

HP ZGX Nano AI Station

SG사업본부 | IDSS사업부



ETEVERS
eBT

목 차

1. 제품 소개
2. HP ZGX Nano AI Station 성능
3. 합리적인 대안
4. HP ZGX Nano AI Station의 장점
5. QnA

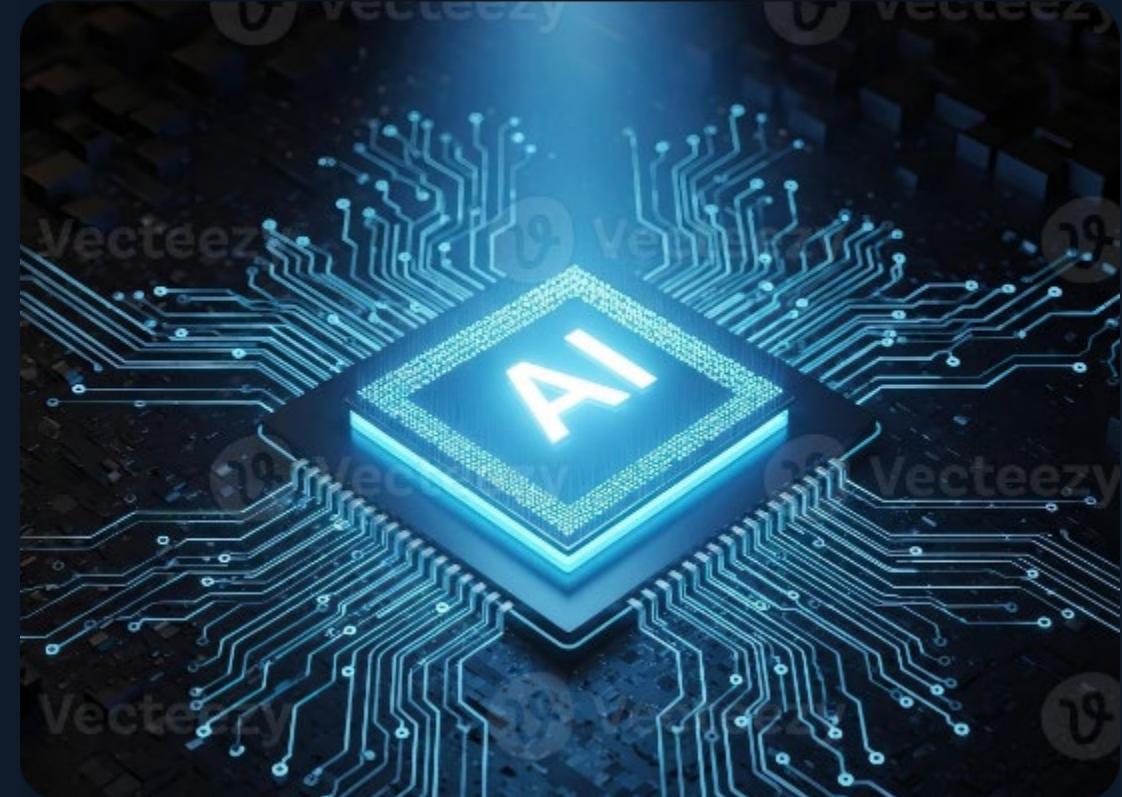
1. 제품 소개

1. 제품 소개

한계를 뛰어넘는 AI 퍼포먼스

HP ZGX Nano AI Station은 공간 제약이 있는 환경에서도 데이터 센터
급의 AI 추론(Inference) 및 데이터 처리가 가능하도록 설계되었습니다.

- ✓ 혁신적인 품팩터 : 기존 타워형 워크스테이션 성능을 1/10 크기로 압축
- ✓ AI 전용 설계 : NVIDIA RTX™ Professional GPU 최적화
- ✓ 무중단 안정성 : 24/7 가동을 위한 군사 표준(MIL-STD) 내구성 테스트
통과
- ✓ 조용한 작업 환경 : 고부하 작업 시에도 정숙함을 유지하는 쿨링 시스템



1. 제품 소개

차세대 AI 컴퓨팅

NVIDIA Grace Blackwell Superchip :

20코어 Arm CPU 및 Blackwell GPU 아키텍처를 탑재한 GB10 칩으로
구동됨으로 탁월한 AI추론 제공 가능.

