



Libérer le potentiel de l'IA : Comment HPE Private Cloud AI accélère le déploiement de l'IA et l'innovation

AUTEUR

Steven Dickens

Vice-président et Chef de service | The Futurum Group

EN PARTENARIAT AVEC



Hewlett Packard
Enterprise



nVIDIA.

JUIN 2024



Résumé

Le potentiel de transformation de l'IA générative (GenAI) a suscité un fort intérêt et stimulé la croissance dans divers secteurs. Sur la seule année 2023, le nombre de projets publics de GenAI sur GitHub a augmenté de 248 %, ce qui reflète l'adoption croissante des technologies IA. Toutefois, cette croissance rapide est associée à plusieurs défis. Confrontées à un paysage technologique complexe, les entreprises ont souvent du mal à effectuer la transition des projets d'IA pilotes à la production. De plus, la protection de la propriété industrielle est une préoccupation majeure pour les entreprises qui s'inquiètent d'exposer leurs données propriétaires à des modèles publics. HPE, en collaboration avec NVIDIA, répond à ces défis en proposant des solutions d'IA sur cloud privé évolutives, sécurisées et aux performances optimisées. Ces solutions permettent aux entreprises d'exploiter la puissance de l'IA tout en gardant le contrôle de leurs données et de leur infrastructure, afin d'accélérer leur passage vers une innovation guidée par l'IA.

Présentation de la solution HPE Private Cloud AI

HPE Private Cloud AI, une composante du portefeuille NVIDIA AI Computing by HPE, est un cloud privé clés en main, évolutif et optimisé pour l'IA conçu pour accélérer le déploiement des projets d'IA tout en assurant que les données restent sous le contrôle de l'entreprise. Cette solution combine le calcul, les réseaux et les logiciels accélérés par NVIDIA avec le calcul et le stockage haute performance HPE, le HPE GreenLake cloud et les capacités de HPE en matière de pipelines de données, d'orchestration et de MPLOps. Elle offre ainsi aux entreprises de toutes tailles une voie rapide et flexible pour le développement et le déploiement des applications GenAI. Les entreprises peuvent exploiter ces capacités pour surmonter les obstacles à l'adoption de l'IA et bénéficier d'un délai de rentabilisation plus court et d'une évolutivité fluide pour répondre à leur croissance future.

Déclaration d'intention du rapport d'étude

Ce rapport d'étude vise à fournir une analyse approfondie des défis auxquels les entreprises sont confrontées en termes de déploiement de charges de travail IA, et à montrer comment HPE Private Cloud AI répond efficacement à ces défis. Il met en lumière plusieurs conclusions et informations essentielles quant aux capacités de la solution en termes de délai de rentabilisation, de reproductibilité et d'évolutivité. Enfin, il propose des recommandations pour les organisations qui souhaitent optimiser leurs déploiements IA et maximiser les avantages des technologies IA tout en gérant leurs coûts et en assurant leur sécurité.

Conclusions et informations essentielles

1. **Délai de rentabilisation :** La solution de HPE et NVIDIA réduit sensiblement le délai de commercialisation des projets IA en offrant des outils, modèles et infrastructures pré-intégrés et testés. Cette approche permet aux entreprises d'effectuer une transition rapide depuis les phases pilotes vers la production, et de bénéficier rapidement des avantages de l'IA.
2. **Des solutions reproductibles :** L'infrastructure clés en main et standardisée proposée par HPE Private Cloud AI permet d'effectuer des déploiements IA cohérents, performants et évolutifs. Grâce à des outils prédéfinis, intégrés et testés, les entreprises peuvent déployer des charges de travail IA rapidement et facilement, le plus souvent en un seul clic.
3. **Le meilleur du cloud :** Conçue pour s'adapter au besoin commercial, la solution HPE offre des options de déploiement flexibles et des modèles financiers de paiement à l'utilisation qui soutiennent la croissance tout en limitant efficacement les coûts. Le modèle CapEx et la facilité de gestion de la solution améliorent encore l'évolutivité et permettent aux entreprises d'étendre leurs capacités IA de manière fluide.

