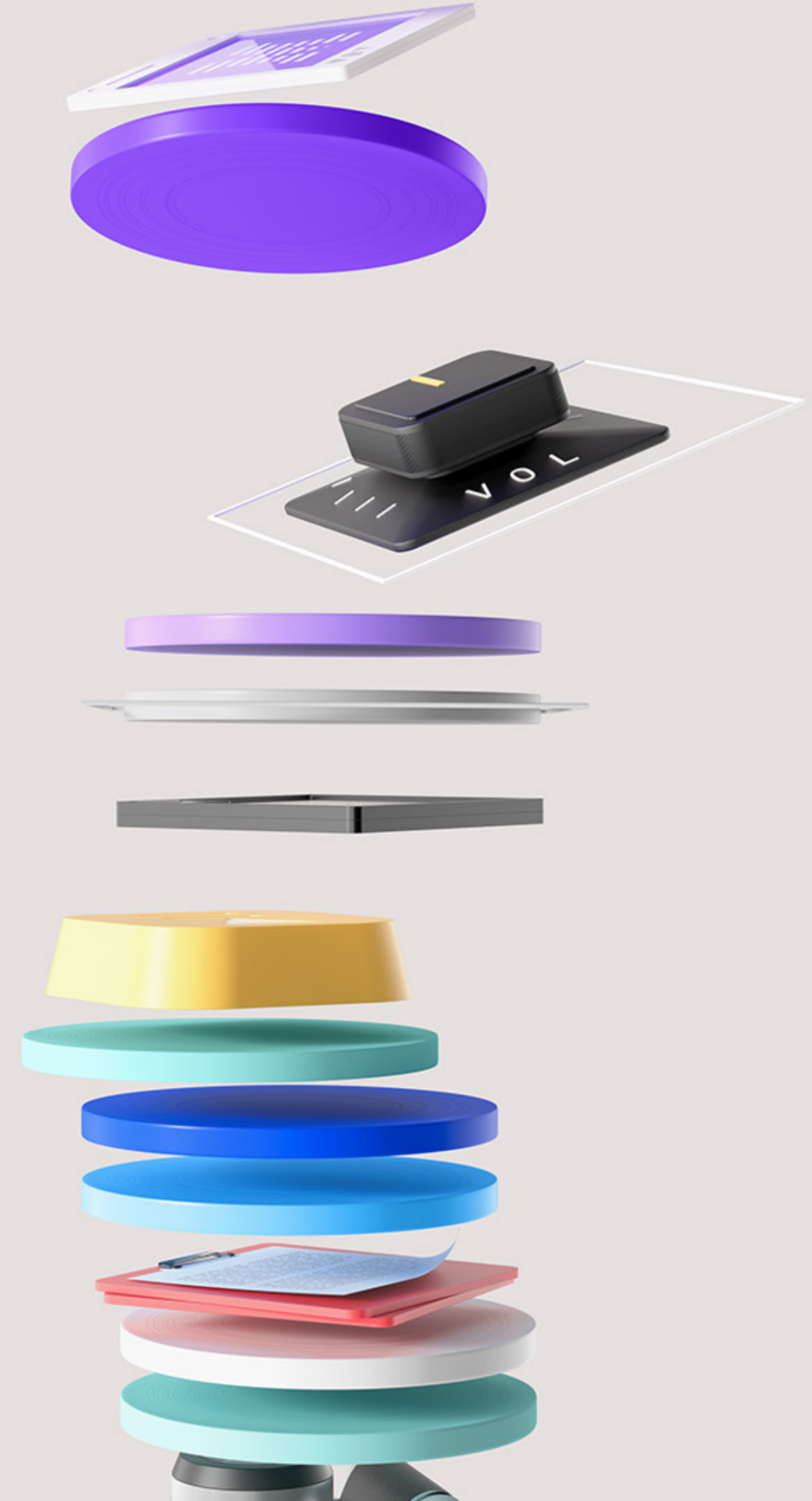


watsonx.data

Le seul data lakehouse  
hybride et ouvert  
pour l'IA d'entreprise  
et l'analytique



IBM

## Contenu

01 →

Pourquoi votre entreprise a besoin d'un data lakehouse prêt pour l'IA générative

02 →

Défis actuels et émergents en matière de données

03 →

Définition du data lakehouse

04 →

Découvrez watsonx.data

05 →

Pourquoi choisir watsonx.data

06 →

Étapes suivantes



# Pourquoi votre entreprise a besoin d'un data lakehouse prêt pour l'IA générative



Des données génériques produisent des résultats génériques. Si les données ne sont pas spécifiques à votre entreprise, les résultats ne le seront pas non plus. Lorsque tout le monde utilise les mêmes données publiques pour entraîner les mêmes modèles d'IA générative, comment pouvez-vous surpasser vos concurrents ?