128GB Unified Memory :

273GB/s 대역폭을 보유한 초고속 LPDDR5x 메모리를 통하여 대규모
LLM(대규모 대형 언어 모델)을 로컬에서 로드 가능.

Extreme Connectivity :

200Gbps QSFP 포트 듀얼과 10GbE 이더넷을 탑재해 방대한 데이터셋과
클러스터 네트워킹 지원.

NVIDIA ConnectX-7 Networking :

ZGX Nano AI Station 2대를 연결하여 최대 4050억(405Billion) 파라미터의
AI모델 작업 가능.

Nano Form Factor :

15cm x 15cm 크기의 컴팩트한 크기는 어떤 책상에도 맞는 워크스테이션급의
크기를 제공.



1. 제품 소개

제품 상세 Specifications



Chassis	Nano desktop, 150mm x 150mm x 51mm, HP ID
SOC/Architecture	NVIDIA GB10 Grace Blackwell Superchip
Memory	128GB LPDDR5x coherent unified system memory, 16 channels, 273 GB/s
Storage	1x PCIe Gen5 NVMe M.2 drive 1TB and 4TB capacities, self-encrypting*
Connectivity	Rear: 4x USB3.2 Type-C, 20Gbps (3 with PD out, 1 with PD in) Bluetooth 5.3
Networking	Wi-Fi 7 Ethernet 1x 10Gb, RJ-45 Connect-X 7: 2x 200G, QS Support for connecting two ZGX Nano units via CX-7 QSFP port
Display Output	1x HDMI 2.1a DP 1.4 over USB-C
Security	TPM 2.0
Peripherals	No keyboard or mouse included with system
Chassis LED	Power LED
Power	240W USB-C external power adapter
Regulatory	TAA Compliant
BIOS	HP Partner BIOS
OS	NVIDIA DGX Base OS, Ubuntu Linux
Preloaded Software	Co-branded setup wizard NVIDIA DGX AI Enterprise Software Stack

