

Un lac de données gouverné pour des informations métier fiables

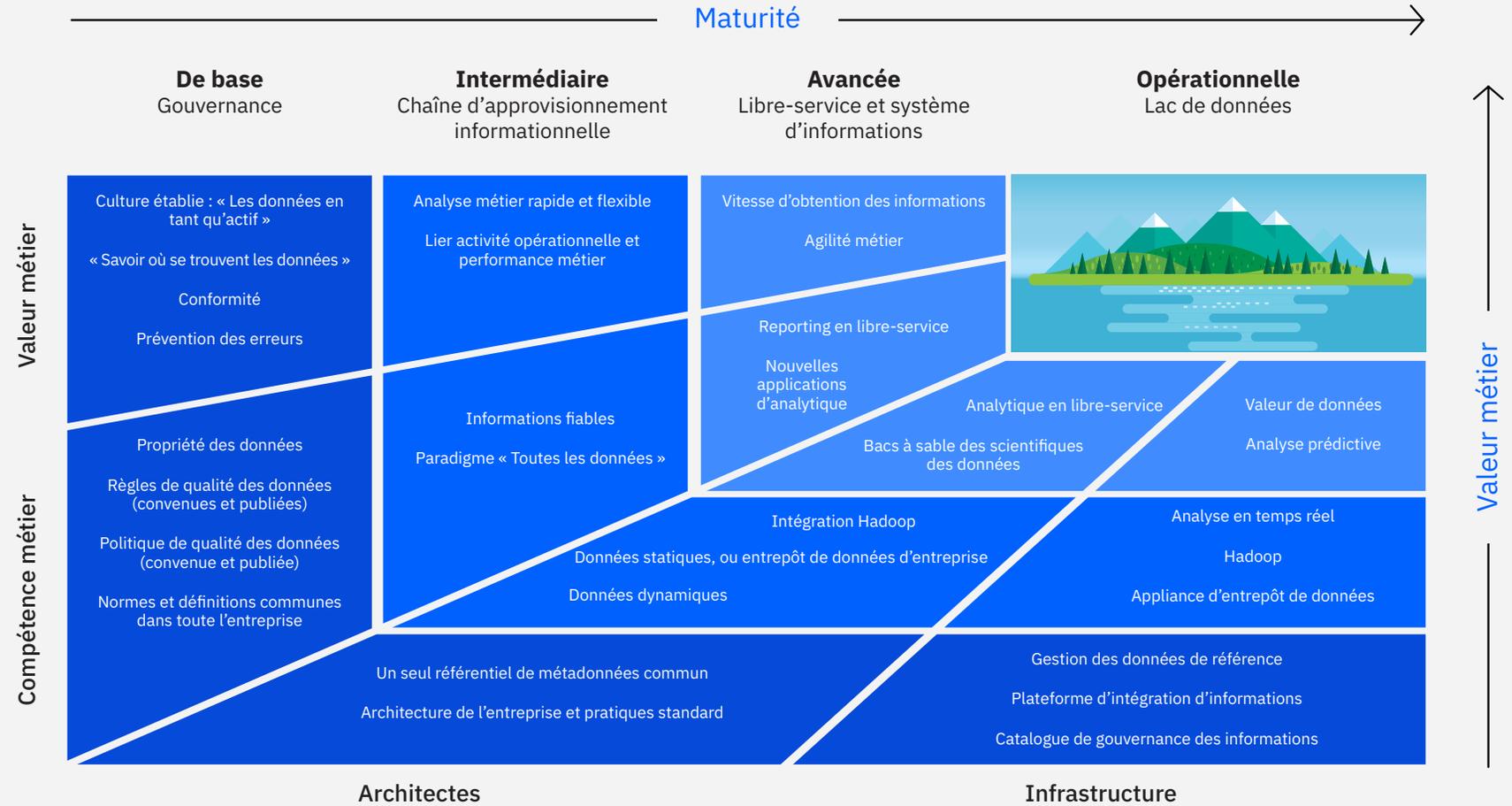
*Explorez les composants clés pour
fournir efficacement des données fiables*

IBM Cloud



Les lacs de données gouvernés créent de la valeur

Les lacs de données sont des solutions idéales pour les organisations qui priorisent les données dans leur stratégie opérationnelle. Le partage sécurisé des données est un facteur crucial lorsque plusieurs équipes ont besoin d'accéder aux données de l'entreprise. Pour les aider, les entreprises peuvent s'appuyer sur un lac de données gouverné contenant des données brutes structurées et non structurées – fiables, sécurisées et gouvernées. Pour les organisations qui cherchent à extraire de la valeur de leurs données, notamment celles sur leurs clients, leurs employés, leurs transactions et d'autres actifs, les lacs de données gouvernés créent des opportunités d'identifier, de comprendre, de partager et d'agir en toute confiance vis-à-vis de ces informations.



