

HPE Cray

5 razones para elegir HPE Cray XD670

Descubre cómo abordar la próxima generación de cargas de trabajo de IA con un rendimiento de vanguardia para el entrenamiento y el ajuste de grandes modelos de IA

HPE Cray XD670 está diseñado y optimizado específicamente para cargas de trabajo de IA que están muy paralelizadas y requieren aceleración por GPU para un rendimiento óptimo. Algunos ejemplos son el entrenamiento y ajuste de la IA, el procesamiento del lenguaje natural y el entrenamiento multimodal y de modelos de lenguaje grandes (LLM). Si eres un proveedor de servicios de IA, o si tu organización ya ha pasado la fase piloto inicial de IA y estás creando y entrenando modelos, o extendiendo la IA a toda la empresa, podrás sacar el máximo provecho de la escala y la potencia de este sistema.

A continuación, enumeramos 5 razones para elegir HPE Cray XD670:

1. Rendimiento revolucionario

MLCommons™, un consorcio líder en ingeniería de IA basado en una filosofía de colaboración abierta para mejorar los sistemas de IA, ha publicado recientemente los resultados de dos conjuntos de estudios comparativos. El estudio comparativo **MLPerf™ Inference v4.0**¹ realiza estudios comparativos de rendimiento de los sistemas de aprendizaje automático (ML) estándar de la industria de una manera representativa, reproducible y neutral en cuanto a la arquitectura. En este estudio comparativo, HPE Cray XD670, basado en las GPU NVIDIA® H100, alcanzó el **primer puesto en procesamiento del lenguaje natural (PLN)** con escenario BERT 99.0 Offline, y obtuvo el mejor desempeño en todas las categorías en las que participó, incluida la IA generativa (GenAI), la visión artificial (VA) y los LLM. El estudio comparativo **MLPerf Training v4.0** mide la rapidez con la que los sistemas pueden entrenar modelos para alcanzar una métrica de calidad objetivo.² En este estudio comparativo, HPE Cray XD670 fue el **segundo sistema de nodo único más rápido** en comparación con otros servidores con 8 GPU NVIDIA H100, tanto en entrenamiento de **PLN** (modelo BERT) y ajuste de **LLM** (modelo Llama2 para LoRA). Estos resultados líderes en una variedad de pruebas comparativas de IA son la demostración del rendimiento superior de HPE Cray XD670 para entornos de IA.

2. Aceleración de GPU de vanguardia

Equipado con **ocho GPU NVIDIA H200** Tensor Core, HPE Cray XD670 aumenta el rendimiento de la IA con una aceleración descomunal por GPU. NVIDIA H200 ofrece 141 GB de memoria HBM3e a 4,8 TB/s. Esto representa casi el doble de la capacidad de su predecesora, la GPU NVIDIA H100 Tensor Core, con 1,4 veces más ancho de banda de memoria.³ Esta memoria, más grande y más rápida, ofrece un rendimiento superior con el mismo perfil de energía que la NVIDIA H100. Esto significa mayor rendimiento con mejor eficiencia energética y menor coste total de la propiedad. Totalmente integrado con la pila NVIDIA, HPE Cray XD670 también está disponible con GPU NVIDIA H100 Tensor Core.

3. Opciones de refrigeración sostenible

HPE Cray XD670 se ofrece como solución de refrigeración por aire o con **refrigeración líquida directa (DLC) plug and play**. Las funcionalidades de refrigeración por aire emplean una transferencia de calor a aire del 100 %, mientras que la refrigeración DLC emplea agua de la instalación para transferir una cantidad sustancial del calor producido al líquido. La opción de DLC proporciona un sistema de TI con refrigeración líquida totalmente contenida e integrada en el rack, que se entrega precargado y listo para funcionar. Cada rack es autónomo y se integra en nuestra estructura con refrigeración líquida. Los racks se pueden rellenar sobre la marcha a un coste lineal. Gracias a su capacidad de refrigeración extremadamente eficiente, la refrigeración líquida es compatible con las últimas tecnologías, lo que permite disponer de más servidores por rack, necesitando menos racks y capturando el calor de forma más eficaz para reducir los requisitos de potencia de refrigeración. Por último, la DLC aumenta la eficiencia energética, lo que te ayuda a alcanzar tus objetivos de sostenibilidad.

