



인공지능 가속화를 위한 딥러닝 솔루션
IBM Power Systems AC922



고객 상담 전문 센터

02-3781-7900

딥러닝의 신기술, AI의 신세계를 열다

AI를 제대로 활용하기 위해서는 AI기술을 뒷받침하는 강력한 하드웨어가 필수입니다.
IBM Power Systems는 AI로의 빠른 여정을 돕기 위한 새로운 딥러닝 솔루션으로,
데이터 과학자가 원하는 최첨단 AI기술과 IT에 필수적인 안정성을 제공합니다.

IBM Power Systems AC922



가속화된 인터커넥트

"CPU-Accelerator" 시대,
최첨단 I/O로 GPU 컴퓨팅의 잠재력 실현



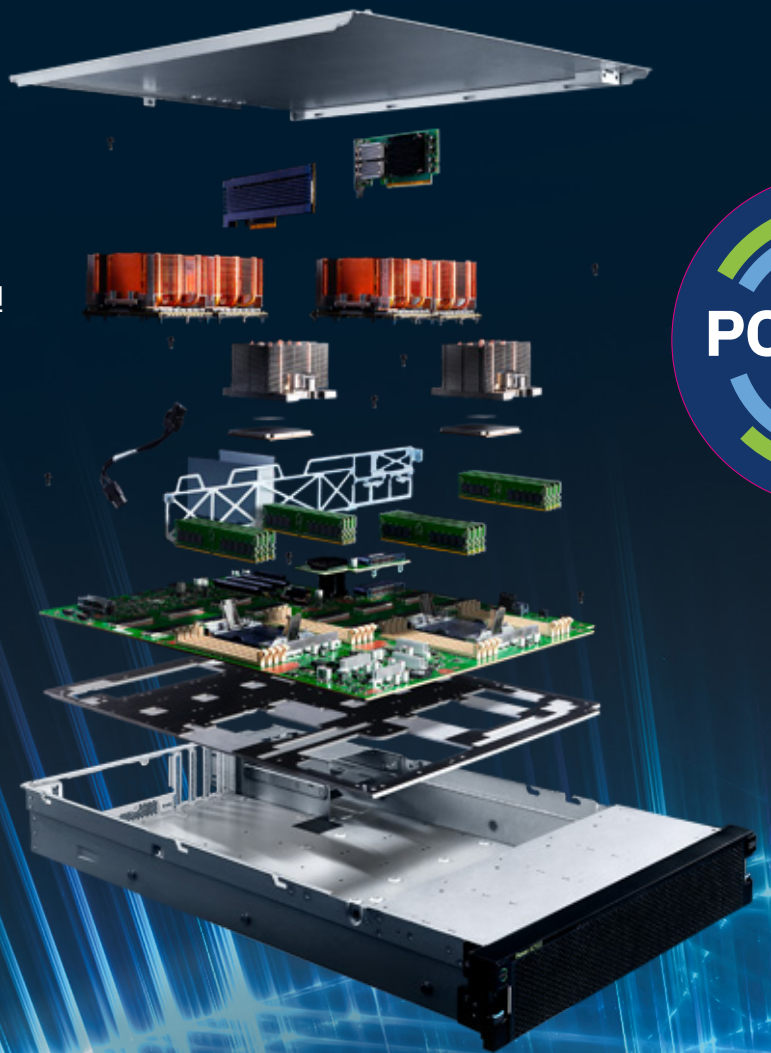
AI 시대에 적합한 설계

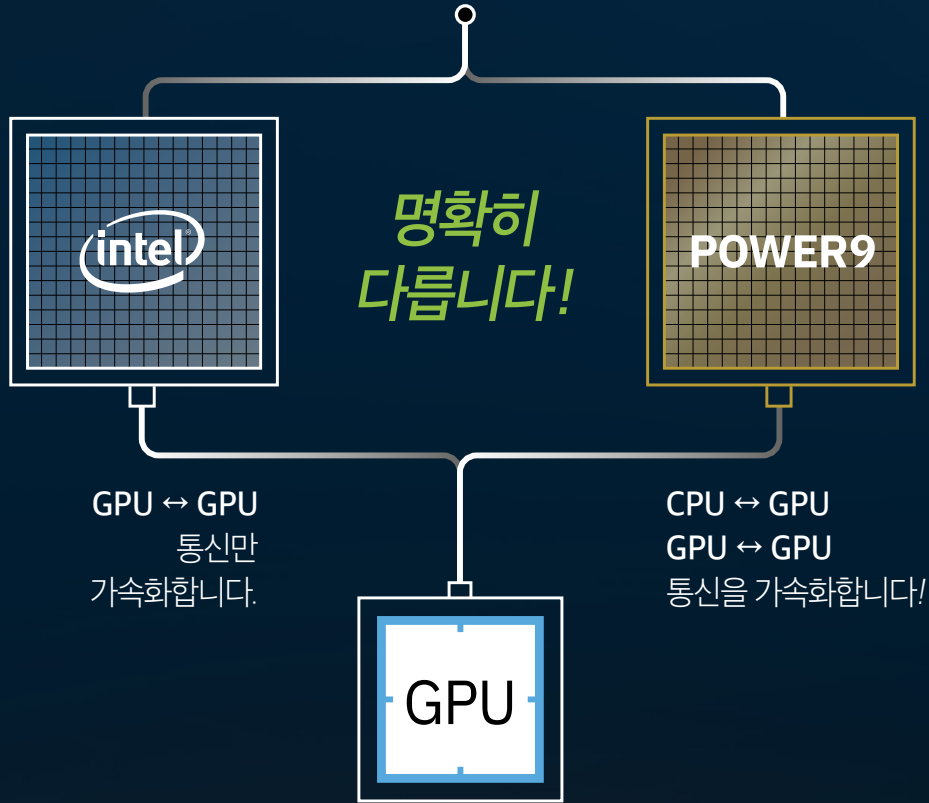
통찰력을 제공하는 최신 분석 및
AI 워크로드를 위한 설계



엔터프라이즈급 AI 제공 (딥러닝 솔루션)

인공신경망을 구축, 훈련, 추론하는
AI 여정을 가속화하여 빠른 비즈니스
가치 실현





뛰어난 슈퍼컴퓨팅 및 AI 리더를 위해 설계된

IBM Power Systems Accelerated Compute(AC922)

x86에서는 지원되지 않았던 기능을 제공하는 새로운 P9 프로세서에
첨단 IO 인터페이스가 통합되어 탁월한 성능과 애플리케이션 이점을 제공합니다.

Power Systems AC922(8335-GTC, 8335-GTW) 개요		표준 요소	
시스템 구성		I/O 포트	USB 3.0 2개, 1GB Eth, VGA 2개
마이크로프로세서	POWER9 NVLink CPU 2개, 16코어, 20코어 또는 18코어, 22코어(액램식)	연결 지원(선택 사항)	
