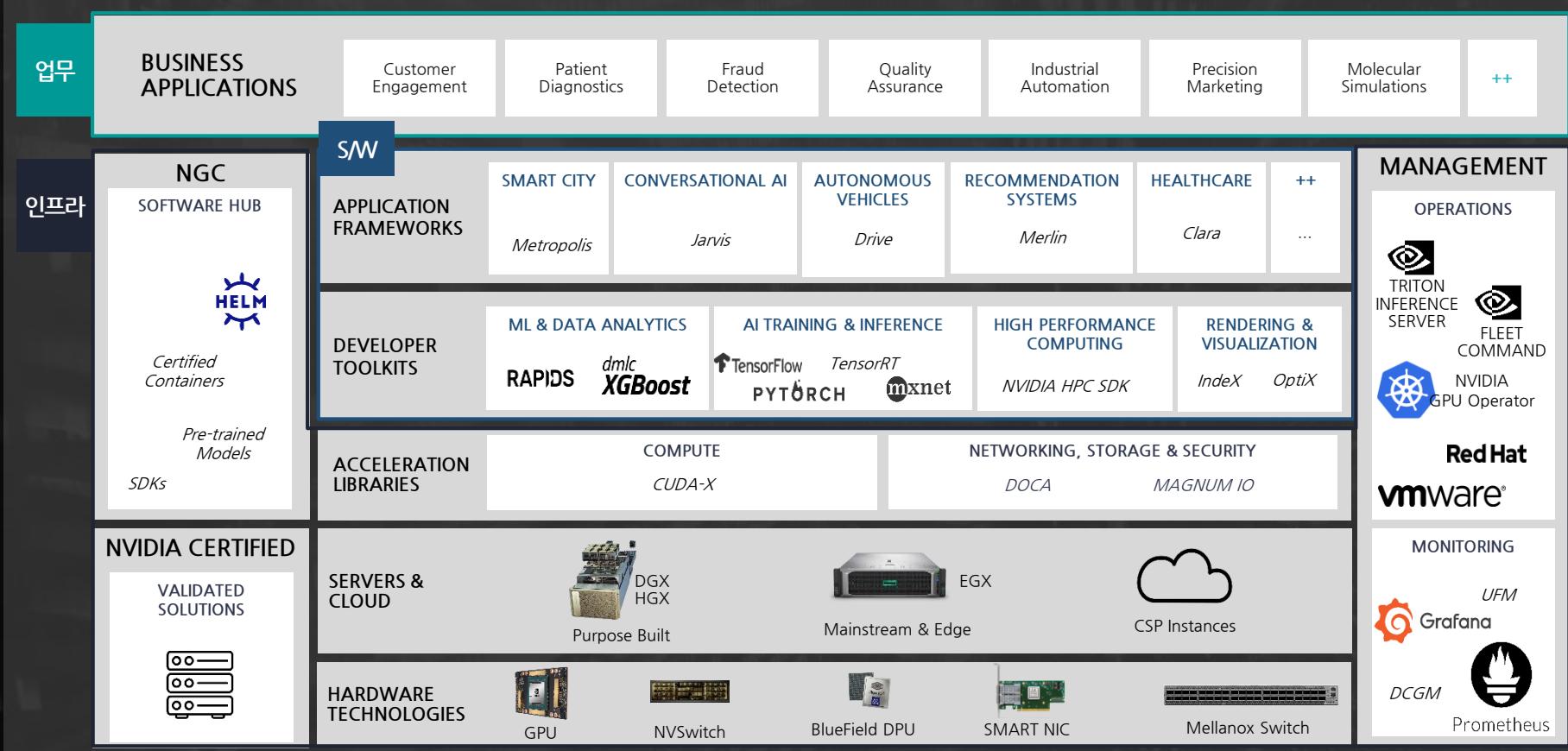


비즈니스 성과를 위한 AI플랫폼 도입 고려사항

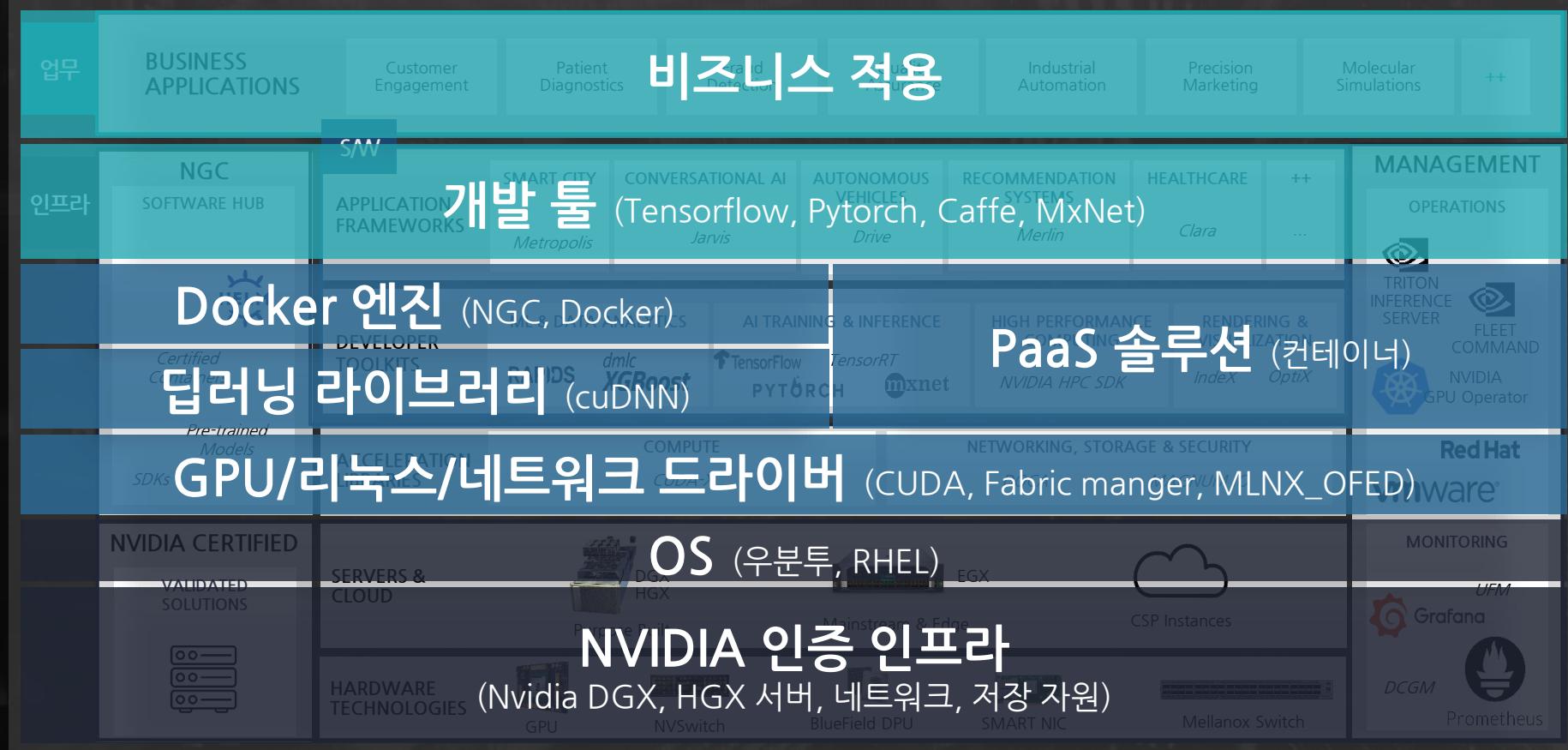
효성인포메이션시스템

김형섭 차장
HPC사업팀

NVIDIA의 AI Datacenter



AI 플랫폼의 단순화



AI플랫폼 도입에 대한 고객의 고민

비즈니스 성과



성능

성능 최적화

- 기존 전통 인프라의 낮은 성능
- 연산 및 I/O 성능 개선



비용

자원 효율

- 한정된 GPU 연산 자원
- 늘어나는 데이터 저장 효율

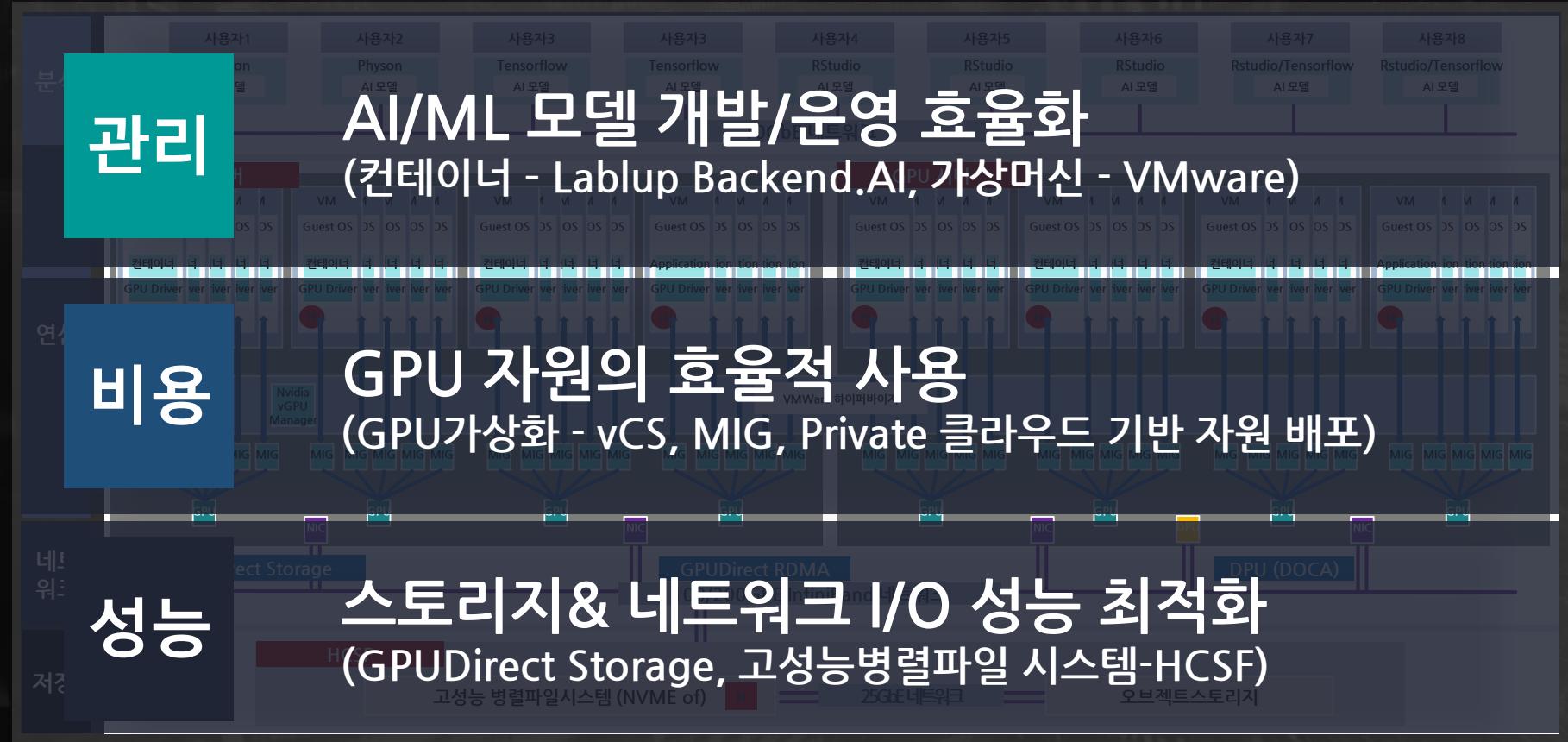


관리

개발/운영

- AI모델 개발 및 운영
- 쉬운 컨테이너 운영 환경

효성의 통합 AI 플랫폼 구현 예시



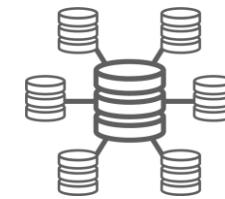
성능을 위한 기본 조건



장치 간 IO 성능 최적화
Magnum IO 구성



NVIDIA NVLink,
NVSwitch 지원 GPU 서버



AI 업무 전용 NVME