Votre avantage concurrentiel réside dans vos données uniques. Les données non structurées, celles qui sont enfouies dans les e-mails, les fichiers PDF, les images et les vidéos, sont particulièrement précieuses, mais très difficiles à rendre opérationnelles. Les entreprises ont besoin d'une architecture de données intelligente pour accéder, préparer et fournir des données structurées et non structurées afin de déverrouiller le plein potentiel de leurs données d'entreprise pour l'IA.

Les données sont le carburant de l'IA, mais on estime que moins de 1 % des données d'entreprise sont exploitées par des modèles d'IA.<sup>1</sup> Et 90 % des données d'entreprise sont non structurées.<sup>2</sup>

Un data lakehouse hybride et ouvert peut vous aider à booster votre IA et à intégrer vos données dans une grande variété de cas d'utilisation modernes. Il unifie tous les types de données, simplifie l'intégration et rend la gestion et la gouvernance des données, structurées et non structurées, plus faciles que jamais. Il n'est donc pas étonnant qu'il s'impose comme l'architecture de données à long terme privilégiée pour l'analytique et les workloads de la nouvelle IA générative.



# 90%

des données d'entreprise  
sont considérés comme  
non structurées.<sup>2</sup>





## Défis actuels et émergents en matière de données

L'importance des données d'entreprise s'est accrue, tout comme les défis qui y sont associés. Le volume de données est devenu écrasant et il est souvent cloisonné au sein de l'entreprise. En outre, de nouvelles variétés de données ont compliqué l'intégration, et des données de mauvaise qualité ont compromis l'impact sur l'IA. L'IA générative peut contribuer à résoudre ces problèmes, mais une architecture de données résiliente et flexible est essentielle à cet effet.

Aujourd'hui, les limites de la génération augmentée de récupération (RAG) empêchent les entreprises de réaliser la valeur des données non structurées pour l'IA générative. Alors, comment intégrer les données non structurées dans l'IA générative et les workloads analytiques traditionnels ? En adoptant un data lakehouse hybride et ouvert.





## Définition du data lakehouse



Les données d'entreprise sont au cœur de toute architecture d'IA. Les entreprises qui souhaitent se doter d'outils d'IA générative en 2025 devront s'assurer que leurs données sont prêtes pour l'IA. Une solution lakehouse qui fournit une base moderne pour déployer les données prêtes pour l'IA.