Cette présentation mise à jour reflète l'importance des arguments en faveur du cloud hybride/privé par rapport au cloud public pour l'IA et décrit les défis spécifiques qui y sont associés :

Présentation des défis opérationnels des implémentations IA

Les entreprises qui envisagent des implémentations IA sont souvent confrontées à des défis spécifiques qui rendent les solutions de cloud hybride/privé plus attrayantes que les options de cloud public :

- 1. Définir, développer et mettre en pratique les cas d'utilisation adéquats :** Les entreprises doivent identifier avec précision les cas d'utilisation de l'IA qui leur apporteront le plus de valeur ajoutée. Ceci nécessite une compréhension fine du besoin commercial et une capacité à le transposer au sein de projets d'IA. Une fois les cas d'utilisation identifiés, leur développement et leur mise en pratique implique une grande complexité, notamment pour intégrer les modèles IA dans les workflows existants et garantir qu'ils fournissent des résultats commerciaux tangibles.
- 2. Maximiser l'impact des données tout en protégeant la propriété intellectuelle :** Pour générer des informations pertinentes, il est essentiel d'exploiter efficacement les données des modèles IA. Néanmoins, ceci doit être fait sans compromettre les données sensibles relevant de la propriété intellectuelle. Les solutions de cloud public sont fréquemment associées à des inquiétudes portant sur la sécurité et le contrôle des données et à la difficulté de protéger les algorithmes propriétaires et les données sensibles contre les accès non autorisés et les failles.
- 3. Adapter les efforts en matière d'IA sans faire exploser les budgets :** Développer des initiatives IA pour répondre à des demandes croissantes peut entraîner des coûts prohibitifs, en particulier si l'on utilise des ressources de cloud public. Les entreprises doivent réaliser des compromis entre leurs besoins en puissance de calcul et en stockage et les implications financières, et s'assurer que leurs déploiements IA conservent un bon rapport coût-efficacité tout au long de leur adaptation.
- 4. Optimiser la technologie pour des cas d'utilisation spécifiques :** Défi : les différentes applications IA nécessitent des technologies et infrastructures différentes. Il est essentiel, mais complexe, de s'assurer que l'ensemble des technologies soit optimisé pour des cas d'utilisation spécifiques. Les solutions de cloud public n'offrent pas toujours le degré de personnalisation ou d'optimisation des performances requis par certaines charges de travail IA.
- 5. Réaliser l'inférence et l'ajustement près des sources de données :** Pour réduire la latence et améliorer les performances, il est essentiel de rapprocher l'analyse des données. Les solutions de cloud public impliquent souvent le transfert des données sur de longues distances, ce qui entraîne des retards et des inefficacités. En vous assurant que l'inférence et l'ajustement sont réalisés près des sources de données, vous pouvez accroître sensiblement la vitesse et la précision des modèles IA.

Ces problématiques montrent bien l'importance d'envisager des solutions de cloud hybride et privé pour les implémentations IA. En effet, ces solutions offrent davantage de contrôle, de personnalisation et de gestion des coûts que les alternatives sur cloud public.



Résumé des conclusions

HPE Private Cloud AI offre des fonctions et capacités essentielles qui répondent aux principaux défis que doivent relever les entreprises pour déployer des charges de travail IA :



1. Délai de rentabilisation : HPE accélère le développement et le déploiement de l'IA et réduit le délai de commercialisation grâce à une infrastructure et des outils pré-intégrés et testés.



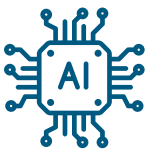
2. Reproductibilité : Cette solution permet d'effectuer des déploiements IA cohérents, performants et évolutifs à l'aide d'un ensemble d'infrastructure standardisé, ce qui garantit fiabilité et efficacité.