Architecture d'un lac de données gouverné

Les principales décisions de conception caractérisent l'architecture d'un lac de données gouverné. Un réservoir de données se compose de trois éléments : des référentiels de lac de données, qui fournissent des plateformes permettant de stocker des données et d'exécuter des analyses au plus près des données ; des services de lac de données, qui permettent de localiser, d'accéder, de préparer, de transformer, de traiter et de déplacer les données vers et hors des référentiels du réservoir de données ; et une matrice de gestion et de gouvernance des informations, qui aide à gouverner et gérer les données stockées dans le lac de données.

Les fonctionnalités de gouvernance valident et améliorent la qualité des données et ont pour but de les protéger contre toute utilisation abusive. Cette mesure veille à ce que les données soient actualisées, conservées puis supprimées à des points appropriés de leur cycle de vie.

La gouvernance, l'organisation de vos données et le fait de pouvoir avoir toute confiance dans leur qualité constituent un aspect important de la gestion d'un lac de données. Bien qu'un lac de données soit conçu pour offrir un accès flexible aux données, vous avez besoin d'un système de gouvernance pour vous assurer que vos données sont sécurisées, protégées et continuent d'être utiles. Le lac de données gouverné peut être illustré par les différentes couches qui le composent, comme suit :

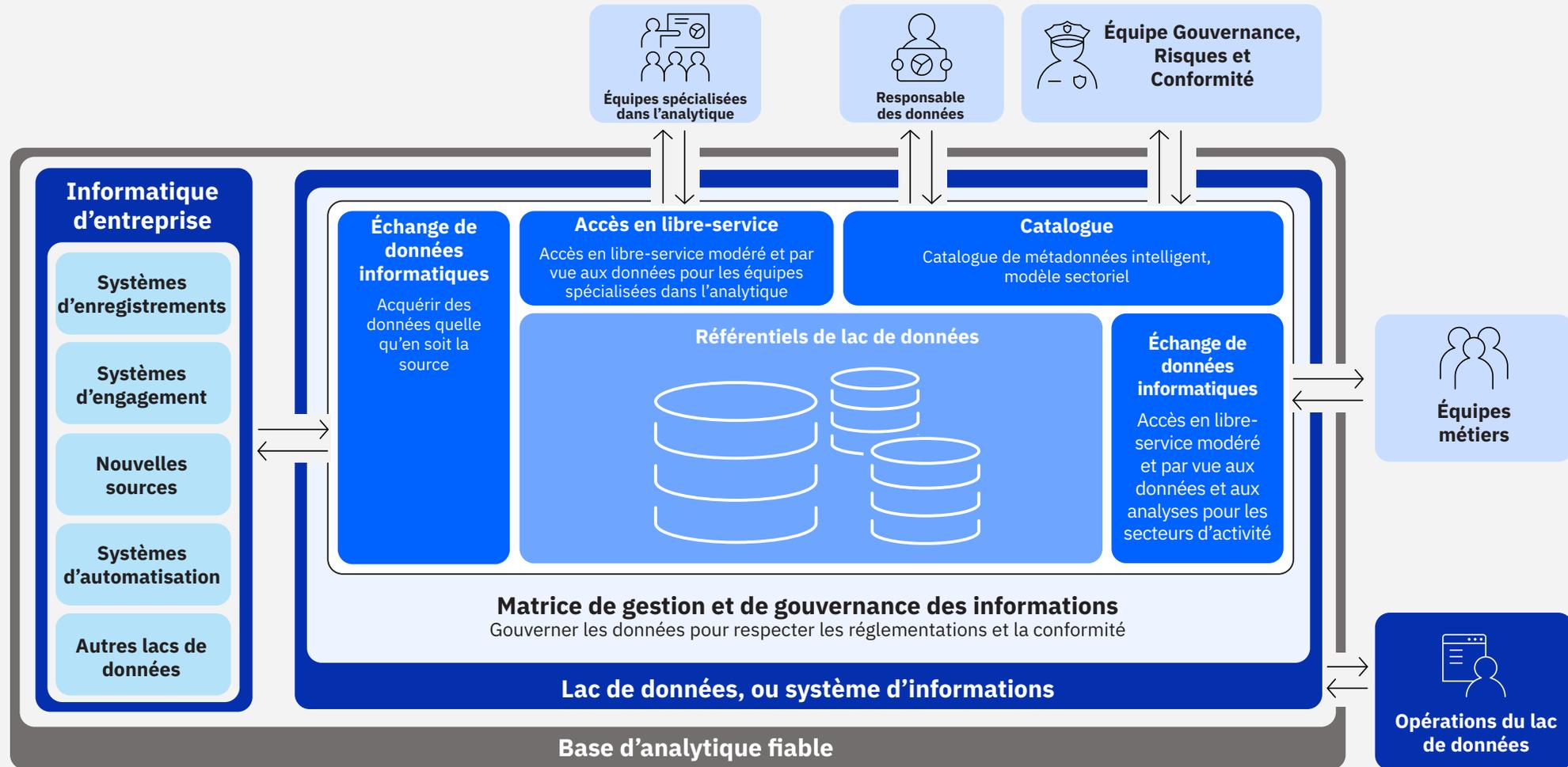
- La couche de base, qui repose principalement sur la gouvernance des données
- La couche intermédiaire, qui étend les référentiels de lac de données initiaux avec de nouveaux types et comportements de données
- La couche avancée, qui permet d'exécuter des analyses en libre-service

