

Impulsione insights de IA eficazes com um ecossistema de código aberto

Servidor Dell PowerEdge XE9680 com acelerador AMD Instinct™ MI300X e plataforma de software de código aberto ROCm™ 6

Com a crescente popularidade do ChatGPT e de outros aplicativos de IA generativa (GenAI), a inteligência artificial (IA) está sendo rapidamente difundida. E um número cada vez maior de organizações quer adotá-la. Um estudo recente da IDC mostra que:



77% das organizações

estão explorando casos de uso de IA ou já estão investindo substancialmente em iniciativas de IA.¹

As organizações que desejam aproveitar os insights acelerados e os outros benefícios da IA precisarão agir logo, caso isso ainda não tenha sido feito. Este é o momento oportuno, principalmente com um ecossistema de código aberto altamente envolvido no avanço da IA e com uma nova solução pronta para impulsionar a inovação em IA e os seus negócios.

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



O trajeto para o futuro está claro

Com a GenAI e outros aplicativos avançados em ascensão, a tecnologia de chip está ficando mais avançada e sofisticada. No entanto, os sistemas proprietários no nível de base criaram barreiras para desenvolvedores na linha de frente.

Isso está prestes a mudar. A Dell Technologies e a AMD estão trabalhando juntas para acelerar seus recursos com soluções de infraestrutura baseadas em tecnologias de código aberto. Executar o acelerador AMD Instinct™ MI300X e a plataforma de software de código aberto AMD Radeon Open Compute (ROCm™) no servidor Dell PowerEdge XE9680 oferece a você uma opção avançada de ecossistema – um caminho claro rumo ao futuro – para estimular oportunidades e liberar seu potencial.

Obtenha desempenho e flexibilidade com a Dell Technologies e a AMD

A Dell Technologies e a AMD estão ajudando empresas em todo o mundo a acelerar os resultados com um sistema que preparará os desenvolvedores para o sucesso.



O servidor Dell PowerEdge XE9680 é uma potência em processamento de dados e foi desenvolvido especialmente para tarefas de IA. Ele oferece suporte a oito aceleradores, sendo ideal para cargas de trabalho de inferência e treinamento de aprendizado de máquina (ML)/aprendizagem profunda (DL), principalmente para os grandes modelos de linguagem (LLMs) de treinamento.

O acelerador AMD Instinct MI300X, habilitado pela arquitetura AMD CDNA™ (Compute DNA) de 3ª geração, entrega a maior capacidade de memória de grande largura de banda do mercado de GPU, fornecendo 1,5 terabyte de memória de aceleração de GPU coerente por servidor para treinar os maiores modelos de IA. Ele também permite que você aumente a capacidade de inferência usando menos servidores.

Possibilite a aceleração com a plataforma ROCm 6

A plataforma ROCm, por sua vez, libera a capacidade computacional do AMD Instinct. A plataforma de software ROCm de código aberto, flexível e escalável para computação acelerada vem equipada com drivers, APIs e outras ferramentas de GPU que dão suporte à otimização de código para os aceleradores AMD Instinct e simplificam a portabilidade do código acelerado por GPU existente.

Ela também simplifica extremamente a sua experiência. A plataforma ROCm é otimizada no nível do dispositivo com uma base de código compartilhada livremente, que pode evoluir para acompanhar a evolução dos sistemas e seus componentes. Essa flexibilidade do código aberto também permite realizar a portabilidade com facilidade e praticamente sem codificação.

Um ecossistema sem fronteiras

Desenvolvidos com padrões abertos e interoperabilidade, o acelerador AMD Instinct MI300X e a plataforma ROCm são nativamente compatíveis com uma variedade de estruturas e bibliotecas de IA. Com isso, as organizações que desejarem aproveitar o poder da AMD e ainda manter a compatibilidade com os sistemas de software existentes terão mais opções e flexibilidade.

