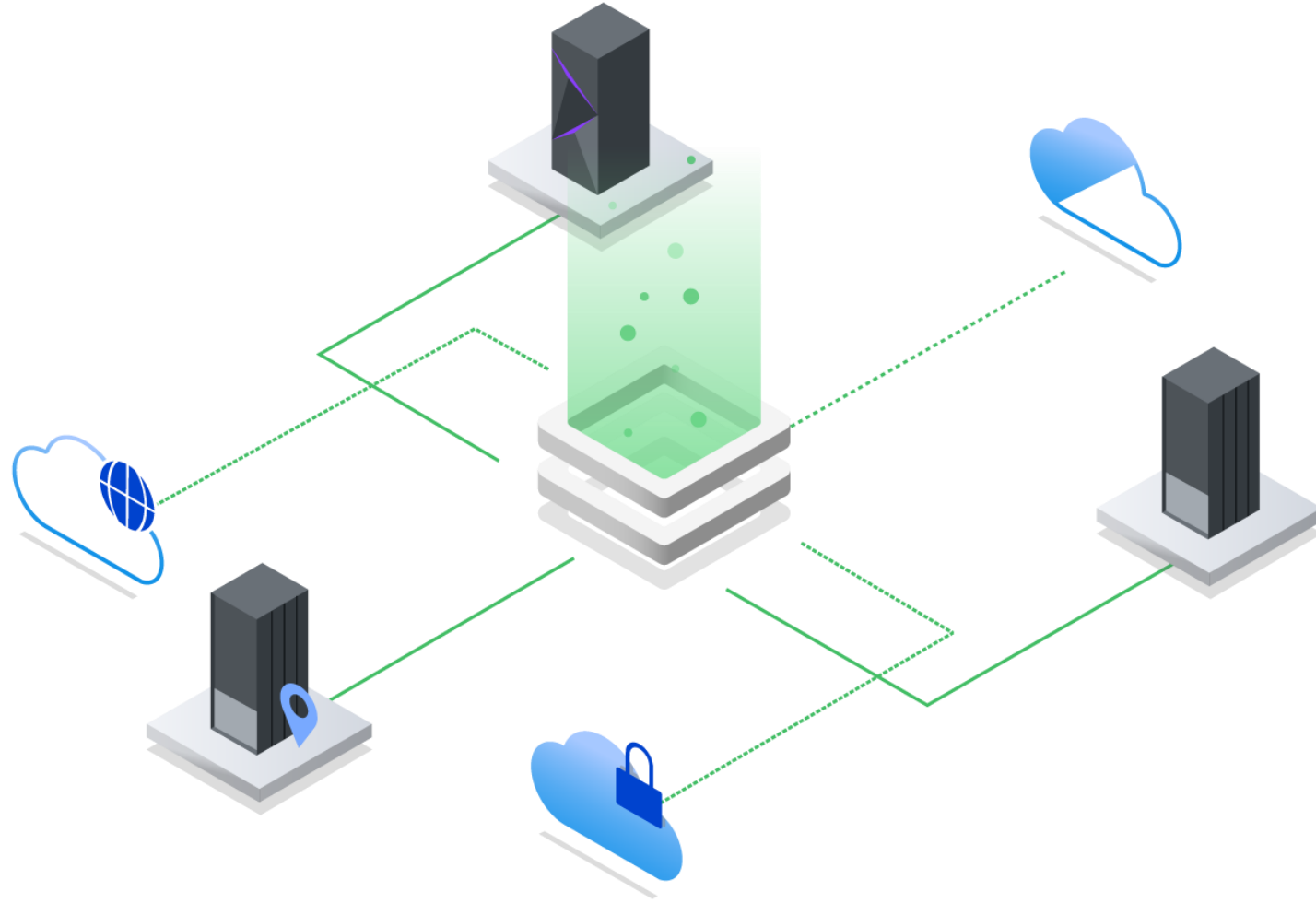


IBM Storage for Data and AI

IBM's family of software defined storage, storage hardware, and storage management software