Chaque couche renferme une valeur spécifique pour les différents groupes de l'organisation. Les architectes bénéficient d'une architecture de référence publiée qui repose sur un référentiel unique et commun de métadonnées. Les scientifiques des données bénéficient d'une zone contrôlée où ils peuvent placer les bacs à sable en cours.

Les principaux avantages d'un lac de données découlent de la gouvernance. La gouvernance favorise une culture qui priorise les données, dans laquelle les utilisateurs métier s'approprient les données et s'accordent sur des règles et des politiques. Les définitions partagées créent une compréhension mutuelle, ce qui permet d'éviter toute confusion au sein des équipes ou entre ces dernières. Grâce à cette base commune, vous pouvez accéder à des données fiables et obtenir plus rapidement de l'information à l'aide de vos applications d'analytique. La valeur métier passe alors de la connaissance des données et de leur importance à la possibilité d'exécuter des [analyses flexibles](#) à tout moment.

Un lac de données modulaire et évolutif se compose de plusieurs éléments qui encouragent l'accès en libre-service dans l'ensemble de votre organisation.

Architecture d'un lac de données gouverné



IBM Cloud / DOC ID / Mars 2018 / © 2018 IBM Corporation

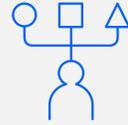
Les quatre types de consommateurs de données

Les utilisateurs qui consomment les données d'un lac de données sont divers et variés. Comprendre les différentes façons dont ils approchent les données est un aspect important d'une gouvernance réussie.



Équipes spécialisées dans l'analytique

- Spécialistes des données, qui gèrent les données et créent des modèles
- Développeurs en analytique, qui transforment ces modèles en applications
- Développeurs d'applications, qui incorporent ces applications d'analytique aux systèmes opérationnels



Responsable des données

- Optimiser la qualité des données et préparer des tâches ETL
- Cataloguer les données et gérer les métadonnées
- Maintenir l'équilibre entre protection des données et confidentialité



Équipe Gouvernance, Risques et Conformité

- Spécialistes de la gouvernance des données qui définissent des politiques de gouvernance et de sécurité des données
- Protéger les données pour garantir l'application de contrôles de confidentialité dans tous les processus
- Compiler les exigences de conservation, d'archivage et de suppression de données, et garantir la conformité des données avec les politiques et réglementations en vigueur



Équipes métiers

- Dirigeants métiers comme les CMO, CFO ou CHRO
- Chief Data Officers qui émergent comme responsables de données
- Dirigeants métiers qui mettent en œuvre des systèmes pour obtenir des résultats métiers spécifiques ou des informations exploitables

