

Accélérer l'adoption de l'intelligence artificielle dans le secteur des services financiers



Le déploiement des technologies d'IA atteint des records dans le monde entier.

- ▶ 94 % des dirigeants d'entreprise pensent que l'IA sera un facteur de réussite déterminant dans les cinq prochaines années².
- ▶ 79 % des entreprises ont déployé au moins trois types d'applications d'IA².
- ▶ 76 % des dirigeants prévoient d'investir davantage dans l'IA².

Les applications intelligentes, un facteur de réussite

Les modèles d'intelligence artificielle (IA) et d'apprentissage automatique (AA) ont besoin de grandes quantités de données pour fournir des informations utiles, automatiser des tâches et améliorer les capacités métier essentielles. Ces technologies ont le potentiel de transformer le secteur des services financiers à tous les niveaux, des clients et du personnel aux équipes de développement et d'exploitation. Les organismes de services financiers utilisent des modèles d'IA/AA pour développer des applications intelligentes et cloud-native qui offrent des résultats mesurables, comme l'augmentation de la satisfaction des clients, la diversification des offres de service, le renforcement de l'automatisation métier et informatique, ou encore l'augmentation de l'efficacité et de la productivité du personnel. En réalité, pour 36 % des organismes de services financiers, les technologies d'IA/AA entraînent une baisse des frais annuels de plus de 10 %¹.

Aujourd'hui, beaucoup d'organismes de services financiers intègrent des modèles d'IA/AA dans leurs solutions déployées au niveau d'une unité commerciale ou d'une division, dans le but d'optimiser des activités jusqu'ici réalisées manuellement. Avec des algorithmes d'AA, il est par exemple possible de détecter des transactions ou des schémas de dépenses inhabituels qu'un analyste humain aurait manqués, puis de les signaler à l'équipe de prévention des fraudes pour qu'elle mène l'enquête. Les technologies d'IA/AA peuvent analyser l'activité d'un compte pour détecter des comportements suspects et révéler des liens entre des comptes qui dissimulent leur appartenance à un système complexe de blanchiment d'argent. Par ailleurs, les chatbots basés sur l'IA sont capables de répondre aux demandes d'assistance courantes rapidement et précisément, voire d'apprendre des interactions passées, ce qui améliore l'expérience client et permet au personnel de se consacrer à des problèmes plus épineux. Enfin, en associant les nouveaux algorithmes d'AA aux technologies classiques de reconnaissance optique des caractères (OCR), les institutions financières peuvent améliorer la netteté des documents scannés et numérisés dans le cadre des processus de connaissance client.

Cependant, il peut s'avérer difficile de créer des applications intelligentes et déployer des pratiques MLOps (Machine Learning Operations) en production. Les organismes de services financiers sont confrontés à de nombreux défis lorsqu'ils souhaitent proposer des services novateurs basés sur des modèles d'IA/AA :

- ▶ Parce que les spécialistes de l'IA/AA se font rares, il est difficile de trouver et fidéliser du personnel compétent pour constituer des équipes de science des données, d'ingénierie, d'ingénierie de l'AA ou encore de développement logiciel.
- ▶ Le manque de liens et de collaboration entre les membres du personnel, notamment entre les équipes de science des données, d'ingénierie de l'AA ou de développement, ralentit les workflows IA/AA et le cycle de vie des modèles.
- ▶ Les différentes étapes du workflow IA/AA (préparation et gestion des données, entraînement et ajustement des modèles, et inférence) nécessitent une infrastructure coûteuse, parfois difficile à déployer et à utiliser de façon automatisée et en libre-service.

1 NVIDIA, « [State of AI in Financial Services: 2023 Trends](#) », 2023

2 Deloitte, « [State of AI in the Enterprise, 5th Edition](#) », octobre 2022



Principaux cas d'utilisation de l'IA dans les services financiers

Les organismes de services financiers utilisent efficacement l'IA/AA pour divers cas d'utilisation¹ :

- ▶ Traitement du langage naturel et grands modèles de langage
- ▶ Systèmes de recommandations et NBA (Next-Best Action)
- ▶ Optimisation des offres
- ▶ Détection des transactions et paiements frauduleux
- ▶ Prévention du blanchiment d'argent et initiatives de connaissance client
- ▶ Commerce basé sur des algorithmes
- ▶ Systèmes d'IA conversationnelle
- ▶ Optimisation marketing
- ▶ Analyse des risques

