

## 백서

# AI 확장 준비가 되었습니까? 코어 기아 현상으로 인한 어려움은 없습니다

후원: IBM

Peter Rutten  
2019년 5월

## IDC 의견

---

인공 지능(AI)과 연계된 비즈니스 기회는 성공 가능성이 매우 높습니다. 기업과 기타 다른 조직에서는 AI를 채택하지 않을 경우 비즈니스가 크게 실패할 수 있다는 것을 잘 알고 있습니다. 경쟁 기업은 과거에는 불가능했던 데이터와 성장 동력을 대량으로 확보하여 고객층을 만족시키고 있기 때문입니다. 이와 관련해서 일부 기업은 “우리는 AI가 필요하지 않아” 혹은 “AI는 그냥 거품일 뿐이야”라고 언급하고 있습니다. 하지만 이와 달리 엄청난 AI 계획이 전 세계적으로 업종과 무관하고 기업 규모와 무관하게 실현되고 있습니다.

수많은 업체의 LOB(현업 부서), IT 담당자, 데이터 과학자 및 개발자가 AI에 대한 학습을 계속 진행하고 있으며 사용 사례를 이해하고 자사의 비즈니스에 맞는 AI 전략을 정의하고 있으며, 초기 AI 계획을 시작하고, 머신 러닝(특히, 딥 러닝) 알고리즘을 통해서 새로운 통찰력과 추진력을 제공하는 최종 AI 애플리케이션을 개발하고 시험합니다.

이제 업체는 계획을 확장할 준비가 완료되어 새로운 난관에 대처하고 있습니다. 업체는 현업을 통해 표준적인 다목적 인프라를 사용하지 못한다는 것을 직접 경험을 통해 알고 있습니다. 또한 AI 학습(AI 모델의 학습) 및 AI 추론(학습된 모델을 사용하여 상황을 파악하거나 예측)에는 완전히 다른 유형의 컴퓨팅이 필요하다는 것을 확인했습니다. 그런데 다른 유형의 컴퓨팅은 어떤 것입니까? 배포 환경은 온프레미스입니까, 클라우드입니까, 아니면 하이브리드 클라우드 모델입니까?

기하 급수적으로 증가하는 엄청난 용량의 데이터를 탐색하는 AI 애플리케이션 특히, 딥 러닝 시스템은 대규모의 코어를 기반으로 한 강력한 병렬 프로세싱 기능이 절대적으로 필요하며 매우 중요한 부분입니다. 표준 CPU로는 이와 같은 AI 작업을 충분히 실행하지 못합니다. IDC 조사에 의하면 코어 용량의 측면에서 향후 몇 년 내에 실제 CPU 용량과 필요한 CPU 용량 간의 차이는 더욱 커질 것으로 확인되었습니다. 이 차이를 극복하기 위해서는 기존 인프라를 사용하여 시도를 해 본 AI 사용자와 확장 준비가 된 사용자는 멀티스레드 CPU와 GPU의 조합, 고속 상호 연결, 대용량 메모리, 첨단 I/O 기능 등을 통해서 필요한 병렬 프로세싱 성능을 확보하기 위해 인프라를 완전히 개조해야 합니다.

## 상황 개요

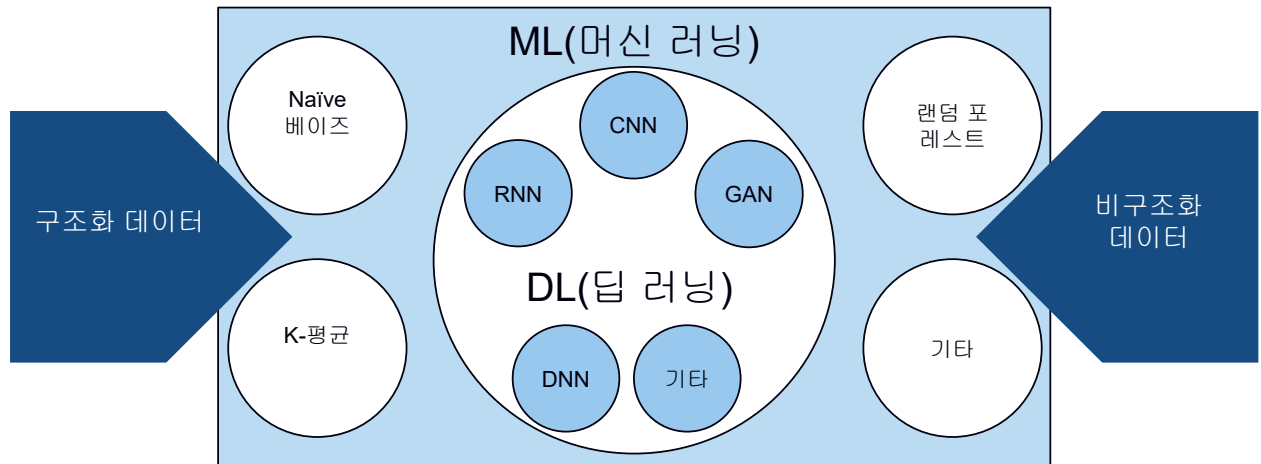
---

전 세계의 기업은 AI 워크로드가 제공하는 새로운 기회에 적극적으로 대응하고 있습니다. IDC에서는 AI를 질문에 응답하고 통찰력을 발휘하며, 권장 사항을 제공하기 위해 자연어 처리(NLP), 이미지/비디오 분석, 머신 러닝, 지식 그래프 및 기타 다른 기술을 사용하는 일련의 기술이라고 정의하고 있습니다. 가정을 세우고 사용 가능한 증거를 기반으로 예상되는 답변을 작성하는 이러한 시스템은 엄청난 물량의 증거를 입력한 후, 재학습 또는 사용자 감독에 따라 실수와 실패를 통해서 학습하고 적응하는 방법으로 학습시킬 수 있습니다.

머신 러닝은 사람이 프로그램 작업을 하지 않아도 컴퓨터 시스템이 지정된 작업에 대해서 행동을 학습하고 향상시키는 것이 가능한 AI 기술의 하부 집합입니다. 머신 러닝 모델은 엄청난 물량의 구조화 및/또는 비정형 데이터를 사용하여 작업(예를 들어, 사람의 얼굴 인식)에 필요한 “학습”이 완료된 것으로 판단될 때까지 반복되는 것으로, 자체 시험을 통해 시간의 흐름에 따라 개선이 가능한 알고리즘입니다. 그림 1은 머신 러닝의 하부 집합인 딥 러닝에 대해서 설명합니다. 일반적인 딥 러닝 구조로는 DNN(심층 신경 네트워크), CNN(컨볼루션 신경 네트워크), RNN(재귀적 신경 네트워크), GAN(생성적 적대 신경 네트워크) 등이 있습니다.

그림 1

## 머신 러닝과 딥 러닝



출처: IDC, 2019

인공 지능(AI) 소프트웨어 플랫폼의 사례는 다음과 같습니다.

- 대화형 AI 소프트웨어(예, 디지털 비서)
- 데이터에서 숨겨진 관계를 찾아내고 예측하는 예측 분석
- 텍스트에서 가치를 인식, 이해, 추출하기 위한 텍스트 분석 및 자연 언어
- 오디오, 음성, 대화에서 정보를 인식, 식별, 추출하기 위한 음성/대화 분석
- 이미지와 비디오에서 정보 인식, 식별, 추출하기 위한 이미지와 비디오 분석. 여기에는 패턴, 물체, 컬러 및 사람, 얼굴, 감정, 차량, 배경 등의 기타 다른 특성에 대한 인식 기능이 포함됩니다.

많은 기업이 AI 계획을 통해 현명하게 대처하고 있으며 제품화 규모 측면에서 AI 배포를 시작할 수 있는 준비가 완료된 단계에 도달했습니다. 다른 기업은 여전히 AI를 시도하는 단계에 있지만, 다른 쪽에서는 자신의 조직에 어떤 AI 애플리케이션이 좋을지 평가하고 있는 단계에 있습니다.

