

# IBM Power Advanced Compute(AC) AC922 Server



엔터프라이즈 AI 환경을 위한 최상의 서버

## 주요 특징

IBM® Power Systems™ Accelerated Compute(AC922) Server는 엔터프라이즈급 AI 환경을 위한 최강의 가속화 솔루션입니다.

- CPU에만 의존하던 시대를 넘어서 더 강력한 가속 컴퓨팅 가능성을 실현하는 초고속 솔루션
- AI 시대에 걸맞게 인사이트를 발굴하는 첨단 분석 및 AI 워크로드를 위한 아키텍처
- 엔터프라이즈급 AI, 즉 데이터 과학자들이 기대하는 첨단 AI 혁신과 IT 팀이 필요로 하는 안정성의 조합
- 소규모 클러스터뿐 아니라 세계 최대 규모의 컴퓨터 환경에서도 구축되어 검증된 성능 및 선형 확장성
- 가장 간소화된 GPU 가속3 —가속화 방식이 아니라 가속화 대상에 초점, 일관성 보장, 더 큰 규모의 모델뿐 아니라 AC922에서도 사용 가능

첨단 AI, HPC, 분석 워크로드가 활성화되면서 데이터 집중적인 작업이 더욱 확대되고 있으며 이는 가속 인프라를 통해서만 해결할 수 있습니다.

이러한 요구 사항을 염두에 두고 IBM Power Systems에서 AI 시대를 위한 AC922, 즉 엔터프라이즈 AI를 위한 최상의 서버를 설계했습니다. AC922는 IBM의 최신 POWER9™ 프로세서와 각종 첨단 연결 기능을 활용하면서 기존 x86의 PCIe Gen 3 버스 대비 최대 5.6배1의 데이터 이동 속도를 실현합니다. IBM Power Systems는 CPU와 GPU 간의 NVLink를 활성화하는 유일한 아키텍처를 기반으로 하여 가속 컴퓨팅의 새로운 가능성을 제시합니다.



Power Systems AC922 내부 구성 요소

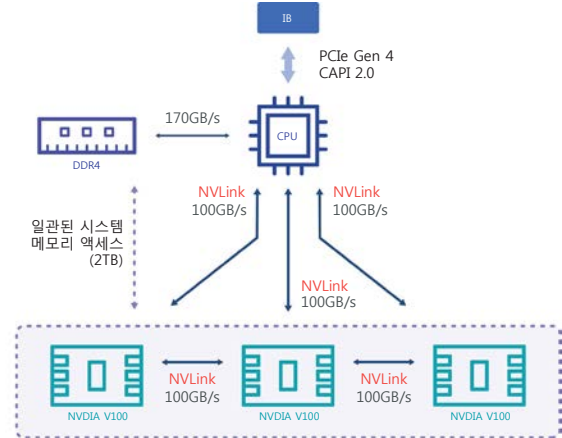


## IBM Power Systems Accelerated Compute(AC922) Server

IBM Power Systems Accelerated Compute(AC922) Server는 엔터프라이즈급 AI 환경을 위한 최강의 가속화 솔루션입니다.

POWER9 기술이 구현된 차세대 IBM Power Systems는 AI 시대에 걸맞은 획기적인 속도를 실현하는 혁신적인 솔루션입니다. 고성능 컴퓨팅을 위한 IBM Power System AC922를 자세히 살펴보겠습니다.

- **더 빠른 I/O** - AC922는 PCIe gen4, CAPI 2.0, OpenCAPI, NVLINK 등 다양한 차세대 I/O 아키텍처를 지원합니다. 이러한 인터커넥트를 통해 기존 x86 서버의 PCIe gen3 대비 최대 5.6배<sup>1</sup>의 대역폭을 제공하면서 현재의 데이터 집중적인 워크로드를 효과적으로 처리할 수 있습니다.
- **최강의 CPU** - POWER9 CPU는 엄청난 속도와 함께 그 주변에 있는 모든 구성 요소의 성능을 십분 활용할 수 있다는 점에서도 뛰어납니다. AI 시대를 위해 설계된 POWER9은 x86 플랫폼 대비 최대 5.6배<sup>1</sup>의 I/O 및 2배의 스레드를 처리할 수 있습니다. POWER9은 16코어, 18코어, 20코어, 22코어 구성으로 제공되며 AC922 서버에서 최대 44코어를 지원합니다.
- **첨단 GPU** - AC922는 엔터프라이즈 AI를 위한 최상의 GPU와 최상의 플랫폼이 만난 솔루션이라 할 수 있습니다. AC922는 POWER9 CPU와 NVIDIA Tesla V100(NVLink GPU)을 페어링합니다. 각 페어링으로 최대 5.6배<sup>1</sup>의 성능을 실현할 수 있습니다. CPU와 GPU 간에 이러한 성능을 제공할 수 있는 유일한 서버입니다. HPC, 딥러닝, AI 워크로드를 위해 광대한 처리 성능을 제공합니다.
- **첫 PCIe 4세대 서버** - AC922는 차세대 산업 표준 PCIe 인터커넥트 기술을 업계 최초로 갖춘 서버입니다. PCIe 4세대는 x86 서버의 PCIe 3세대 인터커넥트 대비 약 2배의 데이터 대역폭<sup>2</sup>을 제공합니다.

