

L'architecture data fabric offre 3 avantages immédiats

Intégrez et interagissez avec l'ensemble des données de votre organisation pour obtenir de meilleurs résultats métier

Sommaire

- 3** Les défis actuels des données
- 5** Les avantages métier d'une data fabric
- 6** Comment IBM fournit une data fabric
- 7** La data fabric en action
- 7** La data fabric est la prochaine étape de la transformation de votre entreprise.

Les défis actuels des données

Les données font partie intégrante de la transformation numérique des entreprises.

Cependant, lorsque les organisations cherchent à tirer parti de leurs données, elles se heurtent à des difficultés résultant de la diversité des sources, des types, des structures, des environnements et des plateformes de données. Ce problème des données multidimensionnelles se complique encore lorsque les organisations adoptent des architectures hybrides et multiclouds. Pour de nombreuses entreprises, les données opérationnelles restent à ce jour largement cloisonnées et cachées, ce qui conduit à une énorme quantité de données sombres.

Prenons l'exemple d'une entreprise nord-américaine du secteur de l'énergie qui a cherché à se réinventer en appuyant son organisation sur les données - pour devenir une entreprise où les capacités de science des données étaient facilement accessibles dans plusieurs lignes métier. Ils ont rapidement réalisé que leur transformation numérique était entravée par le cloisonnement des données, l'incohérence des outils et l'hétérogénéité des compétences, qui provoquaient des écarts importants en matière de données.

Le problème auquel ils étaient confrontés n'était pas propre à leur entreprise. En réalité, il s'agit d'une conséquence courante des paysages de données qui sont sortis du cadre de leur architecture de gestion des données.

C'est finalement l'utilisation d'un nouveau concept d'architecture des données appelé data fabric (fabrique de données) qui a remis l'entreprise d'énergie sur la voie de la transformation numérique.

En quoi consiste exactement une data fabric ? En quoi diffère-t-elle des architectures antérieures ? Que permet-elle pour les entreprises ? Quel est le rôle d'IBM dans sa mise en œuvre ? Ce livre blanc répondra à ces questions.

Qu'est-ce qu'une data fabric ?

Par le passé, les organisations ont déjà tenté de résoudre les problèmes d'accès aux données, soit par une intégration point-à-point, soit par l'introduction de hubs de données. Aucune de ces solutions ne convient lorsque les données sont fortement distribuées et cloisonnées. Les intégrations point-à-point ajoutent un coût exponentiel pour chaque nœud final supplémentaire qui doit être raccordé. Autrement dit, cette approche n'est pas évolutive. Les hubs de données facilitent l'intégration des applications et des sources, mais exacerbent le coût et la complexité du maintien de la qualité et de la confiance dans les données au sein du hub.

La data fabric est une architecture émergente qui vise à relever les défis découlant d'un paysage de données hybride. Son concept fondamental consiste à trouver un équilibre entre décentralisation et globalisation en agissant comme un tissu conjonctif virtuel entre les nœuds finals des données (voir page 4).

Grâce aux technologies telles que l'automatisation et l'augmentation de l'intégration, la gouvernance fédérée ainsi que l'activation des métadonnées, une architecture data fabric permet une orchestration dynamique et intelligente des données dans un paysage distribué, créant un réseau d'informations instantanément disponibles pour alimenter l'entreprise.

La data fabric est indifférente aux plateformes de déploiement, aux processus de données, aux emplacements géographiques et à l'approche architecturale. Elle facilite l'utilisation des données en tant que ressources de l'entreprise. La data fabric garantit que vos différents types de données puissent être combinés, consultés et gérés de manière efficace et performante.

Fonctionnalités et principes de data fabric

Au cœur de l'architecture de la data fabric se trouve une plateforme de gestion des données qui permet d'exploiter l'ensemble des capacités de gestion intégrée des données, notamment la découverte, la gouvernance, l'organisation et l'orchestration.

