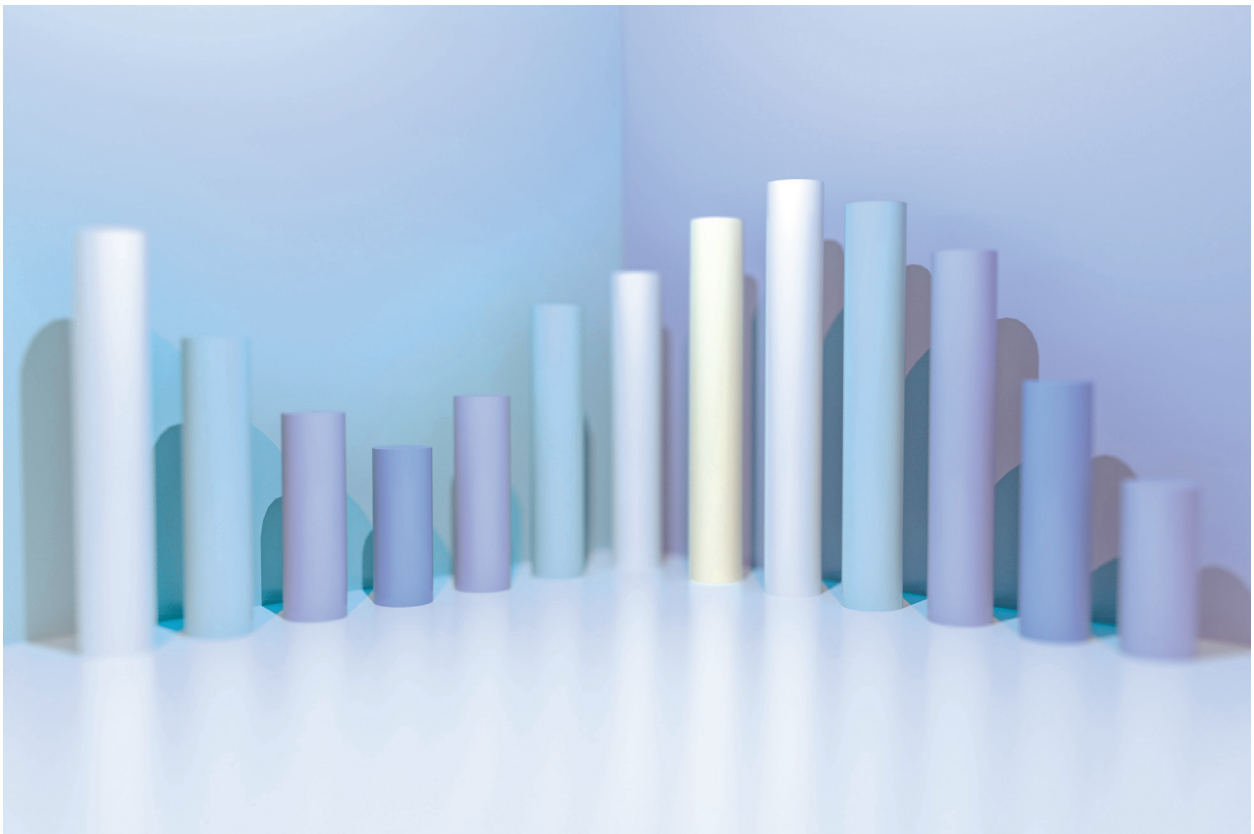


디지털 전환 시대를 앞서는 원스톱 비결 통합 AI 플랫폼을 주목하라

디지털 전환 시대에 AI가 조직 운영의 필수 요소가 되면서 AI를 비즈니스에 적용하려는 기업들이 증가하고 있다. 단순 반복적인 업무에 AI 및 ML(머신러닝) 기반의 자동화를 적용한 기업들도 많으며, AI를 전체 업무로 확장하기 위한 투자도 늘고 있다. 성공적인 AI 도입을 위해서는 자사의 환경에 맞는 최적의 방향 설정이 필수적이다.

효성인포메이션시스템이 제안하는 최적의 AI 플랫폼 구축 전략과 새로운 성장동력의 아이콘으로 자리매김할 'HPC사업팀'을 만나본다.



