

Préparation des données : la nouvelle devise de la ruée vers l'IA

La contrainte cachée de la ruée vers l'IA

L'adoption massive de l'IA est déjà bien amorcée. Les entreprises investissent massivement dans de nouveaux modèles, des accélérateurs spécialisés et des plateformes de développement modernes afin d'obtenir plus rapidement des informations et un avantage concurrentiel. Pourtant, pour bon nombre d'entre elles, les résultats ne progressent pas au même rythme que les investissements.

L'hypothèse généralement admise est que le succès de l'IA dépend principalement des modèles et de la puissance de calcul. Choisissez la bonne architecture, déployez suffisamment de GPU, et la valeur suivra. En pratique, ce point de vue néglige une contrainte plus fondamentale : l'efficacité de l'IA dépend de la qualité des données qui l'alimentent, et dans la plupart des entreprises, la préparation des données est négligée.

À mesure que les organisations déploient l'IA au-delà des pilotes, elles se retrouvent confrontées à des données fragmentées, difficiles d'accès et mal adaptées à l'exploitation par les machines. L'ambition devance rapidement l'exécution. L'IA est par conception une charge de travail hybride, s'étendant aux systèmes sur site, aux clouds et à l'edge. Les données qui l'alimentent sont donc encore plus distribuées. Toute stratégie sérieuse en matière d'IA doit commencer par prendre en compte cette réalité.

La plupart des entreprises ne sont pas réellement prêtes pour l'IA

Malgré des années de modernisation, la plupart des entreprises travaillent encore avec des parcs de données qui n'ont pas été conçus pour l'IA. Les données restent fragmentées entre les systèmes sur site, les clouds privés/publics et les environnements edge, souvent réparties entre différents outils, formats et modèles opérationnels.

Une grande partie de ces données est non structurée, gouvernée de manière incohérente et difficilement accessible aux systèmes d'IA sous une forme utilisable. Il manque le contexte. L'origine est incertaine. Les chemins d'accès varient selon l'environnement. D'après une récente enquête HPE, seulement 45 % des personnes interrogées déclarent pouvoir effectuer des chargements ou des extractions de données en temps réel.¹ Pour compenser, les équipes ont recours à des solutions manuelles : copie des données, préparation des fichiers et reconstitution des étapes préparatoires pour chaque cas d'utilisation.

Ces approches peuvent fonctionner pour les premières expérimentations, mais elles ne sont pas adaptables à grande échelle. À mesure que les initiatives d'IA se développent, les frictions s'accroissent. Une mauvaise préparation des données ralentit les cycles d'itération, entraîne une sous-utilisation de l'infrastructure et sape la confiance dans les résultats. Le schéma est bien connu : les pilotes réussissent isolément, puis stagnent lorsqu'il s'agit de déployer l'IA à l'échelle de l'entreprise.

Le point d'inflexion hybride : quand le calcul moderne expose les données héritées

Pour de nombreuses organisations, le défi de la préparation des données devient impossible à ignorer lors de la modernisation des infrastructures. À mesure que les charges de travail migrent des machines virtuelles vers les conteneurs et Kubernetes, les hypothèses établies de longue date concernant la manière dont les données sont consultées, gérées et déplacées commencent à s'effondrer.

Le calcul devient plus dynamique et éphémère, s'adaptant à la hausse et à la baisse selon les besoins. Cependant, les données restent souvent ancrées dans des architectures héritées et des environnements en silos. Il en résulte un décalage croissant entre le fonctionnement des charges de travail d'IA modernes et la manière dont les données d'entreprise leur sont fournies.

Cet écart révèle de nouvelles exigences auxquelles les approches traditionnelles en matière de données n'étaient pas conçues pour répondre. Les charges de travail d'IA nécessitent des services de données persistants pour le calcul de courte durée, un pipeline de données intelligent avec des métadonnées intégrées et une intelligence vectorielle, une mobilité transparente entre les environnements et une disponibilité constante à grande échelle. Lorsque ces capacités font défaut, les équipes se rabattent sur la copie et la préparation de données, ce qui ajoute de la latence, de la complexité et des risques, au moment même où les initiatives d'IA tentent d'accélérer leur développement.

Le calcul moderne ne crée pas de problèmes au niveau des données : il les révèle.



¹ "One year on—Architecting an AI advantage," HPE, 2025

Définition de la préparation des données pour l'IA

La préparation des données pour l'IA ne se résume pas à la quantité de données dont dispose une organisation ni à la vitesse à laquelle elle peut les stocker. Il s'agit de savoir si les données sont utilisables opérationnellement par les systèmes d'IA, de manière cohérente, à grande échelle et dans différents environnements.

Pour les charges de travail d'IA, les données doivent répondre à des exigences plus élevées que les analyses traditionnelles. Elles doivent être accessibles partout où les modèles sont entraînés et exécutés, et non pas enfermées derrière des barrières spécifiques à un environnement. Elles doivent être structurées et contextualisées afin que les machines (et pas seulement les humains) puissent les interpréter correctement. Elles doivent également être gouvernées en permanence, avec une traçabilité, des politiques et des contrôles qui persistent tout au long du cycle de vie des données plutôt que d'être appliqués a posteriori.

