



# Injectez une dose de confiance dans votre écosystème de partage des données

Août 2021

## À PROPOS DE COMPTABLES PROFESSIONNELS AGRÉÉS DU CANADA

Comptables professionnels agréés du Canada (CPA Canada) travaille en collaboration avec les ordres de CPA des provinces, des territoires et des Bermudes, et représente la profession comptable canadienne sur les scènes nationale et internationale. La profession canadienne peut ainsi faire la promotion de pratiques exemplaires, favorables aux entreprises et à la société en général, et préparer ses membres aux défis posés par un contexte en évolution constante, marqué par des changements sans précédent. Forte de plus de 220 000 membres, CPA Canada est l'une des plus grandes organisations comptables nationales au monde. [cpacanada.ca](http://cpacanada.ca)

La version électronique de ce document est disponible sur le site [cpacanada.ca](http://cpacanada.ca).

© 2021 Comptables professionnels agréés du Canada

Tous droits réservés. Cette publication est protégée par des droits d'auteur et ne peut être reproduite, stockée dans un système de recherche documentaire ou transmise de quelque manière que ce soit (électroniquement, mécaniquement, par photocopie, enregistrement ou toute autre méthode) sans autorisation écrite préalable.

# Table des matières

<b>Un déficit de confiance croissant</b>	<b>3</b>
<b>Virage numérique</b>	<b>4</b>
<b>Les plateformes infonuagiques : de plus en plus répandues</b>	<b>5</b>
<b>La gestion des nouveaux contrôles de partage des données : le rôle des CPA</b>	<b>10</b>
<b>Les opérations d'apprentissage automatique</b>	<b>12</b>
<b>Que nous réserve l'avenir?</b>	<b>14</b>



Ce document d'information porte sur les écosystèmes de partage des données et les contrôles nécessaires pour l'établissement d'un climat de confiance. Les CPA sont bien placés pour participer à la conception et à la gestion des contrôles, qui amélioreront la clarté et la transparence des données, des plateformes de partage et des solutions d'intelligence artificielle (IA). Les systèmes et les approches évoluent rapidement. Or, il n'existe pas encore de normes ou d'indications détaillées sur le maintien de contrôles appropriés à l'égard des écosystèmes de partage des données. Les CPA possèdent une formation technique et des compétences éprouvées en gestion des contrôles sur l'information financière qu'ils peuvent - et même doivent - appliquer aux chaînes de valeur des données pour apporter une dose de confiance qui fait actuellement défaut.

# Un déficit de confiance croissant

Le principal obstacle auquel la plupart des organisations sont confrontées lorsqu'elles entreprennent leur virage numérique n'est pas de nature technique; il s'agit plutôt de préserver la confiance à l'égard des [chaînes de valeur des données](#). Pour ceux qui sont en amont de la chaîne, là où les données sont générées et collectées, il faut faire preuve d'une grande confiance pour permettre le partage de ses données et leur utilisation par des tiers qui se trouvent à l'extérieur de la sphère de contrôle. Ceux qui sont en aval, comme les experts en science des données, ont quant à eux besoin de savoir que les données qu'ils reçoivent sont exactes, complètes et à jour. Au fur et à mesure que les utilisations possibles se multiplient et que les applications s'élargissent, les écosystèmes se complexifient. Il devient donc nécessaire de faire confiance à un plus grand nombre d'unités fonctionnelles, d'intermédiaires qui collectent des données, ainsi que de scientifiques et d'ingénieurs de données.

Pour que les gens aient confiance en l'écosystème de partage des données, votre organisation se doit d'avoir des systèmes et des contrôles solides en place. Même si vos données sont hébergées à différents endroits – plateformes infonuagiques, lacs de données spécialisés gérés par des tiers, serveurs locaux ou appareils en périphérie –, vous devez veiller à ce que *les bons utilisateurs, c'est-à-dire ceux qui disposent des droits d'accès appropriés, puissent accéder aux données voulues*, et ce, au moment opportun. Rapidement, les organisations doivent composer avec toute une série de défis interreliés afin de mettre en place des algorithmes fiables, d'assurer la conformité à la réglementation sur la protection des renseignements personnels, de veiller à la satisfaction des clients et d'éviter d'onéreuses atteintes à la protection des données.

