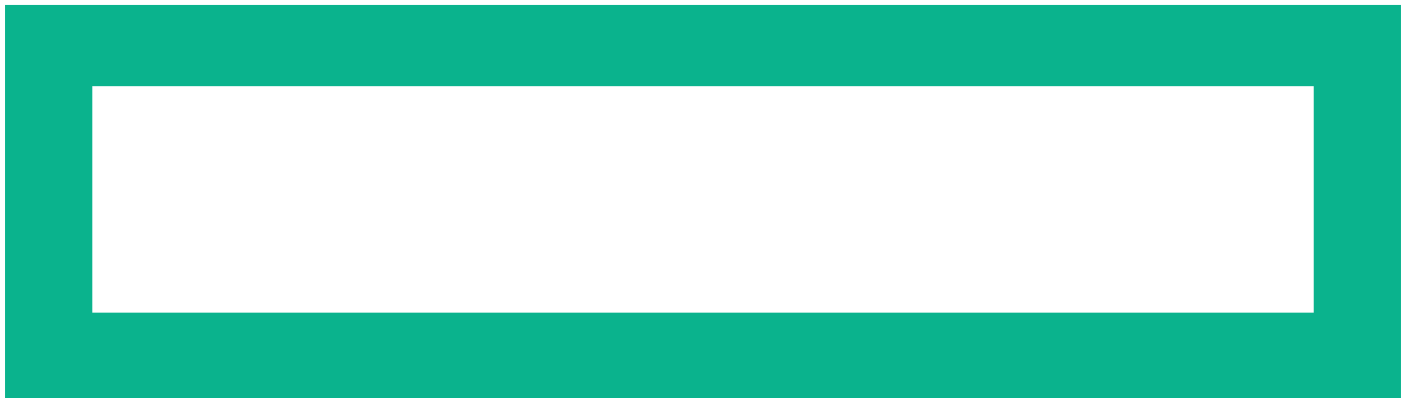




LE RÔLE CRUCIAL DU SUPERORDINATEUR DANS LA PROCHAINE VAGUE DE L'INTELLIGENCE ARTIFICIELLE



Alors que les entreprises avec la plus grande puissance de calcul étaient autrefois considérées comme les plus compétitives, il s'agit aujourd'hui de trouver de nouvelles façons d'exploiter les données pour dévoiler des informations commerciales critiques. Stimulée par le besoin de résoudre des problèmes insolubles et de faire face au déluge de données, l'adoption de l'intelligence artificielle (IA) par les entreprises a explosé ces dernières années. Désormais, quasiment toutes les organisations redoublent d'efforts pour intégrer et déployer des méthodologies d'IA dans leurs principaux cas d'usage. Celles qui y parviennent obtiennent un aperçu de la promesse et de l'opportunité réelles de l'IA ; la capacité à prendre pleinement conscience de leurs données métier et à en extraire de la valeur pour améliorer les unités commerciales actuelles, et à potentiellement dévoiler de tout nouveaux modèles économiques.

Alors que les organisations s'appuient sur l'expérience acquise lors de la première vague de croissance de l'IA, nous sommes maintenant prêts pour la prochaine phase d'adoption de cette technologie. L'explosion des implémentations de démonstration de faisabilité précoces favorisées en partie par l'engouement initial pour l'IA laisse désormais la place à une approche plus mesurée. De plus, des efforts soutenus se portent sur l'application des techniques d'IA à la production pour résoudre des problèmes métiers réels et favoriser des décisions commerciales tout aussi réelles. La vague de croissance suivante de l'IA ne peut toutefois pas se produire et ne se produira pas sans calcul haute performance.

L'IA est désormais étroitement liée au calcul intensif. À mesure que de nombreuses organisations souhaitent en tirer le maximum, elles se rendent compte qu'elles ont besoin d'architectures de superordinateur entièrement repensées pour profiter au mieux de la nouvelle génération d'intelligence artificielle. Les superordinateurs et le stockage haute performance sont des outils cruciaux capables d'offrir les fonctionnalités dont les entreprises auront inévitablement besoin pour surfer sur la nouvelle vague de croissance de l'IA.

LE DÉFI ET L'OPPORTUNITÉ DE L'IA

Les entreprises de toutes tailles et de tous les secteurs appliquent de plus en plus les techniques d'analyse de données, de machine learning et de deep learning à de nombreux cas d'usage. Souvent, c'est une combinaison de plusieurs approches qui est utilisée pour répondre à un cas d'utilisation particulier. Mais pour les projets d'IA appliquée à l'entreprise, le parcours du pilote à la production à la création de valeur pour l'entreprise peut être extrêmement complexe. Lorsqu'elles cherchent à adopter l'IA, les organisations sont confrontées à plusieurs défis qui les forcent à tout réévaluer, du personnel qu'elles emploient aux nouvelles technologies qu'elles achètent.