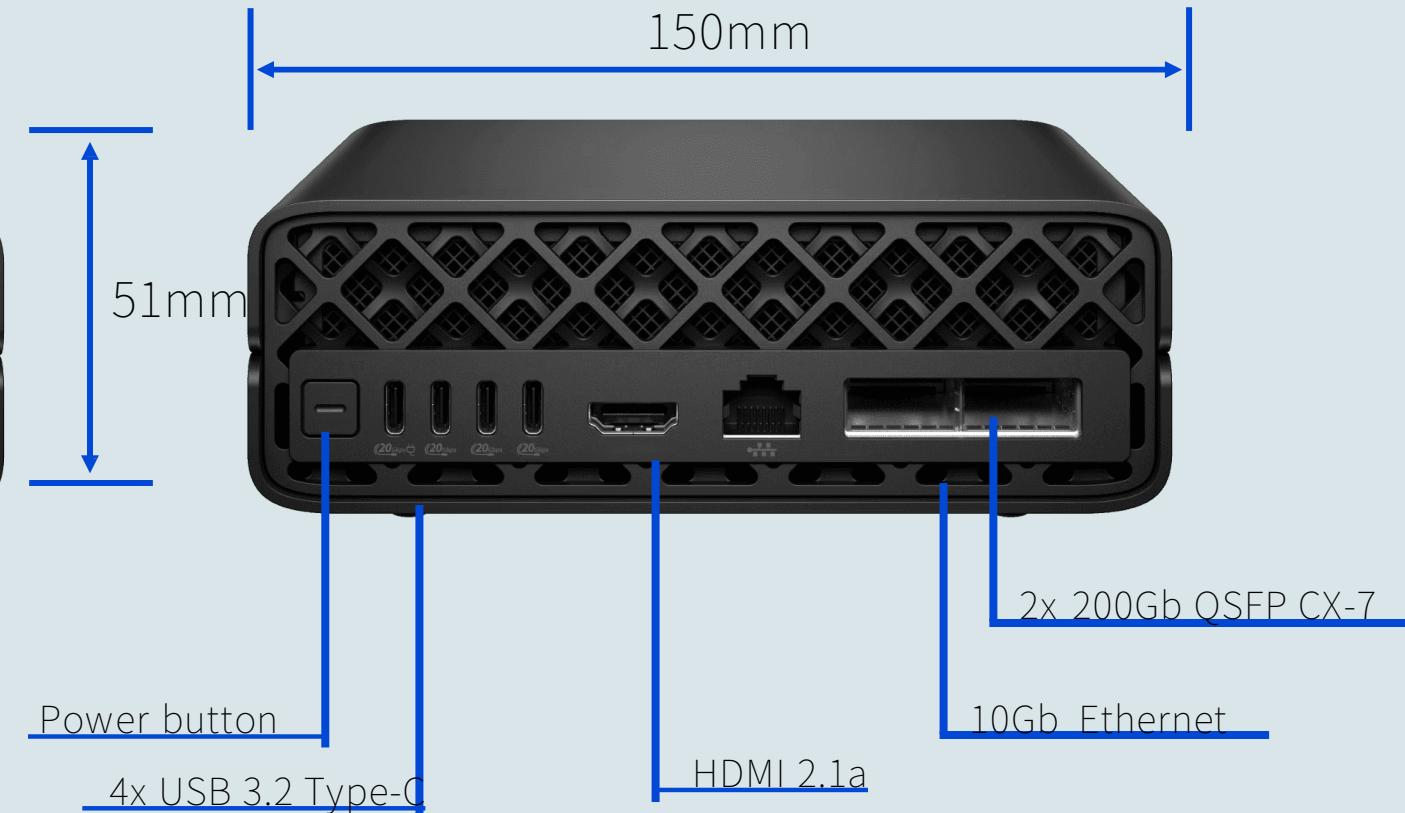
1. 제품 소개

실물 제품 규격

Front View



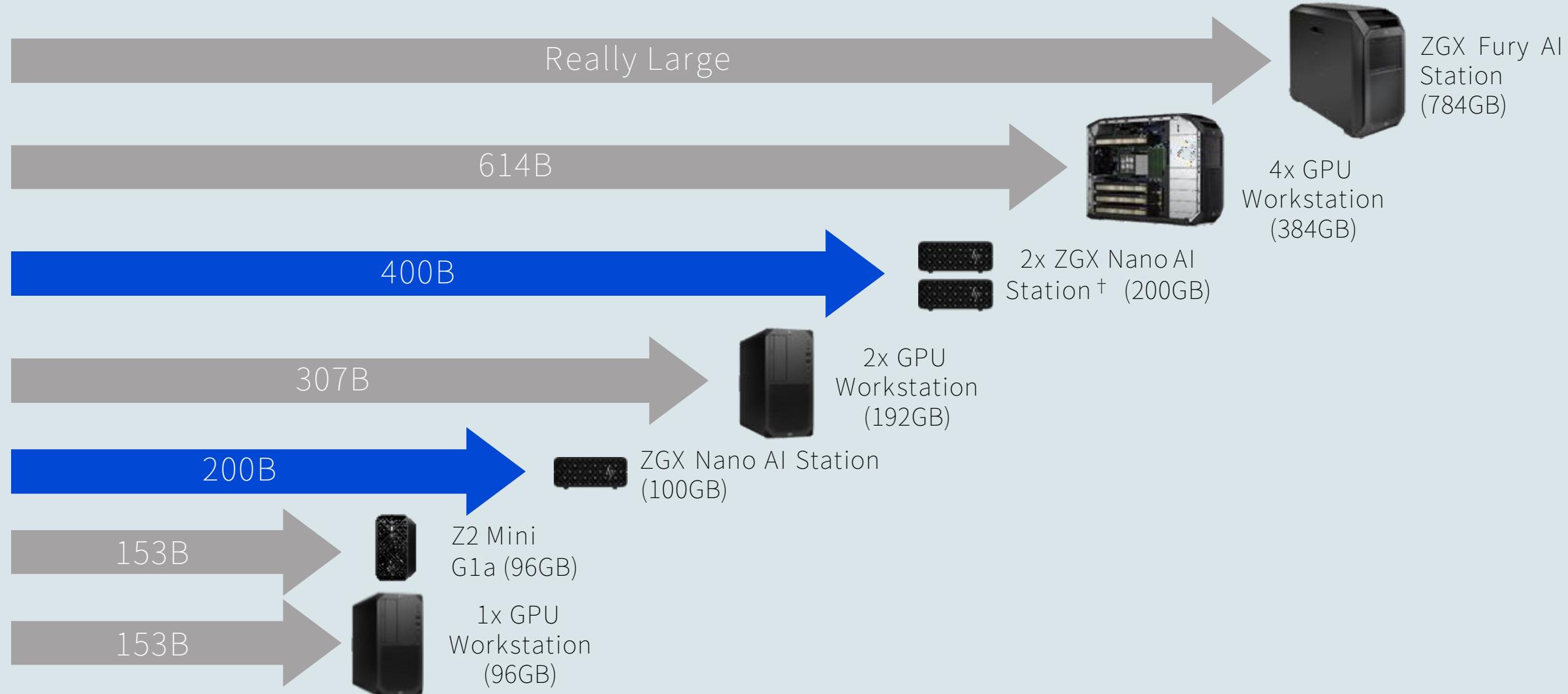
Rear View



2. HP ZGX Nano AI Station 성능

2. HP ZGX Nano AI Station 성능

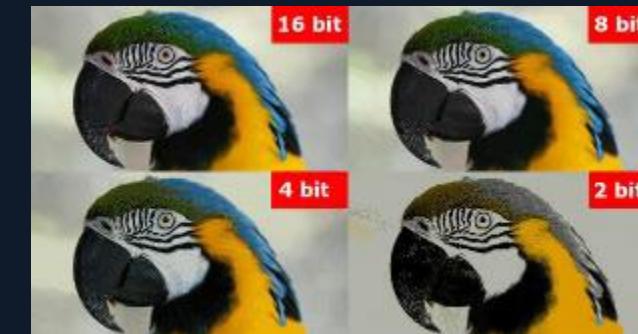
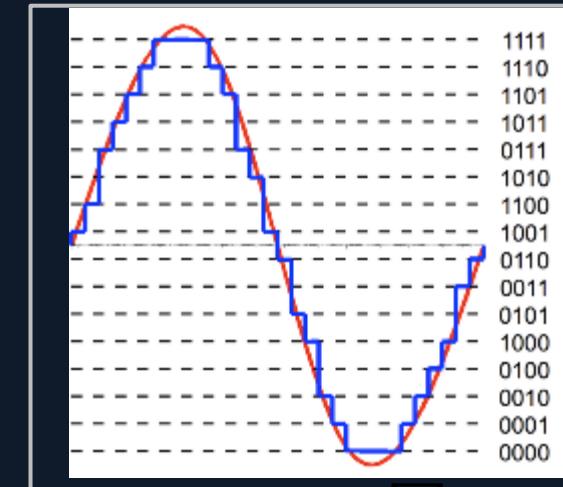
