

Cómo afrontar los desafíos de las cargas de trabajo de IA con la cartera de IA de Dell

Una comparación de la cartera de IA de Dell frente a ofertas similares de HPE

La adopción de la IA presenta nuevos retos para las empresas y el personal de TI del centro de datos, entre otros:

- Abordar las carencias de habilidades existentes en su personal actual mediante formación interna o contratación externa
- Comprender las necesidades de preparación de datos de la IA, lo que incluye la calidad, la cantidad, la ubicación y el estado actual de los datos de la empresa
- Evaluar los objetivos empresariales para determinar qué modelos e implementaciones de IA ofrecerán beneficios
- Evaluar las necesidades de computación, redes y almacenamiento de los sistemas de IA planificados y adquirir estos sistemas

Con carteras preparadas para la IA, los proveedores de infraestructura, como Dell Technologies y HPE, ofrecen soluciones integradas que abarcan todo el ciclo de vida de la IA. Investigamos los datos disponibles públicamente sobre las carteras de IA de Dell y HPE, y examinamos las pruebas de índices de referencia de MLPerf®. Descubrimos que Dell está preparado para ayudar a las empresas a adoptar la IA con una cartera que incluye opciones de almacenamiento y servidores de alto rendimiento, soluciones validadas y servicios profesionales que guían el proceso de la planificación a la producción.

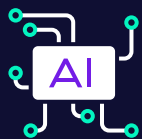
Rendimiento en índices de referencia de modelos de IA: comparación de resultados de MLPerf

Las pruebas de índices de referencia de MLPerf® públicamente disponibles muestran que las ofertas de la cartera de IA de Dell ofrecen un rendimiento coherente y sólido para las cargas de trabajo de IA. MLPerf® prueba el rendimiento tanto para el entrenamiento como para la inferencia en varios modelos de IA. Los datos a los que se hace referencia en este informe se basan en los resultados de MLPerf® v3.1 Inference Datacenter publicados en el sitio web de MLCommons® a partir de noviembre de 2023,¹ Comparamos servidores de 8 y 4 GPU e incluimos un resultado de cada categoría en este resumen. Para verlos todos, lea el informe completo [aquí](#).

* Según análisis de Dell, agosto de 2023. Dell Technologies ofrece soluciones desarrolladas para admitir cargas de trabajo de IA, desde estaciones de trabajo de PC (portátiles y fijos) hasta servidores para computación de alto rendimiento, almacenamiento de datos, infraestructura definida por software nativa de la cloud, switches de red, protección de datos, HCI y servicios. Datos procedentes de: Robert McNeal, "Dell, VMware and NVIDIA Bring AI to Your Data", acceso el 17 de enero de 2024, <https://www.dell.com/en-us/blog/dell-vmware-and-nvidia-bring-ai-to-your-data/>.

1. MLCommons, "MLPerf Inference: Datacenter Benchmark Suite Results", acceso el 12 de diciembre de 2023, <https://mlcommons.org/en/inference-datacenter-31/>.

Nuestro estudio descubrió que Dell tiene:



La cartera de IA generativa más amplia *



Según los resultados de MLPerf®, el servidor Dell PowerEdge XE8640 con cuatro GPU NVIDIA H100 SXM5 logró el mayor rendimiento de IA entre las cuatro GPU presentadas en nueve categorías diferentes



Una cartera más amplia de servicios profesionales adaptados a la IA

Comparación de rendimiento de MLPerf para servidores de ocho GPU

En los resultados de MLPerf® v3.1 para servidores de ocho GPU, Dell PowerEdge XE9680 con GPU NVIDIA SXM5 H100 superó a HPE ProLiant XL675d Gen10 Plus con GPU NVIDIA SXM4 A100 hasta 4,25 veces (véase la figura 1).