Cependant, elle évolue et s'extrait des concepts traditionnels de gestion des données tels que DataOps, qui se concentrent uniquement sur l'établissement de pratiques visant à augmenter le niveau d'opérationnalisation des données. Elle est construite sur une architecture distribuée et une technologie avancée capable de répondre aux besoins qui découlent de l'extrême diversité et de la distribution des données.

Une data fabric peut être divisée logiquement en quatre fonctionnalités (ou composants) :

Connaissances, informations et sémantique

- Fournit une place de marché de données et une expérience d'achat
- Enrichit automatiquement les ressources de données découvertes avec des connaissances et de la sémantique, pour permettre aux consommateurs de trouver et de comprendre les données.

Gouvernance et conformité unifiées

- Permet de gérer et de gouverner localement les métadonnées, mais prend en charge une vue et une application des politiques unifiées.
- Applique automatiquement des politiques sur les ressources de données conformément aux règles mondiales et locales.
- Utilise des capacités avancées pour automatiser la classification et l'organisation des données.
- Établit automatiquement des voies d'accès interrogeables pour toutes les ressources inventoriées afin d'accroître l'activation des données.

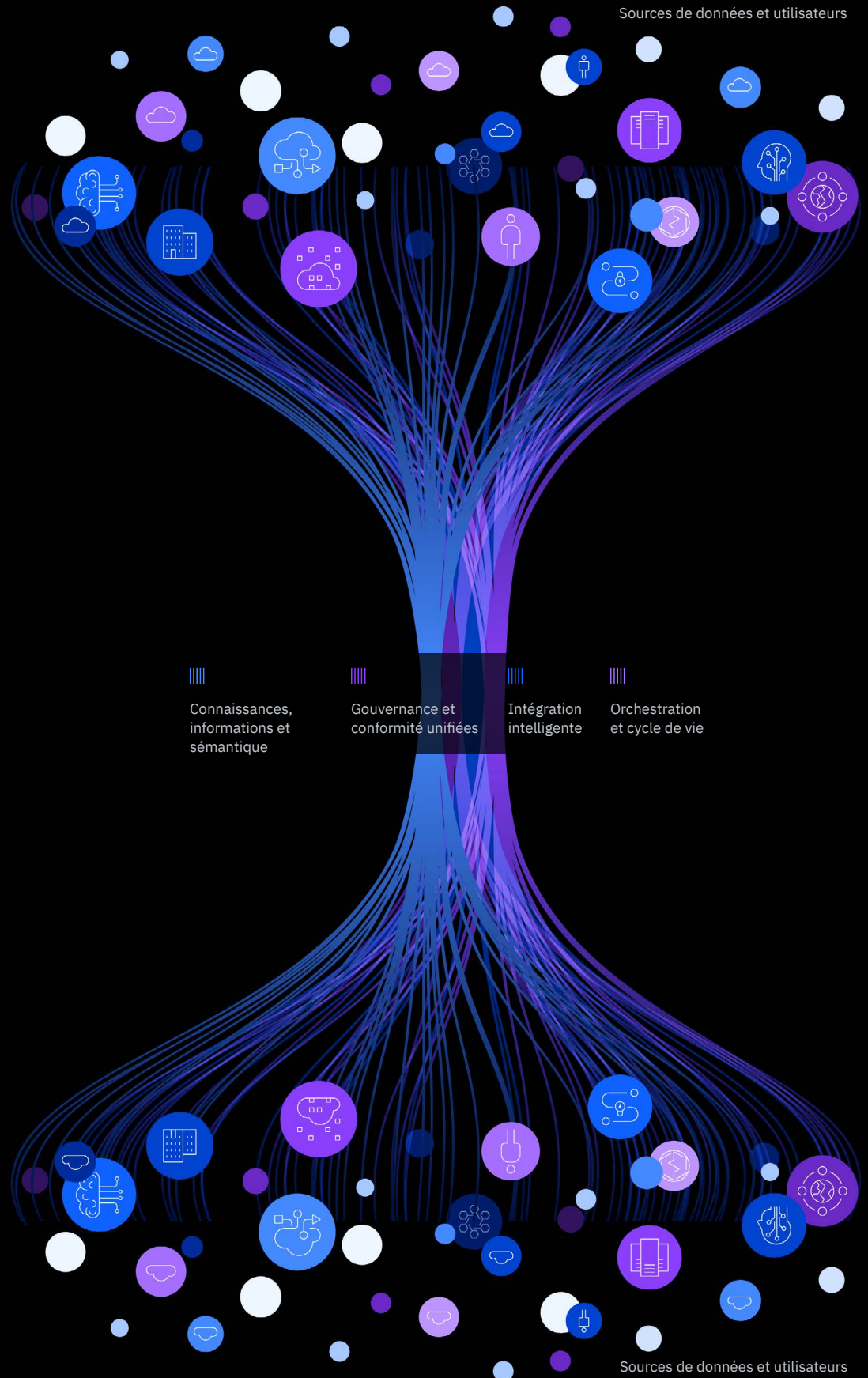
Intégration intelligente

- Accélère les tâches de l'ingénieur de données via un flux automatisé et la création de pipelines à travers les sources de données distribuées.
- Permet l'ingestion et l'accès à n'importe quelles données en libre-service avec une application approfondie des politiques de protection des données au niveau local et global.
- Détermine automatiquement l'exécution la mieux adaptée grâce à une répartition optimisée de la charge de travail, à l'auto-réglage et à la correction des dérives du schéma.

Orchestration et cycle de vie

- Permet de composer, de tester, d'exploiter et de surveiller les pipelines de données.
- Intègre des capacités d'intelligence artificielle dans le cycle de vie des données afin d'automatiser les tâches, de s'auto-régler, de s'auto-réparer et de détecter les modifications des données sources, ce qui facilite les mises à jour automatisées.

Sources de données et utilisateurs



Sources de données et utilisateurs