첫 번째 경우(배포 준비가 된 그룹), IDC는 기업, 정부 및 기타 조직에서 구현하기 시작한 다양한 인공지능 사용 사례를 관찰하고 있습니다. 현재 가장 일반적인 사용 사례 5개는 다음과 같습니다. 기업이 관련된 하드웨어, 소프트웨어, 서비스에 지출하는 비용을 기준으로 순위를 결정했습니다.

- **자동화 고객 서비스 에이전트.** 예를 들어 은행 업계에서는 이 같은 AI 애플리케이션이 학습 프로그램을 통해 고객 서비스를 제공합니다. 즉, 고객의 요구 사항과 문제점을 이해하고 은행을 보조하여 고객의 문제를 해결하는 데 필요한 시간과 리소스를 절감합니다. 이 같은 에이전트가 전체 업종에서 점차 폭 넓게 사용되고 있습니다.
- **판매 프로세스 권장 및 자동화.** 이 같은 AI 애플리케이션은 다양한 업종에서 사용되며, CRM(Customer Relationship Management) 시스템과 협력하여 고객의 상황을 실시간으로 파악한 후 판매 에이전트에게 관련된 조치를 권장합니다.
- **자동화 위협 인텔리전스 및 예방 시스템.** 이 같은 AI 애플리케이션은 정부와 산업계 전반에 위협 예방에 있어서 중요한 일부가 되어 인텔리전스 보고서를 처리한 후 정보를 추출하여 다양한 정보 사이의 관계를 파악하여 데이터베이스, 시스템, 웹 사이트 등에 대한 위협 상황을 식별합니다.
- **사기 분석과 조사.** 다른 업계와 마찬가지로 보험 업계에서도 이 같은 AI 애플리케이션을 다양하게 사용하여 규칙 기반 학습을 통해 사기 트랜잭션을 식별하며, 자동 학습을 통해 다양한 보험 관련 사기 행동을 식별합니다.
- **자동화 예방 유지 보수.** 제조업의 경우에 이 같은 AI 애플리케이션은 예상되는 공장 및 장비 고장을 정확하게 예측하여 가동 중지 시간과 유지 보수 비용을 절감하는 모델을 만드는 머신 러닝 알고리즘을 기반으로 합니다.

관심을 끄는 추가적인 기업의 인공지능 사용 사례는 다음과 같습니다(하드웨어, 소프트웨어, 서비스에 대한 지출 비용을 기준으로 순위를 결정함).

- 프로그램 상담 및 권장 시스템
- 진단 및 치료 시스템
- 지능형 공정 자동화
- 품질 관리 조사 및 권장 시스템
- IT 자동화
- 기업의 지식 노동자를 위한 디지털 비서
- 전문 쇼핑 상담 및 제품 권장
- 공급 및 물류
- 규제 인텔리전스
- 자산/자원 관리
- 자동화 클레임 처리
- 디지털 트윈/첨단 디지털 시뮬레이션
- 공공 안전 및 비상 대응
- 적응형 학습
- 스마트 네트워킹
- 화물 관리
- 제약 연구 및 개발

## 클라우드 대비 온프레미스

이 같은 사용 사례에 사용하는 애플리케이션은 업체에서 맞춤형으로 개발하거나, 상업용 AI 소프트웨어를 기반으로 하거나 AI SaaS 형태로 인도될 수 있습니다. IaaS 에서 클라우드 또는 하이브리드 클라우드인 경우, 맞춤형 개발 및 상업용 소프트웨어를 위한 배포 환경은 온프레미스이지만 온프레미스 환경은 공용 클라우드 환경과 상호 작용합니다.

다양한 배포 시나리오에서 솔루션과 관련해 고려할 사항은 다음과 같습니다.

- AI 모델 학습을 위해 필요한 대량의 데이터를 매우 강력한 성능을 사용하여 안전하게 처리. 딥 러닝 학습의 성능 요구 사항 중에는 GPU 와 고대역 데이터 입력을 사용하는 대규모 병렬 프로세싱 실행 기능도 있습니다.
- AI 모델이 추론 작업을 할 때 사용할 대량의 데이터를 아주 강력한 성능을 사용하여 안전하게 처리. 추론과 관련된 성능은 학습이 완료된 AI 모델로 들어오는 데이터를 처리하여 거의 실시간으로 AI 통찰 또는 의사 결정을 제공하는 기능을 의미합니다.

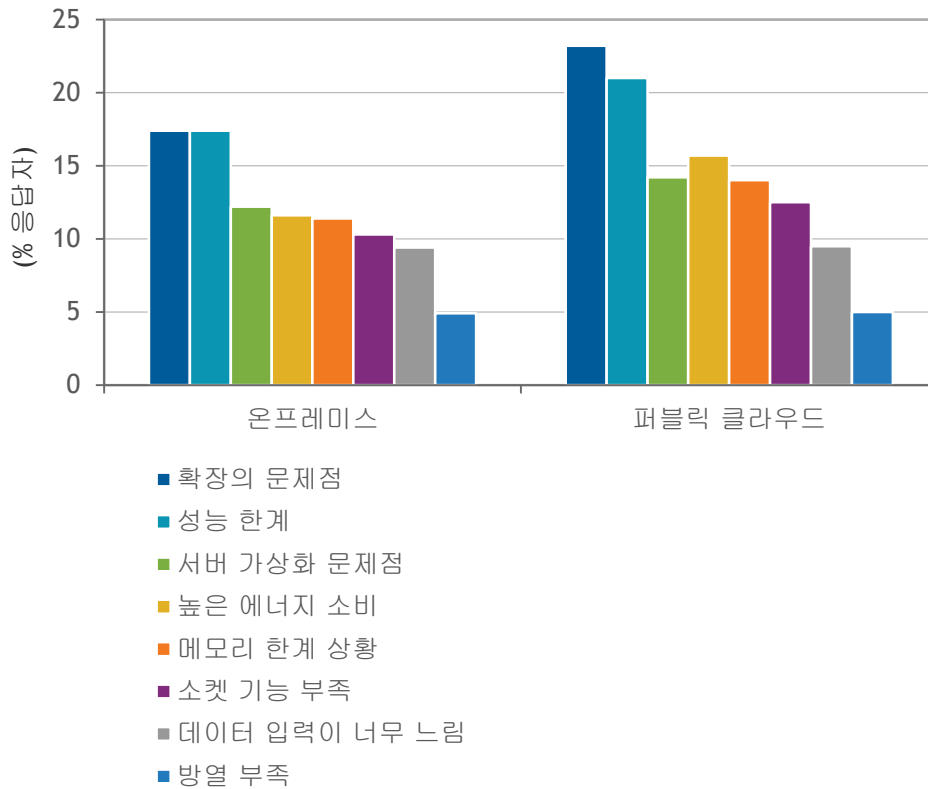
데이터 과학자와 개발자의 경우, AI 작업을 클라우드에서 시작하면 온프레미스 컴퓨팅을 조정하지 않아도 되기 때문에 편리할 수 있으며, 딥 러닝인 경우에는 일반적으로 가속 기능이 필요합니다. 대부분의 공용 클라우드에서는 일반적으로 오픈 소스 AI 스택을 통하여 가속 AI 클라우드 인스턴스를 사용할 수 있습니다. 가속 클라우드 인스턴스에서 AI 학습을 하는 경우에 클라우드 SP 가 최종 사용자에게 프로세서, 코프로세서, 상호 연결, 메모리 용량, I/O 대역폭 등의 사용 가능한 항목을 알려줍니다. 데이터 과학자가 개발할 수 있는 학습 모델의 속도와 품질을 궁극적으로 결정하는 이와 같은 구성 요소의 가장 최적화된 조합을 모든 클라우드 SP 가 지원하는 것은 아닙니다. 따라서 수많은 업체가 온프레미스 환경의 배포를 선택합니다.

