



Technische Studie



# Weltrekordleistung für KI und ML

Untersuchungen von Prowess zeigen den Zusammenhang zwischen der Benchmarkstärke von Dell™ Servern und der realen Leistung für Training und Inferenzierung von Modellen im Bereich künstliche Intelligenz (KI) und maschinelles Lernen (ML).

## Zusammenfassung

Unternehmen aus einer Vielzahl von Branchen sind auf künstliche Intelligenz (KI) und maschinelles Lernen (ML) angewiesen, um Verkäufe zu prognostizieren, ihre Kundenbasis zu segmentieren, Risiken zu identifizieren, komplexe Liefernetzwerke zu managen, Kosten zu optimieren und die Effizienz zu steigern. Allerdings sind alle KI- und ML-Anwendungsfälle, egal, wo und wie sie angewendet werden, von der Rechenleistung abhängig (zusätzlich zur Geschwindigkeit und Menge des Arbeitsspeichers sowie der Bandbreite von Verbindungen und Netzwerken). Und aufgrund der Art der Daten, die Unternehmen für Aktivitäten wie vorausschauende Kundenanalysen oder Betrugserkennung verwenden, ist die Sicherheit häufig genauso wichtig.

Aufgrund der Leistungsanforderungen von KI und ML kann die Infrastruktur (Computing, lokaler Arbeitsspeicher, Netzwerkbandbreite und Datenspeicher) zur Unterstützung dieser Workloads eine erhebliche Investition darstellen, was eine gründliche Bewertung vor dem Kauf notwendig macht. Branchenstandard-Benchmarks können hierfür gut geeignet sein – und Weltrekorde sogar noch besser – wenn sie auf die richtige Art und Weise evaluiert werden.

Unternehmen müssen ihre Daten schnell aufnehmen, speichern und verarbeiten. Diese Benchmarktests können Erkenntnisse darüber liefern, wie schnell Daten erfasst, verarbeitet und nach der Speicherung abgerufen werden können. Um den Zusammenhang zwischen einer hohen Benchmarkleistung und dem potenziellen geschäftlichen Nutzen in der realen Welt zu untersuchen, hat sich Prowess Consulting eingehender damit beschäftigt, was starke Ergebnisse bei Branchenbenchmarks für Unternehmen bedeuten, die Weltrekordserver einsetzen wollen. Aufgrund des großen Marktanteils und der Anzahl der Weltrekorde, die Dell Technologies bei KI- und ML-Szenarien hält, nachgewiesen durch branchenweit anerkannte Benchmarks, hat sich Prowess speziell die Dell™ PowerEdge™-Server angesehen.

Unter den Dell PowerEdge-Servern der neuesten Generation haben wir drei ermittelt, die kürzlich Weltrekorde bei einer Benchmark aufgestellt haben, die speziell entwickelt wurde, um reale Leistungsdaten zu KI- und ML-Workloads zu liefern:

- **Dell PowerEdge R6625-Server (Weltrekord, TPC Express Benchmark™ AI [TPCx-AI])**
- **Dell PowerEdge R7615-Server (Weltrekord, TPCx-AI)**
- **Dell PowerEdge R7515-Server (Weltrekord, TPC Express Benchmark™ IoT [TPCx-IoT])**

Durch die Leistungsoptimierungen von Dell Technologies für KI- und ML-Workloads haben wir außerdem festgestellt, dass AMD EPYC™-Prozessoren und Broadcom®-Netzwerkkarten dazu beitragen können, die Big Data- und Analyseleistung sowohl bei Benchmarks als auch bei realen Anwendungen zu steigern. Beispielsweise ermöglicht die Unterstützung der AMD EPYC-Prozessoren der 4. Generation für das numerische bfloat16-Format die Ausführung größerer KI-Modelle mit größeren Datensets, während die Unterstützung für INT8-Zahlen die Inferenzierung auf KI-Modellen beschleunigen kann. Darüber hinaus ermöglichen zwei auf NVMe Express® (NVMe®) basierende Dell™ PowerEdge™-RAID-Controller-Karten (Dell™ PERC) und die Unterstützung für PCIe® 4.0 durch Broadcom-Netzwerkkarten 100-Gigabit-Ethernet(GbE)-Netzwerkschnittstellen-Controller (NICs) mit zwei Ports und tragen dazu bei, Bandbreiteneinschränkungen zu beseitigen und die Leistung von KI- und ML-Workloads weiter zu beschleunigen.