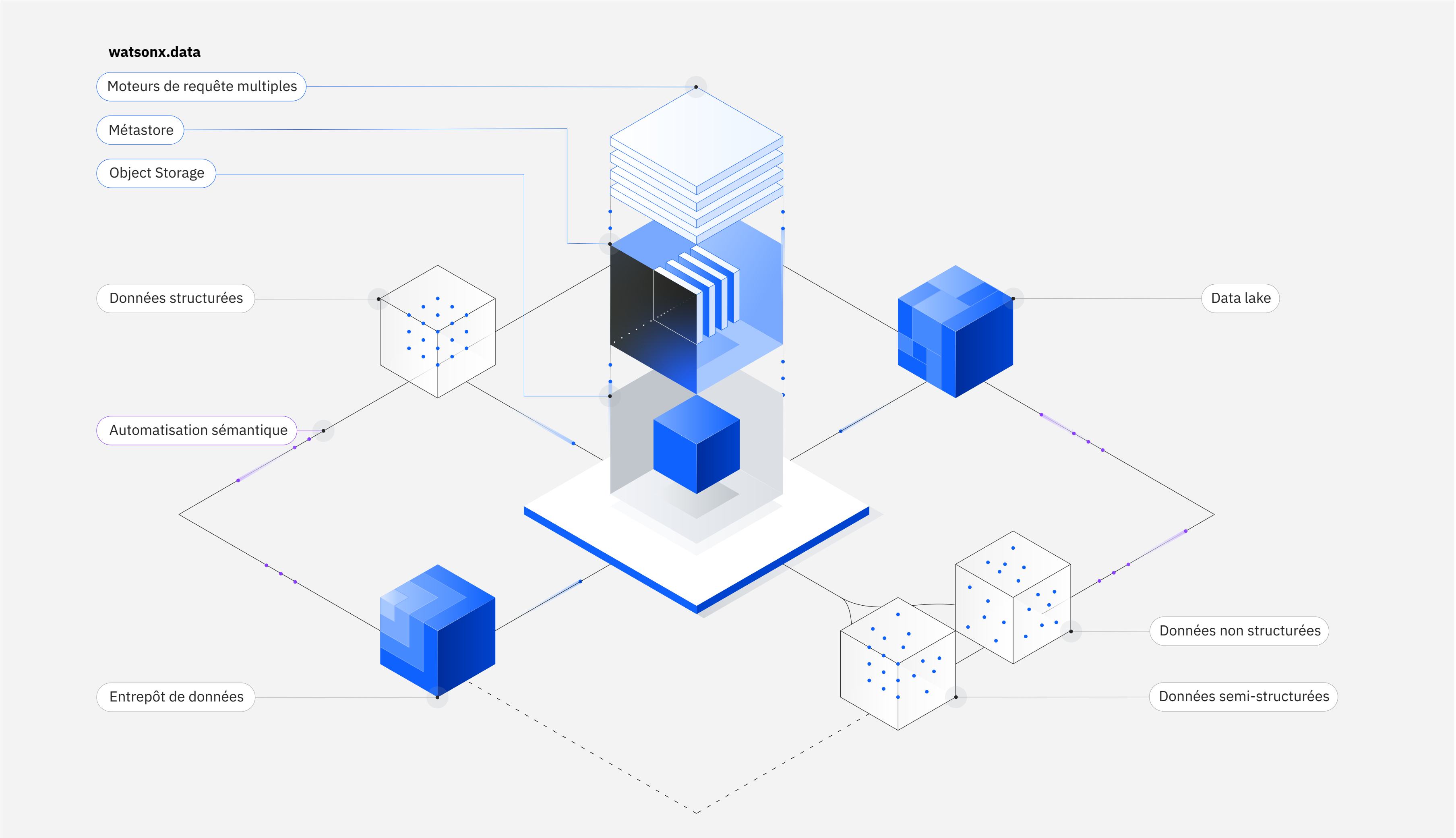
Le data lakehouse est une architecture émergente qui offre la flexibilité d'un data lake avec les performances et la structure d'un entrepôt de données. La plupart des solutions lakehouse proposent un moteur de requête hautes performances sur un stockage à faible coût, associé à une couche de gouvernance des métadonnées. Les couches de métadonnées intelligentes facilitent la catégorisation et la classification des données non structurées, telles que la vidéo et la voix, ainsi que des données semi-structurées, telles que XML, JSON et les e-mails.

Le meilleur data lakehouse offrira des technologies open source qui réduisent la duplication des données et simplifient les pipelines complexes. Cependant, certains centres de données de première génération sont soumis à des contraintes d'extraction, de transformation et de chargement (ETL) qui limitent leur capacité à relever les défis du coût et de la complexité. Par exemple, un moteur de requête unique conçu pour les workloads de business intelligence (BI) ou de machine learning (ML) peut être inefficace lorsqu'il est utilisé pour un autre type de workload.