Part I

최적의 AI 플랫폼 도입을 위한 A to Z

디지털 전환 시대에 자사의 환경에 적합하면서도 지속적인 운영과 개선이 가능한 AI 플랫폼을 찾는 기업들을 위해 효성인포메이션시스템은 '통합 AI 플랫폼'을 제안한다. 통합 AI 플랫폼은 GPU 기반 데이터 연산부터 고성능 컴퓨팅에 최적화된 스토리지를 활용한 대용량 데이터 저장과 자동화된 관리 그리고 운영까지 아우르는 '전천후' 솔루션으로 업계의 시선을 끌고 있다.

떠오르는 AI 플랫폼 시장

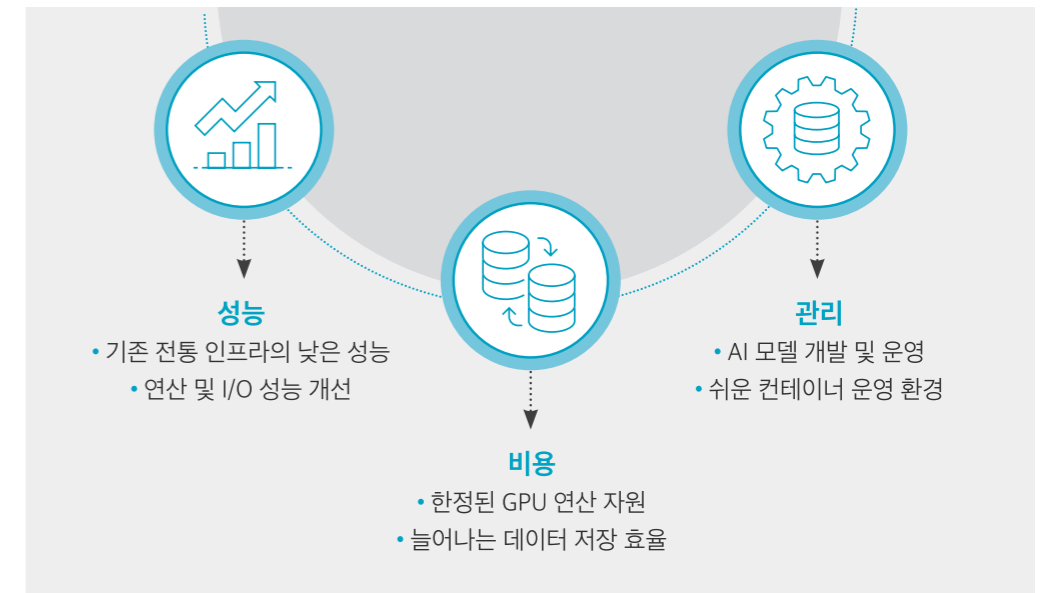
올해 3월 시장조사 전문기관 IDC가 발표한 바에 따르면, 국내 AI 시장은 향후 5년간 연평균 15.1%의 성장률을 기록하고, 2025년에는 규모가 1조 9,074원에 이를 전망이다. 2022년만 하더라도 전년 대비 18.8% 성장하며 1조 1,212억 원의 시장 규모를 형성할 것으로 조사됐다.

IDC의 '전 세계 AI 소프트웨어 전망(2021년-2025년)' 보고서에 따르면, AI 플랫폼은 서비스, 소프트웨어, 하드웨어 인프라 영역으로 구분할 수 있다. IDC는 전체 AI 시장에서 소프트웨어 영역이 88%에 달하지만, 성장률 측면에서는 AI 하드웨어 시장이 향후 몇 년간 가장 빠른 성장을 보일 것이라고 분석했다. 또한 2023년부터는 AI 서비스 시장이 급격히 성장할 것으로 전망했다.

글로벌 AI 하드웨어 시장의 경우, 2021년에는 전년 대비 약 30% 성장세를 보였으며 향후 5년간의 연평균 성장률은 19.4%에 이를 것으로 조사됐다. IDC는 AI 하드웨어 중에서도 스토리지 시장이 5년간 연평균 22.1%의 성장률을 유지하면서 높은 성장세를 보일 것으로 전망했다.