고성능 병렬 파일 스토리지

최적화

연산

저장

GPUDirect Storage

Supernicro HGX

HCSF

성능 - 최적화 (AI 업무를 위한 Extream IO)

- 대형 AI모델의 학습 성능을 위해서는 GPU연산 및 저장 자원의 성능과 함께 IO성능 최적화가 중요
- NVIDIA에서는 MAGNUM IO라는 GPUDirect 기술을 통해 IO성능 최적화를 지원
- NVLINK를 지원하는 GPU서버와 고성능 병렬 파일스토리지 인프라에 GPUDirect 기술을 적용하여 구현

MAGNUM IO

Storage IO

GPUDirect Storage
SNAP

Network IO

GPUDirect RDMA
NVSHMEM

IN-Network Compute

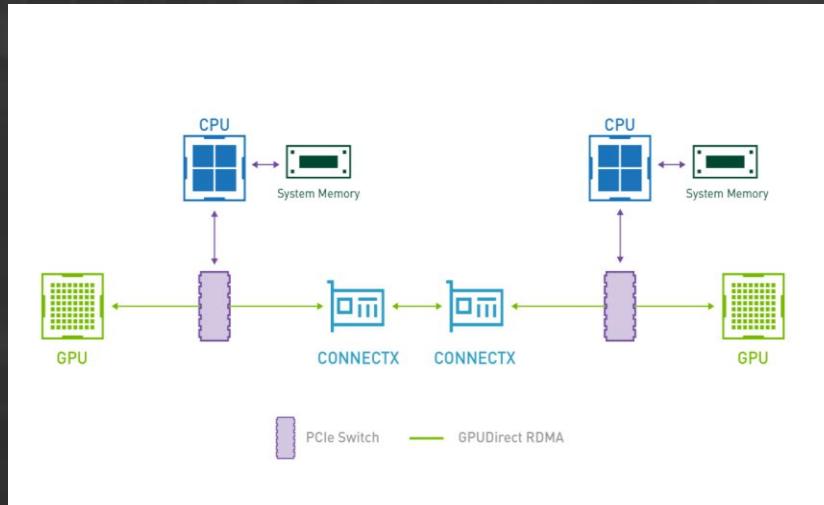
BlueField DPU

IO Management

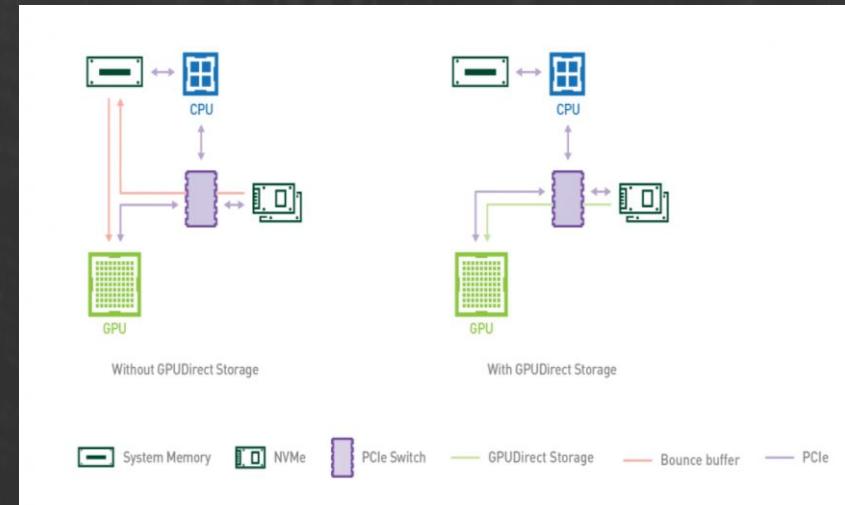
Ethernet NetQ
UFM

성능 - 최적화 (AI 업무를 위한 Extream IO)