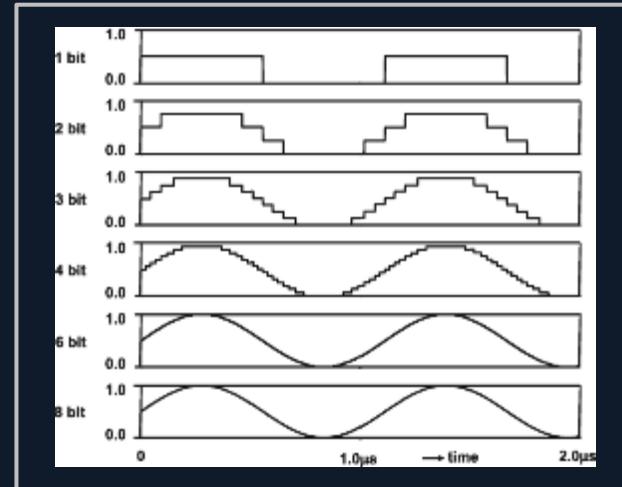
AI 개발자 시스템 : 가장 큰 AI모델 크기 (사용 중 FP4 Data 체재)



2. HP ZGX Nano AI Station 성능

최고의 ZGX Nano 성능 : FP4 (4비트 부동 소수점)

Nano는 GPU에 최대 100GB 시스템 메모리 사용 가능 → 200B 매개변수 LLM 활성화



LLM 양자화는 이미지의 색상 깊이와 매우 유사 합니다. 비트 수를 낮출 경우 용량은 줄어들지만, 그만큼 품질이 떨어지게 됩니다.

