



L'architecture de l'information : le point de départ vers l'intelligence artificielle et des données métier exploitables

Découvrez une solution analytique fiable
centrée sur la gouvernance et le catalogage

Introduction

Table des matières

- Introduction
- Les composantes de base d'une solution analytique fiable
- L'apprentissage automatique, accélérateur de la gouvernance
- Un système unique multi-objectifs
- Conclusion

Points clés

- La conformité incite les entreprises à implémenter des stratégies de gouvernance des données continues et bénéfiques.
- L'apprentissage automatique automatise les initiatives de gouvernance et d'intégration à grande échelle. Il permet ainsi de surmonter les difficultés posées par des volumes énormes de données et par les limites des capacités humaines.
- La gouvernance est efficace à la fois dans les environnements locaux et multicloud.

Les données connaissent une prolifération rapide, tant sur le plan de la quantité que de la variété, dans toutes les catégories d'entreprises. Dans les environnements multicloud, toute une gamme de sources de données est à l'origine d'une croissance exponentielle du flux d'informations reçues : Internet des objets (IoT) et médias sociaux, appareils mobiles, systèmes de réalité virtuelle et suivi optique. Alors même que les entreprises investissent spontanément dans l'intelligence artificielle (IA), la plupart n'ont pas réalisé d'état des lieux sérieux permettant de bien comprendre leurs données ou d'en garantir la qualité pour tirer correctement parti des solutions d'IA. Dans de nombreuses entreprises, les données sont inaccessibles, peu fiables, ou non conformes aux règles de confidentialité et de protection des données.

Les réglementations internationales telles que le RGPD (Règlement général sur la protection des données) dans l'Union européenne, la loi américaine CCPA (California Consumer Privacy Act) en Californie, et la LGPD au Brésil (Lei Geral de Proteção de Dados) sont axées principalement sur les données des clients et des employés. Ces réglementations sont pour les entreprises l'occasion de transformer et de créer de nouveaux modèles de gestion centrés sur les données. C'est une vision positive en dépit des lourdes pénalités qui peuvent découler d'une non-conformité, et qui peuvent ralentir la productivité ou nuire à l'image de marque. Pour répondre à leurs obligations en matière de confidentialité et protéger les informations personnelles, les entreprises doivent d'abord découvrir et classer leurs différents types de données. Celles qui peinent à collecter ou exploiter correctement les données clients risquent de se heurter très vite à des difficultés. Pour relever ce défi, les entreprises implémentent [des architectures d'informations gouvernées](#) qui tiennent compte des réglementations, tout en continuant à favoriser la performance et l'innovation organisationnelles dirigées par les données.

Les composantes de base d'une solution analytique fiable

L'attitude qui consiste à voir les réglementations sur la confidentialité des données comme une obligation, mais aussi comme une opportunité de moderniser les infrastructures de données comporte de nombreux avantages. Elle peut en effet inciter les entreprises à implémenter des stratégies de gouvernance des données qui donnent naissance à de nouveaux modèles métier et permettent d'obtenir de l'information à partir des données. Les initiatives de gouvernance unifiée et d'intégration (UGI) s'appliquent aux données, tant structurées que non structurées, dans les clouds publics et privés. L'implémentation d'une initiative UGI dans un but de conformité est importante en soi, mais a aussi un intérêt pour d'autres secteurs au sein de l'entreprise, en particulier la gouvernance des modèles d'IA pour les scientifiques des données.

Lorsqu'une entreprise a recours à la gouvernance pour garantir la fiabilité de ses données, les utilisateurs ont la certitude que celles-ci proviennent d'une source de qualité. Ils savent comment les données sont utilisées dans toute l'entreprise et en quoi elles permettent d'enrichir un projet d'analyse. Les initiatives analytiques exigent des données fiables pour générer de bons résultats, peu importe le degré de sophistication des outils employés. Des données métier fiables et prêtes à être exploitées semblent présenter des avantages infinis. L'analyse permet d'imaginer de nouveaux concepts de produits, de nouveaux programmes marketing, et d'améliorer les initiatives de ventes, de chaîne d'approvisionnement ou de service à la clientèle. Elle permet même de détecter des insuffisances opérationnelles, et, en les éliminant, d'améliorer l'agilité organisationnelle et de booster les résultats financiers.

L'implémentation de la gouvernance des données et de l'intelligence artificielle dans votre entreprise repose sur plusieurs composantes de base, qui sont décrites ci-après.