- ▶ Les processus d'entraînement des modèles d'IA/AA sont soumis à des exigences de gouvernance et des réglementations en matière de souveraineté des données, qui limitent le partage des données avec certains pays ou certains services d'une institution.
- ▶ Les solutions et pipelines d'IA/AA sont plus complexes à déployer que les applications standard parce qu'ils doivent prendre en charge plus de tâches : le développement, l'entraînement et l'inférence de modèles traitant de gros volumes de données, l'intégration et le déploiement de ces modèles dans des applications intelligentes, et la possibilité de surveiller, mettre à jour et réentraîner les modèles en fonction des comportements observés.

Surmontez les défis en matière d'IA/AA avec les technologies de cloud hybride, de conteneurs et Kubernetes

Pour gérer la complexité du développement et du déploiement des applications intelligentes basées sur l'IA/AA, il faut associer des outils et des meilleures pratiques modernes et cloud-native. Les technologies de conteneurs et Kubernetes offrent l'agilité de déploiement, les capacités de gestion et l'évolutivité indispensables à la distribution et à la gestion d'applications intelligentes. Un conteneur est une unité de logiciel légère et autonome qui contient une application avec toutes ses dépendances (environnement d'exécution, bibliothèques, outils système et paramètres) et produit une image facile à distribuer. Kubernetes est une plateforme Open Source d'orchestration des conteneurs qui crée, déploie, met à l'échelle et gère automatiquement les instances de conteneurs dans l'environnement.

Utilisées comme base pour les solutions d'IA/AA, les plateformes d'orchestration des conteneurs Kubernetes offrent de nombreux avantages. Grâce aux capacités d'automatisation et en libre-service, il est possible de provisionner les environnements MLOps à la demande pour accélérer et simplifier le développement, les tests et le déploiement des modèles d'IA/AA. La portabilité des conteneurs permet une utilisation cohérente des modèles entre toutes les infrastructures (les environnements physiques et virtuels, les clouds publics, privés et hybrides), et ce sans modification. Kubernetes met automatiquement à l'échelle les charges de travail des conteneurs pour garantir la disponibilité des applications, en allouant les ressources uniquement aux charges de travail qui en ont besoin. Avec cette plateforme technologique commune, il est aussi plus facile d'enrichir sa solution d'IA/AA à l'aide de produits commerciaux et Open Source issus d'un vaste écosystème de fournisseurs.

Malgré tout, pour être efficace, une solution d'IA/AA a besoin de plus que de fonctionnalités Kubernetes standard. Les plateformes d'applications modernes complètent Kubernetes avec des outils avancés et des capacités étendues. Les outils d'intégration et de déploiement continus, tels que Tekton et Jenkins, accélèrent la création, les tests, la mise en paquet, la mise à jour et le déploiement des modèles d'IA/AA. Les outils GitOps de distribution continue comme ArgoCD permettent de définir et d'automatiser des déploiements d'applications complexes en tant que code. Comme il est possible de centraliser toutes les informations d'un modèle via des services de surveillance standardisés et interfaces de programmation d'application (API), vous pouvez mesurer la précision des modèles, détecter d'éventuels biais, et lancer des mises à jour ou des réentraînements. L'intégration des conteneurs et des applications traditionnelles qui s'exécutent souvent sur des machines virtuelles permettent d'utiliser les outils existants avec la solution d'IA/AA. En outre, l'entraînement des modèles est rendu plus simple et plus performant par les connexions en libre-service à de nombreux flux de données en temps réel et magasins de données compatibles CSI (Container Storage Interface).

Développez et déployez des applications intelligentes dans un environnement unifié

Les technologies Red Hat® peuvent être associées pour créer une base MLOps unique et flexible qui prend en charge l'ensemble du cycle de vie de l'IA/AA, du développement et de l'entraînement des modèles à l'intégration et au déploiement des applications intelligentes. Vous disposez ainsi des outils nécessaires pour créer et déployer des applications novatrices de manière cohérente et rationalisée, et ce, sur tout type d'infrastructure : matériel physique, machines virtuelles, et clouds privés, publics ou hybrides. Et vous avez accès à un large éventail de produits partenaires certifiés et de composants Open Source pris en charge pour personnaliser votre environnement selon les besoins de votre entreprise.



