

Serveurs : avec Power10, IBM change de braquet sur l'IA

Autres temps, autres mœurs ? Chez IBM, le discours sur l'IA a en tout cas nettement évolué entre deux générations de processeurs.

Fin 2017, le groupe américain lançait son [premier serveur doté de la puce Power9](#). Le découplage des traitements fut un axe majeur de cette annonce. « Dans un sens, c'est un hommage à l'architecture des ordinateurs d'antan », nous [expliquait](#)-on. En ligne de mire, autant les GPU (en particulier avec la technologie NVLink) que les FPGA (avec OpenCAPI).

Avec la [génération Power10](#), on n'est pas loin d'un virage à 180 degrés. IBM insiste en tout cas sur la possibilité d'exécuter des modèles à même le CPU. Et donne, pour son [premier serveur équipé de cette puce](#), un indicateur : par rapport au modèle de référence de la gamme Power9, des performances multipliées par 5 en inférence haute précision (32 bits). Un test effectué en NLP, sur le jeu de données SqUAD 1.1 avec BERT-Large.

À quoi doit-on cette augmentation ? Essentiellement à la présence de quatre accélérateurs matriciels par cœur, contre un sur les Power9.

De l'accélération, il y en a aussi sur la partie chiffrement. C'est en tout cas un autre élément qu'IBM met en avant. Avec un chiffre : des opérations 2,5 fois plus rapides en AES-256 (GCM et XTS). Cela s'entend avec OpenSSL 1.1..1g FIPS sur RHEL 8.4.

Côté sécurité, Big Blue mentionne également le chiffrement des données entre processeur et mémoire vive. Ainsi qu'une protection renforcée contre les attaques de type ROP (technique de dépassement de pile).

Illustrations © IBM