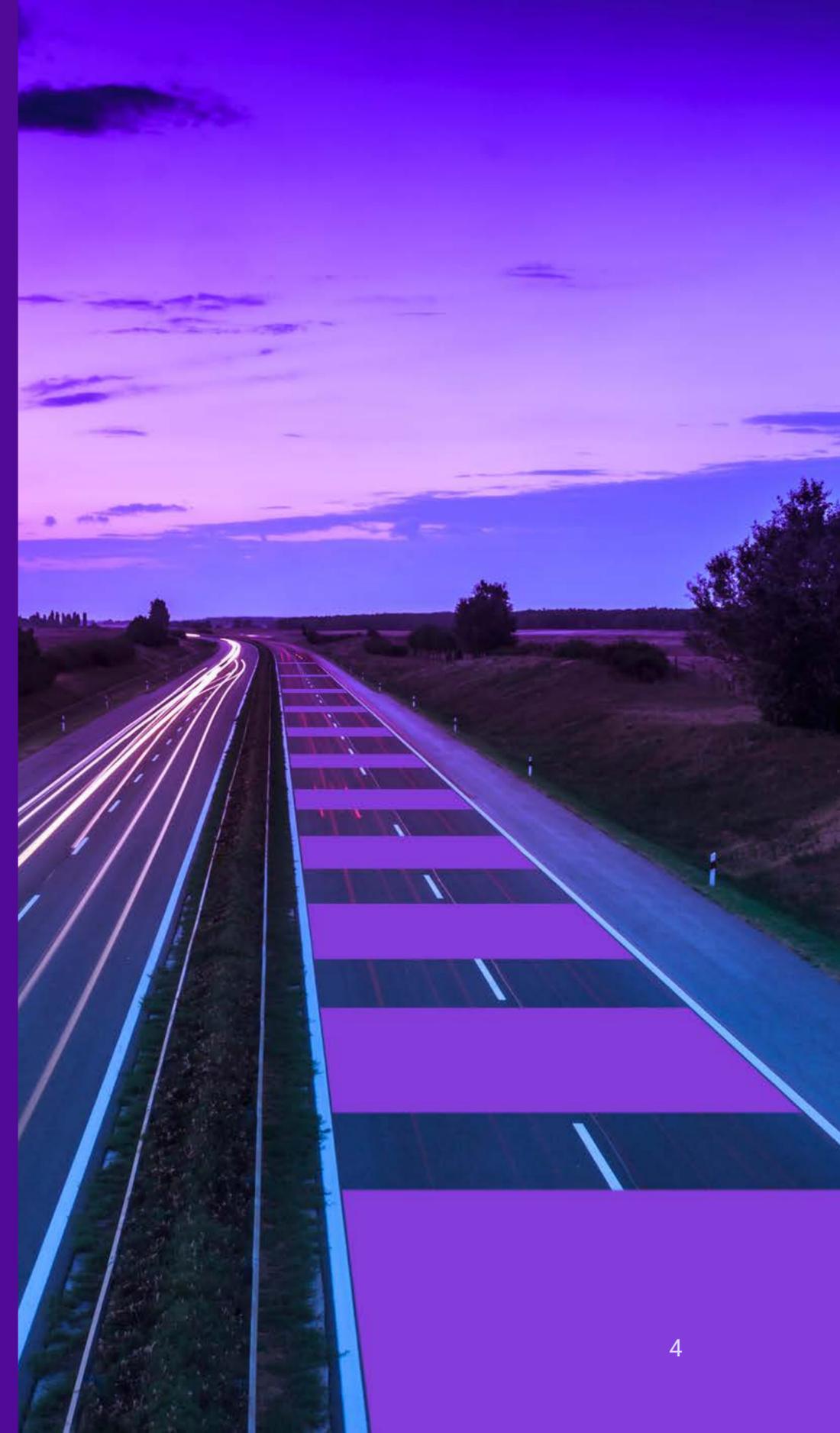


A AMD libera o caminho, proporcionando a liberdade de avançar seus projetos com rapidez.

Vantagens da plataforma ROCm

- Revela o potencial da GPU para acelerar as tarefas computacionais
- Otimizada para cargas de trabalho de ML/DL/GenAI em escala
- Oferece suporte integrado para os modelos e as estruturas mais importantes (inclusive PyTorch®, TensorFlow™ etc.)
- Permite a rápida adoção de inovações recentes (inclusive suporte nativo em JAX, Open Neural Network Xchange (ONNX), OpenAI Triton etc.)
- Impulsiona a inovação, a diferenciação e a colaboração com a conexão com a comunidade de desenvolvimento de OSS
- Fornece uma base flexível para a portabilidade aberta

Consulte [Novidades da plataforma ROCm](#) para saber mais.