Les organismes de services financiers exécutent leurs charges de travail IA/AA dans divers environnements :

- ▶ 44 % utilisent des environnements de cloud hybride¹.
- ▶ 32 % utilisent des environnements de cloud public uniquement¹.
- ▶ 16 % utilisent des datacenters sur site uniquement¹.
- ▶ 5 % utilisent des environnements d'edge computing¹.
- ▶ 4 % utilisent des environnements de cloud privé¹.

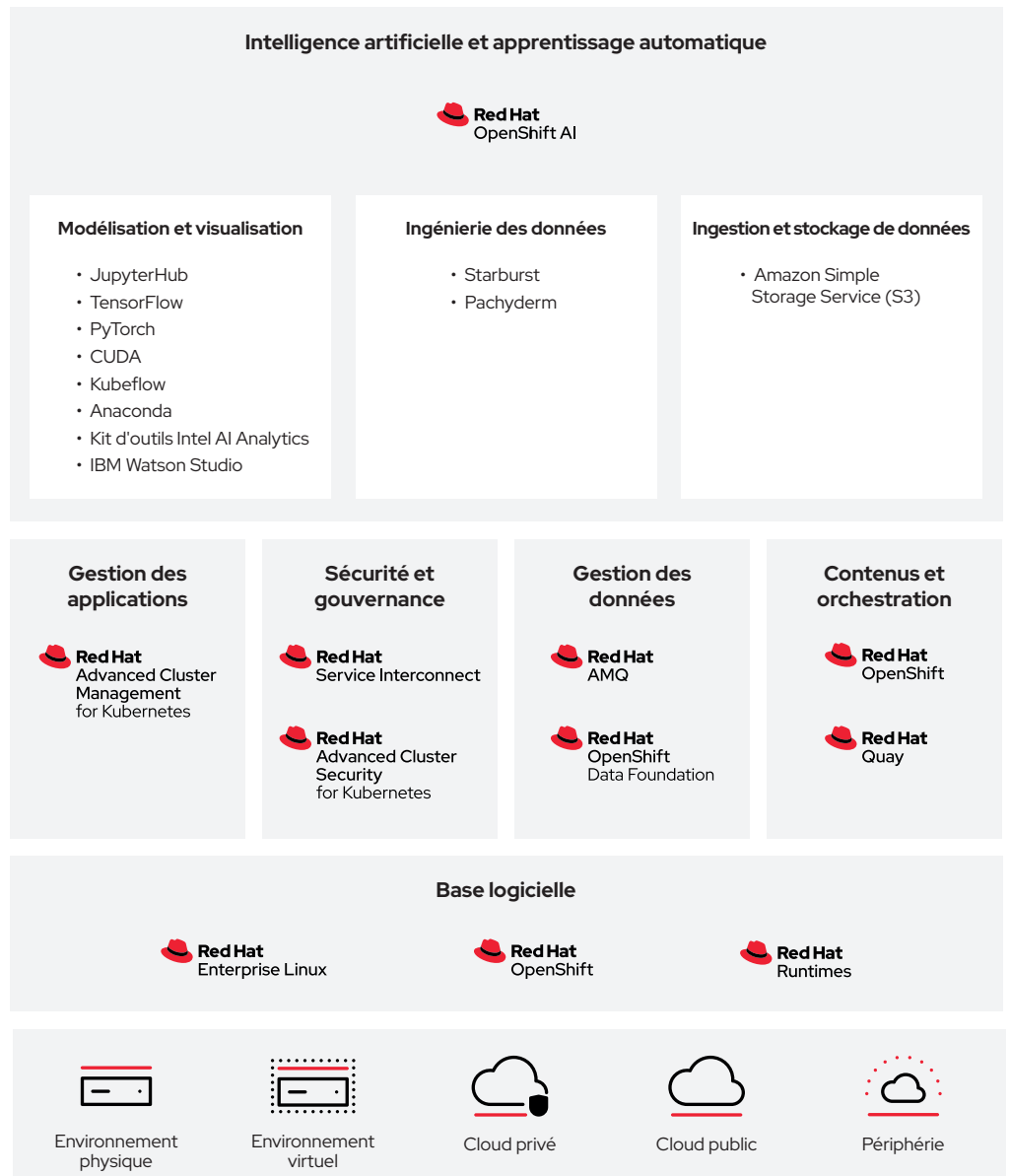


Figure 1 : intégration de technologies clés dans des solutions unifiées, pour un déploiement MLOps cohérent et standardisé



Open Source et IA/AA

Partout dans le monde, des équipes de recherche et développement créent des technologies d'IA/AA parmi les plus avancées en utilisant des modèles Open Source.

