

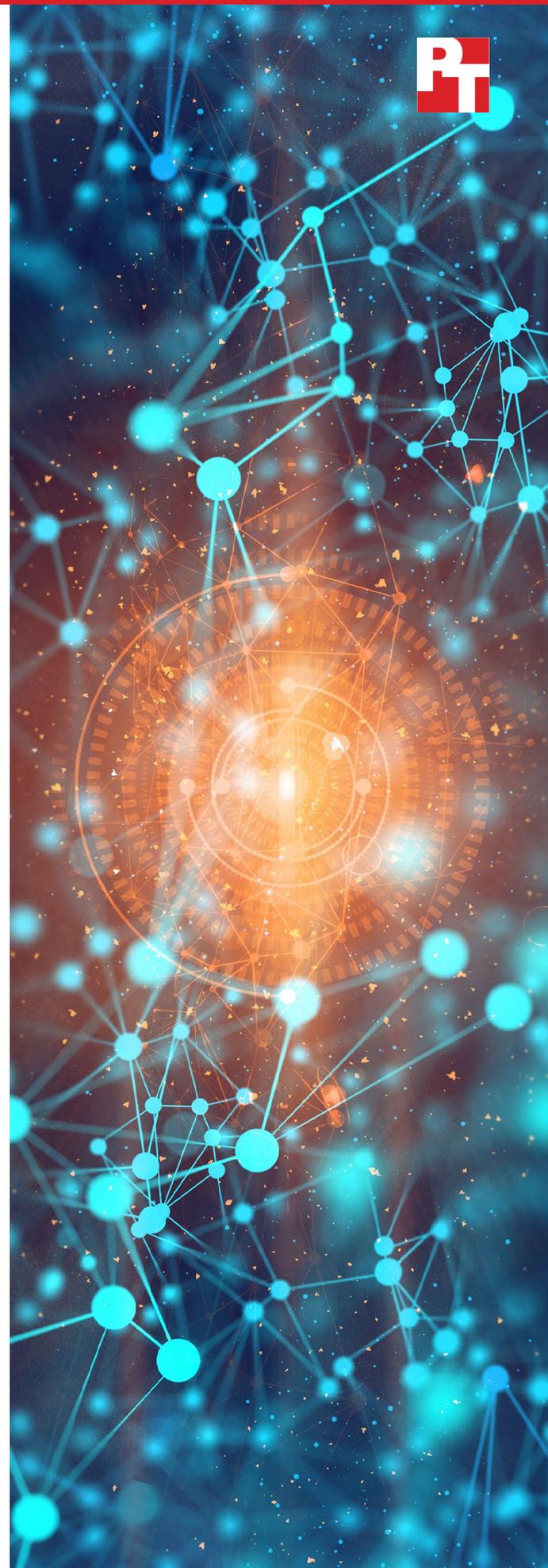


# Den Weg zum KI-Erfolg mit dem KI-Portfolio von Dell finden

## Ein Vergleich des Dell KI-Portfolios mit ähnlichen Angeboten von Supermicro

Künstliche Intelligenz (KI) ist der neue Maßstab, der den Geschäftsbetrieb branchenübergreifend neu gestalten wird. Während Unternehmen im gesamten Spektrum erkunden, wie sie KI nutzen können, um ihren Geschäftsbetrieb zu verbessern, ist vor allem ein Punkt wichtig: Die Implementierung von KI und die Nutzung ihrer Vorteile wird nicht über Nacht geschehen. Da jedes Unternehmen einzigartig ist, müssen die Daten und Geschäftsziele des Unternehmens bewertet werden, um zu ermitteln, wie die Nutzung von KI mit den jeweiligen Daten die spezifischen gewünschten Ergebnisse erzielen kann. Die Zusammenarbeit mit einem Unternehmen wie Dell mit einem umfassenden KI-Portfolio, einschließlich Planung, Datenvorbereitung, geeigneter Hardwareauswahl, KI-Modelldesign, Machbarkeitsstudien, Referenzarchitekturen und End-to-End-Support, kann zu erfolgreichen KI-Projekten führen.

Angesichts der unzähligen auf dem Markt verfügbaren Optionen könnte die Suche nach einem Partner, der bei all diesen Entscheidungen helfen kann, den Unterschied zwischen einer erfolgreichen KI-Implementierung und einem kostspieligen Fehler ausmachen. In diesem Whitepaper werfen wir einen genaueren Blick auf die KI-Portfolios von Dell und Supermicro, um LeserInnen über die Vorteile zu informieren, die Dell Kunden bei der Einführung von KI bieten kann. Wir konzentrieren uns zunächst auf Server- und Compute-Optionen, bei denen beide Unternehmen umfassende und vielfältige Angebote für Kunden bereithalten. Als Nächstes erkunden wir, wie Dell über reine Hardwareüberlegungen für Unternehmen hinausgeht, die sich Schulungen, Planungsservices, Partnernetzwerke und mehr wünschen.



# Server und Performanceergebnisse für KI-Workloads

Server – die grundlegende Computing-Infrastruktur, die KI-Workloads unterstützt – können CPUs, GPUs oder beides je nach Größe oder Art der Workload als Rechenressourcen verwenden. Für größere oder anspruchsvollere Workloads wie HPC oder KI bieten GPUs Spitzenleistung. GPUs sind in einer Vielzahl von Formfaktoren erhältlich, darunter universelle PCIe- und OAM-Systeme (Open Compute Project Accelerator Module) sowie die proprietäre NVIDIA SXM-Architektur, die derzeit höchste Performance bietet.<sup>1</sup> Große Arbeitsspeicherkapazitäten und Serverdesignfunktionen wie Kühlungsarchitektur und Energieeffizienz wirken sich ebenfalls auf die Performance aus. In den meisten Rechenzentren wird weiterhin Luftkühlung verwendet, was bedeutet, dass für KI-Workloads Server erforderlich sind, die so effektiv wie möglich mit Luft gekühlt werden. Im Folgenden werden wir die Dell PowerEdge-Serverangebote in Bezug auf Komponenten, Kühlungsoptionen und mehr sowie die veröffentlichten MLCommons® MLPerf®-Bewertungen vorstellen.

## Testergebnisse

MLPerf® ist eine Benchmarksuite, die die KI-Performance in Bezug auf Training und Inferenz testet. Damit ein Unternehmen offizielle MLPerf®-Ergebnisse veröffentlichen kann, müssen die Ergebnisse den spezifischen Bedingungen entsprechen, die vom Benchmarkentwickler MLCommons® festgelegt wurden.<sup>2</sup> Diese Compliancerichtlinien enthalten Standards, die den Leistungsvergleich vereinfachen. Für Inferenztests verwendet MLPerf® Datacenter-, Edge-, Mobile- und Tiny-Datenvolumen und meldet KI-Bewertungen und Watt des Stromverbrauchs während der Tests. Die Inferenz-Benchmark-Suite umfasst Tests für viele gängige KI-, ML- und DL-Modelle (siehe Tabelle 1).

Tabelle 1: KI-, ML- und DL-Modelle, die in MLPerf® enthalten sind, mit Tests und typischen Anwendungsfällen für jedes Modell. Quelle: Principled Technologies

Gängige KI-Modelle	Typisches Einsatzbeispiel
ResNet	Ein Bildklassifizierungsmodell, mit dem Computer verschiedene Bilder für Anwendungsfälle wie medizinische Bildgebung, Moderation von Social-Media-Inhalten und Gesichtserkennung lernen, sich merken und identifizieren können
RetinaNet	Eine Art der Objekterkennung, die mehr Komplexität als ResNet bewältigen kann. Sie hilft Computern, Objekte in Bildern oder Videoframes zu identifizieren und zu finden und sie nach ihrer Bedeutung zu klassifizieren. Einsatz in Bereichen wie autonomes Fahren, automatische Fahrzeugassistenzsysteme, Überwachung, Gesichtserkennung
3D-UNet	Spezifisch für die Segmentierung medizinischer Bilder
RNN-T	Spracherkennung für Anwendungsfälle wie automatisierte Sprachübersetzung
BERT	Natürliche Sprachverarbeitung für Anwendungsfälle wie Textzusammenfassung, Sprachübersetzung und automatische Fertigstellung von Aufgaben
DLRM-v2-99.9	Empfehlungsmodell für Anwendungsfälle wie gezielte Werbung und personalisierte Produktempfehlungen
GPTJ-99 und 99,9	LLM für die Verarbeitung natürlicher Sprache, die sich bei der Texterstellung für Anwendungsfälle wie Chatbots und chatbasierte KI-Tools auszeichnet