지난 몇 년 동안 AI 시도를 통해 수많은 업체가 자신의 표준 인프라 또는 기본 클라우드 인스턴스로는 “한계에 부딪히는” 경험을 하게 된다는 것을 확인했습니다. 모델 학습 시간이 너무 오래 걸리며 추론 작업 속도가 너무 느렸습니다. IDC 조사에 의하면 77.1%의 응답자가 온프레미스 AI 인프라에서 여러 차례의 한계 상황에 직면하였으며, 90.3%는 클라우드에서 컴퓨팅 한계 상황이 발생했었다고 언급한 것이 확인됩니다.

그림 2 는 온프레미스와 클라우드에서 가장 빈번하게 발생하는 하드웨어 한계 상황의 유형을 보여줍니다. 대부분의 업체에서 이와 같은 난점이 복합적으로 발생하게 됩니다. 가장 빈번하게 나타나는 온프레미스 문제점을 기준으로 답변의 순위를 설정했습니다.

## 그림 2

### AI 사용 사례를 실행하는 서버 인프라에서 발생하는 최상위 하드웨어 한계 상황



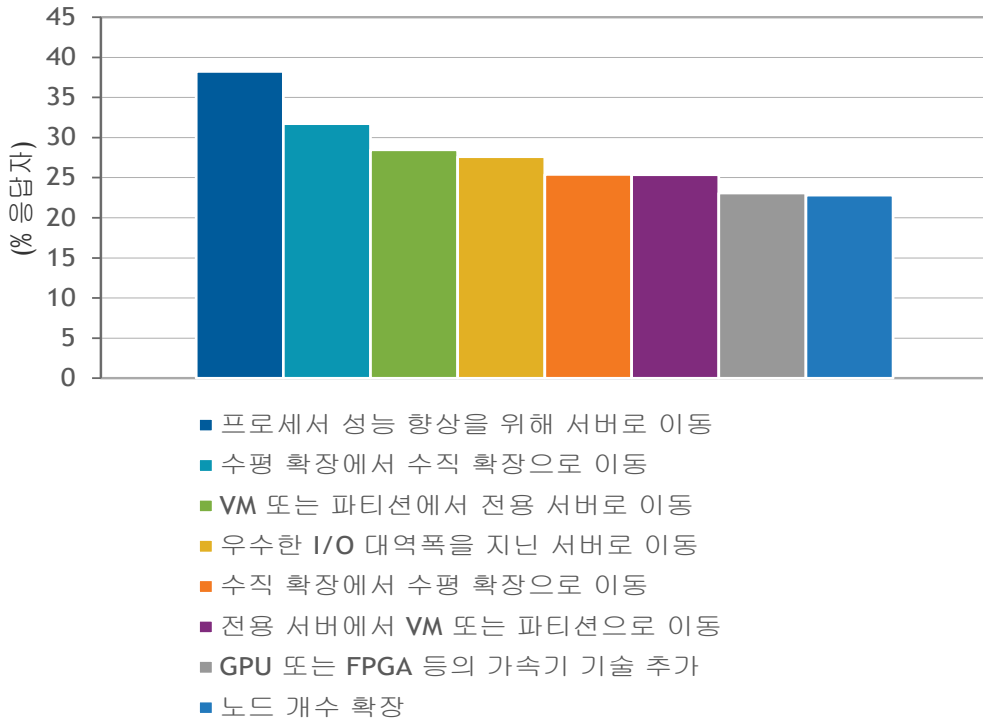
출처: IDC, 2019

IDC는 기업이 하드웨어와 관련된 한계 상황 극복의 일환으로 AI 인프라를 완전하게 개조하는 것을 관찰하였습니다. 일부 기업은 같은 해에 2회 개선 작업을 수행했습니다. 그림 3은 이 같은 인프라 변경의 특성을 보여주고 있으며 일부는 다른 기업의 행동과 완전히 반대되는 매우 모순적인 행동을 취하는 기업을 보여줍니다.

또한 그림 3을 보면 AI 성능 개선을 위해 가장 빈번하게 수행되는 인프라 변경 방법은 가장 왼쪽의 컬럼에서 설명하는 것과 같이 프로세서 성능을 더욱 강화시키는 것으로 나타납니다. 가장 일반적인 4번째 변경 방식은 AI를 위한 데이터 입력을 빠르게 개선하기 위해 I/O 성능을 향상시키는 것입니다. 노드의 개수를 확대하는 것처럼 가속기를 추가하는 것이 일반적인 방식입니다. 차트에서 설명하는 변화는 상호 배타적이지 않다는 점에 유의해야 합니다. 이를테면 일부 응답자는 수평 확장 및 가속 방식으로 인프라를 변경했습니다. 또한 일부 대립적인 인프라 개조 방식(수직 확장에서 수평 확장, 수평 확장에서 수직 확장으로)도 관찰되었으며, 이는 AI가 비즈니스 영역으로 진입한 첫 해부터 진행된 실험 방식 중 일부를 나타냅니다.

### 그림 3

#### AI 인프라 세대 교체의 특성



출처: IDC, 2019

#### 코어 기아 현상

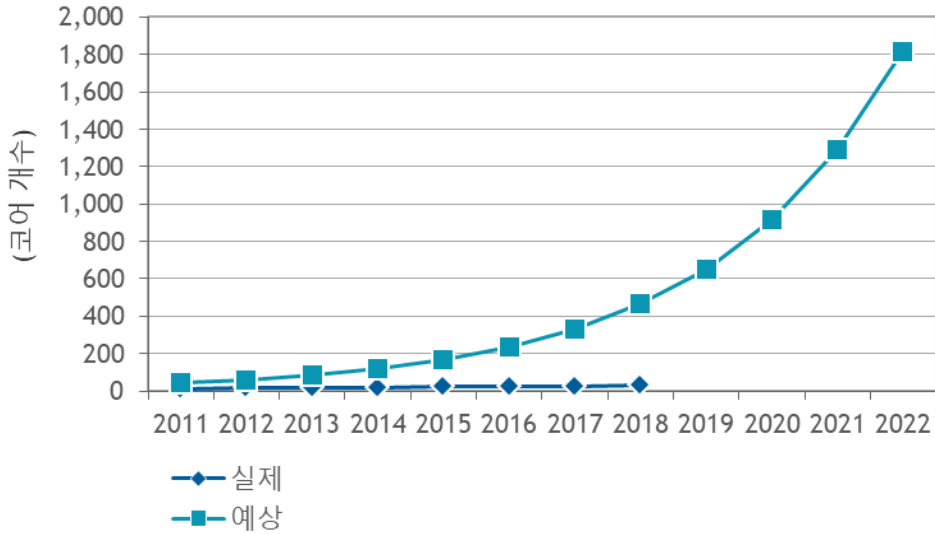
기업이 마주한 한계 상황의 근본 원인은 IDC가 “코어 기아 현상”이라고 일컫는 것 때문입니다. AI는 복잡한 수학과 통계적 컴퓨팅을 기반으로 합니다. 한 예로 이미지와 비디오 분석을 살펴 보겠습니다. 이미지는 행렬로 변환되고 각 픽셀은 수치로 표시됩니다. 수백만 개의 행렬과 해당 분류 상대가 상관 관계 분석을 위해 신경 네트워크로 입력됩니다. 그런 다음에 해당 행렬은 상호간 곱셈 연산을 거쳐서 올바른 결과(예, “개” 또는 “음료수 캔”)를 찾습니다.