Faça a portabilidade e a otimização de aplicativos com facilidade para o AMD Instinct com a plataforma ROCm

- Permite a portabilidade de software com baixo atrito para a AMD. O HIP permite que os desenvolvedores criem aplicativos portáteis para o AMD Instinct MI300X a partir de um único código fonte
- Oferece todos os recursos com as bibliotecas e os operadores mais usados
- Faz rapidamente a portabilidade e a otimização da maioria dos aplicativos para aumentar o desempenho

Aumentando o suporte upstream

A plataforma ROCm oferece suporte upstream para as estruturas atuais mais populares, incluindo, sem limitação:

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



Economize seu tempo valioso com a portabilidade de código aberto

A plataforma ROCm oferece tranquilidade aos desenvolvedores que podem estar receosos com o novo ecossistema e o impacto que ele pode ter sobre o código existente. Uma das principais vantagens da plataforma ROCm é que você pode fazer a portabilidade de aplicativos com o mínimo de retrabalho no código existente. A portabilidade é simples, permitindo que você economize muito tempo.

É possível fazer a portabilidade dos aplicativos existentes para o acelerador AMD Instinct MI300X de quatro maneiras diferentes:

1 Aproveite as estruturas de IA existentes com suporte upstream para a plataforma ROCm.

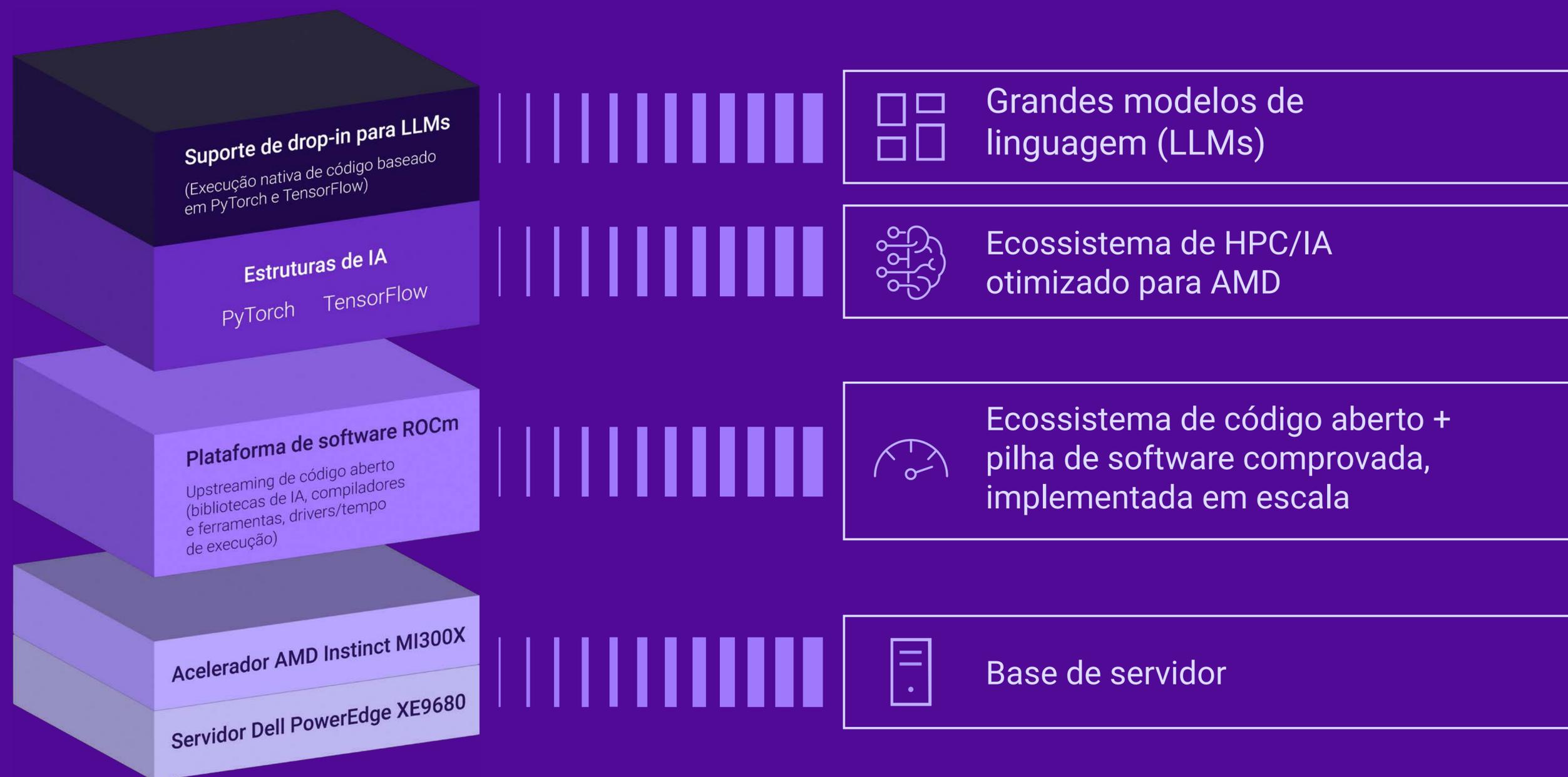
2 Use kernels de ML integrados, otimizados para o AMD Instinct MI300X. Ou aproveite os benefícios da ferramenta de conversão AMD HIP, que permite otimizar aplicativos da arquitetura CUDA em GPUs baseadas na plataforma ROCm.