## Über MLPerf

MLPerf®-Ergebnisse umfassen neben den KI-Modellen auch mehrere Parameter, was dazu führen kann, dass in einem einzigen Diagramm oder einer Tabelle eine große Datenmenge analysiert werden muss. Hier finden Sie eine Kurzübersicht zu diesen Parametern:

- 99,0 und 99,9: Diese Zahlen beziehen sich auf die Genauigkeit, mit der das Modell trainiert wurde. Je genauer die Ausgabe sein muss, desto komplexer ist das Modell und desto länger kann die Verarbeitung der Daten dauern.
- Offline-Samples/Sek.: Modus, in dem die Benchmark alle Abfragen zu Beginn des Tests sendet, um Daten zu simulieren, die bereits auf dem System vorhanden sind.
- Serverabfragen/Sek.: Modus, in dem die Benchmark Abfragen während der Testdauer sendet, um die Analyse eines Livestreams von Daten zu simulieren.

Weitere Informationen zu MLCommons®- und MLPerf®-Ergebnissen finden Sie unter <https://mlcommons.org/benchmarks/inference-datacenter/>.

Die Ergebnisse in diesem Bericht stammen aus MLPerf® v3.1 Inference Datacenter-Ergebnissen, die im November 2023 auf der MLCommons®-Website veröffentlicht wurden.<sup>3</sup> Diese Ergebnisse umfassen Beiträge von Technologieherstellern und Cloud-Serviceanbietern und decken eine Reihe von Konfigurationen ab. Im Vergleich zu öffentlich verfügbaren Daten von Supermicro erzielten Dell PowerEdge-Server ähnliche Ergebnisse. Tabelle 2 enthält Details zu den Servern.

Tabelle 2: Dell und Supermicro-Server sind in den MLCommons® MLPerf® 3.1-Ergebnissen enthalten, die im November 2023 veröffentlicht wurden. Quelle: Principled Technologies

Anfragesteller	Servermodell	Nr. und Modell der GPUs	Beschreibung
Dell <sup>4</sup>	PowerEdge XE9680	8 x NVIDIA H100 SXM	Für KI-Training und Inferenz bei großen Workloads wie großen Sprachmodellen
	PowerEdge XE9640	4 x NVIDIA H100 SXM	Für das Training großer KI-Modelle in Rechenzentren mit hoher Dichte und Flüssigkeitskühlung
	PowerEdge XE8640	4 x NVIDIA H100 SXM	Für den Betrieb herkömmlicher KI-Trainings-, HPC- und Datenanalyseanwendungen in einem 4-HE-Formfaktor für luftgekühlte Rechenzentren
Supermicro <sup>5</sup>	AS-8125GS-TNHR	8 x NVIDIA H100 SXM	Für umfangreiche KI-Trainings und HPC-Workloads mit AMD-Prozessoren
	SYS-821GE-TNHR	8 x NVIDIA H100 SXM	Für umfangreiche KI-Trainings und HPC-Workloads mit Intel Prozessoren
	SYS-421GU-TNXR	4 x NVIDIA H100 SXM	Modulares Design mit Flexibilität zur Unterstützung von HPC- und KI-Workloads

Da Dell und Supermicro Ergebnisse mit sehr ähnlichen GPU-Konfigurationen vorgelegt haben, ist der Leistungsvergleich einfach. Wie in Abbildung 1 und 2 gezeigt, führt die Ähnlichkeit der Konfigurationen beider Anbieter zu Ergebnissen, die weitgehend vergleichbar sind. Bei anderen Konfigurationen wie denen in Abbildung 3 und 4 hat Dell Supermicro beim gptj-99.9-Modell in den 4-GPU-Tests übertroffen. Beachten Sie, dass Dell im Gegensatz zu Supermicro Ergebnisse für alle verfügbaren Modelle bei allen drei Servern übermittelt hat. Wir vergleichen nur die Modelle, bei denen für beide Server Ergebnisse vorhanden sind. Die vollständigen Ergebnisse von Dell finden Sie in den MLCommons® MLPerf®-Ergebnissen.

## Ergebnisse für Server mit acht GPUs

Der Dell PowerEdge XE9680 bietet Support für bis zu acht NVIDIA H100 SXM5-GPUs für KI-Beschleunigung und bis zu zwei skalierbare Intel® Xeon® Prozessoren der 4. Generation. Die PowerEdge XE-Produktreihe verfügt über eine modulare Architektur, die SXM4- oder SXM5-NVIDIA-GPUs oder OAM-GPU-Baugruppen (Open Compute Project Accelerator Module) unterstützt, wodurch die Leistung im Vergleich zu einer standardmäßigen PCIe-GPU gesteigert werden kann. Der Dell PowerEdge XE9680 bietet außerdem den Beschleuniger AMD Instinct™ MI300X.<sup>6</sup> Der PowerEdge XE9680 nimmt nur 6 HE Platz im Rack ein und ist ein kompakter 8-Wege-NVIDIA H100 SXM5-Server.

Im Gegensatz dazu benötigt der Supermicro-8-GPU-Server mit 8 HE 33 % mehr Platz im Rack, bietet aber ebenfalls den SXM-Formfaktor für NVIDIA-GPUs. Der Größenunterschied bedeutet, dass Sie 7 Dell PowerEdge XE9680-Server in einem Rack im Vergleich zu nur 5 Supermicro-Servern einsetzen können. In Abbildung 1 und 2 vergleichen wir die Ergebnisse des Dell PowerEdge XE9680 mit denen von zwei Konfigurationen des Supermicro-8-GPU-Servers: dem SYS-821GE-TNHR mit Intel Prozessoren und dem AS-8125GS-TNHR mit AMD-Prozessoren. Beachten Sie, dass Supermicro keine Ergebnisse für RNN-T auf dem SYS-821GE-TNHR übermittelt hat. Daher schließen wir dieses Modell aus dem Diagramm in Abbildung 1 aus.

### Normalisierte MLPerf®-Ergebnisse: Dell PowerEdge XE9680 mit H100 SXM5 im Vergleich zu Supermicro SYS-821GE-TNHR mit H100 *(größer ist besser)*