Détection et qualité des données

Une entreprise n'a pas toujours connaissance des volumes réels des données qu'elle possède. La première étape de la gouvernance des données consiste à établir l'inventaire des données de l'entreprise. Il est conseillé de commencer par étudier les ensembles de données d'un projet précis, puis de passer aux autres cas d'utilisation métier pour avoir une perspective plus élargie de l'entreprise. Le stockage et la gestion des données redondantes, obsolètes ou triviales sont non seulement coûteux, mais **ralentissent aussi la prise de décision et les opérations**. Ces données compliquent la mise en conformité et nuisent aux projets analytiques. Les données doivent répondre à certains critères de qualité et s'y tenir pour pouvoir être exploitables en aval.

Catalogage

Une fois détectées et profilées, les données sont cataloguées à l'aide de balises de métadonnées qui identifient les types des données, leur utilisation, leur appartenance, leur origine, et ainsi de suite. Dans certains secteurs d'activité, les entreprises ont des exigences communes : le recours à des modèles préconfigurés utilisant des termes métier et une taxinomie déjà existants permet donc d'accélérer le processus de catalogage. Les progrès de l'apprentissage automatique permettent d'établir une correspondance automatique entre les termes et de créer un **catalogue d'entreprise** en quelques heures. L'UGI fournit une base de départ de catalogage qui permet aux entreprises de gouverner leurs modèles d'IA, leurs blocs-notes et d'autres sources de données et de créer ainsi une bibliothèque centrale regroupant les connaissances de l'entreprise. Ce catalogage est une ressource précieuse pour de nombreux utilisateurs de données de l'entreprise : ingénieurs en traitement de données, data stewards et certains utilisateurs métier tels que les analystes, les scientifiques des données et les marketeurs.

Transfert, transformation et synchronisation des données

Les données provenant de sources multiples peuvent être facilement intégrées, transformées et partagées avec d'autres systèmes, physiquement ou virtuellement. Ce processus regroupe les données structurées et non structurées et permet l'intégration avec des technologies ouvertes comme Apache Atlas et Hadoop. L'automatisation et la synchronisation des flux garantissent la disponibilité des données les plus récentes dans les lacs, les entrepôts, les magasins de données et les solutions finales. Au fur et à mesure que les quantités de données augmentent, la réplication prend en charge de gros volumes avec un faible temps de latence. Les entreprises peuvent utiliser la virtualisation sans transférer les données, en fonction de leurs besoins.

En 2019, les utilisateurs de systèmes en libre-service généreront plus d'analyses que les professionnels des données.

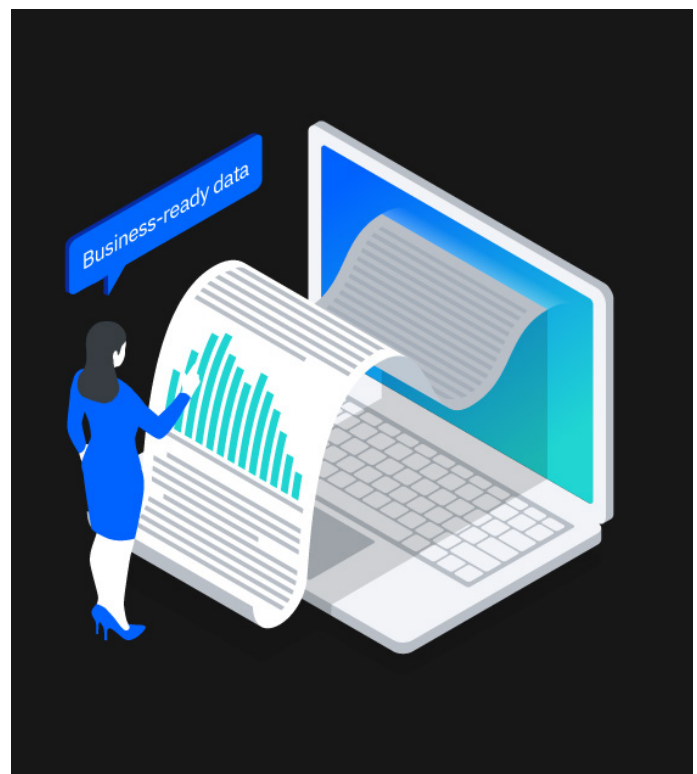
Gestion des données de référence

Selon Gartner, en 2019, les utilisateurs de systèmes en libre-service généreront plus d'analyses que les professionnels des données. Il est vital que les entreprises disposent d'une vue complète, fiable et unifiée d'entités critiques telles que les clients, les produits et les comptes. [Les solutions modernes de gestion des données de référence \(MDM\)](#) comprennent une exploration analytique graphique, un moteur de correspondance de haute précision, une approche privilégiant les données lors de la sélection des algorithmes de correspondance et des processus d'intendance utilisant l'apprentissage automatique. Ces solutions proposent également un accès agile en libre-service, des outils de gouvernance et des fonctions de tableau de bord conviviales.

