



Points clés

- La conformité peut apparaître comme une corvée onéreuse et coûteuse, mais elle s'accompagne d'un avantage notoire : elle pousse les entreprises à mettre en œuvre des stratégies de gouvernance des données.
 - Avec les récents progrès technologiques, l'apprentissage automatique permet de surmonter les limites des capacités humaines.
 - Une fois que le système de gouvernance unifiée et d'intégration des données (UG&I) a été créé pour un projet, il sert de tremplin à d'autres cas d'usage, à l'échelle de l'entreprise ou de ses entités commerciales.
 - Alors que les entreprises opèrent leur transformation numérique, leurs dirigeants ont pris conscience des avantages de la gouvernance des données dans leurs environnements, que ces derniers soient sur site ou sur un cloud public ou privé.
-

Gouvernance unifiée et intégration des données : les fondations pour assurer la conformité et extraire des connaissances

Dans toutes les catégories d'entreprises, les données prolifèrent à toute allure, tant sur le plan de la quantité que de la variété. Dans les environnements multi-cloud, de nouvelles sources font exploser le flux croissant des informations : Internet des Objets (IoT), médias sociaux, appareils mobiles, systèmes de réalité virtuelle et suivi optique.

En parallèle, les réglementations font surgir des obligations de conformité qu'il est impossible d'ignorer. Entré en vigueur au mois de mai 2018, le [Règlement Général sur la Protection des Données](#) (RGPD) de l'Union européenne est devenu une préoccupation majeure des chefs d'entreprise. Les exigences de conformité aux normes PCI-DSS (services financiers et distribution), et à la législation HIPAA (soins de santé), pour n'en citer que quelques-unes, sont tout aussi importantes.

Un système de gouvernance unifiée et d'intégration des données (UG&I) se fonde sur trois préceptes : « bien connaître ses données » « se fier à ses données » et « savoir utiliser ses données », et s'appuie sur sept grands piliers.

La croissance des données a placé plus haut la barre des enjeux de la conformité et de la libre circulation des données. Faute de contrôler leurs données en appliquant une stratégie analytique sûre, basée sur des solutions unifiées de gouvernance et d'intégration, les entreprises risquent non seulement d'être submergées par une avalanche d'informations, mais aussi de s'écarter dangereusement de la conformité réglementaire. Par chance, de nouvelles technologies telles que l'apprentissage automatique et l'intelligence artificielle ont fait leur apparition, et éliminent les difficultés de mise en place d'une gouvernance et d'une conformité solides. Les chefs d'entreprise





visionnaires ont compris que si la gouvernance des données est essentielle pour la conformité, elle est aussi le tremplin vers l'agilité et l'innovation.

Les piliers d'un système analytique fiable

La conformité peut apparaître comme une corvée onéreuse, mais elle s'accompagne d'un avantage notoire : elle pousse les entreprises à mettre en œuvre des stratégies de gouvernance des données. Les initiatives de gouvernance unifiée et d'intégration des données (UG&I) doivent s'appliquer aux données, structurées ou non, dans les clouds publics et privés. Un projet de gouvernance et d'intégration des données est à lui seul très rentable, mais sa valeur ajoutée bénéficie aussi à d'autres domaines de l'entreprise, en particulier l'analytique. Tout projet d'analytique exige de disposer de données fiables pour donner des résultats concluants, indépendamment du degré de sophistication des outils analytiques utilisés. Pour être fiables, les données doivent répondre à trois critères : vous devez être certain qu'elles proviennent d'une source de qualité, savoir comment elles sont utilisées dans toute votre entreprise, et connaître les avantages qu'elles peuvent apporter à vos projets analytiques. L'analytique permet d'imaginer de nouveaux concepts de produits et programmes marketing. Elle peut aussi augmenter les ventes et améliorer la chaîne d'approvisionnement ou les initiatives du service client. Elle permet de détecter les inefficacités opérationnelles qui, une fois supprimées, permettront à l'entreprise d'être plus agile et d'augmenter

ses bénéfices. Un système UG&I se fonde sur trois préceptes : « bien connaître ses données », « se fier à ses données » et « savoir utiliser ses données », et s'appuie sur sept grands piliers :

1. Découverte des données. Au fil du temps, les entreprises ont amassé les données de façon chaotique. Elles ont ainsi parfois accumulé sans le savoir d'énormes volumes de données. La première étape de la gouvernance des données consiste à établir l'inventaire de toutes les données de l'entreprise. Une bonne approche est d'étudier d'abord en détail en projet spécifique, puis d'élargir le champ d'investigation au reste de l'entreprise. Un point de départ très productif consiste à choisir un domaine susceptible d'être rentabilisé rapidement.