이 과정을 빠르게 하려면 CPU가 제공할 수 있는 수많은 복수 코어 이상의 병렬 연산 처리를 해야 합니다. CPU는 직렬 연산 처리에 맞게 설계되면서 소재 물리학적 특성으로 인해 성능의 최대 한계점에 가까워졌습니다. 현재 모든 프로세서 제조 업체는 무어의 법칙에 도달했다는 점을 고려하여 수십 년 동안 일반적으로 향상해왔던 성능을 계속해서 개선하기 위해서는 CPU 개선 이외의 보완적인 접근법이 필요하다는 것에 동의하고 있습니다. 주요 이유는 이 같은 코어의 크기와 비용 조건에 따라 CPU 코어 개수가 제한적(수천 개가 아닌 수십 개)이라는 것입니다.

그림 4는 물리적인 제약으로 인하여 CPU 성능에 어떤 한계도 발생하지 않아 보이는 2011-2018년 사이의 시간에 따른 CPU 성능 개선과 2011-2022년 사이의 대수적 특성을 지닌 CPU 성능 곡선을 보여줍니다. CPU 성능 특성은 다음과 같이 가정합니다(현 상황과 다름). CPU 성능에 있어서 물리적인 한계 상황은 없으므로 CPU 성능 수요의 발전에 대한 통찰력만 제공합니다. 또한 그림 4는 CPU 성능 요구와 사용 가능한 실제 CPU 성능 간의 차이를 설명합니다.

## 그림 4

### 모든 워크로드에 대한 실제 및 예상하는 전 세계 코어 개수 요구 사항



출처: IDC, 2019

GPU(수천 개 코어 장착) 및 특수 제작 프로세서(ASIC, FPGA) 등장 이후 실제와 예상하는 CPU 성능 요건 간의 차이가 없어졌습니다. 가속기 금형은 수백 개 또는 수천 개의 코어로 구성된 대규모 병렬 아키텍처로 구성되어 있어서 필요한 병렬 컴퓨팅 성능을 효율적으로 제공합니다. 이 같은 코프로세서는 각 코프로세서 유형에 따른 다양한 장점과 함께 이산 성능을 강화하는 효과를 발휘합니다. 이와 동시에 엄청난 양의 데이터를 공급하는 이러한 보조 프로세서의 기술이 중요해졌습니다. 이에는 코프로세서 사이, CPU와 코프로세서 사이의 상호 연결, 메모리 용량 확장 및 고속 스토리지 기능 등이 포함됩니다.

IDC는 가속 서버의 전 세계 마켓이 31.6%의 연평균 성장률로 2022년에는 256억 달러로 성장할 것으로 판단합니다. 이들 서버는 온프레미스 및 클라우드 양측에서 GPU, FPGA, ASIC 기술을 사용하여 가속됩니다. 실제로 이 마켓은 매우 빠르게 성장하고 있기 때문에 IDC는 2021년에는 가속 컴퓨팅 마켓이 비가속 컴퓨팅 마켓을 잠식하기 시작할 것으로 예측합니다. 전 세계 서버 가치의 12%가 가속 컴퓨팅 마켓입니다.

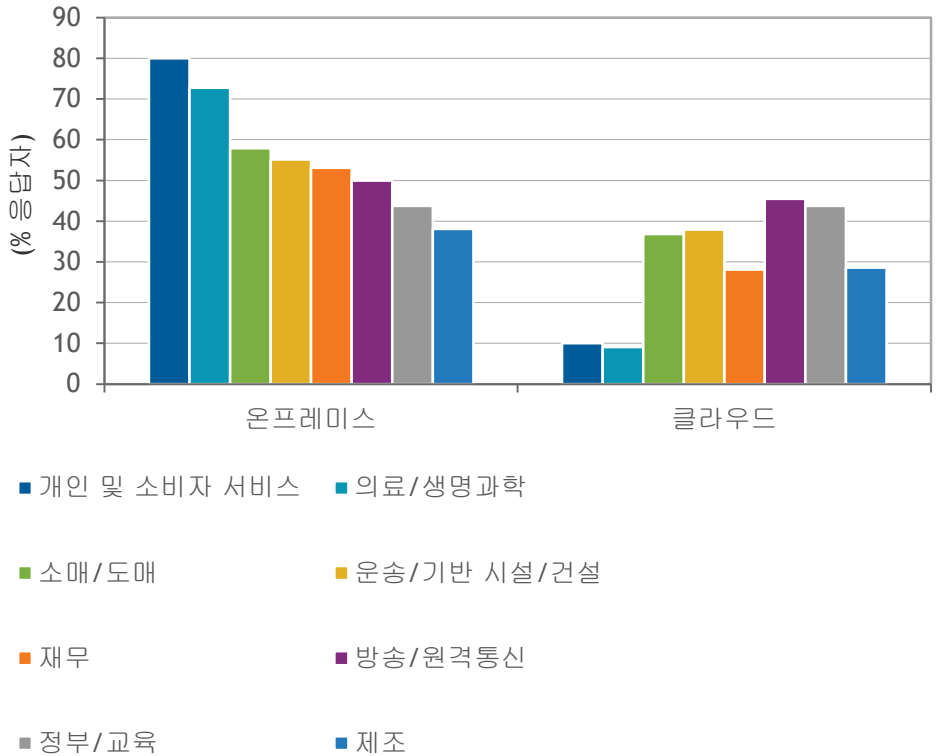
가속 컴퓨팅은 비교적 새롭게 등장한 기술입니다. IDC 조사에서 인프라 성능 강화를 위해 가속기를 활용하는 업체는 성능 강화 작업을 평균적으로 약 2-3년 동안 진행한 것으로 확인되었습니다. IDC는 컴퓨팅에서 이 방식이 계속 유지될 것으로 보고, 양자 컴퓨팅과 같이 완전히 새로운 유형의 컴퓨팅이 주류를 이룰 때까지 다수의 워크로드를 위한 가속 방식이 표준이 될 것으로 예상합니다. 그럼에도 대부분의 워크로드에서 기존의 가속 컴퓨팅 방식이 표준으로 지속될 것입니다.

최대 가속 기능을 요구하는 네트워킹, 암호화, 보안, 압축 등의 워크로드 이외로 가속을 위해 코프로세서 사용이 증가하는 워크로드에는 실시간 분석, AI 딥 러닝, Hadoop 및 데이터베이스, 비디오, 이미지, 음성 인식, 시뮬레이션과 모델링 및 AI 추론 등이 있습니다.

현재까지 온프레미스 환경이 가속 서버를 위해 선호하는 배포 방식입니다. 그림 5는 일부 업종에 대해 GPU 가속 서버의 온프레미스 대비 클라우드 배포 상태를 보여줍니다. 또한 그림 5는 온프레미스 GPU 최대 사용 사례가 개인 및 소비자 서비스 업종(80% 응답자)인 것을 보여주며, 반면에 클라우드 최대 사용 사례는 방송/원격통신 업종(45.5%)인 것을 나타냅니다.

그림 5

업종별 GPU-가속 서버의 온프레미스 대비 클라우드 배포 상태



출처: IDC, 2019

흥미로운 점은 향후 12개월 동안 온프레미스 배포가 증가할 것이며 클라우드 서비스 공급자는 이 같은 문제를 인지해야 합니다. 예를 들어, 가속을 위해 GPU를 사용하는 50.8% 기업은 현재 온프레미스 환경에서 진행한다고 답변했지만, 이 수치는 향후 12개월 이내에 54.2%로 증가할 것입니다. 이와 유사하게 ASIC 및 FPGA를 위한 온프레미스 사용은 약 5% 증가할 것입니다.

클라우드 서비스 공급자에게 더욱 놀라운 것은 상당히 적은 수의 기업이 공용 클라우드에서 실행 중인 특정한 가속 워크로드를 온프레미스로 되돌리는 작업(“송환”이라고 함)을 할 것이라는 사실입니다. 이들 업체가 워크로드를 온프레미스로 되돌릴 것이라는 답변을 통해 제시한 이유는 클라우드에서 경험했던 비용, 보안, 확장, 성능 등의 문제가 있습니다. 66%의 업체가 과거에 가속기가 필요한 워크로드를 클라우드에서 시작했지만 해당 워크로드를 온프레미스 데이터센터로 이동해 처리한다고 말합니다.

