

그래픽 처리 장치(GPU) 확보 사업 (1차 추경, 1.46조원 규모)

참여 사업자 선정 결과

- 참여 희망사, 벤더사 등 간의 경쟁 유도, 적극적 협상 등을 거쳐 당초 예산 계획보다 크게 상회한 첨단 그래픽 처리 장치(GPU) 1.3만장 확보 계획
- 참여사들도 국내 자국 인공 지능(소버린 AI) 생태계 확장 등을 위해 적극 동참
- 속도감 있는 그래픽 처리 장치(GPU) 확보·구축 등 거쳐 연말부터 그래픽 처리 장치(GPU) 순차 지원 추진

과학기술정보통신부(장관 배경훈, 이하 과기정통부)는 그래픽 처리 장치(GPU) 확보 사업(1차추경, 1.46조원) 참여 사업자로 ‘네이버클라우드’, ‘NHN클라우드’, ‘카카오’를 선정 하였으며, 해당 사업자들과 협력하여 첨단 그래픽 처리 장치(GPU) 1.3만장을 확보한다.

이를 통해 국내 인공 지능 컴퓨팅 인프라를 신속 확충하고 필요한 산학연 등에 적기 지원함으로써, 국내 그래픽 처리 장치(GPU) 연구, 서비스 개발의 질적 도약과 인공 지능 생태계 활력 제고의 중요한 출발점이 될 것으로 기대된다.

【그간의 경과】

정부는, 지난 5월 약 1.46조 원 규모의 첨단 그래픽 처리 장치(GPU) 확보 사업 추경 예산을 편성한 데 이어, ‘참여자 공모(5.23~6.23), ‘제안서 평가, ‘데이터센터 현장 실사, ‘사업비 심의·조정 등의 절차를 진행해 왔으며, 국내 인공 지능 생태계에 최대한 많은 그래픽 처리 장치(GPU) 자원 제공과, 이용자 측면에서 경험·역량있는 기업들의 안정적 서비스 제공이 필요함에 따라, 정부와 협상 대상 사업자들 간 긴밀한 협상과 사업자들의 적극적 협력을 통해, ‘네이버클라우드, ‘NHN클라우드, ‘카카오를 최종 참여 사업자로 선정하였다.

【 총 그래픽 처리 장치(GPU) 확보 규모 】

민관의 긴밀한 협력을 통해 확보할 그래픽 처리 장치(GPU)는 총 1.3만장 규모로서, 구체적으로는 엔비디아(NVIDIA) B200 10,080장, H200 3,056장 규모이다.

※ FP(Floating Point) 8, Tensor Core 기준, B200은 H200 대비 연산성능 2.25배 수준

이 중 일부는 참여 사업자(CSP)들의 자체 활용도 지원하며(운영비 성격 고려), 정부가 활용할 그래픽 처리 장치(GPU)는 총 1만장을 상회하는, B200 8,160장, H200 2,296장 규모이다.

특히, 정부가 활용할 그래픽 처리 장치(GPU)는 1개의 B200 510노드(4,080장)와, 2개의 B200 255노드(2,040장), 1개의 H200 255노드(2,040장) 등으로 집적(클러스터링) 함으로써, 대규모 집적(클러스터링) 기반의 인공 지능 작업량(워크로드)에 대응 가능할 것으로 전망된다.

※ 그래픽 처리 장치(GPU) 클러스터링 : 여러 개의 그래픽 처리 장치(GPU)를 연결하여 통합된 컴퓨팅 시스템처럼 작동토록 구성

【 사업자별 그래픽 처리 장치(GPU) 확보 규모 】

참여 사업자별로 확보·구축할 그래픽 처리 장치(GPU) 규모는 아래와 같다.

네이버클라우드 먼저 네이버클라우드는, 총 3,056장의 H200을 확보·구축한다. 전체를 현재 시장의 주력 기종인 H200(Hopper 기반 아키텍처)으로 구성함으로써, 전체 그래픽 처리 장치(GPU) 서비스를 연내 안정적으로 제공할 계획이다.

네이버클라우드의 확보·구축분 중 정부가 활용할 그래픽 처리 장치(GPU)는 H200 2,296장으로, 255노드(2,040장), 32노드(256장)으로 집적(클러스터링)되어 활용될 계획이며, 해당 전체 그래픽 처리 장치(GPU) 자원은 연내부터 산학연 등에 지원·배분할 방침이다.

NHN클라우드 다음으로 NHN클라우드는, 총 7,656장의 B200을 확보·구축한다. 전체를 B200(BlackWell 기반 아키텍처)으로 구성, 수냉식으로 가동할 계획으로, 최신 그래픽 처리 장치(GPU) 클러스터를 친환경·고효율·고성능 방식으로 운용할 것으로 기대된다.

