

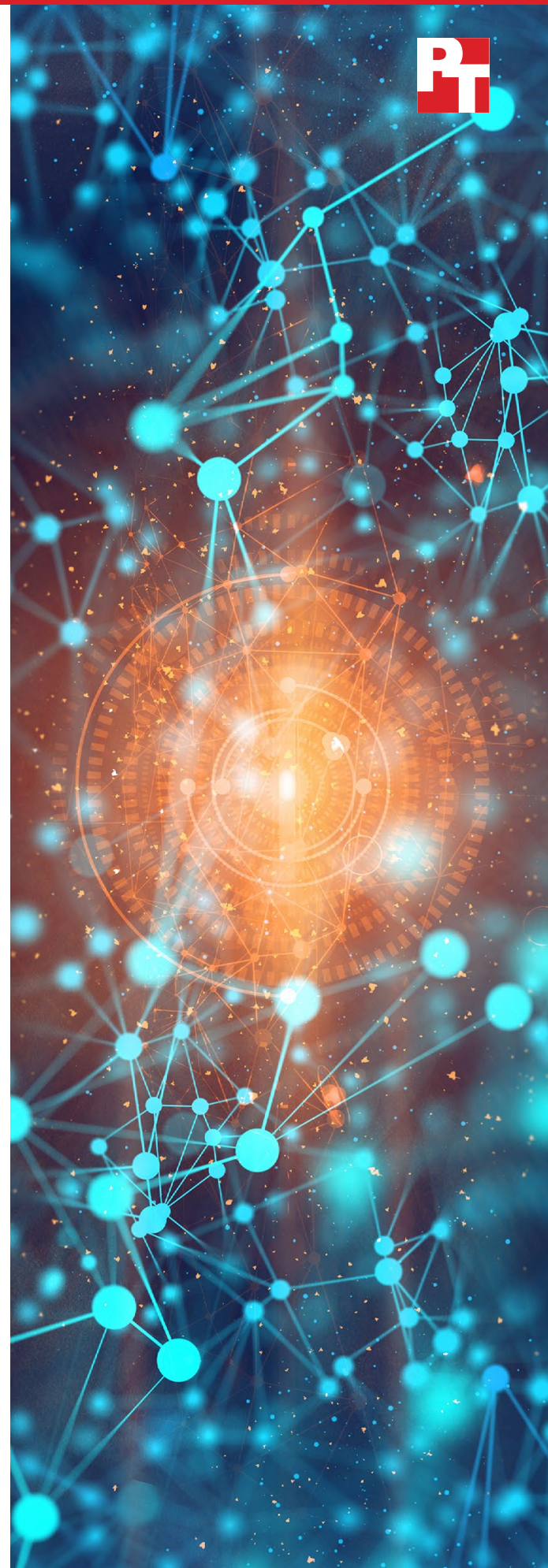


## Finding the path to AI success with the Dell AI portfolio

### A comparison of the Dell AI portfolio vs. similar offerings from Supermicro

Artificial intelligence (AI) is the new frontier, poised to reshape business operations across industries. While organizations across the spectrum explore how they can leverage AI to enhance their business operations, it's important to remember that implementing AI and reaping its benefits doesn't happen overnight. Because every business is unique, each must assess their data and their business goals to see how using AI with their data can bring about their specific desired outcomes. Engaging with a company such as Dell that offers a comprehensive AI portfolio including planning, data preparation, appropriate hardware selection, AI model design, proof-of-concept testing, reference architectures, and end-to-end support can translate to successful AI projects.

With the myriad options available on the market, finding a partner who can assist with all these decisions could mean the difference between a successful AI implementation and an expensive mistake. In this paper, we take a closer look at the AI portfolios of Dell and Supermicro with the aim of educating readers on the advantages Dell can provide throughout a customer's AI journey. We'll first focus on server and compute options, where each company has large and varied offerings for customers. Next, we'll explore how Dell goes beyond hardware considerations for companies seeking education, planning services, partner ecosystems, and more.



## Servers and performance results for AI workloads

Servers, the base computing infrastructure that powers AI workloads, can use CPUs, GPUs, or both as computational resources depending on the size or type of workload. For larger or more demanding workloads like HPC or AI, GPUs provide top performance. GPUs come in a variety of form factors, including universal PCIe, Open Compute Project Accelerator Module (OAM), and proprietary NVIDIA SXM architecture, which currently offers top performance.<sup>1</sup> Large memory capacities and server design features such as cooling architecture and power efficiency also affect performance. Most data centers still use air cooling, which means that AI workloads need servers built to cool with air as effectively as possible. Below, we highlight Dell PowerEdge server offerings in terms of components, cooling options, and more, along with their published MLCommons® MLPerf® scores.

### Test results

MLPerf® is a benchmark suite that tests AI performance for both training and inferencing. For an organization to publish official MLPerf® results, the results must be compliant with specific conditions set by the benchmark developer, MLCommons®.<sup>2</sup> These compliance guidelines provide standards that make it easier to compare performance. For inference testing, MLPerf® uses Datacenter, Edge, Mobile, and Tiny datasets, and reports AI scores and watts of power consumed during testing. The inference benchmark suite includes testing for many common AI, ML, and DL models (see Table 1).

Table 1: AI, ML, and DL models that MLPerf® includes, with testing and typical use cases for each. Source: Principled Technologies.

Common AI models	Typical use cases
ResNet	An image classification model that helps computers learn, remember, and identify different images for use cases such as medical imaging, social media content moderation, and facial recognition
RetinaNet	A type of object detection that can handle more complexity than ResNet. It helps computers to identify and locate objects within images or video frames, and can classify them by importance. Used for things like autonomous driving, vehicle auto-assist technology, surveillance, facial recognition
3D-UNet	Specific to medical image segmentation
RNN-T	Speech recognition for use cases such as automated language translation
BERT	Natural language processing for use cases such as text summarization, language translation, and autocompletion of tasks
DLRM-v2-99.9	Recommendation model for use cases such as targeted ads and personalized product recommendations
GPTJ-99 and 99.9	LLM for natural language processing that excels at text generation for use cases such as chatbots and chat-based AI tools





## About MLPerf

MLPerf® results include several parameters in addition to the AI models themselves, which can make for a lot of data to parse in a single chart or table. Here’s a quick reference to these parameters:

- 99.0 and 99.9: These numbers refer to the accuracy to which the model was trained. The more accurate you need the output to be, the more complex the model and the longer it can take to process data.
- Offline samples/sec: Mode where the benchmark sends all queries at the beginning of the test, simulating data already present on the system.
- Server queries/sec: Mode where the benchmark sends queries throughout the test duration, simulating analyzing a live stream of data.

For more about MLCommons® and MLPerf® results, see <https://mlcommons.org/benchmarks/inference-datacenter/>.

