

Serveurs Arm : Grace, le chaînon manquant de NVIDIA ?

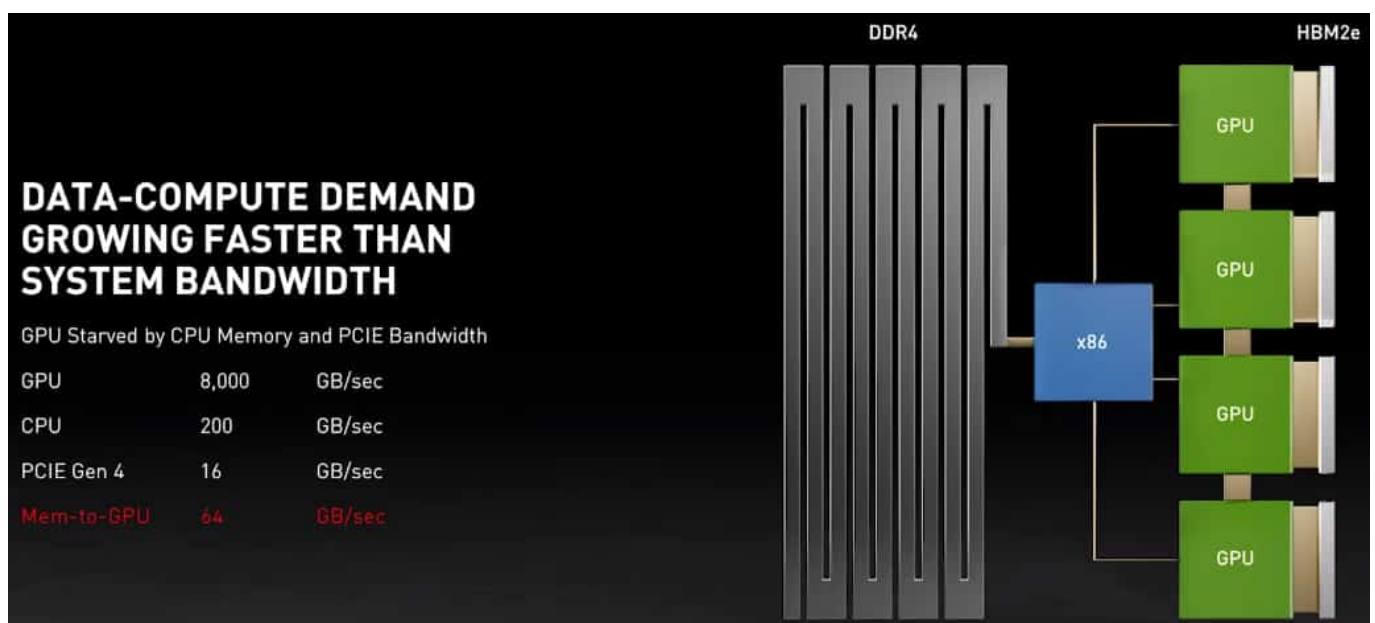
Vous souvenez-vous de Project Denver ? NVIDIA avait [officialisé](#) l'initiative au CES 2011. Sous cette bannière, il entendait concevoir ses propres cœurs Arm. Et les intégrer dans des SoC qui trouveraient place « des PC aux supercalculateurs ». En toile de fond, il y avait, entre autres, le lancement imminent de Windows RT. La démarche ne s'était finalement pas étendue au-delà du mobile, où elle avait notamment engendré la [version 64 bits du Tegra K1](#).



Dix ans plus tard, NVIDIA sonne une nouvelle offensive dans le même esprit. Mais sans cœurs *custom* et avec une visée plus restreinte : **IA et HPC**. [Son bras armé](#) se nomme Grace. On n'en sait encore que peu sur ce CPU dont la disponibilité ne doit intervenir que début 2023. Deux clients se sont toutefois annoncés : le laboratoire national de Los Alamos (dépendant du département américain de l'Énergie) et le Centre suisse de calcul scientifique.

Ce dernier exploitera Grace au travers d'un supercalculateur baptisé Alps. HPE concevra le système sur la base de l'architecture Cray EX et de la plate-forme NVIDIA HGX. Il remplacera [l'actuel](#) Piz Daint, opérationnel depuis 2012 et fondé sur des Xeon associés à des GPU P100.

NVIDIA ne fait pas de Grace un concurrent direct des processeurs x86*. Même s'il annonce des performances au niveau de celles des EPYC 2 d'AMD sur le *benchmark* [SPECrate2017_int_base](#). L'objectif numéro un, c'est l'optimisation du *package* CPU+GPU. Tout particulièrement en tirant parti de la technologie NVLink, face aux limites de bande passante du bus PCIe.



NVIDIA insiste sur les apports d'une telle architecture en matière de deep learning. Dans la

pratique, la durée d'entraînement des plus gros modèles qui existent à l'heure actuelle se compterait en jours et non plus en semaines. La référence : GPT-3 d'OpenAI et ses plus de mille milliards d'hyperparamètres.

* Dont les Xeon Ice Lake, qu'Intel a [présentés](#) la semaine dernière.

Illustrations © NVIDIA