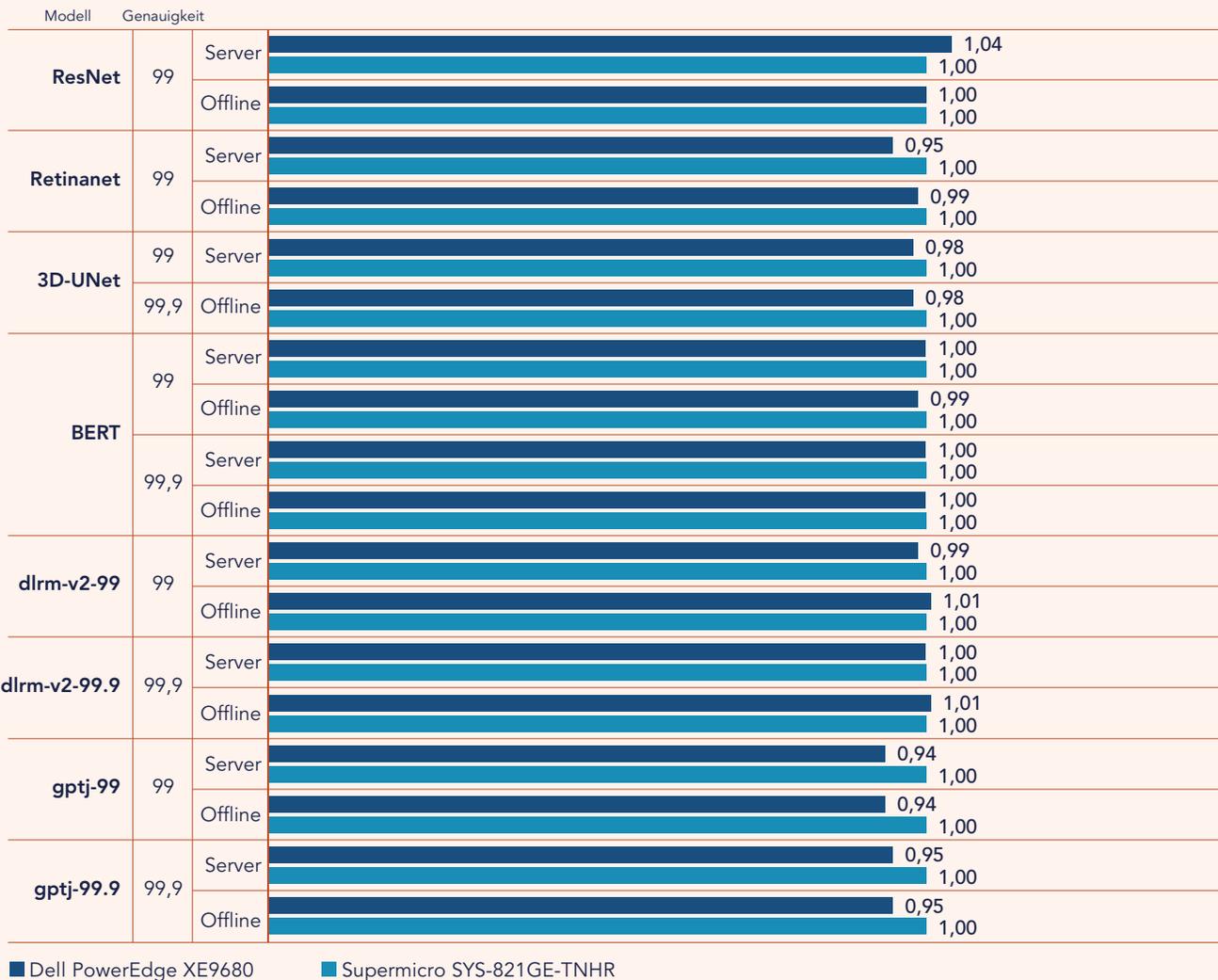


Abbildung 1: Veröffentlichte MLPerf®-Ergebnisse für Dell PowerEdge XE9680 und Supermicro SYS-821GE-TNHR zum 29.11.2023. Beide Systeme verwenden den SXM-Formfaktor der NVIDIA H100-GPUs. Quelle: Principled Technologies unter Verwendung von Daten von MLCommons®.<sup>7, 8</sup>



## Normalisierte MLPerf®-Ergebnisse: Dell PowerEdge XE9680 mit H100 SXM5 im Vergleich zu Supermicro AS-8125GS-TNHR mit H100 SXM5 *(größer ist besser)*

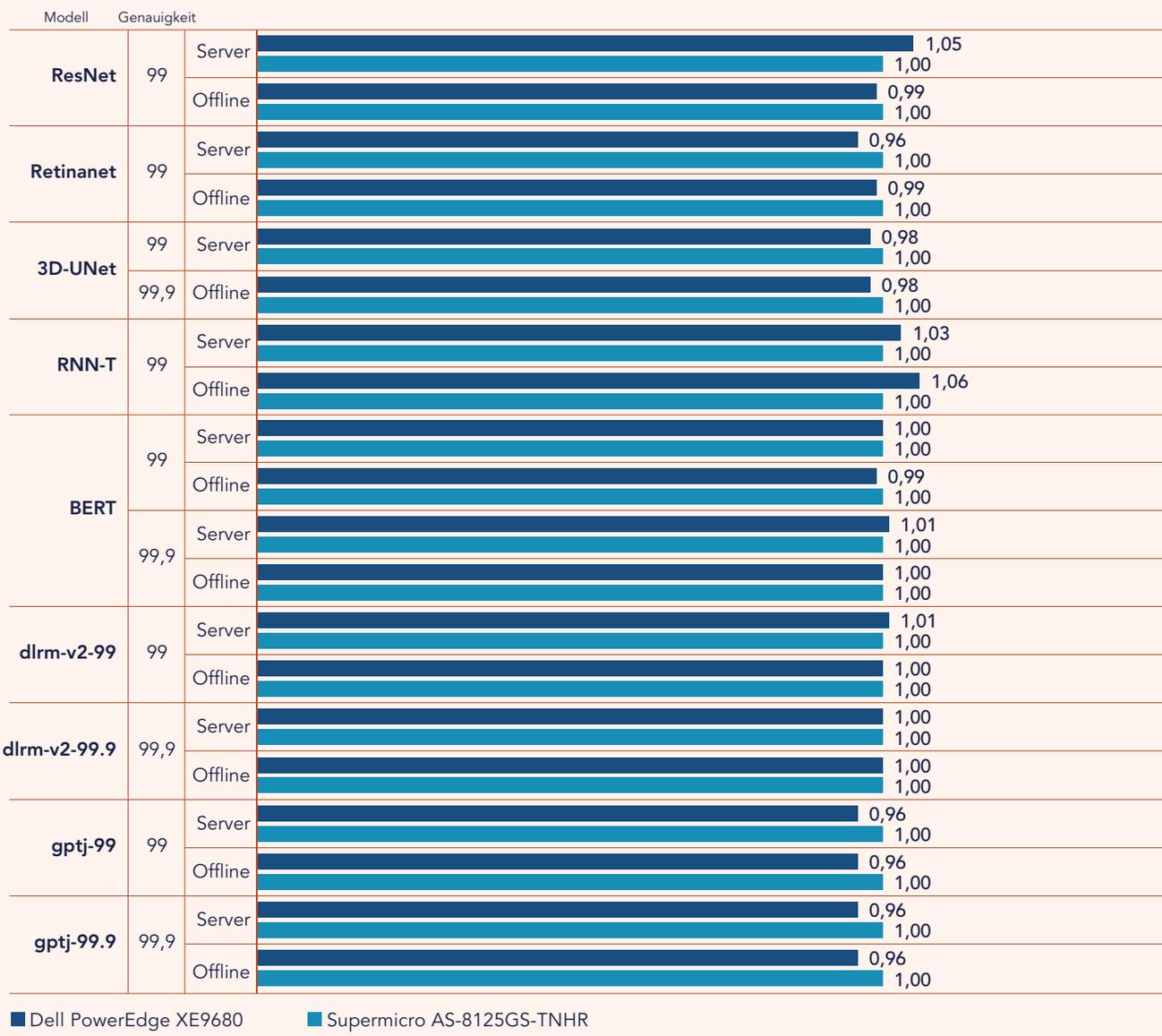


Abbildung 2: Veröffentlichte MLPerf®-Ergebnisse für Dell PowerEdge XE9680 und Supermicro AS-8125GS-TNHR zum 29.11.2023. Beide Systeme verwenden den SXM-Formfaktor der NVIDIA H100-GPUs. Quelle: Principled Technologies unter Verwendung von Daten von MLCommons®.<sup>9, 10</sup>

Wie diese Ergebnisse zeigen, bietet Ihnen die Wahl des Dell PowerEdge XE9680 eine ähnliche Performance bei einer Vielzahl von KI-Inferenz-Workloads und nimmt gleichzeitig weniger Platz im Rechenzentrum ein.