Pendant la première vague d'exploration de l'IA, de nombreuses organisations ont cherché à appliquer une seule technique basée sur l'IA pour leur workflow standard et ces implémentations à petite échelle et ponctuelles représentaient un bon point de départ qui a permis à de nombreuses entreprises de tester la faisabilité de l'IA pour leur activité. Maintenant que cette première vague est arrivée à maturité, la prochaine grosse vague d'adoption et de croissance de l'IA devrait nous apporter des déploiements à grande échelle, un ensemble d'applications plus évoluées et une attention plus forte sur la création de workflows d'IA intégrée qui deviendront critiques pour la prise de décisions commerciales. Cela devrait permettre à davantage d'organisations de considérer l'IA comme plus qu'un outil limité qui ne peut être appliqué qu'à une seule étape du workflow ou utilisé pour automatiser des tâches spécifiques. Désormais, elles seront plus nombreuses à considérer l'IA comme ce qu'elle est vraiment : une fonctionnalité omniprésente qui a un impact profond sur chaque cas d'utilisation et stack technologique.





À mesure que la nouvelle vague d'IA prend de l'ampleur, nous assisterons à une convergence encore plus grande avec la simulation HPC traditionnelle. L'effort nécessaire pour appliquer l'IA de façon plus large sera synonyme d'interdépendances plus fortes entre les applications existantes et les techniques comme le machine learning et le deep learning. De plus, ces techniques seront de plus en plus appliquées à des données multidisciplinaires (par exemple, l'utilisation de données géospatiales, climatologiques et économétriques permettra d'obtenir des prédictions plus précises en matière de production agricole) De la même façon, les approches d'IA dynamiques, telles que les réseaux antagonistes génératifs (GAN)¹ et l'apprentissage par renforcement (RL)² gagneront du terrain.

L'IA n'est pas seulement une destination, c'est une fonctionnalité dont les applications généreront des possibilités révolutionnaires pour l'entreprise, l'ingénierie et la science. Avec la prochaine vague d'IA, nous obtiendrons des résultats encore plus impressionnants, comme la prédiction des modèles climatologiques, la compréhension du fonctionnement interne du cerveau humain, l'utilisation du séquençage génétique pour révolutionner les décisions de soins et même des remèdes pour des maladies incurables (pour n'en nommer que quelques-uns). Chaque nouvelle application de l'IA apporte un impact et un avantage cumulatifs qui sont amplifiés dans toute l'entreprise, de la prise de décision par les cadres dirigeants aux stacks technologiques sous-jacents. L'IA a le potentiel de dévoiler des informations et des recommandations en analysant de grands volumes de données, ce qui permet d'accéder à des décisions commerciales opportunes et éclairées, de la gestion des ressources à la logistique en passant par tous les domaines intermédiaires. Cela signifie que ceux qui maîtriseront réellement l'IA lors de cette prochaine vague constateront des effets positifs dans chaque domaine de leur activité et seront en bonne position pour dominer leur secteur avec un avantage concurrentiel durable.

L'IA EST UN PROBLÈME POUR SUPERORDINATEUR

Selon l'IDC, la taille de l'univers numérique représenté par les données double tous les deux ans et devrait atteindre 175 zettaoctets d'ici 2025.³ Alors que de nombreuses entreprises éprouvent des difficultés à faire face à l'assaut d'informations causé par l'immense croissance des données, cette croissance est positive pour les applications d'IA qui prospèrent grâce aux ensembles de données volumineux. En particulier, les modèles de machine learning et de deep learning améliorent constamment la précision des résultats à mesure que davantage de données sont introduites dans le modèle. À mesure que les ensembles de données du machine learning et du deep learning augmentent, les projets d'IA ont inévitablement besoin de systèmes informatiques plus sophistiqués.

Il existe également une grande convergence entre l'IA et la simulation HPC traditionnelle et de nombreuses organisations qui effectuent des simulations ajoutent de plus en plus le machine learning ou le deep learning à leurs workflows, et vice versa. Ces deux méthodologies deviennent si étroitement liées que nous constatons l'émergence d'une technique totalement nouvelle, la simulation cognitive, où les simulations à grande échelle deviennent plus intelligentes grâce aux algorithmes de machine learning intégrés. Dans l'ensemble, les workflows deviennent de plus en plus hétérogènes, mêlant analyse de données, machine ou deep learning et IA avec une simulation à grande échelle.