In dieser Studie werden die folgenden Themen behandelt:

- **[Branchenlandschaft](#)**
- **[Forschungsmethodik von Prowess](#)**
- **[KI- und ML-Benchmarks](#)**
- **[Hinter den Leistungsergebnissen](#)**



## Branchenlandschaft: KI und ML

KI und ML haben im Geschäftsbetrieb von Unternehmen in einer Vielzahl von Branchen inzwischen eine zentrale Rolle eingenommen. Unabhängig davon, ob es sich um Betrugsüberwachung durch Banken, Verkaufsprognosen bei Einzelhändlern, präzisere Diagnosen in Krankenhäusern oder vorausschauende Wartung von Fließbändern bei mittelständischen Herstellern handelt, Unternehmen jeder Art und Größe verlassen sich auf KI, um Muster zu erkennen, die für Menschen möglicherweise unsichtbar sind. KI-Modelle – und insbesondere Deep-Learning-Modelle (DL) – sind in der Regel präziser, je mehr Daten zum Trainieren zur Verfügung stehen. Dieser Datenhunger erfordert größeren und leistungsfähigeren Storage, schnellere Netzwerke und leistungsstärkere Server, um einen höheren Mehrwert aus den Daten abzuschöpfen – Daten, die ebenfalls sicher aufbewahrt werden müssen.

Unternehmen nutzen aus vielerlei Gründen häufig lokale Server und keine Cloud-Implementierungen. Bei KI und ML haben diese Gründe häufig mit der Datengravitation und der niedrigeren Latenz bei der Verwendung von KI-Modellen zu tun. Es ist oft schneller und einfacher, KI-Trainingsfunktionen näher an den Daten durchzuführen, anstatt große Datenmengen kosten- und zeitaufwendig auf zentralisierte Compute-Ressourcen zu verschieben. Die Verarbeitung von Daten in der Nähe des jeweiligen Standorts kann auch die Latenz für das KI-Training reduzieren, was wiederum den gesamten Prozess beschleunigt. Zudem können behördliche Auflagen und Datenhoheitsgesetze, je nach Branche und Standort eines Unternehmens, es notwendig machen, die Daten vor Ort aufzubewahren. In allen Fällen ist die Leistung eine zentrale Anforderung von Unternehmen für die Verarbeitung und Analyse von Daten durch KI und ML.

Die Leistungsanforderungen von Workloads wie Analysen bedeuten, dass die Dateninfrastruktur optimiert werden muss, um Service-Level Agreements (SLAs) zu erfüllen. Das Zusammenspiel von Prozessor, Arbeitsspeichergröße, Netzwerkbandbreite und Storage-Subsystemen ist dabei von entscheidender Bedeutung. Benchmarkergebnisse sind ein häufig verwendetes Tool, um die Serverleistung aus diesem Zusammenspiel zu vergleichen. Da Benchmarks zahlenbasierte Ergebnisse liefern, erscheinen Vergleiche zwischen konkurrierenden Systemen relativ unkompliziert.

Und genau aus dem Grund, dass Benchmarks klare und scheinbar objektive Ergebnisse liefern, ist es von entscheidender Bedeutung, zu verstehen, was genau gemessen wird – und somit auch, was die Ergebnisse tatsächlich über die Serverplattformen aussagen. Unternehmen, die die Nuancen in den Benchmarks ignorieren und sich blind an den Top-Benchmark-Performern orientieren, erleben eventuell eine Enttäuschung, wenn ihr Return on Investment (ROI) nicht den Erwartungen entspricht.



## Forschungsmethodik von Prowess

Um den Zusammenhang zwischen einer hohen Benchmarkleistung und dem potenziellen geschäftlichen Nutzen in der realen Welt zu untersuchen, hat sich Prowess Consulting eingehender damit beschäftigt, was starke Ergebnisse bei Branchenbenchmarks für Unternehmen bedeuten, die Weltrekordserver einsetzen wollen. Um unsere Untersuchungen zu vereinfachen und uns darauf zu konzentrieren, wie einzelne Benchmarks Einblicke in die Leistung für bestimmte Facetten von KI und ML liefern können, haben wir uns speziell mit den Dell PowerEdge-Servern beschäftigt. Dies liegt sowohl am großen Marktanteil von Dell Technologies Servern als auch an der Anzahl der Weltrekorde, die Dell Technologies bei einer Vielzahl von KI- und ML-Benchmarks hält.