시장 전망뿐 아니라 실제로 AI를 도입하려는 기업들의 움직임도 활발하다. 그러나 AI 플랫폼을 기반으로 자사가 원하는 성과를 창출하기까지는 여러가지를 고려해야 한다. 구체적으로 ▲도입한 인프라의 성능 최적화 ▲한정된 GPU 자원의 효율적인 사용 방법 ▲만들어진 AI 모델에 대한 지속적이고 효율적인 관리 등이 AI 플랫폼을 운영하고자 하는 기업들이 풀어야 할 과제다.

↓ AI 플랫폼 도입 시 기업들의 고민



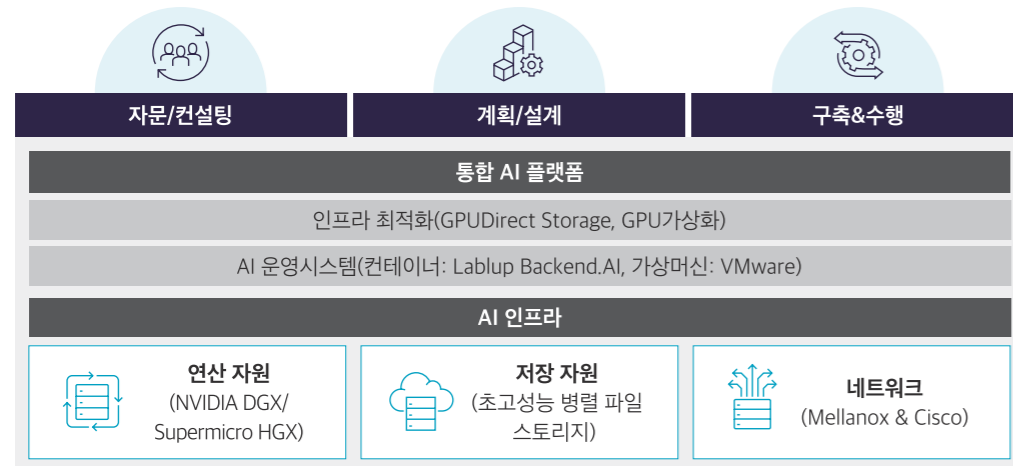
최적의 AI 플랫폼 구축 전략

기업이 AI를 비즈니스에 적용하고 제대로 활용하기 위해서는 클라우드, 컨테이너, 가상머신, 빅데이터, AIops 솔루션, 고성능 스토리지, 네트워크 등 많은 구성요소가 필요하다. 그리고 이는 비용, 업무 효율, 관리 측면에서 큰 변화가 따른다.

기업들은 한정된 GPU 연산 자원의 효율적인 사용 방법과 분석 및 대량의 데이터 저장 방법 등을 끊임없이 고민하고 있다. 전통적 인프라 환경에서는 AI 모델을 개발하고, 이를 효율적으로 운영하기엔 성능 요구사항을 달성하기가 어렵다. 유명한 GPU 서버만 도입한다고 해서 성능이 개선되는 것도 아니다. 기업별로 AI 도입을 통한 비즈니스 요구 사항과 다른 환경에 대한 최적의 방향 설정이 필요하다는 얘기다.

효성인포메이션시스템은 기업이 AI를 업무에 활용함으로써 최적의 가치를 창출할 수 있도록 '통합 AI 플랫폼'을 제공하고 있다. 통합 AI 플랫폼은 최적화된 성능의 GPU 서버 인프라와 AI 모델 운영 관리 시스템을 결합하고, 성능과 자원 효율 그리고 운영관리를 최우선으로 고려하여 고객의 니즈와 운영 환경에 적합하도록 맞춤형 솔루션을 제공한다. 단순히 고속의 인프라 장비를 공급하는 것에서 그치지 않는다. 실질적인 AI 플랫폼 구축을 위해 성능 효율, 비용, 관리 측면에서 최적의 로드맵을 제시하며 '처음부터 끝까지' 함께한다는 점에서 남다르다.