Results in this report come from MLPerf® v3.1 Inference Datacenter results published on the MLCommons® website in November 2023.<sup>3</sup> These results include submissions from technology manufacturers and cloud service providers and cover a range of configurations. Compared to publicly available submissions from Supermicro, Dell PowerEdge servers produced comparable results. Table 2 provides server details.

Table 2: Dell and Supermicro servers included in the MLCommons® MLPerf® 3.1 results published as of November 2023. Source: Principled Technologies.

Submitter	Server model	# and model of GPUs	Description
Dell <sup>4</sup>	PowerEdge XE9680	8x NVIDIA H100 SXM	For AI training and inference with large workloads such as large language models
	PowerEdge XE9640	4x NVIDIA H100 SXM	For training large AI models in high density and liquid cooled datacenters
	PowerEdge XE8640	4x NVIDIA H100 SXM	For driving traditional AI training, HPC, and data analytics apps in a 4U form factor for air-cooled data centers
Supermicro <sup>5</sup>	AS-8125GS-TNHR	8x NVIDIA H100 SXM	For large-scale AI training and HPC workloads with AMD processors
	SYS-821GE-TNHR	8x NVIDIA H100 SXM	For large-scale AI training and HPC workloads with Intel processors
	SYS-421GU-TNXR	4x NVIDIA H100 SXM	Modular design for flexibility for supporting HPC and AI workloads

Because both Dell and Supermicro have submitted results with apples-to-apples GPU configurations, comparing performance is straightforward. As Figures 1 and 2 show, commonality between configurations from both vendors leads to results that are largely on par with one another. For other configurations, such as those we show in Figures 3 and 4, Dell outperformed Supermicro on the gptj-99.9 model in the 4-GPU testing. Note that while Dell has submitted results for all available models with all three servers, Supermicro did not. We compare only the models for which both servers have results. To see the full set of Dell results, visit MLCommons® MLPerf® results.

## Eight-GPU server results

The Dell PowerEdge XE9680 offers support for up to eight NVIDIA H100 SXM5 GPUs for AI acceleration and up to two 4<sup>th</sup> Generation Intel® Xeon® Scalable processors. The PowerEdge XE product family has a modular architecture supporting SXM4 or SXM5 NVIDIA GPUs or Open Compute Project Accelerator Module (OAM) GPU assemblies, which can boost performance compared to a standard PCIe GPU. The Dell PowerEdge XE9680 also offers the AMD Instinct™ MI300X accelerator.<sup>6</sup> Taking up only 6U of rack space, the PowerEdge XE9680 is a compact eight-way NVIDIA H100 SXM5 server.

In contrast, the 8-GPU Supermicro server requires 33 percent more rack space at 8U while also offering the SXM form factor for NVIDIA GPUs. The size difference means you could fit seven Dell PowerEdge XE9680 servers in a rack versus only five Supermicro servers. In Figures 1 and 2, we compare the Dell PowerEdge XE9680 results to those of two configurations of the Supermicro 8-GPU server: the SYS-821GE-TNHR with Intel processors and the AS-8125GS-TNHR with AMD processors. Note that Supermicro did not submit results for RNN-T on the SYS-821GE-TNHR, so we exclude this model from the chart in Figure 1.

### Normalized MLPerf® Results: Dell PowerEdge XE9680 with H100 SXM5 vs. Supermicro SYS-821GE-TNHR with H100 (Larger is better)

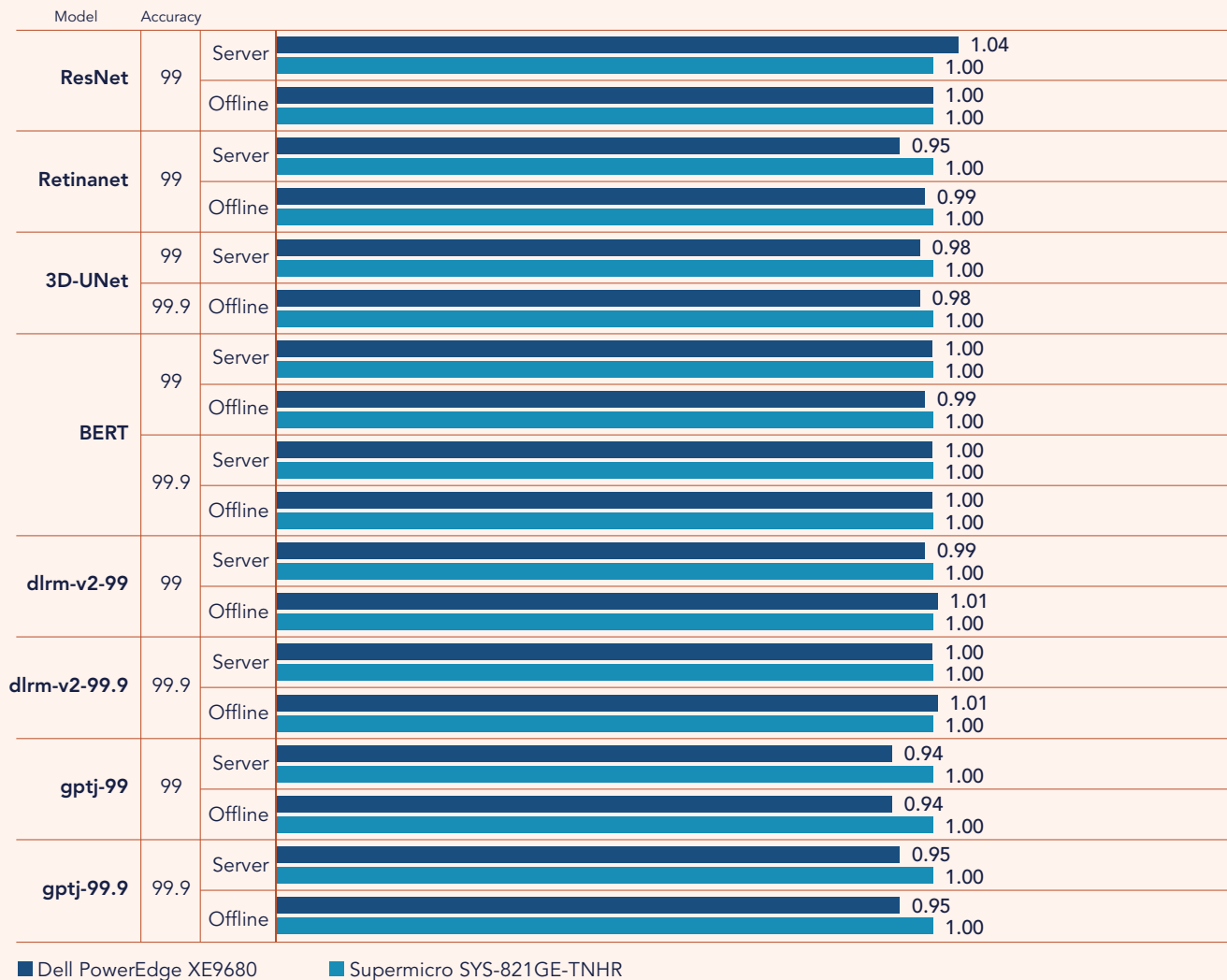


Figure 1: Published MLPerf® results for the Dell PowerEdge XE9680 and Supermicro SYS-821GE-TNHR as of 11/29/23. Both systems equip the SXM form factor of the NVIDIA H100 GPU. Source: Principled Technologies using data from MLCommons®.<sup>7,8</sup>



