 온프레미스 개인 클라우드를 구축 중이거나 향후에 구축할 예정이다.
- **32%**는 SaaS(서비스형 소프트웨어) 및 호스팅형 애플리케이션으로 현재 이동 중이며, **37%**는 향후에 그러할 예정이다.
- **31%** 및 **32%**는 각각 현재와 미래에 호스팅형 개인 클라우드를 사용 중이거나 사용할 예정이다.
- **28%**는 현재 IaaS(서비스형 인프라) 공용 클라우드를 이용 중이며, **33%**는 향후에 이용할 예정이다.
- **22%**는 현재 PaaS(서비스형 플랫폼)를 이용 중이며, **29%**는 향후에 통합할 예정이다.
- **15%**는 호스팅형 비클라우드 환경을 현재 이용 중이거나 향후에 그러할 예정이다.



우선 순위 변화에 대처하기 위한 스토리지 아키텍처 정의

클라우드 및 다양한 클라우드의 변형이 지속적으로 발전함에 따라 스토리지 고려 사항도 변경될 수 있습니다. 최적화된 스토리지 계층화를 사용하면 지정된 애플리케이션, 분석 또는 작업 부하를 위해 적시에 데이터에 액세스할 수 있습니다.

역사적으로 용량의 측면에서 최상의 미디어는 항상 성능 측면에서 가장 비효율적이었고 아니면 그 반대의 경우였습니다. 경제적인 관점에서 AI/ML을 설계하는 아키텍트는 소규모의 매우 빈번히 액세스하는 데이터가 계층의 상위에 위치하고, 대규모의 자주 액세스하지 않는 데이터가 최대한 하위 계층에 위치하도록 해야 합니다.

HDD와 SSD 기술이 통합된 계층형 하이브리드 스토리지는 사실상의 표준으로 자리잡고 있습니다. 이러한 접근 방식을 통해 HDD 및 SSD 모두의 전반적인 시스템 성능이 향상될 뿐만 아니라 IT 아키텍트는 가격, 성능 및 용량의 이상적인 조합을 달성하기 위해 각각의 구성 요소를 선별할 수 있습니다.

일반적으로 대용량 스토리지 기기는 주로 용량 및 대역폭을 위해 AI/ML 작업 부하를 처리하는 데이터 센터의 중심 구성 요소가 될 것입니다. 그러나 AI/ML 애플리케이션의 경우 IOPS 집약적 작업 부하를 위해 소규모의 비휘발성 미디어 장치가 필요할 수도 있습니다. 궁극적으로 솔루션을 혼합하면 예상치 못한 가변 수요와 컴퓨팅 버스트를 처리할 수 있을 뿐만 아니라 비용 효율성을 향상하기 위해서도 도움이 될 수 있습니다.

데이터 센터의 진화에서 또 다른 중요한 요소는 스토리지뿐만 아니라 메모리, CPU 및 GPU와 관련하여 새롭게 등장하고 있는 지속적인 세분화 추세입니다. 가장 기본적인 수준에서, 구성 가능 아키텍처에는 특정 유형, 구성 또는 물리적 구성 요소의 개수를 할당하여 특정 IT 요구 사항에 적합하도록 맞춤형 가상 머신을 생성하는 기능이 포함됩니다.

IT 아키텍트에게 있어 세분화는 CPU 및 마더보드 제조사가 정의하는 폐쇄형 아키텍처에 더 이상 종속되지 않음을 의미합니다. 대규모의 오픈 소스 커뮤니티가 소프트웨어 및 하드웨어 모두를 위한 솔루션을 개발 중이므로 머지않아 편리한 데이터 관리가 달성될 것입니다.

궁극적으로 NVMe-oF 등과 같은 적절한 패브릭이 지원되고 오픈 소스 오케스트레이션 도구로 보완되는 세분화된 컴포저블 아키텍처는 다양한 클라우드 환경과 데이터 센터 유형에서 운영되고 광범위한 AI/ML 애플리케이션을 구축하는 기업에 필요한 유연성을 제공할 가능성이 높습니다.

