



DELL Technologies

AMD
together we advance_

Skapa kraftfulla AI-insikter med ett öppet ekosystem



Servern Dell PowerEdge XE9680 med förstärkaren AMD Instinct™ MI300X och den öppna mjukvaruplattformen ROCm™ 6

I och med att ChatGPT och andra generativa AI-program (GenAI) har blivit allt populärare har artificiell intelligens (AI) snabbt spridits. Och allt fler organisationer vill haka på. I en nyligen genomförd IDC-studie kom man fram till att:



77 % av organisationerna

undersöker antingen AI-användningsfall eller investerar redan avsevärt i AI-initiativ.¹

Organisationer som vill dra nytta av snabbare insikter och andra fördelar med AI måste agera snart, om de inte redan har gjort det. Tiden är inne, särskilt med ett öppet ekosystem som är inblandat i utvecklingen av AI och en ny lösning som är redo att driva AI-innovation – och verksamheten – framåt.

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



Vägen framåt är öppen

Med GenAI och andra avancerade program på uppgång blir chiptekniken mer kraftfull och avancerad. Men tillverkarspecifika system har i grund och botten skapat hinder för utvecklare.

Det håller på att ändras. Dell Technologies och AMD samarbetar för att öka er kapacitet med infrastrukturlösningar som bygger på öppen källa. När ni använder förstärkaren AMD Instinct™ MI300X och den öppna mjukvaruplattformen AMD Radeon Open Compute (ROCm™) på Dells server PowerEdge XE9680 får du ett kraftfullt ekosystemalternativ – en öppen väg framåt – som ger er möjlighet att frigöra er potential.

Få prestanda och flexibilitet med Dell Technologies och AMD

Dell Technologies och AMD hjälper företag världen över att snabba upp resultaten med ett system som hjälper utvecklare att lyckas.



Dells server PowerEdge XE9680 är ett kraftpaket för databearbetning som är särskilt utformat för AI-uppgifter. Den har stöd för åtta förstärkare, vilket är perfekt för träning och inferens med maskininlärning (ML)/djupinlärning (DL), särskilt för dem som tränar stora språkmodeller (LLMs).

Förstärkaren AMD Instinct MI300X, som drivs av tredje generationens AMD CDNA-arkitektur™ (Compute DNA), har den största minneskapaciteten med hög bandbredd på GPU-marknaden, och ger 1,5 terabyte koherent GPU-förstärkarminne per server för att träna de största AI-modellerna. Det gör det också möjligt att öka inferensen med färre servrar.

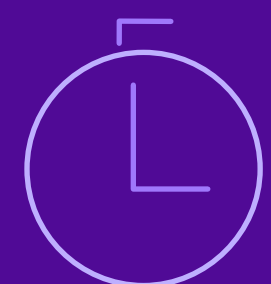
Frigör acceleration med ROCm 6

ROCm släpper i sin tur loss beräkningskraften hos AMD Instinct. ROCm är en öppen, flexibel och skalbar mjukvaruplattform för accelererad beräkning och den har drivrutiner, API:er och andra GPU-verktyg med stöd för kodoptimering för förstärkarna AMD Instinct och förenklar portabiliteten för befintlig GPU-förstärkt kod.

Det förenklar också upplevelsen radikalt. ROCm optimeras på enhetsnivå med en fritt delad kodbas som kan utvecklas i takt med att system och deras delar utvecklas. Med den öppna flexibiliteten blir enklare portering och nästan ingen kodning möjlig.

Ett ekosystem utan gränser

Förstärkaren AMD Instinct MI300X och ROCm bygger på öppna standarder och interoperabilitet och de stöds av en mängd olika AI-ramverk och -bibliotek. Detta ger större valfrihet och flexibilitet för dem som vill utnyttja kraften i AMD och fortfarande upprätthålla kompatibilitet med befintliga mjukvarusystem.

