

# Développement, entraînement, test et déploiement de l'AA dans le cloud hybride

## Avantages clés

Distribuez rapidement des applications intelligentes sur site ou dans le cloud public.

Augmentez l'adoption de l'IA dans votre entreprise en facilitant son adoption et gagnez en flexibilité.

Harmonisez votre stratégie d'IA/AA grâce à une expérience utilisateur cohérente qui favorise une collaboration efficace entre vos équipes de science des données, d'ingénierie des données et de DevOps.

Profitez de la flexibilité du cloud hybride pour créer, entraîner, déployer et surveiller les charges de travail d'IA/AA dans le cloud, sur site ou en périphérie de réseau, au plus près de vos données.

---

[Open Data Hub](#) est un modèle pour la création d'une plateforme d'IA en tant que service sur [Red Hat OpenShift](#). Il découle de projets en amont tels que [Jupyter](#) et [Kubeflow](#), et constitue la base de Red Hat OpenShift AI.

## Accélération des déploiements d'intelligence artificielle et d'apprentissage automatique

L'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond (IA/AA/AP) sont rapidement devenus indispensables pour les entreprises, qui adoptent également l'IA générative à un rythme de plus en plus soutenu. Gartner estime désormais que 85 % des entreprises auront utilisé des interfaces de programmation d'application (API) d'IA générative ou déployé des applications basées sur cette technologie d'ici 2026<sup>1</sup>. Le déploiement de ces technologies présente toutefois certaines difficultés. En effet, les data scientists qui cherchent à développer leurs modèles sont souvent confrontés à un manque d'harmonisation entre des outils en constante évolution. Ces disparités peuvent nuire à la productivité et à la collaboration entre les équipes de science des données, de développement logiciel et d'exploitation. Souvent, la mise à l'échelle des déploiements d'IA/AA représente une tâche difficile à administrer et pour laquelle les entreprises manquent de ressources, alors qu'elle nécessite des processeurs graphiques (GPU) coûteux pour l'accélération matérielle ainsi que des charges de travail distribuées pour l'IA générative. Certaines plateformes cloud très prisées offrent une évolutivité et des outils intéressants, mais ces derniers enferment souvent les utilisateurs dans une dépendance et limitent les options d'architecture et de déploiement.

Basé sur le projet Open Source [Open Data Hub](#), le service [Red Hat® OpenShift® AI](#)<sup>2</sup> permet aux data scientists d'entraîner, de tester, de mettre à disposition et de surveiller rapidement les modèles d'AA/AP, y compris d'IA générative. Ainsi, les utilisateurs peuvent immédiatement se concentrer sur leur modélisation et le développement de leurs applications sans attendre le provisionnement de l'infrastructure. Disponible en tant que module complémentaire pour Red Hat OpenShift, sous forme de service cloud entièrement géré ou de logiciel autogéré, OpenShift AI associe des composants de Red Hat, des logiciels Open Source et des offres de partenaires technologiques à un niveau de flexibilité qui permet de développer et mettre à disposition des modèles sur site, dans le cloud ou sur une infrastructure d'edge computing.

## Red Hat OpenShift AI

Le service OpenShift AI permet aux entreprises de déployer efficacement un ensemble intégré d'outils courants tiers et Open Source pour la modélisation IA/AA. Cette plateforme se propose de remplacer les suites d'IA/AA dites « orientées » (« opinionated ») des autres fournisseurs de cloud. Les entreprises qui l'adoptent bénéficient d'outils Open Source collaboratifs et d'une plateforme pour la création de modèles expérimentaux, sans avoir à se préoccuper de l'infrastructure ou d'une dépendance aux outils propres au cloud public. Elles peuvent même étendre cette plateforme de base à l'aide d'outils de partenaires afin d'augmenter leurs capacités. Les modèles peuvent être mis à disposition de manière cohérente dans différents environnements de production, dans un format compatible avec les conteneurs, que ce soit dans le cloud hybride ou en périphérie du réseau. OpenShift AI fournit aux équipes d'exploitation un environnement facile à gérer, avec des configurations simples sur une plateforme éprouvée, évolutive et sécurisée.