Ein einziger Benchmark-Weltrekord ist schon beeindruckend, aber was bei KI- und ML-Workloads heraussticht, ist die Tatsache, dass Dell Plattformen Weltrekorde über mehrere Benchmarks hinweg erzielen. Jede Benchmark kann als Teil des Workload-Puzzles betrachtet werden und das Erzielen mehrerer Weltrekorde bietet einen guten Einblick darin, wie sich Dell Plattformen in realen Umgebungen verhalten werden.

Wir haben uns die Benchmarkergebnisse angesehen, um zu ermitteln, welche Plattformen die beste Leistung für verschiedene Aspekte von KI- und ML-Workloads bieten. Bei unseren Untersuchungen haben wir uns auf die rackmontierten Server von Dell™ konzentriert. Dell Technologies verfügt weltweit über den größten Marktanteil bei Servern (17,2 Prozent)<sup>1</sup> und Dell PowerEdge-Server sind beliebte „Arbeitstier“-Server, die für standardmäßige bis mittlere Workload-Anforderungen entwickelt wurden. Für diese Studie haben wir speziell rackmontierte 1-HE-Server (Dell PowerEdge R6625-Server) und 2-HE-Server (Dell PowerEdge R7615- und PowerEdge R7515-Server) untersucht.

Um die Benchmarkergebnisse zu evaluieren, müssen diese aus der Perspektive der wichtigsten Leistungsfaktoren betrachtet werden. Für KI- und ML-Workloads gehören dazu:

- Performance
- Bestes Preis-Leistungs-Verhältnis

Dell Technologies hat seine PowerEdge-Plattform, basierend auf AMD®-Prozessoren, für KI- und ML-Lösungen optimiert. KI- und ML-Workloads haben spezifische Anforderungen und PowerEdge-Server ermöglichen Unternehmen ein hohes Maß an Flexibilität, um ihre individuellen Workloads gemäß ihren spezifischen Anforderungen auszuführen. Darüber hinaus hat Dell Technologies die Leistung seiner PowerEdge-Plattformen weiter optimiert, indem neuere Dell PERC-Karten und Broadcom NVMe-Netzwerkadapter integriert wurden, die für erhebliche Bandbreitensteigerungen sorgen.

## KI- und ML-Benchmarks

Branchenweit anerkannte Benchmarks können Einblicke in die gängigen Nutzungsszenarien einer Serverplattform liefern und Unternehmen dabei helfen, herauszufinden, ob die jeweilige Plattform die Anforderungen der auszuführenden Workloads erfüllt. Für diese Studie haben wir uns auf die Benchmarks konzentriert, die für KI- und ML-Workloads am relevantesten sind.

**Hinweis:** Die MLPerf™ Inference Benchmark Suite wird häufig verwendet, um die ML-Leistung zu messen. Prowess hat sich jedoch entschieden, MLPerf für diese Analyse nicht zu verwenden, da es bei dieser Benchmark um die Leistung des Grafikprozessors (GPU) geht, während sich unsere Studie auf die kosteneffiziente CPU-Leistung für die KI-Inferenzierung konzentriert.

TPC Express Benchmark™ AI (TPCx-AI)	Misst die durchgängige Leistung von branchenrepräsentativen KI- und ML-Workloads
TPC Express Benchmark™ IoT (TPCx-IoT)	Misst die Leistung, das Preis-Leistungs-Verhältnis und die Verfügbarkeit von Systemen, die enorme Datenmengen von einer Vielzahl an Geräten aufnehmen

### TPC Express Benchmark™ AI (TPCx-AI)

TPCx-AI misst die Leistung einer durchgängigen ML- bzw. Data-Science-Plattform. Die Benchmark emuliert das Verhalten repräsentativer KI-Branchenlösungen, die in Produktionsrechenzentren und Cloud-Umgebungen eingesetzt werden.