En tant que principal éditeur Open Source au monde, nous sommes convaincus que les modèles de développement Open Source permettent de créer des technologies plus stables, sécurisées et novatrices.

Chez Red Hat, nous collaborons sur des projets communautaires et protégeons des licences Open Source afin de continuer à développer des logiciels qui aident les clients à relever leurs défis les plus complexes.

[En savoir plus](#) sur notre engagement en faveur de l'Open Source

Chaque couche de la solution assure des fonctions clés en lien avec le développement, le déploiement et la gestion des modèles d'IA/AA et des applications intelligentes.

- ▶ La **base logicielle** comprend le système d'exploitation et les outils d'orchestration des conteneurs. Elle assure la cohérence, la stabilité et la sécurité de l'ensemble de la solution d'IA/AA. Grâce à des capacités **DevOps** et d'automatisation intégrées, comme le provisionnement en libre-service, ainsi qu'aux nombreux outils, langages et environnements d'exécution inclus, vous pouvez rapidement développer des modèles et applications d'IA/AA cloud-native et hautement distribués novateurs, et en assurer la maintenance.
- ▶ La **gestion des applications** est une couche constituée des outils de déploiement et de gestion des applications et services cloud-native dans les environnements de cloud hybride. La cohérence et l'efficacité de l'exécution des applications sont assurées par les fonctions de surveillance et d'application des politiques.
- ▶ La **gestion des données** englobe les outils de gestion des flux de données sécurisés et évolutifs, ainsi que le stockage de données structurées et non structurées pour l'entraînement et le stockage des modèles d'IA/AA. Un système de stockage hautement évolutif permet d'entraîner de grands modèles d'IA/AA, avec un niveau de précision élevé.
- ▶ Les outils et technologies **de contenus et d'orchestration** facilitent la gestion de l'environnement et des workflows MLOps. Les registres de conteneurs ainsi que les frameworks et workflows d'intégration et de déploiement continu (CI/CD) garantissent la cohérence et la reproductibilité des processus de création et de déploiement d'applications dans toute l'infrastructure.
- ▶ La couche **Sécurité et gouvernance** inclut des technologies qui protègent les charges de travail d'entraînement et d'inférence, et qui préviennent tout accès non autorisé. Les données sensibles en transit sont mieux protégées avec une sécurité renforcée entre les datacenters et les régions.
- ▶ Les technologies **d'intelligence artificielle et d'apprentissage automatique** incluent des outils et services certifiés pour le développement, l'entraînement et l'utilisation des modèles d'IA/AA. La plateforme sous-jacente offre des capacités de provisionnement en libre-service qui permettent de passer rapidement de l'expérimentation à la production de manière cohérente et collaborative.

Créez une base MLOps ouverte et flexible avec nos produits

Leader dans le domaine des logiciels Open Source d'entreprise, nous proposons une large gamme de produits, une expertise éprouvée et des partenariats stratégiques avec les principaux éditeurs de logiciels indépendants (ISV) afin de vous aider à créer votre architecture de solutions d'IA/AA. Notre base sécurisée pour la création d'environnements MLOps prêts pour la production repose sur une sélection de technologies Open Source certifiées et prises en charge. Grâce à la cohérence entre le datacenter et l'infrastructure cloud, vous pouvez créer des modèles d'IA/AA et des applications intelligentes qui fournissent des informations vraiment utiles et des expériences utilisateur de haute qualité.

Chaque composant de notre base gère une fonctionnalité clé du MLOps.

La gamme **Red Hat OpenShift® AI** associe les capacités éprouvées de Red Hat OpenShift et de Red Hat OpenShift Data Science dans une base cohérente et évolutive, pour un développement d'applications novatrices rapide et un déploiement orchestré entre les infrastructures.

