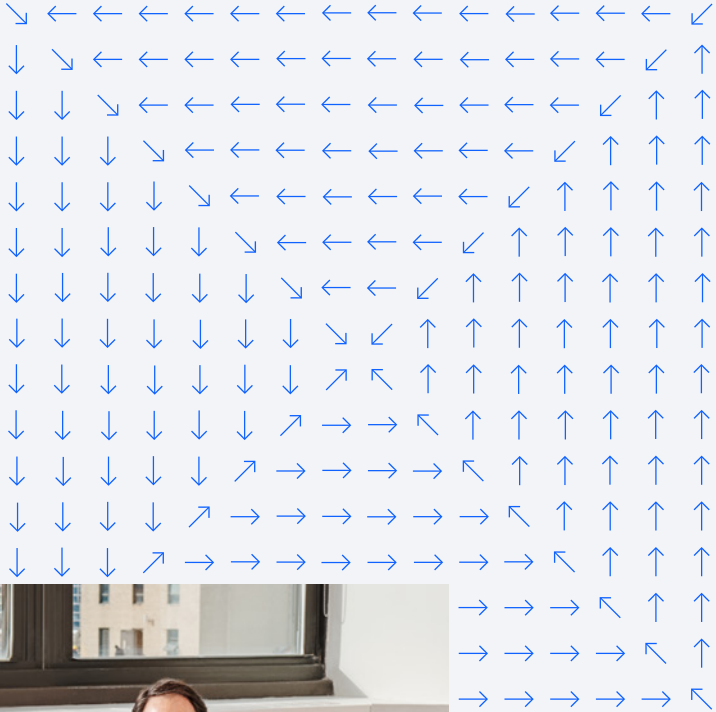




Gouvernance et
confidentialité
des données pour
↳ les responsables
des données



Sommaire



01

Introduction : une approche de la gouvernance et de la confidentialité basée sur la data fabric

02

Pourquoi mettre en place une gouvernance et une confidentialité des données automatisées ?

03

Les éléments constitutifs de la gouvernance et de la confidentialité

04

Data fabric : une approche holistique

05

Exemples de réussite en matière de gouvernance et de confidentialité des données

06

Éléments à prendre en considération

07

Créez votre solution idéale de gouvernance et de confidentialité

01

Introduction : une approche de la gouvernance et de la confidentialité basée sur la data fabric

L'architecture data fabric est une approche technologique qui permet de garantir l'accès à des données de qualité aux bonnes personnes, au moment adéquat.

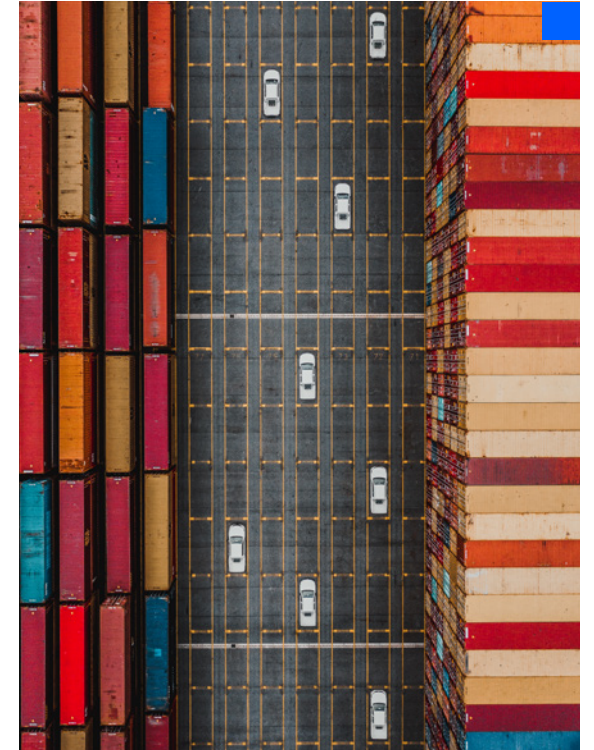
En plus de fournir une base solide pour l'intégration des données multicloud, l'intelligence client à 360 degrés et une IA fiable, la capacité de gouvernance et de confidentialité des données data fabric renforce la conformité grâce à des contrôles automatisés de gouvernance et de confidentialité. Simultanément, elle préserve la conformité avec la réglementation où que résident les données.