Level 2(L2) 캐시	512K	POWER Hypervisor™	KVM
Level 3(L3) 캐시	10MB	Advanced POWER Virtualization(선택 사항)	
RAM(메모리)	최대 2TB, DDR4 RDIMM 소켓 16개	RAS 기능	프로세서 명령 재시도 / 선택형 동적 펌웨어 업데이트 / Chip kill 메모리 / ECC L2 캐시, L3 캐시 / 오류 모니터링이 있는 서비스 프로세서 / 핫스왑 가능 디스크 베이 / 이중 냉각 팬
내부 디스크 스토리지	SFF(2.5") 드라이브 베이 2개, NVMe SSD 지원 옵션 (PCIe 슬롯)	운영 체제	Red Hat Enterprise Linux, Ubuntu Linux
프로세서-메모리 대역폭	소켓당 170GB/s, 시스템당 340GB/s	전원 요구 사항	200V ~ 240V
L2-L3 캐시 대역폭	7TB/s 온칩 대역폭	시스템 크기	너비 : 441.5mm(17.4인치) / 깊이 : 822mm(32.4인치) / 높이 : 86mm(3.4인치) / 무게 : 30kg(65lbs)
내부 SCSI 디스크 베이	해당 없음	보증	- 3년 제한적 보증, 기타 유닛은 CRU(Customer Replaceable Unit)(국가별로 다름) - 익영업일 9am ~ 5pm(휴일 제외) 보증 서비스 업그레이드 및 유지 보수 이용 가능
미디어 베이	해당 없음		
어댑터 슬롯	SXM 2.0 소켓 4개 또는 6개, NVIDIA Tesla V100 GPU 액셀러레이터(NVLink 포함) / PCIe x16 4.0 슬롯 2개 / PCIe16x(x8, x8) 4.0 슬롯 1개(멀티 소켓 호스트, 다이렉트 지원) / PCIe x4 4.0 슬롯 1개		