La fusion de l'IA avec les workflows HPC traditionnels se répand. PGS, une société d'exploration sismique, a récemment appliqué des techniques d'optimisation par machine learning⁴ pour déterminer le modèle de vitesse pour un workload complet d'imagerie sismique d'inversion d'onde. Et Argonne National Laboratory applique le deep learning piloté par superinformatique pour dévoiler rapidement des informations issues de la mine d'informations qui a déjà été extraite des réponses aux médicaments dans tous les différents types de cancer.⁵

Ces témoignages de réussite prouvent que les superordinateurs, qui sont traditionnellement utilisés dans tous les secteurs pour la modélisation et la simulation, sont également très bien positionnés pour traiter de très importants problèmes de machine ou de deep learning.

¹ Un GAN représente une approche dans laquelle deux réseaux neuronaux sont mis en opposition ; l'un génère de nouvelles données candidates et l'autre en évalue l'authenticité.

² Le RL est une approche dans laquelle des algorithmes prennent une série de décisions pour atteindre un objectif défini et apprennent selon l'approche sur ces décisions.

³ [networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html](https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html)

⁴ [pgs.com/media-and-events/news/industry-insights---machine-learning-and-seismic/](https://www.pgs.com/media-and-events/news/industry-insights---machine-learning-and-seismic/)

⁵ [anl.gov/article/automating-the-development-of-deeplearningbased-predictive-models-for-cancer-research](https://www.anl.gov/article/automating-the-development-of-deeplearningbased-predictive-models-for-cancer-research)





On constate également un intérêt croissant pour les méthodes de machine ou de deep learning distribuées ou les algorithmes de formation plus rapides grâce à la mise en parallèle des calculs d'apprentissage sur plusieurs machines. Une puissance de calcul très importante est nécessaire pour former des réseaux neuraux profonds de cette façon, et à mesure que la taille et la complexité des données impliquées dans les problèmes d'IA augmentent, la puissance de calcul nécessaire pour les traiter augmente également constamment. OpenAI, une société de recherche en IA à but non lucratif, a récemment publié une analyse⁶ montrant que la quantité de calcul utilisée pour les grands cycles de formation d'IA a augmenté de façon exponentielle depuis 2012, doublant tous les trois mois et demi. Et dans un récent rapport de recherche, Digital Catapult a indiqué que le prérequis de calcul minimal approximatif pour la formation d'un réseau neural profond sur un ensemble de données de 1,28 million d'images serait de l'ordre d'un exaflop.⁷

Le traitement et l'analyse de volumes de données toujours plus grands, la prise en charge de charges de travail hétérogènes et l'activation de méthodes de formation distribuées nécessitent des architectures de calcul toujours plus puissantes et efficaces. En d'autres termes, l'IA est désormais sans équivoque un problème de superinformatique et de stockage. À mesure qu'un plus grand nombre d'organisations adoptent des technologies IA, les systèmes de calcul/stockage/analyse et les langages de programmation les plus à même de traiter les applications d'IA sont ceux déjà utilisés par le secteur du HPC.

UNE TOUTE NOUVELLE ARCHITECTURE DE SUPERORDINATEUR POUR LA NOUVELLE VAGUE D'IA

Les superordinateurs sont en train de rapidement devenir un ingrédient essentiel pour un déploiement d'IA réussi, et pas seulement à cause de leur puissance de calcul. L'augmentation de la puissance de calcul nécessaire est stimulée non seulement par la croissance des données elle-même, mais également par les manières de plus en plus complexes et sophistiquées qu'ont les organisations d'appliquer les techniques d'IA à leurs données.

Les superordinateurs modernes doivent traiter d'importantes charges de travail de modélisation, de simulation, d'IA et d'analyse. Ce besoin favorise une nouvelle approche de la conception de système qui a engendré une nouvelle génération de systèmes centrés sur les données permettant à diverses charges de travail et workflows de s'exécuter simultanément sur un seul système. Ces machines offrent la prise en charge nécessaire pour les processeurs généralement utilisés pour la modélisation et la simulation comme pour les accélérateurs de processeurs graphiques utilisés pour le machine learning et le deep learning. Le passage à une nouvelle classe de superordinateurs est une étape nécessaire et critique pour optimiser la productivité de la combinaison de plus en plus diverse d'applications que nous pressentons pour la prochaine grosse vague de croissance de l'IA.