Une gouvernance solide permet à ceux qui doivent y avoir accès de trouver plus facilement des données adéquates, de bonne qualité, tout veillant à ce que les données sensibles demeurent cachées, sauf lorsqu'il est approprié de les consulter. Disposer d'informations sur votre entreprise et sur vos clients constitue un avantage concurrentiel. L'enquête Forrester Analytics Business Technographics® Data And Analytics de 2020 révèle que les entreprises s'appuyant sur des informations avancées sont plus susceptibles que des sociétés débutantes et intermédiaires d'avoir une stratégie de gouvernance des données qui implique de définir la conformité, de la mettre en œuvre,

de la surveiller et de dispenser des formations à son sujet. L'enquête indique également que ces entreprises disposent généralement d'un responsable chargé de la gouvernance des données et qu'elles utilisent l'IA pour créer et intégrer l'intendance des données à leurs opérations quotidiennes sur les données.¹

De solides paramètres de confidentialité permettent de mieux se préparer à la conformité et à la protection des données, où qu'elles soient, sur site ou dans plusieurs clouds. Ils permettent aux entreprises de comprendre les politiques réglementaires et les règles de gouvernance propres à leur secteur d'activité et de les appliquer rapidement aux données, où qu'elles se trouvent.

Dans ce guide, nous examinerons les défis les plus courants en matière de gouvernance et de confidentialité auxquels les organisations modernes sont confrontées. Nous présenterons aussi les éléments constitutifs d'une solution ou approche efficace et les composants technologiques dont vous aurez besoin pour créer une couche automatisée et intégrée de gouvernance et de confidentialité des données pour toutes les données de votre entreprise. Nous présenterons également des ressources utiles telles qu'un [essai de solution sur la gouvernance et la confidentialité des données](#).



02

Pourquoi mettre en place une gouvernance et une confidentialité des données automatisées ?

Alors que les organisations s'efforcent d'instaurer une culture de prise de décision basée sur les données, il est essentiel de pouvoir s'appuyer sur des données de qualité conformes à un environnement réglementaire dynamique. Une telle approche permet aux organisations aux impératifs suivants :

Protéger la confidentialité des données à grande échelle

Les risques liés à la non-conformité (comme des sanctions juridiques, la perte de confiance des clients et la perte de réputation) sont réels. Plus de 60 pays ont adopté ou proposé des lois sur la protection de la confidentialité et des données. D'ici 2023, plus de 80 % des entreprises dans le monde devront se conformer à au moins une réglementation de protection des données centrée sur la confidentialité.²

Plutôt que de répondre à chaque défi individuellement, une approche dynamique de la protection de la confidentialité et des données est l'occasion pour les organisations de renforcer la confiance des clients. Cependant, pour y parvenir, les responsables des données doivent mettre en place un programme complet de protection de la confidentialité dans l'ensemble de l'organisation.

Améliorer l'accès aux données

Le partage sécurisé des données est un facteur crucial lorsque plusieurs équipes doivent accéder aux données de l'entreprise. Ces données doivent être traçables et visibles uniquement par ceux qui sont autorisés à les utiliser. Pourtant, 7 organisations sur 10 ne sont pas en mesure de sécuriser les données qui naviguent entre environnements multcloud et sur site.³

Lorsqu'elles ne peuvent pas garantir la conformité à grande échelle, et d'un environnement à l'autre, les équipes hésitent à partager les données entre les différentes unités commerciales, ce qui renforce l'effet Silo. Les équipes informatiques doivent donc protéger chaque référentiel de données individuellement, ce qui peut amener des groupes à créer leurs propres référentiels (IT cachée) et contribue à accroître la complexité.

Selon le cabinet d'avocats DLA Piper, depuis le 28 janvier 2021, les autorités chargées de la protection des données dans l'UE ont imposé des amendes d'un montant total de 1,2 milliard de dollars suite à des violations du RGPD.⁴

02

Pourquoi mettre en place une gouvernance et une confidentialité des données automatisées ?