3. Le meilleur du cloud : La solution HPE soutient la croissance et la flexibilité avec la plateforme HPE GreenLake l'infrastructure évolutive, pour répondre aux variations de besoin commercial.



4. Sécurité : Des mesures de sécurité complètes assurent la protection des données et la conformité et répondent ainsi aux vulnérabilités accrues liées à l'adoption de l'IA.



5. Gouvernance et gestion de l'IA : La solution simplifie la gouvernance de l'IA avec un point d'accès et un contrôle unifiés, ce qui facilite la gestion dans les environnements hybrides.



6. Partenariats stratégiques : HPE tire parti de sa collaboration avec NVIDIA et des intégrateurs système pour améliorer les capacités et fournir un support solide aux déploiements IA.



Recommandations - Les arguments en faveur du cloud hybride/privé par rapport au cloud public pour l'IA

Les entreprises qui envisagent des implémentations IA peuvent tirer de grands avantages des solutions de cloud hybride/privé. L'identification précise des cas d'utilisation de l'IA qui offrent le plus de valeur ajoutée nécessite une compréhension détaillée du besoin commercial et la capacité à les transposer au sein de projets IA. Le développement et la mise en pratique de ces cas d'utilisation implique d'intégrer les modèles IA dans les workflows existants pour garantir qu'ils fournissent des résultats commerciaux tangibles.

Il est essentiel d'exploiter les données efficacement dans les modèles IA pour générer des informations pertinentes sans compromettre les données sensibles relevant de la propriété intellectuelle. En exerçant un contrôle strict sur leurs données, les entreprises peuvent protéger leurs algorithmes propriétaires et informations sensibles contre les accès non autorisés et les failles et répondre aux principales préoccupations de sécurité associées aux solutions de cloud public.

Les initiatives IA doivent pouvoir évoluer pour répondre à la croissance des besoins, sans que les budgets associés deviennent hors de contrôle. Les entreprises doivent réaliser des compromis entre leurs besoins en puissance de calcul et en stockage et les implications financières, et s'assurer que leurs déploiements IA conservent un bon rapport coût-efficacité tout au long de leur adaptation.

L'optimisation de l'ensemble des technologies pour des applications IA spécifiques est cruciale pour la réussite. Les entreprises doivent personnaliser et affiner leurs technologies pour qu'elles répondent aux besoins de diverses charges de travail IA, et veiller à appliquer les meilleures solutions aux différents cas d'utilisation.

La réalisation de l'inférence et de l'ajustement de modèle IA à proximité des sources de données constitue un autre facteur important. Cette approche réduit la latence et améliore la performance : en effectuant l'inférence et l'ajustement près des sources de données, les entreprises peuvent améliorer la vitesse et la précision des modèles IA.

Les solutions d'IA sur cloud privé offrent plusieurs avantages importants aux entreprises qui nécessitent des solutions d'IA déployées de manière sécurisée sur site, dans des annexes, des datacenters ou autres. La proximité des données assure un accès et un traitement plus rapides et améliore l'efficacité et la performance des modèles IA. Une gouvernance complète simplifie la gestion en assurant l'identification des responsabilités et en facilitant des workflows collaboratifs sécurisés. Le contrôle des coûts est amélioré grâce à des modèles financiers flexibles qui permettent d'éviter les dépenses liées aux solutions de cloud public. De plus, l'optimisation de l'ensemble des technologies pour des cas d'utilisation spécifiques assure des déploiements IA sur mesure pour un maximum d'impact et d'efficacité.



