

AWS lance des instances FPGA et low cost

A l'occasion de son évènement Re :Invent à Las Vegas, AWS a dévoilé les nouveautés de sa gamme d'instances disponibles. Cette année, le leader du Cloud public se focalise sur certaines fonctionnalités comme le GPU ou le FPGA. Ainsi, le CEO d'Amazon Web Services, Andy Jassy, a dévoilé « Elastic GPUs for EC2 ». Il ne s'agit pas d'une nouvelle instance, mais une manière pour les développeurs d'intégrer des ressources de traitement graphique au sein des machines virtuelles existantes. « *Cela sera utile dans le cas où les personnes ont simplement besoin d'un peu de GPU et non d'un processus graphique entier* », souligne le dirigeant.

Les instances F1 à base de FPGA

Autre grosse annonce, l'arrivée d'une [instance baptisée F1](#) et qui repose sur des accélérateurs FPGA (field-programmable gate arrays). Un terrain de jeu que [Microsoft a adopté pour repenser l'architecture des datacenters Azure](#). D'autres hébergeurs comme OVH s'en servent pour contrer les attaques DDoS. L'objectif est de s'appuyer sur des puces reprogrammables pour accélérer certains traitements et d'être plus rapide que le couple GPU/CPU. AWS pense notamment à la vidéo 4K ou la réalité virtuelle, mais aussi le machine learning.

Ce service va permettre d'écrire du code, de le packager en image, puis de l'exécuter sur le FPGA. Conscient de la difficulté de programmer pour le FPGA, AWS propose des outils comme un kit de développement (HDK) et des images (AMI). Le fournisseur de Cloud s'appuie sur des puces Xilinx Ultrascale gravée en 16 nm, 64 Go de DDR4 et 16 connexions PCIe. Ces instances F1 seront disponibles dans un premier temps aux Etats-Unis (Région Est pour l'instant) et plus globalement d'ici la fin du mois. Aucune tarification n'a été donnée pour ces instances, AWS indique seulement que l'abonnement se fera à la capacité de calcul à l'heure.

Specifications

F1 Instance Details

Instance Type	FPGA Cards	vCPUs	Instance Memory (GiB)	SSD Storage (GiB)	Enhanced Networking	EBS Optimized
F1.xlarge	1	8	128	488	Yes	Yes
F1.Medium	1	4	64	244	Yes	Yes

Rafraîchissement des petites instances et de l'intensif

AWS en profite pour toiletter certaines instances. C'est le cas de la gamme R, qui voit arriver la série R4 avec deux fois plus de mémoire et de vCPU, ainsi qu'un plus grand cache L3 que la série R3. La famille T pour des petits workload est rafraîchie. La série T2.xlarge comprend 16 Go et 2 vCPUs et la T2.2xlarge embarque 32 Go et 2vCPUs.

Pour une orientation misant sur la performance, il faudra se tourner vers deux nouvelles instances I3 et C5. Les premières s'adressent aux applications gourmandes en I/O en promettant 9 fois plus de traitement d'entrées/sorties par seconde que les instances I2. En complément, elles

embarquent deux fois plus de mémoire et de vCPU. Pour les instances C5 ciblant le calcul intensif, l'innovation réside dans l'usage des puces Intel Xeon Skylake de 7^{ème} génération.

Lightsail à l'assaut de Digital Ocean

Pendant que les grands acteurs du Cloud comme AWS, Microsoft et Google se battaient sur la guerre tarifaire des VM et sur la diversification de leurs offres, des start-ups se sont installées sur un marché de niche, celui de la VM low cost. Parmi elles, on peut citer Digital Ocean, Linode ou Blue Host qui connaissent un certain succès.

Avec Lightsail, AWS tente une percée sur ce marché du VPS (Virtual Private Server) avec un tarif agressif à 5 dollars par mois. A ce prix, la firme américaine vous propose une machine avec 512 Mo de mémoire, un processeur, 20 Go de stockage SSD et 1 To de transfert de données. Elle a repris [la même grille tarifaire que Digital Ocean](#).



A voir si cette incursion d'AWS dans le low cost donne des idées aux autres plateformes Cloud chez Google et Microsoft.

A lire aussi :

[AWS amène sa distribution Linux dans les datacenters des entreprises](#)

[AWS + VMware : un accord qui peut changer la face du Cloud](#)