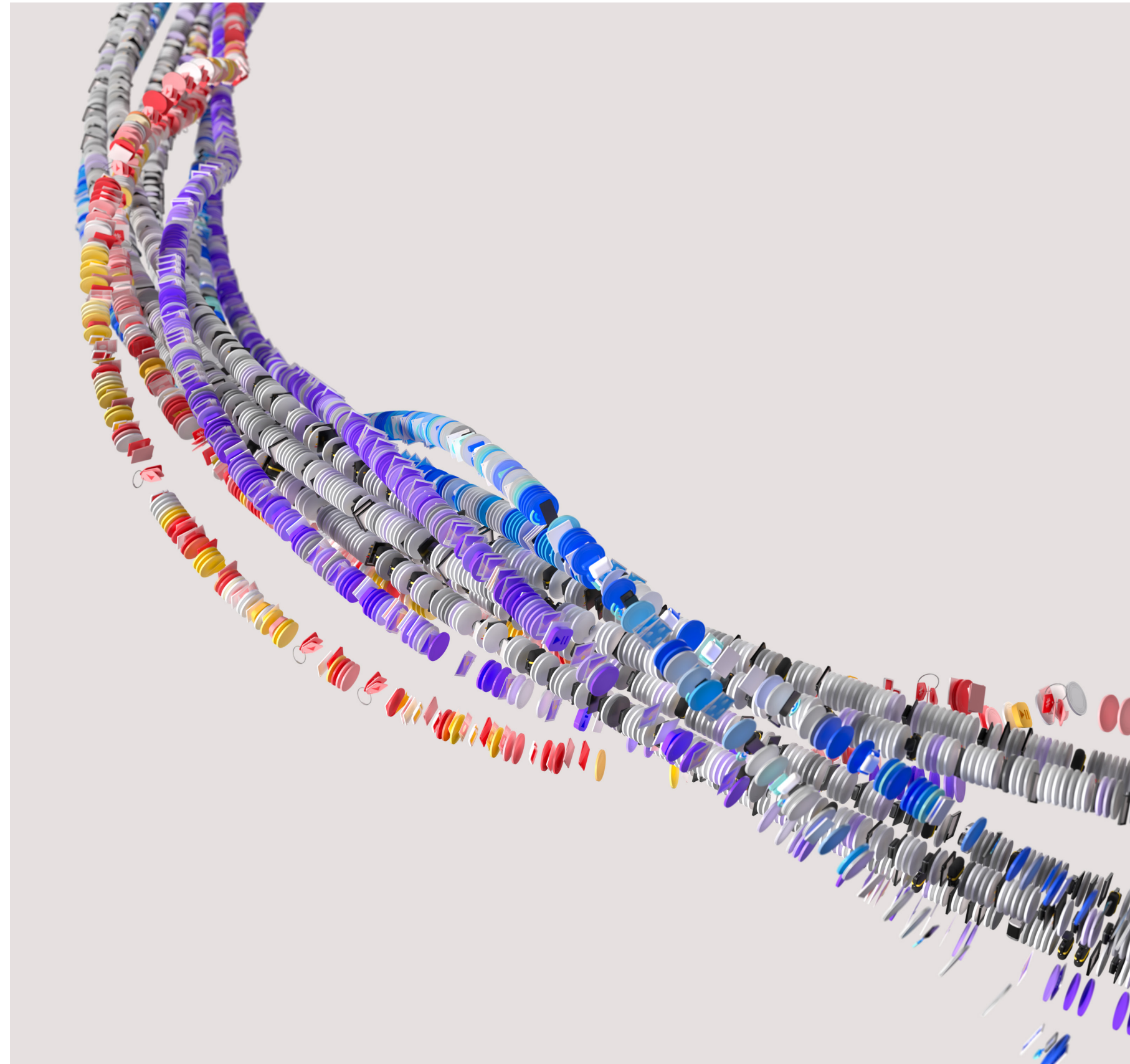
Les équipes données et IA d'IBM estiment que chaque workload est unique et doit être optimisé avec l'environnement le mieux adapté, en maintenant les coûts au minimum et les performances au maximum. En fin de compte, transformer des données brutes en informations exploitables à l'échelle nécessite un centre d'hébergement qui offre une performance d'un bon rapport coût-efficacité et une grande souplesse de format.