Confidentialité et protection des données

Les entreprises doivent sécuriser et protéger proactivement leurs ressources d'informations stratégiques et sensibles. La gestion du cycle de vie des données englobe toutes les étapes, depuis la création à la suppression, et fait appel à des pratiques telles que la gestion des enregistrements, la gestion des conflits et l'archivage. L'application de l'apprentissage cognitif aux documents et à l'historique d'une entreprise permet d'identifier automatiquement les risques en fonction du contexte de l'entreprise.

Une gouvernance dans le cadre des opérations métier et de la conformité permet de disposer de données immédiatement exploitables par l'entreprise à des fins de prise de décisions, d'amélioration ou d'innovation. Au fur et à mesure que les quantités des données augmentent, la réplication prend en charge de gros volumes avec une faible latence. Les entreprises peuvent utiliser la virtualisation sans transférer les données, en fonction de leurs besoins.



L'apprentissage automatique : un accélérateur de la gouvernance

L'apprentissage automatique augmente l'intelligence humaine et, grâce aux progrès technologiques récents, vient compléter les limites des capacités humaines. Il automatise les initiatives de gouvernance et d'intégration à grande échelle et permet de surmonter les difficultés posées par des volumes énormes de données, avec à la clé une gouvernance saine des données dans toute l'entreprise. Pour prendre un exemple, dans une entreprise possédant 20000 termes de

données, il faut en moyenne six mois à une équipe de six personnes pour classifier manuellement ces termes et obtenir une analyse fiable. Avec l'apprentissage automatique, il ne faut que quelques jours, voire quelques heures, selon la quantité des ressources de données. Ce niveau de rapidité retire un lourd fardeau au processus de gouvernance. L'apprentissage automatique simplifie la gestion des obligations de conformité, et ouvre la voie à des initiatives analytiques productives.

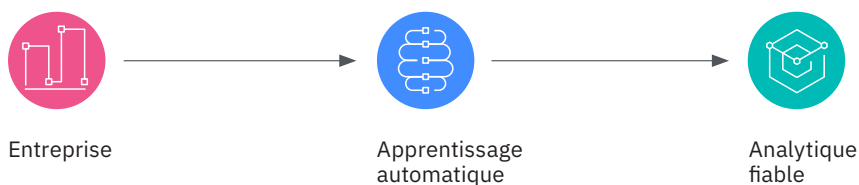
Classification manuelle

Durée nécessaire : 6 mois



Classification par l'apprentissage automatique

Durée : quelques jours ou quelques heures



Un système unique multi-objectifs

Un système gouverné est utilisable dans toutes les unités métier et dans toute l'entreprise, comme le montrent les exemples suivants de cas d'utilisation courants.

Lac de données gouverné

De nombreuses entreprises ont lancé des projets d'analyse prédictive de Big data dans divers secteurs d'activités. Une première étape essentielle consiste à déposer de grandes quantités de données structurées et non structurées dans un lac de données.

Les entreprises ont choisi Hadoop ou des solutions de codage manuel pour leurs lacs de données. Mais l'absence d'une structure de gouvernance, associée à une stratégie inexistante de gestion des normes de données, des termes métier, ainsi que de l'origine, de l'utilisation et de la qualité des données peuvent vite transformer ce lac de données en un marécage. C'est la conséquence lorsque les utilisateurs ne comprennent pas leurs données et n'ont pas confiance en elles. Les entreprises ont donc compris que leur lac de données nécessite une stratégie et une gouvernance pour réussir.

Dans [une implémentation gouvernée](#), les données d'un lac sont mises en correspondance avec des termes métier faciles à comprendre par tous les utilisateurs et homogénéisées dans toute l'entreprise. L'accès à ces données accélère la rentabilisation de quasiment tous les projets en libre-service de sciences des données, d'exploration des données ou d'intelligence artificielle, avec à la clé l'agilité. En outre, un tel accès jette les bases nécessaires à la prise en charge d'environnements multicloud, d'architectures « on premise » et de diverses sources de données.

Modernisation d'applications

Les entreprises investissent beaucoup dans la modernisation des applications afin d'améliorer l'efficacité, de réduire les coûts et de se doter d'un atout concurrentiel. Selon l'entreprise, les solutions de modernisation des applications peuvent prendre des formes très diverses. Les principaux enjeux sont la gestion des données de test, la virtualisation des données et la connectivité. Les entreprises font désormais appel

à une méthodologie agile pour tester et développer simultanément avec un [accès aux données virtuelles](#). Elles choisissent aussi des fonctions d'intégration souples pour connecter les applications métier et les données.