※ 수냉식은 공랭식 대비 에너지 효율성이 높고, 냉각 효율도 높아, 그래픽 처리 장치(GPU) 성능 제고 가능

NHN클라우드의 확보·구축분 중 정부가 활용할 그래픽 처리 장치(GPU)는 B200 6,120장으로, B200 510노드(4,080장)*, B200 255노드(2,040장)으로 집적(클러스터링)되어 활용될 계획이며, 일부는 연내 시범 운영 서비스(베타 서비스)를 추진할 방침이다.

* 255노드부터 순차 집적(클러스터링)하여(`26.1Q), `26년 중 510노드 집적화(클러스터링) 계획

카카오 마지막으로 카카오는 총 2,424장의 B200을 확보·구축한다. 전체를 B200으로 구성하고, 보다 효율적 기반(인프라)을 구동하는 데에 중점을 두었다. 카카오의 확보·구축분 중 정부가 활용할 그래픽 처리 장치(GPU)는 B200 2,040장으로, B200 255노드(2,040장)으로 집적화(클러스터링)되어 활용될 계획이며, 이 역시 일부는 연내 시험 운영 서비스(베타서비스)를 추진할 방침이다.

총 그래픽 처리 장치(GPU) 확보물량	
B200 10,080장, H200 3,056장	
정부 활용분	민간 자체활용분
B200 8,160장, H200 2,296장	B200 1,920장, H200 760장
<ul style="list-style-type: none"> ▶ B200 510노드(4,080장) ※ NHN클라우드 ▶ B200 255노드(2,040장) ※ NHN클라우드 ▶ B200 255노드(2,040장) ※ 카카오 ▶ H200 255노드(2,040장) ※ 네이버클라우드 ▶ H200 32노드(256장) ※ 네이버클라우드 	<ul style="list-style-type: none"> ▶ (네이버클라우드) H200 760장 ▶ (NHN클라우드) B200 1,536 ▶ (카카오) B200 384장

【사업자별 국내 기여계획】

모든 참여사들은 국내 자국 인공 지능(소버린 AI) 생태계 확장에 적극적인 의지를 밝히고, 정부의 그래픽 처리 장치 활용 물량 극대화에 적극 동참하였다. 아울러, 그래픽 처리 장치(GPU) 자체 활용분(운영비 성격 고려)을 이용하여 자체 인공 지능 개발·고도화, 산학연 그래픽 처리 장치(GPU) 저렴 공급 등을 추진하고, 일부 참여사는 국내 인공 지능 기반(인프라) 투자를 강화하는 등 국내 인공 지능 생태계에 다각도로 기여할 계획이다.

또한, ‘가칭 그래픽 처리 장치(GPU) 통합 지원 온라인 체제 기반(플랫폼)’ 구축에 협력키로 합의하여, 향후 필요한 산학연 관계자 등이 온라인으로 자유롭게 그래픽 처리 장치(GPU) 자원을 신청하고, 평가 등을 거쳐 그래픽 처리 장치(GPU) 자원을 지원·배분 받을 수 있는 기반을 마련할 계획이다.

※ 유휴 그래픽 처리 장치(GPU) 자원 발생시 대기 사용자 지원체제 구축 등도 포함될 예정

【향후계획 및 의의】

향후 정부는, 일련의 절차를 거쳐 이르면 8월초부터 사업자 협약과 그래픽 처리 장치(GPU) 구매 발주 등을 속도감 있게 추진할 예정이며, 국가 사업*과 필요한 산학연 등에 그래픽 처리 장치(GPU) 지원을 순차 개시**할 계획이다.

* 독자 인공 지능 기초 모형 사업(AI 파운데이션 모델 프로젝트) 등 / ** 연내부터 그래픽 처리 장치(GPU) 지원 순차 개시

이를 통해, ‘국내 인공 지능 연구, 서비스 개발 역량 강화 ‘인공 지능 초기 창업 기업(스타트업)·중소기업, 대학 등의 인공 지능 컴퓨팅 기반 시설 접근성 증진, ‘대규모 인공 지능 모형 개발 가속화 등에 적극 기여할 수 있을 것으로 기대된다.

과기정통부 배경훈 장관은 “이번 첨단 그래픽 처리 장치(GPU) 확보는 국내에 부족한 인공 지능 컴퓨팅 기반 시설의 마중물이자, 국내 인공 지능 생태계 전반의 혁신을 가속화하는 새 정부 소버린 AI 생태계 확장과 인공 지능 고속도로 구축의 출발점”이라며, “이를 기점으로 더욱 강력한 인공 지능 컴퓨팅 기반 확충, 대한민국의 인공 지능 강국 도약에 정책적 총력을 다하겠다”고 강조하였다.

담당 부서	과학기술정보통신부	책임자	과 장	장기철 (044-202-6360)
	인터넷진흥과	담당자	사무관	이현우 (044-202-6361)

내일을 만드는 과학기술
내일을 채우는 디지털·AI

대한민국
지정책브리핑



