[notasdeprensa.jpg](http://www.mexicopress.com.mx)Publicado en Ciudad de México, México a 16 de noviembre de 2020. el 16/11/2020

# [NVIDIA anuncia la GPU A100 de 80 GB, que potencia la GPU más potente del mundo para Supercómputo con IA](http://www.notasdeprensa.es)

## Los principales proveedores de sistemas, como: Atos, Dell Technologies, Fujitsu, GIGABYTE, Hewlett Packard Enterprise, Inspur, Lenovo, Quanta y Supermicro, ofrecerán sistemas NVIDIA A100 a las industrias

El día de hoy, NVIDIA presentó la GPU NVIDIA A100 de 80 GB, la innovación más actual que potencia la plataforma de súper cómputo NVIDIA HGX™. Cuenta con el doble de memoria que la generación anterior, por lo que les brinda a los ingenieros e investigadores un rendimiento y una velocidad sin precedentes para desbloquear la próxima ola de avances científicos y de IA. La nueva A100 con la tecnología HBM2e duplica la memoria de alto ancho de banda de la GPU A100 de 40 GB a 80 GB. Además, ofrece más de 2 terabytes por segundo de ancho de banda de memoria. Esto permite que los datos se envíen rápidamente a la A100, la GPU de centro de datos más rápida del mundo. De esta manera, los investigadores pueden acelerar sus aplicaciones todavía más y trabajar con modelos y conjuntos de datos aún más grandes. “Lograr resultados de vanguardia en la investigación de HPC e IA requiere el desarrollo de modelos más grandes; sin embargo, éstos exigen más capacidad de memoria y ancho de banda que nunca”, dijo Bryan Catanzaro, Vicepresidente de Investigación de Deep Learning Aplicado de NVIDIA. “La GPU A100 de 80 GB proporciona el doble de memoria que su predecesora, que se lanzó hace solo seis meses, y rompe la barrera de los 2 TB por segundo, lo que permite a los investigadores abordar los desafíos científicos y de big data más importantes del mundo”. La GPU NVIDIA A100 de 80 GB está disponible en los sistemas NVIDIA DGXTM A100 y NVIDIA DGX StationTM A100, que también se anunciaron hoy y se esperan que se envíen este trimestre. Se espera que los principales proveedores de sistemas, como Atos, Dell, Fujitsu, GIGABYTE, Hewlett Packard Enterprise, Inspur, Lenovo, Quanta y Supermicro, comiencen a ofrecer sistemas desarrollados con las placas base integradas HGX A100 con configuraciones de cuatro (4) u ocho (8) GPUs A100 de 80 GB en la primera mitad del 2021. Impulsar cargas de trabajo hambrientas de datos. La GPU A100 de 80 GB se basa en las diversas capacidades de la A100 de 40 GB y es ideal para realizar una amplia gama de aplicaciones con enormes requisitos de memoria de datos. Para el entrenamiento de IA, los modelos de sistemas de recomendación como DLRM tienen enormes tablas que representan miles de millones de usuarios y cientos de millones de productos. El A100 de 80 GB ofrece una aceleración de hasta 3 veces, por lo que las empresas pueden reentrenar rápidamente estos modelos para brindar recomendaciones altamente precisas. El A100 de 80 GB también permite el entrenamiento de los modelos más grandes con más parámetros que se ajustan a un solo servidor con HGX, como GPT-2, un modelo de procesamiento de idiomas naturales con capacidad de texto generativo sobrehumano. Esto elimina la necesidad de datos o modelos de arquitecturas paralelas que pueden llevar mucho tiempo de implementar y ralentizar su ejecución en varios nodos. Gracias a su tecnología de GPU de instancias múltiples (MIG), la A100 se puede particionar en hasta siete instancias de GPU, cada una con 10 GB de memoria. Esto proporciona un aislamiento de hardware seguro y maximiza la utilización de la GPU para una variedad de cargas de trabajo más pequeñas. Para la inferencia de IA de modelos de reconocimiento automático de voz como RNN-T, una sola instancia de MIG A100 de 80 GB puede dar servicio a lotes mucho más grandes, lo que ofrece un rendimiento de inferencia 1.25 veces mayor en producción. En una evaluación del análisis de big data