



Déployez l'intelligence artificielle générative avec Red Hat et NVIDIA

Principaux composants de la plateforme d'IA

Red Hat OpenShift AI : intégré à la plateforme d'applications de cloud hybride de Red Hat, ce module complémentaire optimisé par l'intelligence artificielle fournit aux entreprises des outils d'IA flexibles et évolutifs pour la création et la distribution d'applications basées sur l'IA à grande échelle dans les environnements de cloud hybride.

NVIDIA AI Enterprise : cette plateforme logicielle cloud-native complète facilite la gestion des pipelines de science des données ainsi que le développement et le déploiement de copilotes adaptés à la production et d'autres applications d'IA générative.

NVIDIA NIM : cet ensemble de microservices inclus avec la plateforme NVIDIA AI Enterprise est conçu pour faciliter le déploiement de modèles de fondation.

Red Hat et NVIDIA proposent une plateforme d'IA complète destinée aux entreprises

L'arrivée de l'intelligence artificielle (IA) et des applications basées sur l'IA a révolutionné l'utilisation des données pour l'optimisation de l'exploitation et la création de valeur pour les utilisateurs finaux, qu'il s'agisse d'équipes en interne ou de clients externes.

Pour intégrer efficacement l'IA et l'IA générative dans leurs applications, les entreprises ont besoin d'une plateforme capable de gérer la complexité de leurs charges de travail, d'optimiser l'utilisation du matériel, d'augmenter l'évolutivité et de prendre en charge les différents environnements de déploiement.

Afin de rationaliser le déploiement et l'évolutivité des applications basées sur l'IA générative, Red Hat et NVIDIA proposent une plateforme d'entreprise complète optimisée pour les charges de travail d'IA.

Red Hat et NVIDIA libèrent le potentiel des applications d'IA

L'intégration de la solution Red Hat® OpenShift® AI et des microservices NVIDIA NIM ajoute les fonctionnalités de gestion matérielle de NVIDIA à une plateforme d'applications flexible, évolutive et axée sur la sécurité. Cette association offre aux entreprises une infrastructure d'IA robuste, évolutive et efficace qui contribue à améliorer l'allocation des ressources, augmenter l'efficacité et accélérer l'exécution des charges de travail d'IA.

En exécutant [NVIDIA NIM sur Red Hat OpenShift et NVIDIA AI Enterprise 5.0](#), les entreprises renforcent l'efficacité et les performances de gestion des processeurs graphiques (GPU) de NVIDIA au sein de l'environnement Red Hat OpenShift. De plus, les applications peuvent exploiter le plein potentiel des composants matériels et logiciels basés sur l'IA de NVIDIA pour des performances et une efficacité renforcées.

Rationalisation du déploiement des applications basées sur l'IA. La plateforme Red Hat OpenShift AI fournit des outils de création, de déploiement et de gestion d'applications basées sur l'IA. Les utilisateurs peuvent les associer aux conteneurs préconçus fournis par le logiciel d'inférence NVIDIA intégré à NVIDIA NIM pour réduire considérablement les délais de déploiement.

Amélioration des performances. Déployées dans l'environnement d'orchestration de conteneurs fiable et cohérent de Red Hat OpenShift AI, les capacités d'optimisation logicielle de NVIDIA AI Enterprise et les fonctionnalités de calcul haute performance des GPU NVIDIA renforcent les performances des applications basées sur l'IA.

Réduction de la complexité de la gestion. Red Hat OpenShift AI et NVIDIA NIM sont dotés de solutions préintégrées et d'outils préconfigurés qui aident à simplifier le déploiement et la gestion des applications basées sur l'IA afin de réduire les difficultés liées à l'adoption des technologies d'analyse avancée.

f facebook.com/redhatinc
X [@RedHatFrance](https://twitter.com/RedHatFrance)
in linkedin.com/company/red-hat

Amélioration de l'évolutivité et de la flexibilité. L'association des fonctionnalités de la solution complète de NVIDIA et de la plateforme Red Hat permet la prise en charge d'options de déploiement flexibles qui s'adaptent à l'évolution des besoins métier et du volume de données de toutes les entreprises.

Priorité à la sécurité et à la conformité. Axés sur la sécurité, les environnements et les fonctions de cette solution conjointe aident les entreprises à se concentrer sur la conformité et la sécurité des données et des modèles d'IA dans tous les types d'environnements informatiques.

Rationalisez l'innovation par l'IA avec une plateforme complète, flexible et cohérente

L'IA contribue à améliorer l'analyse des données, à rationaliser l'exploitation ou encore à innover en périphérie du réseau. Quel que soit leur projet, Red Hat et NVIDIA aident les entreprises à transformer leurs données en informations exploitables et en opportunités métier.

Exploitation de l'inférence avancée sur les modèles d'IA de toutes sortes

En exécutant NVIDIA NIM sur Red Hat OpenShift, les entreprises peuvent intégrer l'inférence d'IA avancée dans leurs applications afin d'améliorer la prise de décision en temps réel et l'efficacité des charges de travail. Elles bénéficient ainsi d'un environnement d'exécution flexible pour l'inférence et de la prise en charge de différents modèles d'IA, notamment ceux de la communauté Open Source, les modèles de fondation d'IA NVIDIA ou encore leurs propres modèles personnalisés.

Automatisation du provisionnement des GPU à l'aide du logiciel NVIDIA

Le déploiement de NVIDIA AI Enterprise (et de NVIDIA NIM) sur Red Hat OpenShift AI permet à l'opérateur des GPU NVIDIA d'utiliser le framework de Kubernetes pour automatiser la gestion de tous les composants logiciels NVIDIA nécessaires au provisionnement des GPU, notamment les pilotes NVIDIA, les plug-ins d'appareils Kubernetes pour les GPU, NVIDIA Container Runtime, l'étiquetage automatique des nœuds et la surveillance avec l'outil DCGM.

Cohérence et flexibilité pour le développement, le déploiement et la mise à l'échelle des applications d'IA

La plateforme Red Hat OpenShift AI offre un ensemble complet d'outils de développement et assure la portabilité des applications. Associée aux microservices NVIDIA NIM, elle rationalise la création des applications basées sur l'IA. Cette combinaison fournit également une expérience de développement et de gestion homogène des microservices NVIDIA NIM dans tous les environnements informatiques, ainsi qu'une interface de programmation d'application cohérente basée sur les normes OpenAI pour les modèles d'IA générative.

Optimisez le développement de l'IA générative avec Red Hat et NVIDIA

[Contactez un représentant Red Hat](#) pour découvrir comment la plateforme Red Hat, combinée à l'architecture et aux microservices d'IA NVIDIA, peut vous aider à développer et déployer vos applications basées sur l'IA générative de manière cohérente, flexible et évolutive tout en assurant leur sécurité et leur conformité.



À propos de Red Hat

Premier éditeur mondial de solutions Open Source, Red Hat s'appuie sur une approche communautaire pour fournir des technologies Linux, de cloud hybride, de conteneurs et Kubernetes fiables et performantes. Red Hat aide ses clients à développer des applications cloud-native, à intégrer des applications nouvelles et existantes ainsi qu'à gérer et à automatiser des environnements complexes. [Conseiller de confiance auprès des entreprises du Fortune 500](#), Red Hat propose des services d'assistance, de formation et de consulting [reconnus](#) qui apportent à tout secteur les avantages de l'innovation ouverte. Situé au cœur d'un réseau mondial d'entreprises, de partenaires et de communautés, Red Hat participe à la croissance et à la transformation des entreprises et les aide à se préparer à un avenir toujours plus numérique.