AI/ML 작업 부하의 최적화를 위한 고려 사항 및 솔루션

공용 클라우드를 지지하는 사람들은 동적이며 사용자 친화적이고 대량의 데이터를 수용할 수 있는 등의 여러 이점이 있다고 주장합니다. 이론적으로 기업은 필요한 만큼만 지불하면 됩니다. 그러나 실제 AI/ML 사용 사례에서 공용 클라우드는 비용이 급증할 수 있습니다. “체크아웃은 할 수 있지만, 떠날 수는 없다”는 *Hotel California* 노래 가사와 같은 상황에 직면하게 되는 것입니다. 즉, 성공적인 AI/ML 애플리케이션에 필요한 규모의 데이터를 추출 및 보존하기 위해서는 막대한 비용이 드는 경우가 많습니다.

개인 클라우드, 하이브리드 클라우드 또는 멀티클라우드 환경에서는 막대한 양의 데이터를 저장, 활용 및 검색할 수 있는 능력이 가장 중요합니다. 강력하고 비용 대비 효율적이며 AI/ML 통합을 원활하게 하는 아키텍처의 토대는 스토리지 기기 수준에서 시작됩니다.

그 어느 때보다도 장치의 용량이 커야 합니다. 그러므로 소프트웨어의 기능이 필수적이어서, 매우 많은 수의 장치에서 소프트웨어가 업데이트되어야 하며, 이는 종종 암시적 가정 및 명시적 코딩 패턴 모두에 대한 근본적인 가정을 깨뜨릴 수 있습니다. 또한, 소프트웨어를 지속적으로 업데이트하여 데이터 센터 전체를 검색하고 사용자가 단일 네임스페이스에 더 많은 데이터를 통합할 수 있도록 해야 합니다.

엔터프라이즈 데이터 센터에는 대용량 장치와 확장성이 우수한 시스템이 필요합니다. HDD 중심의 접근 방식은 다양한 AI/ML 작업 부하에 적합하며, 이는 인간/데이터 과학자의 사용자 경험과 라벨링 프로세스 및 소규모 무작위 읽기 모두에서 원활하게 작동합니다.

대규모의 무작위 I/O가 포함된 AI/ML 작업 부하의 경우에는 반드시 장치 용량 이상으로 확장되어야 합니다. HDD 중심의 스토리지 아키텍처의 경우 IOPS 집약적 작업 부하와 관련하여 메타데이터, 대역폭, 용량 검색 등 여러 요구 사항에 적합하지만, 단순히 4K 블록 주소를 지정하는 것과 달리 기존 스토리지와 워드 주소 지정이 가능한 비휘발성 스토리지 기기와 결합해야 합니다.

그러나 이로 인해 데이터 확산, 데이터 복제 및 데이터 사일로의 문제가 다시 나타날 수 있습니다. 이를 완화하려면 데이터 센터에서는 반드시 데이터를 편리하게 검색하는 동시에 적절한 미디어를 활용할 수 있어야 합니다.