2. Catalogage. Une fois les données détectées, vous devez les cataloguer à l'aide de balises de métadonnées qui vous servent à identifier les types des données, leur utilisation, leurs propriétaires, leur origine, et ainsi de suite. Comme certains secteurs ont des besoins communs, l'utilisation de modèles de données pré-configurés permet d'accélérer le catalogage.

3. Qualité. Le stockage et la gestion des données redondantes, obsolètes ou triviales (appelées les données « ROT ») coûtent cher. En outre, ce type de données complique la conformité et peut fausser les résultats des projets analytiques. Les données doivent répondre à certains critères de qualité et se maintenir à ces niveaux pour pouvoir être utilisées correctement en aval.

4. Intégration. Les données issues de sources différentes doivent pouvoir être faciles à combiner, à transformer et à partager avec les autres systèmes qui en dépendent. Vous devez donc regrouper les données structurées et non structurées et faire appel à des technologies ouvertes telles qu'Apache Atlas et Hadoop pour les intégrer.

5. Réplication. Un flux de données synchronisé automatiquement avec les données d'origine vous offre la garantie de disposer de données parfaitement à jour dans vos lacs, entrepôts et magasins de données, et dans vos solutions au point d'impact. Face à l'augmentation de la quantité des données, la réplication doit pouvoir gérer de gros volumes avec une faible latence.

6. Gestion des données de référence (MDM). Il est essentiel que l'entreprise dispose d'une vue complète et fiable des entités critiques telles que les clients ou les produits. Les implémentations de MDM modernes proposent une exploration basée sur des graphiques analytiques, un accès agile en libre-service, des fonctions de gouvernance et un tableau de bord convivial.

7. Protection et gestion du cycle de vie. Les informations sensibles et stratégiques doivent être sécurisées et protégées. Le cycle de vie des données doit être géré depuis la création jusqu'à la suppression, en s'appuyant sur les pratiques de gestion des enregistrements, de contestation et de stockage pour archivage.

L'apprentissage automatique accélère la gouvernance

Suite aux récents progrès technologiques, l'apprentissage automatique permet de surmonter les limites des capacités humaines. Il automatise les initiatives de gouvernance et d'intégration à grande échelle, et résout les problèmes posés par les très grands volumes de données. Il permet ainsi de mettre en place une gouvernance des données solide dans toute l'entreprise. Par exemple, si une entreprise dispose d'environ 20 000 termes de données, il faut environ 6 mois

à une équipe de 6 personnes pour classifier manuellement les termes et permettre à l'entreprise de les exploiter pour exécuter une analyse fiable. L'apprentissage automatique peut effectuer cette tâche en quelques jours, voire quelques heures, selon la quantité des actifs de données. Cette extrême rapidité soulage le processus de gouvernance, simplifie les obligations de conformité et ouvre la voie à des projets analytiques productifs.

Une base unique aux usages multiples

Une fois le système UG&I créé pour un projet, il peut servir de tremplin à d'autres cas d'usage à l'échelle de l'entreprise ou de ses entités commerciales :

Lac de données gouverné. De nombreuses entreprises dans des secteurs d'activité très divers ont lancé des projets d'analyse prédictive sur les Big Data. Une première étape fondamentale consiste à déposer de grandes quantités de données structurées et non structurées dans un lac de données. Par le passé, la qualité médiocre des données remettait souvent en cause les implémentations des lacs de données. **Les données non gouvernées** provoquaient la transformation d'un grand nombre de ces lacs en « marécages », du fait de la présence de données redondantes, obsolètes ou triviales souvent impossibles à utiliser dans leur format chiffré d'origine et non-conformes aux réglementations. Dans une implémentation UG&I, les données d'un lac sont mappées avec des termes métier, cohérents, gouvernés et faciles à comprendre pour tous les utilisateurs. Avec à portée de main des données de cette qualité, tous les projets de science des données, d'exploration des données et d'intelligence artificielle sont rentabilisés beaucoup plus vite et fournissent des informations précieuses.