TPCx-AI bewertet die KI-Leistung anhand einer Reihe von Anwendungsfällen. Die Anwendungsfälle beziehen sich dabei auf einzelne Probleme, die von der DL- und ML-Data-Science-Pipeline im Rahmen der TPCx-AI-Benchmark gelöst werden. Die Pipeline ist Framework- und Syntax-unabhängig und kann auf viele Arten implementiert werden. Zu den Anwendungsfällen von TPCx-AI gehören in der Regel die Datenerzeugungs-, Datenmanagement-, Trainings-, Bewertungs- und Bereitstellungsphasen.

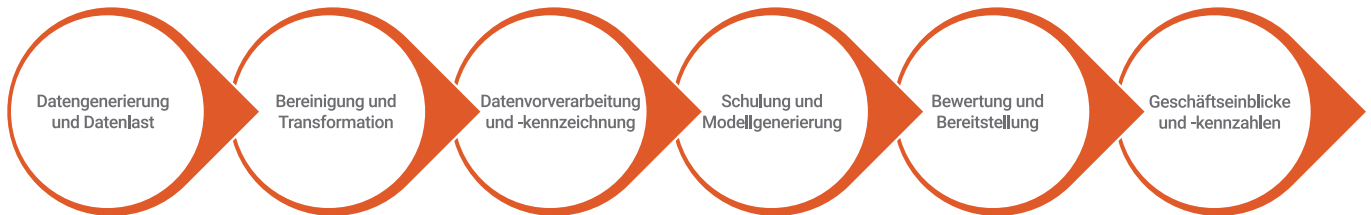


Abbildung 1: Die DL- und ML-Data-Science-Pipeline der TPC Express Benchmark™ AI (TPCx-AI)

Weitere Anwendungsfälle von TPCx-AI sind Kundensegmentierung, Kundengesprächstranskription, Umsatzprognosen, Spamerkennung, Preisprognosen, Klassifizierung und Betrugserkennung.

Dell Technologies hält bei der **TPCx-AI**-Benchmark vier Weltrekorde für Server mit AMD EPYC-Prozessoren der 4. Generation:<sup>2</sup>

- Dell PowerEdge R7615-Server (bei Skalierungsfaktor 3)
- Dell PowerEdge R7615-Server (bei Skalierungsfaktor 10)
- Dell PowerEdge R6625-Server (bei Skalierungsfaktor 30)
- Dell PowerEdge R6625-Server (bei Skalierungsfaktor 100)

Die PowerEdge R7615-Server erzielten einen KI-Pipeline-Durchsatz von 408,36 GB/Sekunde bei Skalierungsfaktor 3 (3 GB an Daten) und 425,31 GB/Sekunde bei Skalierungsfaktor 10 (10 GB an Daten) mit einem Preis-Leistungs-Verhältnis von 118,56 USD/GB/Sekunde bei Skalierungsfaktor 3.<sup>2</sup>

Der PowerEdge R6625-Server erreichte 365,59 GB/Sekunde bei Skalierungsfaktor 30 mit einem Preis-Leistungs-Verhältnis von 196,38 USD/GB/Sekunde, während ein Cluster mit vier PowerEdge R6625-Servern bei Skalierungsfaktor 100 868,49 GB/Sekunde mit einem Preis-Leistungs-Verhältnis von 356,56 USD/GB/Sekunde erreichte.<sup>2</sup>

Diese Server werden von AMD EPYC 9374F-, EPYC 9174F- und EPYC 9354-Prozessoren der 4. Generation unterstützt. Neben einer höheren Core- und Thread-Anzahl als die vorherige Generation bieten AMD EPYC-Prozessoren der 4. Generation auch DDR5-Speicherunterstützung für mehr Arbeitsspeicher, PCIe Gen 5-Unterstützung für einen höheren Datendurchsatz bei Servern und Serverclustern, Unterstützung für die neuesten Advanced Vector Extensions 512 (AVX-512) für einen schnelleren Datendurchsatz im Prozessorregister und Unterstützung für effiziente numerische Datentypen (z. B. bfloat16 und INT8), um das Trainieren und Inferenzieren von Modellen zu beschleunigen. All diese Funktionen haben in jeder Phase der Data-Science-Pipeline der TPCx-AI zu den Ergebnissen der Dell Server beigetragen.