## Normalized MLPerf® Results: Dell PowerEdge XE9680 with H100 SXM5 vs. Supermicro AS-8125GS-TNHR with H100 SXM5 (Larger is better)

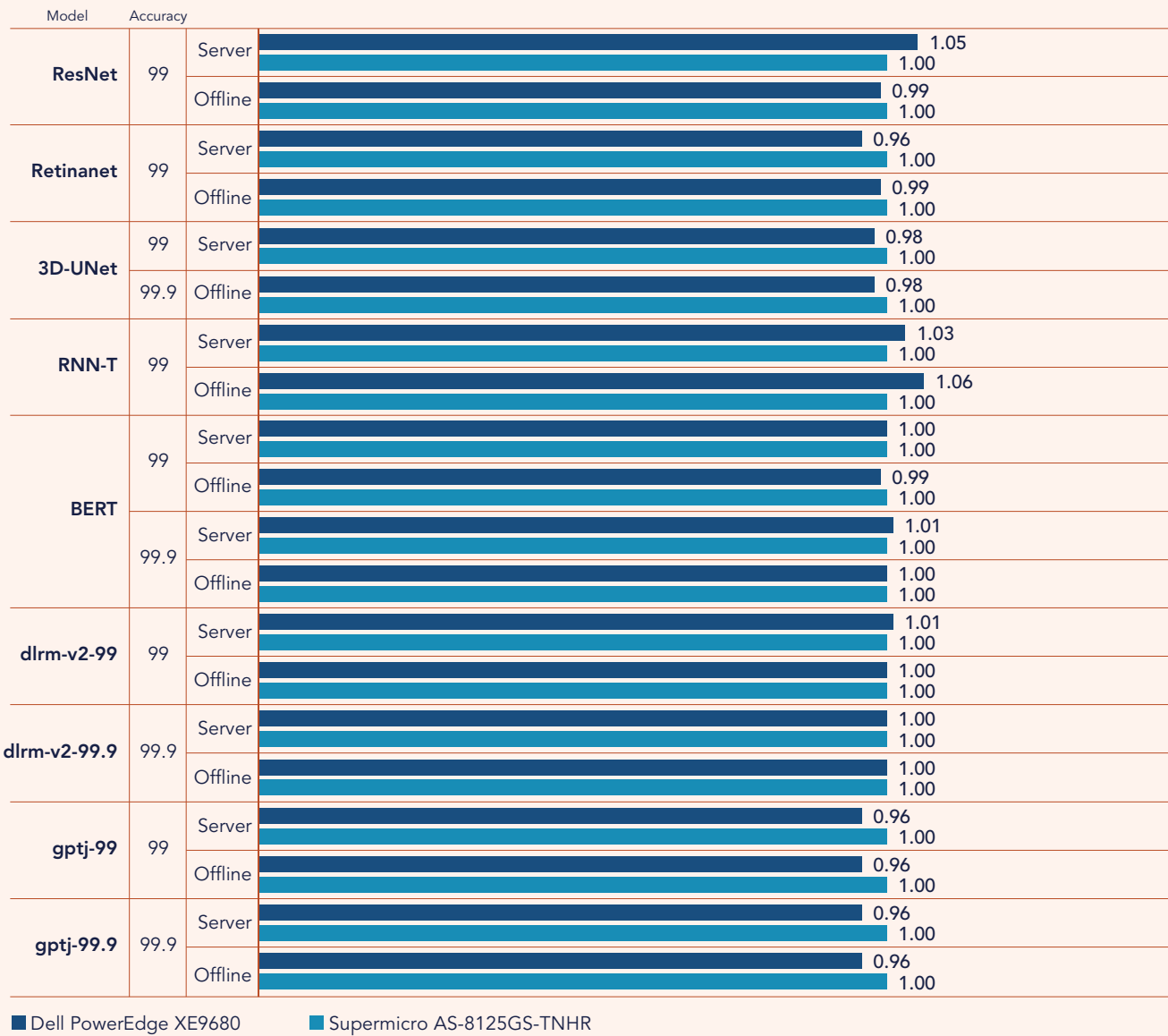


Figure 2: Published MLPerf® results for the Dell PowerEdge XE9680 and Supermicro AS-8125GS-TNHR as of 11/29/23. Both systems equip the SXM form factor of the NVIDIA H100 GPU. Source: Principled Technologies using data from MLCommons®<sup>9,10</sup>

As these results show, choosing the Dell PowerEdge XE9680 will give you similar performance on a myriad of AI inference workloads while taking up less data center space.



## Four-GPU server results

When minimizing data center power usage or space is of top importance, the 2U Dell PowerEdge XE9640 could provide the answer. With up to four NVIDIA H100 SXM GPUs, the PowerEdge XE9640 offers half the GPU computational power of the PowerEdge XE9680 in one-third the space.<sup>11</sup> The densely packed Dell PowerEdge XE9640 server incorporates Dell Smart Cooling technology, providing an array of thermal technology including direct liquid cooling for CPUs and GPUs.<sup>12</sup> The 2U chassis of the PowerEdge XE9640 accommodates improved airflow mechanisms, including larger fans and heatsinks, to help cool the other vital components, such as PCIe cards and memory.<sup>13</sup>

Supermicro offers an older SYS-220GQ-TNAR+ server that provides four NVIDIA A100 HGX GPUs in a 2U form factor, but we were unable to find any 2U Supermicro servers with four of the more recent H100 HGX GPUs that would match the PowerEdge XE9640.<sup>14</sup> The 4-GPU server with HGX H100 NVIDIA GPUs from Supermicro in the MLPerf<sup>®</sup> submissions is the SYS-421GU-TNXR, which is a 4U server. As we mentioned previously, Supermicro submitted MLPerf<sup>®</sup> 3.1 results for only the SYS-421GU-TNXR on the gptj-99.9 AI model, so we cannot compare it to the PowerEdge XE9640 on the other models. However, in the published results, the PowerEdge XE9640 outperformed the Supermicro server in the offline tests, achieving up to 1.37 times the score (see Figure 3).

