



Dell Technologies



AMD
together we advance_

Leistungsstarke KI-Erkenntnisse mit einem offenen Ökosystem

Dell PowerEdge XE9680 Server mit AMD Instinct™ MI300X
Accelerator und der offenen Softwareplattform ROCm™ 6

Mit der wachsenden Popularität von ChatGPT und anderen generativen KI-Anwendungen (GenAI) setzt sich künstliche Intelligenz (KI) immer schneller durch. Und immer mehr Unternehmen möchten hier Schritt halten. Eine aktuelle Studie von IDC ergab Folgendes:



77 % der Unternehmen untersuchen KI-Anwendungsbeispiele oder investieren bereits hohe Summen in KI-Initiativen. ¹.

Unternehmen, die von schnelleren Erkenntnissen und anderen KI-Vorteilen profitieren möchten, müssen schnell handeln, sollten sie es noch nicht getan haben. Jetzt ist der richtige Zeitpunkt, insbesondere dank eines offenen Ökosystems, das stark in die Entwicklung der KI eingebunden ist, und einer neuen Lösung, die die KI und Ihr Unternehmen fördern kann.

¹ IDC, Future Enterprise Resiliency and Spending Survey, Wave 6, Juli 2023.



Der Weiterentwicklung steht nichts im Wege

Mit GenAI und anderen aufstrebenden erweiterten Anwendungen wird die Chiptechnologie noch leistungsfähiger und ausgefeilter. Doch proprietäre Systeme stellen auf der grundlegenden Ebene Hindernisse für Entwickler dar.

Dies wird sich jedoch ändern. Dell Technologies und AMD arbeiten zusammen, um Ihnen mehr Möglichkeiten mit Infrastrukturlösungen auf Basis von Open-Source-Technologien bieten zu können. Dank AMD Instinct™ MI300X Accelerator und der offenen Softwareplattform AMD Radeon Open Compute (ROCm™) auf dem Dell PowerEdge XE9680 Server erhalten Sie eine leistungsstarke Ökosystemoption – eine offene Tür für viele Innovationen –, um die Möglichkeiten zu nutzen und Ihr ganzes Potenzial zu entfalten.

Mehr Leistung und Flexibilität mit Dell Technologies und AMD

Dell Technologies und AMD unterstützen Unternehmen auf der ganzen Welt, Ergebnisse mit einem System schneller zu erreichen, das EntwicklerInnen zum Erfolg verhilft.



Der Dell PowerEdge XE9680 Server ist auf leistungsstarke Datenverarbeitung ausgelegt, der speziell für KI-Aufgaben entwickelt wurde. Er unterstützt acht Accelerators, die sich ideal für das Training von ML (maschinelles Lernen)/ Deep Learning (DL) und Inferenzierung-Workloads eignen, insbesondere für das Training großer Sprachmodelle (LLMs).

Der AMD Instinct MI300X Accelerator mit AMD CDNA™-Architektur (Compute DNA) der 3. Generation bietet die größte Arbeitsspeicherkapazität mit hoher Bandbreite auf dem GPU-Markt und bietet 1,5 Terabyte kohärenten GPU-Beschleunigerspeicher pro Server für das Training der größten KI-Modelle. Außerdem können Sie die Inferenzierungs-Kapazität durch weniger Server erhöhen.

Höhere Geschwindigkeit mit ROCm 6

ROCm entfaltet wiederum die ganze Rechenleistung von AMD Instinct. ROCm ist eine offene, flexible und skalierbare Softwareplattform für beschleunigtes Computing und verfügt über Treiber, APIs und andere GPU-Tools, die die Codeoptimierung für AMD Instinct-Accelerator unterstützen und die Portabilität für vorhandene GPU-beschleunigte Codes vereinfachen.

Außerdem wird Ihre Erfahrung erheblich vereinfacht. ROCm wird auf Geräteebene mit einer gemeinsam genutzten Codebasis optimiert, die sich entsprechend den Systemen und Teilen weiterentwickeln kann. Diese offene Flexibilität ermöglicht auch eine mühelose Portierung, dabei ist eine Programmierung nahezu nicht nötig.