Matthew Klos
Senior Solutions Architect
Americas SWAT Team



IBM Research Scale Update



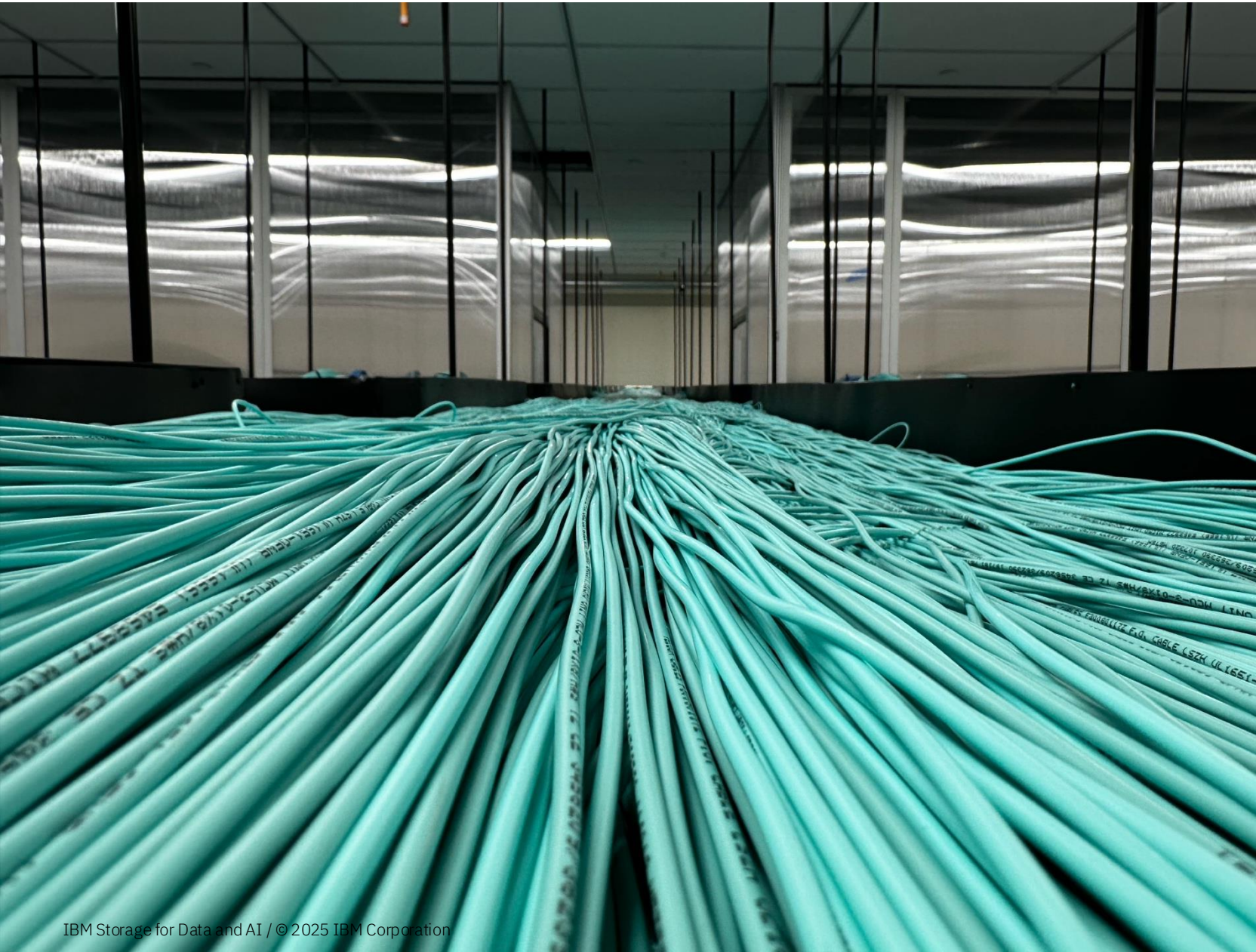
IBM Blue Vela



Blue Vela Compute Pods

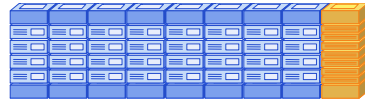


Blue Vela Networking



IBM Blue Vela- HGX “SuperPOD” Storage Fabric (IBM Cloud/ IBM Research/NVIDIA) working together to delivery an AI solution