Composants clés d'un lac de données gouverné

Un lac de données gouverné est une architecture de référence indépendante de toute technologie spécifique qui inclut des processus de gouvernance et de gestion. Il ne s'agit pas d'une solution Hadoop ou d'un entrepôt de données d'entreprise que vous pouvez acheter ou remplacer. Un lac de données gouverné est une solution sur site ou basée sur le Cloud destinée aux organisations qui souhaitent mettre les données au cœur de leurs opérations. Les [composants clés](#) d'un lac de données gouverné sont les suivants :



Échange de données informatiques – pour extraire, analyser, affiner, transformer et échanger des données entre les lacs de données et les systèmes informatiques de l'entreprise. Ce processus permet de nettoyer les données et de surveiller leur qualité de façon continue.



Catalogue – pour décrire les données du lac de données – leur signification, comment elles sont classifiées et les exigences de gouvernance qui en résultent.



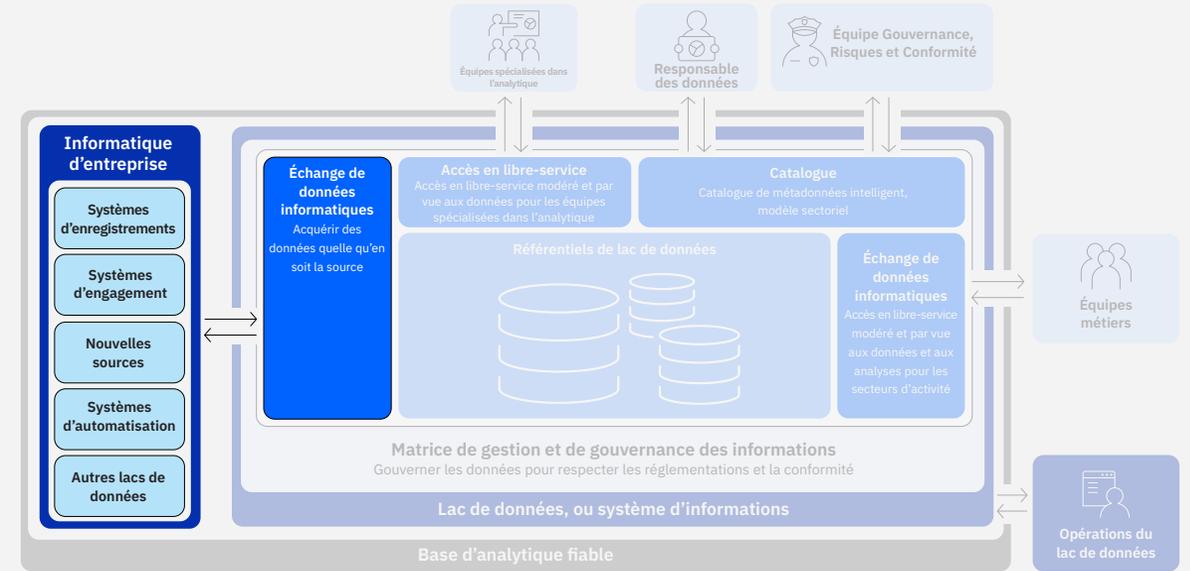
Gouvernance – pour gérer les données du lac de données et appliquer des politiques et des niveaux de sécurité, de qualité et de confidentialité appropriés aux données stockées dans le lac.



Accès en libre-service – trois groupes de services qui offrent un accès à la demande au lac de données. L'accès en libre-service pour les utilisateurs spécialisés dans l'analytique permet d'accéder aux données brutes telles qu'elles sont stockées. Pour les équipes métiers, le service fournit des données normalisées dans des structures de données simplifiées. Pour les équipes Gouvernance, Risques et Conformité, le service fournit des données gouvernées à des fins d'audit.

Ingestion de données issues de différentes sources

Ingestion est un terme qui désigne le processus d'extraction, de transformation, de traitement qualité et d'échange de données entre le lac de données, les systèmes informatiques de l'entreprise et les autres lacs de données existants. Une grande partie des données du lac de données provient des systèmes informatiques de l'entreprise. Ces types de données peuvent être structurés, semi-structurés ou non structurés. Les sources de données peuvent être des systèmes à la base de l'activité, un journal de site Web ou d'autres sources surveillant l'activité. IBM offre une haute évolutivité en termes de volume de données et des fonctionnalités enrichies de transformation et de réplication.



