

De CPU a GPU: construyendo el data center correcto para las necesidades de workloads de IA



¿Está listo para aprovechar la oportunidad de la IA?

Si quiere ejecutar workloads exigentes de inferencia de IA, necesita una plataforma confiable y de alto desempeño, capaz de responder a un conjunto de necesidades en evolución, equilibrando potencia y eficiencia.

Y necesita eso rápidamente.

La IA no va a esperar

96% de los CIOs anticipa un aumento en la inversión en IA.¹

79% de los líderes dicen que sus organizaciones necesitan adoptar IA para mantenerse competitivas.²



Los datacenters comparten capacidades

Ya sea en un entorno interno o contratado, público o privado, las organizaciones dependen de los datacenters para impulsar la IA en toda la operación.

- Modelos públicos
- Servicios públicos, como Chat-GPT y Gemini
- Modelos privados desarrollados y entrenados internamente
- Modelos licenciados, empresariales y sectoriales

¿Cómo va a impulsar una IA cada vez mayor?

CPUs - Los CPUs más recientes están siendo diseñados para IA. Pueden ser más eficientes que los GPUs para el desarrollo, las pruebas y el entrenamiento de modelos de IA pequeños y medianos.



GPUs - Los GPUs pueden ser más rápidos y reducir el costo total de propiedad, o TCO, en el desarrollo y entrenamiento de modelos más grandes.



Tenga lo mejor de ambos mundos
¿Por qué hacer concesiones? Adopte una plataforma flexible que ofrezca todos los beneficios de los CPUs, con la capacidad de escalar de forma eficiente hacia la potencia de los GPUs cuando sea necesario.

Procesadores AMD EPYC™ y aceleradores AMD Instinct™ cubren el espectro de workloads de IA

Implemente con la confianza de estándares abiertos y funciones líderes de seguridad.

Los aceleradores AMD Instinct (GPU) y procesadores AMD EPYC (CPU) fueron diseñados para destacarse en diferentes workloads de IA:

Procesadores AMD EPYC™ CPU son ideales para:

- Implementaciones de inferencia con workloads mixtos
- Machine learning clásico
- Modelos pequeños y medianos
- Inferencia por lotes, offline y en tiempo real a pequeña escala

Aceleradores AMD Instinct™ GPUs son ideales para:

- Entrenamiento de IA
- Implementaciones dedicadas de IA
- Modelos de Gen AI medianos y grandes
- Inferencia en tiempo real a gran escala

ThinkSystem SR685a V3

- Servidor potente de 8U
- Cuenta con dos procesadores AMD EPYC™
- Incluye hasta 8 aceleradores AMD Instinct™

ThinkSystem SR675 V3

- Servidor en rack versátil de 3U con alta densidad de GPU
- Cuenta con uno o dos procesadores AMD EPYC™
- Incluye Lenovo Neptune® hybrid liquid-to-air cooling³

El host de CPU correcto es esencial para maximizar el valor de grandes inversiones en GPU

Socios perfectos

Los AMD EPYC CPUs complementan a los GPUs con:

- Alta frecuencia y desempeño para grandes volúmenes de preprocesamiento de datos
- Alto ancho de banda de memoria y alto desempeño de I/O

Casos de uso de IA basados en CPU

- Sistemas de recomendación
- Reconocimiento y análisis de imágenes con ML
- IA Generativa para sistemas LLM y SLM de clase empresarial, como chatbot de RR. HH. y resumen de documentos
- Aplicaciones mixtas habilitadas por IA, como bases de datos, colaboración y diseño mejorado con IA



Lenovo y AMD ofrecen liderazgo en desempeño y eficiencia para CPUs y GPUs.

Juntos, trabajaremos con usted para ayudarle a identificar el mejor camino para resolver sus desafíos de IA en el data center.

Conozca más



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 los derechos reservados. Lenovo y el logotipo de Lenovo son marcas de Lenovo. AMD, el logotipo AMD Arrow, AMD EPYC, AMD Instinct y sus combinaciones son marcas de Advanced Micro Devices, Inc. Todas las demás marcas pertenecen a sus respectivos propietarios.