Conclusion

L'évolution rapide de l'IA générative est une source d'opportunités mais aussi de défis pour les entreprises. Les entreprises qui cherchent à exploiter tout le potentiel de l'IA doivent naviguer au sein de paysages technologiques complexes, assurer une bonne sécurité et gérer efficacement l'évolutivité et les coûts. NVIDIA AI Computing by HPE fournit une solution complète à ces défis avec une infrastructure IA évolutive, sécurisée et aux performances optimisées. En tirant parti du cloud privé clés en main de HPE pour l'IA et de ses partenariats stratégiques, les entreprises peuvent accélérer leur passage à l'IA et réduire leurs délais de commercialisation tout en conservant le contrôle de leurs données et de leur infrastructure. Dans une perspective d'avenir, HPE reste engagé à stimuler l'innovation et à permettre aux entreprises de libérer le potentiel de transformation offert par l'IA. Grâce à un développement continu et à des collaborations stratégiques, HPE fournit des solutions de pointe qui permettront aux entreprises de prospérer dans un avenir guidé par l'IA.

Informations importantes concernant le présent rapport

CONTRIBUTEURS

Steven Dickens

Analyste-Chercheur | The Futurum Group

ÉDITEUR

Daniel Newman

PDG | The Futurum Group

DEMANDES DE RENSEIGNEMENTS

N'hésitez pas à nous contacter si vous souhaitez commenter ce rapport. Un représentant de The Futurum Group se chargera de vous répondre dans les meilleurs délais.

CITATIONS

Ce document peut être cité par la presse et les analystes accrédités, à la condition d'en indiquer précisément le contexte, le nom et le titre de l'auteur, ainsi que le nom « The Futurum Group ». Pour toute personne n'intervenant pas dans le cadre de la presse ou en tant qu'analyste, toute citation doit faire l'objet d'une autorisation écrite préalable de la part de The Futurum Group.

LICENCE

Ce document, y compris tout support associé, est la propriété de The Futurum Group. Cette publication ne doit pas être reproduite, distribuée ou partagée sous quelque forme que ce soit sans l'autorisation écrite préalable de The Futurum Group.

CONFLITS D'INTÉRÊTS

The Futurum Group fournit des services de recherche, d'analyse et de conseil à de nombreuses sociétés de haute technologie, y compris celles mentionnées dans ce document. Aucun des employés de cette entreprise ne possède des parts dans les sociétés citées dans ce document.



**Hewlett Packard
Enterprise**

À PROPOS DE HPE

HPE Financial Services combine des connaissances technologiques, une expertise financière et une démarche consciente en faveur de la durabilité pour créer des cycles de vie informatiques plus intelligents pour ses clients et partenaires quelle que soit leur taille. Nous travaillons dans tous les domaines technologiques, de l'edge to cloud jusqu'à l'utilisateur final, et notre approche collaborative produit des solutions de gestion des actifs qui non seulement libèrent du capital et optimisent les capacités, mais soutiennent aussi les pratiques de durabilité dans le monde entier en permanence. Pour plus d'informations, consultez : hpe.com/hpefinancialservices



NVIDIA

À PROPOS DE NVIDIA

[NVIDIA](#) fabrique les puces, systèmes et logiciels les plus avancés pour les usines IA du futur. Nous construisons de nouveaux services IA qui aident les sociétés à créer leurs propres usines IA. Depuis plus de 30 ans, scientifiques, chercheurs, développeurs et créateurs utilisent la technologie NVIDIA pour réaliser des choses incroyables. Plus de 4 millions de développeurs créent aujourd'hui des milliers d'applications qui accélèrent l'informatique. Plus de 40 000 sociétés utilisent les technologies IA NVIDIA, et NVIDIA Inception compte plus de 15 000 start-ups dans le monde.



À PROPOS DE THE FUTURUM GROUP

[The Futurum Group](#) est un cabinet indépendant de recherche, d'analyse et de conseil, spécialisé dans l'innovation numérique et dans des technologies et des tendances « disruptives ». Chaque jour, nos analystes, chercheurs et conseillers aident des dirigeants d'entreprise du monde entier à anticiper des changements structurels dans leur secteur d'activité, et à exploiter des innovations radicales pour acquérir ou maintenir un avantage concurrentiel au sein de leurs marchés.



COORDONNÉES

The Futurum Group LLC | futurumgroup.com | (833) 722-5337 |