Scale System 6000s



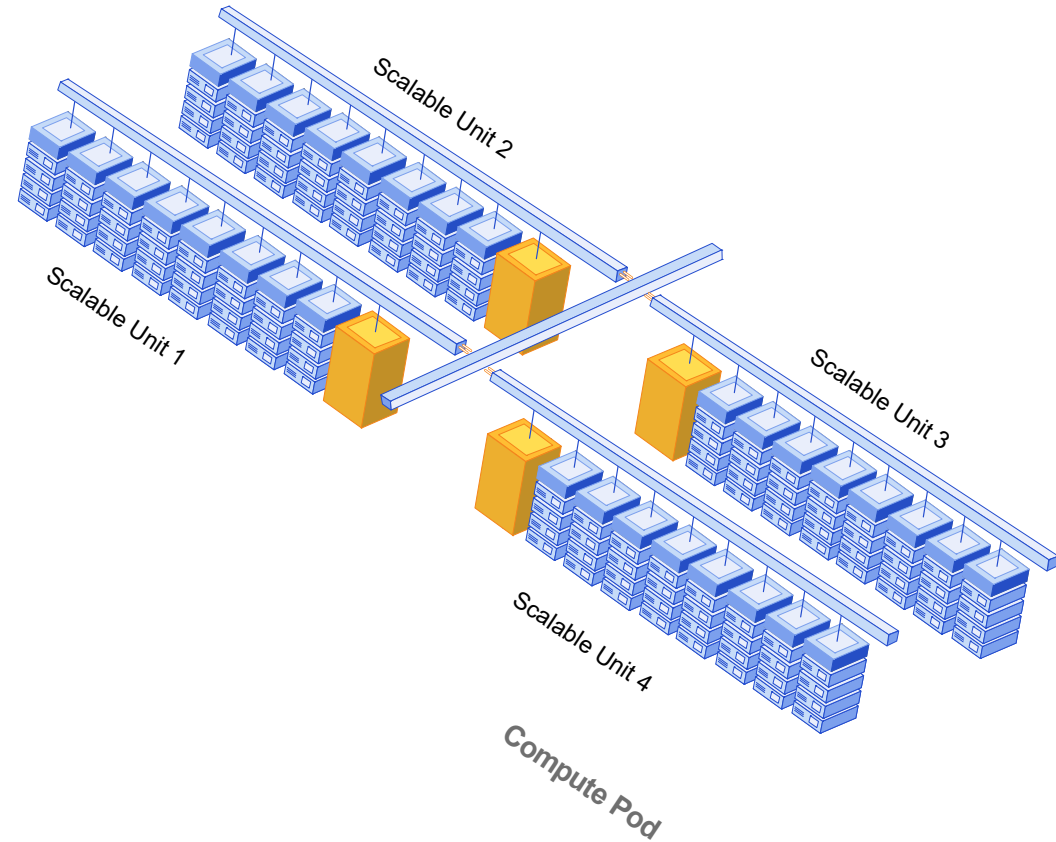
Scalable Unit

- 32 Compute Nodes
- 256 H100 GPUs



Compute Pod

- 4 Scalable Units
- 128 Compute Nodes
- 1024 H100 GPUs
- 82 TB of GPU Ram
- 12,288 Physical Cores
- 256 TB of RAM
- 3481 TB NVME Local Storage



https://www.linkedin.com/posts/dannybarnett_todays-an-incredibly-proud-day-for-my-team-activity-7180654456361910274-4fvU

<https://arxiv.org/pdf/2407.05467>

LLM Data Set Size

Checkpoint time reduced to 1% an hour => 36 seconds

Model Specific Example for Synchronous

Checkpoint

Tensor Model Parallel Size determines how many GPUs participate in the checkpoint.

For example, If Tensor Parallel size is set to 8, 1 out of 8 GPUs will participate in the checkpoint

~14 bytes per Parameter

Example

175B Parameter Model: ~2.4TB Data set size

512B Parameter Model: ~7.2TB Data set size

1T Parameter Model: ~14TB Data set Size

3 x IBM Storage Scale 6000 is 19.4 TB in 36 seconds

Total Number of GPUs is 4000 GPUs

Only 512 GPUs will participate in the checkpoint
(4000_GPUs / 8_Tensor Parallel_Size)