가속화된 인터커넥트

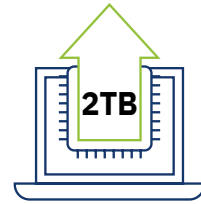
속도와 기술이 선사하는 놀라운 가속화의 힘



**PCIe Gen3 대비
데이터 처리량 5.6배**

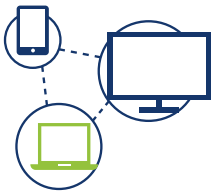
NVIDIA NVLink
최적화 지원

**가속화된
인터커넥트
Accelerated
Interconnect**



**최대 2TB의
시스템 메모리에 액세스**

오직 POWER에서만
가능한 일관성 지원



**여러 대의 디바이스로
뛰어난 데이터 전송**

OpenCAPI GPU 디바이스로
25G 링크



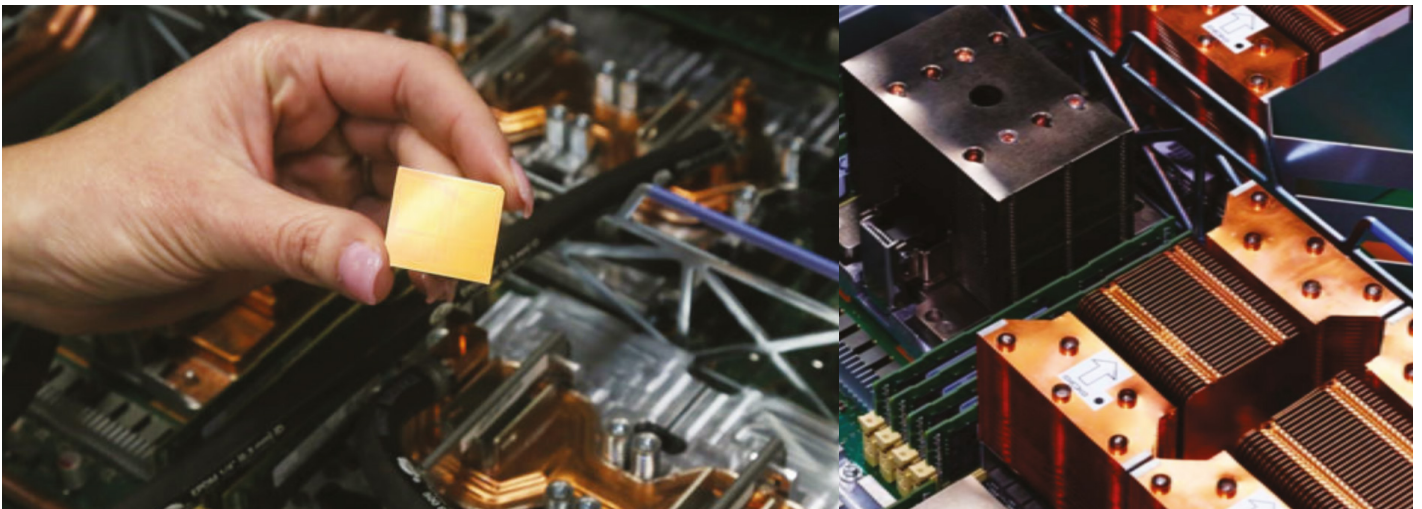
**PCIe Gen3 대비
대역폭 2배**

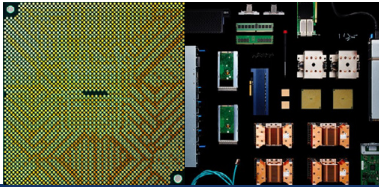
PCIe Gen4 지원



**CPU ↔ GPU 및
GPU ↔ GPU 가속화**

GPU ↔ GPU 이상의
통신 가속화





AI 시대에 적합한 설계

AI에 최적화된 가장 빠르고 유연한 시스템

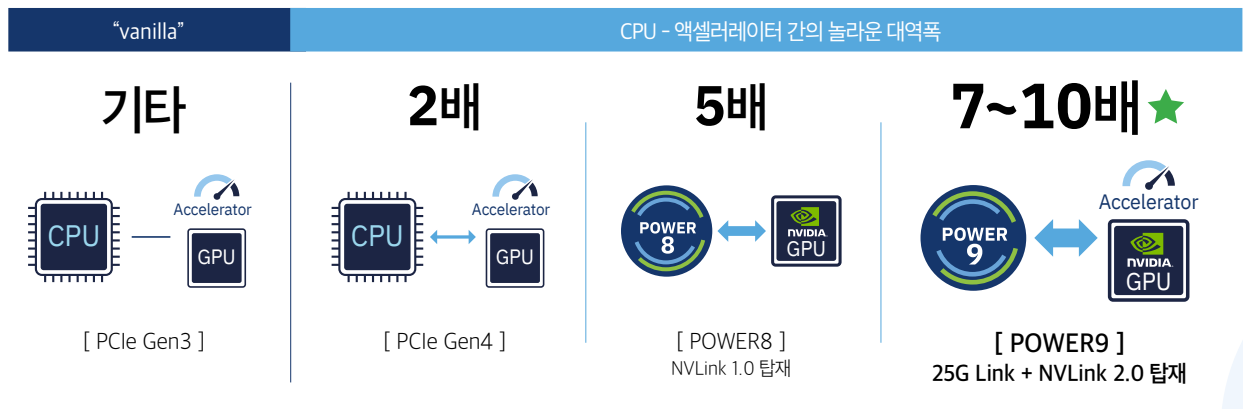
| CPU 및 이기종 컴퓨팅의 효율적인 적용 |

원활한 CPU 및 액셀러레이터 상호 작용

일관된 메모리 공유
향상된 가상 주소 변환

이기종 컴퓨팅의 광범위한 적용

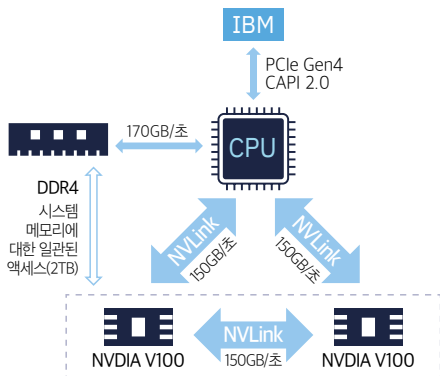
효율적인 프로그래밍 모델을 위한 설계
복잡한 AI 및 분석 앱을 가속화



| CPU ↔ GPU 대역폭 예시 |

