5 considerations when configuring white box servers

Balancing cost, performance and power efficiency

Consider the trends

Data centers are changing significantly due to technological advancements, evolving business needs and the ever-growing appetite for computational power and storage. Adding to the challenge, server budgets are being cannibalized for AI servers, pulling resources away from mission-critical workloads.

In the face of growing pressure from all sides, data centers must find a way to achieve project goals, scale technology, and meet sustainability goals, all while implementing new AI capabilities.

To address these challenges, data center managers are increasingly turning to white box servers[1] — a trend fueled by three key factors:

Cost: White box servers break away from traditional vendor lock in, allowing organizations to choose components based on their specific needs. Shopping for individual deals on hardware and software can lead to substantial cost savings.

Performance: White box servers offer the flexibility to scale up or down, adapting to changing workloads and business growth. Scalability is especially crucial to meet the demands of emerging technologies such as AI.

Power efficiency: As environmental concerns mount and businesses aim to reduce operational costs, energy-efficient solutions become necessary. White box servers, often designed with power efficiency in mind, help data centers maintain performance while minimizing energy consumption.

Despite the advantages, setting up a white box server requires a strategic approach beyond merely adding compute cores or maximizing memory slots. Businesses must find optimal balance among cost, performance, and power efficiency.

White box servers are cheaper than traditional OEM options

White box servers often come with a lower upfront price tag compared to branded servers. In fact, businesses can expect to spend 20% less[2] on white box servers than traditional OEM servers.

It may be tempting to use those savings to max out the servers. However, just because the hardware is powerful, it doesn't mean it's the right choice.

It's all about balance

Choosing the right product line and capacity for your unique workload and server core counts helps minimize compute costs and optimize power efficiency.

- CPU loads over 90% can max out a CPU and result in excessive heat that shortens its lifespan, slow response times or system crashes
- CPU loads under 60% leads to underutilization of costly compute and idling CPUs and GPUs

Finding the right balance of server memory and storage helps ensure that data-hungry (and costly) compute hits a target CPU load between 80-90%.

What about support and warranty?

When purchasing the components for a white box server, you may not receive the same level of support and warranties that come included with branded servers. However, ODMs and SIs often fill this gap by offering maintenance and service packages tailored to custom server setups. These third-party services can be valuable for teams that lack the resources to handle ongoing maintenance and troubleshooting in house.

Align hardware with performance goals

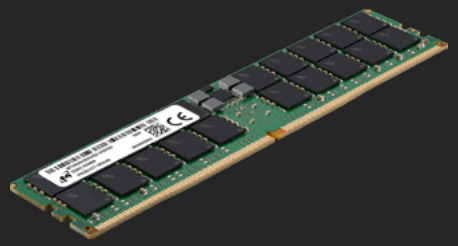
Data-hungry workloads like AI and high-performance computing (HPC) require fast access to massive amounts of data. A prime example is ChatGPT, which requires servers with five times[3] more memory and storage than standard servers.

For AI and HPC, speed takes priority. But for other use cases, reliability remains top of mind. In any case, the goal is to balance performance with your business needs.

For example, if you're using a server that has eight GPUs and two CPUs, you are likely limited to only eight drive slots, which results in a one-to-one relationship between drive slots and GPUs. In this case, you'll most likely want the fastest drive that you can put in that slot.

Conversely, if you're using a more standard 2U CPU-based server with a couple GPUs in it, you can spread your data set across multiple devices to get the optimal performance.

Other situations could require data lakes with hundreds of terabytes or more, putting pressure on the network storage system. In this case, it is imperative to take advantage of a high-capacity storage solution.

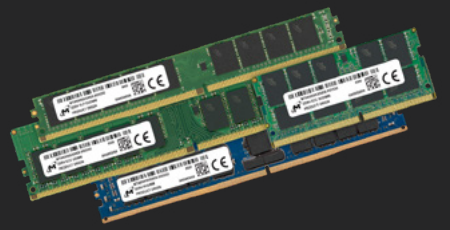


Micron® Next-gen DDR5 Server DRAM

Nearly doubles the performance of DDR4 for servers

Ideal for memory-intensive applications like:

- AI training
- HPC
- Cloud computing
- Virtualized supercomputing
- In-memory database applications



Micron® Mainstream DDR4 Server DRAM

Boosts performance at a fraction of the cost of next-gen server platforms

Ideal for mainstream applications like:

- Virtualization
- Cloud computing
- Big data

Micron® High-performance 9000 Series NVMe SSDs

Sets a new performance standard for PCIe® Gen4 storage and surpasses other major competitors by up to 2.3 times in mixed workload performance[4]

Ideal for mission-critical applications like:

- Caching and database acceleration
- AI/ML training and caching
- High-frequency trading
- Block and object stores
- Massive high-speed OLTP

Micron® Mainstream 7000 Series SSD

Offers the industry's broadest variety of form factors, including multiple U.3, M.2 and E1.S to support all major platform functions

Ideal for mainstream data center applications like:

- Boot
- Caching
- Databases
- Main data storage
- Object storage
- Software-defined storage
- Virtualization solutions

Micron® High-capacity 6000 Series SSD

Massive capacity and purpose-built performance to handle even the most challenging cloud-architecture or delivery networks

Ideal for applications that require massive storage capacity:

- Cloud infrastructure
- Hyperconverged infrastructure
- Content delivery networks
- Big data
- Object storage

Micron® Mainstream 5000 Series SATA

Get more from legacy server platforms with a proven architecture that provides unparalleled peace of mind, reliability and endurance

Ideal for mainstream data center applications:

- Hyperconverged infrastructure
- Cloud infrastructure
- Big data
- Object storage

Prioritize power efficiency

Data centers are facing increasing environmental concerns and rising operational costs. As a response, the industry has shifted its focus toward energy efficiency.