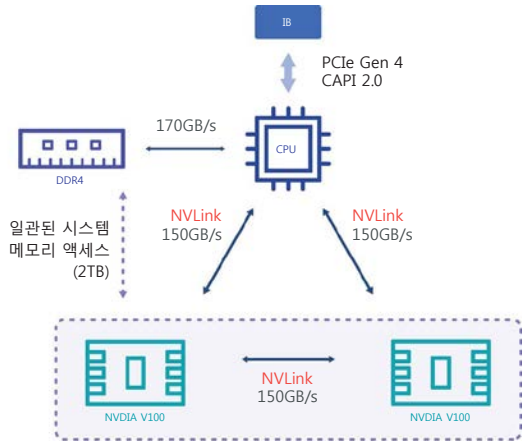


GPU 6개를 갖춘 수랭식 Power Systems

- **가장 간소화된 AI 아키텍처<sup>3</sup>** - AI 모델이 커지면서 x86 서버의 GPU 메모리 용량으로 감당할 수 없는 수준에 이르곤 합니다. AC922의 CPU-GPU 일관성 덕분에 가속 애플리케이션에서 시스템 메모리를 GPU 메모리로 활용하는 것이 가능합니다. 또한 데이터 이동 및 로컬 요건이 사라져 프로그래밍이 간소화됩니다. 게다가 5.6배<sup>1</sup> 더 빠른 NVLink 인터커넥트를 활용하므로 CPU와 GPU 간 메모리 공유 때문에 x86 서버의 PCIe 3 수준으로 속도가 저하되지 않습니다.
- **엔터프라이즈급 솔루션** - 간단하게 딥러닝을 구축하고 우수한 성능을 실현합니다. AI 사용자를 위해 새롭고 더 간편한 엔드 투 엔드 툴체인을 제공합니다. 검증된 API 성능 및 확장성 덕분에 단일 노드에서 시작했다가 단일 랙 또는 노드 수천 개 규모로 확장하면서 선형 확장의 효율성을 누릴 수 있습니다.
- **전 세계의 최대 AI 과제를 해결하도록 설계된 안성맞춤의 솔루션** - AC922는 CORAL Summit 슈퍼컴퓨터의 중추 역할을 하면서 HPC에서 200페타플롭 이상, AI에서 3엑사플롭의 서비스 성능 이정표를 달성합니다. 이와 함께 AI 구축의 효율성 및 편의성도 지원하므로 귀사의 AI 요구 사항을 충족할 수 있습니다.

## Systems

### 데이터시트



GPU 4개를 갖춘 공랭식 Power Systems

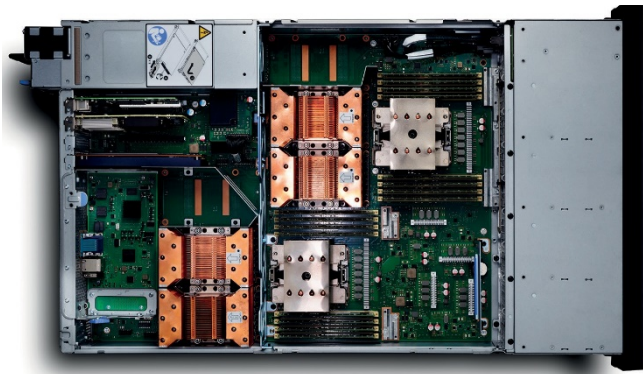
AI, 딥러닝, HPC, 고성능 분석의 선두 주자를 위해 설계된 IBM Power Systems AC922를 자세히 살펴보겠습니다.

