

# Nvidia ouvre la commercialisation de son GPU Tesla V100 pour HPC et IA

Annoncé lors de son événement annuel GTC (GPU Technology Conference) 2017, le [GPU Tesla V100](#) basé sur l'architecture Volta fait ses débuts commerciaux.

Gravée en 12 nm (nanomètres), il s'agit, selon Nvidia, d'une « *solution choisie par tous les grands acteurs Cloud pour l'intelligence artificielle et le calcul haute-performance* ». Parmi les fabricants en lice pour son utilisation figurent ainsi Dell EMC, Hewlett Packard Enterprise (HPE), Huawei, IBM et Lenovo. Ils ont tous annoncé des offres basées sur Volta pour leurs clients.

Les fournisseurs de services cloud ne sont pas en reste avec Alibaba Cloud, Amazon Web Services (AWS), Baidu Cloud, Microsoft Azure, Oracle Cloud et Tencent Cloud, qui ont également exprimé leur intérêt à fournir des services cloud basés sur Volta.

Parallèlement, la firme basée à Santa Clara rend disponible les logiciels et les outils Nvidia GPU Cloud (NGC) destinés aux scientifiques. Ils doivent leur permettre « *d'exploiter pleinement le potentiel de la plateforme informatique accélérée de Nvidia pour la recherche intensive en calcul* ».

## 5120 coeurs Cuda

« *Volta est la plate-forme la plus puissante au monde pour l'IA et le HPC, et permettra aux meilleurs chercheurs du monde de repousser les limites dans les domaines de la découverte de médicaments, des carburants alternatifs et des catastrophes naturelles, se félicite Jensen Huang, P-DG de Nvidia. Avec l'implémentation de Volta dans les centres de calculs et à travers les différents acteurs Cloud du monde entier, une nouvelle vague d'innovation est en marche et aura un impact incroyable sur l'ensemble de la société.* »

La solution graphique Volta embarque 5120 coeurs Cuda cadencés jusqu'à 1455 MHz, avec à la clef 7,8 téraflops de puissance de calcul en double précision. Le double en simple précision.

En plus des coeurs Cuda, 640 Tensor Cores sont embarqués dans la solution en vue de prendre en charge les applications dédiées à l'intelligence artificielle (IA), avec une puissance de calcul de 120 téraflops. Le tout est épaulé par 16 Go de mémoire vive HBM2 (High Bandwidth Memory de seconde génération) avec une bande passante de 900 Go/s. Une puissance sur laquelle AWS (Amazon Web Services) s'est jeté le premier parmi les fournisseur de services Cloud.

Au prochain GTC 2018 prévue fin mars, Volta devrait céder la place à Ampère, la prochaine architecture de GPU signée Nvidia.

*Crédit photo : @Nvidia*