Large companies such as Facebook have been at the forefront of the movement, using the Open Compute Project (OCP) to share information about how the company reduced its energy consumption[5] by 38% by building a custom data center optimized for the task.

At first glance, these results may seem like an outlier. However, the benefits of white box servers aren't limited to hyperscalers like Facebook. On average, businesses see a 25% increase in energy efficiency[2] when implementing white box solutions.

Peak performance for any use case

Compute (CPUs and GPUs) often represent the most expensive components in white box server configurations, especially those designed for high-performance tasks such as data processing, artificial intelligence, or graphical computations. Those costs can compound over time if the system suffers from an inefficient design. Overutilization of these components can lead to a reduced lifespan and an unstable system, while underutilization can result in wasted potential and resources. Take time to plan for future workloads and ensure the server is configured for sustainable and cost-effective performance.

Partner for success

Micron memory and storage solutions provide fast, reliable performance in the most demanding environments. Data centers around the globe rely on Micron's leading DRAM and SSD technology to scale efficiently and optimize white box server configurations.

- Speed up processing times for resource-intensive applications
- Improve reliability with Micron's long-proven architecture
- Increase server and workstation performance to boost critical workloads
- Reduce total cost of ownership (TCO) by balancing compute, memory and storage

Discover your ideal data center solution

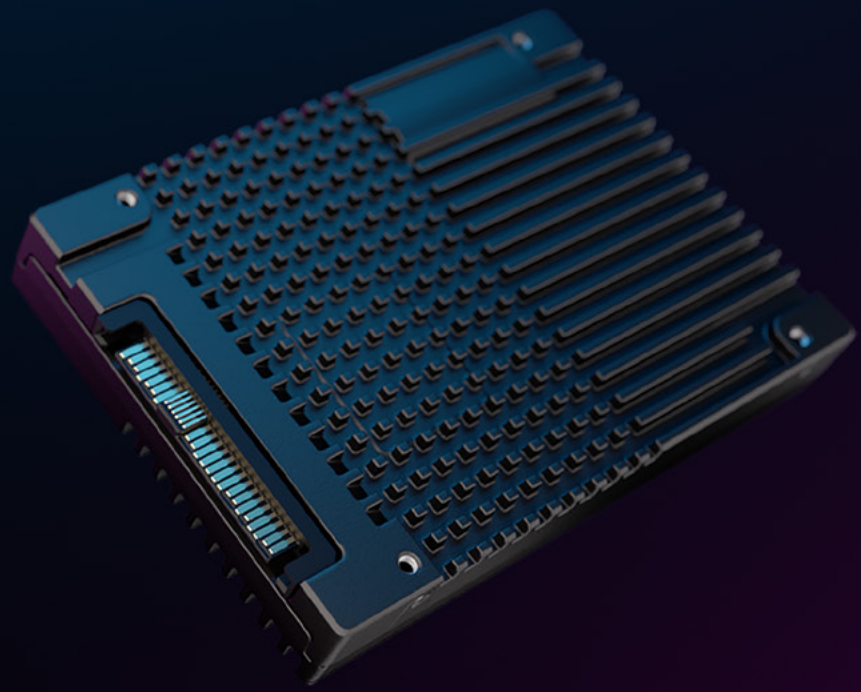
Finding the right balance of memory, storage and compute for your unique workload can be challenging. Micron is here to help.

With more than 45 years of memory and storage innovation and execution, Micron's experts are uniquely qualified to provide guidance on the optimal balance for white box servers. Our experts collaborate with teams across the ecosystem to rigorously assess configurations on a wide range of platforms with an even wider range of workloads. We can use that knowledge to help you choose the right products to meet the needs of your unique workloads, then advise you during validation testing and benchmarking to ensure project goals are met or exceeded.

Connect with an expert at Micron today to find your perfect mix of memory, storage, and compute.

Build for balance.

Learn more at micronpg.com/balance



Sources

[1] Yugandhara R. Y., "White Box Server Market Trend, Size, Share Report 2023-2030," ResearchGate, 2023.

[2] Jeremy Schaller, "Server Upgrade: Should you Purchase White Box Servers?" Exit Technologies, 2022.

[3] Tom Coughlin, "Smarter AI Needs More Memory And Storage," Forbes, 2023.

[4] Comparisons are made based on other leading PCIe Gen4 Data Center U.2/U.3 NVMe SSDs based on data center market share as noted in the Forward Insights SSD Supplier Status Q2/22 report and available on the open market at the time of this document's initial publication. 1GB = 1 billion bytes, formatted capacity is less.

[5] Zachary Oliver, Kyle Clark-Sutton, Sara VanLear, Lindsay Aramayo, Brian Lim, Carson Moss, Stephanie Zayed and Jeffrey Petrusa, "The Impact of Facebook's U.S. Data Center Fleet," RTI International, 2018.

Rev. 03/2024