Ein grenzenloses Ökosystem

AMD Instinct MI300X Accelerator und ROCm basieren auf offenen Standards und Interoperabilität und werden nativ von einer Vielzahl an KI-Frameworks und -Bibliotheken unterstützt. Dies bietet mehr Auswahl und Flexibilität für diejenigen, die die Leistung von AMD nutzen und dennoch die Kompatibilität mit vorhandenen Softwaresystemen aufrechterhalten möchten.

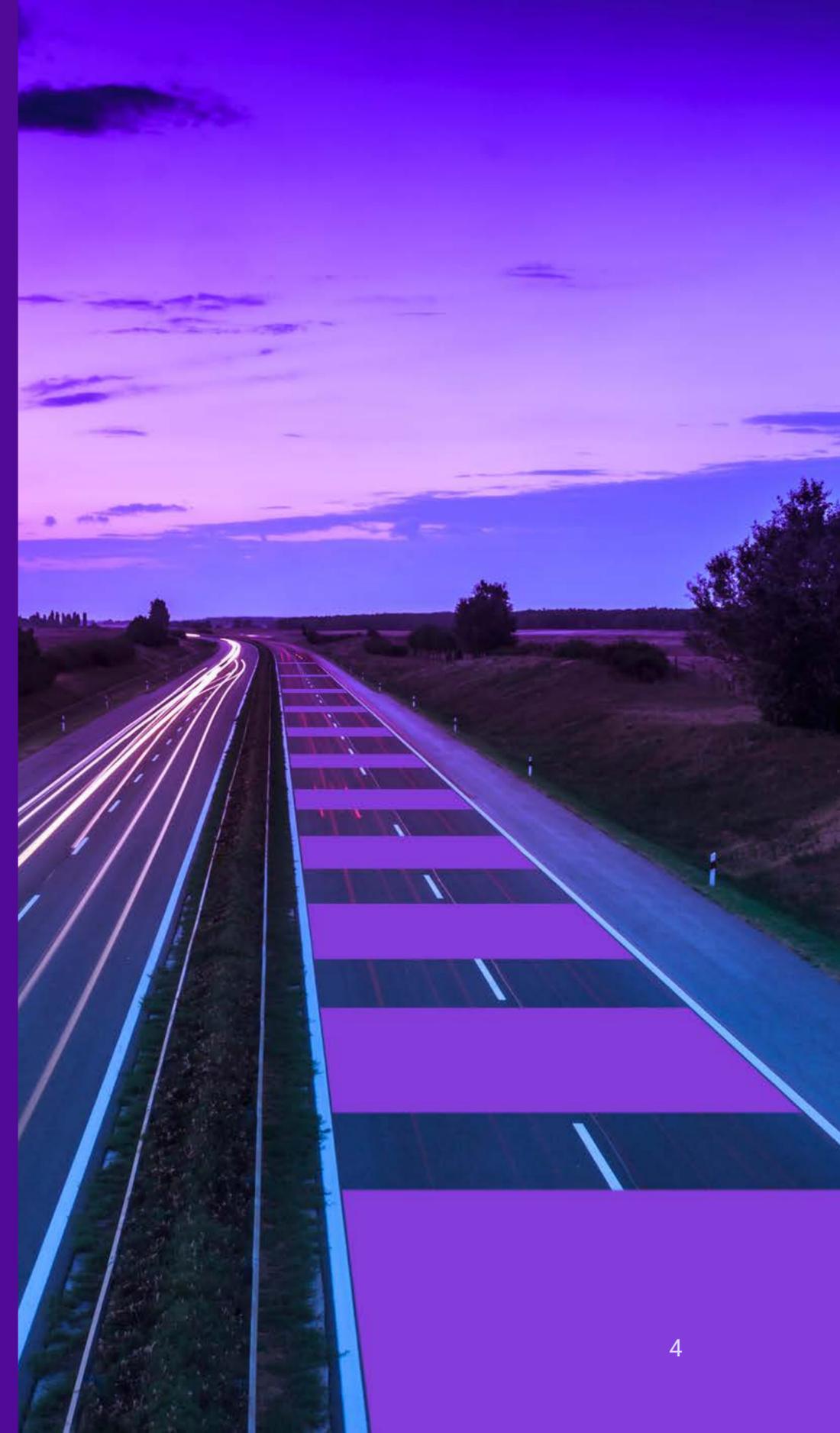


AMD ebnet den Weg und bietet die Freiheit, Ihre Projekte schnell voranzutreiben.