### IoT (Internet der Dinge): TPC Express Benchmark™ IoT (TPCx-IoT)

Das Internet der Dinge (IoT) bietet eine riesige Datenquelle für KI und ML, insbesondere für die Echtzeitinferenzierung von KI-Modellen, z. B. für die Qualitätskontrolle bei Fertigungsstrecken. Das IoT stellt auch eine wichtige Quelle an Trainingsdaten für KI-Modelle bereit. Dies gilt für das anfängliche Training oder neue Modelle, aber auch für das Neutrainieren vorhandener Modelle, die Abweichungen aufweisen.

Obwohl dies nicht Teil der Data-Science-Pipeline der TPCx-AI-Benchmark ist, ist das kontinuierliche Trainieren von KI-Modellen in der Produktion unerlässlich. Im Laufe der Zeit kann sich die Zusammensetzung der Daten in Produktionsumgebungen wie dem Einzelhandel, Finanz- oder Gesundheitswesen auf subtile Weise verändern. Dadurch kann es dazu kommen, dass die Genauigkeit von KI-Modellen kontinuierlich abnimmt. Die effiziente Erfassung aktualisierter Trainingsdaten, um die Abweichungen von KI-Modellen zu korrigieren, ist ein wichtiger Bestandteil jedes KI- oder ML-Workflows.

Was das IoT von anderen Arten unstrukturierter Daten unterscheidet, ist der Umstand, dass IoT-Geräte ihre Daten häufig aus ihrer Umgebung erzeugen, z. B. wenn intelligente Lautsprecher Befehle erhalten oder industrielle Drohnen Landschaftsdaten erfassen. Aufgrund ihres schieren Volumens, das in der Regel größer ist als das anderer Datenquellen wie Einzelhandelstransaktionen oder Gesundheitsakten, müssen IoT-Daten für KI- und ML-Workloads daher effizient erfasst und verwendet werden.

Die TPCx-IoT-Benchmark ermöglicht einen direkten Vergleich verschiedener Software- und Hardwarelösungen für IoT-Gateways. Da sie sich zwischen der Edge-Architektur und dem Back-end-Rechenzentrum befinden, führen Gateway-Systeme Funktionen wie Datenaggregation für Echtzeit-KI und -ML aus. Die TPCx-IoT-Benchmark wurde speziell entwickelt, um überprüfbare Leistungs-, Preis-Leistungs- und Verfügbarkeitskennzahlen für kommerzielle Systeme bereitzustellen, die in der Regel enorme Datenmengen von einer großen Anzahl an Geräten aufnehmen.

Dell Technologies hält bei der **TPCx-IoT**-Benchmark einen Weltrekord für Dell PowerEdge 7515-Server mit AMD EPYC-Prozessoren der 3. Generation. Das Cluster erreichte einen Durchsatz von 1.617.545.000 Datensätzen/Sekunde bei einem Preis-Leistungs-Verhältnis von 329,75 USD/Millionen Datensätze/Sekunde.<sup>3</sup>

Zusätzlich zur Verarbeitungsleistung des AMD EPYC 75F3-Prozessors der 3. Generation zeigen unsere Analysen der Benchmarkergebnisse, dass die Leistung der Dell Server durch die im Cluster verwendeten Broadcom 25-GbE-Netzwerkkarten weiter gesteigert wurde. All diese Funktionen haben zur Leistung der Dell Server bei der TPCx-IoT-Benchmark beigetragen. Die Ergebnisse von Dell Technologies sind dabei besonders für Unternehmen relevant, die – wie der Dell Technologies Rekordhalter – Cloudera® verwenden, um große Datenmengen für Analysen und KI am Edge zu speichern.

## Hinter den Leistungsergebnissen

Da Dell Technologies seine PowerEdge-Plattform mit AMD-Prozessoren optimiert hat, konnten zahlreiche Weltrekorde bei Benchmarks zur Messung der KI- und ML-Leistung erzielt werden. Bereits in einer Vor-Ort-Implementierung zeigten AMD EPYC-Prozessoren der 3. Generation eine starke Leistung, Leistung pro Watt und Leistung pro CPU-Dollar. In der Cloud unterstützen AMD EPYC Systems on Chip (SoCs) HPC-optimierte IaaS-Instanzen (Infrastructure as a Service) für zahlreiche Cloud-Serviceanbieter (CSPs), darunter Amazon Web Services® (AWS®), Microsoft® Azure®, Google Cloud Platform™ und andere.

