



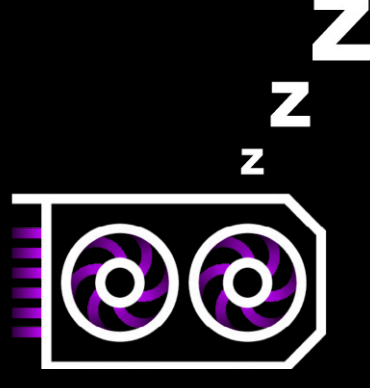
2 expensive mistakes to avoid with data center GPUs

Tips to get the most from your GPUs



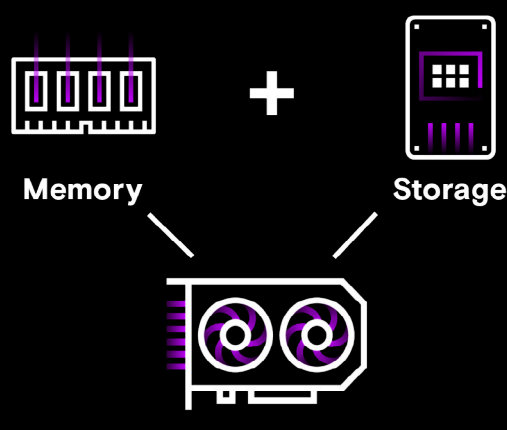
Mistake 1: Under-utilization

Under-utilizing GPUs means you're wasting expensive resources. GPUs are sitting idle when they should be hard at work. In this case, you aren't getting the returns you deserve from your investment, and you're wasting power, space, and potential performance.



Mistake 2: Over-utilization

Over-utilizing GPUs occurs when you push your GPUs too hard. Your power consumption increases, the GPUs are at risk of damage from overheating, and your performance suffers as you encounter bottlenecks, slow processing times, and reduced efficiency.



The solution: Right-sizing memory and storage

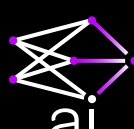
Optimally utilizing GPUs requires finding a balance with your memory and storage resources so the system can work efficiently, delivering high performance without excessive power consumption or risk of overheating.

Memory and storage are the key to achieving optimal GPU usage, especially for resource-intensive tasks like AI, ML, and GNN.



Use high-performance storage to keep GPUs fed

Purpose-built for AI workloads, the Micron 9550 SSD is a breakthrough, high-performance storage device with strong performance, latency, and power efficiency to keep GPUs running at optimal levels during demanding data center workloads.



Optimized for AI and other high-performance application¹



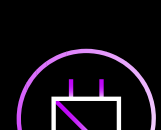
Significantly reduces power consumption¹



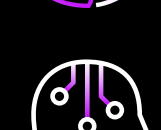
Micron-designed controller ASIC, G8 NAND, and DRAM²



Nearly triple the performance of DDR4³



Reduces memory power draw for better overall TCO⁴



Improves latency during AI training and inference processing⁵

Implement DDR5 DRAM to maintain high throughput

For AI/ML workloads, DDR5 is essential to keep GPUs fed and running at peak efficiency. Micron DDR5 Server DRAM delivers higher bandwidths along with improved reliability, availability, and scaling when compared to DDR4.

Discover purpose-built solutions

Our experts can work with you to keep your GPUs running at optimal levels while reducing power consumption and improving overall efficiency.

Learn more at microncpg.com/datacenter

1. Based on Micron engineering test results in AI training offload, measured SSD-to-GPU direct data transfer rate with a 1TB dataset, and standard AI performance benchmarks.
2. Refer to <https://www.micron.com/products/storage/nand-flash/232-layer-nand>
3. DDR5 MRDIMM data rate of 8,800MT/s are 2.75x faster than DDR4's maximum JEDEC rate of 3,200MT/s.
4. On DDR4 server memory, power management was on the motherboard instead of the module, powering empty slots as well as those in use.
5. Based on publicly available datasheets and JEDEC specifications as of April 2024.