⁶ openai.com/blog/ai-and-compute/

⁷ assets.ctfassets.net/nubxhjwc091/6qDT7u9pzUsq8uukwCAayw/acc7e59350faa88fc504fc990c17deb7/MIG_MachinesforMachineIntelligence_Report_DigitalCatapult-1.pdf





Les organisations qui commencent tout juste le processus d'évaluation de nouvelles technologies pour prendre en charge leurs projets d'IA passent souvent différentes solutions ponctuelles au crible plutôt que d'envisager l'achat d'un système complètement intégré compatible avec l'IA. Cela les oblige à acheter un ensemble de composants informatiques individuels, dont la plupart répondent à une partie du workflow d'IA, sans se préoccuper des besoins de l'ensemble du workflow, et à tenter de les assembler en un système complet. C'est bien sûr possible, mais il existe une façon bien plus efficace et rentable d'aborder l'investissement dans l'IA. Le rythme de l'innovation en IA ces dernières années a été tel que de nombreux modules de base de l'IA sont désormais en place (par exemple, les interconnexions innovantes, les systèmes de processeurs graphiques denses, les stacks logiciels efficaces et une meilleure intégration système globale). Jusqu'à récemment, aucun fournisseur de système n'a rassemblé toutes ces fonctionnalités dans une seule plateforme de façon claire et efficace.

DÉFINIR LE RYTHME DE L'INNOVATION POUR L'IA D'ENTREPRISE

La longue expérience de Hewlett Packard Enterprise dans la fourniture de technologies de calcul intensif, de stockage et d'analyse de données signifie également que nous sommes bien placés pour fournir les outils et l'expertise nécessaire pour résoudre les défis de l'IA. Nous avons aidé nos clients à surfer sur la première vague d'exploration de l'IA et avons pu constater par nous-mêmes leurs succès et les défis liés à la mise en œuvre. Cette expérience a façonné notre réflexion sur la façon d'aider au mieux les organisations à franchir le cap et à réussir avec la nouvelle génération d'IA.

Ces dernières années, notre clientèle vaste et unique nous a donné l'occasion de découvrir certains cas d'utilisation, dans lesquelles des applications d'IA de pointe utilisent de très grands ensembles de données (véritablement des Big Data) pour obtenir des informations. Notre travail nous a offert un aperçu des défis auxquels le reste du secteur du calcul haute performance sera bientôt confronté. Nous résolvons dès aujourd'hui les problèmes insolubles de demain pour assurer un parcours d'IA plus fluide et moins de difficultés à nos clients.

HPE offre un solide portefeuille de solutions et de systèmes d'IA qui aident les clients à mettre en œuvre et à maîtriser l'IA pour leurs cas d'utilisation spécifiques. Depuis un certain temps, nous voyons arriver la prochaine vague d'IA ce qui nous a permis de créer une toute nouvelle architecture de calcul haute performance pour que vos clients puissent exploiter tout le potentiel de la convergence de l'IA et de la simulation. Les clients peuvent également bénéficier de la capacité de s'adapter et d'évoluer dans un paysage complexe et en constante évolution. Cette architecture gère l'hétérogénéité de l'association de l'IA et de la simulation dans la même architecture, ce qui permet à nos clients d'exécuter simultanément des workflows de plus en plus divers sur un seul système et de véritablement favoriser l'avenir de l'IA.

Nous définissons le rythme de l'innovation de l'IA appliquée à l'entreprise en offrant des solutions de calcul haute performance flexibles qui peuvent s'intégrer dans les datacenters standards et pas seulement dans les laboratoires de recherche haut de gamme. L'approche unique d'IA orientée sur le workflow de HPE, qui cible la fourniture de structures et d'outils pour déployer l'ensemble d'un workflow d'IA, renforce la productivité des scientifiques des données tout en réduisant la complexité associée à la réalisation des résultats commerciaux avec de nouvelles applications d'IA. Nos solutions infrastructure IA intégrées aident les sociétés de tous types et les institutions de recherche de toutes tailles à franchir une nouvelle étape pour faire de l'IA une réalité.





CONCLUSION

Nous commençons tout juste à découvrir la façon dont les systèmes d'IA vont changer notre manière de vivre et de travailler. Le véritable envol de la prochaine vague de développement d'IA dépendra de l'existence d'infrastructures de calcul suffisamment puissantes pour prendre en charge la taille et la complexité croissantes des cas d'utilisation de l'IA. À mesure que la nouvelle vague d'IA prend de l'ampleur, nous avons hâte de voir les organisations utiliser l'IA pour affiner la possibilité de déplacer de grands volumes de données au sein des applications et entre elles, de s'adapter aux conditions changeantes du marché et d'évoluer verticalement avec la croissance des données. Ces entreprises assureront ainsi leur réussite avec l'IA. Mais, elles n'y parviendront pas sans architectures de calcul haute performance innovantes et conçues pour une toute nouvelle phase d'IA.

POUR EN SAVOIR PLUS

hpe.com/fr/fr/compute/hpc/supercomputing

Faites le bon achat.
Contactez nos spécialistes.



Live Chat



E-mail



Appel



Mises à jour