Die AMD EPYC-Prozessoren der 4. Generation, die in den Weltrekord-Serverkonfigurationen von Dell bei der TPCx-AI verwendet wurden, bieten Leistungssteigerungen, die auf mehrere Plattformverbesserungen gegenüber der vorherigen Generation zurückgeführt werden können, darunter:

- 50 % mehr Cores<sup>4</sup>, höhere Thread-Anzahl und höhere Frequenzen, wodurch die Verarbeitungsleistung direkt gesteigert werden kann
- 12 DIMMs/Socket (bis zu 8), wodurch Unternehmen den verfügbaren Arbeitsspeicher deutlich erhöhen können. Dies führt zu einer schnelleren Verarbeitung größerer Datasets, insbesondere für In-Memory-Analysen, wie sie von Apache Spark™ durchgeführt werden.
- DDR5-Speicherunterstützung für schnelleren Zugriff auf Daten
- Unterstützung für AVX-512, wodurch AMD EPYC-Prozessoren der 4. Generation mehr gleichzeitige Berechnungen in ihren Registern durchführen können
- Größerer L2-Cache, verdoppelt von 512 KiB auf 1 MiB pro Core, wodurch auch Vorgänge im Arbeitsspeicher beschleunigt werden
- PCIe Gen 5-Unterstützung, die schnellere Verbindungen ermöglicht, um mehr Daten mit niedrigerer Latenz zu übertragen
- Verbesserungen speziell für KI- und ML-Workloads, u. a. Unterstützung für den numerischen Datentyp bfloat16 zum schnelleren Trainieren von KI-Modellen und Unterstützung für INT8-Inferenzierung für eine höhere Leistung bereits trainierter Modelle in der Produktion

Insgesamt arbeiten AMD EPYC-Prozessoren der 4. Generation effizienter als ihre Vorgänger. Die Ergebnisse der SPEC CPU® 2017 Floating Point Rate der Standard Performance Evaluation Corporation zeigen eine Leistungssteigerung von 121 % bei Tests auf einem System mit AMD EPYC-Prozessoren der 4. Generation im Vergleich zu einem System mit AMD EPYC-Prozessoren der 3. Generation.<sup>5</sup> Die Ergebnisse der SPEC CPU2017 Integer Rates zeigten Steigerungen von 102 %.<sup>6</sup> Diese Prozessorleistungsergebnisse spiegeln sich in den Weltrekord-Benchmarkergebnissen wider, die von mehreren der von uns untersuchten PowerEdge-Plattformen erzielt wurden.

Die Anzahl der Cores in diesen Prozessoren ist im Vergleich zur vorherigen Generation um 50 % gestiegen, was die Leistung ebenfalls steigert. Gleichzeitig zeigen die veröffentlichten Spezifikationen von AMD eine Steigerung des standardmäßigen Höchststromverbrauchs von nur 42 % von 280 Watt TDP (Thermal Design Power) auf 400 Watt Maximal-TDP.<sup>7</sup> Verglichen mit den oben genannten SPEC-Leistungsergebnissen zeigen diese Stromverbrauchszahlen, dass die Server, die auf AMD EPYC-Prozessoren der 4. Generation basieren, einen Stromverbrauchs-Leistungsvorteil von bis zu 55 % für Unternehmen mit KI- und ML-Workloads bieten.<sup>8</sup>

Die TPCx-AI-Benchmark, mit der die KI- und ML-Leistung gemessen wird, spiegelt auch die von den Broadcom Netzwerkkarten und Dell PERC-Karten bereitgestellte Leistung wider. Unternehmen mit KI- und ML-Workloads benötigen RAID-Controller für Redundanzen, um interne oder behördliche Auflagen zu erfüllen. Die PowerEdge R6625- und PowerEdge R7615-Server, die die TPCx-AI-Weltrekorde aufgestellt haben, sind mit Dell PERC-Karten ausgestattet, die Unterstützung für schnelles NVMe-RAID bieten.