Gérer l'ensemble du cycle de vie des données pour l'IA



## Découvrez watsonx.data



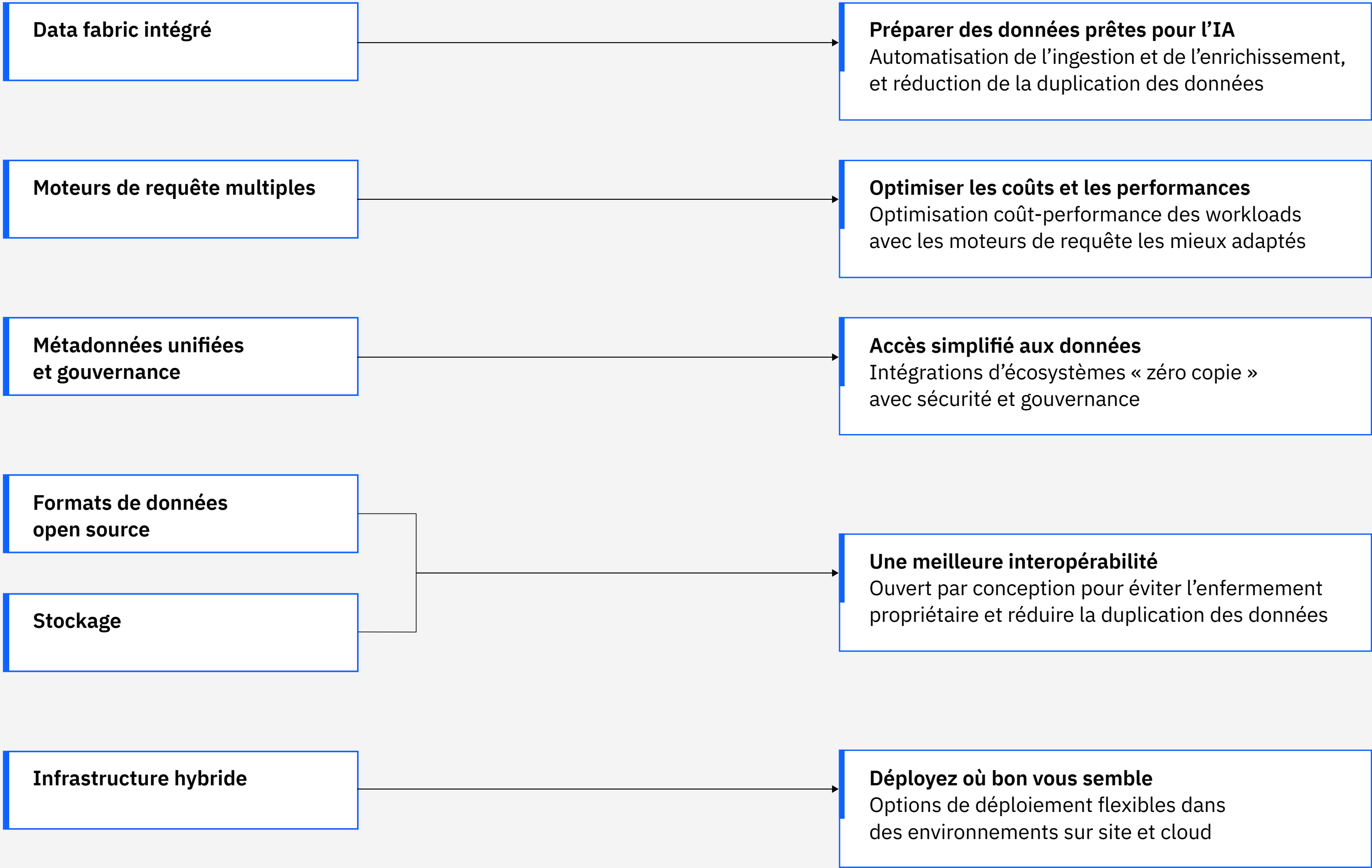
IBM watsonx.data est le seul data lakehouse hybride et ouvert pour l'IA d'entreprise et l'analytique.

Il vous aide à gérer l'ensemble du cycle de vie des données d'entreprise pour l'IA au sein de votre data lakehouse, pour alimenter la prochaine génération d'applications d'IA générative et de BI intégrant l'IA.

Grâce à watsonx.data, vous pouvez simplifier et adapter l'intégration, la gestion et la gouvernance des données structurées et non structurées dans les environnements sur site, cloud et multicloud. Il fait partie d'une pile de données ouverte et moderne utilisant des technologies open source innovantes qui s'intègrent à votre patrimoine de données existant, sans enfermement propriétaire.



IBM watsonx.data intègre de façon fluide les capacités d'un data lakehouse et d'une data fabric.



Quels sont les composants de watsonx.data ?

Une architecture hybride et ouverte de type « data lakehouse » offre une plateforme unifiée pour relever les principaux défis auxquels les entreprises sont confrontées aujourd’hui en matière de données : volumes importants, emplacements et formats multiples, qualité et gouvernance des données.

Voici comment IBM watsonx.data relève ces défis.

Infrastructure

L'infrastructure est l'endroit où votre lakehouse sera déployé, entièrement géré sur n'importe quel cloud ou environnement sur site.

Stockage

La couche de stockage est l’endroit où les données sont physiquement stockées sous forme de fichiers ou dans des formats de données ouverts tels que Apache Parquet et Avro. Les formats de données ouverts sont des spécifications de fichiers et des protocoles mis à la disposition de la communauté open source afin que chacun puisse les ingérer et les améliorer.

Des moteurs de recherche adaptés au besoin

Plusieurs moteurs de requête adaptés à un usage spécifique permettent d’exécuter efficacement les requêtes sur de grands jeux de données. Grâce à l’optimisation des workloads entre plusieurs moteurs de requête et niveaux de stockage, vous pouvez minimiser le coût de votre entrepôt de données tout en offrant un traitement Big Data rapide, fiable et efficace à l’échelle.

Formats de table ouverts

Les formats de table ouverts tels qu’Apache Iceberg contribuent à structurer les données et offrent la fiabilité et la simplicité du langage SQL avec le big data. Ces formats permettent à différents moteurs d’accéder aux mêmes données, en même temps, ce qui permet d’éviter l’enfermement propriétaire. Les données peuvent également être partagées entre plusieurs outils et référentiels de données tels que votre entrepôt de données. Une copie unique des données vous permet de réduire la duplication des données et de supprimer les silos.

