

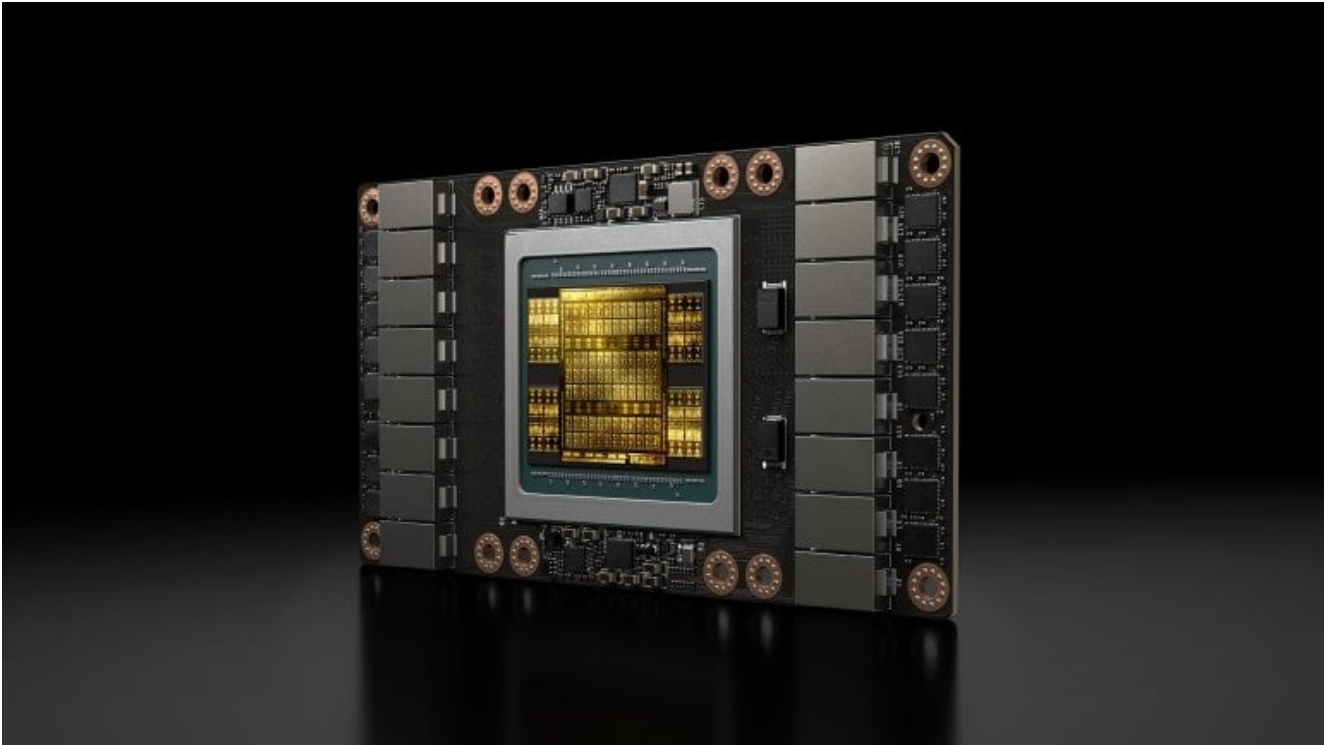
# Summit u OLCF-4, Supercomputadora, superordenador, de IBM, Nvidia

Summit u OLCF-4, Supercomputadora, superordenador, de IBM, Nvidia. El nuevo sistema Summit de DOE presenta una arquitectura única que combina capacidades de computación HPC y AI.

## Introducción

Diseñado por **IBM y Nvidia**, Summit funciona con los **procesadores Power 9** del primero y las **GPU Tensor Core** de este último. La máquina ocupa una gran área del Laboratorio Nacional Oak Ridge del DOE en Tennessee y puede extraer hasta 13 MW de potencia cuando dispara con todos los cilindros.