가속으로 인한 비용 대비 성능 이점은 업종에 따라 차이가 있습니다. 일부 업종에서 성능은 평균적으로 2배가 되지만 온프레미스 설비 투자(가속 서버 취득 비용) 또는 클라우드 운영 비용(가속 인스턴스 실행 비용)이 업종에 따라 26-33%로 증가하게 됩니다. 일부 가속화된 서버를 사용하는 실제 인프라에 따라 성능 이점이 향상될 수 있습니다. 그럼에도 불구하고 하드웨어 신생 기업은 물론이고 IBM®, Intel, AMD, Xilinx, NVIDIA 등에서도 진행하는 상당히 많은 수의 새로운 프로세서 및 코프로세서 연구에 따라 시간이 지나면서 AI가 두 배 이상의 성능을 요구하게 되는 시점에 투자 수익률이 크게 향상될 것으로 판단해야 합니다. IDC 조사에서 대부분 업체는 서버 공급자가 가속 기술을 설치하여 시스템 통합자 또는 VAR을 적게 사용하고, 일부 업체는 IT팀에서 가속기를 설치하는 것을 선호하는 것으로 확인되었습니다.

GPU를 위한 CUDA, OpenACC, OpenMP 등과 같은 가속기 프로그래밍에는 숙련된 직원이 있어야 합니다. 업체에서는 이 같은 목적을 위해 업종에 따라 평균 2.4-3.7명의 FTE 직원이 필요하다고



말합니다(가속기 유형은 무관). GPU 프로그램 작업은 간단하게 진행되는 경향이 있으며, FPGA는 조금 더 어렵고 ASIC는 개발 시간이 아주 오래 걸립니다. 대부분 기업은 가속기로 프로그램할 수 있는 속도에 대해서 상당히 만족한다고 말합니다. 따라서 이 부분은 업체에 부담 요소가 될 수 없습니다.

궁극적으로 기업은 가속 서버를 선택할 때 성능 강화 확보가 가장 중요한 목표라고 말합니다. 성능은 데이터 과학자가 AI 모델의 학습 완료를 얼마나 오래 기다려야 하는지, AI 모델이 얼마나 깊고 정확할지, 학습이 완료된 AI 모델이 얼마나 빠르게 추론 작업을 할 수 있는지 등에 직접적으로 연관됩니다. 따라서 성능이 강화되면 보다 빠르고 더 정확한 결과가 나옵니다. 새로운 기술 수요에 대한 중요도가 매우 낮기는 하지만 비용은 부차적입니다. 이 모든 것을 종합해 보면 기업은 AI 추진력이 필요하며 그에 따른 비용을 지불할 것이고, 해당 비용이나 새로운 기술 획득을 부담스럽게 생각하지 않습니다.

## 추론

AI 모델의 학습 대비 그 모델의 추론 요구 사항 사이의 차이점에 대해 일부 논란이 있습니다. IDC 조사에서 장기적으로 추론 작업에 사용하는 서버 인프라는 학습을 위한 서버 인프라보다 빠르게 성장할 것으로 확인됩니다. 여기에는 모든 유형의 머신 러닝을 위한 인프라가 포함됩니다. 이는 “전통적인” 머신 러닝(예, 최근접 또는 Naïve 베이즈 구분기)과 딥 러닝(예, 음성/대화 또는 이미지와 비디오 분석) 등이 있습니다. 전 세계 서버의 가치 측면에서 보면 2020년에는 추론이 학습을 추월할 것입니다. 추론 작업 역시 고성능 컴퓨팅이 필요하며 많은 경우에 가속 컴퓨팅이 필요합니다.

AI 추론 워크로드가 AI 학습의 경우와 다른 것에는 다양한 요인이 있습니다. 추론은 학습이 완료된 AI 모델에 입력되는 데이터에 적용합니다. 대부분의 애플리케이션은 거의 실시간으로 결과를 제공하기 위해 추론 작업을 최소한의 지연 시간으로 처리합니다. 데이터 용량이 작거나(일회성 이미지 인식 작업) 큰 경우(일반 도로에서 연속, 실시간 안면 인식 작업)도 있습니다. 추론의 전체 서버 인프라 요구 사항은 학습의 경우에 필요한 가속 유형(GPU 성능이 낮음) 또는 엄청난 반복 추론 작업을 수행하는 최대 규모 배포(FPGA 또는 ASIC)의 경우와도 다를 수 있다는 것이 일반적인 생각입니다.

강력한 호스트 프로세서에서 매우 간단하게 추론 작업을 실행하는 일부의 경우 코프로세서를 사용하는 가속은 필요하지 않습니다. AI를 사용할 수 있는 애플리케이션이 이런 경우에 해당됩니다(완전한 AI 애플리케이션과 다름). AI를 사용할 수 있는 애플리케이션의 경우, 애플리케이션의 일부분에서만 AI 기능을 수행합니다. 예를 들면 대화형 인터페이스를 사용하는 비즈니스 소프트웨어의 구매 기능입니다. 이와 같은 애플리케이션에서 강력한 CPU를 사용하는 경우, GPU 등의 코프로세서에서 가속하기 위해 AI 처리 부하를 분배할 필요는 없습니다. AI 추론 과정은 클러스터링으로부터 도움을 받을 수 있으나 멀티소켓 수직 확장 플랫폼에서 실행할 수도 있습니다.

## AI 인프라 고려 사항

지난 2-3년의 경우에는 AI 애플리케이션에 적합한 인프라는 시도하려는 목적에 따라 크게 좌우되었습니다. 기업은 하이퍼컨버지로부터 수직 확장과 수평 확장에 이르는 모든 것을 시도합니다. 그 후 AI 구현이 성숙되고 확장이 시작되면서 IDC는 클러스터링을 통해서 우선적으로 최대한 이득을 얻는 AI 개념에 대해서 상당한 동기가 구축된 것을 확인했습니다. AI 처리의 병렬적인 특성상 한 가속기에서 수백 개 코어를 활용하고 하나의 서버 노드에서 여러 개의 가속기를 활용하며, 하나의 서버 클러스터에서 여러 대의 서버를 활용하는 능력은 성능 측면에서 장점이 됩니다.

AI 인프라 확장을 시작할 준비가 된 기업은 일반적으로 (수준의 차이는 있으나) 온프레미스 배포 및 클라우드에서 달성될 수 있는 고려 사항을 적용합니다. 이에 따라 가속 AI 인프라 확장을 온프레미스로 할 것인지, 클라우드로 할 것인지 또는 하이브리드 클라우드로 할 것인지 결정하기 전에 어떤 사항이 가장 중요한지 생각하는 시간을 갖습니다.

고객에게 AI를 위한 가능한 최선의 인스턴스를 제공하기를 원하는 클라우드 서비스 공급자는 이 같은 고려 사항이 중요합니다. 이는 가속이 필요한 애플리케이션의 온프레미스 배포 증가가

확인되었기 때문이기도 합니다. 클라우드 서비스 공급자가 AI 클라우드 공급자로 선택되어 가속 제품을 서비스로 제공하고자 한다면 가능한 최선의 온프레미스 제품과 동등한 제품을 제공할 수 있어야 합니다.

표 1은 가속 시스템을 위한 여러 가지 하드웨어, 소프트웨어, 데이터센터 고려 사항에 대한 개요를 나타냅니다. 이 내용은 중요도의 관점에서 살펴보거나(IDC 조사에서 가속 시스템 사용자에게 대해 적용한 기본 방법), 온프레미스 배포나 주요 클라우드 서비스 공급자를 사용하는 방법에 대해 달성 가능성 측면으로 검토할 수 있습니다.

