

Google en dit plus sur le TPU, son composant spécial Machine Learning

Il y a presque un an, **Google** levait le voile sur le **TPU**, un module hardware dédié spécifiquement aux applications de **Machine Learning**. La firme indiquait alors que cette solution était en production dans ses datacenters depuis 2015.

Nous en savons maintenant plus sur le **Tensor Processing Unit**. [PCWorld](#) donne en effet quelques précisions concernant les performances de ce composant. Il s'avère **15 à 30 fois plus rapide** pour des tâches de Machine Learning que des CPU Intel Haswell ou des GPU Nvidia K80. Le rapport performance par watt des TPU se veut également de **25 à 80 fois meilleur** que celui des CPU et GPU.

Ce composant est un **ASIC**, un composant dédié à une tâche donnée. Google ne précise pas dans quelle mesure l'utilisation d'un **FPGA**, une puce reprogrammable, pourrait aller chasser sur les terres du TPU. Les FPGA, c'est la voie qu'a choisie Microsoft pour accélérer les travaux de Machine Learning dans ses datacenters.

Le TPU répond à un impératif économique

Google explique avoir été contraint de créer cette offre pour répondre au défi de la montée en puissance des applications mettant en œuvre du Machine Learning.

Si les utilisateurs des solutions de la firme venaient à employer la reconnaissance vocale ne serait-ce que 3 minutes par jour chacun, Google aurait dû **doubler le nombre de ses datacenters**. Le TPU permet donc de réduire fortement ses besoins en infrastructures.

À lire aussi :

[Google I/O : de l'Android, des assistants et de la technologie](#)

[Des smartphones boostés à coup de machine learning en 2017](#)

[AMD se projette dans le Machine Learning avec les accélérateurs Radeon Instinct](#)