## Ergebnisse für Server mit vier GPUs

Wenn die Minimierung des Stromverbrauchs oder der Stellfläche im Rechenzentrum ein zentrales Anliegen ist, könnte der Dell PowerEdge XE9640-2-HE-Server die Antwort sein. Mit bis zu vier NVIDIA H100 SXM-GPUs bietet der PowerEdge XE9640 die Hälfte der GPU-Rechenleistung des PowerEdge XE9680 auf einem Drittel weniger Platz.<sup>11</sup> Der dicht gepackte Dell PowerEdge XE9640-Server ist mit der Dell Smart Cooling-Technologie ausgestattet, die eine Reihe von Wärmetechnologien einschließlich direkter Flüssigkeitskühlung für CPUs und GPUs bietet.<sup>12</sup> Das 2-HE-Gehäuse des PowerEdge XE9640 verfügt über verbesserte Luftstrommechanismen, einschließlich größerer Lüfter und Kühlkörper, um die anderen wichtigen Komponenten wie PCIe-Karten und Arbeitsspeicher zu kühlen.<sup>13</sup>

Supermicro bietet einen älteren SYS-220GQ-TNAR+-Server, der vier NVIDIA A100 HGX-GPUs in einem 2-HE-Formfaktor bereitstellt. Wir konnten jedoch keine 2-HE-Supermicro-Server mit vier der neueren H100 HGX-GPUs finden, die dem PowerEdge XE9640 entsprechen würden.<sup>14</sup> Der 4-GPU-Server mit HGX H100 NVIDIA-GPUs von Supermicro in den MLPerf®-Einreichungen ist der SYS-421GU-TNXR, ein 4-HE-Server. Wie bereits erwähnt, hat Supermicro MLPerf® 3.1-Ergebnisse nur für SYS-421GU-TNXR auf dem gptj-99.9-KI-Modell eingereicht. Daher können wir diesen nicht mit dem PowerEdge XE9640 auf den anderen Modellen vergleichen. In den veröffentlichten Ergebnissen hat der PowerEdge XE9640 den Supermicro-Server in den Offlinetests jedoch übertroffen und erzielte eine Bewertung von bis 1,37 (siehe Abbildung 3).

### Normalisierte MLPerf®-Ergebnisse: Dell PowerEdge XE9640 mit H100 SXM5 im Vergleich zu Supermicro SYS-421GU-TNXR mit H100 SXM5 (größer ist besser)

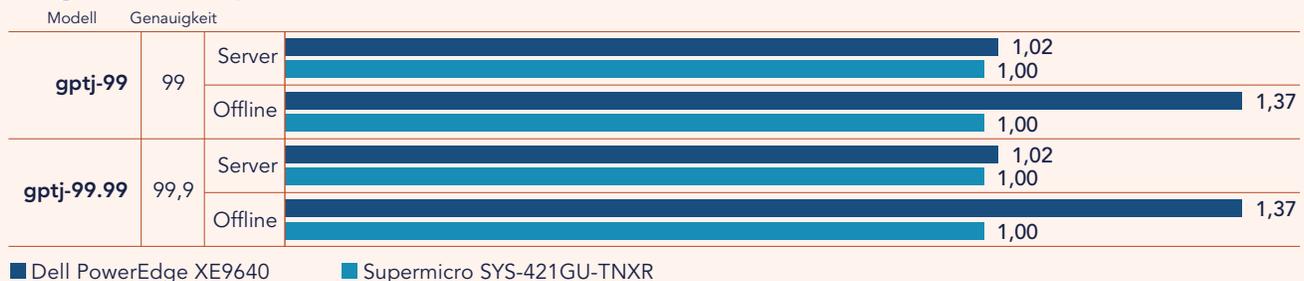


Abbildung 3: Veröffentlichte MLPerf®-Ergebnisse für Dell PowerEdge XE9640 und Supermicro SYS-421GU-TNXR zum 29.11.2023. Beide Systeme verwenden den SXM-Formfaktor der NVIDIA H100-GPUs. Quelle: Principled Technologies unter Verwendung von Daten von MLCommons®.<sup>15, 16</sup>

Wir können die Ergebnisse des Supermicro SYS-421GU-TNXR-Servers außerdem mit dem Dell PowerEdge XE8640 vergleichen, einem 4-HE-Server mit 4 GPUs, der auch NVIDIA H100 HGX-GPUs unterstützt. Der PowerEdge XE8640 ist zwar größer als der PowerEdge XE9640, benötigt jedoch keine direkte Flüssigkeitskühlung, was ihn zu einem Kompromiss zwischen Dichte und Kühltechnologien für Rechenzentren macht, in denen keine Wasserkühlung verfügbar ist. Der PowerEdge XE8640 bietet Luftkühlung für Prozessoren und einen flüssigkeitsgestützten Luftkühlungsradialator für die GPUs, der keine Wasserverfügbarkeit im Rack erfordert.<sup>17</sup> Der Dell PowerEdge XE8640 verfügt über die neuesten skalierbaren Intel Xeon Prozessoren der 4. Generation und bis zu 4 TB Arbeitsspeicher für die großen Datenvolumen und komplexen Berechnungen, die in KI und Datenanalysen üblich sind.<sup>18</sup> Mit seinem 4-HE-Formfaktor ähnelt der PowerEdge XE8640 dem Supermicro SYS-421GU-TNXR in Bezug sowohl auf Dichte als auch auf GPU-Funktionen. Wie wir jedoch schon beim PowerEdge XE9640 gesehen haben, erzielt auch der Dell PowerEdge XE8640 in den Offlinetests bessere gptj-99-Bewertungen als der Supermicro-Server (Abbildung 4).

## Normalisierte MLPerf®-Ergebnisse: Dell PowerEdge XE8640 mit H100 SXM5 im Vergleich zu Supermicro SYS-421GU-TNXR mit H100 SXM5 (größer ist besser)

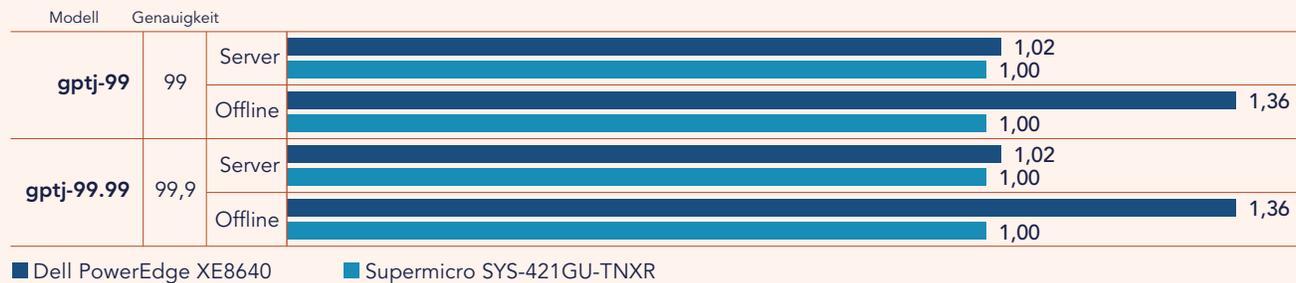


Abbildung 4: Veröffentlichte MLPerf®-Ergebnisse für Dell PowerEdge XE8640 und Supermicro SYS-421GU-TNXR zum 29.11.2023. Beide Systeme verwenden den SXM-Formfaktor der NVIDIA H100-GPUs. Quelle: Principled Technologies unter Verwendung von Daten von MLCommons®.<sup>19, 20</sup>

Wie wir gesehen haben, sind die MLPerf®-Performanceergebnisse der GPU-basierten Serverangebote von Supermicro und Dell insgesamt ähnlich, allerdings mit einem bemerkenswerten Vorteil für die Dell PowerEdge XE8640- und PowerEdge XE9640-Server auf einem KI-Modell.

Da Performance nur ein Aspekt der KI-Implementierung ist, haben wir uns auch die restlichen KI-Angebote von Dell und Supermicro angesehen, von Workstations, Storage und Netzwerken bis hin zu Services, Support, Schulung und mehr. Wir haben festgestellt, dass das Dell KI-Portfolio breiter ist als das Supermicro-Angebot und Lösungen für viele der Hindernisse bietet, die Unternehmen über die Compute-Performance hinaus überwinden müssen.

## Client-Workstations und Storage-Angebote

### Zusätzliche Compute-Optionen mit Workstations

Manche KI-Anwendungsfälle benötigen einen anderen Compute-Ansatz. Nicht alle, die KI-fähige Hardware benötigen, können an ein Rechenzentrum gebunden bleiben. WissenschaftlerInnen, die in Laboren arbeiten, haben möglicherweise keinen Platz für ein Rack mit Servern. Für diejenigen, die am Edge arbeiten, ist es kaum praktikabel, einen großen, schweren Desktop-PC herumzutragen. Hier kommen Workstations ins Spiel. Das KI-Portfolio von Dell umfasst mehrere KI-fähige Precision-Workstations – einschließlich Tower-, mobile und Rack-Konfigurationen –, die eine Vielzahl von Anforderungen erfüllen.<sup>21</sup> Im Gegensatz dazu bietet Supermicro keine mobilen Workstations für einfachen Zugriff unterwegs. Die GPU-Workstation-Angebote von Supermicro bestehen aus mehreren verschiedenen Tower-Konfigurationen, von denen NutzerInnen einige gemäß Angaben des Unternehmens in Racks montieren können.<sup>22</sup> Unsere Recherche ergab zwei im Rack montierbare Workstations, die jedoch älter zu sein scheinen, nur mit NVIDIA A100-GPUs verfügbar sind und möglicherweise nicht mehr verkauft werden.<sup>23</sup> Wenn Ihr Unternehmen Flexibilität in Bezug auf den Typ und die Mobilität der bereitgestellten GPU-Workstations benötigt, ist das Dell KI-Portfolio besser für diese Anforderungen geeignet.