para la industria del comercio minorista, en el rango de tamaño de terabytes, la A100 de 80 GB aumenta el rendimiento hasta 2 veces, lo que la convierte en una plataforma ideal para brindar información rápida sobre los conjuntos de datos más grandes. Las empresas pueden tomar decisiones clave en tiempo real, ya que los datos se actualizan de forma dinámica. Para aplicaciones científicas, como la previsión meteorológica y la química cuántica, la A100 de 80 GB puede ofrecer una enorme aceleración. Quantum Espresso, una simulación de materiales logró un rendimiento cerca de 2 veces con un solo nodo de A100 de 80 GB. “La capacidad y el ancho de banda de memoria amplios y rápidos son vitales para lograr un alto rendimiento en aplicaciones de supercomputación”, dijo Satoshi Matsuoka, Director del Centro RIKEN de Ciencias Computacionales. "La NVIDIA A100 con 80 GB de memoria de GPU HBM2e, que proporciona el ancho de banda de 2 TB por segundo más rápido del mundo, ayudará a ofrecer un gran impulso en el rendimiento de las aplicaciones". Características clave de A100 de 80 GB. La A100 de 80 GB incluye las diversas e innovadoras características de la arquitectura NVIDIA Ampere: - Tensor Cores de 3.ª generación: Proporciona un rendimiento de IA hasta 20 veces superior, en comparación con la generación anterior de Volta, gracias a un nuevo formato TF32, 2.5 veces superior en FP64 para HPC y 20 veces superior en INT8 para la inferencia de AI. También admite el formato de datos BF16. - Memoria de GPU HBM2e más grande y rápida: Duplica la capacidad de memoria y es el primer dispositivo de la industria en ofrecer más de 2 TB por segundo de ancho de banda de memoria. - Tecnología MIG: Duplica la memoria por instancia aislada, para proporcionar hasta siete MIGs con 10 GB cada uno. - Baja densidad estructural: Ofrece hasta el doble de aceleración en las inferencias de modelos con baja densidad. - NVLink y NVSwitch de tercera generación: Proporciona el doble de ancho de banda de GPU a GPU que la tecnología de interconexión de la generación anterior, lo que acelera las transferencias de datos a la GPU para cargas de trabajo con uso intensivo de datos a 600 gigabytes por segundo. Plataforma de Súper Cómputo NVIDIA HGX AILa GPU A100 de 80 GB es un elemento clave de la plataforma de supercomputación NVIDIA HGX AI, que reúne toda la potencia de las GPUs de NVIDIA, NVIDIA NVLink®, las redes NVIDIA InfiniBand y una pila de software NVIDIA de IA y HPC totalmente optimizada para proporcionar el mayor rendimiento de las aplicaciones. Permite a los investigadores y científicos combinar HPC, análisis de datos y métodos de computación de deep learning para avanzar en el progreso científico. Más información sobre NVIDIA A100 de 80 GB en la Charla especial de NVIDIA SC20 en vivo, que se realizará a las 5:00 p. m. Hora de la CDMX del día de hoy. # # # Acerca de NVIDIAEl invento de NVIDIA (NASDAQ: NVDA) en 1999 generó el crecimiento del mercado de juegos para PC, redefinió los gráficos por computación modernos y revolucionó la computación en paralelo. Recientemente, el deep learning de GPU impulsó la inteligencia artificial moderna, la próxima era de la computación, ya que la GPU ocupa el papel del cerebro de las computadoras, los robots y los automóviles de conducción autónoma que pueden percibir y comprender al mundo. Obtén más información en http://nvidianews.nvidia.com/.

**Datos de contacto:**

Carlos Valencia

NVIDIA PR Agency.

55 39 64 96 00

Nota de prensa publicada en: [https://www.mexicopress.com.mx/nvidia-anuncia-la-gpu-a100-de-80-gb-que](http://www.notasdeprensa.es/educalivecom-disfruta-de-cursos-presenciales-y-clases-particulares-a-traves-de-internet)

Categorías: Nacional Inteligencia Artificial y Robótica Programación Hardware Software Ciberseguridad

[notasdeprensa.jpg](http://www.mexicopress.com.mx)

[**http://www.mexicopress.com.mx**](http://www.notasdeprensa.es)