Déchargement des entrepôts de données d'entreprise.

Avec le temps, de nombreuses entreprises ont poussé des jobs ETL (Extraction, Transformation et Chargement) dans leurs entrepôts de données. Il en a résulté un environnement coûteux pour l'exécution des tâches ETL. Par ailleurs, les entrepôts n'ont pas la capacité d'élargir leurs fonctionnalités aux données non structurées ou semi-structurées pour les tâches d'exploration de données.

Le **déchargement des données** stockées dans les entrepôts permet de réduire les coûts, mais aussi de les combiner avec d'autres types de données, dans des environnements tels que les lacs de données qui permettent une exploration dynamique.

Expérience client. Lorsque les représentants du service clientèle interagissent avec les clients, il est important qu'ils disposent d'**informations de référence donnant une vue unique** des clients, des produits ou des autres entités. Des informations erronées ou obsolètes peuvent altérer les interactions avec les clients et éroder la confiance de ces derniers, avec un risque de désertion. Grâce aux données traitées par les processus UG&I, chaque interaction permet d'approfondir la relation avec le client et de lui inspirer davantage confiance, avec à la clé une fidélisation accrue et un meilleur capital de marque.

Conformité

Un système analytique fiable est également un outil qui renforce et accélère la conformité aux obligations réglementaires. Chaque composante a son rôle à jouer lorsqu'il s'agit de gérer les réglementations comme celles du RGPD. Surtout, la protection des données personnelles commence par en donner une définition. L'entreprise peut ainsi déterminer celles qu'elle détient. C'est là que prend son importance le catalogage initial des données, qui comprend toutes les règles de gouvernance en matière de qualité, d'enrichissement et d'analyse des données, et de conformité.

D'autres réglementations, notamment les normes PCI-DSS dans le secteur des services financiers et la législation HIPAA dans le secteur de la santé, sont en vigueur depuis plusieurs années. Les deux ont pour mission de protéger toute atteinte aux informations personnelles. Les mesures de protection qui définissent les droits d'accès en fonction des rôles sont particulièrement importantes, car elles empêchent les données de tomber entre de mauvaises mains.

Conclusion

Alors que les entreprises opèrent leur transformation numérique, leurs dirigeants ont pris conscience des avantages de la gouvernance des données dans leurs environnements, que ces derniers soient sur site ou sur un cloud public ou privé. En donnant la priorité aux sept piliers de l'UG&I (reconnaissance, catalogage, qualité, intégration, réplication, gestion des données de référence, protection/gouvernance du cycle de vie), les entreprises préparent leurs données à l'analyse et à l'extraction de connaissances, mais aussi à la conformité aux réglementations qu'elles doivent respecter. Même si les volumes de données sont considérables, l'apprentissage automatique et l'intelligence artificielle permettent de surmonter les limites des capacités humaines pour certaines tâches : mappage et catalogage de données, mise en concordance de très grands volumes de données et préservation de la qualité des données.

Les chefs d'entreprise visionnaires ont compris qu'ils ont tout intérêt à prendre le temps de créer un système UG&I solide, car ils savent qu'il sera rentabilisé immédiatement et dans un proche avenir. Ils ont su percevoir les multiples avantages dont ils bénéficieront s'ils adoptent la gouvernance comme outil d'optimisation métier, d'innovation et de conformité positive.

Pour plus d'informations, visitez le site :

ibm.com/analytics/unified-governance-integration



© Copyright 2018 International Business Machines Corporation. IBM, le logo IBM et ibm.com sont des marques d'IBM Corporation, enregistrées auprès de nombreuses juridictions dans le monde. Les autres noms de sociétés, 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