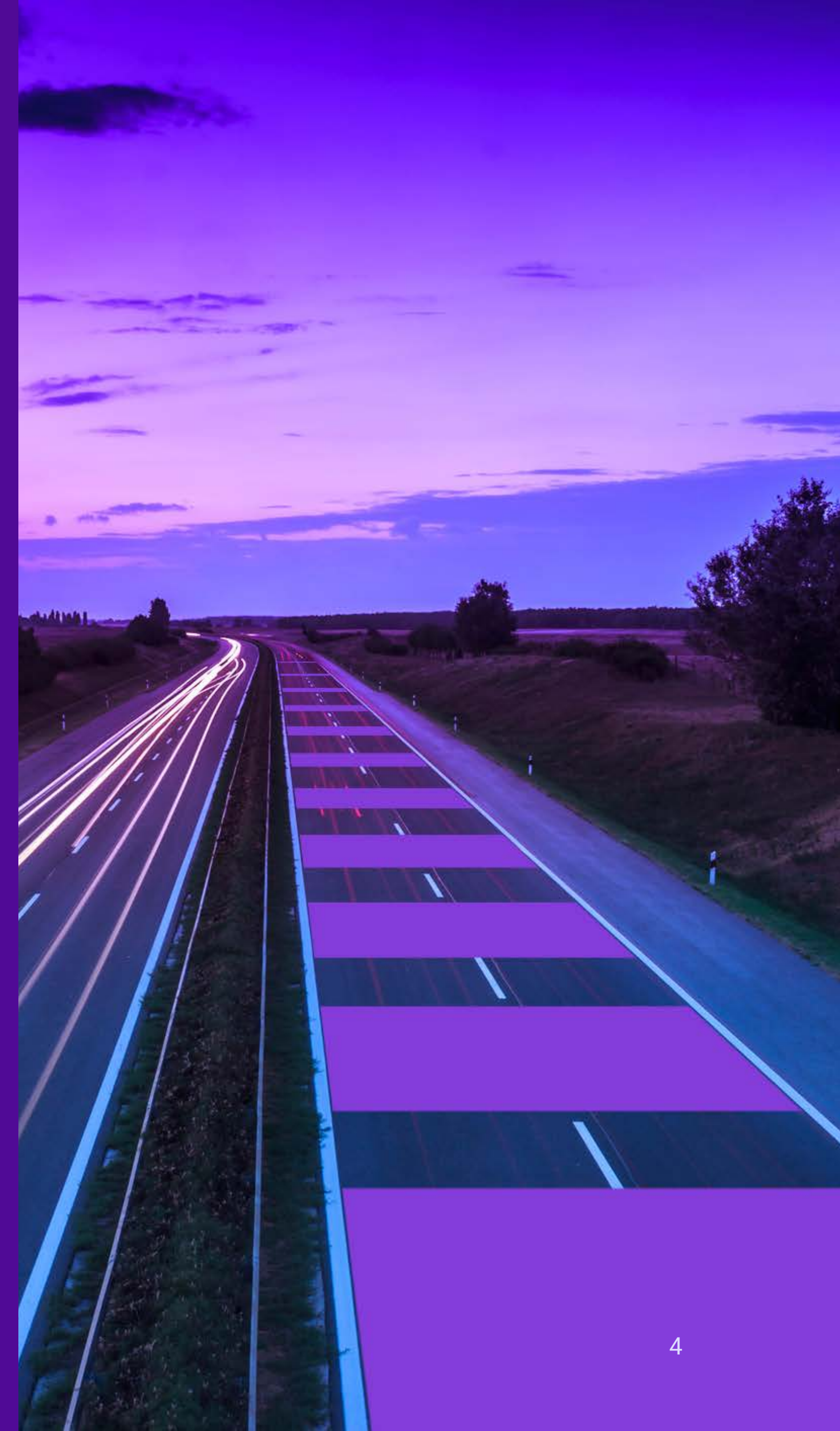


AMD tar bort hinder och ger er friheten att utveckla projekt snabbt.

Fördelar med ROCm

- Frigör GPU-kraft för att snabba upp beräkningsuppgifter
- Optimerad för ML-/DL-/GenAI-arbetsbelastningar i skala
- Ger heltäckande stöd för ledande modeller och ramverk (inklusive PyTorch®, TensorFlow™ och andra)
- Möjliggör snabbt införande av nya innovationer (inklusive inbyggt stöd i JAX, Open Neural Network Xchange (ONNX), OpenAI Triton, etc.)
- Driver innovation, differentiering och samarbete med kontakt med OSS-utvecklarecommunityn
- Ger en flexibel grund för öppen portabilitet

Mer information om [nyheter i ROCm](#).



Portera och optimera enkelt program för AMD Instinct med ROCm

- Möjliggör mjukvaruportering med låg friktion till AMD. HIP gör det möjligt för utvecklare att skapa flyttbara program för AMD Instinct MI300X från en enda källkod.
- Erbjuder alla funktioner med vanliga bibliotek och aktörer
- Porterar snabbt och optimerar de flesta program för prestanda

Växande uppströmssupport

ROCm har uppströmsstöd för dagens mest populära ramverk, inklusive men inte begränsat till:

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



Spara värdefull tid med öppen portabilitet

ROCm ger sinnesro till utvecklare som kan vara på sin vakt mot nya ekosystem och den inverkan de kan ha på befintlig kod. En viktig fördel med ROCm-plattformen är att du kan anpassa program med minimal omarbetning av befintlig kod. Portering är enkelt, och du sparar otroligt mycket tid.