### Normalized MLPerf<sup>®</sup> Results: Dell PowerEdge XE9640 with H100 SXM5 vs. Supermicro SYS-421GU-TNXR with H100 SXM5 (Larger is better)

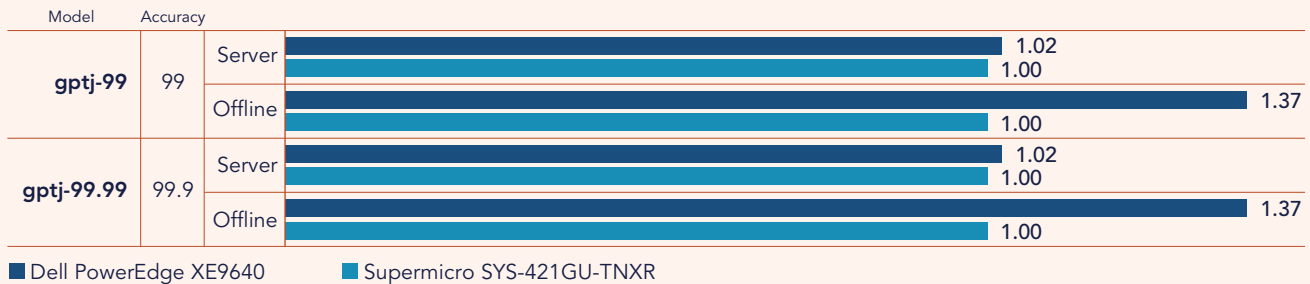


Figure 3: Published MLPerf<sup>®</sup> results for the Dell PowerEdge XE9640 and Supermicro SYS-421GU-TNXR as of 11/29/23. Both systems equip the SXM form factor of the NVIDIA H100 GPU. Source: Principled Technologies using data from MLCommons<sup>®</sup>.<sup>15,16</sup>

We can also compare the Supermicro SYS-421GU-TNXR server results to the Dell PowerEdge XE8640, a 4U 4-GPU server that also supports NVIDIA H100 HGX GPUs. While larger than the PowerEdge XE9640, the PowerEdge XE8640 does not require direct liquid cooling, making it a compromise between density and cooling technologies for data centers that don't have access to water cooling. The PowerEdge XE8640 has air cooling for processors and a Liquid-Assisted Air Cooling Radiator for the GPUs, which does not require facility water-to-rack availability.<sup>17</sup> The Dell PowerEdge XE8640 features the latest 4<sup>th</sup> Generation Intel Xeon Scalable processors and up to 4 TB of memory to handle the large datasets and complex computations common in AI and data analytics.<sup>18</sup> With its 4U form factor, the PowerEdge XE8640 is similar to the Supermicro SYS-421GU-TNXR in both density and GPU capabilities. However, much as we saw with the PowerEdge XE9640, the Dell PowerEdge XE8640 gets better gptj-99 scores in the offline tests than the Supermicro server (Figure 4).

## Normalized MLPerf® Results: Dell PowerEdge XE8640 with H100 SXM5 vs. Supermicro SYS-421GU-TNXR with H100 SXM5 (Larger is better)

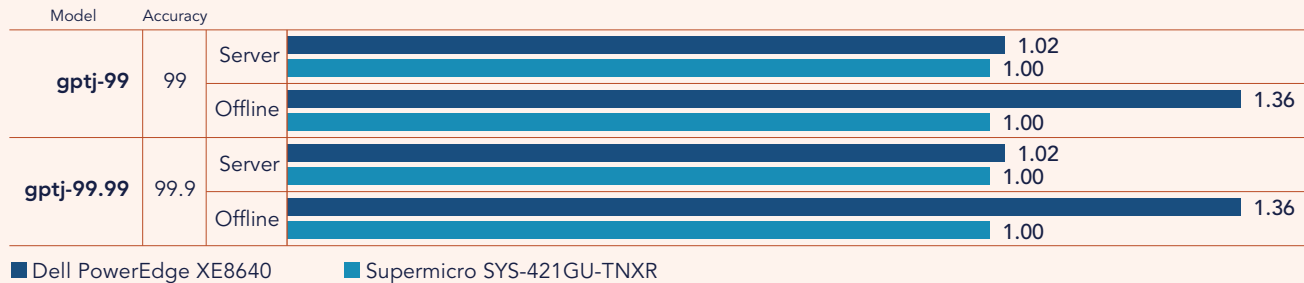


Figure 4: Published MLPerf® results for the Dell PowerEdge XE8640 and Supermicro SYS-421GU-TNXR as of 11/29/23. Both systems equip the SXM form factor of the NVIDIA H100 GPU. Source: Principled Technologies using data from MLCommons®.<sup>19,20</sup>

As we have seen, the MLPerf® performance results of the Supermicro and Dell GPU-based server offerings are similar overall, with a notable advantage for the Dell PowerEdge XE8640 and PowerEdge XE9640 servers on one AI model.

Because performance is only one aspect of the AI implementation journey, we also looked at the rest of the AI offerings from Dell and Supermicro, from workstations, storage, and networking to services, support, education, and more. We found that the Dell AI portfolio is broader than the Supermicro offering and offers solutions to many of the obstacles companies face beyond compute performance.

## Client workstations and storage offerings

### Additional compute options with workstations

Some AI use cases require a different compute approach, and not everyone who needs access to AI-capable hardware can remain tied to the data center. Scientists working in labs may not have room for a rack of servers, and it's impractical for those working at the edge to lug around a large, heavy desktop system. That's where workstations come into play. The Dell AI portfolio includes several AI-ready Precision workstations—including tower, mobile, and rack configurations—to meet a variety of needs.<sup>21</sup> In contrast, Supermicro does not offer mobile workstations for easy, on-the-go access. Its GPU workstation offerings consist of several different tower configurations, some of which Supermicro claims users can rack mount.<sup>22</sup> Our research uncovered two rack-mountable workstations, but they appear to be older, available with only NVIDIA A100 GPUs, and possibly no longer for sale.<sup>23</sup> If your company needs flexibility in the type and mobility of GPU workstations it deploys, the Dell AI portfolio is better suited to those needs.

### Storage considerations

Storage may be just as vital as compute when running AI workloads. Having more data improves AI model accuracy, but storing and managing massive datasets can challenge the capabilities of many data centers. Additionally, because models typically train using unstructured data, AI-ready storage systems must handle many different data types with ease.<sup>24</sup> To provide capacity and scaling for AI, ML, and DL datasets, Dell offers the PowerScale™ series for file storage and Elastic Cloud Storage (ECS) or software-defined ObjectScale for object storage.