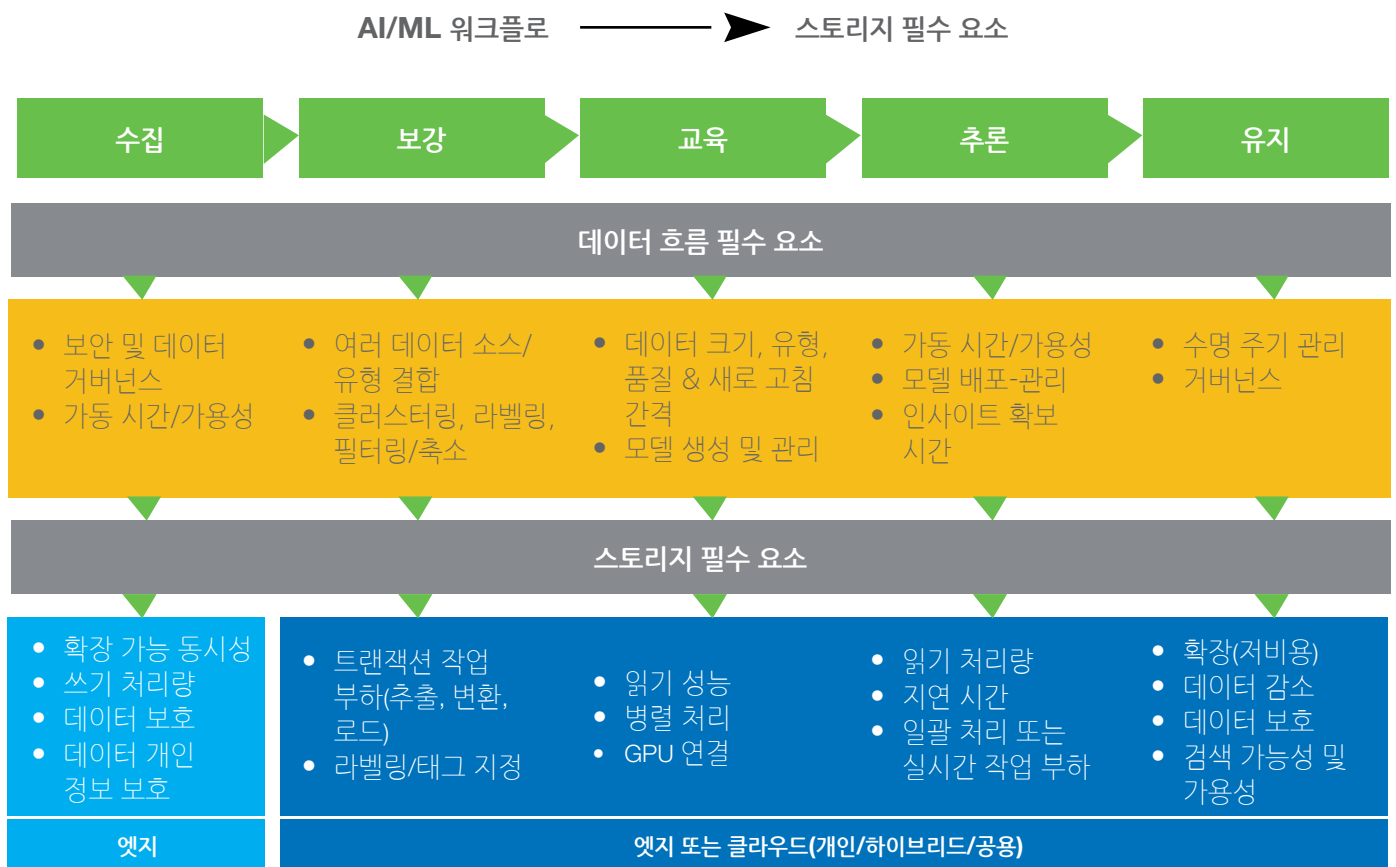
AI/ML 워크플로의 다양한 단계를 위한 여러 데이터 흐름 필수 요소

AI/ML 작업 부하는 선형적으로 진행되지 않으며, 작업 부하라기 보다는 워크플로라는 용어로 사용하는 것이 보다 적절할 수 있습니다. 데이터 수집에서 의사 결정, 장기 또는 단기 스토리지에 이르기까지 데이터가 이동하는 광범위한 단계 사이에 점을 연결하는 것은 단순한 문제가 아닙니다.

아키텍처 및 데이터 관리 요구 사항은 AI/ML 시스템의 기능 및 지침에 따라 매우 달라집니다. 예를 들어, 기본 이미지 인식 시스템은 자율 주행 자동차의 요구 사항과 요구 사항이 매우 다릅니다.

모든 AI/ML 워크플로에 적합한 단일 권장 사항은 없지만, 여러 광범위한 단계로 나눠 IT 아키텍트가 적절한 패브릭, 오케스트레이션 소프트웨어 및 스토리지 솔루션에 관한 주요 결정을 내리는 데 도움이 될 수 있습니다.