175B: ~4.8GB data set / GPU

512B: ~15GB data set / GPU

1T: ~28GB data set / GPU



Model Load

Using the same Tensor Parallel Size of 8 from the checkpoint

8x the data set size will need to be loaded across all GPUs

Example

175B Parameter Model: ~19TB Data set size

512B Parameter Model: ~58TB Data set size

1T Parameter Model: ~110TB Data set Size

3 x IBM Storage Scale 6000 load in < 2 min

Total Number of GPUs is 4000 GPUs

Data set per GPU is the same, but now all GPUs participate in the Model Load

175B: ~4.8GB data set / GPU

512B: ~15GB data set / GPU

1T: ~28GB data set / GPU

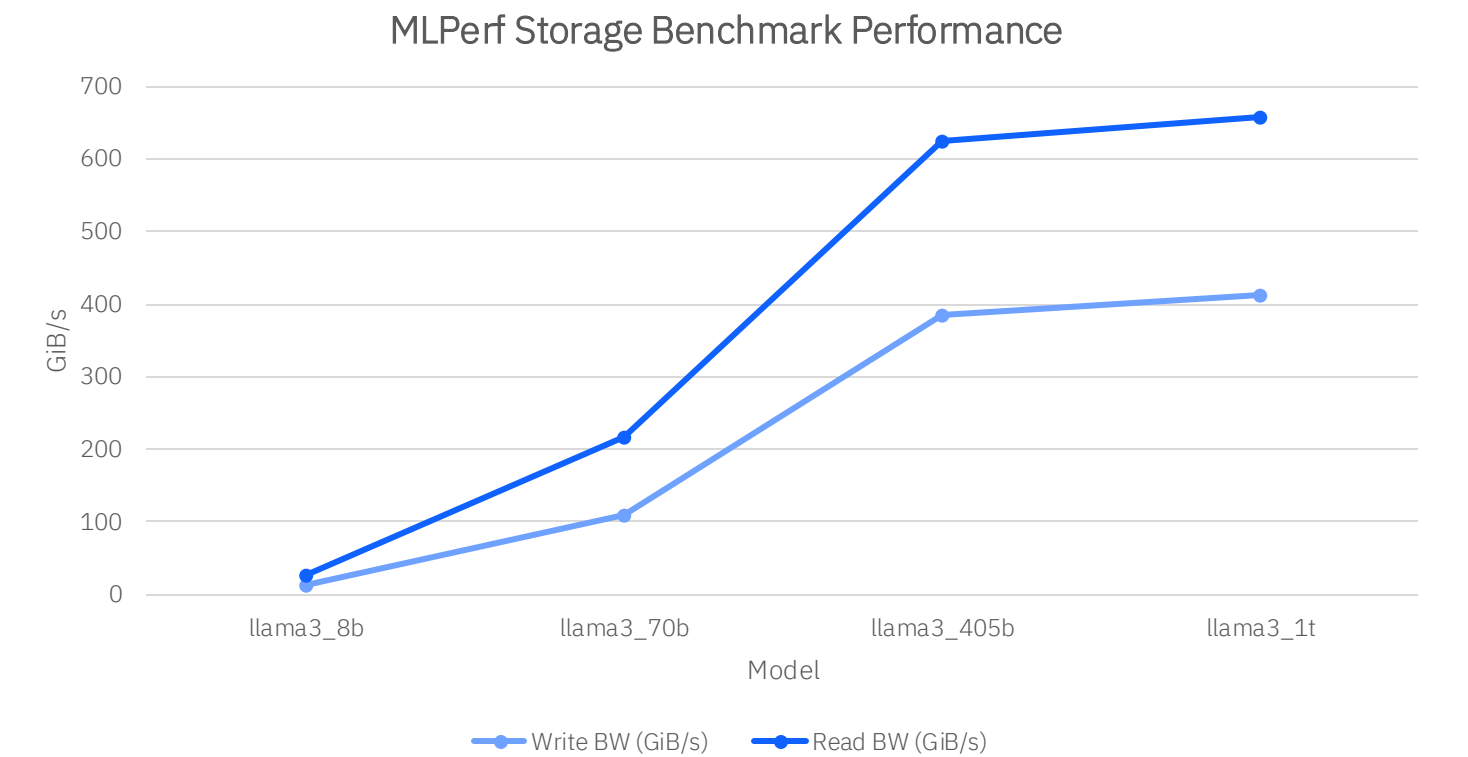
- ✓ READ: 320 GB/s
- ✓ WRITE: 155 GB/s

MLPerf Storage Benchmarks

Checkpointing Results

The results from our benchmarks were impressive; our tests with the Llama 3.1 1T model demonstrated a read bandwidth of 656.7 GiB/s and a write bandwidth of 412.6 GiB/s. To put this into perspective, this translates to roughly 23 seconds required to load a model checkpoint and about 37 seconds to save it. Similarly, the Llama 3 405B model showed nearly equivalent performance, with a read bandwidth of 624.7 GiB/s and a write bandwidth of 384.7 GiB/s, achieving approximately 8.5 seconds for loading and 14 seconds for saving