IBM Cloud / DOC ID / Mars 2018 / © 2018 IBM Corporation



Quand c'est fait et bien fait

- Les données affluent dans le lac de données sans interruption
- Possibilité d'analyser des données transformées, normalisées et enrichies
- Coûts de stockage réduits même lorsque les volumes de données augmentent
- Possibilité d'utiliser un bac à sable à des fins d'analyse exploratoire



Quand ce n'est pas ou mal fait

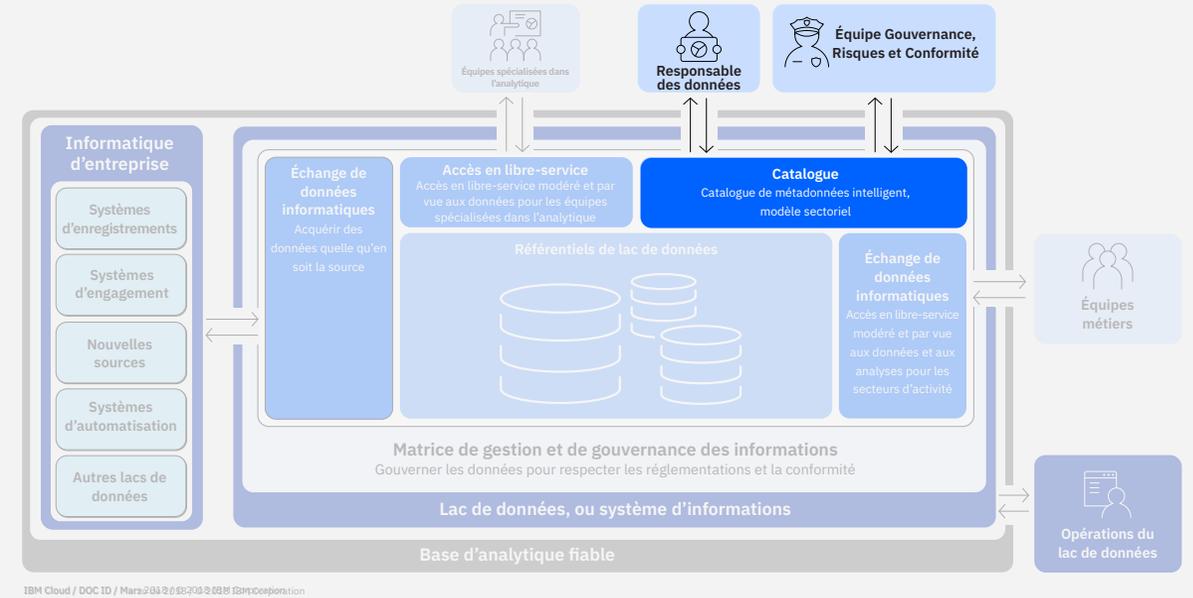
- Difficulté à actualiser les données lorsque les volumes augmentent
- Impossible d'utiliser des actifs informationnels non structurés
- Coûts de stockage plus importants
- Difficulté à nettoyer les données, ce qui entraîne une augmentation des coûts de traitement des données

Catalogage

Le catalogage aide à étiqueter les données dans le lac de données et à créer un inventaire de vos actifs informationnels. Les interfaces de catalogue fournissent aux utilisateurs du lac de données des informations sur les données : leur classification, leur lignée, comment elles sont gouvernées, etc. Les offres IBM en matière de catalogage de lac de données incluent les fonctionnalités suivantes :

- Possibilité de capturer des actifs informationnels non structurés dans le catalogue
- Intégration à un écosystème ouvert pour n'importe quel actif informationnel
 - Catalogue d'entreprise unique pour tous les actifs informationnels de l'organisation
 - Données et termes métier propres au secteur d'activité
 - Fonctions d'évaluation et d'étiquetage social dans le cadre des métadonnées

Les données envoyées dans le [pipeline de gouvernance](#) doivent être comprises, afin que les données techniques fassent sens d'un point de vue commercial. Par exemple, un numéro à neuf chiffres peut représenter un numéro de sécurité sociale ou un identifiant d'employé, ou les deux. L'étape de classification et d'attribution de termes métier ajoute une signification métier aux données techniques. L'automatisation est une étape essentielle pour adapter ce processus au volume et à la variété des données stockées dans le lac. Puis, après une série de flux de travaux de traitement, d'évaluations de qualité et de contrôles, les données peuvent être cataloguées afin d'être accessibles dans l'ensemble de l'entreprise.



