목차

NVIDIA GPU 분류

세대 / 가로 구분	소비자용 PC GPU (GeForce)	워크스테이션 및 서버 그래픽 (RTX / L-시리즈)	AI 및 데이터센터 연산 가속기 (Compute)
각 세그먼트 특징	개인 게이밍, 크리에이터 작업, 로컬 AI 모델 테스트 및 가벼운 학습 중심. 공식 하이퍼바이저 가상화(vGPU, SR-IOV) 드라이버 지원 안 됨(소프트웨어 락 적용). 일반 엔비디아 홈페이지 배포.	3D CAD 설계, 미디어 렌더링, 고밀도 VDI(가상 데스크톱) 및 원격 가상 워크스테이션 인프라 가속. NVIDIA vGPU (GRID) 라이선스 스택 기반 작동. NVIDIA Licensing Portal 배포.	대규모 LLM(거대언어모델) 학습 및 추론, HPC(고성능 연산), 데이터 사이언스 전용 인프라. 디스플레이 출력 포트 없음. NVIDIA AI Enterprise (NVAIE) 라이선스 및 NVIDIA NGC Catalog 독점 배포.
Blackwell 세대 (2025 ~ 2026)	GeForce RTX 5090 GeForce RTX 5080 GeForce RTX 5070 Ti GeForce RTX 5070 GeForce RTX 5060 Ti GeForce RTX 5060	RTX PRO 6000 Server Edition (96GB) RTX PRO 4500 Server Edition (32GB) RTX PRO 6000 6000 Max-Q 5000 (48GB/72GB) 4500 4000 4000 SFF 2000	B200 (SXM/PCIe) B100 GB200 NVL72 GB200 NVL36 GB200 NVL2 B200X
Hopper 세대 (2022 ~ 2024)	존재하지 않음 (소비자 라인업은 Ada Lovelace 아키텍처로 우회 설계됨)	존재하지 않음 (순수 연산 및 고성능 인프라 전용 아키텍처로 그래픽 라인업 분리됨)	H100 NVL 94GB H100 (SXM5 80GB / PCIe 80GB) H200 (141GB HBM3e) GH200 Grace Hopper Superchip H800 H20
Ada Lovelace 세대 (2022 ~ 2024)	GeForce RTX 4090 GeForce RTX 4080 Super GeForce RTX 4080 GeForce RTX 4070 Ti Super GeForce RTX 4070 Ti GeForce RTX 4070 Super GeForce RTX 4070 GeForce RTX 4060 Ti (16GB/8GB) GeForce RTX 4060	L40S (48GB, AI/그래픽 혼합) L40 (48GB, VDI 최적화) L4 (24GB, 유니버설) RTX 6000 Ada 5000 Ada 4500 Ada 4000 Ada 4000 SFF Ada 2000 Ada	존재하지 않음 (해당 세대의 고성능 데이터센터 연산 포지션은 Hopper 아키텍처가 전담)
Ampere 세대 (2020 ~ 2022)	GeForce RTX 3090 Ti GeForce RTX 3090 GeForce RTX 3080 Ti GeForce RTX 3080 GeForce RTX 3070 Ti GeForce RTX 3070 GeForce RTX 3060 Ti GeForce RTX 3060 (12GB/8GB) GeForce RTX 3050	A40 (48GB) A16 (64GB, VDI 전용 쿼드 GPU) A10 (24GB) RTX A6000 A5500 A5000 A4500 A4000 A2000	A100 80GB (SXM4 / PCIe) A100 40GB A30 (MIG 특화) A2 (저전력 에지) A800 (규제 우회용)

세대 / 가로 구분	소비자용 PC GPU (GeForce)	워크스테이션 및 서버 그래픽 (RTX / L-시리즈)	AI 및 데이터센터 연산 가속기 (Compute)
Turing / Volta 세대 (2017 ~ 2020)	GeForce RTX 2080 Ti GeForce RTX 2080 Super GeForce RTX 2080 GeForce RTX 2070 Super GeForce RTX 2070 GeForce RTX 2060 Super GeForce RTX 2060 GTX 1660 Ti GTX 1660 Super GTX 1660 GTX 1650 Super GTX 1650 GTX 1630 NVIDIA TITAN V (Volta) NVIDIA TITAN RTX (Turing)	Tesla T4 (16GB, 싱글슬롯 유니버설) Quadro GV100 (Volta 기반) Quadro RTX 8000 (48GB) Quadro RTX 6000 (24GB) Quadro RTX 5000 Quadro RTX 4000 T1000 T600 T400	Tesla V100 32GB (SXM2 / PCIe) Tesla V100 16GB Tesla V100S
Pascal 세대 (2016 ~ 2017)	GeForce GTX 1080 Ti GeForce GTX 1080 GeForce GTX 1070 Ti GeForce GTX 1070 GeForce GTX 1060 GeForce GTX 1050 Ti GeForce GTX 1050 GT 1030 NVIDIA TITAN X (Pascal) NVIDIA TITAN Xp	Quadro P6000 Quadro P5000 Quadro P4000 Quadro P2200 Quadro P2000 Quadro P1000 Quadro P620 Quadro P600 Quadro P400	Tesla P100 (SXM2 16GB / PCIe 12GB, 16GB) Tesla P40 (24GB, 초기 대용량 추론) Tesla P4 (8GB, 저전력 가속)
Maxwell 세대 (2014 ~ 2015)	GeForce GTX 980 Ti GeForce GTX 980 GeForce GTX 970 GeForce GTX 960 GeForce GTX 950 GeForce GTX 750 Ti GeForce GTX 750 NVIDIA TITAN X (Maxwell)	Tesla M60 (Dual-GPU, 고성능 vWS용) Tesla M10 (Quad-GPU, 고밀도 vPC용) Quadro M6000 M5000 M4000 M2000 K1200 M620	Tesla M40 (24GB/12GB 연산 전용) Tesla M4 (8GB 슬롯형 연산)
Kepler 세대 (2012 ~ 2014)	GeForce GTX 780 Ti GeForce GTX 780 GeForce GTX 770 GeForce GTX 760 GeForce GTX 690 GeForce GTX 680 GeForce GTX 670 GeForce GTX 660 Ti GeForce GTX 660 GeForce GTX 650 Ti GeForce GTX 650 GT 640 NVIDIA GTX TITAN TITAN Black TITAN Z	GRID K2 (Dual-GPU, 하이엔드용) GRID K1 (Quad-GPU, 고밀도 VDI용) Quadro K6000 K5200 K5000 K4200 K4000 K2200 K2000 K620 K600 Quadro K420	Tesla K80 (Dual-GPU, 24GB) Tesla K40 (12GB 연산용) Tesla K20X Tesla K20 Tesla K10