

Big Sur, le serveur de Facebook dédié à l'intelligence artificielle

Facebook a dévoilé Big Sur, un design de serveur fonctionnant sur le principe des réseaux neuronaux au service de l'intelligence artificielle (IA) et autre Machine Learning (ML). La solution présentée sous forme de rack intègre huit GPU haute performance jusqu'à 300 Watt chacun avec la possibilité de configurer plusieurs topologies PCI-e. Elle s'appuie sur la plateforme Tesla Accelerated Computing de Nvidia dédiée au HPC (High Performance Computing). Dans le cas présent, Big Sur a été construit autour des Tesla M40 mais le système peut supporter un large choix de cartes PCI-e, souligne les concepteurs de Big Sur. Si, au même titre que Google ou Microsoft, Facebook a déjà mis un pied dans la partie logicielle de l'AI avec le programme FAIR (Facebook IA Research), Big Sur en constitue la première réalisation matérielle.

Deux fois plus rapide

« Big Sur est deux fois plus rapide que les précédentes générations [de hardware dédié à l'IA], assure [Serkan Piantino](#), directeur de recherche sur l'IA chez Facebook, ce qui signifie que nous pouvons travailler deux fois plus rapidement et explorer des réseaux deux fois plus grands. Et la distribution de l'exécution à travers huit GPU nous permet de multiplier la taille et la vitesse de nos réseaux par un autre facteur de deux. »

Une telle puissance pourrait, chez Facebook, être utilisée pour exploiter à grande échelle des applications d'IA développées en interne comme la capacité de « comprendre » des histoires, mais aussi analyser le contenu d'une image [et répondre à des questions sur son contexte](#), jouer ou encore apprendre des tâches non spécifiées à travers l'observation d'exemples.

Une conception simplifiée

Big Sur se distingue aussi par une conception visant à simplifier au maximum les éventuelles interventions de maintenance. La solution intègre son propre système de refroidissement, les éléments peu utilisés en sont absents, et les disques durs et barrettes mémoire sont remplaçables en quelques secondes sans outils. Y compris la carte-mère qui peut être changée en une minute, contre une heure habituellement sur les systèmes informatiques dédiés à l'IA. « Les dissipateurs de chaleur du processeur sont les seules pièces qui nécessitent un tournevis », indique Serkan Piantino. Une conception qui devrait optimiser les coûts de construction d'une telle machine mais aussi son exploitation du fait de



l'absence de systèmes de refroidissement supplémentaires, gros consommateurs d'énergie dans

les centres de calculs.

N'importe quel fabricant pourrait proposer son propre Big Sur à l'avenir. Facebook compte en effet placer son nouveau design dans Open Compute Project (OCP), le programme de hardware Open Source visant l'amélioration de la conception et du fonctionnement des datacenters. « *Nous pensons que cette collaboration ouverte contribue à favoriser l'innovation pour les conceptions futures, nous permettant tous de faire un pas de plus vers la construction de systèmes d'IA complexes qui apportent ce type d'innovation à nos utilisateurs et, en fin de compte, nous aider à construire un monde plus ouvert et connecté* », espère le chercheur.

Lire également

[Yann LeCun, Facebook : l'intelligence artificielle est son amie](#)

[Nvidia booste l'intelligence artificielle](#)

[TensorFlow, le machine learning de Google passe en Open Source](#)

[IBM passe sa solution de machine learning en Open Source](#)