↓ 효성인포메이션시스템이 제공하는 '통합 AI 플랫폼'



AI 플랫폼이 최적의 성능을 발휘하려면 ▲성능 최적화 ▲연산 자원 ▲저장 자원의 세 가지가 완벽한 조화를 이뤄야 한다.

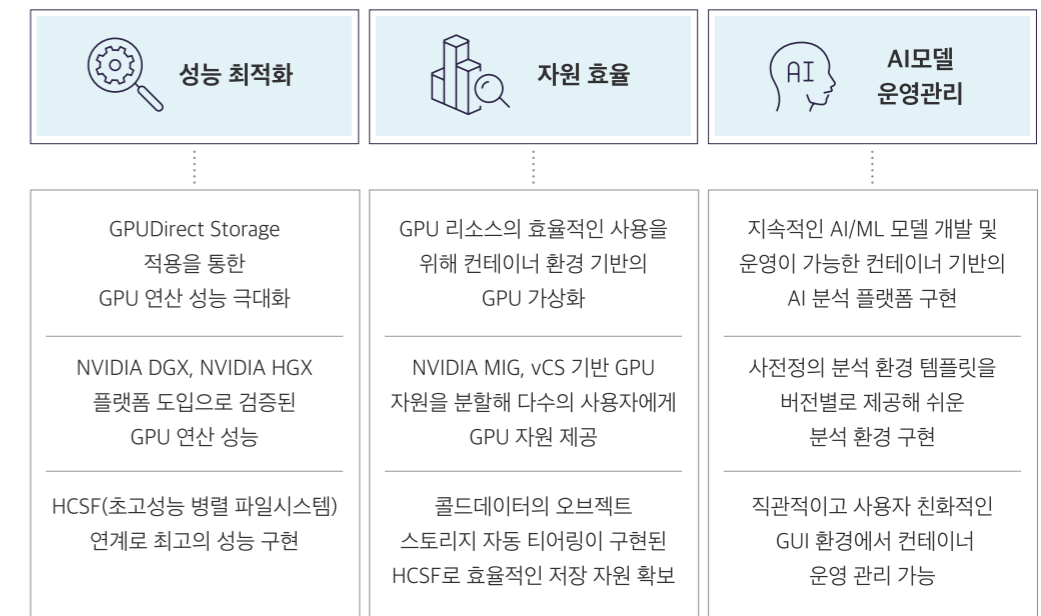
GPU 서버 리소스가 충분하다고 성능을 최적화할 수는 없다. 대형 AI 모델의 학습 성능을 높이기 위해서는 GPU 연산 및 저장 자원의 성능과 함께 I/O 성능 최적화가 매우 중요하다. 따라서 NVLink를 지원하는 GPU 서버와 고성능 병렬 파일 스토리지 인프라에 GPUDirect 기술을 적용하고 이를 실제로 구현해야 한다.

GPUDirect는 CPU와 CPU 메모리로 인한 병목현상을 제거하여 데이터 프로세스를 개선함으로써 성능 최적화가 가능하다. 즉 GPUDirect 기술을 적용하면 GPU 서버의 최대 성능 대역폭 I/O 처리가 가능하다.

또한 AI 분석 환경에서 간과하면 안 되는 부분이 바로 스토리지 성능이다. 기존 NFS NAS의 I/O는 속도에 한계가 있기 때문에 반드시 HCSF(Hitachi Content Software for File)와 같은 초고성능 병렬 파일시스템이 필요하다.

효성인포메이션시스템의 HCSF는 NVIDIA의 인증 GPU 서버와 함께 GPUDirect 스토리지 구성을 통해 저장 성능을 최적화한다. 특히 HCSF는 오브젝트 스토리지와 연계하기 때문에 비용 효율적으로 운용할 수 있으며, 정책 기반의 티어링 파일 시스템 운영으로 비용 대비 높은 성능과 용량 구성이 가능하다.

↓ '통합 AI 플랫폼'의 강점



기업별 맞춤 솔루션으로 가치 극대화

GPU 서버의 대표 주자인 NVIDIA의 DGX A100은 대규모 AI 모델의 학습과 추론, 분석에 최적화된 성능을 제공한다. 최근에는 성능이 더욱 업그레이드된 DGX H100이 출시됐다. DGX H100은 NVLink, NVSwitch의 업그레이드와 2장의 DPU 장착 등으로 A100에 비해 네트워크와 GPU 간 대역폭이 확장되면서 성능이 약 6배 정도 향상됐다.