Ce service prend en charge les modèles d'IA générative les plus répandus et vous permet de régler, d'ajuster et de mettre à disposition ces modèles préentraînés en fonction de vos cas d'utilisation uniques et de vos propres données. Vous pouvez même distribuer les charges de travail sur plusieurs clusters Red Hat OpenShift, indépendamment de leur emplacement. La plateforme facilite également l'exploitation de l'accélération matérielle de l'IA. Elle prend en charge l'infrastructure matérielle basée sur un processeur et un processeur graphique (GPU), notamment les GPU NVIDIA et les XPU Intel, et vous épargne la mise en place et la gestion de votre propre plateforme de science des données.

---

<sup>1</sup> Communiqué de presse de Gartner, « [Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026](#) » 11 oct. 2023

<sup>2</sup> Anciennement Red Hat OpenShift Data Science

Les [services de consulting Red Hat](#) proposent le contrat [OpenShift AI Pilot](#) pour aider les entreprises à adopter le service OpenShift AI et à l'intégrer à leur environnement de travail existant.

Ils proposent également un service de consulting [MLOps Foundation](#) aux entreprises qui souhaitent aller au-delà de l'expérimentation de modèles pour développer des stratégies de déploiement de modèles en production.

## Outils commerciaux de partenaires technologiques et outils Open Source en amont

Le service Red Hat OpenShift AI fournit un ensemble d'outils (voir Tableau 1) issus du projet communautaire Open Data Hub. Les entreprises peuvent développer, tester et déployer des modèles dans tous les environnements cloud et sur une plateforme Red Hat OpenShift entièrement gérée ou autogérée, et centraliser la surveillance de leurs performances. En outre, Red Hat publie régulièrement des mises à jour pour les outils Open Source (comme Jupyter, PyTorch et TensorFlow) et vous dispense ainsi des tâches d'intégration, de test et de maintenance. L'offre inclut également plusieurs produits d'IA/AA proposés par des partenaires technologiques (voir Tableau 1). Par ailleurs, il est possible d'ajouter d'autres offres commerciales de plus de 30 partenaires technologiques pour l'IA dont les produits ont été certifiés sur Red Hat OpenShift.

**Tableau 1. Écosystème Red Hat OpenShift AI**

<b>Outils de modélisation et de visualisation d'IA/AA</b>	<a href="#">JupyterLab UI</a> avec des images de notebook préconçues ainsi que des bibliothèques et paquets Python courants ; <a href="#">TensorFlow</a> ; <a href="#">PyTorch</a> , CUDA ; Kubeflow, contrôleur de notebooks pour la gestion de plusieurs sessions de notebooks ; <a href="#">Anaconda</a> (souscription Professional facultative) ; <a href="#">outils d'IA d'Intel</a>
<b>Ingénierie des données</b>	<a href="#">Starburst</a> (souscriptions Galaxy et Enterprise facultatives) ; <a href="#">Pachyderm</a> (facultatif)
<b>Ingestion et stockage de données</b>	<a href="#">Red Hat AMQ</a> (module complémentaire facultatif) ; <a href="#">Amazon Simple Storage Service (S3)</a>
<b>Prise en charge GPU</b>	<a href="#">NVIDIA</a> (avec opérateur GPU), XPU Intel (avec <a href="#">processeurs Intel Xeon</a> , <a href="#">Habana Gaudi</a> et <a href="#">Intel Data Center GPU Flex Series</a> )
<b>Mise à disposition et surveillance de modèles</b>	Mise à disposition de modèles (KServe avec interface utilisateur), surveillance de modèles, OpenShift <a href="#">Source-to-Image (S2I)</a> , Red Hat OpenShift API Management (module complémentaire facultatif), <a href="#">Intel Distribution of OpenVINO Toolkit</a>
<b>Pipelines de science des données</b>	Les <a href="#">pipelines de science des données</a> (Kubeflow Pipelines) regroupent certains processus tels que la préparation des données ainsi que la création et la mise à disposition de modèles.



## À propos de Red Hat

Red Hat aide ses clients à standardiser leurs environnements, à développer des applications cloud-native et à intégrer, automatiser, sécuriser et gérer des environnements complexes en offrant des services d'assistance, de formation et de consulting [primés](#).

[f](#) facebook.com/redhatinc  
[t](#) @RedHatFrance  
[in](#) linkedin.com/company/red-hat

EUROPE, MOYEN-ORIENT  
ET AFRIQUE (EMEA)  
00800 7334 2835  
europe@redhat.com

FRANCE  
00 33 1 41 91 23 23  
fr.redhat.com