성능, 메모리, 보안, 고가용성, 가상화, 상호 연결 대역폭 등은 중요한 것으로 판단되는 서버 특성이며, 현재 클라우드보다 온프레미스에서 성능이 우수하며 이상적인 접근법으로 하이브리드 클라우드 모델을 지향하고 있습니다.

표 1

가속 서버 고려 사항의 중요도 및 온프레미스와 클라우드에서 달성 가능성

|   | 중요도 | 온프레미스 달성 가능성 | 클라우드에서 달성 가능성 |
|---|-----|--------------|---------------|
| <b>하드웨어</b>   |     |              |               |
| 호스트 CPU의 성능   | ●   | ●            | ●             |
| 가속기에서 사용 가능한 메모리 용량                                 | ●   | ●            | ●             |
| 가속 서버의 보안성  | ●   | ●            | ●             |
| 가속 서버의 고가용성   | ●   | ●            | ●             |
| 가속 시스템의 서버 가상화                                      | ●   | ●            | ●             |
| 서버 노드 내에서 가속기 확장                                    | ●   | ●            | ●             |
| 가속기와 호스트 CPU 사이 상호 연결의 대역폭                          | ●   | ●            | ●             |
| 가속기에서 성능 개선   | ●   | ●            | ●             |
| 가속 서버의 전력 요구 사항                                     | ●   | ●            | ●             |
| 가속 서버 노드의 수평 확장                                     | ●   | ●            | ●             |
| 가속 서버의 방열   | ●   | ●            | ●             |
| <b>소프트웨어</b>  |     |              |               |
| 가속기 프로그래밍 용이성                                       | ●   | ●            | ●             |
| 가속 서버 진단  | ●   | ●            | ●             |
| API, 라이브러리, 소프트웨어 개발 키트, 툴킷, 프레임워크, 프로그래밍 언어 등의 가용성 | ●   | ●            | ●             |
| 가속기 프로그램에 필요한 시간                                    | ●   | ●            | ●             |
| 가속기 프로그램에 필요한 비용                                    | ●   | ●            | ●             |
| OpenCL, OpenMP, OpenACC 등의 오픈 소스 지원                 | ●   | ●            | ●             |
| <b>데이터센터</b>  |     |              |               |
| 인프라의 나머지 부분에 대한 가속 서버의 상호 운용성                       | ●   | ●            | ●             |
| 운영 환경에 필요한 기술 수준                                    | ●   | ●            | ●             |
| 가속 서버의 관리 효율  | ●   | ●            | ●             |

높음: ●; 약간 높음: ●; 중간: ●

출처: IDC, 2019

IBM POWER SYSTEM AC922

IBM Power® System AC922는 Oak Ridge National Laboratory가 소유한 전 세계에서 가장 빠른 슈퍼 컴퓨터인 Summit의 구성 요소입니다. IDC는 IBM이 2018년 중반에 발표된 이 같은 성과를 너무 가볍게 여긴다고 판단합니다. Summit은 200petaflops 기능을 갖추고 있으며, 10<sup>18</sup> 연산/초를

의미하는 **exaops** 에 도달한 최초의 슈퍼 컴퓨터입니다. 또한 **Summit** 은 전 세계 환경 친화도에서 **3** 위를 차지한 슈퍼 컴퓨터입니다. 흥미로운 것은 **Summit** 은 시뮬레이션 및 모델링 목적으로 구축된 다른 수많은 슈퍼 컴퓨터와 달리 **AI** 를 고려하여 구축되었습니다. **IBM** 은 초대형 **AI** 워크로드 확장을 고려하는 기업을 위해 “**mini-Summits**”라고 하는 가속 컴퓨팅 플랫폼을 제공합니다.

이와 같은 규모를 운영하지 않는 업체의 경우에는 단일 **Power System AC922** 또는 **Power System AC922** 의 중소 규모 클러스터를 통해 놀라운 성능을 활용할 수 있습니다. **IBM** 의 설명에 의하면 밀집 설계와 **PCIe Gen4** 및 **InfiniBand** 조합의 장점을 통해, 업체에서는 단일 **Power System AC922** 노드로 시작할 수 있으며 이후 랙 또는 준선형 확장 효율성을 지닌 수천 개의 노드로 확장할 수 있습니다.

수 년 전 **IBM** 이 **Linux®** 기반 수평 확장 **Power Systems** 구축을 시작했을 때, 엄청난 데이터 집약적 컴퓨팅 용량과 시장을 앞서 나가는 이 같은 전략으로 인해 빌딩 시스템에 특히 관심이 집중되었으며, 당시 **AI** 및 분석은 새롭게 등장하는 워크로드였습니다. 이와 동시에 **IBM** 은 새로운 **Linux** 기반 수평 확장 시스템에서 핵심 업무 데이터를 위한 안전하면서 신뢰성 있는 **Power Systems** 플랫폼의 명성을 확인해 주었습니다.

이 같은 노력의 결과는 업계 최고의 코어 성능 중 일부를 차지한 **1** 및 **2** 소켓 **Linux** 서버의 광범위한 포트폴리오였습니다. **Power System AC922** 는 **4** 또는 **6** 개의 **NVIDIA Tesla V100 GPU** 를 추가로 제공하는 **2** 소켓 **2U** 시스템으로 이 부문에서 최고의 제품입니다. 이 제품은 **POWER9** 프로세서와 **NVIDIA** 의 상호 연결 **NVLink2(2 세대 NVLink)**를 독특하게 통합하여 완벽한 **CPU-GPU** 상호 작용과 일관성 있는 메모리 공유가 가능합니다. 시스템에 일관성을 부여하여 시스템 메모리를 **GPU** 메모리와 같이 사용하기 때문에 프로그래밍이 손쉽고 보다 큰 규모의 **AI** 모델에 적합합니다. 현재 다른 어떤 서버 플랫폼에도 초고속 **CPU-GPU** 연결을 위해 **NVLink** 가 프로세서에 직접 내장된 것은 없습니다. 이 링크를 통해 **GPU** 는 **DRAM** 에 고대역으로 액세스할 수 있습니다.

**Power System AC922** 에는 다른 몇 가지 주목할 만한 세부 사항이 있습니다. 코어당 **Level 2** 캐시는 **512KB** 이며, **Level 3** 캐시는 **10MB** 를 초과하고 시스템 메모리는 **256- 2,048GB** 사이에서 제공됩니다. **8** 개 스레드를 사용하는 **POWER** 코어는 **x86** 솔루션 대비 **4** 배 많은 스레드를 제공합니다. **AI** 와 같은 병렬화 워크로드의 경우에는 이것이 중요한 사양입니다. **Power System AC922** 에는 높은 처리량의 온 칩 환경이 있으며, **7TBps** 속도로 데이터를 이동할 수 있는 온 칩 스위치가 제공됩니다. 각 코어의 데이터 입출력 이동 속도는 **256GBps** 입니다. **NVLink2** 이외로 **CAPI 2.0** 이 제공되어 **ASIC**, **FPGA** 또는 외부 플래시 스토리지와 일관성 있는 대규모 고대역 활용이 가능하고, 업계 최초의 **PCIe Gen 4** 제공을 통해 **PCIe** 장치와 연결됩니다.

**4** 개 **GPU** 로 구성된 **Power System AC922** 제품 냉각은 공랭식 또는 수랭식이며 **6** 개 **GPU** 버전은 수랭식만 사용합니다. 이 부분에서도 **IBM** 은 일부 공급 업체와 함께 시장을 선도하고 있습니다. 수랭식은 현대의 가속 컴퓨터를 위해 새롭게 등장(또는 재등장)하는 기술로서, 고밀도가 가능하므로 노드당 보다 많은 **GPU** 및 랙당 더 많은 노드를 제공합니다.