Les avantages métier d'une data fabric

Les données n'ont de valeur métier qu'à condition d'être contextualisées et accessibles à tout utilisateur ou application de l'organisation. Lorsqu'elle est mise en œuvre correctement, une data fabric permet de garantir que ces valeurs soient disponibles dans l'ensemble de l'organisation de la manière la plus efficace et la plus automatisée possible. En tant que telle, la data fabric présente trois avantages essentiels :

1. Permettre la consommation des données et la collaboration en libre-service
2. Automatiser la gouvernance, la protection et la sécurité, grâce à des métadonnées actives.
3. Automatiser les tâches d'ingénierie des données et améliorer l'intégration des données des ressources du cloud hybride.

Permettre la consommation des données et la collaboration en libre-service

En intégrant des données provenant de sources multiples et en analysant une plus grande partie de l'énorme quantité de données générées quotidiennement, les organisations obtiennent de meilleures informations et répondent plus rapidement aux demandes changeantes des entreprises. La data fabric met rapidement les données à la disposition de ceux qui en ont besoin. Le libre-service permet à l'ensemble de l'organisation de trouver plus rapidement les données appropriées et de consacrer davantage de temps à l'utilisation de ces données pour fournir des informations tangibles.

Avantages de la data fabric pour la consommation de données en libre-service :

- Les utilisateurs professionnels disposent d'un point d'accès unique pour trouver, comprendre, façonner et consommer les données dans l'ensemble de l'organisation.
- La gouvernance et le lignage centralisés des données aident les utilisateurs à comprendre ce que signifient les données, d'où elles viennent et comment elles sont liées à d'autres ressources.
- La gestion étendue et personnalisable des métadonnées évolue facilement et est accessible via des API.
- L'accès en libre-service à des données fiables et contrôlées permet une collaboration entre les lignes métiers et les autres utilisateurs.

Une étude de Forrester sur l'impact économique total¹ a révélé ce que ces fonctionnalités pouvaient apporter :

USD

5,8 Mio 459 %
de gains de retour sur investissement

Automatiser la gouvernance, la protection et la sécurité des données, grâce à des métadonnées actives.