4비트 양자화: 원본 데이터 \otimes FP4 (4비트 부동 소수점)

Use case	ZGX Nano가 적합할까요?	설명
고속 NLP 추론	<input checked="" type="checkbox"/> 예	FP4는 이 부분에서 탁월합니다
생성적 이미지 생성	<input checked="" type="checkbox"/> 예	FP4는 이를 잘 지원합니다.
과학 등급 방사선학 AI	<input checked="" type="checkbox"/> 주의 요망	DGX H100이나 그 이상의 워크스테이션을 사용하세요
원거리 산업용 OCR/LPR	<input checked="" type="checkbox"/> 주의 요망	FP4는 OCR 정확도를 유지하지 못할 수 있습니다. 철저히 테스트하세요.

2. HP ZGX Nano AI Station 성능

첫 파인튜닝(Fine-Tuning) Experience 비교

HP ZGX Nano G1n AI Station VS DGX Spark 기반 솔루션

	HP ZGX Nano & Toolkit ¹	DGX Spark + NVIDIA® SW stack	Public Cloud instances / 서비스
 설정 시간:	Minutes	Variable (minutes to hours)	Variable (minutes to hours)
 미세 조정 속도:	Fast (local optimized)	Moderate (manual setup, less optimization)	Variable (Cloud latency, data transfer)
 스택 유연성:	High	Medium	Medium
 동반자 워크플로:	Seamless	Manual	Not available
 비용 효율성:	High	Medium (licenses fees)	Low (unpredictable ongoing fees)
 데이터/자산 제어:	Full control	Full control	Cloud Service EULAs / shared risk models

2. HP ZGX Nano AI Station 성능

HP ZGX Nano : 경쟁사 대비 우위



자제

Full Metal Chassis

발열제어 기능

솔루션

HP Tool Kit

Z by HP AI Studio 연동

업계 유일 전용 SW제공

Toolkit을 통한 LLM설치

타 경쟁사

일부 Plastic 자재 활용

자체 서버에 데이터 전송 SW만 제공

ConnectX-7 연결 시 구성 이슈 발생

별도 시스템을 통한 LLM설치 필요

3. 합리적인 대안

3. 합리적인 대안

HP Z2 Mini G1a Desktop Workstation



"비용 효율적인 AI 입문용 워크스테이션"

ZGX의 고성능이 부담스러운 고객을 위해, [HP Z2 Mini G1a](#)를 제안합니다. 일반 데스크탑보다 강력하지만, ZGX보다 가벼운 워크로드에 최적화 되어 있습니다.

- ✓ **탁월한 가성비**: 예산 내에서 AI 추론 환경 구축 가능
- ✓ **공간 절약**: 모니터 뒤, 책상 아래 어디든 설치 가능
- ✓ **충분한 성능**: 기본적인 머신러닝 모델링 및 3D 디자인 작업 원활



3. 합리적인 대안

HP ZGX VS HP Z2 Mini G1a 비교 제안

HP ZGX Nano AI

Premium Performance

- ✓ 타겟 : AI 연구소, 데이터 사이언티스트
- ✓ 용도 : 딥러닝 학습, 대용량 렌더링
- ✓ GPU : NVIDIA Blackwell
- ✓ 확장성 : 매우 높음 (NVMe 4슬롯 등)

HP Z2 Mini G1a

Cost Effective

- ✓ 타겟 : 소규모 연구조직, 일반 디자이너
- ✓ 용도 : AI 추론(Inference), CAD/BIM
- ✓ GPU : AMD Radeon 8060S 그래픽
- ✓ 비용 : 도입 부담이 낮은 합리적 가격



예산이 제한적이거나 단순 추론(Inference) 위주의 고객에게는 Z2 Mini G1a가 최적의 선택입니다.