The Dell PowerScale all-flash NAS portfolio offers capacity options ranging from 3.84 TB up to 720 TB of raw capacity per node, with clustered all-flash capacities reaching 186 PB of raw capacity. The flexibility and scale of PowerScale can support a wide variety of customers and AI use cases.<sup>25</sup> All three all-flash PowerScale models—F200, F600, and F900—include inline data compression and deduplication to improve storage efficiency.<sup>26</sup> Each PowerScale storage model uses the Dell OneFS™ file system, which takes advantage of advanced tiering policies to ensure the most frequently accessed data resides on the highest performing storage tiers.<sup>27</sup> Dell also offers OneFS software in the Amazon Web Services (AWS) marketplace with APEX File Storage for AWS. Customers can leverage OneFS with their AWS compute instances for a consistent user experience with the same features available in on-premises OneFS arrays.<sup>28</sup>

Supermicro storage offerings consist of storage servers—rack-mounted, storage-dense servers at various sizes and densities.<sup>29</sup> To obtain file storage from Supermicro, customers must choose from various third-party software-defined storage offerings such as WekaIO, Scality RING, or OSNEXUS.<sup>30</sup> While Scality RING and OSNEXUS include file storage options as part of their platform descriptions, WekaIO appears to be the primary option for customers seeking basic file storage. Supermicro offers several reference architectures covering a wide array of use cases, but customers must have a WekaIO software subscription or license, which can increase the overall cost of the solution.<sup>31</sup>

Object storage options from Dell include Dell ECS Enterprise Object Storage, which is “purpose built to store unstructured data at public cloud scale.”<sup>32</sup> Along with built-in compatibility with Amazon S3 object storage for hybrid cloud functionality, ECS storage nodes deliver capacities up to 14PB per rack.<sup>33</sup> As with file storage, the Supermicro object storage offerings require third-party configurations. The OSNEXUS platform is a combined storage platform for file, block, and object storage, while the Scality RING solution combines file and object storage; both require licenses with the third-party vendors.<sup>34,35</sup> For just object storage, customers can buy the Supermicro solution for Quantum ActiveScale for private cloud object storage with a Quantum software subscription.<sup>36</sup> As of this writing, we could not find any flexible consumption/operating expenditure options that Supermicro offers.

Because Supermicro storage offerings require customers to engage with third-party software vendors, they are likely to incur additional licensing or subscription costs and may experience difficulties in support, troubleshooting, and more. The Dell AI portfolio gives Dell storage customers a single, reliable services-and-support solution across every aspect of their storage solution.

## About Dell APEX

For customers who wish to consume their file storage as a service, Dell offers APEX Data Storage Services, which includes file, block, and backup storage. Using the Dell APEX Console, customers can order new subscriptions, adjust and monitor storage capacity, and more. Per Dell, this solution allows you to “get the ease and agility of the cloud experience with more control over your applications and data.”<sup>37</sup>

To read more about Dell APEX, visit <https://www.dell.com/en-us/dt/apex/storage/data-storage-services/index.htm>







## Networking options

Networking is another vital component of AI infrastructure. With many AI workloads running on large clusters of servers requiring constant communication between each other and to storage, AI workloads need robust networking to avoid bottlenecks. If your networking is insufficient for the AI workload, training and inference times will increase, slowing data processing and time to insights. Dell offers PowerSwitch Data Center top-of-rack (ToR) switches and PowerEdge MX I/O modules for Ethernet and fabric networks.<sup>38</sup> The PowerSwitch offerings range from 1GbE to 400GbE to meet a range of needs. Additionally, the Dell PowerSwitch Z Series switches offer 100GbE and 400GbE connections optimized for leaf/spine fabrics.<sup>39</sup>

Supermicro also offers switches with Ethernet ports up to 400GbE for ToR and other applications such as Data Center Spine and leaf.<sup>40</sup> However, Dell networking services provide several ease-of-use and flexibility benefits that Supermicro does not offer. Services such as the Dell Fabric Design Center can help avoid networking mismatches, gaps, or inefficiencies, helping customers plan and deploy network fabrics with automation.<sup>41</sup> For specific environments such as VMware VxRail, VMware ESXi, and Dell PowerStore configurations, Dell offers SmartFabric services, which enable software-defined infrastructure deployment and lifecycle management. With PowerStore, SmartFabric Services can automate "up to 99% of LAN connectivity tasks with a plug-and-play fabric."<sup>42</sup> Services such as these, providing automation, guidance, and more for networking design and implementation, support customers as they navigate the AI adoption process.

## Services, training, and more

The primary non-hardware challenge in implementing AI is the need for in-house expertise for strategy, planning, data preparation, and management. Managing and maintaining AI workloads requires a unique knowledge set, including both more traditional hardware expertise and machine learning operations and data science. The people designing and implementing the AI strategy also need a deep understanding of the company's unique operational goals to ensure the new AI workloads meet those goals.<sup>43</sup>

Another significant hurdle can be integrating AI seamlessly into existing operational systems. This integration demands a strategic alignment of new AI technologies with current business processes, ensuring that the introduction of AI keeps established workflows intact. Partnering with a company such as Dell that provides several optimized, validated solution reference architectures, training courses, management options, and a large partner ecosystem can ease your AI adoption journey.

## Professional services for AI

To cover education and planning, Dell offers a variety of services specifically for AI.<sup>44</sup> Dell services supporting AI implementation encompass consulting, data preparation, deployment, support, and education, each targeting specific aspects of AI adoption. Dell Advisory Services for Generative AI help customers create a road map identifying use cases and helping companies streamline their processes.<sup>45</sup> Similarly, Adoption Services for Generative AI provide workshops with Dell professionals to review your needs and unique challenges to determine a pretrained model for your business and conduct knowledge-transfer sessions to train your IT staff.<sup>46</sup> Dell also offers Implementation, Scale, and Managed Services for Generative AI that provide varying levels of support and training, up to a fully managed AI infrastructure, that free your IT staff to focus on the models and data while Dell manages the hardware.<sup>47</sup> ProSupport services ensure optimal system performance and provide essential hardware and software assistance for ongoing AI operations, addressing technical issues.<sup>48</sup>

Educational services are integral in fostering the necessary skills and knowledge for AI utilization. Dell training offerings include comprehensive training programs in data science, advanced analytics certifications, and workshops on specific AI technologies, such as machine learning.<sup>49</sup>

Supernano services, on the other hand, are mostly limited to troubleshooting, manuals, return merchandise authorizations (RMAs), and warranties.<sup>50</sup> We could not find any design, implementation, management, or educational services in the Supernano AI portfolio. For companies that seeking a training partner in navigating the complexities of AI adoption, Dell is the clear choice of these two.

## Third-party partnerships for AI workloads

Dell Technologies and NVIDIA collaborate to offer Dell Validated Designs, which aim to provide a comprehensive solution for generative AI in business settings. This project creates a scalable, high-performance infrastructure based on Dell and NVIDIA technologies and software along with an AI model framework that enables businesses to build and run custom AI models. The solution enables customers to get GenAI workloads up and running quickly.<sup>51</sup> To learn more, read the Dell Validated Designs section below.