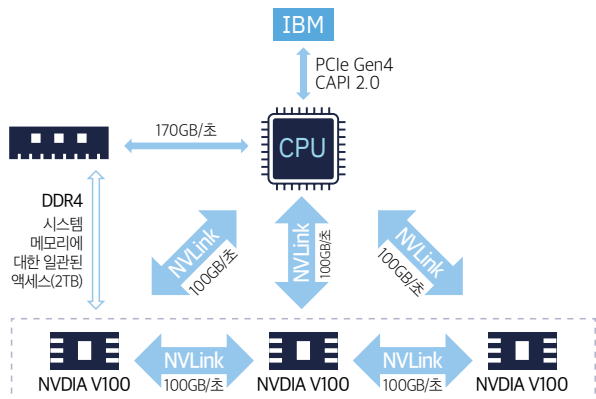
4 GPU @150GB/초

시스템 메모리에 대한 일관된 액세스
PCIe Gen4와 CAPI 2.0을 통해 InfiniBand로 연결
공기 냉각 및 수냉각 옵션



6 GPU @100GB/초

시스템 메모리에 대한 일관된 액세스
PCIe Gen4와 CAPI 2.0을 통해 InfiniBand로 연결
수냉각 옵션만 제공



컴퓨팅 시스템에서 코그너티브 시스템으로 진화

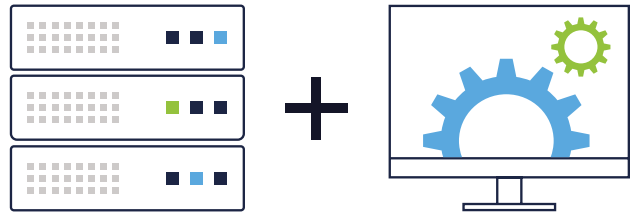
개발 에코시스템

- 업계 지원
- 파트너십
- 개방형 프레임워크
- IBM 소프트웨어



- 개방형 액셀러레이터 인터페이스
- 액셀러레이터 로드맵

하드웨어 설계뿐만 아니라
**유연한 에코시스템과
최적화된 지원이 필수입니다.**



[하드웨어]

[소프트웨어]

**머신러닝, 딥러닝 및 인공지능에서
효율적으로 작동합니다!**

POWER9

POWER9은 AI에 최적화된 시스템으로
뛰어난 학습 속도와 정보 수집 능력을 보여 줍니다.

Caffe

GoogLeNet - 1000 epochs
숫자가 작을수록 좋습니다

POWER9 NVIDIA 4xTesla V100 GP NVLink 2.0 IBM PowerAI
= 2,940 초

IBM 3.7배 더 빠름!

intel NVIDIA 4xTesla V100 GPU PCIe3 = 11,215 초



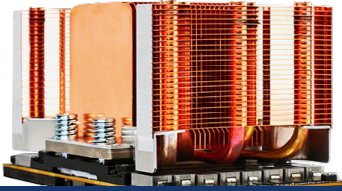
TensorFlow

GoogLeNet - 1000 epochs
숫자가 클수록 좋습니다