### Überlegungen zum Speicher

Storage kann für die Ausführung von KI-Workloads genauso wichtig sein wie die Rechenleistung. Mehr Daten verbessern die Genauigkeit des KI-Modells, aber das Speichern und Managen riesiger Datenvolumen kann die Kapazitäten vieler Rechenzentren überfordern. Da Modelle in der Regel mit unstrukturierten Daten trainiert werden, müssen KI-fähige Storage-Systeme viele verschiedene Datentypen problemlos verarbeiten können.<sup>24</sup> Zur Bereitstellung von Kapazität und Skalierung für KI-, ML- und DL-Datensätze bietet Dell die PowerScale™-Serie für Datei-Storage und Elastic Cloud Storage (ECS) oder softwarebasiertes ObjectScale für Objektspeicher an.

Das Dell PowerScale-All-Flash-NAS-Portfolio bietet Kapazitätsoptionen von 3,84 TB bis zu 720 TB Rohkapazität pro Node mit geclusterten All-Flash-Kapazitäten von bis zu 186 PB Rohkapazität. Die Flexibilität und Skalierung von PowerScale kann eine Vielzahl von Kunden- und KI-Anwendungsfällen unterstützen.<sup>25</sup> Alle drei All-Flash PowerScale-Modelle – F200, F600 und F900 – beinhalten Inline-Datenkomprimierung und -Deduplizierung, um die Storage-Effizienz zu verbessern.<sup>26</sup> Jedes PowerScale-Storage-Modell verwendet das Dell OneFS™-Dateisystem, das erweiterte Tiering-Richtlinien nutzt, um sicherzustellen, dass sich die am häufigsten aufgerufenen Daten auf den Storage Tiers mit der höchsten Performance befinden.<sup>27</sup> Dell bietet außerdem OneFS-Software auf dem AWS Marketplace (Amazon Web Services) mit Dell APEX File Storage for AWS an. Kunden profitieren von einem konsistenten Nutzererlebnis, da sie OneFS mit ihren AWS-Compute-Instanzen mit denselben Funktionen nutzen können, die in On-Premise-OneFS-Arrays verfügbar sind.<sup>28</sup>

Supermicro-Storage-Angebote bestehen aus Storage-Servern – im Rack montierten Servern mit hoher Storage-Dichte in verschiedenen Größen und Dichten.<sup>29</sup> Um Datei-Storage von Supermicro zu erhalten, müssen Kunden aus verschiedenen softwarebasierten Storage-Angeboten von Drittanbietern wie WekaIO, Scality RING oder OSNEXUS wählen.<sup>30</sup> Während Scality RING und OSNEXUS Datei-Storage-Optionen als Teil ihrer Plattformbeschreibungen enthalten, scheint WekaIO die primäre Option für Kunden zu sein, die grundlegenden Datei-Storage suchen. Supermicro bietet mehrere Referenzarchitekturen für eine Vielzahl von Anwendungsfällen an. Kunden müssen jedoch über ein WekaIO-Softwareabonnement oder eine WekaIO-Lizenz verfügen, was die Gesamtkosten der Lösung erhöhen kann.<sup>31</sup>

Zu den Objektspeicheroptionen von Dell zählen Dell Enterprise Object Storage (ECS), der „speziell für die Speicherung unstrukturierter Daten in Public-Cloud-Skalierung entwickelt wurde“.<sup>32</sup> Neben der integrierten Kompatibilität mit Amazon S3-Objektspeicher für Hybrid-Cloud-Funktionen bieten ECS-Storage-Nodes Kapazitäten von bis zu 14 PB pro Rack.<sup>33</sup> Wie bei Datei-Storage sind auch bei den Supermicro-Objektspeicherangeboten Konfigurationen von Drittanbietern erforderlich. Die OSNEXUS-Plattform ist eine kombinierte Storage-Plattform für Datei-, Block- und Objektspeicher, während die Scality-RING-Lösung Datei- und Objektspeicher kombiniert. Für beide müssen Lizenzen bei Drittanbietern erworben werden.<sup>34,35</sup> Für reinen Objektspeicher können Kunden die Supermicro-Lösung für Quantum ActiveScale für Private-Cloud-Objektspeicher mit einem Quantum-Softwareabonnement erwerben.<sup>36</sup> Zum Zeitpunkt der Veröffentlichung dieses Artikels konnten wir keine flexiblen Optionen für die Nutzung/Betriebsausgaben von Supermicro finden.

Da Kunden bei den Supermicro-Storage-Angeboten mit Softwaredrittanbietern zusammenarbeiten müssen, fallen wahrscheinlich zusätzliche Lizenzierungs- oder Abonnementkosten an. Zudem kann es zu Problemen beim Support, beim Troubleshooting und mehr kommen. Das Dell KI-Portfolio bietet Dell Storage-Kunden eine einzige, zuverlässige Service- und Supportlösung für jeden Aspekt ihrer Storage-Lösung.

## Informationen zu Dell APEX

Für Kunden, die ihren Datei-Storage als As-a-Service-Lösung nutzen möchten, bietet Dell APEX Data Storage Services an, die Datei-, Block- und Backup-Storage umfassen. Mit der Dell APEX Console können Kunden neue Abonnements bestellen, die Storage-Kapazität anpassen und überwachen und vieles mehr. Laut Dell können Sie sich mit dieser Lösung „die Einfachheit und Flexibilität von Cloud-Umgebungen sichern und dabei gleichzeitig die volle Kontrolle über Ihre Anwendungen und Daten behalten“.<sup>37</sup>

Weitere Informationen über Dell APEX finden Sie unter <https://www.dell.com/en-us/dt/apex/storage/data-storage-services/index.htm>.



## Netzwerkoptionen

Netzwerke sind eine weitere wichtige Komponente der KI-Infrastruktur. Da viele KI-Workloads auf großen Serverclustern ausgeführt werden, die eine ständige Kommunikation untereinander und mit dem Storage erfordern, sind KI-Workloads auf robuste Netzwerke angewiesen, um Engpässe zu vermeiden. Wenn Ihr Netzwerk für die KI-Workload nicht ausreicht, erhöhen sich die Trainings- und Inferenzzeiten, wodurch sich die Datenverarbeitung und die Zeit bis zum Gewinnen von Erkenntnissen verzögern. Dell bietet ToR-Switches (Top-of-Rack) für PowerSwitch Data Center und PowerEdge MX-I/O-Module für Ethernet- und Fabric-Netzwerke.<sup>38</sup> Die PowerSwitch-Angebote reichen von 1 GbE bis 400 GbE und erfüllen so eine Reihe von Anforderungen. Darüber hinaus bieten die Switches der Dell PowerSwitch Z Serie 100-GbE- und 400-GbE-Verbindungen, die für Leaf-/Spine-Fabrics optimiert sind.<sup>39</sup>

Supermicro bietet ebenfalls Switches mit Ethernetanschlüssen bis zu 400 GbE für ToR und andere Anwendungen wie Data Center Spine und Leaf.<sup>40</sup> Dell Netzwerkservices bieten jedoch mehrere Vorteile in Bezug auf Benutzerfreundlichkeit und Flexibilität, die bei Supermicro fehlen. Services wie das Dell Fabric Design Center können dazu beitragen, Abweichungen, Lücken oder Ineffizienzen bei Netzwerken zu vermeiden und Kunden bei der Planung und Bereitstellung von Netzwerk-Fabrics mit Automatisierung zu unterstützen.<sup>41</sup> Für bestimmte Umgebungen wie VMware VxRail-, VMware ESXi- und Dell PowerStore-Konfigurationen bietet Dell SmartFabric-Services an, die eine softwarebasierte Lösung für Infrastrukturbereitstellung und Lebenszyklusmanagement ermöglichen. Mit PowerStore können SmartFabric Services „bis zu 99 % der LAN-Konnektivitätsaufgaben mit einer Plug-and-Play-Fabric automatisieren“.<sup>42</sup> Services wie diese, die Automatisierung, Anleitungen und mehr für das Design und die Implementierung von Netzwerken bieten, unterstützen Kunden beim KI-Einführungsprozess.