Mis en œuvre et correctement mis en œuvre

- Résultats plus rapides et plus de temps pour analyser les données
- Capture d'actifs contextuels et amélioration de l'utilité des données
- Meilleur suivi de la lignée des données et fiabilité accrue des données
- Commercialisation des actifs informationnels pour une consommation à plus grande échelle
- Aide à la conformité des données



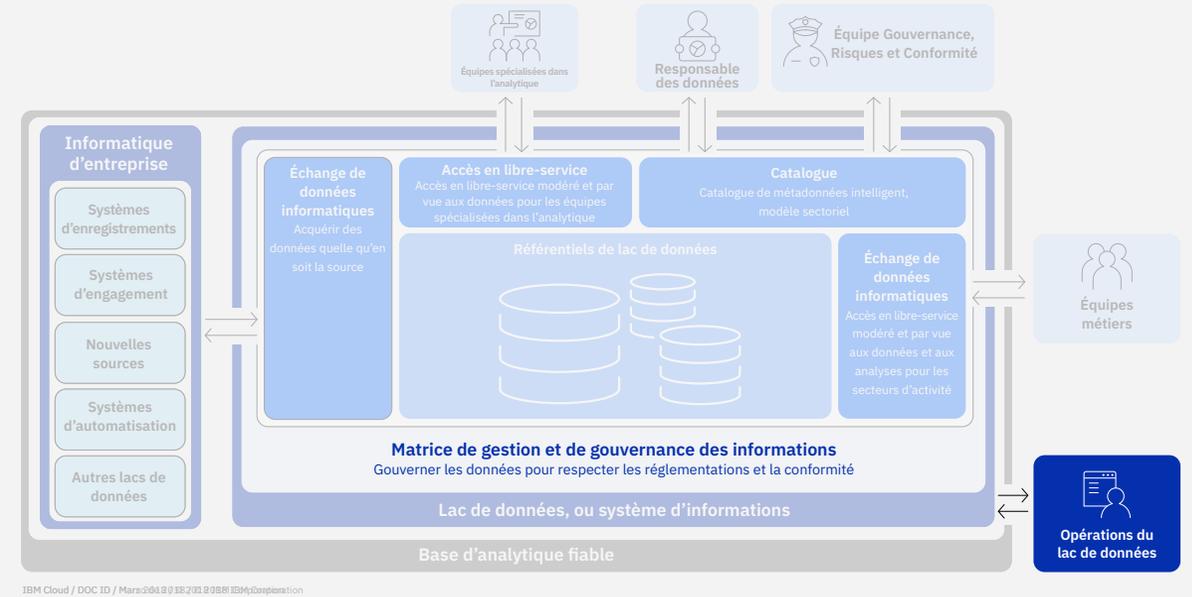
Non mis en œuvre ou incorrectement mis en œuvre

- Perte de temps à rechercher et étiqueter les données
- Difficulté à localiser vos données et impossibilité de trouver des collègues qui comprennent ces données
- Impossibilité de savoir qui a accès aux données
- Non-respect des exigences de conformité et de gouvernance

Gouverner et gérer vos données

La matrice d'intégration et de gouvernance des informations permet au système de suivre efficacement votre lac de données afin que les informations entrantes soient bien comprises et que les stratégies de gouvernance que vous avez définies soient automatiquement appliquées. L'infrastructure de gouvernance vous aide à documenter vos politiques de gouvernance et à établir des règles pour définir la façon dont les informations doivent être structurées, stockées, transformées et déplacées.

Les exigences en matière de gouvernance des informations sont documentées dans le catalogue sous la forme de politiques, de règles et de classifications. Avec les solutions IBM, les actifs non structurés peuvent faire partie intégrante du lac de données, et le volume, la variété et la rapidité des niveaux de données sont préservés.



