

Le nouveau serveur IBM Power étend les applications d'IA « on-premises » vers le Cloud et l'« Edge » pour une valeur business accrue dans tous les secteurs d'activité



Le 07 mai 2024, par Steve Sibley, Vice President IBM Power Product Management: Alors que de plus en plus d'organisations adoptent les promesses de l'intelligence artificielle pour accroître la valeur de l'entreprise, nous voyons des clients dans des secteurs tels que la vente au détail, l'industrie manufacturière, la santé et le transport déployer des applications à la périphérie (at the edge) afin de capitaliser sur les données là où elles sont produites. Dans son Market Guide for Edge Computing (Guide du marché de l'Edge Computing) de mars 2024, Gartner® note : « En plaçant les données, les capacités de gestion des données et les applications analytiques à des points optimaux, allant jusqu'aux appareils périphériques (endpoints), les entreprises peuvent mettre en place davantage de cas d'usage en temps réel. En outre, la flexibilité de déplacer les applications de gestion des données de manière ascendante ou descendante du continuum depuis des datacenters centralisés ou depuis des appareils périphériques Cloud/Edge permettra une plus grande optimisation des ressources^[1] ».

Pour contribuer à cet effort, IBM a annoncé aujourd'hui l'extension de sa gamme de serveurs avec l'introduction de **IBM® Power® S1012**. Ce système à demi-largeur à base de processeurs Power10 à 1 socket offre jusqu'à 3 fois plus de performances par cœur que le Power S812^[2]. Il est disponible en format monté en rack 2U ou tour de bureau et est optimisé pour l'informatique de type Edge. Il offre également le prix d'entrée le plus bas de la gamme Power pour exécuter les principales applications des petites et moyennes entreprises. L'IBM Power S1012 offre aux clients la flexibilité d'exécuter des applications d'inférence d'IA dans des bureaux et back-offices distants (ROBO) en dehors des datacenters traditionnels, et en connexion directe avec des services Cloud tels qu'**IBM® Power® Virtual Server** pour la sauvegarde et la reprise après sinistre.

² 2 serveurs monosocket, hauteur 2U, largeur de demi-rack équipés d'un module eSCM à 1, 4 ou 8 cœurs, jusqu'à 256 Go de mémoire chacun.

Châssis de type tour à 1 socket, équipé d'un module eSCM à 1, 4 ou 8 cœurs, avec jusqu'à 256 Go de mémoire.

Pour découvrir les dernières avancées de la plateforme IBM Power,[inscrivez-vous ici](#).

Faire plus avec moins d'espace à la périphérie (at the edge)

IBM Power S1012 est conçu pour améliorer les capacités de gestion à distance pour les clients qui cherchent à étendre des applications telles que l'inférence d'IA « on-premises » vers le Cloud et le « Edge ». L'Edge computing peut également fournir un avantage concurrentiel grâce à des informations en temps réel dans tous les secteurs, avec des exemples tels que l'analyse du comportement des clients dans le commerce de détail, le contrôle et l'optimisation des processus de production dans l'industrie manufacturière, et bien d'autres encore.

Les clients communs d'IBM et d'Equitus, partenaire de l'écosystème dans le domaine de l'analytique améliorée, utilisent IBM Power pour exécuter des modèles d'IA à la périphérie (at the edge) afin de fournir une classification des objets à des fins de défense.

*« Equitus Corp. avait besoin de plateformes matérielles critiques pour les environnements de « deep edge », d'opérations avancées, d'isolation physique (Air Gap) et de Cloud traditionnel. Nous avons constaté qu'IBM Power10 et son accélérateur mathématiques matriciel (Matrix Math Accelerator : MMA) offraient la meilleure technologie pour l'inférence en périphérie aussi facilement que dans le datacenter », a déclaré **Matt Niessen, President, Equitus Federal Corp.** « Aujourd'hui, les clients peuvent déployer nos systèmes d'IA Equitus Video Sentinel (EVS) et Knowledge Graph Neural Network (KGNN) sur les serveurs IBM Power10 et Red Hat OpenShift Container Platform pour de nombreux cas d'usage, y compris les plus cruciaux comme l'aide à la protection de la sécurité nationale. IBM Power S1012 fournira les capacités les plus récentes pour prendre en charge l'inférence de l'IA là où les données elles-mêmes sont générées. »*