Stockage de métadonnées

Le magasin de métadonnées dans watsonx.data est basé sur le métastore open source Hive, qui garantit l’interopérabilité et permet à des outils externes d’accéder facilement à vos métadonnées.



**Gouvernance**

Les métadonnées sont également stockées dans des formats de table ouverts. Elles servent à définir les formats de fichiers pour tout outil capable de lire ou d’écrire des formats de données ouverts.

**Service de métadonnées techniques**

Le composant de service de métadonnées techniques est nécessaire pour comprendre quelles données sont disponibles dans la couche de stockage. Le moteur de requête utilise des métadonnées pour fournir des informations complètes sur la provenance des données, afin que vous sachiez où elles se trouvent, à quoi elles ressemblent et comment les lire.

**Catalogue de données**

Les catalogues de données aident les utilisateurs à trouver les bonnes données pour leur travail et fournissent des informations sémantiques pour les politiques et les règles. Les métadonnées d’entreprise telles que les terminologies d’entreprise et les balises sont stockées pour soutenir la recherche et la protection des données.

**Moteur de politiques**

Ce moteur de politiques permet aux utilisateurs de définir des politiques de protection des données et permet au moteur d’appliquer ces politiques. Pour créer un cadre de gouvernance évolutif, un moteur de politiques est souvent déployé avec le service de métadonnées techniques et le catalogue de données.

**Moteurs de requête**

Le moteur de requête, souvent appelé composant Compute, est au cœur du data lakehouse ouvert. Il peut être open source ou propriétaire, et il accède aux données dans des formats de table ouverts.

Dans une architecture de data lakehouse ouverte, le moteur de requête est entièrement modulaire, ce qui signifie qu’il peut être mis à l’échelle dynamiquement pour répondre aux demandes de workload et à la concurrence. Les moteurs de requête peuvent également être connectés à n’importe quel catalogue et système de stockage.

Qu'est-ce qui différencie  
IBM watsonx.data des autres ?

**Hybride et ouvert**

watsonx.data est conçu à partir de technologies open source fondamentales pour favoriser l'innovation et le soutien communautaires, sans enfermement propriétaire. Vous pouvez accéder aux données et les déployer n'importe où dans des environnements de cloud hybride, dans des VPC clients, dans des environnements multicloud et sur site, et en tant que SaaS.

**Optimisé pour les workloads**

Soutenu par les capacités intégrées de data fabric, y compris la gouvernance des données, la qualité et la traçabilité, watsonx.data vous aide à gérer des volumes de données croissants grâce à une architecture multi-moteur unique optimisée pour le rapport prix-performance.

**Prêt pour l'IA générative**

watsonx.data permet une autorisation cohérente de l'utilisateur de bout en bout, depuis les systèmes sources de documents jusqu'à leur extraction.

**Comment tout cela s'articule-t-il ?**

watsonx.data permet aux entreprises de déverrouiller les informations critiques de leurs données d'entreprise non structurées pour une IA générative et une analyse plus précises et plus puissantes.

40%

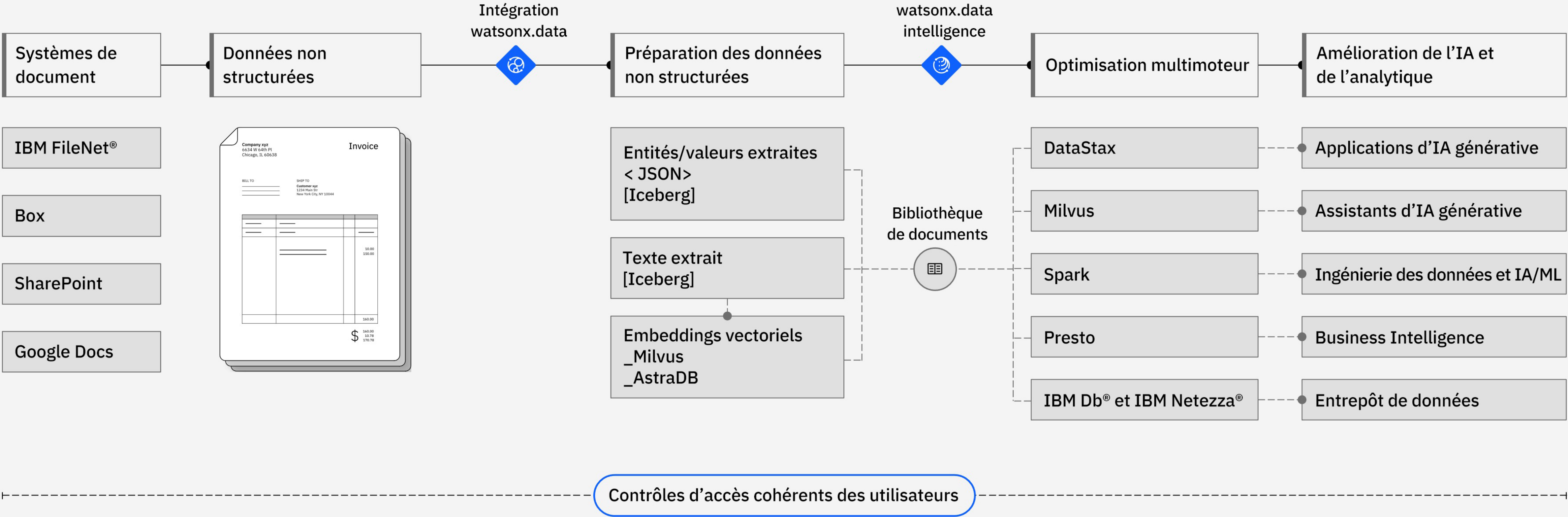
plus précise que la  
RAG conventionnelle  
avec watsonx.data.<sup>3</sup>