IBM POWER9 NVIDIA 4xTesla V100 GP NVLink 2.0
= 4,763 이미지/초

IBM 2.3배 더 빠름!

intel NVIDIA 4xTesla V100 GPU PCIe3 = 2,042 이미지/초



엔터프라이즈급 AI 제공(딥러닝 솔루션)

쉽고 빠른 딥러닝 연구를 위한 IBM 딥러닝 솔루션



코딩 등의 전문 지식 없이도 가능한
모두를 위한 AI 훈련!

IBM Spectrum Deep Learning Impact (DLI)

모델 최적화 및 GPU서버 효율성
극대화를 위한 AI 개발 플랫폼!

딥러닝 프레임워크	Caffe-BVLC	Caffe-IBM	Tensorflow	PyTorch	SnapML	etc...
라이브러리 지원	OpenBLAS	Bazel	OpenCV	HDF5	Leveldb	etc...
소프트웨어 인프라	NVIDIA CUDA	NVIDIA NCCL	NVIDIA cuDNN	Anaconda		
확장성과 안정성을 위한 GPU 서버 인프라스트럭처	NVLink 서버 클러스터 	Spectrum Scale : 고속 병렬 파일시스템 	Scale to Cloud(optional) 			

최적화된 툴 사용으로 쉽고 빠르게 적용된 딥러닝 솔루션 적용 사례

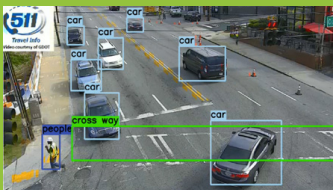


“Point-and-Click” 간단하지만 강력한 이미지 & 비디오 분석 AI Tool

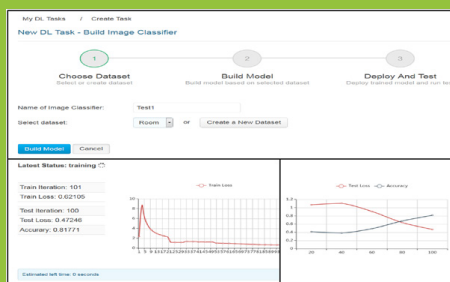
End-to-End 딥러닝을 지원하는 GUI 기반 Toolkit!