네트워크 I/O : GPUDirect RDMA



스토리지 I/O : GPU Direct Storage



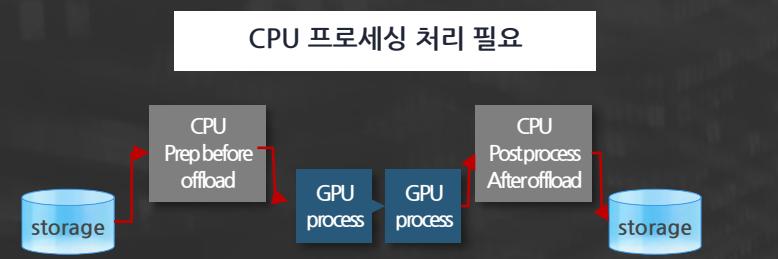
- RDMA를 통해 PCIe 가 GPU 메모리에 직접 액세스
- 원격 시스템에서 NVIDIA GPU 간의 직접 통신을 제공
- CPU와 메모리 데이터 버퍼를 제거, 10배 성능 향상

- NVMe /NVMe-oF 스토리지와 GPU 메모리 간 직접 연결
- CPU 메모리의 바운스 버퍼를 제거
- 스토리지와 GPU 메모리의 데이터 로드 프로세스 IO 개선

성능 - 최적화 (GPUDirect Storage)

- GPUDirect 기술을 적용하여, CPU 대비 빠르고 많은 대역폭의 IO 처리 가능
- NVIDIA GPU 서버와 HCSF 구성 시, 8 EDR LINKS & DGX서버 1대에서 약 80GB/s 고성능 제공

I/O 흐름 : CPU vs GPU



GPU서버 + HCSF 동시 구성 → 고성능 제공

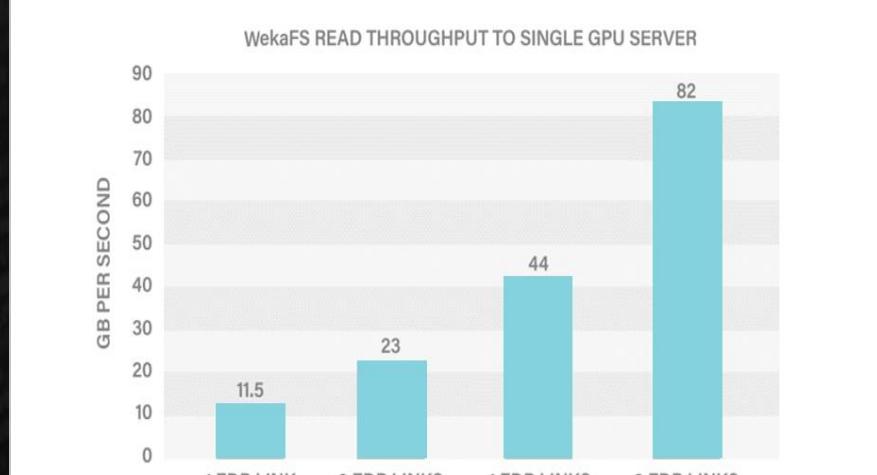
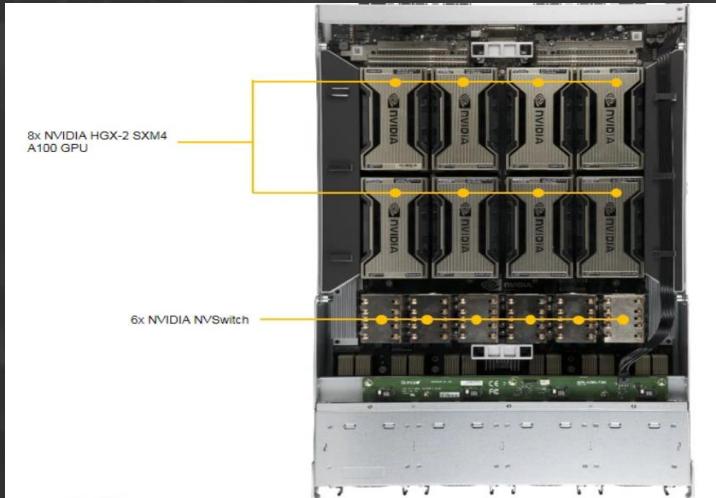


Figure 20: Performance Scaling Inside a Single NVIDIA DGX-2 GPU Server with GPUDirect Storage

성능 - 연산 (HGX, NVLink)

Nvidia 인증 Supermicro HGX



- 3세대 NVIDIA NVLink는 600GB/s GPU 대역폭으로 PCIe Gen4 대비 10배 고성능
- NVIDIA A100 Tensor 코어 GPU의 고속 상호 연결구현

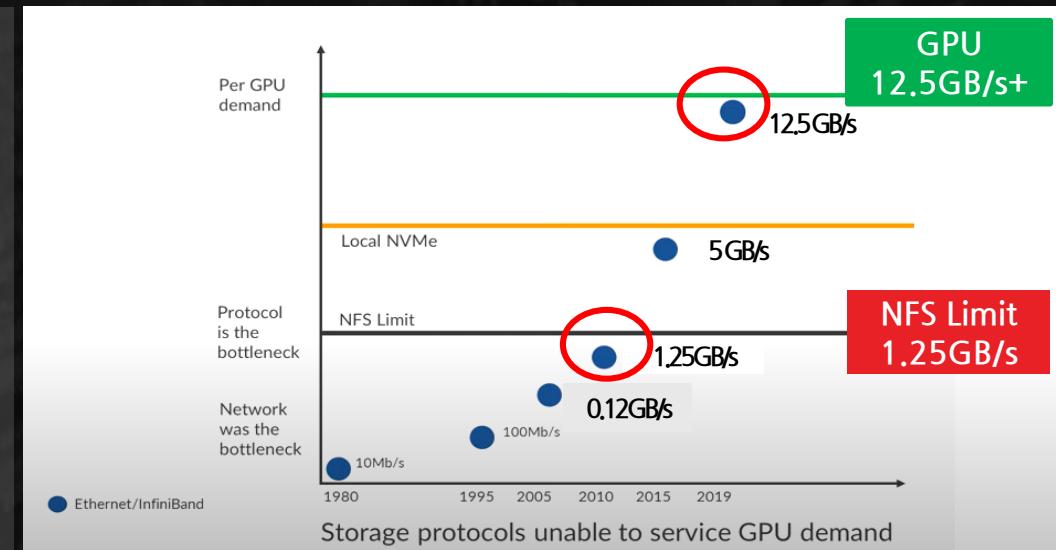
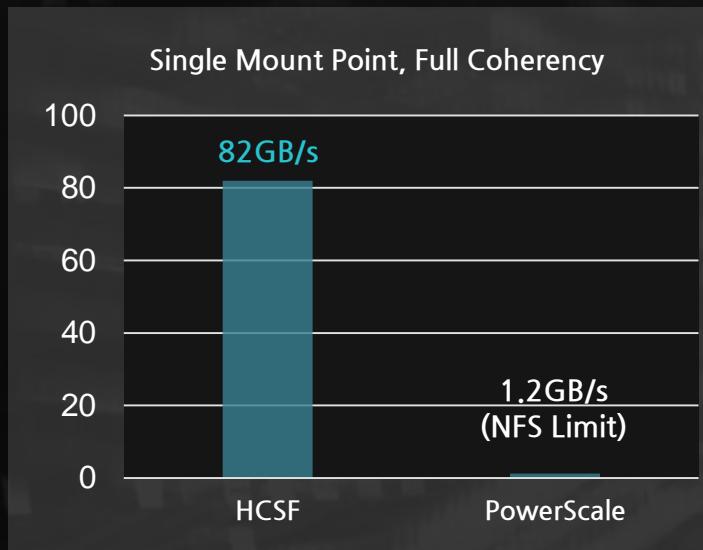
NVLink와 PCIe 비교