AI/ML 워크플로에는 일반적으로 데이터 수집, 보강, 추론 및 보존 단계가 포함됩니다. IT 아키텍트는 이러한 각 단계의 데이터 흐름 및 스토리지 필수 요소를 고려해야 합니다.



수집 단계에서 데이터 흐름 관점의 주요 초점은 보안 및 거버넌스, 가용성 및 가동 시간입니다. 이 단계에서 데이터 소스는 인프라 및 스토리지를 결정하는 데에서 가장 중요합니다. 예를 들어, IoT 애플리케이션에서는 다양한 출처에서 유입되는 매우 작은 파일을 집계해야 합니다. 이미지 데이터와 센서의 시계열 데이터가 형식이 다른 것처럼, 데이터 형식이 보통 다를 것입니다. 인프라는 다양한 센서에서 대량의 정보 흐름을 가져와 이를 이해할 수 있는 기능을 지원해야 합니다.

보강 단계에서 데이터는 여러 소스와 여러 유형에서 결합된 후 클러스터링, 라벨링 및 필터링이 수행되고 축소되어 가장 관련성이 높은 정보의 하위 집합을 생성한 후 AI 학습에서 사용됩니다. 이 단계에서 스토리지 시스템은 다음 단계를 위한 태그 지정 및 라벨링을 수행하는 동시에 데이터 추출, 변환 및 로드와 관련된 트랜잭션 작업 부하를 관리할 수 있어야 합니다.

다음으로 학습과 추론은 별도의 작업 부하로 수행됩니다. 학습을 위해서는 데이터를 가져와(일부 경우 대량의 데이터가 필요) 해당 정보를 사용하여 모델을 학습시켜야 합니다. 모델이 구축 및 배포되며 아마도 엣지 장치에서 추론 단계가 시작됩니다. 이를 위해서는 실시간 데이터를 활용하여 인사이트를 생성하고 의사결정을 수행해야 합니다. 여기에서는 인사이트 확보 시간, 가동 시간 및 가용성이 특히 필수적인 요구 사항입니다.

보존 요소는 완전히 다른 요구 사항으로, 재교육을 위해 데이터를 검색하거나 규정 및 규제 준수 목적으로 몇 달 또는 몇 년 동안 데이터를 보관 또는 저장해야 할 수 있습니다. 이 단계는 아마 비용 측면에서 가장 어려운 단계일 겁니다. 데이터 수집 및 학습에서 클라우드를 사용하는 것이 중요하지만, 데이터를 추출하거나 개인 클라우드로 이동하기 위해서는 *Hotel California* 문제와 같이 비용이 급증할 수 있습니다. 사실, 데이터 보관 요구 사항은 기타 아키텍처 구성 요소에 대한 사고 프로세스의 핵심이 되어야 하며, 그렇지 않으면 솔루션 비용이 증가할 수 있습니다.

AI/ML의 당면 과제는 이러한 각 단계의 요구 사항이 상대적으로 고유하다는 점입니다. 아래는 다양한 작업 부하를 설계할 때 고려해야 하는 몇 가지 이점과 트레이드 오프를 설명하는 두 가지 샘플 사용 사례입니다.

사용 사례: 스마트 제조

일반적으로 스마트 제조 애플리케이션용 학습 모델에 대한 요구 사항은 그렇게 높지 않습니다. 이 경우 아키텍처의 지배적인 측면은 데이터 과학자의 시간입니다.

학습 단계에서의 주요 우려 사항은 학습 프로세스 전체에서 데이터 반복을 최대한 빠르게 수행하는 것입니다. 학습에서 클라우드를 사용하는 것은 매력적일 수 있습니다. 탄력적인 클라우드를 사용하면 과학자들은 원하는 정도의 컴퓨팅을 매우 빠르게 실행할 수 있기 때문입니다.