3 Use versões espelhadas de repositórios existentes de software de biblioteca da arquitetura CUDA, que são otimizados e mantidos para ROCm.

4 Acesse contêineres e aplicativos de IA predefinidos da AMD.

Comece hoje mesmo

Acelere a IA existente e os novos projetos de GenAI com uma estrutura de código aberto da Dell Technologies e da AMD. Potencialize suas cargas de trabalho em uma base comprovada do Dell PowerEdge, compatível com os aceleradores AMD mais recentes, inclusive com o acelerador AMD Instinct MI300X. [Visite o hub de desenvolvedores da plataforma ROCm.](#)



O acelerador AMD Instinct MI300X está pronto para ser implementado

Oferecendo desempenho inovador, o acelerador AMD Instinct MI300X entrega a velocidade, a escalabilidade e o ecossistema de código aberto flexível que são necessários atualmente, sobretudo no momento em que as empresas enfrentam desafios de recurso, complexidade, latência e arquitetura. O acelerador AMD Instinct MI300X oferece uma vantagem distinta em termos de:



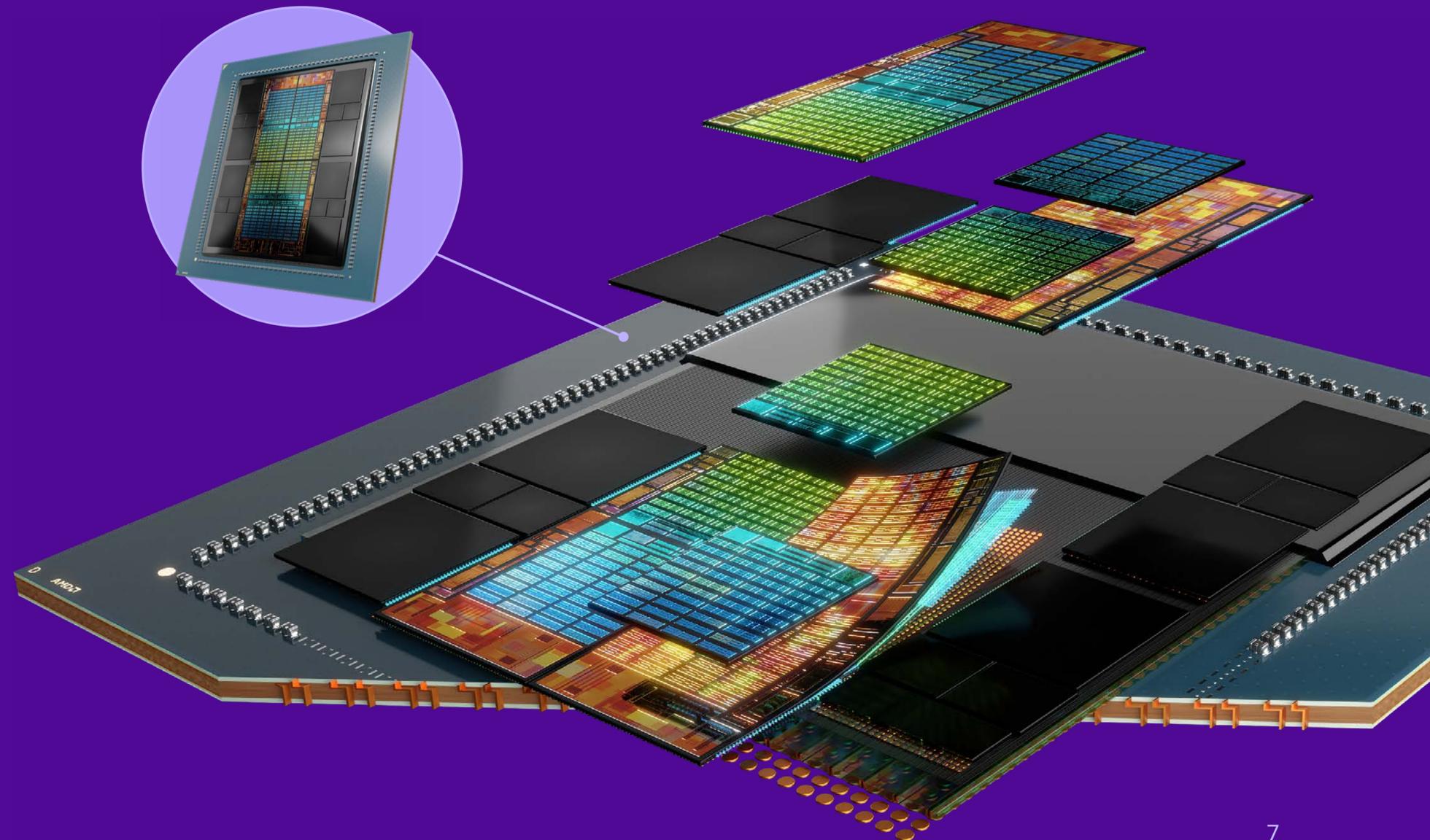
Desempenho: Habilite novos fluxos de trabalho de IA com os 192 GB sem precedentes de memória HBM3 empilhada por GPU operando a 5,3 TB por segundo.