Ingérer à l'échelle

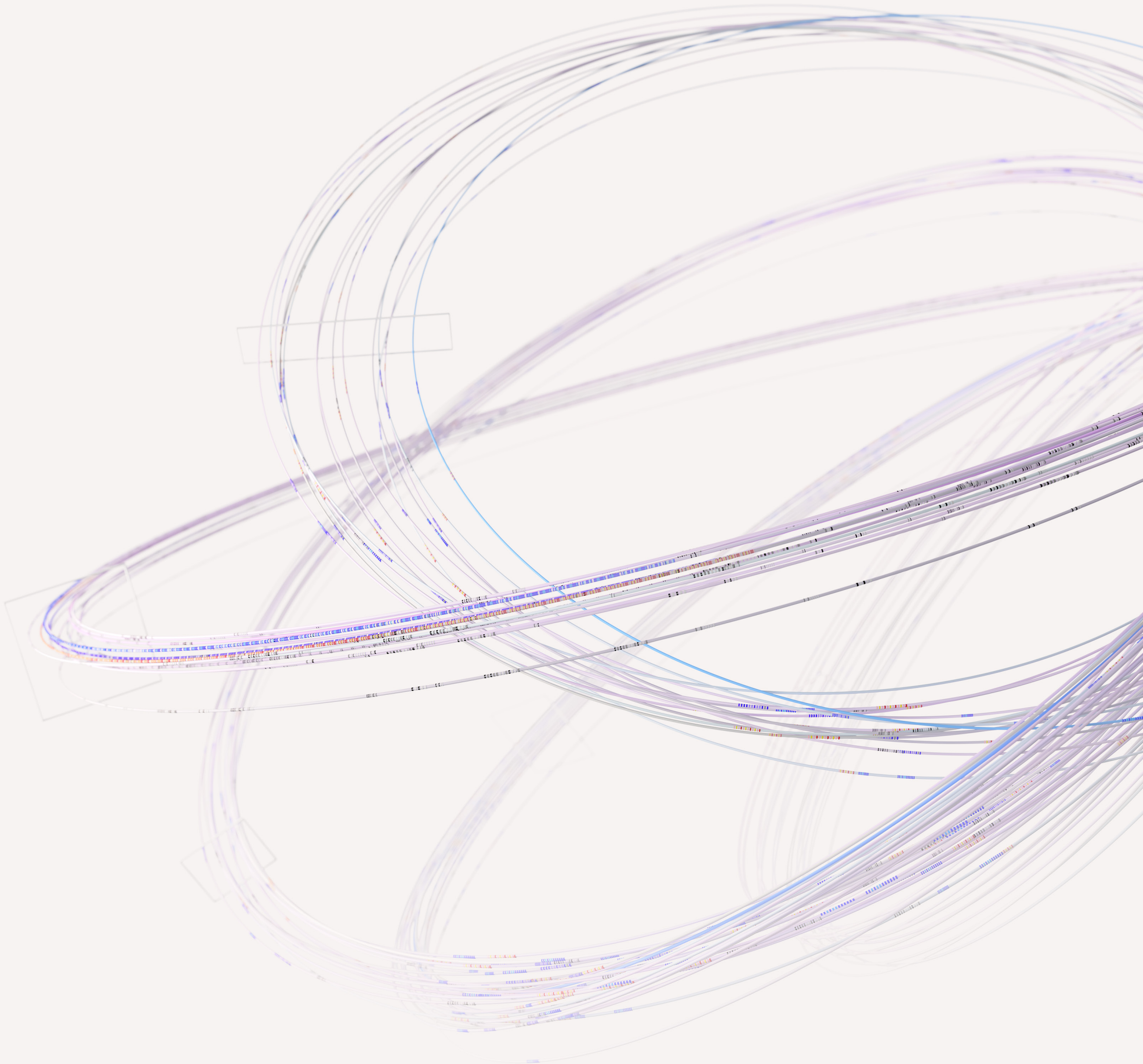
Organiser et gérer

Transmettre à l'IA



# Pourquoi choisir watsonx.data

-  Pour unifier les données structurées et non structurées pour améliorer la précision de l'IA et du ROI
-  Pour déverrouiller les données inexploitées pour l'analyse conversationnelle
-  Pour simplifier et adapter l'intégration, la gestion et la gouvernance des données structurées et non structurées
-  Pour permettre à tous les utilisateurs de données de disposer de données de haute qualité et fiables
-  Pour protéger vos investissements dans les données existantes grâce à une pile de données ouverte et moderne





## Étapes suivantes

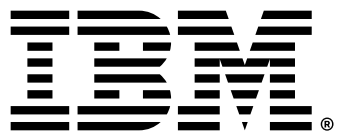
Profitez de l'expertise d'IBM en matière de gestion et d'optimisation des données, acquise grâce à des décennies de gestion des workloads de données les plus exigeantes au monde. Découvrez à quelle vitesse vous pouvez maximiser le ROI de vos données avec watsonx.data.

Commencez votre essai gratuit →

Demander une démo →







- 1. [L’avenir de l’IA est ouvert, IBM, 2024](#)
- 2. [IDC, UNTAPPED VALUE: What Every Executive Needs to Know About Unstructured Data, 2023](#)
- 3. Reposant sur des tests internes comparant l’exactitude des réponses des sorties des modèles d’IA à l’aide de la couche de récupération watsonx.data Premium ; RAG vectorielle disponible uniquement sur trois cas d’utilisation courants avec des jeux de données IBM, à l’aide du même ensemble de modèles d’inférence open source, d’évaluation et d’embedding et variables supplémentaires. Les résultats peuvent varier.

© Copyright IBM Corporation 2025