Dell has partnered with several companies to enhance AI technology applications. With Hugging Face, Dell makes setting up large language models (LLMs) on site easier. This partnership combines Hugging Face AI expertise with Dell servers and storage systems. A Dell-specific Hugging Face portal will provide tools for straightforward, secure deployment of Hugging Face open-source AI models. The ongoing goal is to keep improving these models for Dell systems, boosting performance and supporting new AI applications.<sup>52</sup>

Dell and Starburst are working on a high-performance, scalable data lake that integrates Starburst analytics with Dell computing and storage technology, looking to offer a single point of access to all data sources for AI and ML tools. Customers will be able to take advantage of this partnership to help eliminate data silos.<sup>53</sup>

According to our research, Supernano has much more limited partnerships for AI. SiMa.ai and Supernano have collaborated to develop the Supernano SYS-E300-13AD, a compact Edge ML Server designed for multi-stream video analytics processing. This server, equipped with SiMa.ai ML pipeline on a chip, efficiently handles multiple video channels, reduces the total cost of ownership, and improves reliability and security. The server offers a computational setting designed for the processing and analysis of numerous video streams, delivering edge intelligence suitable for various enterprise applications.<sup>54</sup>

## Dell Validated Designs

To help take the guesswork out of AI hardware solutions, Dell offers lab-validated reference architectures optimized for several AI and other workloads. These Validated Designs include architectural concepts, full solution overviews, and performance and other lab validations proving the solution's capabilities on the workload for which it was designed. These workloads include virtualized environments, MLOps, machine learning, conversational AI, GenAI Inference, GenAI model tuning, NVIDIA Fleet Command, and OpenShift AI.<sup>55</sup>

As one example, the AI for Virtualized Environments Validated Design combines VMware-enabled AI with NVIDIA AI Enterprise on Dell infrastructure, optimizing AI in virtual settings.<sup>56</sup> The Validated Design Guide includes performance results showing ResNet model training proving to customers that the design works and showing what kind of performance they could expect to see.<sup>57</sup> These validations provide customers value beyond simply listing hardware that work together by explaining concepts, recommending configurations, and walking customers through considerations and performance expectations.<sup>58</sup>

Supermicro does offer solutions on a use-case basis, but stops short of the level of reference architecture we saw with the Dell Validated Designs. Rather than recommend a purpose-built solution for specific workloads, Supermicro organizes its servers and GPUs into categories, such as AI inference and training, HPC/AI, visualization and design, and more.<sup>59</sup> In their brochures and datasheets, these categories consist of a few servers and GPUs they recommend as best suited for the task, several use cases, and lists of key technologies at play and software suggestions.<sup>60</sup> Unlike the Dell Validated Designs, they do not appear to include network architectures and performance or validation data. Supermicro also provides several Reference Designs that offer a more detailed reference architecture for some AI solutions, such as a large-scale AI training with liquid cooling solution brief Supermicro published with NVIDIA.<sup>61</sup> Customers may be able to find a more in-depth reference architecture for a specific scenario, but at the time of our research, we found only three: the liquid cooling architecture we previously mentioned, an AI workstation architecture, and a RedHat OpenShift architecture.<sup>62</sup>

Overall, we found that Dell Validated Designs covered more AI workloads and offered more in-depth guidance than the Supermicro offerings.



## Management services and iDRAC

According to an [April 2023 report](#) by Principled Technologies, the Integrated Dell Remote Access Controller (iDRAC) offers several advanced features over the Supermicro Intelligent Platform Management Interface (IPMI), particularly in automation, security, and configuration.<sup>63</sup> Table 3 presents a comparison of Dell and Supermicro management features from that report that shows how iDRAC could provide easier deployment, easier firmware updating, and more security features than Supermicro IPMI. Note that some findings may have changed since original publication.

Table 3: Summary of April 2023 Principled Technologies comparison between Dell and Supermicro management tools. Some findings may have changed since publication. Source: Principled Technologies <https://facts.pt/V5fDf06>.

	What's different with Dell management tools	How much better
<b>Easier firmware updates</b> <i>iDRAC9 vs. Supermicro IPMI</i> <i>OME vs. Supermicro SSM</i>	<ul style="list-style-type: none"> <li>Automated online updates with iDRAC9, with scheduling options</li> <li>OME allows for the creation of custom firmware repositories and can update firmware of BIOS, BMC, and other server components without additional tools or agents</li> </ul>	<ul style="list-style-type: none"> <li>We set up automatic updates in iDRAC in <b>just 74 seconds</b></li> <li>Supermicro IPMI has <b>no automatic update function</b> available, so admins must update manually</li> <li>SSM supports only BIOS and BMC firmware updates and requires SUM to update other components</li> </ul>
<b>More security features</b> <i>iDRAC9 vs. Supermicro IPMI</i> <i>OME vs. Supermicro SSM</i>	<ul style="list-style-type: none"> <li>iDRAC9 offers MFA and dynamic disabling of USB ports with no system downtime</li> <li>OME offers both role-based access control (RBAC) and scope-based access control (SBAC) to restrict device management to a subset of device groups</li> </ul>	<ul style="list-style-type: none"> <li>Supermicro IPMI has <b>no MFA features</b></li> <li>Supermicro IPMI requires a system reboot and entering BIOS configuration to disable USB ports</li> <li>Supermicro SSM offers RBAC but not the more restrictive SBAC</li> </ul>
<b>Easier lifecycle management</b> <i>OME vs. Supermicro SSM</i>	<ul style="list-style-type: none"> <li>Full, agentless lifecycle management via OME to ease management and monitoring</li> </ul>	<ul style="list-style-type: none"> <li>SSM requires the SuperDoctor5 agent for detailed local system health metrics and Supermicro Update Manager (SUM) to update additional components</li> </ul>
<b>Easier server deployment</b> <i>iDRAC9 vs. Supermicro IPMI</i>	<ul style="list-style-type: none"> <li>Import a complete Dell server profile in just 12 steps using iDRAC9</li> <li>Robust BIOS configuration options with iDRAC9 with 52 BIOS features and support for components configuration such as RAID NIC &amp; iDRAC</li> </ul>	<ul style="list-style-type: none"> <li>Supermicro IPMI allowed us to save and restore only the IPMI configuration rather than the whole server profile</li> <li>iDRAC9 has 52 BIOS features, while IPMI offers <b>no BIOS configuration options</b></li> </ul>
<b>More options for reporting and analytics</b> <i>iDRAC9 vs. Supermicro IPMI</i> <i>OME vs. Supermicro SSM</i>	<ul style="list-style-type: none"> <li>iDRAC9 offers telemetry streaming, which lets users easily send server data to analytics tools such as Splunk</li> <li>OME sends telemetry data directly to CloudIQ for easier monitoring</li> </ul>	<ul style="list-style-type: none"> <li>IPMI offers only a SYSLOG feature admins can use to send messages for aggregation and eventual analysis</li> <li>SSM has no cloud-based management solution equivalent to Dell CloudIQ</li> </ul>
<b>More sustainability features</b> <i>OME vs. Supermicro SSM</i>	<ul style="list-style-type: none"> <li>More metrics for monitoring in OME Power Manager, including carbon footprint data</li> </ul>	<ul style="list-style-type: none"> <li>SSM has less robust utilization metrics and <b>no way to track carbon footprint</b></li> </ul>
<b>More ways to monitor</b> <i>OME vs. Supermicro SSM</i>	<ul style="list-style-type: none"> <li>Manage Dell servers from anywhere via the OpenManage mobile app</li> <li>Monitor third-party devices with OME using server IPs and credentials with support to import 3<sup>rd</sup> party SNMP MIBS</li> </ul>	<ul style="list-style-type: none"> <li>SSM has <b>no mobile app</b></li> <li>SSM <b>does not allow</b> monitoring of third-party devices with server IPs</li> </ul>