### Hardwarebasierte Sicherheit von AMD®

Für alle in dieser Studie evaluierten Workloads sind Sicherheitsüberlegungen von entscheidender Bedeutung. AMD EPYC™-Prozessoren der 3. Generation und AMD EPYC-Prozessoren der 4. Generation bieten hardwarebasierte Sicherheit für KI- und ML-Workloads. AMD® Secure Memory Encryption (AMD® SME) verschlüsselt den System Speicher, um verwendete Daten zu schützen. AMD® Secure Encrypted Virtualization (AMD® SEV) schützt laufende virtuelle Maschinen (VMs), sodass diese verschlüsselt und voneinander sowie vom Hostsystem-Hypervisor isoliert werden. AMD® Secure Encrypted Virtualization-Encrypted State (AMD® SEV-ES) verschlüsselt die CPU-Registerinhalte angehaltener VMs, um die darin gespeicherten Daten zu schützen. Und AMD® Secure Boot schützt Server während des Startvorgangs und bietet Abwehrmechanismen gegen Rootkits, Bootkits und Firmware, wenn Server am anfälligsten sind.

Netzwerkkarten von Broadcom beschleunigen den Datenfluss für KI/ML-Workloads. Da die AMD EPYC-Prozessoren der 4. Generation und die Broadcom NICs Unterstützung für PCIe Gen 5 bieten, können 100-GbE-Netzwerkkarten von Broadcom eingesetzt werden, die auf dem NIC-3.0-Formfaktor des Open Compute Projects (OCP) basieren. Diese modernen Designs spiegeln den schnellen Übergang der Branche hin zu 100-GbE-Adaptoren wider, die auf einem effizienteren Formfaktor basieren und durch PCIe 4.0 und PCIe 5.0 ermöglicht werden. Darüber hinaus kann die Unterstützung für PCIe 4.0 und PCIe 5.0 Leistungszahlen von einem einzigen NIC hervorbringen, die zwei 100-Gbit/s-NICs in nichts nachstehen. Die NIC-3.0-Spezifikation von OCP ermöglicht es Serverherstellern wie Dell Technologies, kompaktere Designs zu verwenden, die Unterstützung für leistungsfähige Adapter mit erweiterten Hardwarebeschleunigungsfunktionen bieten, um KI- und ML-Workloads weiter zu beschleunigen.<sup>9</sup>

### Der Dell™ PowerEdge™-RAID-Controller (Dell™ PERC) schützt Daten und steigert die Storage-Leistung

Moderne PCIe® Gen 4-RAID-Schnittstellen sorgen zusammen mit NVM Express®-(NVMe®-)Solid-State-Laufwerken (SSDs) mit hoher Bandbreite für eine deutliche Steigerung der Storage-Leistung. Zwei Dell PowerEdge-RAID-Controller-11- und -12-Karten (PERC 11 und PERC 12) sowie NVMe-Adapter mit PCIe Gen 4-Host- und PCIe Gen 4-Storage-Schnittstellen können dazu beitragen, Bandbreiten- und Latenzeinschränkungen zu beseitigen.



## Fazit

Benchmarkergebnisse im Allgemeinen (und Weltrekordergebnisse im Besonderen) sind nicht nur dafür da, dass Serverhersteller sich damit brüsten können. Richtig interpretiert, können Best-in-Industry-Ergebnisse bei Benchmarks Einblicke in die Leistung von Servern für reale Anwendungsfälle bieten. Aufgrund des Marktanteils von Dell Technologies und der Anzahl der Weltrekorde haben die PowerEdge-Server eine natürliche Gelegenheit geboten, zu untersuchen, was die Benchmarkergebnisse für Leistungsvorteile für Unternehmen in der Produktion bieten. Auch wenn die Benchmarkleistung (Weltrekord oder anderweitig) nie 1:1 übertragen werden kann, zeigt unsere Untersuchung, dass die in einer branchenweit anerkannten Benchmark erzielte Leistung der Dell PowerEdge-Server auf eine starke KI- und ML-Leistung für verschiedene Anwendungsfälle in unterschiedlichen Branchen hinweist. Darüber hinaus beweist die Anzahl der Weltrekorde von Dell Technologies über verschiedene Benchmarks hinweg, dass das Unternehmen Plattformen entwickelt hat, die die Stärken der einzelnen Komponenten nutzen, um KundInnen einen echten Mehrwert für eine Vielzahl von Workloads zu bieten.