항목	서브 링크 속도	서브 링크 수	전체 속도	GPU 아키텍쳐
PCIe 3.x	16GB/s	1	16GB/s	Pascal, Volta, Turing
PCIe 4.0	32GB/s	1	64GB/s	Volta, Ampere
NVLink 1.0	20GB/s	4	160GB/s	Pascal
NVLink 2.0	25GB/s	6	300GB/s	Volta
NVLink 3.0	25GB/s	12	600GB/s	Ampere
NVLink 4.0	25GB/s	18	900GB/s	Hopper

- GPU-GPU, CPU-GPU간 전송 기술로 GPU메모리 직접 통신
- NVLink 1.0(SXM), NVLink 2.0(SXM2), NV Switch (SXM3)로 NVLink 방식 발전

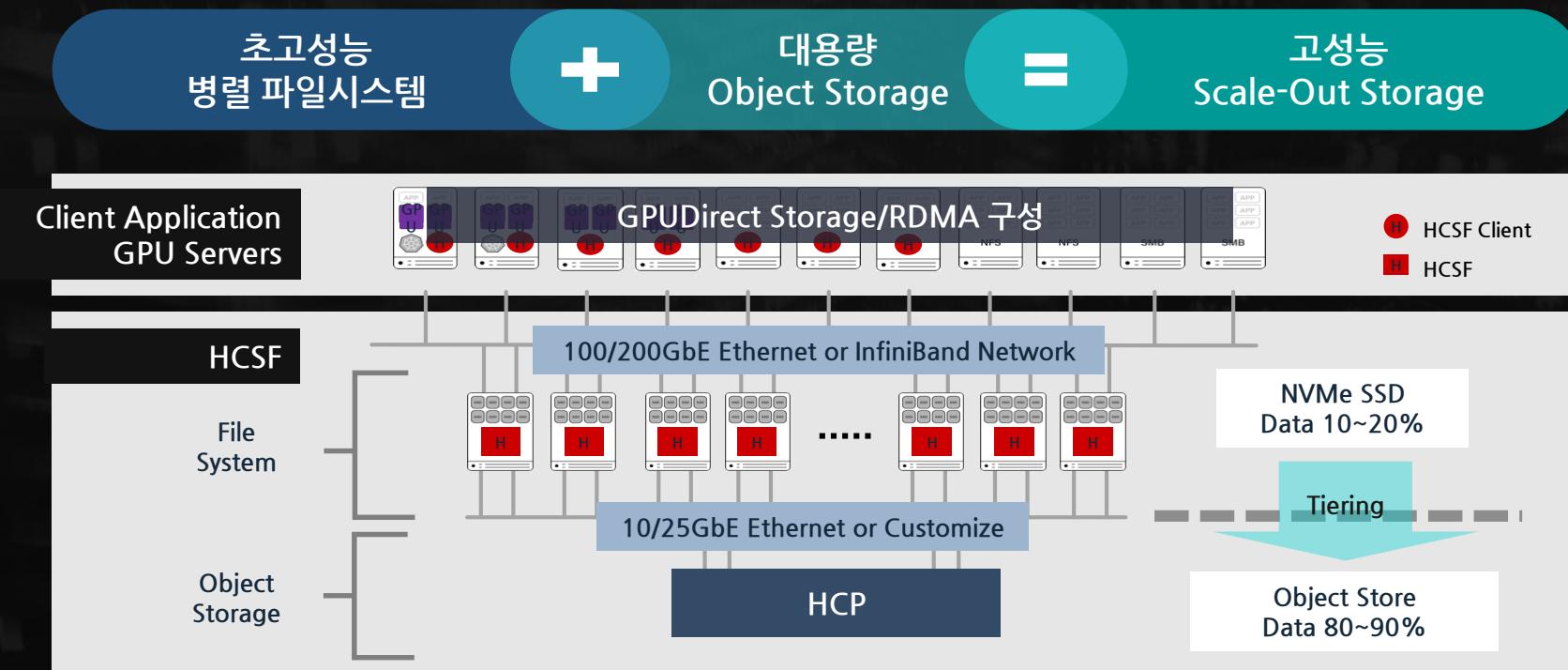
성능 - 저장 (HCSF)

- AI 분석 환경에 GPU 서버 연동은 필수이며, GPU 서버는 대부분 Nvidia사 주도
- 4 GPU 탑재 서버 경우 단일 서버에서 50GB/s 데이터 처리 필요
- NFS 기반 Legacy NAS는 1.2GB/s로 성능 병목 발생
- GPU기반 분석은 혼합 워크로드(throughput과 IOPs), 작은 파일에 대한 성능 그리고, 파일 수 제한 고려 필요



성능 - 저장 (HCSF)

- 효성의 HCSF는 NVIDIA 인증 GPU서버와 함께 GPUDirect Storage 구성으로 저장 성능을 최적화 합니다.



성능 - 저장 (HCSF)

HCSF는

GPUDirect Storage - Early Access Program

부터 지원한 파일시스템을 탑재

- WekaFS
- Lustre

HCSF는 다양한 GPU연동 사례 보유,
타사의 경우, 최근 지원 시작

Partner Company	Partner Product Version	Compatible GDS Version	Date
NetApp	ONTAP 9.10.1	1.0 and higher	Jan 2022
NetApp	BeeGFS Tech Preview	1.1.1 and higher	TBA
ThinkParQ			
System Fabrics Works			
IBM	Spectrum Scale 5.1.2	1.1 and higher	Nov 2021
DDN	EXAScaler 6.0	1.1 and higher	Nov 2021
VAST	Universal Storage 4.1	1.1 and higher	Nov 2021
WekaIO	WekaFS 3.13	1.0	June 2021
DellEMC	PowerScale 9.2.0.0	1.0	Oct 2021
Hitachi Vantara	HCSF	1.0	Oct 2021

NASA 고성능 스토리지 사례

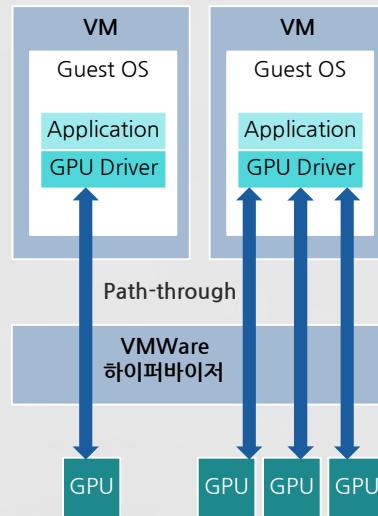


지구에서 화성으로 프로젝트에서 착륙을 위한
시뮬레이션

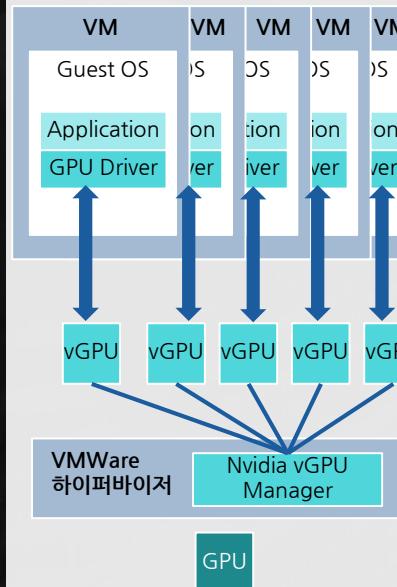
3,312 NVIDIA V100 Tensor Core GPU 연산으로
시뮬레이션 처리
(2019년, NASA Langley research center)