Ces attributs sont d'autant plus importants pour l'IA que celle-ci amplifie les incohérences. De petites lacunes en matière de contexte ou d'accès ne font pas que ralentir l'analyse : elles propagent les erreurs, réduisent la confiance dans les résultats et compromettent la reproductibilité à mesure que les systèmes évoluent. Lorsque les données ne sont pas prêtes, les équipes compensent par des corrections manuelles et des pipelines fragiles qui cèdent sous la pression de la production.

Le coût d'une mauvaise préparation des données ne se limite pas à la dette technique, mais entraîne également un manque de fiabilité au niveau de l'IA, des retards de déploiement et une perte de confiance dans les résultats.

Les services de données modernes doivent fonctionner dans des environnements hybrides

Les architectures de données traditionnelles n'ont pas été conçues pour l'IA. Elles s'appuient sur la copie, la préparation et le déplacement manuel des données entre les systèmes pour les rendre utilisables pour différentes charges de travail. Dans les environnements hybrides, ces pratiques deviennent rapidement un handicap, entraînant de la latence, des frais opérationnels et des incohérences au moment même où les initiatives d'IA tentent de se développer.

L'IA exige une approche différente. Les services de données modernes appliquent des principes cloud-native, tels que l'automatisation, les API et l'orchestration, mais de manière à fonctionner dans des environnements hybrides. L'objectif n'est pas de transférer toutes les données vers le cloud. Il s'agit de rendre les données utilisables pour l'IA, quel que soit leur emplacement actuel.

Lorsque les services de données sont cohérents d'un environnement à l'autre, les organisations peuvent privilégier la mobilité à la réplication, fournir un accès fiable sans reconstruire les pipelines, maintenir une gouvernance cohérente et aligner les opérations de données sur les pratiques modernes DevOps et MLOps.

Des expérimentations en matière d'IA à l'impact commercial

Lorsque les données sont préparées et disponibles par défaut, la trajectoire des initiatives d'IA change. Les équipes peuvent travailler plus rapidement car elles passent moins de temps à manipuler les données et plus de temps à entraîner, tester et affiner les modèles. Les cycles d'itération sont raccourcis. L'évolutivité devient plus prévisible. Les déploiements en production cessent d'être perçus comme des prouesses d'ingénierie ponctuelles.

En outre, et c'est tout aussi important, l'IA devient reproductible. Au lieu de reconstruire les pipelines de données pour chaque nouveau cas d'utilisation, les organisations établissent une infrastructure qui prend en charge plusieurs modèles, équipes et charges de travail. La confiance dans les résultats s'améliore à mesure que l'accès, le contexte et la gouvernance restent constants dans le temps.

Ce changement marque la différence entre l'IA en tant qu'expérimentation et l'IA en tant que capacité commerciale durable. Le succès à long terme ne s'obtient pas en poursuivant constamment de nouveaux modèles ou de nouvelles architectures. Il découle d'investissements dans la préparation des données qui permettent à l'IA de passer de succès isolés à un impact durable à l'échelle de l'entreprise.



Conclusion : les données sont la véritable devise

Chaque ruée vers l'or récompense ceux qui investissent dans les infrastructures adéquates. À l'ère de l'IA, cette infrastructure ne se limite pas aux modèles ou au calcul : elle englobe la préparation des données dans des environnements hybrides. Les organisations qui modernisent leurs infrastructures de données peuvent agir plus rapidement, déployer l'IA à plus grande échelle de manière plus fiable et transformer l'expérimentation en impact commercial durable. Celles qui résisteront au changement continueront à rencontrer des difficultés, peu importe le degré de perfectionnement de leurs modèles. À mesure que l'adoption de l'IA s'accélère, la différence entre ceux qui rencontreront le succès et les retardataires deviendra évidente. La véritable richesse de la ruée vers l'IA ne réside pas uniquement dans l'innovation, mais aussi dans la préparation des données prêtes à la soutenir.

Visiter [HPE.com](https://www.hpe.com)

Pour en savoir plus, rendez-vous sur :

[HPE.com/storage](https://www.hpe.com/storage)



[Live Chat](#)

© Copyright 2026 Hewlett Packard Enterprise Development LP. Les informations figurant dans ce document sont susceptibles d'être modifiées sans préavis. Les seules garanties relatives aux produits et services Hewlett Packard Enterprise sont stipulées dans les déclarations de garantie expresses accompagnant ces produits et services. Aucune partie du présent document ne saurait être interprétée comme offrant une garantie supplémentaire. Hewlett Packard Enterprise décline toute responsabilité quant aux éventuelles erreurs ou omissions techniques ou rédactionnelles qui pourraient être constatées dans le présent document.

a00155954FRE

HEWLETT PACKARD ENTERPRISE

[hpe.com](https://www.hpe.com)