## Services, Schulungen und mehr

Die primäre Herausforderung bei der Implementierung von KI ist die Notwendigkeit von internem Fachwissen für Strategie, Planung, Datenvorbereitung und Management. Management und Wartung von KI-Workloads erfordern ein einzigartiges Wissen, das sowohl herkömmliches Hardwarefachwissen als auch Kenntnisse rund um maschinelles Lernen und Data Science umfasst. Die MitarbeiterInnen, die die KI-Strategie entwickeln und implementieren, benötigen außerdem ein umfassendes Verständnis der einzigartigen betrieblichen Ziele des Unternehmens, um sicherzustellen, dass die neuen KI-Workloads diesen Ziele entsprechen.<sup>43</sup>

Eine weitere wichtige Hürde kann die nahtlose Integration von KI in vorhandene Betriebssysteme sein. Diese Integration erfordert eine strategische Ausrichtung neuer KI-Technologien an den aktuellen Geschäftsprozessen, um sicherzustellen, dass die Einführung von KI etablierte Workflows intakt hält. Die Partnerschaft mit einem Unternehmen wie Dell, das mehrere optimierte, validierte Lösungsreferenzarchitekturen, Schulungskurse, Managementoptionen und ein großes Partnernetzwerk anbietet, kann Ihren Weg zur KI-Einführung erleichtern.

## Dienstleistungen für AI

Um Schulung und Planung abzudecken, bietet Dell eine Vielzahl von Services speziell für KI an.<sup>44</sup> Dell Services zur Unterstützung der KI-Implementierung umfassen Beratung, Datenvorbereitung, Bereitstellung, Support und Schulung, wobei jeweils spezifische Aspekte der KI-Einführung berücksichtigt werden. Dell Advisory Services for Generative AI unterstützen Kunden bei der Erarbeitung einer Roadmap, die Anwendungsfälle identifiziert und Unternehmen hilft, ihre Prozesse zu optimieren.<sup>45</sup> In ähnlicher Weise bieten Adoption Services for Generative AI Workshops mit Dell ExpertInnen, um Ihre Anforderungen und einzigartigen Herausforderungen zu überprüfen, ein vorab trainiertes Modell für Ihr Unternehmen zu ermitteln und Wissensübertragungssitzungen zur Schulung Ihrer IT-MitarbeiterInnen durchzuführen.<sup>46</sup> Dell bietet außerdem Implementation, Scale und Managed Services for Generative AI an, die ein unterschiedliches Maß an Support und Schulung bis hin zu einer vollständig gemanagten KI-Infrastruktur bieten, sodass Ihre IT-MitarbeiterInnen sich auf die Modelle und Daten konzentrieren können, während Dell die Hardware managt.<sup>47</sup> ProSupport Services sorgen für eine optimale Systemleistung und bieten grundlegende Hardware- und Softwareunterstützung für den laufenden KI-Betrieb zur Behebung technischer Probleme.<sup>48</sup>

Schulungsservices sind ein wesentlicher Bestandteil, um die erforderlichen Kompetenzen und Kenntnisse für die KI-Nutzung zu vermitteln. Dell Schulungsangebote beinhalten umfassende Schulungsprogramme im Bereich Data Science, Zertifizierungen für erweiterte Analysen und Workshops zu bestimmten KI-Technologien wie maschinellem Lernen.<sup>49</sup>

Supermicro-Services sind im Vergleich dazu meist auf Troubleshooting, Handbücher, RMAs (Return Merchandise Authorizations) und Service beschränkt.<sup>50</sup> Im Supermicro-KI-Portfolio konnten wir keine Design-, Implementierungs-, Management- oder Schulungsservices finden. Für Unternehmen, die einen Schulungspartner für die Bewältigung der Komplexität der KI-Einführung suchen, ist Dell klar die erste Wahl unter diesen beiden Anbietern.

## Partnerschaften mit Drittanbietern für KI-Workloads

Dell Technologies und NVIDIA bieten gemeinsam Dell Validated Designs an, die eine umfassende Lösung für generative KI in geschäftlichen Umgebungen bereitstellen sollen. Dieses Projekt schafft eine skalierbare, leistungsstarke Infrastruktur, die auf Technologien und Software von Dell und NVIDIA basiert, sowie ein KI-Modell-Framework, mit dem Unternehmen kundenspezifische KI-Modelle entwickeln und ausführen können. Mit der Lösung können Kunden GenAI-Workloads schnell in Betrieb nehmen.<sup>51</sup> Weitere Informationen finden Sie im Abschnitt „Dell Validated Designs“ unten.

Dell hat mit mehreren Unternehmen zusammengearbeitet, um KI-Technologieanwendungen zu optimieren. Mit Hugging Face erleichtert Dell die Einrichtung großer Sprachmodelle (Large Language Models, LLMs) vor Ort. Diese Partnerschaft kombiniert Hugging Face KI-Fachwissen mit Dell Servern und Storage-Systemen. Ein spezifisches Hugging Face-Portal von Dell bietet Tools für die einfache und sichere Bereitstellung von Hugging Face-Open-Source-KI-Modellen. Das fortlaufende Ziel besteht darin, diese Modelle für Dell Systeme weiter zu verbessern, die Leistung zu steigern und neue KI-Anwendungen zu unterstützen.<sup>52</sup>

Dell und Starburst arbeiten an einem leistungsfähigen, skalierbaren Data Lake, der Starburst-Analysen in Dell Computing- und Storage-Technologie integriert und einen zentralen Zugriff auf alle Datenquellen für KI- und ML-Tools bietet. Kunden können mithilfe dieser Partnerschaft Datensilos beseitigen.<sup>53</sup>

Laut unserer Studie sind die Partnerschaften von Supermicro im Bereich KI wesentlich begrenzter. SiMa.ai und Supermicro haben gemeinsam den Supermicro SYS-E300-13AD entwickelt, einen kompakten Edge ML-Server, der für die Verarbeitung von Multi-Stream-Videoanalysen entwickelt wurde. Dieser Server, der mit SiMa.ai ML-Pipeline auf einem Chip ausgestattet ist, verarbeitet effizient mehrere Videokanäle, reduziert die Gesamtbetriebskosten und verbessert die Ausfallsicherheit und Sicherheit. Der Server bietet eine Recheneinstellung für die Verarbeitung und Analyse zahlreicher Videostreams sowie Edge-Intelligenz, die für verschiedene Unternehmensanwendungen geeignet ist.<sup>54</sup>

## Dell Validated Designs

Um Ungewissheiten bei KI-Hardwarelösungen auszuräumen, bietet Dell laborvalidierte Referenzarchitekturen an, die für mehrere KI- und andere Workloads optimiert sind. Diese Validated Designs umfassen Architekturkonzepte, vollständige Lösungsübersichten sowie Performance- und andere Laborvalidierungen, die die Fähigkeiten der Lösung für die spezifische Workload belegen, für die sie entwickelt wurde. Zu diesen Workloads gehören virtualisierte Umgebungen, MLOps, maschinelles Lernen, dialogorientierte KI, GenAI-Inferenz, GenAI-Modelltuning, NVIDIA Fleet Command und OpenShift AI.<sup>55</sup>

Ein Beispiel: Das Validated Design „AI for Virtualized Environments“ kombiniert VMware-fähige KI mit NVIDIA AI Enterprise auf der Dell Infrastruktur und optimiert KI in virtuellen Umgebungen.<sup>56</sup> Der Validated Design Guide enthält Performanceergebnisse mit ResNet-Modelltrainings, die Kunden zeigen, dass das Design funktioniert und welche Performance sie erwarten können.<sup>57</sup> Diese Validierungen bieten Kunden einen Mehrwert, der über die einfache Auflistung von zusammenarbeitender Hardware hinausgeht, indem Konzepte erläutert, Konfigurationen empfohlen und Kunden durch Überlegungen und Performanceerwartungen geführt werden.<sup>58</sup>