비용 - GPU자원 효율

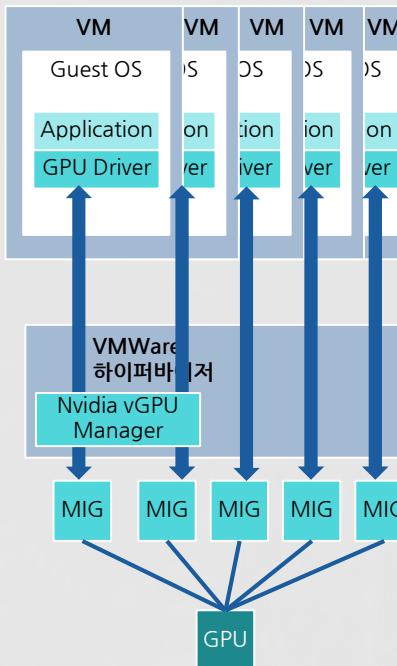
DirectPath I/O



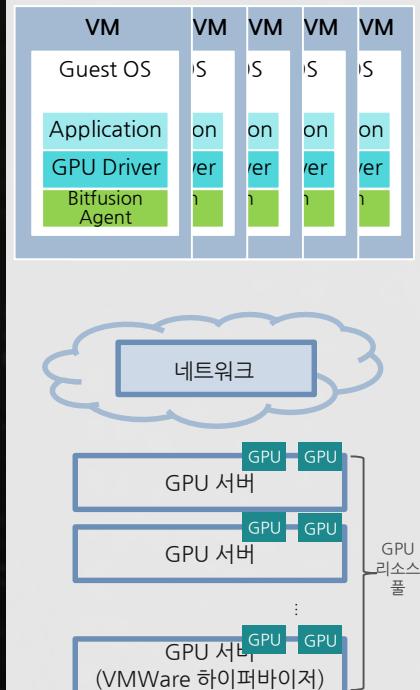
vGPU



MIG



Bitfusion



비용 - GPU자원 효율 (NVIDIA MIG)

MIG 지원 GPU 리스트

제품명	아키텍처	메모리	GPU최대 분할
A100-SXM4	NVIDIA Ampere	40GB	7
A100-SXM4	NVIDIA Ampere	80GB	7
A100-PCIe	NVIDIA Ampere	40GB	7
A100-PCIe	NVIDIA Ampere	80GB	7
A30	NVIDIA Ampere	24GB	4

구성 예



※ 선택 블록이 수직으로 겹치지 않는 구성 가능

3g. 40gb → GPU 메모리



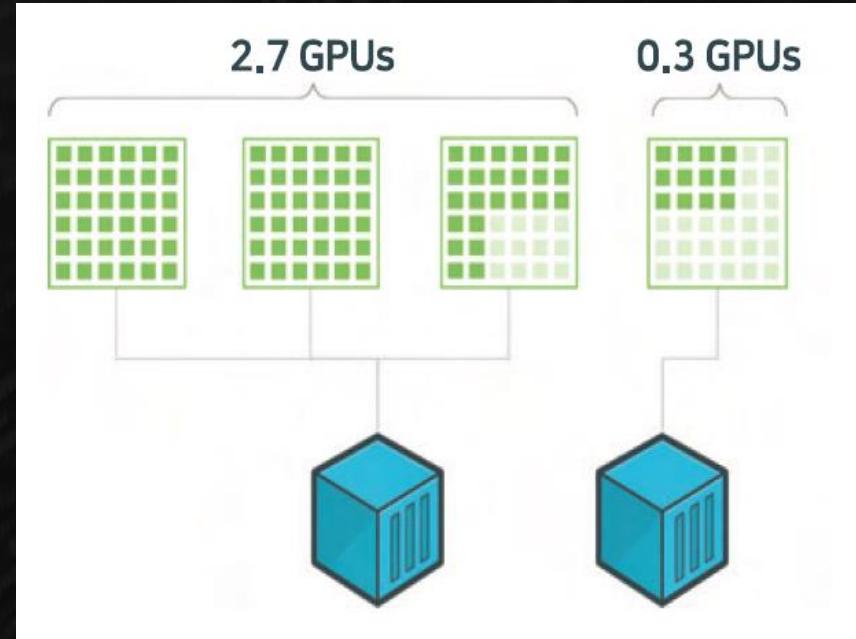
GPU Compute

비용 - GPU자원 효율 (컨테이너 기반 GPU분할)

Lablup Backend.AI 의 GPU 분할 가상화

컨테이너 기반 GPU 스케일링

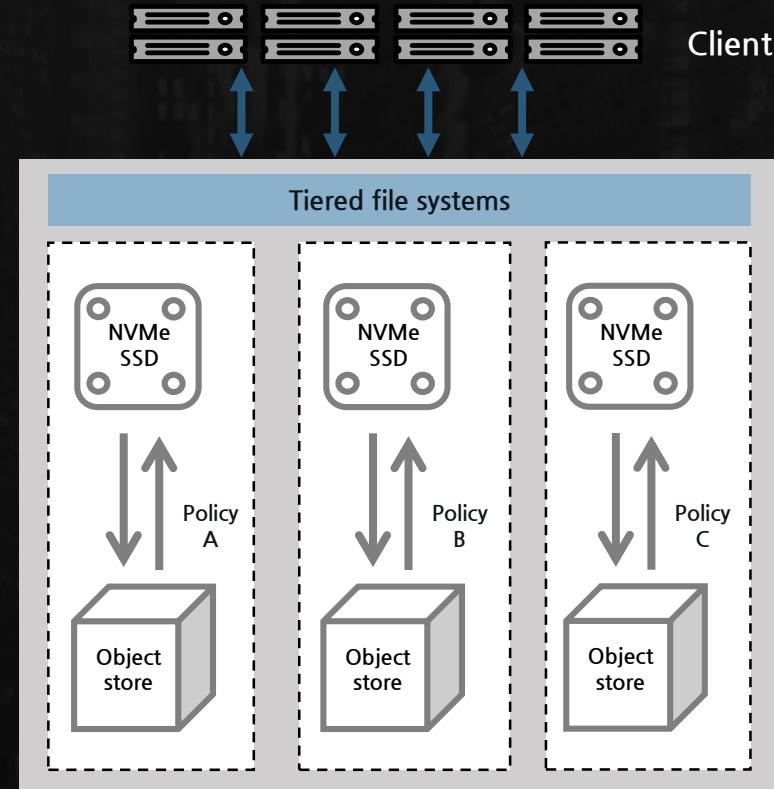
- ✓ 교육 및 추론 워크로드를 위한 단일 GPU 공유
 - ✓ 모델 훈련 등 대규모 워크로드를 위한 다중 GPU 할당
 - ✓ 자체 개발한 CUDA 가상화 계층으로 구현
- [Lablup Backend.AI 대한민국, 미국, 일본 등록 특허]



비용 - 저장자원 효율(오브젝트 스토리지 Tiering)

HCSF의 정책 기반 Tier 관리

- 정책 기반 Tiering으로 비용 대비 고성능, 대용량 제공
 - Client는 Tier 관련없이 파일시스템 엑세스
 - Tiering 정책 운영
 - 1) 최신 데이터는 HOT Tier에 저장/운영
 - 2) 정책에 의해 Cold data는 Object Store로 Auto Tiering
 - 3) 사용량에 따른 Tiering 정책 운영, 임계치 도달 시, 미사용 데이터 Object Store로 동
 - 각 Tier 운영에 대한 데이터 관리, 모니터링 지원



관리 - AI업무를 위한 엔터프라이즈 클러스터 운영 시스템

- Lablup Backend.AI는 아태지역 최초 NVIDIA DGX-Ready S/W로 검증된 AI 연구&개발 플랫폼으로 GPU분할 가상화를 제공하여, 데이터 과학자 및 AI플랫폼 담당자의 효율적 연산 자원 사용이 가능하게 합니다.