Vorteile von ROCm

- Bietet GPU-Leistung zur Beschleunigung von Rechenaufgaben
- Optimiert für skalierbare ML-/DL-/GenAI-Workloads
- Bietet sofortige Unterstützung für führende Modelle und Frameworks (einschließlich PyTorch®, TensorFlow™ und andere)
- Ermöglicht die schnelle Einführung neuer Innovationen (einschließlich nativer Unterstützung für JAX, Open Neural Network Xchange (ONNX), OpenAI Triton usw.)
- Fördert Innovation, Differenzierung und Zusammenarbeit gemeinsam mit der OSS-Entwicklungscommunity
- Bietet eine flexible Grundlage für offene Portabilität

Mehr erfahren über [Neuheiten in ROCm](#).



Einfache Portierung und Optimierung von Anwendungen für AMD Instinct mit ROCm

- Ermöglicht eine reibungslose Softwareportierung auf AMD. Mit HIP können EntwicklerInnen portable Anwendungen für AMD Instinct MI300X aus einem einzigen Quellcode erstellen.
- Bietet alle Funktionen mit häufig verwendeten Bibliotheken und Operatoren.
- Schnelle Portierung und Optimierung der meisten Anwendungen für die Performance

Wachsende Upstream-Unterstützung

ROCm bietet Upstream-Unterstützung für die beliebtesten Frameworks von heute, einschließlich, aber nicht beschränkt auf:

- PyTorch
- TensorFlow
- Open Neural Network Xchange (ONNX)
- JAX



Dank offener Portabilität wertvolle Zeit sparen

Mit ROCm müssen sich EntwicklerInnen, die möglicherweise vor einer neuen Umgebung und den Auswirkungen, die es auf den vorhandenen Code haben kann, verunsichert sind, keine Sorgen machen. Ein wesentlicher Vorteil der ROCm-Plattform ist die Tatsache, dass Sie Anwendungen mit minimalen Nacharbeiten an Ihrem vorhandenen Code portieren können. Die Portierung ist einfach und Sie sparen enorm viel Zeit.

Portieren Sie vorhandene Anwendungen auf den AMD Instinct MI300X Accelerator auf vier verschiedene Weisen:

1 Nutzen Sie vorhandene KI-Frameworks mit Upstream-Unterstützung für ROCm.

2 Verwenden Sie Drop-in-ML-Kernels, die für AMD Instinct MI300X optimiert sind. Oder nutzen Sie das AMD HIP-Konvertierungstool, mit dem Sie CUDA-Anwendungen auf ROCm-basierten GPUs optimieren können.

3 Verwenden Sie gespiegelte Versionen vorhandener CUDA-Bibliothekssoftware-Repositorys, die für ROCm optimiert und gewartet werden.

4 Greifen Sie auf von AMD vorgefertigte KI-Anwendungen und -Container zu.

Machen Sie noch heute den ersten Schritt

Beschleunigen Sie vorhandene KI- und neue GenAI-Projekte mit einem offenen Framework von Dell Technologies und AMD. Führen Sie Ihre Workloads auf einer bewährten Dell PowerEdge-Grundlage aus, die die neuesten AMD-Accelerators unterstützt, einschließlich des AMD Instinct MI300X Accelerator. [ROCm Developer Hub besuchen.](#)



Der AMD Instinct MI300X Accelerator kann bereitgestellt werden

Mit seiner bahnbrechenden Leistung bietet der AMD Instinct MI300X Accelerator die Geschwindigkeit, Skalierbarkeit und flexible, offene Umgebung, die heutzutage benötigt wird, insbesondere wenn Unternehmen Herausforderungen rund um Ressourcen, Komplexität, Latenz und Architektur bewältigen müssen. Der AMD Instinct MI300X Accelerator bietet einen deutlichen Vorteil in Bezug auf folgende Punkte:



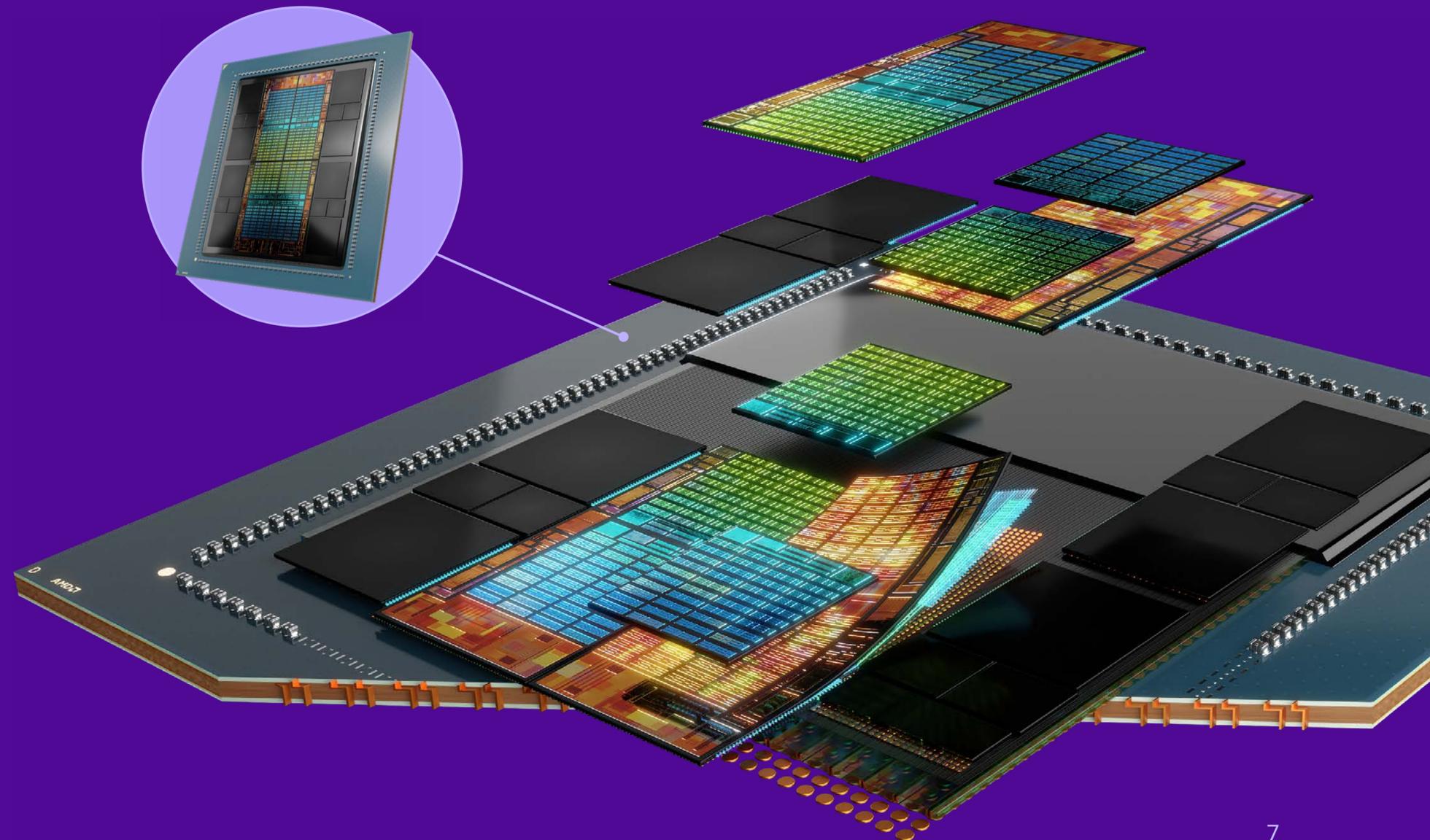
Performance: Ermöglichen Sie neue KI-Workflows mit 192 GB gestapeltem HBM3-Arbeitsspeicher pro GPU, der mit 5,3 TB pro Sekunde betrieben wird.



Dichte: Weniger Datenverschiebung dank modernstem Die-Stacking. Chiplet-Technologien für mehr Compute- und Arbeitsspeicherkapazität pro Server.



Frei zugänglich: Einfache Ergebnisse dank eines Ökosystems, das auf Zugänglichkeit und Anpassungsfähigkeit ausgelegt ist.