Maintenir des normes de qualité des données dans toute l'organisation

Seuls 20 % des cadres d'entreprise font entièrement confiance aux données qu'ils obtiennent.⁵ Selon un récent rapport de Gartner, la mauvaise qualité des données coûte 12,9 millions de dollars en moyenne aux organisations chaque année. Gartner prévoit que, d'ici 2022, 70 % des entreprises suivront rigoureusement les niveaux de qualité des données au moyen d'indicateurs et qu'ils les amélioreront de 60 % afin de réduire considérablement les risques et les coûts d'exploitation.⁶

Pour que tous les utilisateurs d'une organisation soient en mesure de comprendre pleinement les données qu'ils s'approprient à utiliser et d'avoir confiance en elles, il est essentiel de mettre en place une gouvernance des données fondée sur les métadonnées et sur une définition propre à l'entreprise. Les conditions commerciales, les classifications de données, les données de référence, les métadonnées associées, ainsi que l'établissement et l'application de politiques et de règles de gouvernance des données, constituent la base de cette approche.

Suivi et traçabilité des données

Une fois que les équipes d'analystes ont créé et déployé ces produits de données (tableaux de bord, rapports et modèles de machine learning), elles doivent pouvoir remonter l'historique pour comprendre comment les données ont été produites. Pour les cas d'utilisation liés à l'audit et à la conformité (souvent présents dans les secteurs réglementés), une équipe d'analystes peut être amenée à exposer toutes les étapes de la vie des données, depuis le système transactionnel où elles ont été initialement créées jusqu'à leur forme finale, lorsqu'elles sont utilisées pour aider à la prise de décisions commerciales. En ce qui concerne les utilisateurs finaux, connaître la source des données et les transformations subies par ces dernières peut offrir un gain de temps précieux lorsqu'ils créent leur propre version personnalisée du tableau de bord.

Faciliter la consommation des données

Pour exploiter le pouvoir d'innovation et de transformation des données, les entreprises doivent permettre la consommation des données en libre-service. Pour simplifier l'accès aux données et leur utilisation, il faut un cadre et une architecture solides garantissant que les utilisateurs de données d'une organisation peuvent facilement trouver et utiliser les bonnes données grâce à un index des actifs catalogués riche en métadonnées. La gouvernance et la confidentialité des données permettent aux entreprises de répondre de manière dynamique à la nécessité d'innover et d'atteindre les objectifs commerciaux.

03

Les éléments constitutifs de la gouvernance et de la confidentialité

En définitive, l'objectif de la gouvernance est de savoir d'où viennent les données, quelle est leur nature, qui peut y accéder et quand elles doivent être supprimées. Il existe plusieurs éléments technologiques clés pour répondre à la nécessité d'intégrer et d'améliorer la confidentialité, la qualité et la traçabilité de toutes les données d'une organisation, et l'accès à ces données.

Voyons ce dont vous avez besoin.

N°1

Catalogage des données

La qualité de vos données détermine la confiance avec laquelle vous pouvez agir en fonction des informations réunies. Si des données de mauvaise qualité sont introduites dans les modèles d'IA, cela peut conduire à des résultats inexacts, non conformes ou discriminatoires. Pour obtenir les meilleures informations, il faut pouvoir accéder à des données récentes, propres et pertinentes, avec une taxonomie cohérente. Le catalogage des données peut aider les utilisateurs à trouver et à utiliser facilement les bonnes données grâce à un index riche et axé sur les métadonnées des actifs catalogués.

[Lire le « Guide complet pour le catalogue de données moderne » →](#)

N°2

Génération automatisée de métadonnées

Les métadonnées permettent de suivre l'origine, le niveau de confidentialité, l'âge et les utilisations potentielles de vos données. La création manuelle de métadonnées est fastidieuse, mais avec le machine learning, les données peuvent être automatiquement marquées avec des métadonnées pour limiter les erreurs humaines et les oublis de données. Le marquage automatique des métadonnées permet de faire respecter les politiques au point d'accès, de sorte que davantage de données sensibles peuvent ensuite être utilisées de manière non identifiable et conforme. En outre, les métadonnées sont utilisées pour établir un vocabulaire commun de termes propres à l'entreprise qui fournissent un contexte aux données et permettent de relier des données provenant de sources différentes. Ce contexte ajoute une signification sémantique aux données qui deviennent ainsi plus faciles à trouver et à utiliser, et plus cohérentes dans l'organisation. Il s'agit d'un élément décisif pour qui recherche des données pour l'analyse et l'IA.