대형 기업이나 대량의 데이터를 보유한 기업이라면 NVIDIA의 GPU 도입이 필수가 될 수 있다. 하지만 강력한 성능이 필요하지 않고, 가용 예산이 충분하지 않은 기업도 있다. 특히 AI 프로젝트를 처음 시작하는 기업의 경우 인프라 성능이 낮거나 제한적인 예산 등 여러 이유로 투자가 쉽지 않다.

비용 효율적인 GPU 도입을 원하는 기업들을 위해 효성인포메이션시스템은 다양한 GPU를 보유한 슈퍼마이크로와 파트너십을 맺고 기업의 환경에 맞는 GPU 서버를 제공한다.





효성인포메이션시스템은 슈퍼마이크로의 다양한 GPU 서버 라인업 중에서, NVLink를 지원하는 2개의 모델과 PCIe 방식의 GPU를 장착할 수 있는 2개의 모델을 전략적으로 제공한다. NVLink가 지원되는 제품은 HPC와 AI, 딥러닝을 위한 고성능 인프라이며, PCIe 방식은 최소 1개부터 최대 10개까지 GPU를 장착할 수 있어 고객의 예산과 환경에 맞춰 유연하게 대응할 수 있다.

현재 슈퍼마이크로는 인텔 및 AMD CPU 기반의 다양한 NVIDIA GPU 서버를 보유하고 있으며, 새롭게 출시될 인텔과 AMD의 GPU에 대한 지원시스템도 제공할 예정이다.

이 외에도 효성인포메이션시스템은 하이퍼스케일 AI 개발 플랫폼 전문 기업 래블업과도 파트너십을 맺고 협력 중이다. 래블업의 AI 개발 및 서비스를 위한 엔터프라이즈 클러스터 플랫폼인 Backend.AI(백엔드닷에이아이)는 컨테이너 기반의 GPU 분할 가상화 기술, 딥러닝 모델 분산 훈련 및 초거대 모델 개발 플랫폼 기술, GPU 기반의 초저비용/초저지연 AI 모델 서비스 기술을 포함해 AI 개발 및 서비스에 필요한 기능을 갖추고 있다.

백엔드닷에이아이는 아태지역 최초로 엔비디아 DGX-ready software로 검증된 AI 연구 개발 플랫폼이다. GPU 분할 가상화를 제공함으로써 데이터 과학자 및 AI 플랫폼 담당자들이 효율적으로 연산 자원을 사용할 수 있게 해준다. 효성인포메이션시스템은 래블업이 제공하는 AI 개발 플랫폼과 GPU 분할 가상화 기술을 기반으로 복잡한 AI 업무를 단순화하는 솔루션을 제공한다.

↓ Backend.AI의 엔터프라이즈 클러스터 운영 시스템

 <p>GPU 활용 극대화</p> <ul style="list-style-type: none"> • 컨테이너수준 GPU 분할 가상화 지원 • NVIDIA GPU MIG 지원 	 <p>직관적인 관리 및 사용자 경험</p> <ul style="list-style-type: none"> • GUI 기반 컨테이너 운영관리 지원 • 웹 UI와 데스크탑 앱 동시 지원 	 <p>AI 및 HPC 성능 최적화</p> <ul style="list-style-type: none"> • 독자적 엔진으로 최적의 GPU 연산 자원 배치 구현 • 다중노드 워크로드 및 데이터 I/O 병렬화 지원 	 <p>쉬운 워크로드 확장</p> <ul style="list-style-type: none"> • 다중노드/다중 컨테이너세션 • 모델훈련 및 데이터 I/O 파이프라인 분리
--	--	--	---

기업별 환경에 맞는 '통합 AI 플랫폼' 제공을 위해 효성인포메이션시스템은 발빠른 행보를 보이고 있다. 파트너와 협업해 로드맵 제안부터 수행, 유지보수까지 안정적인 비즈니스를 펼치는 한편, 기업들이 통합 AI 플랫폼을 직접 시연해볼 수 있도록 DX센터 고객 초청 세미나도 기획 중이다. DX센터에는 GDS, 컨테이너, GPU 가상화, AI 모델 등이 구현돼 있다. 이 외에도 파트너사와 AI 모델의 개발 및 운영 관리를 위한 구체적인 데모 시연과 사례를 소개하는 공동 세미나 준비에도 박차를 가하고 있다.