Blog: <https://research.ibm.com/blog/ibm-storage-scale-mlperf>



Model	Write BW (GiB/s)	Write Duration (Sec)	Read BW (GiB/s)	Read Duration (Sec)	Checkpoint Size (GiB)
llama3_8b	13.9908	7.48381	26.3972	3.96661	104.70
llama3_70b	109.999	8.28386	216.527	4.20892	911.22
llama3_405b	384.728	13.7709	624.705	8.47203	5298.05
llama3_1t	412.612	37.4345	656.689	23.4938	15445.92

*other workloads actively running

IBM Storage Scale System 6000 Now a Certified NVIDIA Cloud Partner



<https://community.ibm.com/community/user/storage/blogs/mike-kieran/2025/01/10/ibm-storage-scale-system-6000-now-a-certified-nvid>

IBM Storage Scale System 6000 is now a certified NVIDIA Cloud Partner (NCP) for HGX H100/H200/B200 systems. As a certified high performance storage partner for NCP, IBM Storage Scale System 6000 has demonstrated that it can deliver scalable high-performance IO to the most demanding AI training and inferencing workloads deployed on NVIDIA HGX GPUs in the cloud.



“The supercomputer will leverage **IBM Storage Scale System 6000** technology to deliver high-performance storage for AI, data analytics, and other demanding workloads.

As part of this agreement, CoreWeave customers can access the IBM Storage platform within CoreWeave’s dedicated environments and AI cloud platform.”

CoreWeave Partners with IBM to Deliver New AI Supercomputer for IBM Granite Models



NEWS PROVIDED BY
CoreWeave →
Jan 15, 2025, 08:00 ET

<https://www.prnewswire.com/news-releases/coreweave-partners-with-ibm-to-deliver-new-ai-supercomputer-for-ibm-granite-models-302351465.html>

- One of the first deployments of NVIDIA GB200 NVL72 at supercomputing scale
- Supercomputer will leverage IBM Storage Scale System to power AI research and development

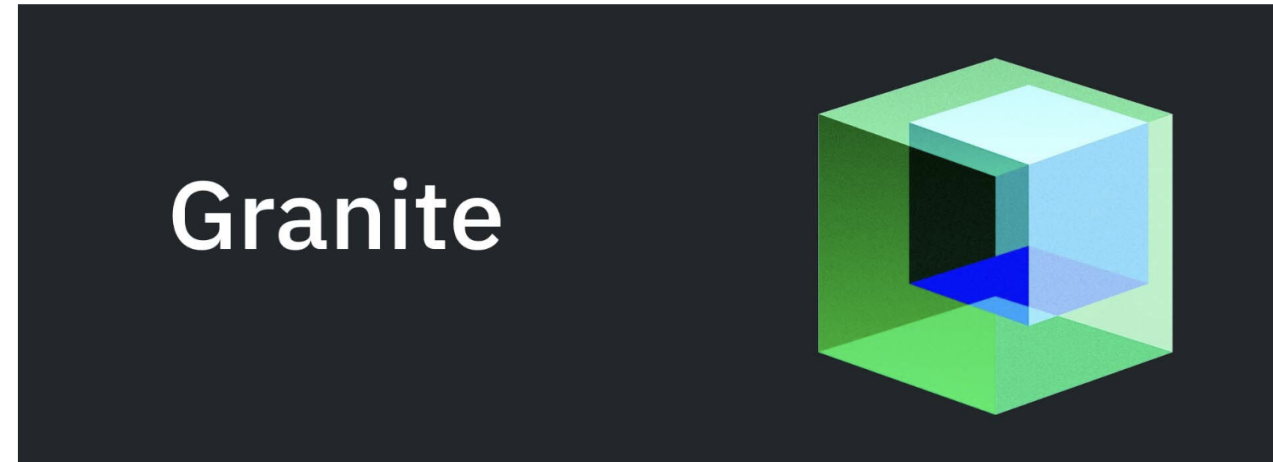




CoreWeave Partners with IBM to Deliver New AI Supercomputer for IBM Granite Models

- One of the first deployments of NVIDIA GB200 NVL72 at supercomputing scale
- Supercomputer will leverage IBM Storage Scale System to power AI research and development