Volles KI-Potenzial mit dem Dell PowerEdge XE9680 Server entfalten

Mit einer speziell für KI entwickelten Serverplattform können Sie schnellere und intelligentere Ergebnisse schneller erreichen. Verbessern Sie datengesteuerte Geschäfte, bieten Sie visuelle Ergebnisse und beschleunigen Sie den geschäftsweiten KI-Betrieb mit einer Lösung, die Erkenntnisse beschleunigt, vertrauenswürdige KI bereitstellt und den Betrieb vereinfacht.



Vertrauenswürdige KI

Minimieren Sie Risiken und sparen Sie Zeit, während Sie gleichzeitig Ihren KI-Lebenszyklus mit einer vertrauenswürdigen, hochwertigen Lösungsinfrastruktur beschleunigen. PowerEdge XE9680 bietet integrierten Schutz von der chipbasierten Sicherheit (Silicon Root of Trust) bis hin zur Stilllegung von Beständen.



Schnellere Erkenntnisse

Erzielen Sie eine bessere Performance über den gesamten KI-Lebenszyklus hinweg und für die Bereitstellung von KI-Vorgängen im Tempo des Unternehmens. Dell Technologies bietet innovative Computelösungen, mit denen Sie KI mit maximaler Leistung nutzen können.



Vereinfachter Betrieb

Mit KI-Infrastrukturautomatisierung können Sie Ihre KI-Infrastruktur- und Performance-Workloads überall effektiv kontrollieren und managen. Dell Technologies bietet intelligente Systeme, die zusammen und unabhängig voneinander arbeiten, um Aufgaben zu beschleunigen und Ihre Erfahrung zu vereinfachen.

Konsequente Innovation

Dell Technologies und AMD arbeiten hart daran, Innovationen für unsere geschätzten KundInnen zu fördern. Gemeinsam bringen wir Chipvielfalt in einer praktikablen Lösung auf den Markt, die EntwicklerInnen noch mehr Möglichkeiten bietet. Unsere gemeinsame Lösung bietet Ihnen folgende Vorteile:



- **Schnellere Markteinführung** mit mehr Möglichkeiten und weniger Hindernissen

- **Gewinnen Sie schnellere Erkenntnisse** mit einem offenen Ökosystem, das Sie nicht ausbremst.

- **Entscheiden Sie sich für die beste Lösung.** Verlassen Sie sich auf die Vorteile der Dell Lieferkette, die stets innovative Technologien einführt, wenn sie am dringendsten benötigt werden.

Grenzenlose Möglichkeiten mit Dell Technologies und AMD

KI ist jetzt da, aber ihr Einflussbereich wird sich stets erweitern und beschleunigen. Für Unternehmen, die diese Funktionen nutzen möchten, scheinen die Möglichkeiten endlos zu sein.

Aus diesem Grund soll Ihr Potenzial nicht durch technologische Hindernisse eingeschränkt werden. Mit leistungsstarken Lösungen, die für KI-Workloads vorgesehen sind, halten Sie sich alle Optionen offen.

Die optimale Lösung: AMD Instinct MI300X Accelerator und die ROCm-Software auf dem Dell PowerEdge XE9680 Server. Schnellere Beschaffung von neuen Anwendungen, schnellere Markteinführung und Förderung von Ergebnissen, die für Ihr Unternehmen und die Welt von großer Bedeutung sind.

Weitere Informationen unter [Dell.com/GPU](https://www.dell.com/gpu)

Copyright © 2023 Dell Inc. oder Tochtergesellschaften. Alle Rechte vorbehalten. Dell und andere Marken sind Marken von Dell Inc. oder deren Tochtergesellschaften. AMD, Instinct™, ROCm™ und deren Kombinationen sind Marken von Advanced Micro Devices, Inc. PyTorch® ist eine Marke oder eingetragene Marke von PyTorch oder den Lizenzgebern von PyTorch. TensorFlow™ und alle damit zusammenhängenden Marken sind eingetragene Marken von Google Inc. CUDA® ist eine Marke und/oder eingetragene Marke der NVIDIA Corporation in den USA und anderen Ländern. Alle anderen Marken können Marken ihrer jeweiligen Inhaber sein. Veröffentlicht in Deutschland. 12/23 E-Book.

Dell Technologies ist der Ansicht, dass die Informationen in diesem Dokument zum Zeitpunkt der Veröffentlichung korrekt sind. Die Informationen können jederzeit ohne vorherige Ankündigung geändert werden.