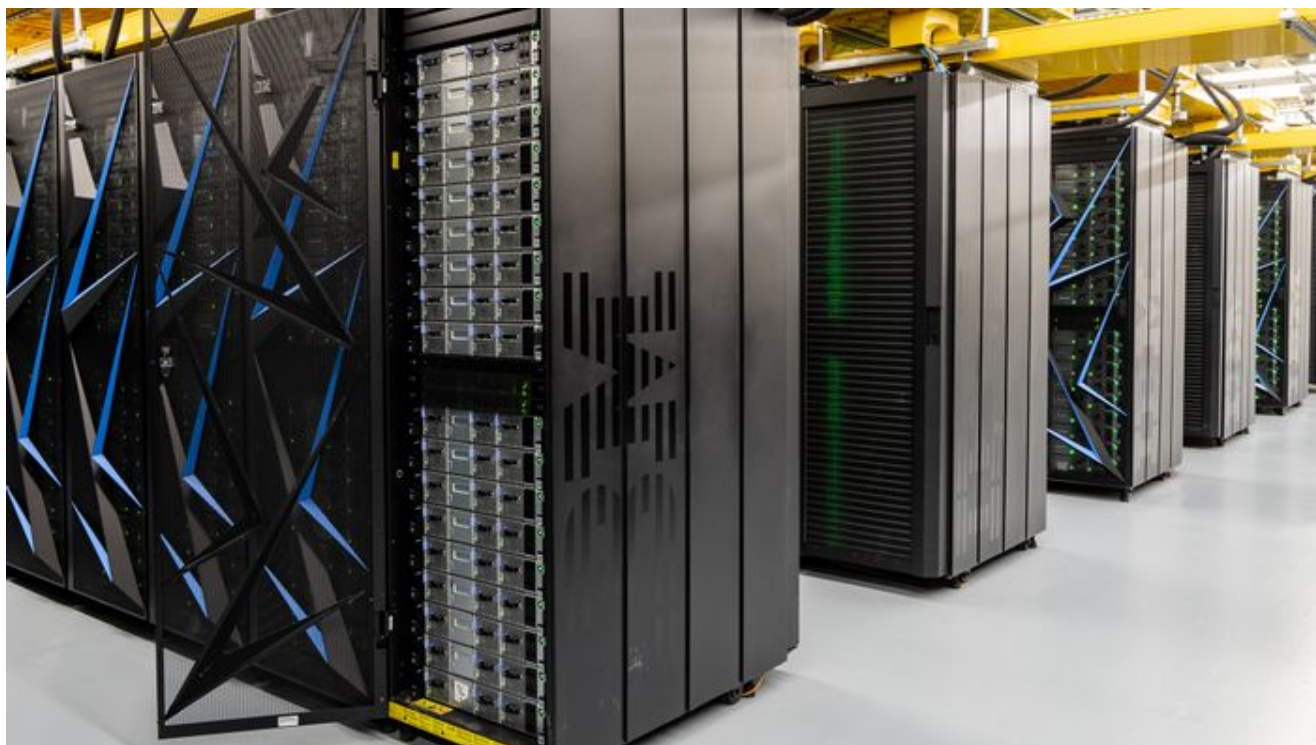
La nueva arquitectura informática desarrollada para el sistema combina las capacidades informáticas de inteligencia artificial con las tradicionales de computación de alto rendimiento. Eso significa que puede usarse para construir modelos científicos y simulaciones, mientras que al mismo tiempo procesa cargas de trabajo de aprendizaje automático, abriendo en gran medida las posibilidades en la investigación científica asistida por supercomputadora.



Tomada de internet

## Algunas características

Los 9,000 pies cuadrados (2.743 m<sup>2</sup> y más de 340 toneladas de peso, algo así como un avión comercial grande) de espacio del centro de datos que Summit ocupa son parte del edificio de ciencias computacionales del laboratorio, renovado específicamente para acomodar el nuevo sistema. El sistema Summit, con 9.216 procesadores IBM impulsados por 27.648 chips gráficos Nvidia, ocupa tanto espacio como dos canchas de tenis y tanta potencia como una ciudad pequeña. Se utilizará para la investigación civil en temas como la ciencia de los materiales, el cáncer, la energía de fusión, la astrofísica y el clima cambiante de la Tierra



amiento bombea más de 4,000 galones de agua a través de sus tuberías para disipar el calor que produce la supercomputadora, según IBM.

Summit puede realizar 200 cuatrillones (200,000 billones) de cálculos por segundo, o 200 petaflops (Hay una carrera mundial hacia un umbral de rendimiento llamado un exaflops: un quintillón de operaciones matemáticas de coma flotante por segundo. IBM cree que puede alcanzar ese nivel, pero aún queda camino por recorrer). Summit está conectado por 185 millas de cables de fibra óptica, que es la distancia que separa a Knoxville de Nashville, Tennessee.

Se trata del superordenador más poderoso e inteligente del mundo, así como de la supercomputadora acelerada por GPU más grande del mundo.

Las GPU se han usado durante mucho tiempo para acelerar las cargas de trabajo HPC tradicionales. En los últimos años, sin embargo, también han surgido como el chip de aceleración para el entrenamiento de modelos de aprendizaje profundo para aplicaciones de [inteligencia artificial](#). Siendo el proveedor número uno de GPU, Nvidia ha sido el principal beneficiario de esta tendencia.

## 28,000 GPU

El fabricante de chips con sede en Santa Clara, California, gastó más de \$ 3 mil millones en el desarrollo de la GPU Volta Tensor Core que impulsa Summit, dijo un vocero de Nvidia. Según Kharya, las GPU son responsables del 95 por ciento del rendimiento del sistema.

