IBM Power System AC922

Engineered to be the most powerful training platform

Highlights

- Breakthrough data bottlenecks with high IO bandwidth: 5.6x vs compared x86

- Experience faster iterations and training vs compared x86

- Build and train AI models with flexible deployment options

- Enterprise-class support for your AI journey

- Proven scale from one node to the world's most powerful supercomputers

As AI initiatives shift from backroom experiments to boardroom imperatives, many organizations find their existing traditional infrastructure ill-suited for accelerated and AI workloads. Organizations that do not evaluate their existing IT infrastructure, ensuring it is appropriate for Enterprise AI, risk not fully deploying AI.

Model building and training are the foundation of AI. Engineered to be the most powerful training platform, the IBM Power System AC922 provides the data and compute-intensive infrastructure needed to deliver faster time to insights.

The Power AC922 offers a blend of innovative technologies making it a robust AI platform. IBM POWER9 is the first chip with PCIe Gen4 (2x the bandwidth of PCIe Gen3)[1,] the Power AC922 has PCIe Gen4 and other advanced I/O interconnects including CAPI 2.0, OpenCAPI and NVIDIA® NVLink™. Unlike x86-based servers, on the Power AC922 the NVIDIA® NVLink™ enables CPU to GPU connectivity delivering 5.6x[2] the data throughput for today's data-intensive and AI workloads. The Power AC922 supports up to 6 NVIDIA® Tesla V100 GPUs (16GB or 32GB).

With a co-optimized hardware and software stack, the Power AC922 with Watson Machine Learning Accelerator delivers faster deep learning training iterations, up to 3.7x faster[3] for Caffe. Additionally, the Power AC922 can speed up machine learning iterations by 46x[4] with SnapML. Faster training iterations allows Data Scientists to spend less time in model development and get to production and to result quicker.

Every organization's AI journey is different. Flexible deployment options offer organizations ways to implement the Power AC922 in the way that best matches their infrastructure and business needs, whether on-premises or in the cloud, all with enterprise-class support. As the backbone of the world's most powerful supercomputers[5], the Power AC922 has proven to be scalable from one to thousands of nodes to meet business needs for both large and small organizations.
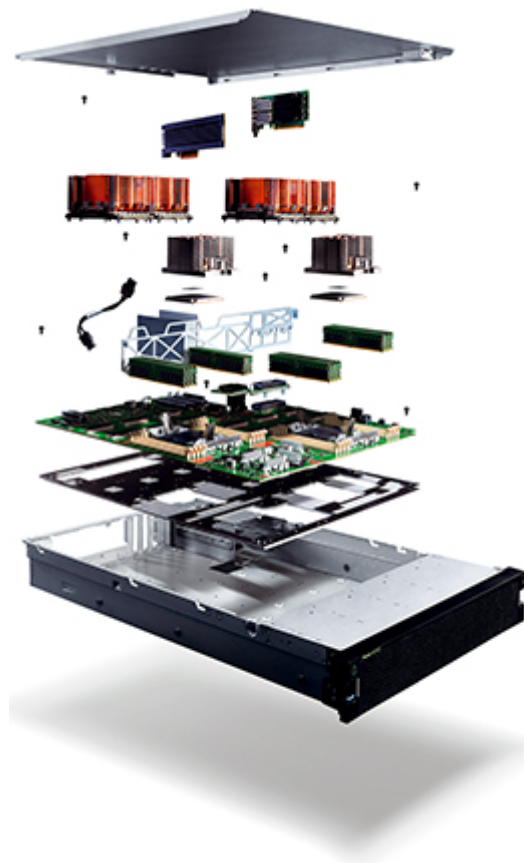
IBM Power System AC922 components

[1] PCIe Generation 4 provides 2x data throughput vs. PCIe gen 3 (31.5 GB/s vs 15.8 GB/s x16), based on publicly available material

[2] 5.6x I/O bandwidth claim based on CUDA H2D Bandwidth Test conducted on a Xeon E5-2640 V4 +P100 vs Power9 + V100 (12 GB/s vs 68 GB/s rated)