- ▶ **Red Hat OpenShift** est une plateforme d'applications unifiée pour les entreprises, conçue pour le développement, le déploiement et l'orchestration cloud-native. Les tâches de modélisation et d'inférence sont accélérées par les ressources de calcul à la demande et la prise en charge de l'accélération du matériel et des processeurs graphiques (GPU), notamment via les GPU NVIDIA et la technologie Intel® Deep Learning Boost (Intel DL Boost). La cohérence dans les environnements sur site, de cloud public et d'edge computing fournit la vitesse et la flexibilité nécessaires au succès des équipes. Vous pouvez par exemple créer une plateforme MLOps



Témoignage client : Banco Galicia

Avec l'aide de l'équipe de consulting Red Hat, Banco Galicia a créé une plateforme de traitement du langage naturel (TLN) intelligente et basée sur l'IA qui repose sur les solutions Red Hat OpenShift, Red Hat Integration et Red Hat Single Sign-On (SSO).

Principaux résultats :

- ▶ Réduction des délais de vérification de **plusieurs jours à quelques minutes**, avec **90 %** de précision
- ▶ Diminution de **40 %** des temps d'arrêt des applications
- ▶ Multiplication de l'agilité par **4**

Lire le [témoignage client](#)

en libre-service pour les équipes de science des données, d'ingénierie des données et de développement afin de produire rapidement des modèles, de les incorporer aux applications et de réaliser des tâches d'inférence. Les fonctions de collaboration permettent aux équipes de créer et partager des résultats de modélisation dans des conteneurs de manière cohérente.

Red Hat OpenShift contient plusieurs outils essentiels pour le développement et la distribution d'applications intelligentes. L'opérateur [Red Hat OpenShift Virtualization](#) permet d'intégrer, d'exécuter et de gérer les applications virtualisées existantes en tant qu'objets Kubernetes natifs. Avec [Red Hat OpenShift Pipelines](#), vous pouvez concevoir des pipelines d'IA/AA dans un framework CI/CD natif pour Kubernetes, et exécuter chaque étape de ces pipelines dans des conteneurs isolés dont la mise à l'échelle s'effectue à la demande. Enfin, l'opérateur [Red Hat OpenShift GitOps](#) fournit un workflow de déploiement continu qui gère automatiquement l'environnement MLOps en fonction des configurations que vous créez et stockez dans Git.

Inclus dans Red Hat OpenShift, [Red Hat Enterprise Linux®](#) est un système d'exploitation cohérent et évolutif aux performances élevées pour les datacenters, et les environnements cloud et d'edge computing. Ses fonctions de sécurité intégrées, comme [Security-Enhanced Linux \(SELinux\)](#), vous protègent contre les menaces et assurent la conformité avec les exigences sectorielles et réglementaires. De plus, grâce à l'interopérabilité de nos plateformes, ces fonctions et certifications couvrent toute la pile logicielle.

- ▶ [Red Hat OpenShift Data Science](#) est une plateforme d'IA basée sur le projet Open Source [Open Data Hub](#). Elle offre aux équipes de science des données et de développement une puissante plateforme d'IA/AA pour rassembler les informations et créer des applications intelligentes, et inclut des outils et des frameworks tels que Jupyter, TensorFlow et PyTorch. Le service comprend également plusieurs composants de partenaires technologiques comme Starburst, Anaconda, IBM et Intel, qui viennent accélérer davantage le développement d'applications intelligentes pour les services financiers.

L'offre [Red Hat Integration](#) est un ensemble complet de technologies d'intégration et de messagerie qui permettent de connecter les applications et les données au sein des infrastructures hybrides. Voici ses principaux composants :

- ▶ [Red Hat Runtimes](#) est un ensemble de produits, d'outils et de composants (notamment des environnements d'exécution et frameworks légers) pour le développement et la maintenance des applications d'IA/AA cloud-native et hautement distribuées.
- ▶ [Red Hat AMQ](#) est une plateforme de messagerie flexible, évolutive et distribuée qui transmet les données aux modèles d'IA/AA avec un débit élevé et une faible latence.

[Red Hat OpenShift Data Foundation](#) est un système de stockage logiciel en modes bloc, fichier et objet pour stocker les grands ensembles de données d'IA/AA de manière persistante et évolutive.

[Red Hat Service Interconnect](#) est un protocole de communication OTT qui simplifie la connectivité entre les applications et les services répartis dans plusieurs datacenters ou régions. Il peut être utilisé en toute sécurité par tous les membres de votre équipe de développement, quels que soient leurs privilèges.