Une vue à 360 degrés des données clients

Il est important que les employés disposent d'[une vue unique des informations de référence](#) sur les clients, les produits ou d'autres entités, et que ces données soient fiables, actualisées et exactes. Des informations erronées ou périmées peuvent nuire aux interactions avec les clients et éroder la confiance, provoquer une fuite des comptes clients ou augmenter les coûts de la chaîne d'approvisionnement. Les données obtenues grâce aux processus UGI permettent aux interactions des clients de se transformer en outils susceptibles d'accroître la confiance, la fidélité à la marque, la valeur de la marque et d'optimiser l'agilité de la chaîne d'approvisionnement.

Optimisation de l'entrepôt de données

Lorsqu'un entrepôt de données est utilisé, l'optimisation de l'architecture améliore et modifie considérablement la façon dont les données sont consultées, stockées, préparées, gouvernées et analysées. L'une des meilleures stratégies [d'optimisation efficace](#) consiste à décharger les tâches ETL (extraction, transformation et chargement), les données inutilisées et les données nécessaires aux modèles d'exploration. Ce processus réduit les coûts et permet aussi de combiner les données avec d'autres types de données dans les environnements tels que les lacs gouvernés, afin de mettre en place une exploration dynamique.

Conformité réglementaire

Un système d'analyse fiable renforce et [accélère la conformité](#) aux obligations réglementaires. Surtout, la mise en œuvre de la protection des données personnelles commence par une définition de ces données. Cette étape permet à l'entreprise de savoir de quelles données à caractère personnel elle dispose. Le catalogue des données de départ contient des règles de gouvernance concernant la qualité, l'enrichissement et l'analyse des données, et les politiques de conformité.

Conclusion

Alors que les entreprises opèrent leur transformation numérique, leurs dirigeants se sensibilisent aux avantages d'une gouvernance globale de leurs modèles de données et d'intelligence artificielle, dans des environnements « on premises » ou multicloud. En se concentrant sur des pratiques de gouvernance essentielles, les entreprises préparent leurs données et leurs solutions d'IA non seulement en vue du traitement analytique et de l'obtention de connaissances, mais aussi à la mise en conformité aux réglementations auxquelles elles sont assujetties. Dans un contexte de volumes de données considérables, les pratiques liées à l'apprentissage automatique et à l'intelligence artificielle contribuent à augmenter les capacités et l'intelligence humaines pour les tâches telles que le mappage des données, le catalogage, la mise en correspondance de gros volumes de données et la préservation de la qualité des données.

Les dirigeants d'entreprise visionnaires ont compris que le fait de prendre du temps pour créer une solide base UGI se traduit par une rentabilisation importante, dans l'immédiat et dans un proche avenir. Ils savent également que leurs entreprises se doteront d'atouts majeurs en adoptant la gouvernance, qui permet de mettre en œuvre l'optimisation métier, l'innovation et la conformité dans toutes les initiatives de données et d'intelligence artificielle. Il est désormais critique d'utiliser des solutions qui englobent de A à Z la gestion de l'exploitation des données, depuis leur création jusqu'à leur consommation. La rationalisation de ces opérations exige des économies en termes de gamme, d'échelle et de partage.

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

La page d'accueil d'IBM se trouve à l'adresse :
ibm.com

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corp., déposées dans de nombreuses juridictions réparties dans le monde entier. Les autres noms de produit et de service 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 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 ne sont pas disponibles dans tous les pays dans lesquels IBM est présent.

Les exemples cités concernant des clients ne sont présentés qu'à titre d'illustration. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation spécifiques.

LES INFORMATIONS DE CE DOCUMENT SONT DISTRIBUÉES « TELLES QUELLES » SANS AUCUNE GARANTIE NI EXPLICITE NI 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.

Il est de la responsabilité de chaque client IBM de s'assurer qu'il respecte la législation et la réglementation applicables. IBM ne donne aucun avis juridique et ne garantit pas que ses services ou produits sont conformes aux lois applicables.

© Copyright IBM Corporation 2019

En savoir plus

Les solutions UGI d'IBM vous aident à créer une base analytique fiable pour mettre en œuvre l'IA à grande échelle.

[Visitez le site Web
ibm.com/unified-governance-integration](https://ibm.com/unified-governance-integration)

[Explorez pour en savoir plus](#)
Continuez à vous informer sur la gouvernance cognitive des données et découvrez les solutions IBM d'apprentissage automatique et d'intelligence artificielle, leaders sur le marché. [Lisez le livre blanc.](#)

[Dialoguez avec un expert](#)
Communiquez avec des leaders d'opinion, des ingénieurs émérites et des experts en UGI qui ont collaboré avec des milliers de clients pour élaborer des stratégies gagnantes de données, d'analyse et d'IA. [Planifiez une consultation de 30 minutes](#)