4. HP ZGX Nano AI Station의 장점

4. HP ZGX Nano AI Station의 장점

4.1. Why HP ZGX Nano AI Station?



경쟁사는 단순히 “Linux / Ubuntu” 컴퓨터를 판매하는 반면,
HP에서는 윈도우 개발자가 바로 사용할 수 있는 “AI개발 환경”을 제공 합니다.
Visual Studio Code에서 HP Toolkit을 통하여 PC세팅에 시간을 낭비하지 않고, 바로 “Hands-on”을 할 수 있습니다.



HP의 다양한 제품을 Bundle하여 하나의 묶음 상품으로 판매가 가능합니다.
- HP는 워크스테이션 뿐만 아닌 디스플레이, 키보드, 마우스 및 도킹스테이션 등 다양한 제품 라인업을 보유하고 있습니다.



Only from HP
- HP Toolkit 제공.
- Market Share 1등 HP 제품 사용을 권장 합니다. (워크스테이션 부문 국내 1위)
- 뛰어난 AS, 장애지원 및 Customer Service를 제공합니다.
(전 세계에서 가장 많은 AS지원센터 보유)

4. HP ZGX Nano AI Station의 장점

4.2. Why HP ZGX Nano AI Station?



Cloud 비용 절감

- AWS, Azure 등 클라우드에서 GPU인스턴스를 임대하여 튜닝을 하거나, 개발자가 코딩을 안하고 있는 시간에도 Cloud 비용은 초당 과금 됩니다.
- HP ZGX nano를 통하여 무제한 튜닝 및 개발 QC가 가능한 환경 구성이 가능합니다.
모든 QC가 종료 되었을 때, Cloud를 통한 배포를 진행하면 용이 합니다.
(SI 및 SW 개발사 그리고 LLM을 활용한 연구를 진행하는 사내 연구진에는 비용 절감을 할 수 있습니다)



데이터 보안

- 오픈 소스 LLM을 활용하여 사내 챗봇 또는 문서 파싱(Document Parcing) 기능을 개발하는 것은 보안에 매우 취약하고, 위반되는 행위입니다.
- HP ZGX nano AI Station을 활용한다면, 오픈소스 LLM을 AI Station에 구축하고 민감한 사내 문서 및 데이터로 테스트가 가능하게 됩니다.

4. HP ZGX Nano AI Station의 장점

4.3. 적용 가능 분야

Proto-Typing



Fine-Tuning



추론



데이터 사이언스



엣지 응용



로봇 공학



컴퓨터 비전



VLM



LLM,
생성 AI

서버 및 클라우드와
의 경쟁

5. QnA

QnA



네트워크 대역폭에 대한 이슈?

- 사용자가 많을 때 네트워크 대역폭 보다는 연산대기 “Queuing”으로 인한 지연 발생 가능성이 높습니다.
- 물리적인 네트워크 대역폭은 매우 넉넉합니다. (RJ45 포트 / 200GbE 포트 탑재).
따라서, 물리 텍스트나 코드를 주고 받는 일반적인 회선량으로는 대역폭 이슈가 발생하지 않습니다.

LLM 동시접속자가 많을 경우에는?

- 다수의 인원이 무거운 AI모델의 튜닝을 진행하는 경우 회선량이 문제가 된다기 보다는, 연산 순서를 기다리느라 응답이 느려질 수는 있습니다.
- ZGX Nano 제품은 대규모(대국민)서비스의 사용 보다는 전문가를 위한 개인 장비 또는 소규모 팀의 형태에서 가장 쾌적한 성능을 보여줍니다.

Windows OS 지원이 가능한지?

- ZGX Nano는 Linux 기반의 Ubuntu만 지원함으로 Windows OS는 지원이 불가합니다.
- 단, 사용자는 Windows 11에 설치된 PC, Workstation 또는 노트북에서 VS Code를 활용하여 HP Toolkit을 활용하여 사용할 수 있습니다.

기존 장비를 어떻게 그대로 사용 한다는 건지?

- ZGX Nano는 HP Toolkit을 제공하고 있습니다. 해당 Toolkit에는 Visual Studio Code 전용 확장 프로그램을 제공합니다.
따라서 Windows 사용자가 VS Code를 설치한다면, 복잡한 IP설정 없이 네트워크 상의 ZGX 장비를 바로 검색하여 찾아주고 연결해줍니다.

Thank you

에티버스아이비티 | SG사업본부