03

Les éléments constitutifs de la gouvernance et de la confidentialité

N°3

Gouvernance automatisée de l'accès aux données et de leur traçabilité

La traçabilité des données indique comment les données ont été consultées et utilisées, et par qui. Savoir d'où proviennent les données est utile non seulement pour les rapports de conformité, mais aussi pour créer des modèles d'IA fiables et explicables. Et cela peut être automatisé sans compliquer l'accès. Grâce aux restrictions intégrées directement aux points d'accès, seules les données auxquelles les utilisateurs sont autorisés à accéder sont visibles. En outre, les données sensibles peuvent être masquées de manière dynamique afin que les modèles et les ensembles de données soient partagés sans présenter de données privées à des utilisateurs non autorisés. Cette définition claire des données qui peuvent ou non être utilisées soutient les demandes de données en libre-service et permet aux organisations de répondre avec flexibilité aux besoins métier.

N°4

Virtualisation des données

La virtualisation des données relie les données de tous les sites, de sorte que des sources de données disparates apparaissent comme une base de données unique. Vous pouvez ainsi garantir un accès conforme aux données grâce à un accès gouverné, quel que soit l'endroit où se trouvent les données, sans déplacement. Grâce à la couche unique de gouvernance virtualisée, l'accès des utilisateurs aux données est défini en un seul endroit et non plus au niveau de chaque source, ce qui réduit la complexité de la gestion des accès.

N°5

Rapports et audits

Les entreprises doivent se conformer à une grande variété de réglementations changeantes et disparates selon le pays, le secteur d'activité et le type de données. Ces réglementations doivent être réparties dans un catalogue d'exigences assorti d'un ensemble clair d'actions que les entreprises doivent entreprendre. Les informations réglementaires doivent être automatiquement intégrées, dédoublonnées et appliquées aux flux de travail.

Le secret pour harmoniser tous ces besoins en matière de confidentialité et de gouvernance des données avec les opportunités commerciales consiste à aligner les composants technologiques sur une stratégie globale de données et une architecture ouverte et holistique.

04

Data fabric : une approche holistique

Pour exploiter les données à des fins de compréhension et de croissance de l'entreprise (et, en fin de compte, pour créer une culture dans laquelle les décisions sont fondées sur les données), vous devez adopter une approche holistique de l'architecture et de la stratégie des données qui est efficace et ne repose pas sur l'assemblage manuel de plusieurs solutions. C'est pourquoi de nombreuses organisations adoptent une architecture data fabric.

Une architecture **data fabric** est une approche visant à simplifier l'accès aux données et leur consommation au sein d'une organisation. Cette architecture est indépendante des environnements de données, des processus, du fonctionnement et de la géographie. Elle associe des systèmes de données disparates à des capacités de gestion et de gouvernance des données complètes, simplifiant l'accès aux données en libre-service et la collaboration.

Avec une architecture data fabric, les entreprises valorisent davantage leurs données en fournissant les bonnes données, au bon moment, quel que soit leur emplacement. Elle rassemble des capacités telles que celles énumérées précédemment dans une architecture unifiée, évitant ainsi le coût et la complexité de l'intégration de multiples solutions indépendantes. Au lieu d'un groupe de produits disparates, la data fabric offre une solution unique et holistique, conçue pour fonctionner en toute transparence.

En outre, une architecture data fabric peut répondre à trois cas d'utilisation distincts, au-delà de la gouvernance et de la confidentialité des données. Il s'agit notamment de l'intégration des données multicloud, des vues à 360 degrés des clients, ainsi que des MLOps et de l'IA fiable (traités dans des eBooks distincts).

■ Avec la data fabric, les entreprises valorisent davantage leurs informations en fournissant les bonnes données, au bon moment, quel que soit leur emplacement.

D'ici 2024, les déploiements d'infrastructures data fabric quadrupleront l'efficacité de l'utilisation des données tout en réduisant de moitié les tâches de gestion des données effectuées par les humains.⁷

05

Exemples de réussite en matière de gouvernance et de confidentialité des données

Services financiers : ING ↻

ING est une banque néerlandaise qui compte plus de 57 000 employés au service de près de 39,3 millions de clients, entreprises et établissements financiers dans plus de 40 pays. Pour concrétiser sa vision de la gouvernance des données chez ING, Ferd Scheepers, architecte en chef d'ING, souhaitait mettre en œuvre une architecture data fabric dans l'environnement de cloud hybride de l'entreprise.