1 GPU 활용 극대화

- ✓ 컨테이너수준 GPU분할 가상화 지원
- ✓ NVIDIA GPU MIG 지원

2 직관적인 관리 및 사용자 경험

- ✓ GUI 기반 컨테이너 운영관리 지원
- ✓ 웹UI와 데스크탑 앱 동시 지원

3 AI 및 HPC 성능 최적화

- ✓ 독자적 엔진으로 최적의 GPU 연산 자원 배치 구현
- ✓ 다중노드 워크로드 및 데이터 I/O 병렬화 지원

4 쉬운 워크로드 확장

- ✓ 다중노드/다중 컨테이너 세션
- ✓ 모델훈련 및 데이터 I/O 파이프라인 분리

관리 - AI업무를 위한 엔터프라이즈 클러스터 운영 시스템

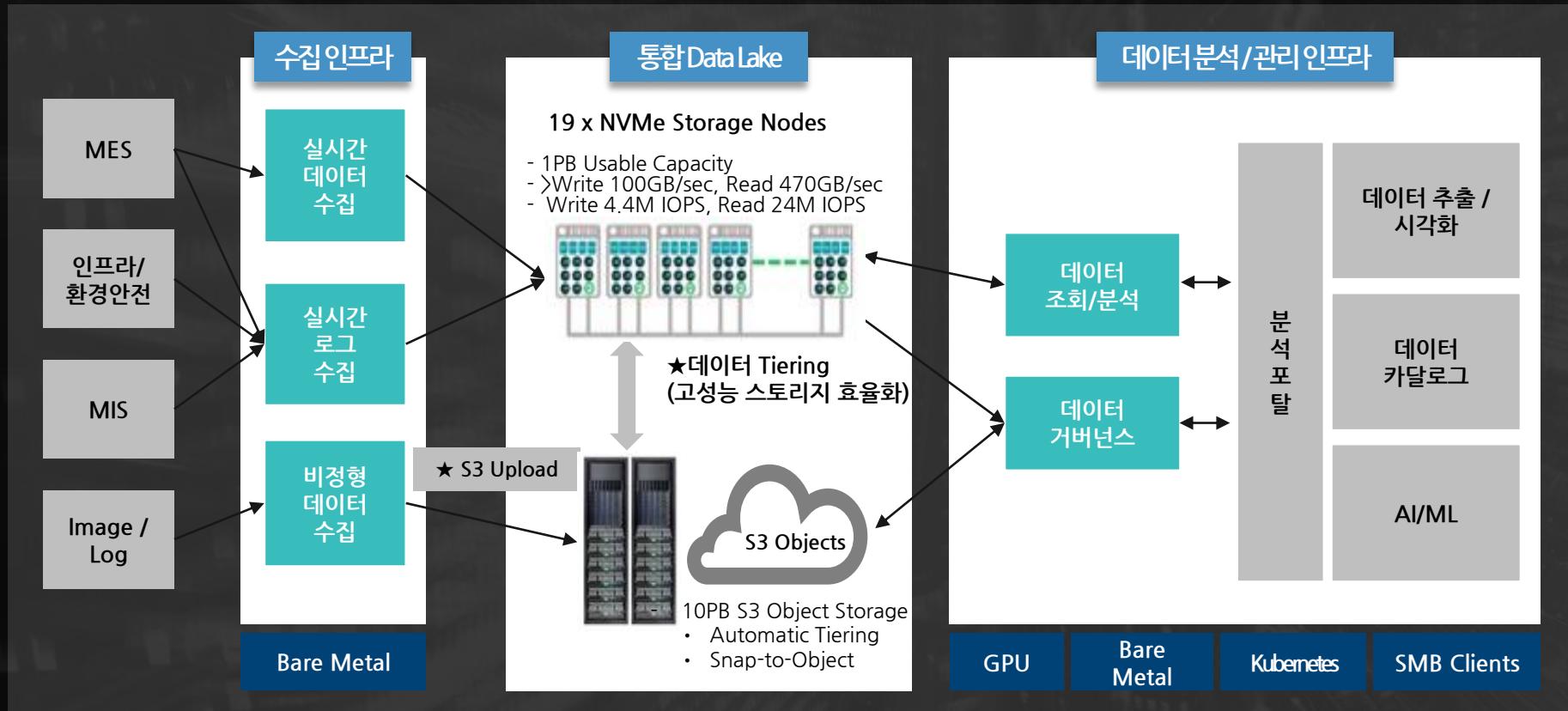
Lablup Backend.AI 경쟁 기술 플랫폼 비교

기술 사양		Nvidia-docker	Docker Swarm	OpenStack	Kubernetes	Apache Mesos	Lablup Backend.AI
GPU지원	컨테이너 별 GPU 할당	○			○	○	○
	이기종 가속기 지원 (AMD, Google TPU 등)						○
	GPU 부분 공유 (fractional scaling)						○
보안	Hypervisor 및 컨테이너에 의한 가상화	○	○	○	○	○	○
	프로그래밍 가능한 샌드 박싱						○
가상화	VM (Hypervisor)			○	○*	○*	○*
	Docker 컨테이너	○	○	○	○	○	○
스케줄링	슬롯 기반		○	○	○	○	○
	고급 (DRF 등)				○**	○	○
통합	최신 AI 프레임워크의 파이프라인 통합						○

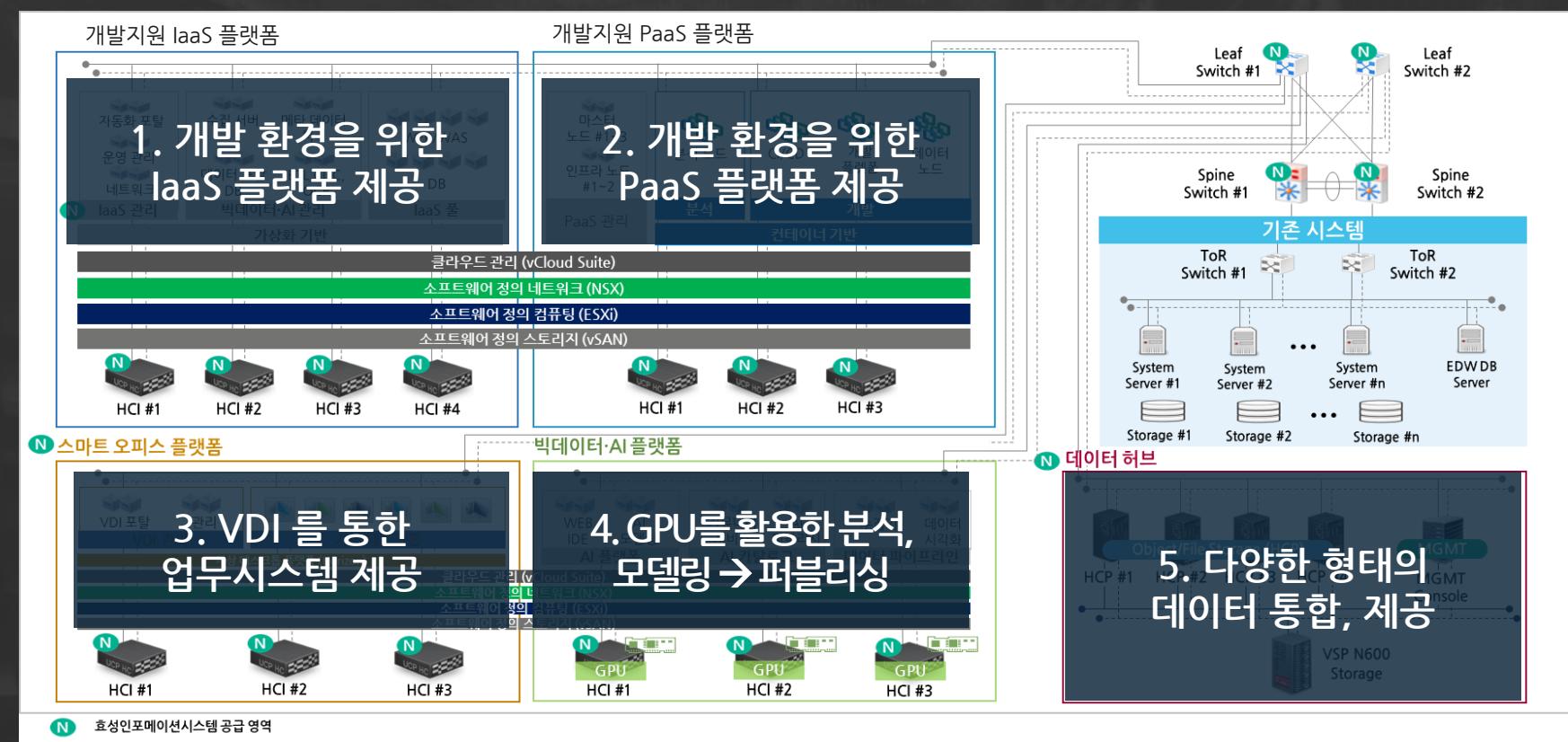
* 설치 된 클라우드 벤더나 VM환경에서 간접 지원

** 슬롯 기반의 한계가 있으나 label을 활용한 다양한 사용자화 지원

사례1- 전사 데이터 분석 체계의 통합 저장소 (제조)