## Anhang A: Links zur Benchmarkleistung

- TPCx-AI-Leistungsergebnisse: [www.tpc.org/tpcx-ai/results/tpcxai\\_perf\\_results5.asp](http://www.tpc.org/tpcx-ai/results/tpcxai_perf_results5.asp)
- TPCx-IoT V2-Leistungsergebnisse: [www.tpc.org/tpcx-iot/results/tpcxiot\\_perf\\_results5.asp?version=2](http://www.tpc.org/tpcx-iot/results/tpcxiot_perf_results5.asp?version=2)

## Anhang B: Links zu Dell Technologies Systemspezifikationen

- Technische Datenblätter zu Dell PowerEdge-Servern: [www.dell.com/en-us/dt/servers/poweredge-rack-servers.htm](http://www.dell.com/en-us/dt/servers/poweredge-rack-servers.htm)

<sup>1</sup> History-Computer, „The 10 Largest Server Companies In The World, And What They Do“, September 2022, <https://history-computer.com/largest-server-companies-in-the-world-and-what-they-do/>.

<sup>2</sup> TPC, „TPCx-AI Top Performance Results“, abgerufen am 1. November 2022, [www.tpc.org/tpcx-iot/results/tpcxiot\\_perf\\_results5.asp?version=2](http://www.tpc.org/tpcx-iot/results/tpcxiot_perf_results5.asp?version=2).

<sup>3</sup> TPC, „TPCx-IoT V2 Top Performance Results“, abgerufen am 1. November 2022, [www.tpc.org/tpcx-iot/results/tpcxiot\\_perf\\_results5.asp?version=2](http://www.tpc.org/tpcx-iot/results/tpcxiot_perf_results5.asp?version=2).

<sup>4</sup> Tom's Hardware, „Zen 4 Madness: AMD EPYC Genoa With 96 Cores, 12-Channel DDR5 Memory, and AVX-512“, August 2021, [www.tomshardware.com/news/zen4-madness-amd-epyc-genoa-with-96-cores-12-channel-ddr5-memory-and-avx-512](http://www.tomshardware.com/news/zen4-madness-amd-epyc-genoa-with-96-cores-12-channel-ddr5-memory-and-avx-512).

<sup>5</sup> Bis zu 121 % höhere SPEC® Floating Point-Leistung beim Vergleich von AMD EPYC™-Prozessoren der 4. Generation mit AMD EPYC-Prozessoren der 3. Generation, basierend auf der SPEC Floating Point Rate-Punktzahl von 1.410 auf einem Dell™ PowerEdge™ R7625-Server mit AMD EPYC 9654-Prozessoren verglichen mit der Punktzahl von 636 auf einem Dell PowerEdge R7525-Server mit AMD EPYC 7763-Prozessoren. Bewertungen am 10. November 2022 abgerufen, siehe Benchmarkergebnisse der Standard Performance Evaluation Corporation, <http://spec.org/benchmarks.html>.

<sup>6</sup> Bis zu 102 % höhere SPEC® Integer Rate-Leistung beim Vergleich von AMD EPYC™-Prozessoren der 4. Generation mit AMD EPYC-Prozessoren der 3. Generation, basierend auf der SPEC Integer Rate-Punktzahl von 1.660 auf einem Dell™ PowerEdge™ R7625-Server mit AMD EPYC 9654-Prozessoren verglichen mit der Punktzahl von 821 auf einem Dell PowerEdge R7525-Server mit AMD EPYC 7763-Prozessoren. Bewertungen am 10. November 2022 abgerufen, siehe Benchmarkergebnisse der Standard Performance Evaluation Corporation, <http://spec.org/benchmarks.html>.

<sup>7</sup> AMD, Website mit den technischen Daten der AMD EPYC-Prozessoren der 7003-Serie, [www.amd.com/en/processors/epyc-7003-series](http://www.amd.com/en/processors/epyc-7003-series).

<sup>8</sup> 55 % verbesserte CPU-Leistung pro Watt, berechnet anhand der SPEC® Floating Point-Punktzahl von 1.410 auf einem Dell™ PowerEdge™ R7625-Server mit AMD EPYC 9654-Prozessoren mit einem Prozessor-cTDP von 400 Watt verglichen mit der Punktzahl von 636 auf einem Dell PowerEdge R7525-Server mit AMD EPYC 7763-Prozessoren und einem Prozessor-cTDP von 280 Watt.

<sup>9</sup> Broadcom, NetXtreme E-Series OCP NIC 3.0 Ethernet Adapters Product Brief, 2021, <https://docs.broadcom.com/doc/12395120>.