Portera befintliga program till förstärkaren AMD Instinct MI300X på fyra olika sätt:

1 Utnyttja befintliga AI-ramverk som har uppströmsstöd för ROCm.

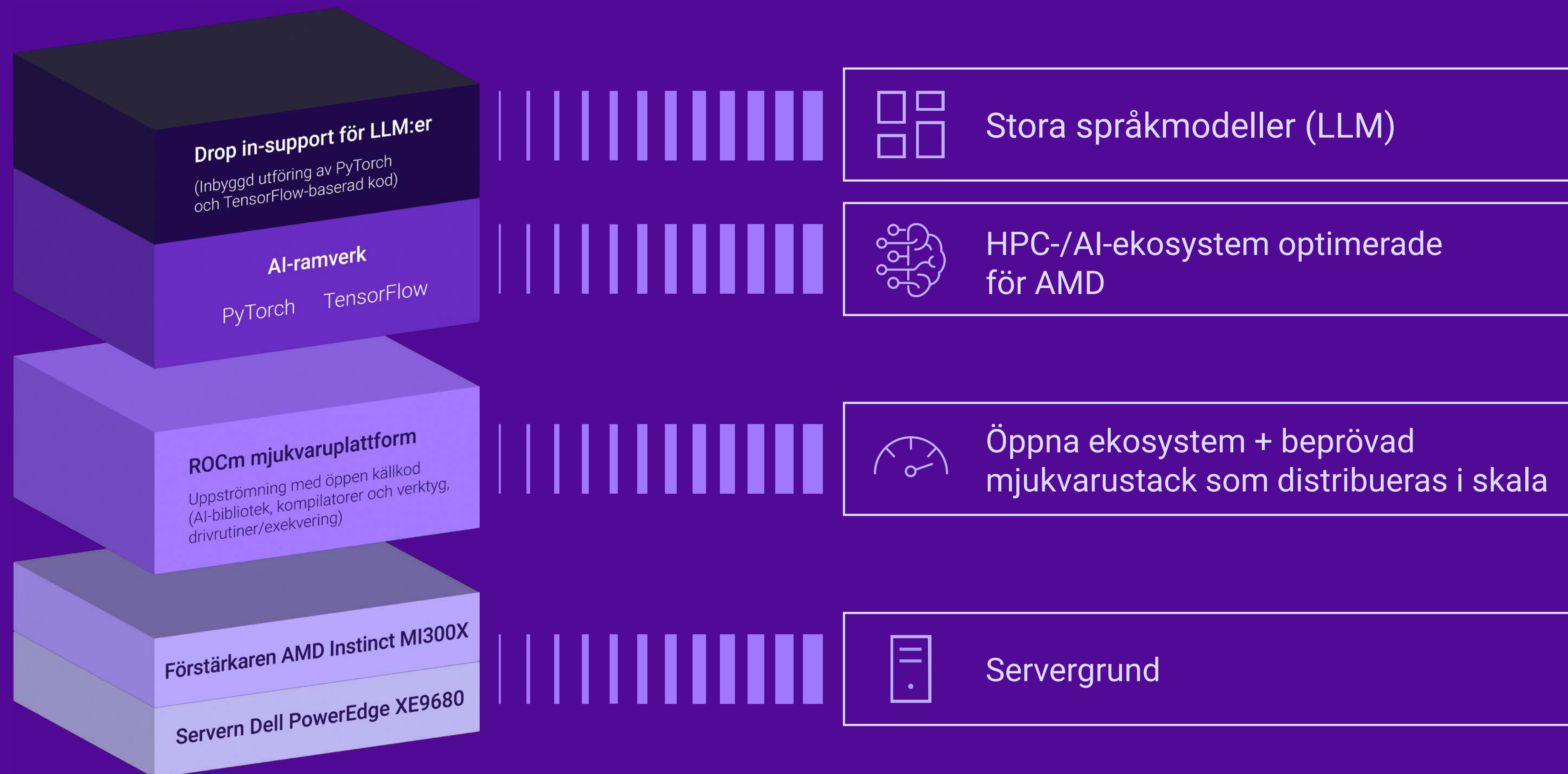
2 Använd drop-in maskininlärningskärnor som är optimerade för AMD Instinct MI300X. Eller använd konverteringsverktyget AMD HIP som gör att du kan optimera CUDA-program på ROCm-baserade GPU:er.