사례2 - 통합 AI 서비스 환경의 기획/설계/구현 (공공)



사례3 - AI연구를 위한 GPU시스템 (해외 연구기관)

요구사항



- 의료영상 분석 (심장 MRI, 흉부&치과 CT), 코로나 증상 연구, 자율주행 및 로봇, 언어 모델 학습 업무
- GPU연산 환경 필요
NVIDIA의 P2P GPU 연결을 지원하는 고성능 연산 자원 필요

해결방안



- 2개의 AMD CPU와 8개의 NVIDIA HGX A100 GPU가 포함된 Supermicro GPU 서버 도입
- NVLink 및 NVSwitch와 연결된 여러 GPU를 사용
최대 600GB/s 양방향 대역폭 통신 GPU인프라 제공

성과



- 신규 알고리즘을 신속 구현으로 연구원의 연구 속도 개선
- 이전 시스템에 비해 알고리즘 연산 성능 약 20배 개선

A+ 4124GO-NRT

- 2 x 2nd Gen AMD EPYC™ CPUs 7F72 3.2 GHz
- GPU-A100-SXM4-8

효성이 가지고 있는 것

서비스	데이터 센터 설계 (SDDC, 클라우드, AI 분석 인프라 설계)	데이터 컨설팅 (데이터 처리 프로세스/모델 설계)	개발 (모델/ 데이터 처리/화면 개발)	
S/W	AI/ML/DL 개발툴 (Tensorflow, Pytorch, RAPIDS)	데이터 가공/분류 솔루션 (Pentaho/Lumada)	저장 솔루션 (데이터 마트, 하둡)	
인프라	인프라 최적화 (CUDA, vGPU, MIG, Bitfusion)	GPU 최적화 (CUDA, vGPU, MIG, Bitfusion)	GPU 인프라 성능 최적화 역량 (Magnum I/O)	DPU 최적화 (DOCA)
인프라	운영 시스템	Lablup 컨테이너 Backend.Ai (Lablup Backend.Ai, Nyng.Cube, Tanzu)	가상머신 기반 클라우드 (VM 구축 역량)	
H/W	Nvidia DGX & HGX (GPU 서버, Smart NIC, DPU)	Hitachi HCSF 저장 (Weka IO) (HCSF, 오브젝트 스토리지, SAN)	네트워크 (네트워크 스위치)	

효성의 AI플랫폼



비즈니스 성과를 위한 AI플랫폼 도입 고려사항 - 슈퍼마이크로 GPU 플랫폼

효성인포메이션시스템

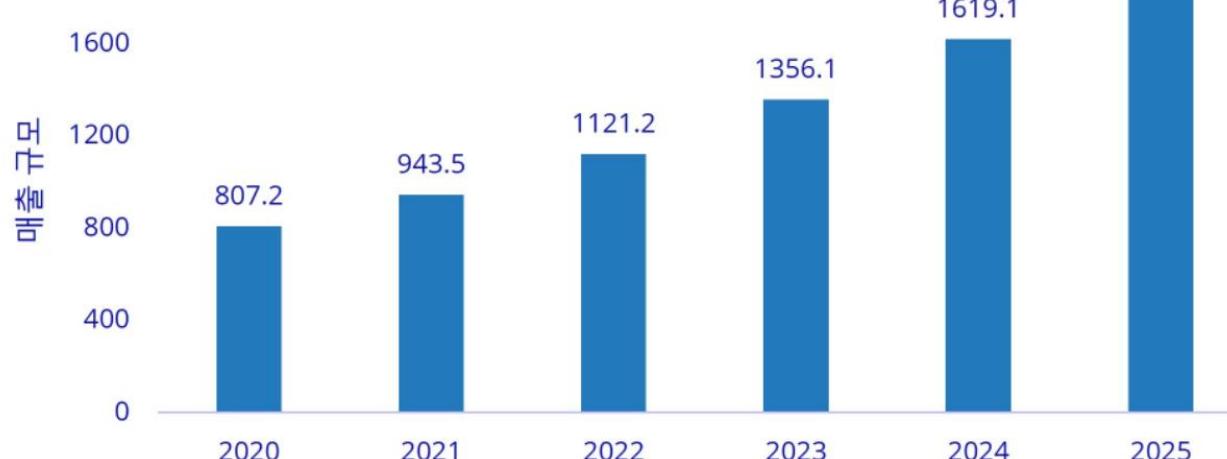
정문종 차장
HPC사업팀

국내 인공지능(AI) 시장 전망



(단위: 십억원)
2000

국내 인공지능(AI) 시장 전망 (2020년-2025년)



Source: IDC Semmiannual Artificial Intelligence Tracker, September 2021

GPU서버의 대표 주자

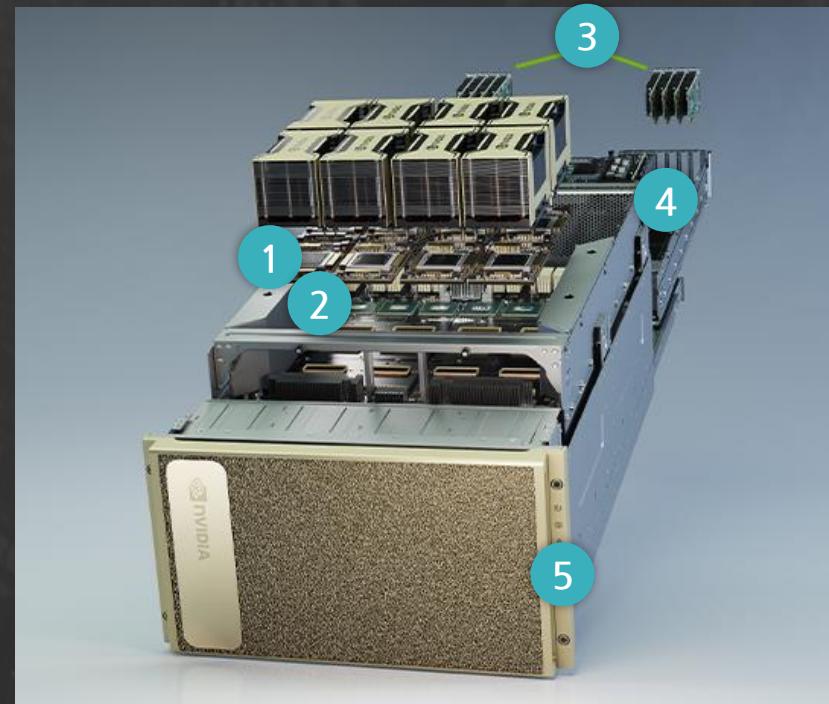
NVIDIA DGX A100

1 총 640GB의 GPU 메모리를 탑재한 NVIDIA A100 GPU 8개
GPU당 NVLink 12개, GPU 간 양방향 대역폭 600GB/s