[3] Results are based IBM Internal Measurements running 1000 iterations of Enlarged GoogleNet model (mini-batch size=5) on Enlarged Imagenet Dataset (2240x2240). Power AC922; 40 cores (2 x 20c chips), POWER9 with NVLink 2.0; 2.25 GHz, 1024 GB memory, 4xTesla V100 GPU ; Red Hat Enterprise Linux 7.4 for Power Little Endian (POWER9) with CUDA 9.1/ CUDNN 7;. Competitive stack: 2x Xeon E5-2640 v4; 20 cores (2 x 10c chips) / 40 threads; Intel Xeon E5-2640 v4; 2.4 GHz; 1024 GB memory, 4xTesla V100 GPU, Ubuntu 16.04. with CUDA .9.0/ CUDNN 7.

Software: IBM Caffe with LMS Source code https://github.com/ibmsoe/caffe/tree/master-lms

[4] 46x SnapML (https://www.zurich.ibm.com/snapml/) In a newly published benchmark, using an online advertising dataset released by Criteo Labs (http://labs.criteo.com/2013/12/download-terabyte-click-logs/) with over 4 billion training examples, we train a logistic regression classifier in 91.5 seconds. This training time is 46x faster than the best result that has been previously reported (https://cloud.google.com/blog/products/gcp/using-google-cloud-machine-learning-to-predict-clicks-at-scale), which used TensorFlow on Google Cloud Platform to train the same model in 70 minutes.

[5] https://www.top500.org/list/2018/11/

Power System AC922 (8335-GTH, 8335-GTX) at a glance	
System configurations	
Microprocessors	2x POWER9 with NVLink CPUs 16, 20 cores, or 18, 22 cores (with water cooling)
Level 2 (L2) cache	512 K
Level 3 (L3) cache	10 MB
RAM (memory)	Up to 2 TB, from 16 DDR4 RDIMM Sockets
Internal disk storage	2x SFF (2.5") drive bays, optional NVMe SSD support in PCIe slots
Processor-to-memory bandwidth	170 GB/s per socket, 340 GB/s per system
L2 to L3 cache bandwidth	7 TB/s on chip bandwidth
Adapter slots	4 or 6 SXM 2.0 sockets, for NVIDIA Tesla V100 GPU Accelerators with NVLink. 2x PCIe x16 4.0 slots 1x PCIe16x (x8,x8) 4.0 slot (multi-socket host direct supported) 1x PCIe x4 4.0 slot
Standard features	
I/O ports	2x USB 3.0, 2x 1 GB Eth, VGA
POWER Hypervisor	KVM
RAS features	Processor instruction retry Selective dynamic firmware updates Chip kill memory ECC L2 cache, L3 cache Service processor with fault monitoring Hot-swappable disk bays Redundant cooling fans
Operating systems	Red Hat Enterprise Linux, Ubuntu Linux
Power requirements	200 V to 240 V
System dimensions	Width: 441.5 mm (17.4 in.) Depth: 822 mm (32.4 in.) Height: 86 mm (3.4 in.) Weight: 30 kg (65 lbs.)
Warranty	3-year limited warranty, CRU (customer replaceable unit) for all other units (varies by country) next business day 9am to 5pm (excluding holidays), warranty service upgrades and maintenance are available.

IBM®

Why IBM?

For over a century, IBM has pioneered technologies and provided services that help companies manage and mine their valuable business data. And now IBM is helping customers adopt AI. From early machine learning systems in IBM Research, to Watson on Jeopardy, AI isn't just a buzzword. IBM provides a flexible and comprehensive range of technology and services needed for an enterprises entire journey to AI.

For more information

To learn more about the IBM Power System AC922 please contact your IBM representative or IBM Business Partner, or visit the following website:
https://www.ibm.com/us-en/marketplace/power-systems-ac922

Additionally, IBM provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit IBM Global Financing.

IBM Maintenance and Technical Support solutions can help you get the most out of your IT investment by reducing support costs, increasing availability and simplifying management with integrated support for your multiproduct, multivendor hardware and software environment. For more information on hardware maintenance, software support, solution support and managed support, visit:
ibm.com/services/maintenance