

Cloud : OVH lance des instances optimisées GPU

Ce n'est encore qu'un galop d'essai, mais il signe l'arrivée du Français OVH sur le terrain des instances Cloud accélérées par des GPU. En l'occurrence, le Roubaisien s'est tourné vers des composants Nvidia Geforce GTX 1060. Les tarifs des instances bâties sur cette base vont de 0,306 euro HT par heure pour 15 Go de Ram (ou 110 euros par mois) à 0,472 euro HT de l'heure (170 euros par mois) pour 30 Go. Notons que, dans un message envoyé à ses clients, OVH précise que ces instances baptisées g1-15 et g1-30 sont pour l'instant disponibles sur un seul datacenter du groupe (GRA3), mais qu'elles seront étendues à l'ensemble des régions du Cloud maison si l'offre rencontre son public. D'autres modèles de cartes graphiques devraient également être proposés à l'avenir, ainsi que des workflows optimisés pour les GPU, assure l'hébergeur.

AWS, Google et Azure ont déjà des GPU

Rappelons que les cartes graphiques permettent d'accélérer des tâches demandant d'importants calculs en parallèle, comme le transcodage vidéo, la prévisualisation de séquences, mais aussi les recherches dans de grosses bases de données, les applications prédictives ou le Deep Learning. Spécialiste du collaboratif autour de vidéos, la start-up Weaverize a pu tester les instances GPU d'OVH en avant-première, dans le cadre du programme Digital Launch Pad du Roubaisien, dédié aux jeunes pousses. « *L'utilisation du GPU nous a permis de traiter un volume de données jusqu'à 1 000 fois plus rapidement qu'avec une instance Public Cloud classique ; il y a là un gain considérable en matière de performances — et donc de coûts !* », explique Rémi Auguste, fondateur de Weaverize.

Dans la pratique, les grands Cloud américains offrent tous des instances GPU. Chez AWS, le leader mondial du laas, l'offre est même déjà très diversifiée (avec les familles P2 et G2) et monte jusqu'à 16 GPU pour l'instance la plus musclée. Chez Google, l'offre, encore en bêta, est également disponible sur la zone Europe de l'Ouest. Elle repose sur les cartes Tesla K80 de Nvidia. Microsoft Azure n'est pas en reste avec ses familles NC (à base de Tesla K80) et NV (Tesla M60). Celles-ci, encore en bêta, ne sont toutefois pas encore disponibles sur la zone Europe de l'Ouest.

A lire aussi :

[FPGA : l'arme secrète spéciale IA du Cloud Azure de Microsoft](#)

[Microsoft et Facebook lancent des serveurs GPU dédiés l'IA](#)

[FPGA : l'arme secrète d'OVH pour parer les attaques DDoS](#)