## Conclusion

When it comes to designing, implementing, managing, and maintaining AI solutions in your business, there are many factors to consider. To help you invest wisely and get the most out of your AI solution, you may wish to look for a vendor that can be more than a hardware supplier. Our research indicates that Dell offers services that can help you as a partner throughout the entire journey, so consider investing with Dell as you dive into AI.

- 
1. Vipera, "NVIDIA's H100 and A100 GPU Cards: Exploring the Intricacies of SXM and PCI-E Connections," accessed January 5, 2024, <https://www.viperatech.com/unraveling-the-mysteries-sxm-vs-pci-e-connections-in-nvidias-high-end-h100-and-a100-gpus/>.
  2. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results," accessed January 5, 2024, <https://mlcommons.org/en/inference-datacenter-31/>.
  3. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results."
  4. Dell, "PowerEdge XE Servers," accessed January 5, 2024, <https://www.dell.com/en-us/dt/servers/specialty-servers/poweredge-xe-servers.htm>.
  5. Supermicro, "Next Leap of AI Infrastructure is Here," accessed January 5, 2024, <https://www.supermicro.com/en/accelerators/nvidia>.
  6. Dell, "PowerEdge XE9680," accessed January 5, 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9680-spec-sheet.pdf>.
  7. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0069. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.
  8. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0135. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.
  9. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0069. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.
  10. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0132. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.

- 
11. Dell, "PowerEdge XE9640 Rack Server," accessed January 5, 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe9640>.
  12. Accelsius, "Enabling the AI Revolution with Liquid Cooling," accessed January 5, 2024, <https://www.accelsius.com/blog/enabling-the-ai-revolution-with-liquid-cooling>.
  13. Dell, "PowerEdge XE9640 Technical Guide," accessed January 5, 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9640-technical-guide.pdf>.
  14. Supermicro, "GPU Server Systems," accessed January 5, 2024, [https://www.supermicro.com/en/products/gpu?pro=pl\\_grp\\_type%3D1](https://www.supermicro.com/en/products/gpu?pro=pl_grp_type%3D1).
  15. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0067. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.
  16. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0133. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.
  17. Dell, "PowerEdge XE8640," accessed January 5, 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe8640-spec-sheet.pdf>.
  18. Dell, "PowerEdge XE8640 Rack Server," accessed January 5, 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe8640>.
  19. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0066. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.
  20. Verified MLPerf® score of v3.1 Inference Closed. Retrieved from <https://mlcommons.org/benchmarks/inference-datacenter/> 5 December 2023, entry 3.1-0133. The MLPerf name and logo are registered and unregistered trademarks of MLCommons Association in the United States and other countries. All rights reserved. Unauthorized use strictly prohibited. See [www.mlcommons.org](http://www.mlcommons.org) for more information.
  21. Dell, "Artificial Intelligence (AI) technologies powered by Dell Precision workstations," accessed January 5, 2024, <https://www.dell.com/en-us/dt/ai-technologies/index.htm?hve=explore+dell+precision+for+ai#tab0=0>.
  22. Supermicro, "Super Workstations," accessed January 5, 2024, <https://www.supermicro.com/en/products/superworkstation>.
  23. Supermicro, "Rackmount Workstations," accessed January 5, 2024, <https://www.supermicro.com/en/products/rackmount-workstations>.
  24. Stephen Pritchard, "Storage requirements for AI, ML and analytics in 2022," accessed January 5, 2024, <https://www.computerweekly.com/feature/Storage-requirements-for-AI-ML-and-analytics-in-2022>.
  25. Dell, "PowerScale AI-Ready Data Platform," accessed January 5, 2024, <https://www.dell.com/en-us/shop/powerscale-family/sf/powerscale>.
  26. Dell, "Dell PowerScale All-Flash," accessed January 5, 2024, <https://www.delltechnologies.com/asset/en-us/products/storage/technical-support/h15963-ss-powerscale-all-flash-nodes.pdf>.
  27. Dell, "Dell PowerScale OneFS Software Features," accessed January 5, 2024, <https://www.delltechnologies.com/asset/en-us/products/storage/technical-support/h18275-onefs-software-features-data-sheet.pdf>.
  28. Dell, "Dell ECS Enterprise Object Storage," accessed January 5, 2024, <https://www.dell.com/en-us/dt/storage/ecs/>.
  29. Supermicro, "Accelerating AI Data Pipelines," accessed January 5, 2024, <https://www.supermicro.com/en/products/storage>.
  30. Supermicro, "Supermicro Software-Defined Storage and Memory Solutions," accessed January 5, 2024, <https://www.supermicro.com/en/solutions/software-defined-storage>.
  31. Supermicro, "Supermicro WEKA Distributed Storage Solution," accessed January 5, 2024, <https://www.supermicro.com/en/solutions/wekaio>.
  32. Dell, "Dell ECS Enterprise Object Storage," accessed January 5, 2024, <https://www.dell.com/en-us/dt/storage/ecs/>.
  33. Dell, "Dell ECS Enterprise Object Storage."