IBM Power S1012 est conçu pour :

- **Permettre l'inférence d'IA à la périphérie** : en déployant Power S1012 dans le « Edge », les clients peuvent exécuter l'inférence d'IA au niveau des données, éliminant ainsi les transferts de données. Chaque Power S1012 comprend quatre accélérateurs mathématiques matriciels par cœur pour prendre en charge l'inférence d'IA. Pour s'assurer que les informations restent un avantage concurrentiel et ne tombent pas entre de mauvaises mains, le chiffrement transparent de la mémoire avec le Power10 sécurise les données qui entrent et sortent des modèles d'IA exécutés localement, ce qui permet de remédier aux fuites de données. De plus, grâce aux capacités avancées de gestion à distance et aux

caractéristiques de fiabilité Power10, les meilleures de leur catégorie^[3], IBM Power S1012 permet aux entreprises de gérer et de surveiller efficacement leurs environnements informatiques à distance, afin d'améliorer la réactivité et de minimiser les temps d'arrêt. Les fonctionnalités de haute disponibilité telles que le matériel redondant et les mécanismes de basculement peuvent contribuer à assurer un fonctionnement continu, le tout dans un encombrement physique minimal.

- **Réduire l'empreinte informatique et augmenter les principales performances pour les clients AIX et IBM i:** pour les clients **IBM i** de petite et moyenne taille, le Power S1012 permet d'accéder à la dernière technologie basée sur le processeur Power10. Disponible avec 1, 4 ou 8 cœurs de processeur pour améliorer la rentabilité et l'efficacité informatiques pour les applications IBM i critiques, les clients peuvent améliorer les performances jusqu'à 3 fois par rapport au Power S814^[4]. De plus, la conception du Power S1012 2U demi-largeur peut réduire l'espace alloué à l'empreinte physique informatique d'un client jusqu'à 75% par rapport au serveur rack Power S1014 4U et permet la convergence des applications et de la gestion informatique avec l'option d'utiliser un serveur pour la production côté à côté avec l'autre hébergeant une console de gestion matérielle virtuelle (vHMC) ou d'être utilisé pour le développement, les tests ou la haute disponibilité.

Assistance et disponibilité

Il est essentiel de maintenir une haute disponibilité tout au long de la durée de vie des systèmes tels que l'IBM Power S1012. **IBM Power Expert Care** permet d'associer immédiatement des services et une assistance par le biais d'une approche à plusieurs niveaux. Les clients peuvent bénéficier d'un niveau optimal de support pour les exigences critiques de leur infrastructure informatique avec des options allant de 3 à 5 ans de couverture en fonction du niveau de support. En outre, il existe des niveaux de service optionnels, en fonction des besoins du client, qui peuvent fournir une personnalisation et une assistance supplémentaires.

IBM Power S1012 sera disponible sur le marché auprès d'IBM et des partenaires commerciaux certifiés le 14 juin 2024.

Contacts Presse :

Weber Shandwick pour IBM

IBM

Gaëlle Dussutour

Tél. : + 33 (0)6 74 98 26 92

dusga@fr.ibm.com

Louise Weber

Tél. : + 33 (0)6 89 59 12 54

ibmfrance@webershandwick.com

[1] Gartner, Market Guide for Edge Computing, ID G00780388, Published 12 March 2024

[2] Based on the CPW Benchmark results for 1-core of 29,000 on Power S1012 compared to 9,360 on Power S812

[3] ITIC 2023 Global Server Hardware Server OS Reliability Report, August/September 2023, p.3

[4] Based on the CPW Benchmark results for 4-core of 111,300 on Power S1012 compared to 37,440 on Power S814 and 8-core of 203,100 on Power S1012 compared to 81,050 on Power S814