Une couche de gouvernance active distribuée pour toutes les initiatives en matière de données réduit les risques de conformité et de réglementation en assurant la confiance et la transparence. Elle permet l'application automatique de politiques pour tous les accès aux données, garantissant ainsi un niveau élevé de protection des données et de conformité.

L'utilisation de l'IA et des technologies d'apprentissage automatique permet aux utilisateurs d'une data fabric d'augmenter leur niveau d'automatisation, par exemple en extrayant automatiquement les règles de gouvernance des données en fonction du langage et des définitions dans les documents réglementaires. Les organisations peuvent ainsi appliquer en quelques minutes des règles de gouvernance propres à leur secteur d'activité, afin d'éviter des amendes coûteuses et de garantir une utilisation éthique des données, où qu'elles se trouvent.

Avantages de la data fabric pour la virtualisation gouvernée :

- L'agilité, la sécurité et la productivité sont accrues pour les ingénieurs de données, les scientifiques de données et les analystes métier.
- Plusieurs sources de données globales apparaissent comme une seule base de données.
- La découverte d'informations personnelles identifiables (IIP) et d'éléments de données critiques est possible à grande échelle.

Ces fonctionnalités peuvent générer de :

USD

2,4 Mio 430 %
de gains¹ d'amélioration des performances²

Automatiser les tâches d'ingénierie des données et renforcer l'intégration des données

L'ingénierie avancée des données signifie que pratiquement tous les processus d'accès ou de livraison des données sont automatisés et ne nécessitent aucun processus de codage fastidieux ou sujet aux erreurs. L'amélioration de l'intégration utilise les métadonnées pour optimiser la livraison et l'accès aux données.

Avantages de la data fabric pour l'ingénierie et l'intégration des données :

- L'automatisation automatique de l'intégration des données permet d'accélérer la livraison des données.
- Grâce à l'équilibrage automatique de la charge de travail et à la mise à l'échelle élastique, les tâches sont prêtes quels que soient l'environnement et le volume de données.
- La résilience et l'automatisation CI/CD sont intégrées.
- Le processus automatisé de saisie des changements en temps réel prend en charge la fourniture de données de qualité pour les processus opérationnels.
- L'apprentissage automatique permet d'automatiser et d'étendre les processus personnalisés de découverte, de classification et d'organisation des données, ce qui permet d'accélérer le retour sur investissement.
- L'analyse continue peut être effectuée automatiquement et en temps réel, quel que soit l'endroit où se trouvent les données.

Les résultats d'une grande enseigne³ montrent :

Accélération de
x60
du délai de livraison des
données

Accélération de
x20
des analyses des préférences
des clients

Comment IBM fournit une data fabric

IBM Cloud Pak® for Data rend possible ce concept de data fabric. IBM Cloud Pak for Data est une plateforme de connaissances qui simplifie et automatise la collecte, l'organisation et l'analyse des données et accélère le déploiement de l'IA dans votre entreprise.

Grâce à ses fonctionnalités permettant de relier les données où qu'elles se trouvent, d'exécuter des charges de travail n'importe où et de créer, déployer et gérer l'IA à grande échelle dans des environnements de cloud hybride, IBM Cloud Pak for Data est le catalyseur de la transformation numérique des entreprises.

Cette plateforme permet une intégration transparente dans une entreprise hybride pour:

- Les services IBM Cloud Pak for Data
- Les applications et sources de données externes
- Les capacités avancées basées sur l'IA pour la gestion et la gouvernance des données

Ce socle met les données conservées à la disposition des consommateurs avec un équilibre optimal entre coût, performance et conformité, et avec l'intelligence nécessaire pour orchestrer et optimiser le traitement des données en fonction des charges de travail, de la localisation des données et des politiques en la matière.

IBM Cloud Pak for Data exploite les fonctionnalités suivantes pour fournir les données prêtes à l'emploi qu'exige une data fabric. Toutes ces fonctionnalités jouent un rôle dans la prise en charge de l'architecture de la data fabric.

1. Noyau de connaissances basé sur des métadonnées

Facilite la découverte de sources de données, inventorie et enrichit les ressources de données et effectue des analyses pour extraire des informations en vue d'une plus grande automatisation grâce à l'IA. Le noyau de connaissances est utilisé pour alimenter la place de marché grâce à la recherche sémantique.