Part II

새로운 성장동력 AI 비즈니스 개척자 'HPC사업팀'

국내 AI 시장이 급격히 성장하고, 기업들의 니즈가 증가하면서 효성인포메이션시스템은 AI 비즈니스를 본격적으로 추진하기 위한 발판을 마련했다. 그 일환으로 신설된 'HPC사업팀'을 만나 비즈니스를 위한 각오를 들어본다.

데이터 시대 신규 비즈니스의 중추

디지털 혁신의 시대, 즉 데이터 시대가 열리면서 AI는 강력한 주역으로 부상했다. 시장의 가속화와 기술 경쟁력 강화에 따라, AI는 고객의 새로운 경험에 큰 영향을 미치는 혁신적이고 효과적인 기술로 진화하고 있다. 특히 코로나 팬데믹으로 비대면 시대가 활성화되면서 AI는 전 산업 계로 빠르게 퍼져나가고 있다.

전 세계 AI 시장은 빠르게 성장하고 있으며, 국내 기업들도 AI 도입을 적극적으로 추진하고 있다. 효성인포메이션시스템 역시 시장의 변화를 감지했다. 'AI 통합 인프라' 비즈니스의 가능성에 주목한 것이다. 이에 AI 전문 인력들을 모아 특화된 비즈니스를 위한 별도의 팀을 구성했다. 작년 하반기에 AI 사업 TF를 조직해 AI 비즈니스의 시장 환경과 사업화를 검토했고, 올해 초 'HPC사업팀'을 신설해 본격적인 비즈니스의 물꼬를 텄다.

"어떤 업무에 AI를 도입해야 효율적으로 사용할 수 있을지 고민하는 고객들이 많습니다. AI를 시작하는 것은 더욱 어려운 일이고요. 바로 HPC사업팀의 출발 지점이자 출발 이유이기도 합니다. AI 시작 단계부터 고객과 함께 AI 업무를 그려 가는 것이 우리의 목표입니다." HPC사업팀 송병진 팀장이 말했다.

HPC사업팀은 연산 처리를 담당하는 GPU와 네트워크, 저장장치로 구성되는 AI 하드웨어 플랫폼과 이를 운영·관리하고 GPU를 가상화해 효율성을 높이는 PaaS 솔루션을 모두 제공한다. 구체적으로 AI 모델링이 필요한 사업들, 예컨대 AI와 ML 예측분석 업무를 위한 슈퍼컴퓨팅 사업, AI 클라우드 GPU 구축사업, 데이터 레이크 연계 AI 업무용 GPU 서버가 필요한 시장에 솔루션을 공급하는 것이다.

도전의 즐거움을 아는 전문가들

송병진 팀장을 필두로 김형섭 차장, 정문종 차장, 김창겸 과장의 4명으로 구성된 HPC사업팀은 효성인포메이션시스템에서도 잔뼈가 굵은 베테랑들이다. 통합 인프라 총판 업무를 오래 해온 송병진 팀장은 HPC사업팀장으로서 고객의 업무 향상 방안을 수립하기 위해 영업팀과 협력을 진행 중이다. 3명의 팀원도 각 솔루션/인프라 분야에서 쌓은 컨설팅, 기술지원 역량 등 남다른 전문성을 갖췄다.

“10년 동안 총판 비즈니스를 담당해왔습니다. 이제 성장 가능성이 높은 AI 비즈니스를 통해 새로운 솔루션으로 신규 고객을 만날 수 있다는 점에서 기대가 매우 큽니다. AI 인프라 도입을 목전에 둔 고객들에게 실질적인 도움이 되기를 바랍니다.” 정문종 차장이 HPC사업팀 합류에 대한 기대감을 드러냈다.