Mis en œuvre et correctement mis en œuvre

- Capacité à suivre l'évolution des volumes de données et à les gouverner
- Respect des exigences réglementaires à l'aide d'outils de conformité propres au secteur d'activité
- Adoption plus rapide des données de référence
- Informations plus précises avec des données de haute qualité
- Meilleure réactivité vis-à-vis des audits de conformité
- Meilleure capacité à protéger vos données



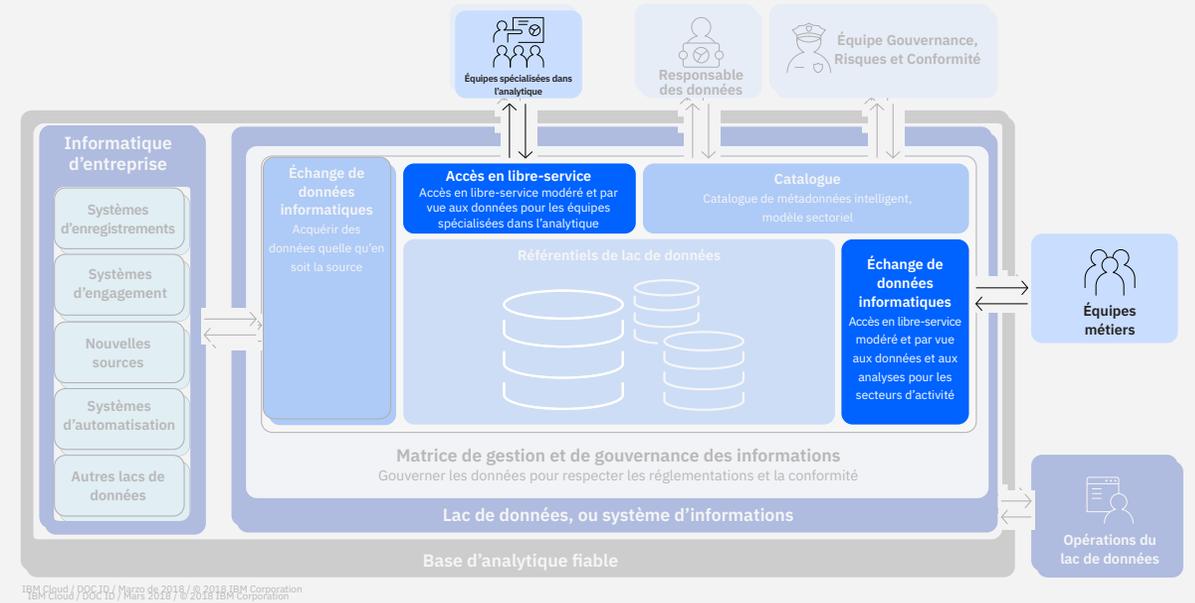
Non mis en œuvre ou incorrectement mis en œuvre

- Impossibilité de gérer l'augmentation des volumes de données issues de sources structurées comme non structurées
- Perte de temps à trouver des données, ce qui impacte l'état de préparation de l'entreprise aux audits
- Opportunités manquées de respecter les exigences de conformité et de gouvernance

Libre-service ou reporting

L'accès en libre-service vous aide à trouver des informations pertinentes à partir de vos données via de simples interfaces de recherche. Il fournit des données fiables de haute qualité aux utilisateurs autonomes qui peuvent utiliser ces données pour créer des modèles analytiques dans le cadre de leurs initiatives de science des données. Il permet également aux utilisateurs sans connaissances techniques de transformer les données avant de créer et de déployer des modèles.

Un accès direct aux données aide les informaticiens à préparer et transformer les données. Il permet aux équipes de gouvernance et de conformité de traiter les données à des fins d'audit. Il aide également les utilisateurs de solutions à créer des rapports personnalisés répondant à leurs besoins métier, et à accéder à des données prêtes à l'emploi pour prendre des décisions rapides et extraire des informations précieuses de leurs données.