- POWER9 & NVLink 2.0 CPU 2개, DIMM 소켓 16개, 최대 2TB 메모리
- 차별화된 GPU 가속 플랫폼
  - POWER9 & NVLink 2.0 기술: 최대 5.6배의 성능을 제공하는 링크(100GB/s 공랭식 또는 150GB/s 수랭식)로 각 NVIDIA V100(NVLink GPU포함)에 연결
  - 놀라운 CPU-GPU 통신 및 GPU-GPU 통신: PCI-E Gen3 x16 솔루션의 동일 소켓에서 인접 NVIDIA® Tesla® V100 GPU 액셀러레이터 간에 최대 5.6배의 데이터 플로우(100GB/s 또는 150GB/s) 실현
- 쉬운 GPU 프로그래밍 - 완전한 일관성 및 시스템 메모리 액세스
- 첨단 Mellanox ConnectX-5 InfiniBand Fabric, 업계 최고의 PCIe Gen4 인터페이스
- CAPI 연결 NVMe 스토리지 옵션으로 초고속 스토리지 I/O 및 버스트 버퍼 데이터 스테이징 지원
- 모든 액셀러레이터를 위한 최신 인터페이스: NVLink, OpenCAPI™, CAPI 2.0, PCIe Gen4

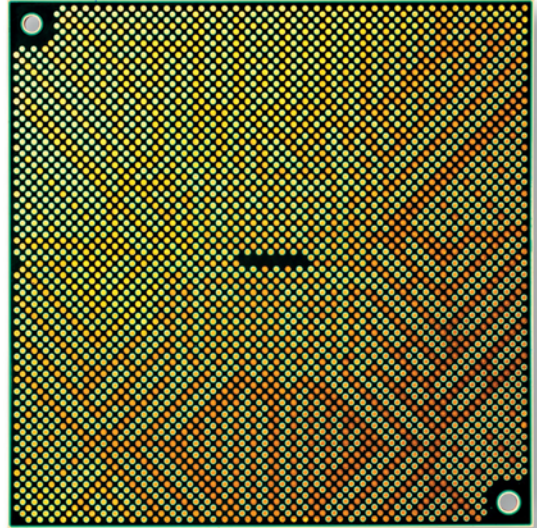
## Systems

데이터시트

검증된 API 성능 및 확장성 덕분에  
단일 노드에서 시작했다가 단일 랙  
이상의 규모로 확장하면서 선형 확장의  
효율성을 누릴 수 있습니다.



Power Systems AC922 내부 모습



Power Systems AC922 POWER9 CPU



Power Systems AC922 전면 모습

## Systems

### 데이터시트

#### Power Systems AC922(8335-GTC, 8335-GTW) 개요

##### 시스템 구성

마이크로프로세서	POWER9 NVLink CPU 2개, 16코어, 20코어 또는 18코어, 22코어(액랭식)
Level 2(L2) 캐시	512K
Level 3(L3) 캐시	10MB
RAM(메모리)	최대 2TB, DDR4 RDIMM 소켓 16개
내부 디스크 스토리지	SFF(2.5") 드라이브 베이 2개, NVMe SSD 지원 옵션(PCIe 슬롯)
프로세서-메모리 대역폭	소켓당 170GB/s, 시스템당 340GB/s
L2-L3 캐시 대역폭	7TB/s 온칩 대역폭
내부 SCSI 디스크 베이	해당 없음
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

##### 표준 요소

I/O 포트	USB 3.0 2개, 1GB Eth, VGA 2개
연결 지원(선택 사항)	
POWER Hypervisor™	KVM
Advanced POWER Virtualization1(선택 사항)	
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
운영 체제	Red Hat Enterprise Linux, Ubuntu Linux
전원 요구 사항	200V ~ 240V
시스템 크기	너비: 441.5mm(17.4인치) 깊이: 822mm(32.4인치) 높이: 86mm(3.4인치) 무게: 30kg(65lbs.)
보증	3년 제한적 보증, 기타 유닛은 CRU(customer replaceable unit)(국가별로 다름) 익명업일 9am ~ 5pm(휴일 제외). 보증 서비스 업그레이드 및 유지 보수 이용 가능.