2. Place de marché de données en libre-service

Le catalogue de données de nouvelle génération qui aide les consommateurs de données, par exemple les analystes commerciaux, à récupérer des données dans l'ensemble du paysage de données de l'entreprise.

3. Intégration intelligente

Permet de consommer des données en les extrayant, en les virtualisant, en les transformant et en les diffusant en continu. Elle est intégrée au noyau de connaissances pour automatiser l'intégration des données, et possède l'intelligence nécessaire pour décider de l'approche d'intégration la mieux adaptée en fonction des charges de travail et des politiques de données. Elle peut également être utilisée pour la préparation des données dans le cadre de charges de travail d'ingénierie des données ou pour créer des produits de données. Enfin, elle offre la possibilité de publier les mises à jour des produits de données.

4. Gouvernance

Inventorie et conserve les métadonnées, définit les politiques de confidentialité des données, conserve les données, enregistre le lignage des données et effectue d'autres tâches liées à la sécurité et à la conformité. Cette couche comprend le format des données (par exemple, structurées vs non structurées) et leur importance (par exemple, publiques vs protégées). Elle applique les politiques appropriées à chaque élément de données et à chaque utilisateur potentiel. Au lieu d'appliquer manuellement des normes et des règles aux données, cette fonctionnalité de plateforme intégrée permet de les appliquer au niveau organisationnel et de les propager dans les diverses ressources du site en fonction des besoins. Les modèles analytiques de différents outils peuvent communiquer entre eux ; l'application des politiques de données au niveau granulaire peut être largement automatisée.

5. Développement et opérations unifiés

Permet la mise en place d'un cycle de vie unifié pour configurer et exécuter tous les aspects de la plateforme de données en production.