El clúster de cómputo consiste en aproximadamente 4,600 nodos, con dos CPU y seis GPU por nodo, para un total de 27,648 GPU, según Nvidia. Las GPU están interconectadas con las CPU Power9 más recientes de IBM, diseñadas específicamente para cargas de trabajo HPC y AI , mediante la nueva tecnología de interconexión NVLink de Nvidia.

Nvidia e IBM trabajaron juntas para unir sus chips a través de NVLink, que entrega velocidades de hasta 300 gigabytes por segundo, o 10 veces más rápido que PCIe, dijo Kharya.

## Dentro de un nodo de Summit

Summit promete entregar 5-10 veces más rendimiento que su predecesor, Titan, pero genera mucha más energía en una huella más pequeña. Titan presentó 18,688 nodos, pero Summit lo dominará con «solo» ~ 4,600 nodos. Esa capacidad se deriva de un mayor rendimiento del nodo; Summit ofrecerá más de 40 TeraFLOPS por nodo, mientras que cada nodo Titán pesa 1,4 TeraFLOPS.

El empaque de toda esa potencia en un solo nodo comienza con el nodo Power Systems AC922 refrigerado por agua de IBM. Cada nodo está equipado con dos procesadores IBM POWER9 y seis GPU Nvidia Volta GV100 . Los nodos también cuentan con un agregado de 512 GB de DDR4 y HBM2 (High Bandwidth Memory) coherentes junto con 1.600 GB de RAM no volátil.

Los supercomputadores tienen que ver con el cálculo en paralelo y el movimiento de datos entre las CPU, GPU, memoria y redes, por lo que Summit proporciona numerosas capas de

ancho de banda extremo.



Tomada de internet

El sistema presenta 96 carriles de PCIe 4.0 que son útiles para el adaptador Mellanox EDR InfiniBand de doble puerto, que tiene un rendimiento teórico máximo de 400 Gb / s. IBM ha medido el rendimiento en 392 Gb / s, que es dos veces el ancho de banda de un adaptador PCIe 3.0.

El Volta GV100 se conecta a través de PCIe 3.0 y NVLink 2.0. La interfaz de NVLink proporciona 100 GB / s de rendimiento para el tráfico de CPU a GPU y de GPU a GPU. Las GPU están dispuestas en un diseño de doble malla.

Curiosamente, IBM también produce un modelo con cuatro GPU que alimentarán la supercomputadora Sierra de CORAL. El modelo de cuatro GPU (la última imagen en el álbum anterior) ofrece 150GBps para la comunicación entre GPU / CPU. Debido al reducido número de GPU, IBM puede proporcionar más enlaces («ladrillos» en el lenguaje de NVLink) a las CPU y GPU, lo que aumenta el rendimiento.

Los procesadores POWER9 tienen ocho canales de memoria, para un total de 16 canales por servidor que proporcionan 340 GB /



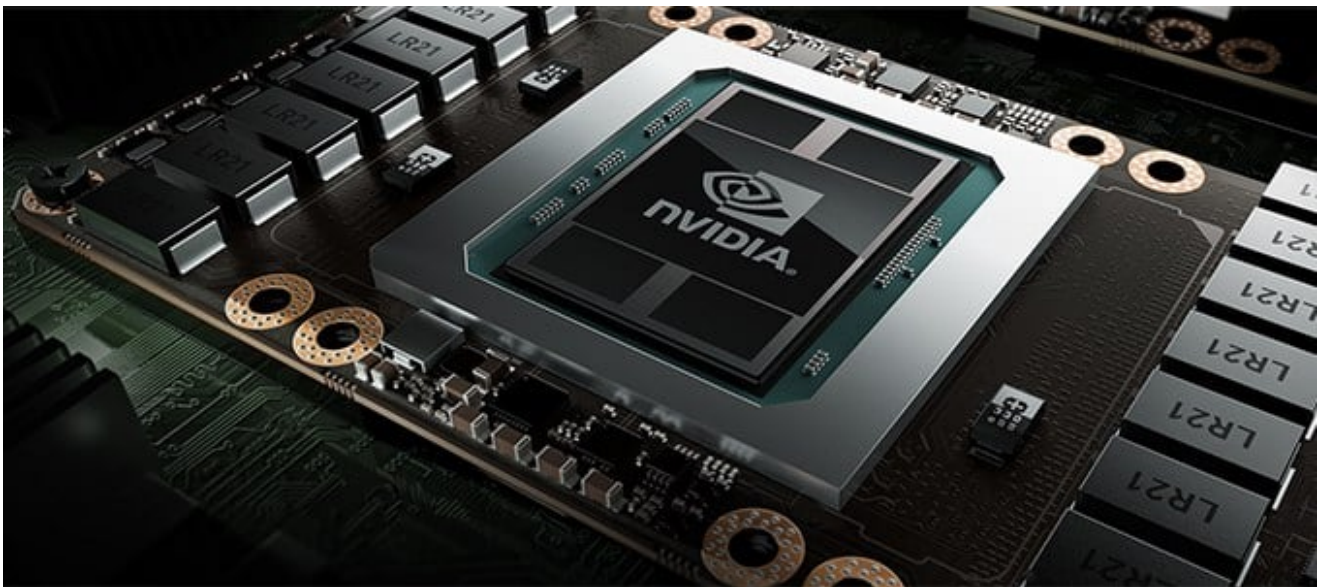
s de ancho de banda agregado. Cada nodo Summit manejará un máximo de 2TB de memoria DDR4-2666.

