



NVIDIA DGX-2 LE SYSTÈME DEEP LEARNING LE PLUS PUISSANT AU MONDE POUR LES PROJETS IA LES PLUS COMPLEXES

Repoussez toutes les limites de rapidité et d'évolutivité pour l'IA et le Deep Learning

Les réseaux de neurones profonds gagnent en taille et en complexité afin de répondre aux demandes les plus pressantes des nouveaux workflows d'entreprise et de recherche scientifique. Les architectures traditionnelles sur Data Center ne sont plus en mesure d'offrir une capacité de calcul suffisante pour prendre en charge les systèmes IA modernes. Le parallélisme se systématisait mais se heurte aux limitations de bande passante inter-GPU, alors que les développeurs s'appuient sur des clusters de calcul accéléré de plus en plus intensifs. Par conséquent, les Data Centers doivent être à même d'offrir toujours plus d'évolutivité. Une nouvelle approche du calcul informatique est désormais requise. Pour obtenir des résultats concrets et transformer le monde, il faut pouvoir compter sur une puissance de calcul IA virtuellement infinie.

Des performances d'entraînement auparavant inimaginables

Les workflows de calcul IA sont de plus en plus complexes et requièrent une puissance sans précédent. NVIDIA® DGX-2™, tout premier système à 2 pétaflops de l'industrie, vous permet d'interconnecter 16 GPU à hautes performances pour accélérer des modèles Deep Learning auparavant impossibles à entraîner. Cette solution offre une évolutivité exceptionnelle pour vous aider à entraîner des modèles jusqu'à quatre fois plus complexes sur un nœud unique. En comparaison avec les architectures x86 existantes, la capacité de NVIDIA DGX-2 à entraîner des réseaux ResNet-50 impliquerait de déployer 300 serveurs équipés de doubles CPU Intel Xeon Gold – pour un coût cumulé d'environ 2,7 millions de dollars.

NVIDIA NVSwitch : structure réseau révolutionnaire basée sur l'IA

La recherche scientifique de pointe, qui exploite le parallélisme de manière intensive, requiert des niveaux de bande passante inter-GPU encore jamais vus. NVIDIA a créé NVSwitch pour répondre à ce besoin. Aussi importante que l'évolution des lignes RTC aux connexions à très haut débit, la technologie NVSwitch vous offre dès maintenant une structure réseau performante et pérenne. Grâce à NVIDIA DGX-2, la complexité et la taille des modèles ne sont plus limitées par les architectures traditionnelles. Vous pouvez désormais paralléliser l'entraînement de vos modèles en tirant profit de la structure réseau du DGX-2, qui délivre une bande passante bisection à 2,4 téraoctets/s avec des performances jusqu'à 24 fois plus rapides que les solutions de génération précédente. Cette nouvelle "autoroute d'interconnexion" offre des possibilités illimitées pour des modèles pouvant désormais exploiter le plein potentiel d'un entraînement distribué sur 16 GPU simultanément.

SPÉCIFICATIONS SYSTÈME

GPU	16x NVIDIA® Tesla V100
Mémoire GPU	512 Go total système
Performances	2 pétaflops
Cœurs NVIDIA CUDA®	81 920
Cœurs NVIDIA Tensor	10 240
Modules NVSwitch	12
Consommation maximale	10 kW
CPU	Double CPU 24 cœurs Intel Xeon Platinum 8168 à 2,7 GHz
Mémoire système	1,5 To
Réseau	8x 100Gb/sec Infiniband/100GigE Dual 10/25Gb/sec Ethernet
Stockage	Système d'exploitation : 2x 960 Go NVME SSD Stockage interne : 30 To (8x 3,84 To) NVME SSD
Logiciels	OS hôte Ubuntu Linux Voir environnement logiciel pour plus d'infos
Poids	154,2 kg
Dimensions système (mm)	440 mm (H) 482,3 mm (L) 795,4 mm (P) - Sans bezel avant 834,0 mm (P) - Avec bezel avant
Températures de fonctionnement	5°C à 35°C (I)