ING avait besoin de gérer ses données dans le cloud de manière cohérente avec son environnement sur site. En tant que responsable des données, M. Scheepers avait des objectifs précis :

- Donner aux utilisateurs de données d'ING un accès rapide et simple aux données et aux outils de gouvernance
- Garantir des paramètres de gouvernance et de confidentialité solides dans un écosystème mondial complexe
- Se conformer à la politique de l'entreprise et aux différentes évolutions des réglementations nationales

ING a créé une solution data fabric pour aider à mettre en œuvre un modèle opérationnel d'entreprise unique et à rationaliser la gestion des données et des applications dans tous les pays d'implantation.

Cette solution s'exécute dans un environnement de cloud hybride ouvert qui s'adapte au paysage multiplateforme et hétérogène d'ING. En appliquant la virtualisation des données aux investissements existants sur site, elle supprime les silos de données, permettant ainsi un accès en temps opportun aux bonnes données dans n'importe quel cloud et sur site, pour un coût optimal, avec le niveau de gouvernance approprié.

Grâce à son architecture data fabric, ING peut offrir une expérience utilisateur cohérente pour accroître la collaboration, rationaliser la gestion des applications et optimiser les coûts informatiques et d'octroi de licence.

[Lire l'étude de cas →](#)



06

Éléments à prendre en considération

Pour mettre en place une solide capacité de gouvernance et de protection de la confidentialité, il faut un ensemble de technologies conçues pour assurer une gouvernance complète, fournir des données de qualité et, enfin, accélérer la collaboration. Dans le cadre d'une entreprise, la valeur de la gouvernance des données est amplifiée lorsque cette capacité est associée à l'intégration des données, ce qui permet d'avoir une vue d'ensemble des clients et d'utiliser au maximum les données pour obtenir des résultats commerciaux.

Dans le cadre d'une architecture data fabric moderne, la capacité à assurer la gouvernance et la confidentialité des données crée une expérience utilisateur complète, fondée sur les métadonnées et la gestion active des politiques, qui permet aux utilisateurs de visualiser, de consulter, de manipuler et d'analyser les données sans avoir à en comprendre le format physique ou l'emplacement, et sans avoir à les déplacer ou à les copier.

Les composants technologiques de l'approche data fabric d'IBM permettent aux entreprises d'appliquer automatiquement des règles et des politiques réglementaires spécifiques à leur secteur d'activité et à leurs ressources de données, en sécurisant l'ensemble de l'entreprise, grâce à :

- Un catalogue de données enrichi par l'IA permettant aux utilisateurs professionnels de comprendre et d'enrichir les données, de collaborer et d'accéder facilement aux bonnes données
- Une couche de métadonnées et de gouvernance pour toutes les données, les analyses et les initiatives d'IA qui augmente la visibilité et la collaboration sur n'importe quel cloud
- La possibilité de masquer les données de manière dynamique et cohérente à un niveau de granularité défini par l'utilisateur
- La possibilité de créer des données de formation et des ensembles de tests anonymes tout en préservant l'intégrité des données



IBM Cloud Pak® for Data

IBM Cloud Pak® for Data est une plateforme construite spécifiquement pour une architecture data fabric visant à prévoir les résultats plus rapidement pour vous permettre de collecter, d'organiser et d'analyser vos données, quel que soit leur emplacement. La plateforme permet ainsi d'améliorer la productivité et de réduire la complexité en construisant une architecture data fabric qui relie les données en silo réparties dans un environnement de cloud hybride.

[En savoir plus sur IBM Cloud Pak for Data →](#)

06

Éléments à prendre en considération



IBM Watson Knowledge Catalog

IBM Watson® Knowledge Catalog offre un catalogage intelligent, avec collecte automatisée des métadonnées et gestion des politiques, pour veiller à ce que les détails d'un modèle soient automatiquement collectés et stockés afin d'atteindre une transparence et une répétabilité maximales. Il garantit que les modèles sont impartiaux, qu'ils ne sont pas biaisés, qu'ils sont explicables et qu'ils s'adaptent à l'évolution des paramètres du modèle.

[En savoir plus sur IBM Watson Knowledge Catalog →](#)



IBM Watson Query

L'application de règles de gouvernance de portée générale à l'ensemble des data lakes, des bases de données et des entrepôts de données prend beaucoup de temps. En outre, les utilisateurs doivent souvent attendre longtemps avant d'avoir accès aux bonnes données. Watson Query applique des politiques de gouvernance lorsque les données sont réparties sur des sources multiples. Les données sont rapidement communiquées à vos applications finales via une vue unique, sans modification manuelle, déplacement de données ou réplication.