Jan 15, 2025



f X in B M R

ROSELAND, N.J. and YORKTOWN HEIGHTS, N.Y. – January 15, 2025 – [CoreWeave](#), the AI Hyperscaler™, today announced its plans to deliver one of the first NVIDIA GB200 Grace Blackwell Superchip-enabled AI supercomputers to IBM (NYSE: [IBM](#)), equipped with NVIDIA [GB200 NVL72 systems](#), interconnected with NVIDIA [Quantum-2 InfiniBand](#) networking. IBM will use CoreWeave’s highly performant, reliable, and resilient cloud platform to train the next generations of its Granite models, IBM’s open source, enterprise-ready series of AI models that deliver state-of-the-art performance relative to model size while maximizing safety, speed, and cost-efficiency for enterprise use cases.

“We are thrilled to partner with IBM, a long-time pioneer of innovative technology solutions, to push the boundaries of artificial intelligence,” said Michael Intrator, CoreWeave CEO and co-founder. “This collaboration is a testament to CoreWeave’s ability to deliver some of the world’s most advanced AI cloud solutions and will combine our strengths in engineering and product development. We look forward to deepening our relationship with IBM to drive transformative innovation together.”

CoreWeave’s Cloud Platform is purpose-built to deliver industry leading performance, reliability, and resiliency, with enterprise-grade security. Its proprietary software and cloud services deliver the software and software intelligence needed to manage the most complex AI infrastructure at scale, and are trusted by some of the world’s leading AI labs and AI enterprises.

The supercomputer will leverage [IBM Storage Scale System](#), which is combined with NVMe flash technology to deliver high-performance storage for AI, data analytics, and other demanding workloads. As part of this agreement, CoreWeave customers can access the IBM Storage platform within CoreWeave’s dedicated environments and AI cloud platform.



CoreWeave, NVIDIA and IBM Submit Largest-Ever MLPerf Results on NVIDIA GB200 Grace Blackwell Superchips

NEWS PROVIDED BY
CoreWeave →
Jun 04, 2025, 11:08 ET

SHARE THIS ARTICLE



Submission with nearly 2,500 NVIDIA GB200 GPUs achieved breakthrough results on most complex benchmarking model

LIVINGSTON, N.J., June 4, 2025 /PRNewswire/ -- [CoreWeave](#) (Nasdaq: [CRWV](#)), in collaboration with NVIDIA and IBM, delivered the largest-ever MLPerf® Training v5.0 submission on [NVIDIA Blackwell](#), using 2,496 [NVIDIA Blackwell](#) GPUs running on CoreWeave's AI-optimized cloud platform. This submission is the largest NVIDIA GB200 NVL72 cluster ever benchmarked under MLPerf, 34x larger than the only other submission from a cloud provider highlighting the large scale and readiness of CoreWeave's cloud platform for today's demanding AI workloads.

The submission achieved a breakthrough result on the largest and most complex foundational model in the benchmarking suite—Llama 3.1 405B—completing the run in just 27.3 minutes. When compared against submissions from other participants across similar cluster sizes, CoreWeave's GB200 cluster achieved more than 2x faster training performance. This result highlights the significant performance leap enabled by the GB200 NVL72 architecture and the strength of CoreWeave's infrastructure in delivering consistent, best-in-class AI workload performance.

"AI labs and enterprises choose CoreWeave because we deliver a purpose-built cloud platform with the scale, performance, and reliability that their workloads demand," said Peter Salanki, Chief Technology Officer and Co-founder at CoreWeave. "These MLPerf results reinforce our leadership in supporting today's most demanding AI workloads."

https://www.prnewswire.com/news-releases/coreweave-nvidia-and-ibm-submit-largest-ever-mlperf-results-on-nvidia-gb200-grace-blackwell-superchips-302473361.html?utm_source=adwords&utm_medium=cpc&utm_campaign=MDF_BlackwellDSA_Broad&utm_term=blackwell%20ai&gc

Scale System in CoreWeave

Rack Layout

Overview:

- 7PB of Usable Flash using IBM Scale System 6000
- Estimated performance using Ethernet Up to 3.5TB/s Read and 1.75TB/s Write
- Estimated performance using RoCE
 - Up to 4.2TB/s Read and 2.1TB/s Write

Challenges:

- Link Layer Discovery Protocol (LLDP)
- ECMP + BGP Networking (FRR)
- CSI on ARM CPU
- CNSA on ARM on Ubuntu on Vanilla K8s
- Stateless nodes with lots of churn