**AI** 분야에서 **Power System AC922** 의 장점은 시스템에서 **AI** 학습 시간이 보다 빠르다는 것입니다. **IDC** 는 **AI** 솔루션의 생산 단계에 있는 업체에서는 학습 시간에 상당히 부담을 느끼는 것으로 파악하고 있습니다. 모델 학습이 완료되려면 데이터 과학자는 종종 수 일 또는 수 주를 기다려야 합니다. 그 후에 모델을 수정하거나 전체 작업을 반복해야 합니다. **Power System AC922** 를 사용하면 보다 빠르게 해당 모델을 반복시킬 수 있습니다. 또한 **GPU** 와 시스템 메모리가 거의 직접 액세스에 가까우므로 **IBM** 의 **LMS(Large Model Support)** 기능을 사용하여 학습시킬 수 있습니다. 이는 **GPU** 메모리에만 의존하는 것보다 더 크고 보다 복잡한 모델 및/또는 더 높은 정밀도 수준을 사용할 수 있다는 것을 의미합니다.

확장성을 고려하여 **IBM** 은 **DDL(Distributed Deep Learning)**이라는 라이브러리를 개발했습니다. 이는 데이터 과학자를 위해 완벽하게 끊임 없는 방식으로 수백 개의 서버 전체에서 단일 작업을 거의 선형으로 수평 확장할 수 있습니다. 이 작업에서 필요한 것은 간단한 호출과 필요한 **GPU** 개수를 요청하는 것입니다. **DDL** 은 **TensorFlow**, **Caffe**, **Torch**, **Chainer** 와 같은 **ML** 프레임워크에

내장되기 때문에 이 같은 프레임워크를 여러 개의 GPU로 확장할 수 있습니다. 이를 통해 데이터 과학자는 전체 수십 개의 GPU에서 학습 프로젝트를 손쉽게 확장할 수 있으며 시스템이 이것을 동적으로 관리합니다. 이 부분은 다른 제품에서는 수행하기 어려운 작업입니다.

표 2는 Power System AC922 기능을 앞에서 최종 사용자가 중요하게 생각하는 하드웨어 기능과 비교해 설명합니다.

표 2

가속 서버로서 IBM Power System AC922의 중요한 하드웨어 기능

| 하드웨어                       |   |
|----------------------------|---|
| 호스트 CPU의 성능                | ✓ |
| 가속기에서 사용 가능한 메모리 용량        | ✓ |
| 가속 서버의 보안성                 | ✓ |
| 가속 서버의 고가용성                | ✓ |
| 가속 시스템의 가상화                | ✓ |
| 서버 노드 내에서 가속기 확장           | ✓ |
| 가속기와 호스트 CPU 사이 상호 연결의 대역폭 | ✓ |
| 가속기에서 성능 개선                | ✓ |
| 가속 서버의 전력 요구 사항            | ✓ |
| 가속 서버 노드의 수평 확장            | ✓ |
| 가속 서버의 방열                  | ✓ |

출처: IDC, 2019

Power System AC922는 온프레미스이든 IBM Cloud 이든 무관하게 IBM의 새로운 오픈 소스 기반 AI 스택을 위한 하드웨어 환경을 제공합니다(IBM에서 이 스택의 대부분을 x86 기반의 하드웨어에서 사용할 수 있게 구성함). Power System AC922에서 해당 스택이 가장 최적화되어 있습니다. 딥 러닝은 가속기에 크게 의존하며 IBM은 소프트웨어를 최적화하여 Power System AC922의 4개 또는 6개 GPU, NVLink, Power System AC922 서버의 클러스터 등을 적극 활용합니다. 이 스택은 IBM Cloud를 포함한 사설 클라우드뿐만 아니라 공용 클라우드에서도 사용할 수 있으며 이 구성은 다음과 같습니다.

- **Watson™ Studio**에는 Jupyter 및 RStudio가 포함되어 데이터 준비와 모델 개발을 지원합니다. 이제 Watson Studio에는 WML CE(Watson Machine Learning Community Edition)가 포함되어 PowerAI가 Watson Studio로 통합됩니다. 본 제품은 무료로 제공합니다.
- **Watson Machine Learning**은 사설 클라우드 및 공용 클라우드 어디에서든 AI 모델을 학습, 배포, 관리하는 런타임 환경입니다. 이 제품은 모델 관리를 위한 공용 클라우드 솔루션과

다른 IBM 의 특징을 보여줍니다. 사설 클라우드 옵션으로 IBM Cloud Private 또는 Kubernetes 기반의 접근법을 사용합니다. Watson Machine Learning 에는 다른 여러 가지와 함께 Spark, TensorFlow, PyTorch, Chainer, Keras 와 더불어 (로지스틱 회귀, 의사 결정 트리, 랜덤 포레스트 등의) 관련 데이터 과학 작업에서 매우 유명해진 IBM 의 새로운 “기존” 머신 러닝 성능 부스터 SnapML 이 포함됩니다. 또한 이제는 Watson ML Accelerator 라고 일컫는 WML Community Edition 이 제공됩니다. PowerAI Enterprise 브랜드를 사용하는 Watson ML Accelerator 는 IBM 의 딥 러닝 모델 학습용 소프트웨어입니다. Watson ML Accelerator 는 여러 명의 데이터 과학자가 동일한 인프라를 사용하려는 경우에 리소스 관리를 위해 주로 사용하며, 단일 작업을 인프라 전체에 탄력적으로 확장할 때에도 사용합니다.

- **Watson OpenScale™은 AI 모델 지표와 AI 모델의 가중치 및 적합도 모니터링을 제공합니다.** OpenScale 은 애플리케이션 성능 관리와 유사한 AI 모델 성능 모니터링 도구입니다. 이 제품은 모델의 정밀도를 추적하거나 가중치 및 적합도에 대해 지정된 일군의 지표를 구현하거나 이에 관련된 모델을 점검할 수 있습니다.

업체는 이 스택을 사용해 매우 다양한 AI 애플리케이션을 실행할 수 있습니다. 흥미로운 한 가지 사례는 IBM PowerAI Vision 입니다. 그 핵심은 컴퓨터 비전 딥 러닝으로 업체에서는 다양한 종류의 이미지를 분류 및 감지하는 신경 네트워크 모델을 아주 손쉽게 신속하게 개발할 수 있습니다. PowerAI Vision 은 광범위한 워크플로 지원을 제공합니다. 여기에는 설치 및 구성, 데이터 레이블링, 모델 학습, 추론, 모델의 제품화 이동 등의 전체 라이프 사이클 관리가 포함됩니다. PowerAI Vision 은 드론 감시, 현장 및 시설의 안전 규제 집행, 제조 품질 감시, 도시 교통 관리 및 기타 수많은 사용 사례에서 사용할 수 있습니다.

## 향후 전망

IDC 는 AI 의 미래, 발표 예측, 드라이버, IT 에 대한 영향 및 권장 사항 등에 대해서 이른바 IDC FutureScapes 및 기타 문서 유형으로 방대한 조사를 수행했습니다. 본 문서는 이와 같은 예측의 일부를 소개하기에는 공간이 부족합니다. 다만 2024 년에 이르면 AI 는 우리의 삶, 기업의 활동, 데이터센터 운영 등에 엄청난 변화를 초래할 것이라고 짧게 말씀드릴 수 있습니다.