꾸준히 성장 중인 국내 AI 플랫폼 시장에서 HPC사업팀은 그동안 쌓은 전문성을 기반으로 확고한 위치를 차지하고, 새로운 비즈니스로 회사의 중심에 서기를 희망한다.

“새로운 시장에서 고객에게 가치를 제공하기 위한 방안을 찾는 과정이 신생팀인 우리에게 큰 도전이자 즐거움이 될 것입니다. 현재는 HPC사업팀 소개, 대외 세미나 활동을 통해 AI 통합 플랫폼 사업을 적극적으로 알리고 있습니다.” 팀의 리더, 송병진 팀장이 HPC사업팀의 현주소를 전했다.

토탈 AI 솔루션 제공 벤더로 도약

통합 AI 인프라 솔루션 제공을 위해서는 효성인포메이션시스템 내 타 사업부서와의 긴밀한 협력이 중요하다. AI 연산(GPU 서버)과 데이터 레이크(HCSF)를 통합 제안하기 위해 데이터사업팀과 힘을 모으고 있으며, 클라우드사업팀과는 협업을 통해 다양한 GPU 포트폴리오를 갖추고 시너지 효과를 극대화하고 있다.

“AI 플랫폼 시장의 특성상 HPC사업팀이 담당하고 있는 연산용 GPU 서버와 데이터사업팀의 솔루션 HCSF를 통합해 서비스할 때 고객의 AI 비즈니스 가치가 상승할 것이라고 생각합니다. 하이퍼컨버지드 인프라(HCI) 시장을 맡고 있는 클라우드사업팀과의 협업도 아주 중요하고요. 팀 간 완벽한 협업으로 최적화된 인프라를 제공할 수 있는 효성인포메이션시스템만의 강점을 제대로 활용할 계획입니다.” 자신감에 찬 김형섭 차장의 말이다.

GPU 및 X86 서버 기술지원을 담당해오며 최근 효성인포메이션시스템에 합류한 김창겸 과장도 HPC사업팀의 비전에 힘을 실어주고 있다. “AI, HPC 워크로드 성능을 고려할 때, 스토리지 병목현상을 해결하기 위해 기업들이 GPU 서버와 스토리지 서버 구매를 별도로 고민하고 있습니다. 효성인포메이션시스템은 CPU에 의존하는 전통적인 I/O 처리방식을 해결하기 위해 GPUDirect 스토리지와 RDMA를 지원하는 동시에 초고성능 병렬파일시스템 HCSF를 함께 제안할 수 있는 큰 이점을 가지고 있습니다.”고 밝혔다. 김 과장은 진화하는 GPU 기술에 발빠르게 대응하며 효성인포메이션시스템이 AI 플랫폼 시장을 선도할 수 있도록 역량을 발휘하겠다는 포부를 드러냈다.

HPC사업팀은 통합 AI 플랫폼 비즈니스의 일환으로 AI DX센터 오픈도 준비하고 있다. AI DX센터에는 AI 플랫폼 시장에 성공적으로 진입할 수 있는 방안의 하나로, 통합 AI 플랫폼 시연을 통해 고객들이 쉽게 AI를 도입하고 활용할 수 있는 최적의 방법을 제시할 계획이다.

“지금은 AI 시장에 첫 걸음을 뗐지만, 머지 않아 효성인포메이션시스템이 AI 관련 토털 솔루션 공급 벤더로 자리매김할 수 있도록 비즈니스 활성화에 주력할 계획입니다. 효성인포메이션시스템의 역량을 재확인할 수 있는 좋은 기회가 되도록 모두가 노력하고 있습니다.” 정문종 차장이 AI 시장의 리더로 우뚝 서겠다는 바람을 드러냈다.

시장의 열기와는 별개로 AI나 머신러닝이 다소 어렵다는 기업들의 인식이 적지 않다. HPC사업팀은 AI, ML, GPU와 같은 단어를 접하는 순간 ‘효성인포메이션시스템 HPC사업팀’을 떠올릴 때까지 도전을 멈추지 않을 것이다.