[Red Hat Advanced Cluster Management for Kubernetes](#) est une console unifiée pour le contrôle, l'automatisation et la surveillance du déploiement des applications, de la gestion des clusters et de l'application des politiques à grande échelle. Elle se base sur l'emplacement géographique de votre environnement cloud.

[Red Hat Advanced Cluster Security for Kubernetes](#) permet de protéger les charges de travail Kubernetes conteneurisées dans le cloud hybride, renforçant ainsi la sécurité des applications.

Plateforme de registre privé à la sécurité renforcée, [Red Hat Quay](#) permet de gérer le contenu cloud-native dans les environnements de cloud hybride.

66

« Red Hat nous a aidés à mettre en œuvre notre plateforme de TLN en nous fournissant l'architecture et les technologies dont nous avons besoin. Les spécialistes Red Hat nous ont bien expliqué comment tirer parti de Red Hat OpenShift. Nous avons également commencé à concevoir une architecture qui pourrait être cloud-native. »

Matias Lorusso

Architecte de solutions,
Banco Galicia

Enfin, notre **écosystème de partenaires certifiés** permet d'intégrer des outils d'IA/AA, d'analyse des données, de gestion, de stockage, de sécurité et de développement à cette architecture. Chez Red Hat, nous travaillons en étroite collaboration avec nos partenaires afin de certifier leurs logiciels sur nos plateformes, pour plus de facilité de gestion, une sécurité renforcée et une meilleure prise en charge. En outre, de nombreux partenaires proposent des **opérateurs Red Hat OpenShift certifiés** qui simplifient la gestion du cycle de vie des logiciels.

Tirez parti des grands modèles de langage

Dans le secteur des services financiers, les applications de numérisation des documents, d'analyse des rapports et de service conversationnel s'appuient souvent sur de grands modèles de langage (LLM) tels que GPT-4 (Generative Pre-trained Transformer 4). Parce que la création de LLM demande beaucoup d'efforts et de ressources de calcul, les entreprises utilisent généralement des modèles pré-entraînés dans ces applications. Toutefois, afin de fournir des résultats précis pour le domaine ou l'entreprise concernés, ces modèles ont quand même besoin d'être entraînés et ajustés avec un ensemble réduit de données locales. De plus, chaque application nécessite un modèle entraîné et ajusté spécialement pour elle.

La plateforme Red Hat OpenShift AI est idéale pour l'entraînement et l'ajustement des LLM. Avec une plateforme unique, cohérente et évolutive pour le développement, l'entraînement et l'inférence des modèles, ainsi que pour l'intégration et le déploiement des applications, vous évitez d'effectuer deux fois la même tâche, avec à la clé une amélioration de l'efficacité de l'ensemble du cycle de vie d'IA/AA. Les capacités avancées de gestion des clusters et en libre-service permettent de créer des pipelines d'IA/AA reproductibles et réutilisables dans plusieurs modèles et applications. D'ailleurs, les entreprises qui ont déployé Red Hat OpenShift AI constatent un gain d'efficacité de 20 % chez les data scientists³.

En savoir plus

Avec notre gamme complète de technologies, notre expérience éprouvée et nos partenariats stratégiques, nous pouvons vous aider à atteindre vos objectifs. Nous vous proposons une base solide pour la création d'environnements d'IA/AA prêts pour la production, ainsi que des services et des formations pour accélérer l'adoption de notre solution. Quelle que soit votre progression dans votre parcours vers l'IA/AA, nous pouvons vous aider à créer un environnement MLOps prêt pour la production qui accélère le développement et la distribution des applications intelligentes.

- ▶ [Apprenez-en plus](#) sur les solutions d'IA/AA pour le secteur des services financiers.
- ▶ [Organisez une session de découverte gratuite](#) pour découvrir comment nous pouvons vous aider à déployer une base pour le MLOps.

³ Forrester Consulting, étude commissionnée par Red Hat, « [The Total Economic Impact™ of Red Hat Hybrid Cloud Platform for MLOps](#) », mars 2022. Ces résultats concernent une entreprise mixte et représentent les clients interrogés.



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 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.

f facebook.com/redhatinc
t @RedHatFrance
in linkedin/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