IBM, le logo IBM, Db2, FileNet, Netezza et watsonx.data sont des marques d’International Business Machines Corporation déposées aux États-Unis et/ou dans d’autres pays. Les autres noms de produits et de services peuvent être des marques d’IBM ou d’autres sociétés. La liste actualisée de toutes les marques d’IBM est disponible à l’adresse [ibm.com/fr-fr/legal/copytrade](https://ibm.com/fr-fr/legal/copytrade).

Les informations contenues dans le présent document étaient à jour à la date de publication initiale et sont susceptibles d’être modifiées à tout moment par IBM.

Certaines offres mentionnées dans le présent document ne sont pas disponibles dans tous les pays où la société IBM est présente.

Il incombe à l’utilisateur d’évaluer et de vérifier le fonctionnement de tout autre produit ou programme avec les produits et programmes IBM. IBM n’est pas responsable des produits et programmes non IBM.

LES INFORMATIONS CONTENUES DANS LE PRÉSENT DOCUMENT SONT FOURNIES « EN L’ÉTAT », SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE, NOTAMMENT SANS AUCUNE GARANTIE DE QUALITÉ MARCHANDE, D’ADÉQUATION À UN USAGE PARTICULIER ET TOUTE GARANTIE OU CONDITION D’ABSENCE DE CONTREFAÇON. Les produits IBM sont garantis conformément aux dispositions des contrats qui régissent leur utilisation.

Aucun système ou produit informatique ne doit être considéré comme entièrement sécurisé, et aucun produit ou mesure de sécurité ne peut être totalement efficace pour empêcher les accès ou l’utilisation non autorisés. IBM ne garantit pas qu’un système, produit ou service, quel qu’il soit, est à l’abri, ou mettra votre entreprise à l’abri, de la conduite malveillante ou illégale de quelque partie que ce soit.

Le client est tenu de respecter l’ensemble des lois et des réglementations applicables. IBM ne fournit pas de conseils juridiques et ne déclare ni ne garantit que ses services ou produits garantiront que le client est en conformité avec la législation ou la réglementation en vigueur.