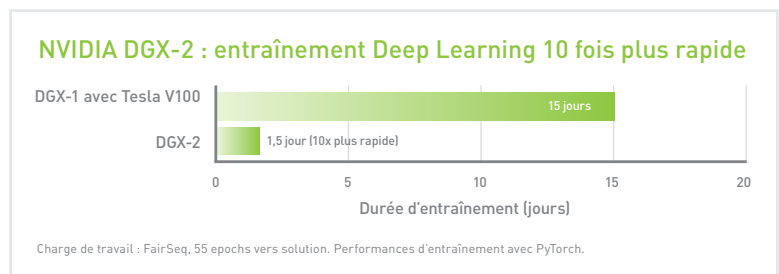
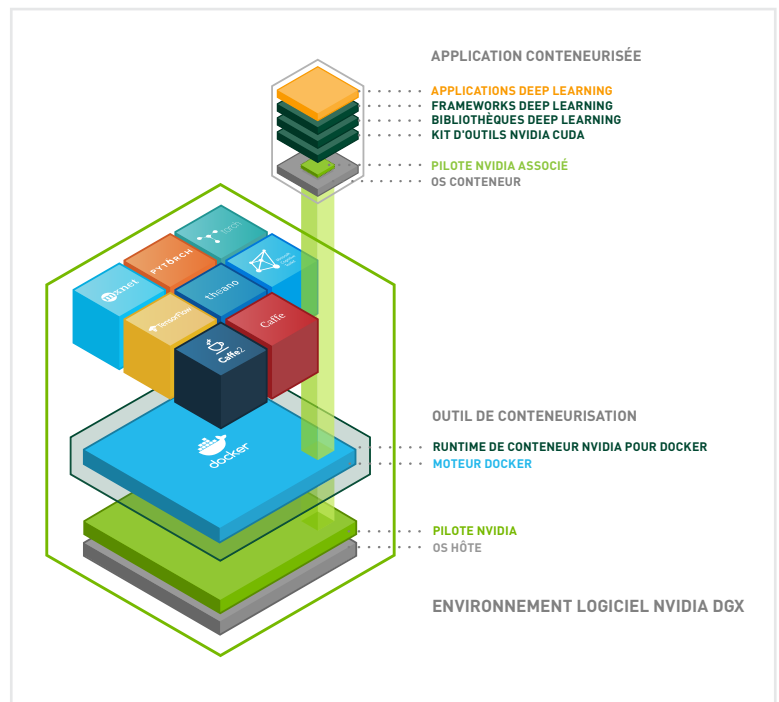
Déployez vos projets IA à une échelle sans précédent

Pour répondre à leurs impératifs commerciaux, les entreprises doivent être en mesure de déployer leurs projets IA plus rapidement que jamais - sans hausse des coûts d'exploitation ni impact sur l'infrastructure existante. Le système DGX-2 et l'environnement logiciel DGX accélèrent le déploiement et simplifient les opérations avec une mise à l'échelle exceptionnelle des performances. Solution modulaire, évolutive et clé-en-main, DGX-2 constitue la voie la plus rapide vers l'intelligence artificielle. Exploitez des fonctionnalités avancées de virtualisation IA pour configurer librement votre cloud d'entreprise privé. Profitez d'une puissance IA illimitée et accélérez l'ensemble de vos ressources de calcul grâce à une solution qui s'adapte à vos besoins via une infrastructure évolutive et compacte. Ce modèle de déploiement rapide et cette architecture évolutive permettent à vos équipes informatiques d'accélérer les procédures de déploiement pour se concentrer sur la réalisation de vos projets.

Infrastructure IA d'entreprise à disponibilité permanente

L'IA est essentielle à votre entreprise, c'est pourquoi vous devez pouvoir travailler en toute tranquillité d'esprit avec une plateforme fiable, disponible et fonctionnelle. La solution NVIDIA DGX-2, qui offre une stabilité à toute épreuve pour les workflows IA, a été spécifiquement conçue pour réduire les risques d'interruption, rationaliser le fonctionnement réseau et préserver la continuité opérationnelle.

Accélérez les procédures d'installation et d'optimisation pour vous focaliser sur vos projets de recherche. Les services d'assistance de NVIDIA vous accompagnent dans votre déploiement et vous aident à configurer rapidement votre matériel et vos logiciels open-source. Grâce à une infrastructure entièrement intégrée qui repose sur les outils, les logiciels et l'expertise de NVIDIA, vous bénéficiez d'une mise en service rapide et d'un entraînement encore plus rapide pour vous permettre de réaliser vos projets les plus importants.



En savoir plus : www.nvidia.fr/data-center/dgx-2