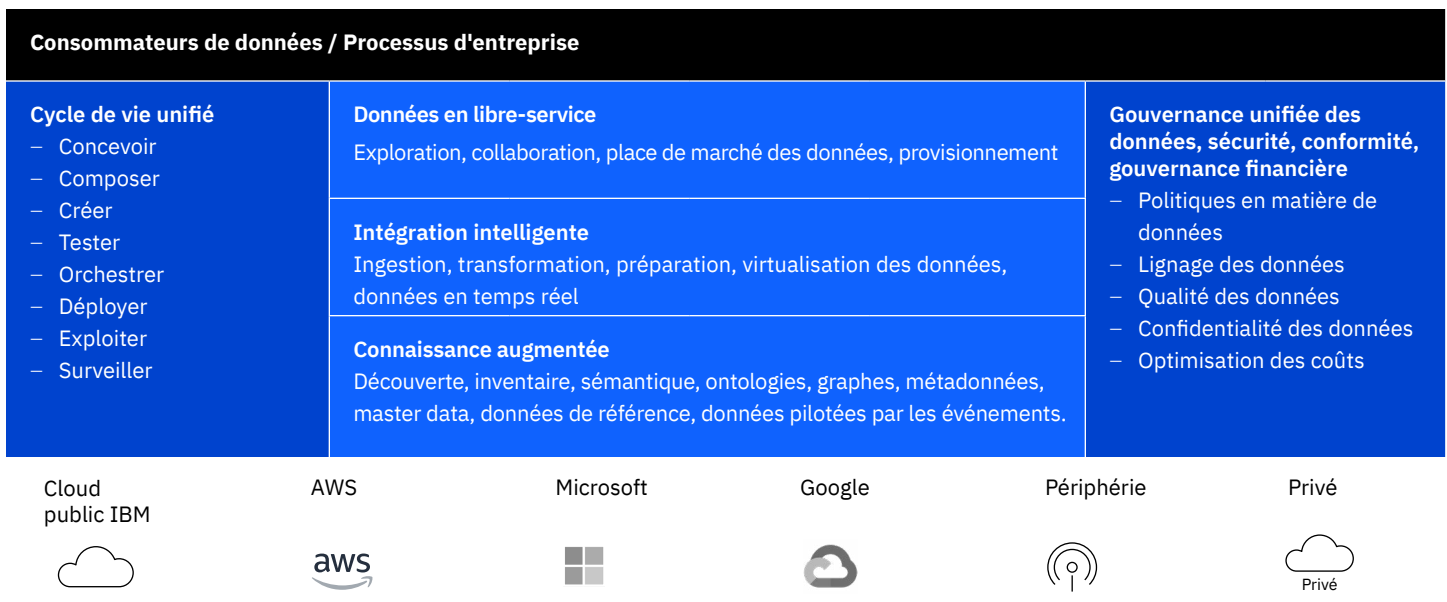


Figure 1. Les fonctionnalités de la plateforme IBM Cloud Pak for Data - qui prennent toutes en charge le développement de l'IA et de la data fabric.

La data fabric en action

La meilleure façon de comprendre la valeur d'une data fabric est d'étudier les avantages commerciaux que des organisations réelles ont obtenus grâce à la mise en œuvre d'une data fabric au-dessus d'une plateforme d'analyse.

Entreprise d'énergie

La section consacrée aux défis au début de ce livre blanc a présenté le cas d'une entreprise nord-américaine du secteur de l'énergie en pleine transformation numérique.

Ce client a fait appel à IBM, qui a mis en œuvre une architecture de données basée sur IBM Cloud Pak for Data.

Grâce à la mise en place d'une structure de données flexible et intégrée, l'entreprise a pu mettre en œuvre une série de projets de données importants dans différentes unités, notamment :

- eMobilité
- Découverte de documents sur les opérations gazières, y compris l'extraction de l'écriture manuscrite
- Segmentation de la clientèle électrique et prévision de la charge
- Gestion des ressources
- Impacts de la COVID-19 sur la charge
- Modèle de risque du retour au travail

Avec une data fabric, cette organisation est capable de :

- Fournir un accès direct aux données à de multiples lignes métier via une plateforme d'analyse en libre-service.
- Utiliser une capacité de calcul adaptée pour exécuter efficacement des modèles sur des milliards de lignes de données.
- Développer des modèles en collaboration et les déployer facilement afin d'insuffler des connaissances dans l'ensemble de l'entreprise.

Réduire les mouvements et améliorer la supervision

Un grand fournisseur de services industriels a amélioré la gouvernance de ses données et facilité la conformité réglementaire.

Cette organisation a eu du mal à transférer de grandes quantités de données vers son lac de données basé sur le cloud. Elle avait une double exigence : appliquer une forte couche de gouvernance à chaque instance d'accès aux données et déterminer la qualité des données avant de donner accès aux utilisateurs métier.

En mettant en œuvre une data fabric basée sur IBM Cloud Pak for Data, cette organisation a apporté d'énormes améliorations à ses processus de gouvernance, de conformité et de transformation des données. Avec une data fabric en place, elle peut fournir à ses utilisateurs professionnels un accès facile et sécurisé à des centaines de sources de données dans son lac de données dans le cloud et parmi les sources de données SAP sur site.

Elle s'appuie également sur les outils d'accélération de la réglementation du secteur pour analyser les sources de données à la recherche d'éventuelles données IIP qui seraient soumises à la réglementation RGPD ou CCPA. Le marquage automatique des données permet d'identifier toutes les sources de données qui doivent être gérées pour le traitement des IIP, tant sur site que dans le cloud.

Grâce à un meilleur accès et à un filtrage plus rapide des IIP, les utilisateurs professionnels peuvent exploiter les données pertinentes pour obtenir des informations importantes sans avoir à attendre longtemps pour accéder aux données ou à courir le risque d'exposer des données protégées.

Avec une data fabric, cette organisation est capable de :

- Assurer une bonne gouvernance des données, tout en exploitant simultanément les données de l'ensemble de l'entreprise.
- Obtenir des données fiables et réduire le temps consacré à la préparation des données.
- Contribuer à la conformité avec les réglementations relatives à la protection de la vie privée, telles que RGPD et CCPA.