Densidade: Reduza a necessidade de movimentação de dados com as tecnologias de chiplet e empilhamento de matriz de última geração, que acomodam mais capacidade de computação e memória por servidor.



Abertas: Facilite o caminho até a obtenção de resultados com um ecossistema projetado para acessibilidade e capacidade de adaptação.



Potencialize sua vantagem de IA com o servidor Dell PowerEdge XE9680

Você pode acelerar seu caminho para a obtenção de resultados mais rápidos e mais inteligentes com uma plataforma de servidor desenvolvida especialmente para a IA. Aprimore os negócios orientados por dados, entregue resultados visíveis e agilize as operações de IA em toda a empresa com uma solução que acelera os insights, oferece uma IA confiável e simplifica as operações.



IA confiável

Reduza os riscos e economize tempo conforme acelera o ciclo de vida de IA com uma infraestrutura de soluções confiável e de alta qualidade. O PowerEdge XE9680 oferece proteção integrada desde a raiz de confiança de silício até a retirada de serviço dos ativos.



Operações simplificadas

Controle e gerencie com eficiência sua infraestrutura de IA e suas cargas de trabalho de desempenho, em qualquer lugar, com a automação da infraestrutura de IA. A Dell Technologies oferece sistemas inteligentes que funcionam em conjunto e de forma independente para acelerar as tarefas e simplificar sua experiência.