[En savoir plus sur IBM Watson Query →](#)

07

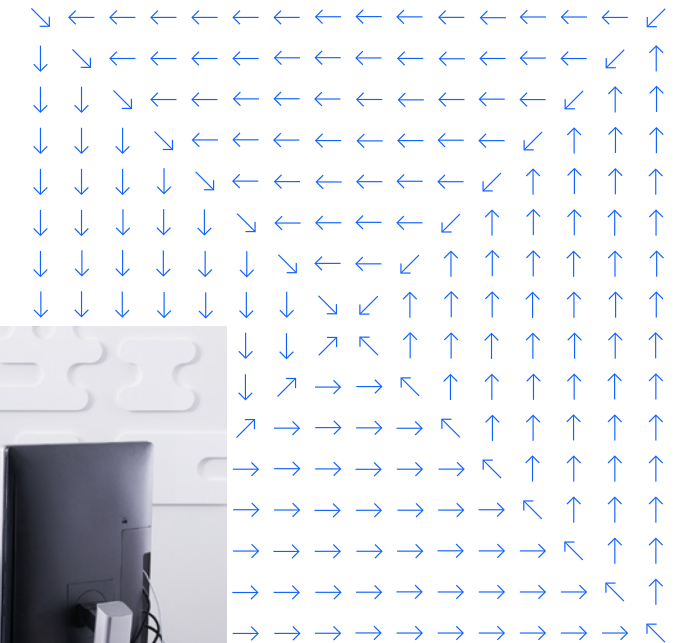
Créez votre solution idéale de gouvernance et de confidentialité

Si vous êtes prêt à adopter une stratégie et une architecture unifiées pour améliorer l'accessibilité, la sécurité et la conformité de vos données de tous types et provenant de toutes sources, nous vous encourageons à profiter de quelques ressources.

En premier lieu, l'[essai de solution de gouvernance et de confidentialité des données](#), gratuit, qui permet d'acquérir une expérience pratique d'une architecture data fabric conçue pour ce cas d'utilisation. Ensuite, consultez les informations sur notre [site Web IBM consacré à la gouvernance et la confidentialité des données](#) pour en savoir encore plus sur les avantages. Enfin, vous pouvez contacter l'un de nos spécialistes en prenant [rendez-vous en ligne](#), en parlant avec votre représentant IBM ou en vous adressant à l'un de nos partenaires commerciaux.

Consultez les trois autres eBooks sur les cas d'utilisation d'une architecture data fabric :

- [Intégration de données multicloud](#)
- [Vue client à 360°](#)
- [MLOps et IA fiable](#)





© Copyright IBM Corporation 2022

Compagnie IBM France
17 avenue de l'Europe
92275 Bois-Colombes Cedex

Produit aux États-Unis d'Amérique
Mai 2022

IBM, le logo IBM et ibm.com sont des marques commerciales d'International Business Machines Corp., déposées dans de nombreux 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 sur la page Web « Copyright and trademark information » à l'adresse ibm.com/legal/copytrade.html.

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 la société IBM est présente.

Les données de performance et les exemples de clients cités sont uniquement présentés à titre indicatif. Les résultats de performance réels peuvent varier en fonction des configurations et des conditions d'exploitation spécifiques.

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 NI AUCUNE GARANTIE OU CONDITION D'ABSENCE DE CONTREFAÇON. Les produits IBM sont garantis conformément aux dispositions des contrats qui régissent leur utilisation.

- 01 Forrester, "Break Through Data Governance Fatigue, A Framework For Effectiveness And Sustainability," February 25th, 2021.
- 02 Gartner, "Hype Cycle™ for Data Privacy," 2021.
- 03 McCurdy, Chris, Shue-Jane Thompson, Lisa-Gaine Fisher, and Gerald Parham. "Getting started with zero trust security: A guide for building cyber resilience." IBM Institute for Business Value. August 2021. <https://ibm.co/zero-trust-security>
- 04 CNBC, "Fines for breaches of EU privacy law spike sevenfold to \$1.2 billion, as Big Tech bears the brunt," January 2022
- 05 The data-powered enterprise, Capgemini Research Institute
- 06 Gartner, "How to Improve Your Data Quality," July 14, 2021 <https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality>
- 07 Gartner "Top Strategic Technology Trends for 2022: Data Fabric," Mark Beyer, Ehtisham Zaidi, Guido De Simoni, October 18, 2021.