2 NVIDIA NVSwitch 6개
양방향 대역폭 4.8TB/s, 이전 세대 NVSWITCH보다 2배 증가

3 NVIDIA CONNECTX-7 200GB/s 네트워크 인터페이스 10개
양방향 대역폭 최대 500GB/s

4 30 TB GEN4 NVME SSD
최대 500GB/s의 대역폭 Gen3 NVME SSD보다 2배 빠른 속도



AI 인프라의 최적 표준

NVIDIA DGX H100

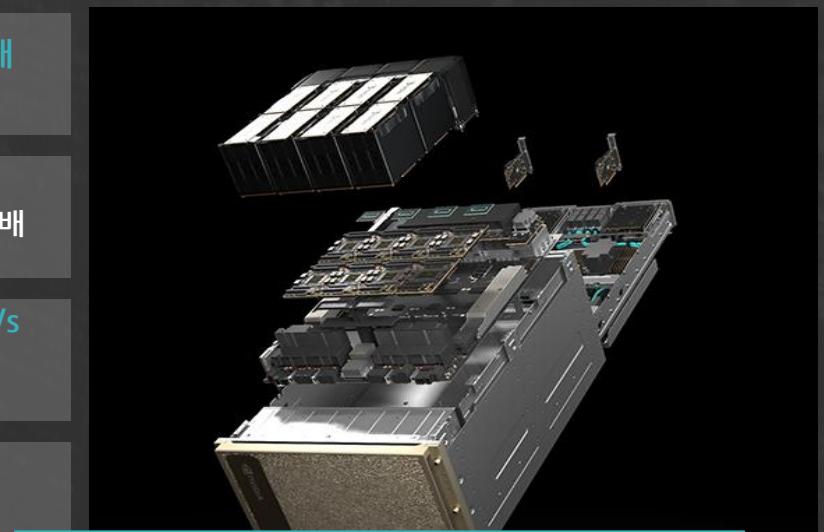
1 최대 640GB의 총 GPU 메모리를 탑재한 NVIDIA H100 GPU 8개
GPU당 NVLink 12개, 900GB/s의 GPU 간 양방향 대역폭

2 NVIDIA NVSwitch 4개
초당 7.2 테라바이트의 양방향 GPU 간 대역폭으로 이전 세대 대비 1.5배
향상

3 NVIDIA CONNECTX-7 8개 및 NVIDIA BLUFIELD DPU 400Gb/s
네트워크 인터페이스 2개
1TB/s의 최대 양방향 네트워크 대역폭

4 듀얼 x86 CPU 및 2TB 시스템 메모리
초고도 AO 작업을 위한 강력한 CPU

5 30 TB NVME SSD
최고의 성능을 위한 고속 스토리지



NVIDIA H100	FP8	4,000 TFLOPS	6X
	FP16	2,000 TFLOPS	3X
	TF32	1,000 TFLOPS	3X
	FP64/FP32	60 TFLOPS	3X

우리에게 적합한 GPU 서버는?

?



HIS에서 고민을 해결해 드립니다.

- 다양한 라인업을 갖춘 슈퍼마이크로를 통해 최적의 솔루션을 제공해 드리겠습니다



슈퍼마이크로 GPU Systems

- Intel & AMD CPU 기반의 다양한 NVIDIA GPU Servers를 보유하고 있으며
- Intel Gaudi 와 AMD MI250 에서 출시할 GPU를 지원하는 시스템도 곧 출시 예정입니다.

Universal GPU Systems (6)

4U GPU Lines (2)

2U GPU Lines (1)

1U GPU Lines (1)

4U GPU with NVLink (5)

2U GPU with NVLink (4)

2U 2-Node Multi-GPU (2)

GPU Workstation (1)

슈퍼마이크로 GPU Systems

- 4개의 주력 모델 선정

GPU with NVLink



AS -4124GO-NART+

AS -2124GQ-NART+

8 x A100 GPU
2 x AMD CPU
Max 8TB Memory
6 x 2.5" Drive bays
8 x PCIe 4.0 x16 LP
AIOM support
4 x 3000W PSU

4 x A100 GPU
2 x AMD CPU
Max 8TB Memory
4 x 2.5" Drive bays
4 x PCIe 4.0 x16 LP
1 x PCIe 4.0 x8 LP
Dual 1GbE NIC
2 x 3000W PSU

PCIe GPU



AS -4124GS-TNR

SYS-220GP-TNR

8(10) x GPU
2 x AMD CPU
Max 8TB Memory
24 x 2.5" Drive bays
Dual 1GbE NIC
AIOM support
4 x 2000W PSU

6 x GPU
2 x Intel CPU
Max 4TB Memory
10 x 2.5" Drive bays
2 x PCIe 4.0 x8 LP
AIOM support
2 x 2600W PSU

SYS-220GP-TNR

- Scientific Virtualization
- High Performance Computing(HPC)
- VDI
- AI/Deep Learning Training

SYS-220GP-TNR



① Processor Support

- Dual 3rd Gen Intel® Xeon® Scalable Processors upto 270W TDP

② Memory Capacity

- 16 DIMM slots, up to 4TB DDR4 memory 3200 MHz DIMMs

③ GPU

- 6 PCIe GPUs Double Width FHFL

④ PCI-E Expansion Slots

- 6 PCIe 4.0 x16 FHFL, 2 PCIe 4.0 x8 LP, AIOM support

⑤ I/O ports

- 1 BMC LAN port ,1 VGA port, 2 USB 3.0 ports

⑥ Drive bays

- 10x 2.5" drive bays, Up to 6 NVMe drives, 2x M.2

⑦ Power Supply

- Two 2600W High-efficiency (Titanium level) power supply

AS -4124GS-TNR

- AI / Deep Learning
- High Performance Computing(HPC)
- Cloud Gaming
- Molecular Dynamics Simulation

AS -4124GS-TNR



① Processor Support

- Dual AMD EPYC™ 7002, 7003 Series Processors

② Memory Capacity

- 32 DIMM slots, up to 8TB DDR4 memory 3200 MHz DIMMs

③ GPU

- Supports up to 10 A100 PCIe GPUs (Default with up to 8 GPUs)

④ PCI-E Expansion Slots

- 9 PCIe 4.0 x16 (Option: 10 PCIe 4.0 x16 slots without NVMe devices)

⑤ I/O ports

- 1 BMC LAN port, 1 VGA port, 2 USB 3.0 ports

⑥ Drive bays

- 24x 2.5" drive bays (Default with 4 SATA and 4 NVMe drives)

⑦ Power Supply

- 4 2000W Redundant Power Supplies Titanium Level (96%)

AS -2124GQ-NART+

- AI /ML, Deep Learning Training and Inference
- High Performance Computing
- Research Laboratory/ National Laboratory
- Autonomous Vehicle Technologies

AS -2124GQ-NART+



① Processor Support

- Dual AMD EPYC™ 7002, 7003 Series Processors

② Memory Capacity

- 32 DIMM slots, up to 8TB DDR4 memory 3200 MHz DIMMs

③ GPU

- Supports 4x A100 80GB SXM4 GPUs with NVLink

④ PCI-E Expansion Slots

- 4 PCIe 4.0 x16 LP, 1 PCIe 4.0 x8 LP

⑤ I/O ports

- Dual RJ45 10GbE LAN, RJ45 1GbE IPMI

⑥ Drive bays

- 4x 2.5" drive bays

⑦ Power Supply

- 2 3000W Redundant Power Supplies Titanium Level (96%+)

AS -4124GQ-NART+

- High Performance Computing(HPC)
- AI / Deep Learning

AS -4124GQ-NART+



① Processor Support

- Dual AMD EPYC™ 7002, 7003 Series Processors

② Memory Capacity

- 32 DIMM slots, up to 8TB DDR4 memory 3200 MHz DIMMs

③ GPU

- Supports 8x A100 80GB SXM4 GPUs with NVLink

④ PCI-E Expansion Slots

- 8 PCIe 4.0 x16 LP, 1 PCIe 4.0 x16 AIOM

⑤ I/O ports

- RJ45 1GbE for IPMI, AIOM for selectable network options

⑥ Drive bays

- 6x 2.5" drive bays

⑦ Power Supply

- 4 3000W Redundant Power Supplies Titanium Level (96%+)

서버는 결정! 그 다음은?

HIS 통합 AI 플랫폼



자문/컨설팅



계획/설계



구축&수행

통합 AI 플랫폼

인프라 최적화 (GPUDirect Storage, GPU가상화)

AI 운영시스템 (컨테이너: Lablup Backend.AI, 가상머신: VMware)

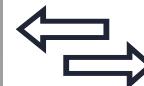
AI 인프라



연산 자원
(NVIDIA DGX/
Supermicro HGX)



저장 자원
(초고성능 병렬 파일
스토리지)



네트워크
(Cisco&Mellanox)



AI의 시작과 끝 효성이 함께 합니다.

-효성인포메이션시스템-

