

Soluciones Ethernet abiertas de Dell Technologies para IA generativa

Navegar por las nuevas fronteras de la infraestructura de TI

"Se prevé que el tejido de IA (conmutación de backend para conectividad de GPU a GPU) crezca de 1200 millones de USD (2022) a 15 200 millones de USD (2027), con una TCAC a cinco años del 65 %.

Se espera que Ethernet alcance una cuota de ingresos del 32 % y un 37 % de envíos de puertos para tejidos de IA (2027)".

Investigación de Dell'Oro¹

El rendimiento de la GPU depende en gran medida del rendimiento de la red. Dado que muchas cargas de trabajo de IA se ejecutan en grandes clústeres de servidores que requieren una comunicación constante entre los nodos de computación y el almacenamiento, necesitan redes sólidas para evitar cuellos de botella. Si el rendimiento de la red es insuficiente para la carga de trabajo, las GPU quedarán inactivas y aumentarán los tiempos de entrenamiento e inferencia, lo que ralentizará el procesamiento de datos y el tiempo de obtención de información.

Introducción: las demandas de red de la IA generativa

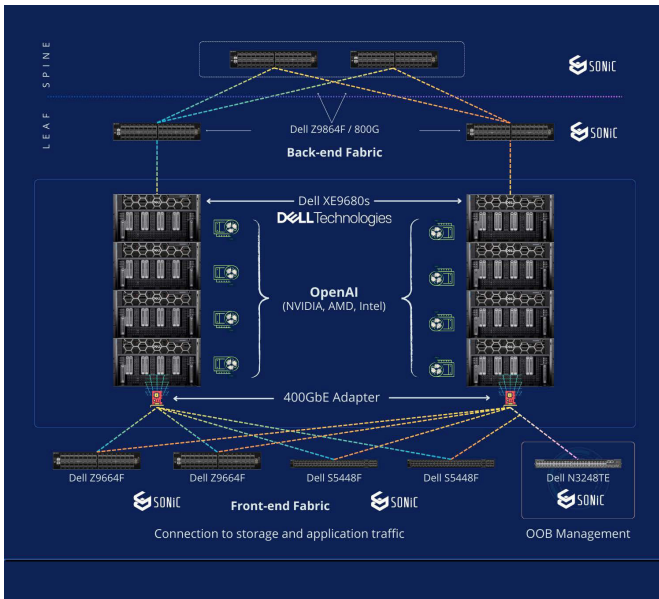
A medida que las soluciones de IA generativa (GenAI) continúan evolucionando, ampliando los límites del procesamiento de datos y las necesidades computacionales, las infraestructuras de TI se enfrentan al reto de encontrar formas de satisfacer los inmensos requisitos de estos entornos. Estos modelos, sobre todo en el caso de los modelos grandes de lenguaje (LLM), no solo exigen mayor infraestructura, sino también sistemas cuidadosamente diseñados para satisfacer las enormes necesidades de conectividad en los clústeres de GPU. Las soluciones de red tradicionales se están convirtiendo rápidamente en cuellos de botella, amenazando la viabilidad y el éxito de las iniciativas de IA generativa. Los tejidos de IA requieren baja latencia, rendimiento sin pérdidas y el máximo ancho de banda. Los requisitos de procesamiento masivo de datos y aplicaciones incrementan los requisitos tanto para los tejidos de front-end como para los de back-end.

Retos y necesidades de la implementación de infraestructura de IA generativa

El despliegue de tecnologías de IA generativa presenta una serie de desafíos, desde las complejidades técnicas asociadas con las nuevas arquitecturas hasta la escasez de profesionales capacitados para gestionar implementaciones de estas características. Las soluciones que se basan en tecnologías propietarias, como InfiniBand, agregan otra capa de complejidad, lo que limita la disponibilidad de recursos y complica la integración con las plataformas de supervisión o coordinación existentes. Además, los altos costes, los largos tiempos de evaluación y la dependencia del proveedor asociados a las soluciones propias suponen importantes barreras, especialmente en una era caracterizada por la incertidumbre en la cadena de suministro. Estos retos subrayan la necesidad urgente de contar con soluciones de infraestructura de IA generativa abiertas, flexibles y sólidas que puedan adaptarse a las demandas únicas de las cargas de trabajo de la IA generativa.