- 
34. Supermicro, "Supermicro OSNEXUS Software-Defined Storage Solution," accessed January 5, 2024, <https://www.supermicro.com/en/solutions/osnexus>.
  35. ASBIS, "Supermicro Solution for Scalify RING," accessed January 5, 2024, <https://news.asbis.com/news/suppliers/supermicro-renewed-the-line-of-scality-ring-solution/>.
  36. Supermicro, "Supermicro solution for Quantum ActiveScale," accessed January 5, 2024, <https://www.supermicro.com/en/solutions/activescale>.
  37. Dell, "Scalable and elastic Storage as-a-Service," accessed January 5, 2024, <https://www.dell.com/en-us/dt/apex/storage/data-storage-services/>.
  38. Dell, "Flip the Switch to Open Networking with PowerSwitch," accessed January 5, 2024, <https://www.dell.com/en-us/dt/networking/>.
  39. Dell, "Dell PowerSwitch Data Center Switches," accessed January 5, 2024, <https://www.dell.com/en-us/dt/networking/data-center-switches/>.
  40. Supermicro, "SSE-T7132S - 400Gb Ethernet Switch," accessed January 5, 2024, <https://www.supermicro.com/en/products/accessories/Networking/SSE-T7132SR.php>.
  41. Dell, "Dell EMC Networking SmartFabric Services Deployment with VxRail 4.7—Fabric Design Center," accessed January 5, 2024, <https://infohub.delltechnologies.com/l/dell-emc-networking-smartfabric-services-deployment-with-vxrail-4-7-1/fabric-design-center-26/>.
  42. Dell, "Dell SmartFabric Services," accessed January 5, 2024, <https://www.dell.com/en-us/dt/networking/smartfabric/>.
  43. Penny Madsen, "Scaling Skills for AI: Lessons from Early Adopters," accessed January 5, 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/industry-market/idc-brief-importance-of-skills-for-ai-dell.pdf>.
  44. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Overview," accessed January 5, 2024, <https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/overview-5381/>.
  45. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Advisory Services for Generative AI," accessed January 5, 2024, <https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/advisory-services-for-generative-ai-1/>.
  46. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Adoption Services for Generative AI," accessed January 5, 2024, <https://infohub.delltechnologies.com/l/design-guide-generative-ai-in-the-enterprise-model-customization/adoption-services-for-generative-ai-1/>.
  47. Dell, "Design Guide—Generative AI in the Enterprise – Model Customization—Adoption Services for Generative AI."
  48. Dell, "Artificial Intelligence (AI) Ready Solution Services," accessed January 5, 2024, <https://www.dell.com/en-us/dt/services/solutions/artificial-intelligence-services.htm>.
  49. Dell, "Comprehensive AI Training Modules Tailored for You," accessed January 5, 2024, <https://education.dell.com/content/emc/en-us/home/training/aiml.html>.
  50. Supermicro, "Services and Support," accessed January 5, 2024, <https://www.supermicro.com/en/support>.
  51. Travis Vigil, "Dell and NVIDIA: Bringing Generative AI to the Enterprise," accessed January 5, 2024, <https://www.dell.com/en-us/blog/dell-and-nvidia-bringing-generative-ai-to-the-enterprise/>.
  52. Dell, "Dell Technologies and Hugging Face to Simplify Generative AI with On-Premises IT," accessed January 5, 2024, <https://www.dell.com/en-us/dt/corporate/newsroom/announcements/detailpage.press-releases~usa~2023~11~20231114-dell-technologies-and-hugging-face-to-simplify-generative-ai-with-on-premises-it.htm>.
  53. Richard DeMare, "Starburst and Dell expand partnership to accelerate AI efforts with more intelligent data collection," accessed January 5, 2024, <https://www.starburst.io/blog/starburst-and-dell-expand-partnership-to-accelerate-ai-efforts-with-more-intelligent-data-collection/>.
  54. Business Wire, "SiMa.ai and Supermicro Announce Partnership to Accelerate Power Efficient ML at the Edge," accessed January 5, 2024, <https://www.businesswire.com/news/home/20231129609794/en/SiMa.ai-and-Supermicro-Announce-Partnership-to-Accelerate-Power-Efficient-ML-at-the-Edge>.
  55. Dell, "Dell AI Solutions," accessed January 5, 2024, <https://www.dell.com/en-us/dt/solutions/artificial-intelligence/index.htm#accordion0&tab0=0>
  56. Dell, "Unlock the power of AI in virtualized environments," accessed January 5, 2024, <https://www.delltechnologies.com/asset/en-us/products/ready-solutions/briefs-summaries/ai-vxrail-powerscale-brief.pdf>.

- 
57. Dell, "Design Guide—Virtualizing GPUs for AI with VMware and NVIDIA Based on Dell Infrastructure—Performance Results," accessed January 5, 2024, <https://infohub.delltechnologies.com/l/design-guide-virtualizing-gpus-for-ai-with-vmware-and-nvidia-based-on-dell-infrastructure-1/performance-results-15/>.
  58. Dell, "Design Guide—Virtualizing GPUs for AI with VMware and NVIDIA Based on Dell Infrastructure—Design Considerations," accessed January 5, 2024, <https://infohub.delltechnologies.com/l/design-guide-virtualizing-gpus-for-ai-with-vmware-and-nvidia-based-on-dell-infrastructure-1/design-considerations-105/>.
  59. Supermicro, "Accelerate Every Workload," accessed January 5, 2024, <https://www.supermicro.com/en/solutions/ai-deep-learning>.
  60. Supermicro, "Supermicro Enterprise AI Inference & Training," accessed January 5, 2024, [https://www.supermicro.com/datasheet/Datasheet\\_AI-Workloads\\_Enterprise\\_AI\\_Inferencing\\_and\\_Training.pdf](https://www.supermicro.com/datasheet/Datasheet_AI-Workloads_Enterprise_AI_Inferencing_and_Training.pdf).
  61. Supermicro, "SUPERMICRO RACK SCALE SOLUTIONS: LARGE SCALE AI TRAINING WITH LIQUID COOLING," accessed January 5, 2024, [https://www.supermicro.com/solutions/Solution-Brief\\_Rack\\_Scale\\_AI.pdf](https://www.supermicro.com/solutions/Solution-Brief_Rack_Scale_AI.pdf).
  62. Supermicro, "Accelerate Every Workload," accessed January 5, 2024, <https://www.supermicro.com/en/solutions/ai-deep-learning>.
  63. Principled Technologies, "Dell management tools made server deployment and updates easier, offered more comprehensive security, and provided more robust infrastructure analytics," accessed January 5, 2024, <https://www.principledtechnologies.com/Dell/Management-tools-vs-Supermicro-0423.pdf>.

This project was commissioned by Dell Technologies.



**Facts matter.®**

Principled Technologies is a registered trademark of Principled Technologies, Inc. All other product names are the trademarks of their respective owners.

**DISCLAIMER OF WARRANTIES; LIMITATION OF LIABILITY:**

Principled Technologies, Inc. has made reasonable efforts to ensure the accuracy and validity of its testing, however, Principled Technologies, Inc. specifically disclaims any warranty, expressed or implied, relating to the test results and analysis, their accuracy, completeness or quality, including any implied warranty of fitness for any particular purpose. All persons or entities relying on the results of any testing do so at their own risk, and agree that Principled Technologies, Inc., its employees and its subcontractors shall have no liability whatsoever from any claim of loss or damage on account of any alleged error or defect in any testing procedure or result.

In no event shall Principled Technologies, Inc. be liable for indirect, special, incidental, or consequential damages in connection with its testing, even if advised of the possibility of such damages. In no event shall Principled Technologies, Inc.'s liability, including for direct damages, exceed the amounts paid in connection with Principled Technologies, Inc.'s testing. Customer's sole and exclusive remedies are as set forth herein.