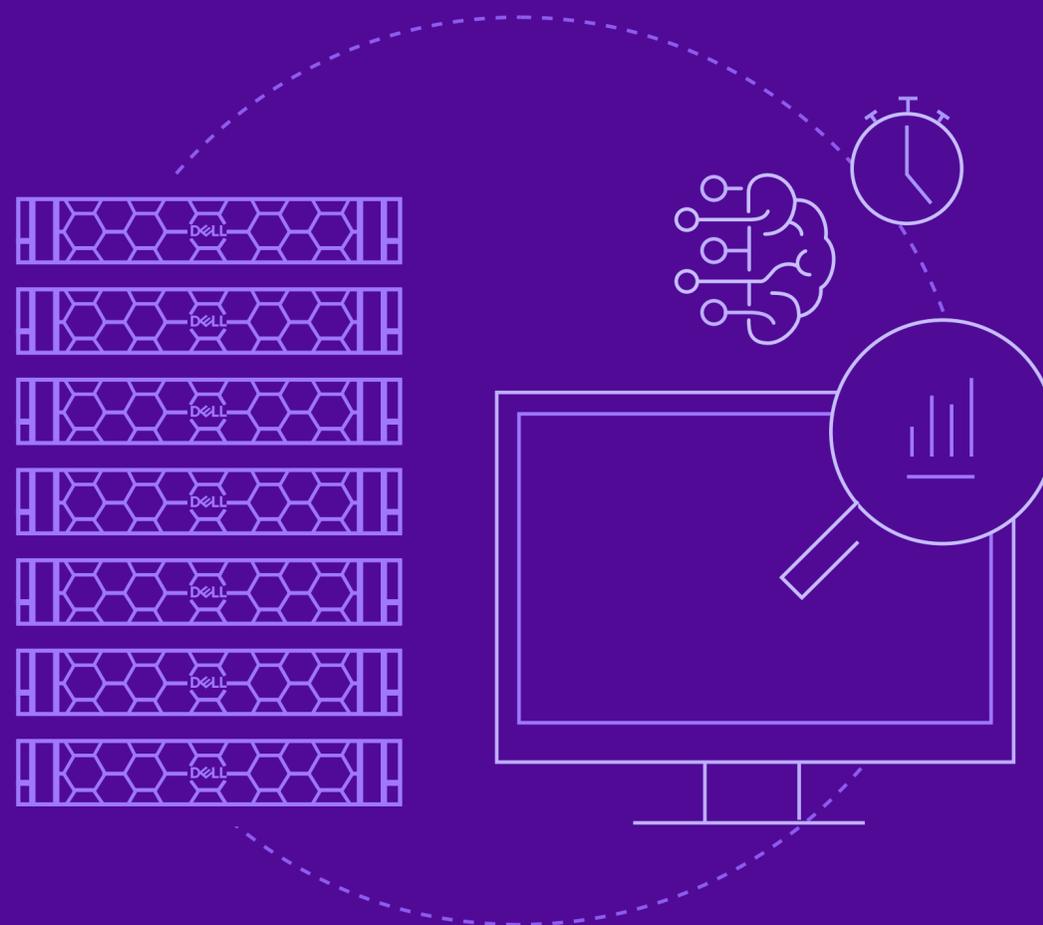


Percepções aceleradas

Aprimore o desempenho em todo o ciclo de vida da IA e a entrega das operações de IA na velocidade dos negócios. A Dell Technologies entrega computação inovadora para ajudar você a aproveitar a IA com o desempenho máximo.

Impulsione a inovação

A Dell Technologies e a AMD estão trabalhando arduamente para impulsionar a inovação aos seus valiosos clientes. Juntas, elas trazem a diversidade do silício para o mercado em uma solução viável, que capacita os desenvolvedores. Essas soluções conjuntas permitem que você:



- **Entre no mercado mais rapidamente** com mais opções e menos obstáculos.
- **Impulsione insights mais rapidamente** com um ecossistema de código aberto sem limitações.
- **Escolha o que é melhor para você.** Conte com a vantagem da cadeia de suprimentos Dell para colocar continuamente tecnologia inovadora no mercado quando for mais necessário.

Avance em velocidade máxima com a Dell Technologies e a AMD

A IA está aqui agora, mas a influência dela vai expandir e acelerar apenas avançando. Para as empresas que desejam aproveitar os recursos dela, as possibilidades parecem infinitas.

É por isso que você não deseja que o seu potencial fique limitado por causa de obstáculos tecnológicos que podem prejudicá-lo. Em vez disso, mantenha suas opções abertas com um pacote de soluções avançadas dedicado às cargas de trabalho de IA.

O acelerador AMD Instinct MI300X e a plataforma de software ROCm no servidor Dell PowerEdge XE9680 é a resposta. Abra o caminho para novos aplicativos, entre no mercado com mais rapidez e comece a gerar resultados importantes para seus negócios e para o mundo todo.

Saiba mais em Dell.com/GPU

Copyright © 2023 Dell Inc. ou suas subsidiárias. Todos os direitos reservados. Dell e outras marcas comerciais pertencem à Dell Inc. ou às suas subsidiárias. AMD, Instinct™, ROCm™ e as respectivas combinações são marcas comerciais da Advanced Micro Devices, Inc. PyTorch® é uma marca comercial ou marca registrada da PyTorch ou dos licenciadores da PyTorch. TensorFlow™ e quaisquer marcas relacionadas são marcas comerciais do Google Inc. CUDA® é uma marca comercial e/ou marca registrada da NVIDIA Corporation nos EUA e em outros países. Outras marcas comerciais pertencem a seus respectivos proprietários. eBook publicado nos EUA em dezembro de 2023.

A Dell Technologies assegura que as informações apresentadas neste documento estão corretas na data da publicação. As informações estão sujeitas a alterações sem aviso prévio.