El pool de almacenamiento 250 PB de Summit, que utiliza el GPFS de IBM (General Parallel File System), distribuye 2.5TB / s de ancho de banda en los 4,600 nodos. Cada nodo presenta 1.600 GB de RAM no volátil que sirve como un búfer de ráfaga. El almacenamiento intermedio absorbe ráfagas de datos antes de transferirlos al grupo de almacenamiento primario remoto. Eso ayuda a eludir la asignación de rendimiento limitado para cada nodo.

## Nvidia Volta GV100

Cada nodo Summit presenta seis GPU Nvidia GV100. En pocas palabras, el Volta GV100 cuenta con una matriz masiva de 815 mm<sup>2</sup> Volta que maneja 21 mil millones de transistores incorporados en el proceso FFN de 12 nm de TSMC. Se complementa con cuatro pilas de HBM2 (16 GB en total). Nvidia envía la GPU con 80 SM activados (5.120 núcleos CUDA).

E  
l  
n  
o  
d  
o  
S  
u  
m  
m  
i  
t



presenta hasta 2TB de memoria coherente a la que se puede acceder ya sea mediante la GPU o la CPU, por lo que el ancho de banda es un requisito fundamental. Nvidia equipó la parte inferior del GV100 con dos conectores mezzanine. Un conector proporciona energía y se adapta al tráfico PCIe y de banda lateral, mientras que el otro está dedicado a la conexión

NVLink 2.0. NVLink es una interfaz independiente que proporciona 6X el ancho de banda de una conexión PCIe 3.0 x16. Los procesadores POWER9 de IBM tienen una conexión integrada NVLink 2.0 en el procesador.

El paquete Volta extrae un promedio de 300W a un poco menos de 1V, por lo que más de 300A fluyen hacia el molde. Eso requiere una solución de refrigeración carnosa. Al igual que con todos los demás componentes informáticos primarios dentro del nodo, los GV100 están refrigerados por agua. Podemos ver el enorme bloque de agua y los agujeros (junto a los elementos de sujeción equipados con resortes) donde el puente, que abarca grupos de tres GV100, se conecta al bloque.

## El procesador IBM POWER9

Las cargas de trabajo de AI cambian la mayor parte del cálculo a la GPU, pero la CPU aún sirve como el mariscal de campo para el procesamiento y movimiento de datos. El POWER9 de IBM está excepcionalmente bien equipado para el puesto. El procesador POWER9-S0 (Scale-Out) de IBM se basa en el proceso FinFET de 14nm de Global Foundries y cuenta con 8 mil millones de transistores. Ese dado grande requiere un paquete masivo de CPU, que hemos destacado con nuestra tarjeta de visita de tamaño estándar. El procesador presenta hasta 24 núcleos que ejecutan 96 hilos. Cada núcleo ejecuta cuatro hilos en paralelo (SMT4), mientras que el modelo de 12 núcleos ejecuta ocho hilos por núcleo (SMT8).

Cada procesador ofrece 48 carriles de PCIe 4.0 que ofrecen 192GB / s de ancho de banda dúplex, y ocho canales de memoria admiten hasta 4TB de memoria DDR4-2666 por socket con 120GB / s de ancho de banda sostenido.

El procesador también cuenta con 120 MB de caché LDR eDRAM compartida (bloques de 10 MB compartidos entre grupos de dos núcleos) con hasta 256 GB / s de rendimiento por núcleo.

El procesador también tiene soporte para Nvidia's NVLink 2.0 y

CAPI 2.0 integrados en el dado. El tráfico de memoria caché, memoria, PCIe y NVLink (entre otros) fluye a través del procesador, lo que requiere una estructura sólida. IBM infundió el procesador con un tejido en chip de 7TB / s.

Leer también: post de supercomputadoras en blog Argentina; [GPU Nvidia Tesla P100](#); [¿qué es la Ley de Moore? ¿De que trata?](#), [Las supercomputadoras más poderosas del mundo](#)