3 Använd speglade versioner av befintliga mjukvaruarkiv för CUDA-bibliotek som är optimerade och underhållna för ROCm.

4 Få åtkomst till AMD-färdigpaketerade AI-program och -behållare.

Kom igång idag

Snabba upp befintliga AI-projekt och nya GenAI-projekt med ett öppet ramverk från Dell Technologies och AMD. Ge dina arbetsbelastningar kraft på en beprövad Dell PowerEdge-grund som har stöd för de senaste AMD-förstärkarna, inklusive förstärkaren AMD Instinct MI300X. [Besök ROCm Developer Hub.](#)



Förstärkaren AMD Instinct MI300X är redo att installeras

Genom att leverera banbrytande prestanda ger förstärkaren AMD Instinct MI300X den hastighet, skalbarhet och det flexibla, öppna ekosystem som behövs i dag, särskilt när företag hanterar utmaningar med resurser, komplexitet, latens och arkitektur. Förstärkaren AMD Instinct MI300X ger en tydlig fördel när det gäller:



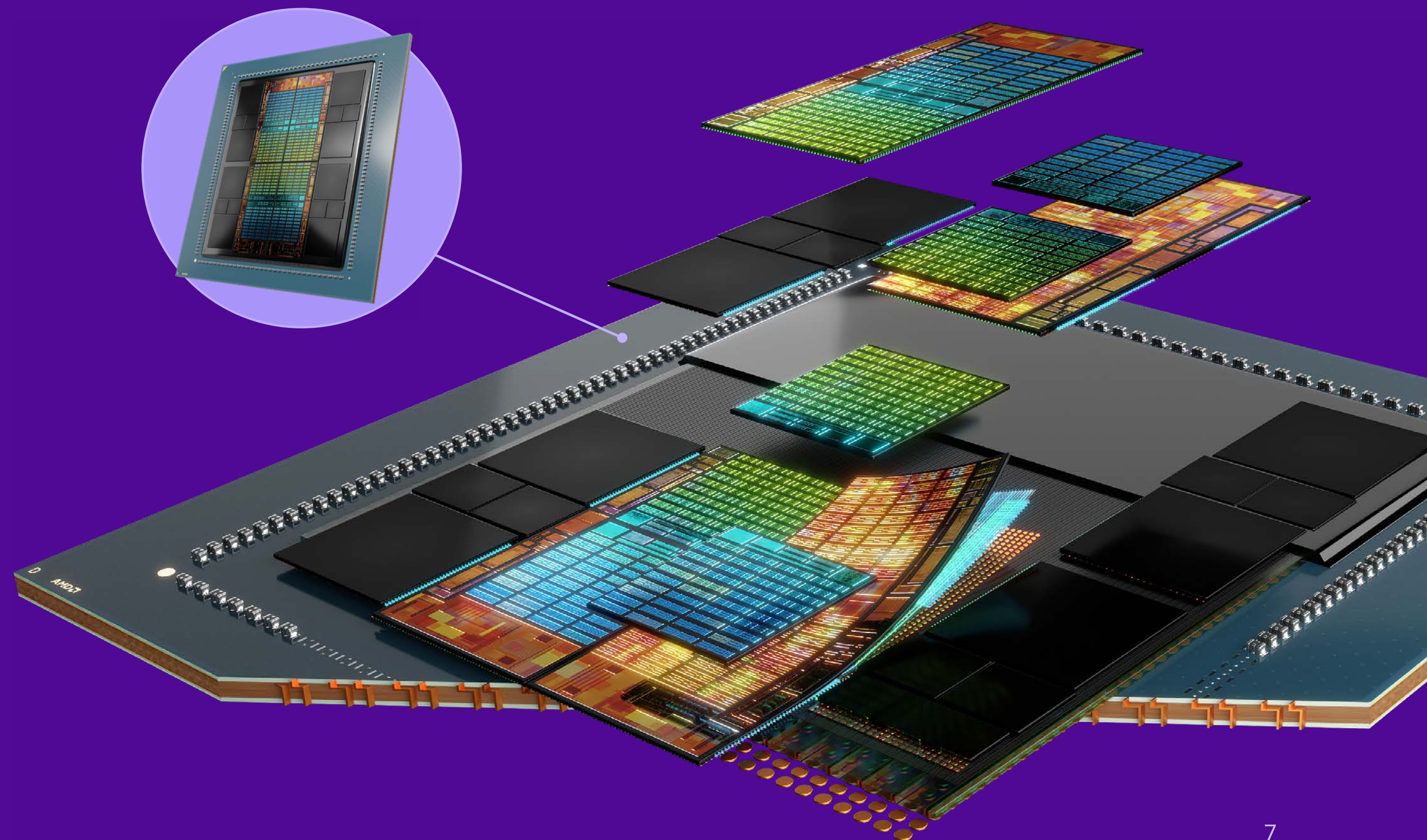
Prestanda: Möjliggör nya AI-arbetsflöden med ett aldrig tidigare skådat 192 GB stackat HBM3-minne per GPU som arbetar med 5,3 TB per sekund.