Enfoque de Dell Technologies para las redes de IA generativa

En respuesta a estos desafíos, Dell Technologies se ha erigido como pionera en soluciones integrales y abiertas basadas en Ethernet diseñadas para satisfacer las complejas demandas de la infraestructura de IA generativa. Gracias a su amplia experiencia en entornos de IA, modelado e informática de alto rendimiento (HPC), Dell Technologies ofrece un conjunto de soluciones que cumplen los requisitos tanto de front-end como de back-end. Desde sistemas informáticos modulares optimizados para la aceleración, como los servidores Dell PowerEdge XE, hasta soluciones de almacenamiento centradas en la IA, como PowerScale, Dell Technologies proporciona los componentes esenciales para una implementación de la IA generativa con plenas garantías. Un elemento central de este enfoque es la implementación de tejidos Ethernet de última generación con tecnología avanzada de chips de red. Con **Dell PowerSwitch Z9864-ON**, que ofrece **800 GbE** de rendimiento de red sin bloqueos esencial para las aplicaciones de IA generativa, los clientes pueden implementar clústeres de IA con baja latencia y alto rendimiento utilizando conmutación de alto ancho de banda y nuevas funciones que se encuentran en **Enterprise SONiC Distribution by Dell Technologies**, como Advanced Routing, RoCEv2, Enhanced Hashing y Priority Flow Control, para mejorar el rendimiento del tejido y una mejor supervisión de la congestión.



Ejemplo de arquitectura de tejido con IA generativa

Innovaciones en redes con IA generativa de Dell Technologies

Dell Technologies se encuentra a la vanguardia de la innovación en redes con IA generativa y ofrece soluciones que cumplen los requisitos de los entornos de IA generativa actuales y futuros, del perímetro al núcleo y a la cloud. Al centrarse en soluciones abiertas y ampliables, que hacen uso de chips comerciales y software basado en código abierto, Dell Technologies garantiza un rendimiento y una flexibilidad máximos.

El uso de sistemas operativos de red de código abierto disponibles comercialmente como SONiC, junto con la participación activa y la contribución de Dell Technologies al **Ultra Ethernet Consortium (UEC)**, subraya su compromiso con los estándares abiertos y el desarrollo colaborativo en el espacio Ethernet. Estos esfuerzos tienen como objetivo garantizar que Ethernet siga desempeñando un papel fundamental en el soporte de la próxima generación de entornos de IA.

La serie Z de Dell PowerSwitch, que utiliza chips de vanguardia, proporciona la estructura central para una red ampliable de alto rendimiento capaz de admitir miles de nodos, abordando así los desafíos de conectividad inherentes a las aplicaciones de IA generativa.

Aceleración de la implementación de la IA generativa con Dell Technologies

El auge de la IA generativa ha traído consigo una serie de retos para las infraestructuras informáticas, lo que exige un nuevo enfoque de las redes que sea a la vez innovador y flexible. Dell Technologies responde a esta llamada con soluciones abiertas con tecnología Ethernet que no solo satisfacen las necesidades inmediatas de las implementaciones de IA generativa, sino que también sientan las bases para futuros avances.

Para ayudar a erradicar las conjeturas de las soluciones de hardware de IA, Dell ofrece arquitecturas de referencia validadas en laboratorio y optimizadas para cargas de trabajo de IA. Estos diseños validados incluyen validaciones de conceptos arquitectónicos, descripciones generales completas de la solución y de rendimiento, junto a otras validaciones de laboratorio que demuestran las capacidades de la solución aplicadas a la carga de trabajo para la que se diseñó. Pase de una IA posible a una IA probada con soluciones validadas que facilitan una entrega más rápida de información más útil.

Al elegir Dell Technologies, las organizaciones obtienen un socio con la experiencia, la pila de soluciones de punto a punto y el compromiso necesarios para garantizar el éxito de sus iniciativas de IA generativa. Con Dell Technologies, las empresas están equipadas para sortear las complejidades de las arquitecturas de IA generativa, garantizando así que sus proyectos no solo sean viables, sino que estén posicionados para alcanzar el éxito.

Acelere el tiempo de implementación y rentabilización de sus entornos de IA generativa y reduzca el riesgo y la complejidad operativa con Dell Technologies. Le invitamos a descubrir cómo una solución de red abierta, flexible y sostenible puede transformar sus iniciativas de IA generativa e impulsar su negocio hacia una nueva era de innovación y eficiencia.



Más información
acerca de Dell
Networking



Póngase en
contacto con un
experto de Dell
Technologies



Lea el Informe
del analista de
ESG



Lea el Informe
del analista de
IDC