## IBM을 선택하는 이유

IBM은 AI 분야를 개척해왔습니다. IBM 연구소의 초기 머신러닝 시스템부터 Jeopardy에 출연한 Watson®까지 IBM에게 AI는 단순한 유행어가 아닙니다. 또한 IBM은 이 혁신을 코그너티브 인프라에 접목하여 고객의 AI 여정을 적극 지원합니다.

IBM은 신뢰할 수 있는 엔터프라이즈 환경에서 최첨단 혁신을 선보입니다. IBM은 105년 넘게 끊임없는 혁신으로 고객의 비즈니스 요구 사항을 해결해왔습니다.

IBM은 AI 시대에 귀사의 동반자가 되겠습니다. IBM은 AI의 가능성을 탐구하기 위해 몇 개의 노드를 구축하거나 엑사플롭 규모의 AI를 서비스 형태로 구축하려는 기업 모두를 위해 AI 여정 전반에 필요한 모든 기술 및 서비스를 가장 유연하고 포괄적인 구성으로 제공합니다.

## 추가 정보

Power Systems AC922에 대한 자세한 내용은 IBM 담당자 또는 IBM 비즈니스 파트너에게 문의하거나 다음 웹 사이트에서 확인하십시오.

[ibm.com/marketplace/ai-hpc-server-power-ac922](http://ibm.com/marketplace/ai-hpc-server-power-ac922)

또한 IBM Global Financing에서는 고객이 비즈니스 성장에 필요한 기술을 부담 없이 도입할 수 있도록 다양한 결제 옵션을 제공합니다. IT 제품 및 서비스의 도입부터 폐기까지 라이프사이클 전반을 관리합니다. 자세한 내용은 다음 사이트에서 확인하십시오. [ibm.com/financing](http://ibm.com/financing)



© Copyright IBM Corporation 2017

IBM Systems  
New Orchard Road  
Armonk, NY 10504

Produced in the United States of America  
December, 2017

IBM, IBM logo, ibm.com, Power Systems, POWER는 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보" ([ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml))에 있습니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

NVIDIA, NVIDIA Volta, NVIDIA NVLink는 미국 또는 기타 국가에서 사용되는 NVIDIA Corporation의 상표입니다.

이 문서는 최초 발행일을 기준으로 하며, 통지 없이 언제든지 변경될 수 있습니다. IBM이 영업하는 모든 국가에서 모든 오퍼링이 제공되는 것은 아닙니다.

인용된 성능 데이터와 고객 예제는 예시 용도로만 제공됩니다. 실제 성능 결과는 특정 구성과 운영 조건에 따라 다를 수 있습니다.

이 문서의 정보는 상품성, 특정 목적에의 적합성에 대한 보증 및 타인의 권리 침해에 대한 보증이나 조건을 포함하여(단, 이에 한하지 않음) 명시적이든 묵시적이든 일체의 보증 없이 "현상대로" 제공됩니다. IBM 제품에 대한 보증은 제품의 준거 계약 조항에 의거하여 제공됩니다.

사용 가능한 실제 스토리지 용량은 압축 및 비압축 데이터 모든 경우에 대해, 언급한 것과 다르거나 더 작을 수 있습니다.

- 5.6배의 I/O 대역폭** – IBM 내부에서 CUDA H2D 대역폭 테스트 하드웨어를 사용하여 테스트하고 측정한 결과입니다. 테스트 하드웨어는 Power AC922, 32코어(16c 칩 2개), POWER9 & NVLink 2.0, 2.25GHz, 1024GB 메모리, Tesla V100 GPU 4개, Ubuntu 16.04. S822LC for HPC, 20코어(10c 칩 2개), POWER8 & NVLink, 2.86GHz, 512GB 메모리, Tesla P100 GPU로 구성되었습니다. 비교 대상 하드웨어는 Xeon E5-2640 v4 2개, 20코어(10c 칩 2개)/40스레드, Intel Xeon E5-2640 v4, 2.4GHz, 1024GB 메모리, Tesla V100 GPU 4개, Ubuntu 16.04로 구성되었습니다.
- PCIe Generation 4는 PCIe gen 3 대비 2배의 데이터 처리량을 지원합니다(31.5 GB/s vs 15.8 GB/s x16).
- 가장 간소화된 AI 아키텍처 - 일관성으로 개발자를 위해 데이터 이동 및 로컬성을 추상화하여 코딩 작업을 간소화합니다.



재활용 하십시오.