그러나 추론의 측면에서는 여러 가지 이유로 온프레미스를 선택하는 것이 더 좋을 수 있습니다. 효율적인 공장 운영을 위해서는 실시간 데이터 흐름이 필요하며, 이를 위해 클라우드 아키텍처만 이용하는 것에는 상당한 문제가 있습니다. 한 가지 문제는 가동 시간으로, 클라우드가 추론 또는 의사결정에 연료를 공급하고 어떤 이유로든 클라우드로의 연결이 끊어지면 공장이 가동을 멈추게 됩니다. 온프레미스 인프라를 사용하면 (최소한으로 하드웨어와 관련해) 어쩔 수 없이 벌어지는 상황에 대한 제어가 향상됩니다. 또한, 일반적으로 추론과 관련된 컴퓨팅 요구 사항은 훨씬 낮으므로 하드웨어 비용도 더 저렴합니다.

사용 사례: 자율 주행 자동차

이 사용 사례에서는 데이터 요구 사항이 훨씬 더 높습니다.

수집 단계를 살펴보면, 실시간 비디오를 포함한 데이터를 클라우드로 가져오기 위한 경제성을 달성하기가 매우 까다롭습니다.

또한, 차량에서 실수를 식별하고 실수로부터 학습하기 위해서는 데이터가 지속적으로 재순환되어야 합니다. AV 시스템에서 추론 하드웨어는 일반적으로 엣지의 차량 자체에 위치합니다. (단순히 이 단계에 관련된 볼륨으로 인해, 데이터 학습을 클라우드 또는 엣지에서 수행해야 하는지에 대해서는 아직 논란이 있습니다.)

이 사용 사례에서 엣지 위치와 클라우드 또는 코로케이션 간의 대역폭을 고려할 때 유선을 따라 데이터를 이동하는 것은 비용과 시간이 매우 많이 소요될 수 있습니다. 엣지에 위치한 장치의 물리적인 모바일 어레이는 다음을 해결하는 데 도움이 될 수 있습니다. 차량의 경우 데이터 집계를 위해 엣지에 플래시 기반 솔루션 또는 SSD가 탑재될 수 있으며, 이후에 셔틀 장치를 통해 데이터가 전송된 후 보다 효율적인 수집을 위해 데이터 센터로 이동될 수 있습니다.

해당 사용 사례와 관련해서는 수집 및 유지해야 하는 데이터의 양, 추론의 측면에서 인사이트 확보 시간, 모델 재학습에 필요한 주기와 관련된 전체 워크플로를 이해하는 것이 중요합니다. 이러한 구성 요소를 따로 생각해서는 안 됩니다.



객체 기반 소프트웨어 아키텍처 선택

모든 AI/ML 시스템의 공통점 하나는 데이터 클러스터링, 필터링, 축소 또는 라벨링을 수행하여 데이터를 의미 있는 방식으로 수정해야 한다는 것입니다. ML 세트는 로우 이미지만을 사용해서는 학습할 수 없으며, 초기 라벨링에서는 여전히 사람이 개입해야 하는 요소가 있습니다.

이러한 이유로 소프트웨어 아키텍처는 라벨 기능을 강력하게 지원할 뿐만 아니라 대용량을 위한 경제성이 달성되어야 합니다. 개인 및 하이브리드와 공용 클라우드에 걸친 객체 기반 또는 객체 스토리지는 대규모 비정형 데이터의 경제적인 스토리지를 위한 최상의 인터페이스로 AI/ML에 적합한 추상화를 제공합니다. 기본적으로 보다 확장적인 특성과 작업 부하에 적합한 기저 데이터 구조, 추상화 및 가정 때문에 데이터 집약적 AI/ML 시스템을 보유하는 기업들의 객체 스토리지로의 이동이 증가하고 있습니다.

또한, 객체 데이터는 메타데이터에서 정보를 수집하기 위한 매우 강력한 도구이기도 합니다. 데이터가 객체 스토리지 시스템에 저장되면 검색 및 필터링이 매우 편리하며, HTML과 호환성이 매우 높은 데이터 관련 작업을 수행하기 위한 여러 기능이 제공됩니다.