Dataset 관리와 전처리, 훈련 과정의 모니터링, 분석, 최적화 + Inference를 위한 모델 배포까지 관리

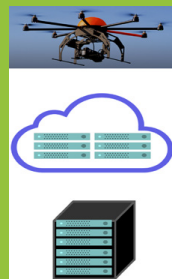
- 데이터 라벨링 (Image, Video, Semi-auto)



- AI 모델 자동 훈련 및 관리 (Dataset, Label, Training, Model 관리)



- AI 모델 배포 (Inference API)



IBM Spectrum Deep Learning Impact (DLI)

인공지능 모델 개발 및 운용을 위한 분산 딥러닝 플랫폼

사용자에게는!

▷ 훈련 단계

- 훈련 정확성 향상
- 모델 디자인 이슈 확인 및 조기에 중단 제안(효율성 향상)
- 동시 훈련을 통한 모델 개발 주기 단축

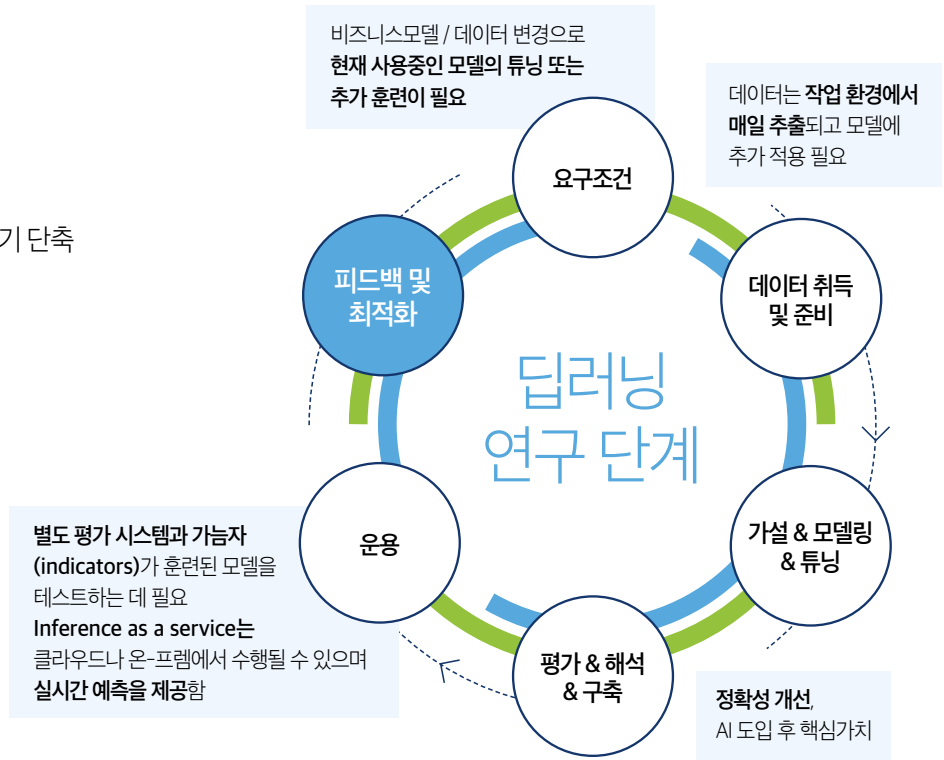
▷ 테스트 단계 (인퍼런스 단계)

- 실시간 온라인 서비스 지원

IT 담당자에게는!

▷ 단일화된 AI 개발 플랫폼

▷ 하드웨어 클러스터 활용 극대화



전문가와 함께하는 찾아가는 세미나 안내

딥러닝의 인프라부터 소프트웨어까지, 고객의 환경에 맞춰 지원해 드리는 IBM 딥러닝 솔루션!

IBM 딥러닝 솔루션에 대해 보다 자세한 설명을 원하세요?

IBM 전문가가 직접 고객사를 방문하여 맞춤 세미나를 제공해 드립니다.

맛있는 도시락이 함께 제공되는 찾아가는 세미나를 신청해 보세요.

문의 | 한국IBM 마케팅 총괄 본부 [✉mktg@kr.ibm.com](mailto:mktg@kr.ibm.com)



© Copyright IBM Corporation 2018 한국아이비엠주식회사
(07326) 서울시 영등포구 국제금융로10 서울국제금융센터 (Three IFC)
TEL : (02)3781-7800 www.ibm.com/kr

2018년 7월 Printed in Korea All Rights Reserved