La data fabric est la prochaine étape de la transformation de votre entreprise.

Avec une data fabric basée sur la technologie IBM Cloud Pak for Data, vous pouvez hyper-automatiser la découverte, la gouvernance et la consommation des données dans un paysage de données hybride et multicloud. Employez une data fabric pour accélérer la création de valeur pour les utilisateurs, accroître la productivité de l'ingénierie et de l'exploitation des données, et améliorer la fidélité de la gouvernance et de la conformité.

[Pour commencer, essayez gratuitement IBM Cloud Pak for Data.](#)

[Fixez un rendez-vous gratuit avec un représentant IBM spécialiste des données et l'IA.](#)



© Copyright IBM Corporation 2021

IBM France
17, avenue de l'Europe
92275 Bois-Colombes Cedex
Mai 2021

IBM, le logo IBM et IBM Cloud Pak sont des marques ou des marques déposées d'International Business Machines Corporation, aux Etats-Unis et/ou dans d'autres pays. Les autres noms de services et de produits peuvent être des marques d'IBM ou d'autres sociétés. Une liste actualisée des marques commerciales IBM est disponible sur le Web à l'adresse suivante : ibm.com/trademark.

L'information contenue dans ce document était à jour à la date de sa publication initiale, et peut être modifiée sans préavis par IBM. Les offres mentionnées dans le présent document ne sont pas toutes disponibles dans tous les pays où IBM est présent.

Les exemples cités concernant des clients et les performances ne sont présentés qu'à titre d'illustration. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même le fonctionnement des produits ou logiciels non IBM avec les produits ou logiciels IBM. LES INFORMATIONS CONTENUES DANS LE PRÉSENT DOCUMENT SONT FOURNIES "EN L'ÉTAT", SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DÉCLINE NOTAMMENT TOUTE RESPONSABILITÉ RELATIVE À CES INFORMATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DÉFAUT D'APTITUDE À L'EXÉCUTION D'UN TRAVAIL DONNÉ. Les produits IBM sont garantis conformément aux dispositions des contrats.

Déclaration de bonnes pratiques de sécurité : La sécurité des systèmes informatiques implique la protection des systèmes et des informations via la prévention, la détection et la réponse en cas d'accès incorrect, de l'intérieur ou de l'extérieur de votre entreprise. Un accès non autorisé peut entraîner la modification, la destruction, le détournement ou l'utilisation impropre des informations, ou une détérioration ou une utilisation impropre de vos systèmes, notamment en vue de les utiliser pour attaquer autrui. Aucun système ou produit informatique ne doit être considéré comme étant complètement sécurisé et aucun produit, service ou mesure de sécurité ne peut être entièrement efficace contre une utilisation ou un accès non autorisé. Les systèmes, les produits et les services IBM sont conçus pour s'intégrer à une approche de sécurité complète, qui implique nécessairement des procédures opérationnelles supplémentaires, et peuvent avoir besoin d'autres systèmes, produits ou services pour optimiser leur efficacité. IBM NE GARANTIT PAS QUE TOUTS LES SYSTÈMES, PRODUITS OU SERVICES SONT À L'ABRI DES CONDUITES MALVEILLANTES OU ILLICITES DE TIERS OU QU'ILS PROTÈGERONT VOTRE ENTREPRISE CONTRE CELLES-CI.

Il incombe au client de s'assurer de la conformité avec la législation et les réglementations applicables. IBM ne donne aucun avis juridique et ne garantit pas que ses services ou produits sont conformes aux lois applicables.

- 1 Étude Forrester New Technology : Projected Total Economic Impact™ commandée par IBM, février 2020.
- 2 Les mesures de performance ont été recueillies dans un environnement d'essai contrôlé dans les laboratoires IBM de la Silicon Valley en utilisant la virtualisation des données IBM avec diverses sources de données de 100 To. Les mesures effectuées en mai 2019 et les gains de performance sont comparés à la fédération IBM.
- 3 Basé sur l'expérience d'un client IBM.