Supermicro bietet ebenfalls Lösungen auf Anwendungsfallbasis an, die aber nicht bis zu dem Niveau der Referenzarchitekturen gehen wie bei den Dell Validated Designs. Anstatt eine speziell entwickelte Lösung für bestimmte Workloads zu empfehlen, organisiert Supermicro seine Server und GPUs in Kategorien wie KI-Inferenz und -Training, HPC/KI, Visualisierung und Design und mehr.<sup>59</sup> In den Broschüren und Datenblättern des Unternehmens bestehen diese Kategorien aus einigen Servern und GPUs, die sie als für die jeweilige Aufgabe empfehlen, mehreren Anwendungsfällen sowie Listen wichtiger Technologien und Softwarevorschläge.<sup>60</sup> Im Gegensatz zu den Dell Validated Designs scheinen sie keine Netzwerkarchitekturen und Performance- oder Validierungsdaten zu enthalten. Supermicro bietet außerdem mehrere Referenzdesigns, die eine ausführlichere Referenzarchitektur für einige KI-Lösungen bieten, z. B. ein umfangreiches KI-Training mit Flüssigkeitskühlung, das Supermicro mit NVIDIA veröffentlicht hat.<sup>61</sup> Kunden können möglicherweise eine ausführlichere Referenzarchitektur für ein bestimmtes Szenario finden, aber zum Zeitpunkt unserer Recherche haben wir nur drei gefunden: die bereits erwähnte Architektur mit Flüssigkeitskühlung, eine KI-Workstation-Architektur und eine RedHat OpenShift-Architektur.<sup>62</sup>

Insgesamt haben wir festgestellt, dass Dell Validated Designs mehr KI-Workloads abdecken und ausführlichere Anleitungen als die Supermicro-Angebote bieten.



## Managementservices und iDRAC

Laut einem Principled Technologies-Bericht vom April 2023 bietet der Integrated Dell Remote Access Controller (iDRAC) mehrere erweiterte Funktionen über die Supermicro Intelligent Platform Management Interface (IPMI) hinaus, insbesondere in den Bereichen Automatisierung, Sicherheit und Konfiguration.<sup>63</sup> Tabelle 3 enthält einen Vergleich der Managementfunktionen von Dell und Supermicro aus diesem Bericht, der zeigt, inwiefern iDRAC eine einfachere Bereitstellung, reibungslosere Firmwareupdates und mehr Sicherheitsfunktionen als Supermicro IPMI bieten könnte. Beachten Sie, dass sich einige Ergebnisse seit der ursprünglichen Veröffentlichung möglicherweise geändert haben.

Tabelle 3: Zusammenfassung des Principled Technologies-Vergleichs der Managementtools von Dell und Supermicro vom April 2023. Einige Ergebnisse haben sich seit der Veröffentlichung möglicherweise geändert. Quelle: Principled Technologies <https://facts.pt/V5fDf06>.

	Was ist der Unterschied bei den Managementtools von Dell	Wie viel besser
<b>Einfachere Firmwareupdates</b> <i>iDRAC9 im Vergleich zu Supermicro IPMI</i> <i>OME im Vergleich zu Supermicro SSM</i>	<ul style="list-style-type: none"> <li>Automatisierte Onlineupdates mit iDRAC9 mit Planungsoptionen</li> <li>OME ermöglicht die Erstellung kundenspezifischer Firmware-Repositories und kann die Firmware von BIOS, BMC und anderen Serverkomponenten ohne zusätzliche Tools oder Agents aktualisieren</li> </ul>	<ul style="list-style-type: none"> <li>Die Einrichtung automatischer Updates in iDRAC dauerte <b>nur 74 Sekunden</b>.</li> <li>Supermicro IPMI verfügt über <b>keine automatische Updatefunktion</b>, sodass AdministratorInnen manuell aktualisieren müssen.</li> <li>SSM unterstützt nur BIOS- und BMC-Firmwareupdates und erfordert SUM, um andere Komponenten zu aktualisieren.</li> </ul>
<b>Mehr Sicherheitsfunktionen</b> <i>iDRAC9 im Vergleich zu Supermicro IPMI</i> <i>OME im Vergleich zu Supermicro SSM</i>	<ul style="list-style-type: none"> <li>iDRAC9 bietet eine Multi-Faktor-Authentifizierung (MFA) und die dynamische Deaktivierung von USB-Anschlüssen, ohne Systemausfallzeiten zu verursachen.</li> <li>OME bietet sowohl rollenbasierte Zugriffskontrolle (Role-Based Access Control, RBAC) als auch bereichsbasierte Zugriffskontrolle (Scope-Based Access Control, SBAC), um das Gerätemanagement auf eine Teilmenge von Gerätegruppen zu begrenzen.</li> </ul>	<ul style="list-style-type: none"> <li>Supermicro IPMI bietet <b>keine MFA-Funktionen</b>.</li> <li>Supermicro IPMI erfordert einen Systemneustart und das Aufrufen der BIOS-Konfiguration, um USB-Anschlüsse zu deaktivieren.</li> <li>Supermicro SSM bietet RBAC, aber nicht die restriktivere SBAC.</li> </ul>
<b>Einfacheres Lebenszyklusmanagement</b> <i>OME im Vergleich zu Supermicro SSM</i>	<ul style="list-style-type: none"> <li>Vollständiges Lebenszyklusmanagement ohne Agents über OME zur Vereinfachung von Management und Monitoring</li> </ul>	<ul style="list-style-type: none"> <li>SSM benötigt den SuperDoctor5-Agent für detaillierte lokale Systemintegritätsmetriken und Supermicro Update Manager (SUM), um zusätzliche Komponenten zu aktualisieren.</li> </ul>
<b>Einfachere Serverbereitstellung</b> <i>iDRAC9 im Vergleich zu Supermicro IPMI</i>	<ul style="list-style-type: none"> <li>Importieren eines vollständigen Dell Serverprofils in nur 12 Schritten mit iDRAC9</li> <li>Robuste BIOS-Konfigurationsoptionen mit iDRAC9 mit 52 BIOS-Funktionen und Unterstützung für Komponentenkonfigurationen wie RAID NIC und iDRAC</li> </ul>	<ul style="list-style-type: none"> <li>Mit Supermicro IPMI konnten wir nur die IPMI-Konfiguration speichern und wiederherstellen, anstatt das gesamte Serverprofil.</li> <li>iDRAC9 verfügt über 52 BIOS-Funktionen, IPMI bietet <b>keine BIOS-Konfigurationsoptionen</b>.</li> </ul>
<b>Mehr Optionen für Reporting und Analysen</b> <i>iDRAC9 im Vergleich zu Supermicro IPMI</i> <i>OME im Vergleich zu Supermicro SSM</i>	<ul style="list-style-type: none"> <li>iDRAC9 bietet Telemetrie-Streaming, mit dem NutzerInnen Serverdaten einfach an Analysetools wie Splunk senden können.</li> <li>OME sendet Telemetriedaten direkt an CloudIQ, um das Monitoring zu vereinfachen.</li> </ul>	<ul style="list-style-type: none"> <li>IPMI bietet nur eine SYSLOG-Funktion, die AdministratorInnen verwenden können, um Meldungen für die Aggregation und eventuelle Analyse zu senden.</li> <li>SSM verfügt über keine cloudbasierte Managementlösung, die Dell CloudIQ entspricht.</li> </ul>
<b>Mehr Nachhaltigkeitsfunktionen</b> <i>OME im Vergleich zu Supermicro SSM</i>	<ul style="list-style-type: none"> <li>Mehr Metriken für das Monitoring in OME Power Manager, einschließlich Daten zum CO<sub>2</sub>-Fußabdruck</li> </ul>	<ul style="list-style-type: none"> <li>SSM verfügt über weniger robuste Auslastungsmetriken und bietet <b>keine Möglichkeit, den CO<sub>2</sub>-Fußabdruck nachzuverfolgen</b>.</li> </ul>
<b>Mehr Überwachungsmöglichkeiten</b> <i>OME im Vergleich zu Supermicro SSM</i>	<ul style="list-style-type: none"> <li>Dell Servermanagement von überall aus über die mobile OpenManage-App</li> <li>Monitoring von Drittanbietergeräten mit OME mithilfe von Server-IPs und Zugangsdaten mit Unterstützung für den Import von Drittanbieter-SNMP-MIBS</li> </ul>	<ul style="list-style-type: none"> <li>SSM hat <b>keine mobile App</b>.</li> <li>SSM <b>ermöglicht kein</b> Monitoring von Drittanbietergeräten mit Server-IPs.</li> </ul>