Mis en œuvre et correctement mis en œuvre

- Les utilisateurs de données ont accès à des données contextuelles
- Les consommateurs de données peuvent faire confiance aux données à leur disposition grâce aux connaissances, aux balises sociales et aux évaluations qualitatives associées aux actifs informationnels
- Les données deviennent un actif organisationnel accessible à tous les consommateurs de données
- ROI plus rapide
- Rapidité d'innovation
- Exploration et analyse agiles et itératives des données



Non mis en œuvre ou incorrectement mis en œuvre

- Plus de temps passé à rechercher et préparer des données plutôt qu'à analyser ces dernières
- Impossibilité de trouver ou d'accéder à des actifs informationnels non structurés
- Prise de décision plus longue en raison d'un accès limité à des données fiables
- Innovation au ralenti

Pourquoi IBM

Selon une étude menée par Radiant Advisors, la gouvernance et la sécurité sont considérées par 72 % des dirigeants comme les principaux défis à relever, mais aussi les principaux facteurs de réussite de leur organisation. Faire de la gouvernance et de l'architecture d'informations une priorité est la première étape. Cela permet d'engager la conversation au sein de l'organisation pour définir clairement ce que tous les utilisateurs de données exigent de leurs données. Dans un monde où posséder des données de mauvaise qualité équivaut à distribuer des données de mauvaise qualité, chaque utilisateur de données est au cœur de la conversation.

Déployer une plateforme unique à l'échelle de l'entreprise dédiée à l'intégration, au traitement qualité et à la gouvernance des données est essentiel au succès de vos initiatives d'analytique. Vous pourrez ainsi acquérir des données, en assurer la qualité et les gouverner afin d'alimenter vos processus d'analyse. En abordant ces défis avec une approche gouvernée de votre lac de données, vous créez une base solide pour distribuer des données fiables pouvant être utilisées à de nombreuses fins.

Aucun autre fournisseur ne peut rivaliser avec l'exhaustivité et la richesse de la [Plateforme IBM Unified Governance and Integration](#). Que vous recherchiez une meilleure évolutivité pour gérer d'immenses volumes de données ; des accélérateurs spécifiques à un secteur ; des fonctions pour rendre des données structurées, non structurées et semi-structurées utilisables ; ou des compétences éprouvées en matière d'apprentissage automatique et d'intelligence artificielle, IBM possède la solution complète qui vous permettra de créer un lac de données gouverné et fiable.

Pour en savoir plus, visitez le site ibm.com/governed-data-lake.



Compagnie IBM France

17 avenue de l'Europe
92275 Bois-Colombes Cedex

L'adresse de la page d'accueil IBM est :

ibm.com

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corp. dans de nombreux pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. Une liste actualisée de toutes les marques IBM est disponible sur la page Web « Copyright and trademark information », à l'adresse suivante : ibm.com/legal/copytrade.shtml

Le présent document contient des informations qui étaient en vigueur et valides à la date de la première publication et qui peuvent être modifiées par IBM à tout moment. Toutes les offres mentionnées ne sont pas distribuées dans tous les pays où IBM exerce son activité.

LES INFORMATIONS DU PRÉSENT DOCUMENT SONT FOURNIES « EN L'ÉTAT » ET SANS GARANTIE EXPLICITE OU IMPLICITE D'AUCUNE SORTIE. 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 avec lesquels ils sont fournis.

Il incombe au client de s'assurer qu'il respecte les lois et réglementations applicables. IBM ne donne aucun avis juridique et ne garantit pas que ses produits ou services permettent au client de se conformer aux lois applicables.

Instructions relatives aux bonnes pratiques de sécurité : la sécurité des systèmes d'information implique la protection des systèmes et des informations par la prévention, la détection et la réponse aux accès non autorisés effectués depuis l'intérieur comme depuis l'extérieur de votre entreprise. Un accès non autorisé peut se traduire par la modification, la destruction ou une utilisation inadéquate ou malveillante de vos données et de vos systèmes, ainsi que par leur utilisation pour attaquer d'autres systèmes. Aucun système ou produit informatique ne doit être considéré comme totalement sécurisé, et aucun produit, service ou mesure de sécurité ne peut empêcher complètement les utilisations ou les accès inappropriés. Les systèmes, produits et services IBM sont conçus pour s'intégrer à une approche de sécurité légale et complète, qui implique nécessairement des procédures opérationnelles supplémentaires, et peuvent avoir besoin d'autres systèmes, produits ou services pour offrir une efficacité maximale. IBM NE GARANTIT EN AUCUNE MANIÈRE QU'UN SYSTÈME, PRODUIT OU SERVICE EST IMMUNISÉ, OU IMMUNISERA VOTRE ENTREPRISE, CONTRE LES AGISSEMENTS MALVEILLANTS OU ILLÉGAUX DE QUI QUE CE SOIT.