Densitet: Minska behovet av dataflyttning med den senaste kretsstackingen och chiplet-teknikerna som har mer beräknings- och minneskapacitet per server.



Öppet: Enkel väg till resultat med ett ekosystem som är utformat för tillgänglighet och anpassningsbarhet.



Släpp loss fördelarna med AI med servern Dell PowerEdge XE9680

Du kan snabba upp vägen till snabbare och smartare resultat med en serverplattform som är särskilt utformad för AI. Förbättra datadriven verksamhet, leverera visuella resultat och påskynda AI-verksamheten i hela företaget med en lösning som snabbar upp insikter, levererar betrodd AI och förenklar driften.



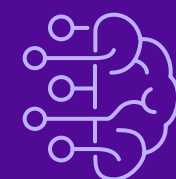
Betrodd AI

Minska riskerna och spara tid när du accelererar AI-livscykeln med en tillförlitlig lösningssinfrastruktur av hög kvalitet. PowerEdge XE9680 har ett inbyggt skydd i form av en kiselbaserad förtroenderot hela vägen till kassering av uttjänta tillgångar.



Förenklad drift

Styr och hantera AI-infrastrukturen och dina prestandaarbetsbelastningar effektivt var som helst med automatisering av AI-infrastrukturen. Dell Technologies tillhandahåller smarta system som arbetar tillsammans och oberoende för att snabba upp uppgifter och förenkla upplevelsen.



Förstärker insikter

Förbättra resultaten under AI-livscykeln och för leverans av AI-drift med samma hastighet som verksamheten. Dell Technologies levererar innovativ beräkning som hjälper dig att utnyttja AI med maximal prestanda.

Driv innovationen framåt

Dell Technologies och AMD arbetar hårt för att driva innovation framåt för våra uppskattade kunder. Tillsammans tar vi kisel mångfald till marknaden i en livskraftig lösning som stärker utvecklare. Med vår gemensamma lösning kan du:



- **komma ut på marknaden snabbare** med fler alternativ och färre hinder
- **få snabbare insikter** med ett öppet ekosystem som inte håller dig tillbaka
- **välja det som är bäst för dig.** Du kan lita på fördelarna med Dells återförsäljarkedja för att konsekvent få ut innovativ teknik på marknaden när den behövs som mest.

Öka produktiviteten med Dell Technologies och AMD

AI är här nu, men dess påverkan kommer bara att öka och accelerera framåt. För företag som vill dra nytta av dess kapacitet verkar möjligheterna oändliga.

Därför vill du inte att potentialen ska begränsas av tekniska hinder som ibland kan vara i vägen för dig. Håll istället alternativen öppna med en kraftfull lösningsstack avsedd för AI-arbetsbelastningar.

Förstärkaren AMD Instinct MI300X och mjukvaran ROCm på Dells server PowerEdge XE9680 är svaret. Snabba upp vägen till nya program, kom ut på marknaden snabbare och börja förbättra resultaten som är viktiga för ert företag och hela världen.

Mer information finns på [Dell.com/GPU](https://www.dell.com/gpu)

Upphovsrätt © 2023 Dell Inc. eller dess dotterbolag. Med ensamrätt. Dell och andra varumärken är varumärken som tillhör Dell Inc. eller dess dotterbolag. AMD, Instinct™, ROCm™ och kombinationer av dessa är varumärken som tillhör Advanced Micro Devices, Inc. PyTorch® är ett varumärke eller registrerat varumärke som tillhör PyTorch eller PyTorchs licensgivare. TensorFlow™ och alla relaterade märken är varumärken som tillhör Google Inc. CUDA® är ett varumärke eller registrerat varumärke som tillhör NVIDIA Corporation i USA och andra länder. Andra varumärken kan vara varumärken som tillhör sina respektive ägare. Publicerat i USA 12/23 eBook.

Dell Technologies anser att informationen i detta dokument är korrekt på publiceringsdagen. Informationen kan komma att ändras utan föregående meddelande.