¹ [«New MLPerf Inference Benchmark Results Highlight The Rapid Growth of Generative AI Models»](#) («Los resultados del nuevo estudio comparativo MLPerf Inference destacan el rápido crecimiento de los modelos de IA generativa»), MLCommons, 27 de marzo de 2024

² [Resultados de MLCommons MLPerf Training v4.0](#), junio de 2024

³ [«GPU NVIDIA H200 Tensor Core»](#), NVIDIA, noviembre de 2023



4. Escalabilidad para satisfacer las crecientes cargas de trabajo de IA

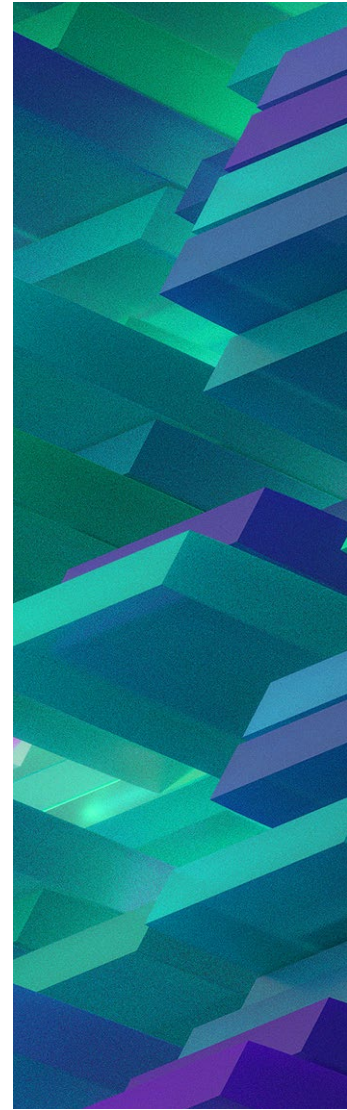
Con el espectacular aumento de los requisitos de computación para cargas de trabajo de IA a gran escala, los proveedores de servicios de IA y las empresas más ambiciosas necesitan entornos de GPU a gran escala. HPE Cray XD670 ofrece una solución completa para IA que puede escalarse sin problemas con una configuración extremadamente densa. Los entornos pueden escalarse desde uno hasta miles de nodos para poder admitir entornos de IA en crecimiento. Los principales proveedores de nube de IA, tanto en EE. UU. como en Europa, están implementando entornos basados en HPE Cray XD670 con decenas de miles de GPU, lo que demuestra la gran escalabilidad de la plataforma.

5. Los mayores conocimientos de su categoría

Los sistemas HPE Cray XD670 cuentan con el respaldo de **HPE Services** y sus décadas de experiencia en la industria relacionada con la planificación, la implementación y la gestión de soluciones. Ofrecemos servicios profesionales, de asesoramiento y operativos que te ayudarán a crear una hoja de ruta hacia tus objetivos específicos e implementar la solución ideal para tus iniciativas.

El servicio de soporte técnico recomendado para HPE Cray XD670 es **HPE Complete Care Service**, un servicio para entornos informáticos modular, desde el extremo hasta la nube e impulsado por IA, diseñado para ayudarte a optimizar todo tu entorno de TI y alcanzar los resultados de TI y los objetivos empresariales establecidos a través de una experiencia personalizada. También puede ayudarte a cuantificar, informar y optimizar el consumo de energía y el impacto ambiental de TI. El servicio HPE Complete Care Service, suministrado por un equipo asignado de expertos de HPE Services, ofrece un enfoque de cobertura completa, una experiencia de gestión de incidencias mejorada con acceso prioritario y una experiencia de cliente digital e impulsada por IA.

Además, obtendrás mayor circularidad en tu entorno de IA con soluciones de gestión de activos y financiación de HPE Financial Services, que te ayudan a ampliar el uso de la tecnología y reducir el impacto medioambiental a largo plazo.



Más información en

[HPE.com/HPE-Cray-XD670](https://www.hpe.com/HPE-Cray-XD670)

Visita **HPE GreenLake**



Chat con Ventas