즉, 가장 중요한 단계는 데이터 과학자가 소프트웨어와 하드웨어 도구를 이용하여 보다 나은 고품질의 AI 솔루션을 더 빠르게 만들도록 하는 것입니다. AI 의 미래를 충족시키기 위해 데이터 과학자가 요구하는 컴퓨팅 작업량은 엄청난 수준입니다. 가속 컴퓨팅은 데이터 중심 워크로드에서 예외가 아닌 표준이 될 것으로 예상됩니다. GPU(및 FPGA 와 ASIC)가 지금까지 상상도 하지 못할 처리 성능을 보여주고 있으나, 신생 기술 기업과 IBM 을 포함하는 수많은 기존 기술 기업은 AI 프로세서를 새롭게 선보일 것입니다. 이 같은 새로운 프로세서는 AI 처리에 있어서 10 배 심지어 100 배의 성능을 제공할 것입니다. 이와 동시에 새로운 소프트웨어 모델이 개발되어 모든 것에 AI 를 손쉽게 적용할 수 있게 됩니다.

AI 모델 학습은 여전히 좋지 않을 것이며 모델은 더욱 커지고, 보다 복잡해지며 가중치를 배제하는 등 보다 더 정확하게 될 것입니다. 이 같은 모델에서 추론 작업은 곧바로 에지 단계에서 최대 워크로드가 되고, 이 단계에서 재학습이 이미 이루어지고 있습니다. 이 모든 것의 핵심은 사용 사례 설계와 모델 개발의 측면에서 (지원할 적합한 인프라 구축 및 운영에서 실패하지 않는 한) AI 추진력이 강화된다는 것입니다. AI 인프라 시도를 위해서 “난관에 직면해야 하는” 시대는 지나갔습니다. 업무상 중요한 AI 애플리케이션 확장을 시도하는 수많은 업체의 최고 우선 순위는 이제 인프라입니다.

## 당면 과제/기회

### 기업에 대한 안내

이 문서는 제품화를 위해 AI 애플리케이션 확장을 준비하는 업체가 당면하게 되는 다양한 인프라 문제에 대해 설명합니다. 데이터 준비로부터 모델 개발과 학습을 위한 런타임 환경, 배포, AI 모델

관리에 이르는 기본 인프라에 대한 요건은 범용 하드웨어를 사용하는 기존의 모델에 적용하지 못합니다. 데이터 집약적 워크로드, 탁월한 코어 성능, 여러 개의 GPU, 고속 상호 연결, 대용량의 일관성 있는 메모리, 최고의 I/O 대역폭을 고려하여 설계된 인프라만이 딥 러닝 학습 워크로드를 충분히 빠르게 실행할 수 있습니다. 기업은 기존의 범용 하드웨어를 교체하거나 AI 특정 처리 작업 전용 하드웨어를 사용하여 보완하는 의사 결정을 해야 합니다. 이에 관련된 기회는 최첨단 AI 애플리케이션을 개발 및 실행할 수 있도록 해주는 처리 역량을 구축하는 것입니다.

## IBM Power Systems 에 대한 안내

IBM Power Systems 의 과제는 항상 시장에서 인정받는 제품 중 하나가 되는 것입니다. IBM 은 우수하게 설계된 AI 소프트웨어 스택이 내장된 최고의 AI 인프라 솔루션을 제공하지만, 잠재적 고객은 불행하게도 “이기종” 제품 또는 비용이 더 들어가는 제품으로 잘못된 인식을 하고 있습니다. 대용량 데이터 집약적 워크로드를 제공하는 대형 상용 하드웨어 공급자를 대할 때 자동적으로 나타나는 소비자 반응 덕분에 기업은 많은 이점을 얻을 수 있는 매우 우수한 AI 인프라 솔루션을 사용하지 못하게 됩니다. 예를 들면 Power System AC922 는 모든 데이터센터에 위한 슈퍼 컴퓨팅 구성 요소입니다. IBM 은 Watson 브랜드를 통해 AI 인프라와 AI 소프트웨어 브랜딩의 효율성을 조절하고 있습니다. 이것은 장기적인 관점에서 보면 IBM 과 고객 양측에 효과가 있으며 도움이 될 것입니다. 단기적으로 보면 새로운 서브 브랜드에 대해 더 많은 설명을 해야 하며, 고객과 업계 전문가의 혼동을 없애기 위해 노력해야 합니다. IBM 의 기회는 기술적인 역량에 달려 있습니다. 이제 새로운 AI 워크로드가 실행되는 온프레미스 및 클라우드 인프라 분야에 대한 도전을 진지하게 시도함에 따라 IBM 에게는 잠재적 고객이 있기 때문에 이것이 힘찬 도약의 기회입니다!

## 결론

지난 몇 년 동안 IDC 는 많은 기업이 광범위한 AI 기능 개발을 시작하는 것을 목격했습니다. 처음에는 비교적 경험이 부족한 직원과 함께 실험적으로 시작하다가 점점 사용 가능한 모든 인프라에서 실행을 하였으며, 이제 원하는 효과를 얻기 시작했습니다. 수많은 업체가 광범위한 AI 전문 지식을 개발해 왔으며 AI 기능이 비즈니스에서 중요한 측면이 되는 속도와 관련된 직접적인 경험을 가지고 있습니다.

이와 동시에 IT 부서 역시 AI 를 실행하는 인프라 측면에서 학습 곡선을 경험하게 되었습니다. 이제 딥 러닝 학습 또는 추론 작업을 위한 인프라 요구 사항 및 제품화 환경을 위한 확장 방법은 매우 명백합니다. 딥 러닝 학습에는 다른 워크로드와 구별되는 특정한 인프라가 필요합니다. 딥 러닝 학습에는 강력한 프로세서, 대용량 코프로세서, 고속 상호 연결, 대규모 I/O 대역폭, 대규모 메모리가 준비된 클러스터드 노드가 필요합니다.

이제 IT 부서가 해야 할 최선의 의사 결정은 AI 워크로드에 사용할 노드당 성능을 어떻게 할지 결정하는 것입니다. 이때 성능 이득을 달성하기 위해 앞에서 설명한 요소와 이들의 연결 방법과 관련된 기능 최적화를 고려해야 합니다. 대량의 GPU 만 있으면 되는 것이 아니라 이런 GPU 사용을 최대화하는 플랫폼이 있어야 합니다. IDC 는 IBM 의 Power System AC922 가 AI 학습을 위한 최대 노드당 성능을 달성하는 탁월한 선택이라고 판단합니다. 무엇보다 Power System AC922 는 전 세계에서 가장 빠른 슈퍼 컴퓨터의 구성 요소입니다.

## IDC 안내

IDC(International Data Corporation)는 정보 기술, 통신 및 소비자 기술 시장에 마켓 인텔리전스, 자문 서비스 및 이벤트 서비스를 제공하는 세계적인 공급 업체입니다. IDC는 IT 전문가, 비즈니스 임원 및 투자 커뮤니티가 기술 구매 및 비즈니스 전략에 대한 사실 기반 의사 결정을 내릴 수 있도록 지원합니다. 1,100명 이상의 IDC 분석가가 전 세계 110여 개 국가의 기술 및 업계 기회와 동향에 대한 글로벌, 지역 및 현지 전문 지식을 제공합니다. IDC는 50년 동안 고객이 주요 비즈니스 목표를 달성하는 데 도움이 되는 전략적 통찰력을 제공해 왔습니다. IDC는 세계 최고의 기술 미디어, 리서치 및 이벤트 서비스 제공업체인 IDG의 자회사입니다.

## 본사

5 Speen Street  
Framingham, MA 01701  
USA  
508.872.8200  
트위터: @IDC  
idc-community.com  
www.idc.com

---

### 저작권 통지

IDC 정보 및 데이터의 외부 발행 - 광고, 보도 자료 또는 홍보 자료에 사용되는 모든 IDC 정보는 해당 IDC 부사장 또는 국가 관리자의 사전 서면 승인이 필요합니다. 제안된 문서의 초안에는 이러한 요청이 수반되어야 합니다. IDC는 어떠한 이유로든 외부 사용 승인을 거부할 권리를 갖습니다.

Copyright 2019 IDC. 서면 허가 없이 복제하는 행위는 엄격히 금지되어 있습니다.