보다 원활한 AI/ML을 간소화할 미래의 패브릭

일반적으로 AI/ML 시스템에는 대규모의 DRAM 풀이 필요하지만, 싱글 서버에서 최근 DRAM의 최대 용량은 약 2TB입니다. 자율 주행 자동차의 예를 다시 살펴보면, 차량이 하루에 해당하는 운전을 마치고 6TB 크기의 데이터를 가지고 돌아오면, 데이터는 여러 서버에 걸쳐 있으므로 궁극적으로 하이퍼컨버전스 상황이 발생합니다.

현재는 시스템에서 적극적으로 작동하는 모든 CPU에서 공유할 수 있도록 해당 데이터를 세분화된 메모리에 저장하는 것이 가장 쉽습니다. CPU는 액세스를 공유할 수 있어, DRAM에서 스토리지급 메모리로 오가며 페이지를 매길 수 있습니다. 그러나 그 사이를 오가며 이동하는 데에 비용이 발생합니다. 미래의 새로운 저대기 시간 패브릭은 가용 메모리를 확보하고 그러한 장치에 매우 가까운 풀에 배치하여 장치 간에 공유하는 것이 가능할 것입니다. 이러한 진화는 보다 신속하고 유능한 머신 러닝에서의 다음 단계가 될 수도 있습니다.

그러나 이러한 반복에는 아직 몇 년이 더 걸릴 수 있습니다. 새로운 패브릭은 개발하기가 매우 복잡합니다. CPU 중심 컴퓨팅에서 메모리 중심 컴퓨팅으로의 Z세대 혁명은 약 2012년부터 나타나고 있습니다. 향후 10년 이내에 Z세대 패브릭은 하이퍼바이저 및 일부 기업 솔루션의 일부일 뿐만 아니라 에코시스템의 표준으로 자리 잡을 것입니다. CXL 2.0 및 3.0은 2025년 중반까지 연구소에서 나타나기 시작하고, 향후 10년 내에 CXL 2.0 및 3.0이 데이터 센터로의 배포가 시작될 수 있습니다.



유연성, 보안 및 최적화에 매우 중요한 개방형 환경

개방형 환경은 혁신 및 성장을 위한 유연성과 기회를 추가적으로 제공합니다. 이는 오래 시간에 걸쳐 소프트웨어에서 사실로 증명되고 있으며 하드웨어의 경우에도 그렇습니다. 그리고 기업에서 지속적으로 AI/ML 기술을 실험하기에 특히 그러할 것입니다.

개방형 아키텍처에는 컴포저블 컴퓨팅, 네트워킹 및 스토리지 리소스의 통합 컬렉션이 포함됩니다. 확장 가능한 개방형 하드웨어 인프라를 통해 기존 시스템 내에서의 보다 쉬운 협업과 혁신이 가능합니다. 동료 평가를 통해 오픈 소스 프로젝트를 다양한 관점으로 검토할 수 있으며, 사용 사례와 목적을 공유하는 다양한 작성자가 함께 모여 솔루션, 보안 취약성 등을 논의 및 식별할 수 있으므로 궁극적으로는 최적화를 가속화합니다.

최근의 IT 전문가들은 높은 수준으로 확장이 가능하고 공급업체 고유 구성 요소에 대한 종속성이 감소한 하드웨어 아키텍처로 서서히 이동하고 있습니다. OCP(Open Compute Project)와 같은 단체는 협력을 통해 예를 들어 새시의 높이 또는 길이에 대해 논의하고 궁극적으로 이상적인 솔루션을 식별하므로 하드웨어 표준이 진화하는 프로세스를 민주화합니다.

소프트웨어 및 하드웨어 모두의 관점에서 오픈 소스는 기업이 AI/ML 및 기타 빅 데이터 애플리케이션과 관련하여 변동이 임박한 데이터 흐름을 관리하는 데 도움이 될 수 있습니다. 대규모 데이터 세트 관련 작업을 공개적으로 논의할 수 있으며, 아키텍트들은 최적화 활동과 관련한 공통적인 문제를 공유할 수 있습니다.