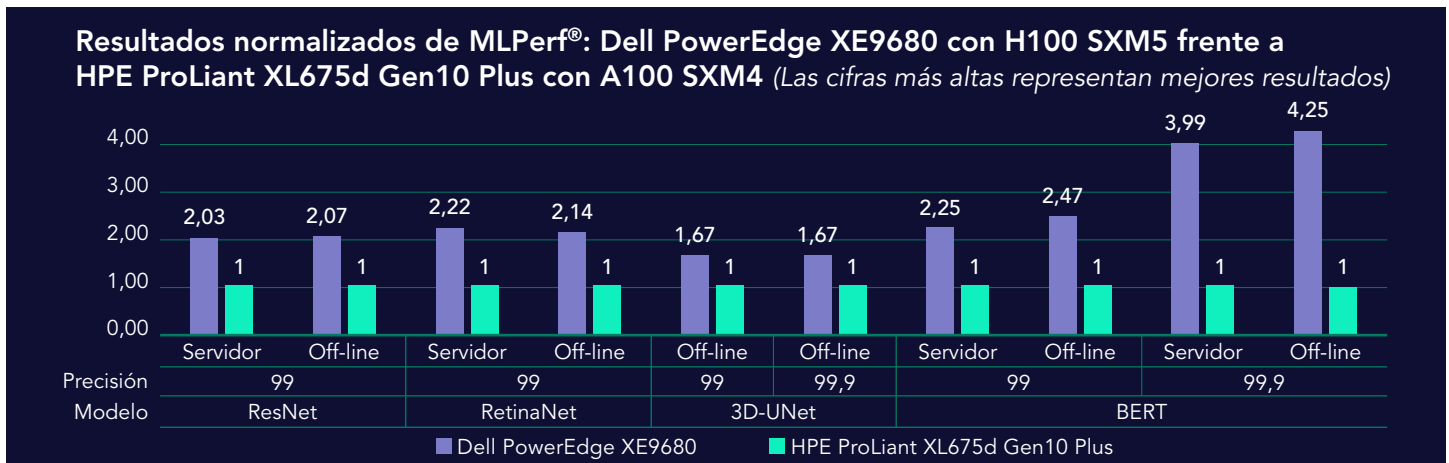


Figura 1: Resultados publicados de MLPerf® para Dell PowerEdge XE9680 y HPE ProLiant XL675d Gen10 Plus a fecha de 29/11/23. El sistema Dell utiliza la GPU NVIDIA H100, mientras que las GPU del sistema HPE son de una generación más antiguas. Fuente: Principled Technologies utilizando datos de MLCommons®.*

Comparación de rendimiento de MLPerf para servidores de cuatro GPU

El servidor Dell PowerEdge XE8640 con cuatro GPU NVIDIA H100 SXM5 logró el mayor rendimiento de IA entre las cuatro GPU presentadas en nueve categorías diferentes. Como se muestra en la figura 2, en comparación con el servidor HPE ProLiant DL380a obtuvo una puntuación hasta 2,07 veces superior en las pruebas de índices de referencia de MLPerf® en comparación con otros modelos evaluados.

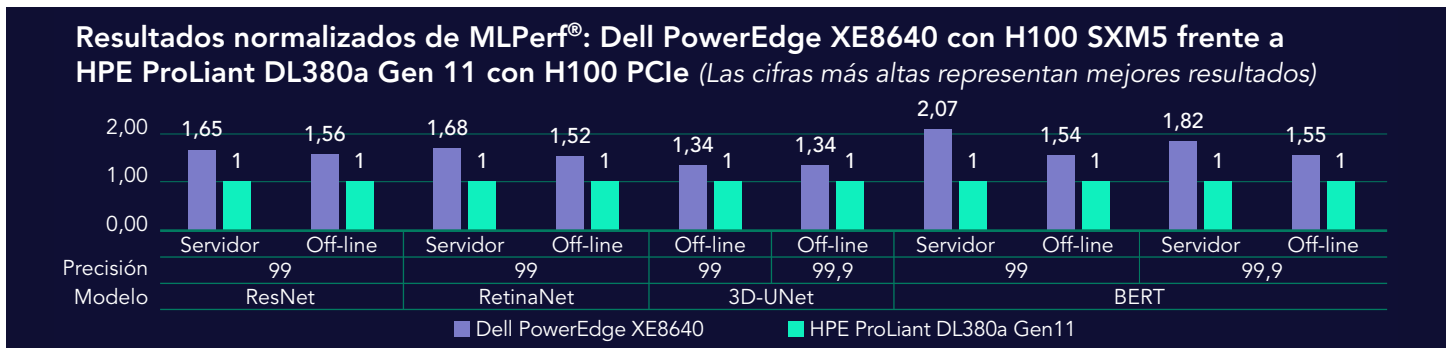


Figura 2: Resultados publicados de MLPerf® para Dell PowerEdge XE8640 y HPE ProLiant DL380a Gen11 a fecha de 29/11/23. El sistema Dell utiliza el factor de forma NVIDIA H100 SXM, mientras que el sistema HPE utiliza el factor de forma PCIe menos potente. Fuente: Principled Technologies utilizando datos de MLCommons®.*

En resumen: Contribuyentes adicionales al éxito de la IA

Los modelos de IA requieren más que servidores de alto rendimiento para el éxito. Para planificar, preparar, implementar y gestionar su solución de IA, tendrá que considerar también el almacenamiento para datos no estructurados y los servicios profesionales. La cartera de Dell ofrece almacenamiento para conjuntos de datos de IA con la serie PowerScale para almacenamiento de archivos y Elastic Cloud Storage u ObjectScale Storage para almacenamiento de objetos.

Las organizaciones pueden aprovechar las ventajas de los servicios profesionales y de consultoría de Dell para IA como la preparación de datos, ofrecida por algunos servicios pero no por HPE. La cartera de Dell también incluye diseños validados para IA, lo que elimina las conjeturas de diseño e implementación de este modelo.

Lea el informe ►

*Puntuación de MLPerf. verificada de v3.1 Inference Closed. Recuperado de <https://mlcommons.org/benchmarks/inference-datacenter/> 5 de diciembre de 2023, apartados 3.1-0069, 3.1-0085, 3.1-0067 y 3.1-0084. El nombre y el logotipo MLPerf son marcas registradas y no registradas de MLCommons. Asociación en los Estados Unidos y en otros países. Todos los derechos reservados. El uso no autorizado está estrictamente prohibido. Consulte www.mlcommons.org para obtener más información.

► Consulte la versión original en inglés de este informe en <https://facts.pt/PiSgyY7>



Facts matter.®

Principled Technologies es una marca comercial registrada de Principled Technologies, Inc. El resto de los nombres de productos son las marcas comerciales de sus respectivos propietarios. Para obtener más información, consulte el informe.