## Entscheidung

Wenn es um das Design, die Implementierung, das Management und die Wartung von KI-Lösungen in Ihrem Unternehmen geht, gibt es viele Faktoren, die zu berücksichtigen sind. Um klug zu investieren und das Beste aus Ihrer KI-Lösung herauszuholen, sollten Sie nach einem Anbieter suchen, der mehr als nur ein Hardwareanbieter sein kann. Unsere Recherche zeigt, dass Dell Services anbietet und Sie als Partner auf dem gesamten Weg unterstützt. Ziehen Sie daher in Betracht, bei Ihrer Einführung von KI in Dell zu investieren.

1. Vipera, „NVIDIA's H100 and A100 GPU Cards: Exploring the Intricacies of SXM and PCI-E Connections“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 2024, <https://mlcommons.org/en/inference-datacenter-31/>.
3. MLCommons, „MLPerf Inference: Datacenter Benchmark Suite Results“.
4. Dell, „PowerEdge XE Server“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/servers/specialty-servers/poweredge-xe-servers.htm>.
5. Supermicro, „Next Leap of AI Infrastructure is Here“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/accelerators/nvidia>.
6. Dell, „PowerEdge XE9680“, abgerufen am 5. Januar 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9680-spec-sheet.pdf>.
7. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0069. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).
8. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0135. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).
9. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0069. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).

- 
10. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0132. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).
  11. Dell, „PowerEdge XE9640 Rack Server“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe9640>.
  12. Accelsius, „Enabling the AI Revolution with Liquid Cooling“, abgerufen am 5. Januar 2024, <https://www.accelsius.com/blog/enabling-the-ai-revolution-with-liquid-cooling>.
  13. Dell, „PowerEdge XE9640 Technical Guide“, abgerufen am 5. Januar 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe9640-technical-guide.pdf>.
  14. Supermicro, „GPU Server Systems“, abgerufen am 5. Januar 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. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0067. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).
  16. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0133. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).
  17. Dell, „PowerEdge XE8640“, abgerufen am 5. Januar 2024, <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-xe8640-spec-sheet.pdf>.
  18. Dell, „PowerEdge XE8640 Rack Server“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/shop/ipovw/poweredge-xe8640>.
  19. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0066. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).
  20. Verifizierter MLPerf®-Score von v3.1 Inference Closed. Abgerufen von <https://mlcommons.org/benchmarks/inference-datacenter/> 5. Dezember 2023, Eintrag 3.1-0133. Der Name und das Logo von MLPerf sind eingetragene und nicht eingetragene Marken der MLCommons Association in den USA und anderen Ländern. Alle Rechte vorbehalten. Unerlaubte Nutzung ist strengstens untersagt. Weitere Informationen siehe [www.mlcommons.org](http://www.mlcommons.org).
  21. Dell, „Artificial Intelligence (AI) technologies powered by Dell Precision workstations“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/ai-technologies/index.htm?hve=explore+dell+precision+for+ai#tab0=0>.
  22. Supermicro, „Super Workstations“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/products/superworkstation>.
  23. Supermicro, „Rackmount Workstations“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/products/rackmount-workstations>.
  24. Stephen Pritchard, „Storage requirements for AI, ML and analytics in 2022“, abgerufen am 5. Januar 2024, <https://www.computerweekly.com/feature/Storage-requirements-for-AI-ML-and-analytics-in-2022>.
  25. Dell, „PowerScale AI-Ready Data Platform“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/shop/powerscale-family/sf/powerscale>.
  26. Dell, „Dell PowerScale All-Flash“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/storage/ecs/>.

- 
29. Supermicro, „Accelerating AI Data Pipelines“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/products/storage>.
  30. Supermicro, „Supermicro Software-Defined Storage and Memory Solutions“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/solutions/software-defined-storage>.
  31. Supermicro, „Supermicro WEKA Distributed Storage Solution“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/solutions/wekaio>.
  32. Dell, „Dell ECS Enterprise Object Storage“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/solutions/osnexus>.
  35. ASBIS, „Supermicro Solution for Scality RING“, abgerufen am 5. Januar 2024, <https://news.asbis.com/news/suppliers/supermicro-renewed-the-line-of-scality-ring-solution/>.
  36. Supermicro, „Supermicro Solution for Quantum ActiveScale“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/solutions/activescale>.
  37. Dell, „Scalable and Elastic Storage as a Service“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/apex/storage/data-storage-services/>.
  38. Dell, „Flip the Switch to Open Networking with PowerSwitch“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/networking/>.
  39. Dell, „Dell PowerSwitch Data Center Switches“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/networking/data-center-switches/>.
  40. Supermicro, „SSE-T7132S - 400Gb Ethernet Switch“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/networking/smartfabric/>.
  43. Penny Madsen, „Scaling Skills for AI: Lessons from Early Adopters“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 2024, <https://www.dell.com/en-us/dt/services/solutions/artificial-intelligence-services.htm>.
  49. Dell, „Comprehensive AI Training Modules Tailored for You“, abgerufen am 5. Januar 2024, <https://education.dell.com/content/emc/en-us/home/training/aiml.html>.
  50. Supermicro, „Services and Support“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/support>.
  51. Travis Vigil, „Dell and NVIDIA: Bringing Generative AI to the Enterprise“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 2024, <https://www.supermicro.com/en/solutions/ai-deep-learning>.
60. Supermicro, „Supermicro Enterprise AI Inference & Training“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 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“, abgerufen am 5. Januar 2024, <https://www.principledtechnologies.com/Dell/Management-tools-vs-Supermicro-0423.pdf>.

► Lesen Sie die Originalversion dieses Berichts in englischer Sprache unter <https://facts.pt/q9p46K9>.

Dieses Projekt wurde in Auftrag gegeben von Dell Technologies.



Facts matter.®

Principled Technologies ist eine eingetragene Marke von Principled Technologies, Inc. Alle anderen Produktnamen sind Marken der jeweiligen Inhaber.

**GEWÄHRLEISTUNGS-AUSSCHLUSS, HAFTUNGSEINSCHRÄNKUNG:**

Principled Technologies, Inc. hat angemessene Anstrengungen unternommen, die Genauigkeit und Richtigkeit der Tests sicherzustellen. Principled Technologies, Inc. schließt jedoch jegliche ausdrückliche und implizite Gewährleistung aus, die sich auf die Testergebnisse und Analysen, deren Genauigkeit, Vollständigkeit oder Qualität bezieht, einschließlich jeglichen impliziten Eignungsversprechens für einen bestimmten Zweck. Alle natürlichen oder juristischen Personen, die sich auf die Ergebnisse der Tests verlassen, tun dies auf eigenes Risiko und stimmen zu, dass Principled Technologies, Inc., seine Mitarbeiter und Auftragsnehmer keinerlei Haftung für Verlust oder Beschädigung jeglicher Art aufgrund vermeintlicher Fehler oder Mängel in einem Testverfahren oder Testergebnis übernehmen.

In keinem Fall haftet Principled Technologies, Inc. für indirekte, spezielle, zufällige oder Folgeschäden in Verbindung mit den Tests, auch wenn das Unternehmen auf die Möglichkeit solcher Schäden hingewiesen wurde. In keinem Fall geht die Haftung von Principled Technologies, Inc. über den in Verbindung mit den Tests von Principled Technologies, Inc. gezahlten Betrag hinaus. Dies gilt auch für direkte Schäden. Die alleinigen und ausschließlichen Rechtsmittel, die dem Kunden zur Verfügung stehen, sind die in diesem Dokument beschriebenen Rechtsmittel.