또한, 개방형 표준 하드웨어는 궁극적으로 하드웨어 아키텍처의 책임과 보안을 향상할 수 있습니다. 그리고 보안을 최우선으로 하여 시스템을 설계하는 것도 가능하게 해줍니다. 예를 들어, RISC-V International은 기본 보안 기능이 제공되는 차세대 프로세서를 위한 개방형 표준을 개발했습니다⁶. 이러한 RISC-V 프로세서의 오픈 소스 구현은 보안을 평가하고 보안이 향상된 시스템을 설계하기 위한 기준으로 사용할 수 있습니다. 또 다른 예는 OpenTitan⁷인데, 이는 오픈 소스, 실리콘 기반 RoT(Root of Trust) 프로젝트로 RoT에 대한 투명한 고품질 레퍼런스 설계 및 통합 지침을 개발한 최초의 프로젝트입니다. RISC-V 및 OpenTitan과 같은 개방형 하드웨어 표준 및 오픈 소스 하드웨어 프로젝트를 통해 데이터 저장을 위한 하드웨어 보안을 재고할 수 있습니다.

보다 세분화된 수준에서 AI/ML 워크로드에 적합한 오픈 소스 객체 스토리지 소프트웨어, 특히 이를 위해 필요한 대용량 장치에 최적화된 소프트웨어는 효율성을 유지하면서 하드웨어의 특성을 처리하기 위한 비용과 유연성이라는 두 가지 주된 이유로 필수적입니다.



결론

AI/ML의 발전은 혁신의 기회를 제공하는 동시에 데이터 관리 및 저장과 관련한 당면 과제를 제시합니다. 사용 가능한 클라우드 및 멀티클라우드 환경의 수가 증가함에 따라 AI/ML 작업 부하를 위해 설계할 때 고려해야 하는 복잡성이 더욱 증가하고 있습니다.

이미 막대한 양의 데이터가 생성되고 있으며 향후 몇 년 동안 계속 증가할 것이지만, 비용 제약으로 인해 일반적으로 상대적으로 적은 양의 데이터만 저장되고 있습니다. 방정식의 소프트웨어 측면에서 쿠버네티스와 같은 시스템을 통한 자율 관리의 발전으로 필요에 따른 애플리케이션의 확장 및 축소가 가능하지만, 전통적으로 하드웨어 측면은 유연하지 않습니다.

컴포저블 기능, 세분화 및 오픈 소스/개방형 표준의 새로운 동향은 이러한 흐름을 바꾸는 데에 도움이 될 수 있습니다. 컴포저블 기능을 이용할 수 있는 미래에는 스토리지, 메모리 및 컴퓨팅 풀을 필요에 따라 확장하여 진행 중인 소프트웨어 혁신을 보완해야 합니다. AI/ML의 맥락에서, 아키텍트는 서버를 장치에 대한 가용성 경로로 고려하는 것이 아닌, 성능과 용량을 위해 필요한 서버와 스토리지의 정확한 비율을 구현할 수 있어야 합니다.

Seagate는 하이브리드 클라우드 옵션 및 통합의 모든 부분에 걸쳐 설계하는 방법을 잘 이해하고 있습니다. 귀사의 AI/ML 요구 사항에 맞춤형 솔루션에 대해 자세히 알아보십시오.



참조 자료

1. https://www2.deloitte.com/content/dam/insights/us/articles/6462_state-of-ai-in-the-enterprise/DI_State-of-AI.pdf
2. <https://www.idc.com/getdoc.jsp?containerId=prUS45213219>
3. <https://www.seagate.com/our-story/rethink-data/>
4. 451 Research, "Balancing Hybrid Storage as Part of a Hybrid Cloud Data Management Strategy"
5. 451 Research, "Cloud Confidence Report," 2020년 3월
6. <https://riscv.org/>
7. <https://opentitan.org/>

리소스

고성능 컴퓨팅 솔루션: <https://www.seagate.com/solutions/high-performance-computing/>

빅 데이터 분석 솔루션: <https://www.seagate.com/solutions/data/big-data/>

더 자세히 알아보시겠습니까?

[seagate.com](https://www.seagate.com)에서 확인하십시오.

