

Da CPU à GPU: construindo o data center certo para as necessidades de workloads de IA



Você está pronto para aproveitar a oportunidade da IA?

Se você quer executar workloads exigentes de inferência de IA, precisa de uma plataforma confiável e de alto desempenho, capaz de atender a um conjunto de necessidades em evolução, equilibrando potência e eficiência.

E você precisa disso rapidamente.

A IA não vai esperar

96%

dos CIOs antecipam aumento no investimento em IA.¹

79%

dos líderes dizem que suas organizações precisam adotar IA para se manter competitivas.²



Data centers compartilham capacidades

Seja em ambiente interno ou contratado, público ou privado, as organizações dependem dos datacenters para impulsionar a IA em toda a operação.

- Modelos públicos
- Serviços públicos, como Chat-GPT e Gemini
- Modelos privados desenvolvidos e treinados internamente
- Modelos licenciados, empresariais e setoriais

Como você vai alimentar uma IA cada vez maior?

CPUs - Os CPUs mais recentes estão sendo projetados para IA. Eles podem ser mais eficientes do que os GPUs para desenvolvimento, testes e treinamento de modelos de IA de pequeno e médio porte.



GPUs - Os GPUs podem ser mais rápidos e reduzir o custo total de propriedade, ou TCO, no desenvolvimento e treinamento de modelos maiores.



Tenha o melhor dos dois mundos

Por que fazer concessões? Adote uma plataforma flexível que ofereça todos os benefícios dos CPUs, com capacidade de escalar de forma eficiente para o poder dos GPUs conforme necessário.

Processadores AMD EPYC™ e aceleradores AMD Instinct™ atendem ao espectro de workloads de IA

Implemente com a confiança de padrões abertos e recursos de segurança líderes.

Os aceleradores AMD Instinct (GPU) e processadores AMD EPYC (CPU) foram projetados para se destacar em diferentes workloads de IA:

Processadores AMD EPYC CPU são ideais para:

- Implantações de inferência com workloads mistos
- Machine learning clássico
- Modelos de pequeno e médio porte
- Inferência em lote, offline e em pequena escala em tempo real

Aceleradores AMD Instinct GPUs são ideais para:

- Treinamento de IA
- Implantações dedicadas de IA
- Modelos de Gen AI de médio e grande porte
- Inferência em tempo real em grande escala

ThinkSystem SR685a V3

- Servidor potente de 8U
- Conta com dois processadores AMD EPYC™
- Inclui até 8 aceleradores AMD Instinct™

ThinkSystem SR675 V3

- Servidor em rack versátil de 3U com alta densidade de GPU
- Conta com um ou dois processadores AMD EPYC™
- Inclui Lenovo Neptune® hybrid liquid-to-air cooling³

O host de CPU certo é essencial para maximizar o valor de grandes investimentos em GPU

Parceiros perfeitos

Os AMD EPYC CPUs complementam os GPUs com:

- Alta frequência e desempenho para grandes volumes de pré-processamento de dados
- Alta largura de banda de memória e alto desempenho de I/O

Casos de uso de IA baseados em CPU

- Sistemas de recomendação
- Reconhecimento e análise de imagem com ML
- IA Generativa para sistemas LLM e SLM de classe empresarial, como chatbot de RH e sumarização de documentos
- Aplicações mistas habilitadas por IA, como banco de dados, colaboração e design aprimorado com IA

A Lenovo e a AMD oferecem liderança em desempenho e eficiência para CPUs e GPUs.

Juntas, vamos trabalhar com você para ajudar a identificar o melhor caminho para resolver seus desafios de IA no data center.

Saiba mais



¹ Microsoft and LinkedIn, "2024 Work Trend Index Annual Report," May 2024

² Microsoft and LinkedIn, "2024 Work Trend Index Annual Report," May 2024

³ NVIDIA HGX H200 4-GPU offering with NVLink and Lenovo Neptune hybrid liquid-to-air cooling

© 2025 Lenovo. Todos os direitos reservados. Lenovo e o logotipo Lenovo são marcas da Lenovo. AMD, o logotipo AMD Arrow, AMD EPYC, AMD Instinct e suas combinações são marcas da Advanced Micro Devices, Inc. Todas as outras marcas pertencem aos seus respectivos proprietários.