# Virage numérique

Un écosystème de partage des données bien rodé est un outil essentiel qui aidera votre organisation à prendre le virage numérique. Cet écosystème doit idéalement reposer sur un ensemble de règles de gouvernance des données bien établies. C'est pourquoi l'une des premières choses à faire est d'adopter une [politique de gestion des données](#). Il convient également d'élaborer une [stratégie numérique](#), d'établir un budget et de mettre sur pied une équipe hybride composée de spécialistes, d'experts des données et de professionnels des TI. La clé du succès consiste d'abord à cerner les problèmes à résoudre ou les occasions à saisir par l'organisation. Ensuite, il faut créer et documenter des cas d'utilisation, puis sélectionner les bons modèles d'IA à former. Une fois ces premières étapes franchies, l'organisation devrait tourner son attention vers la [création de données de qualité](#). En effet, il est crucial que les données soient épurées, exactes, complètes et annotées convenablement, ce dont l'organisation peut s'assurer lors de la collecte et de la préparation des données. Une fois que les ensembles de données sont prêts, l'organisation peut les transmettre à son équipe hybride.

# Les plateformes infonuagiques : de plus en plus répandues

Il y a quelques années, lorsqu'une organisation entamait son virage numérique, elle commençait généralement par assembler des ensembles de données convenables, puis procédait à des essais avec l'IA en interne. La gestion de l'accès aux données s'avérait relativement simple, puisque les données étaient habituellement hébergées sur des serveurs locaux. Mais, un an après le début de la pandémie de COVID-19, il est devenu clair que les organisations devaient adopter des modèles hybrides. Beaucoup d'entre elles ont dû avoir recours à de l'aide extérieure pour augmenter leur capacité de mémoire, fournir de l'assistance à un grand nombre d'utilisateurs, mettre à jour leurs modèles d'apprentissage automatique en fonction de la nouvelle situation et appuyer leurs effectifs en ligne. Voilà qui explique le gain de popularité fulgurant qu'ont connu les plateformes en ligne d'apprentissage automatique en 2020. Le taux d'adoption est tel que selon Gartner, un cabinet-conseil de premier plan, les services d'infonuagique publics joueront un rôle essentiel dans 90 % des initiatives liées aux données et à l'analytique d'ici 2022<sup>1</sup>.

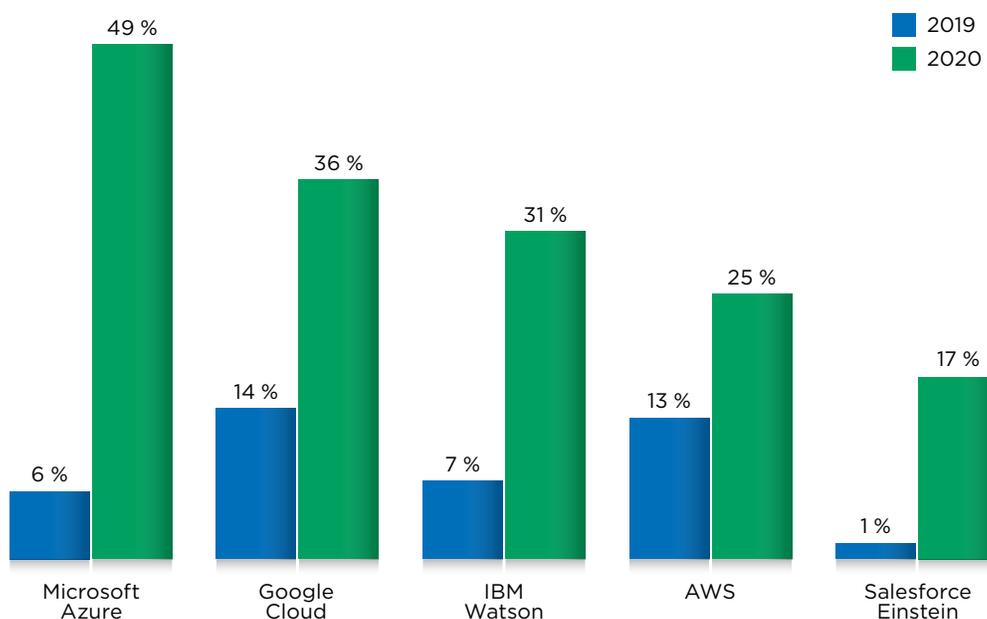
L'utilisation de plateformes infonuagiques présente des avantages manifestes pour les organisations. En effet, une grande partie des outils, des logiciels et du matériel requis pour la préparation des données, la formation des modèles et la mise en œuvre de ces derniers est offerte avec l'abonnement à ces plateformes. Qui plus est, lors de la formation des modèles, les organisations ne paient que le temps de traitement dont elles ont réellement besoin. De plus en plus d'organisations prennent conscience que les plateformes mondiales

---

<sup>1</sup> Gartner (2020). « Top 10 Trends in Data and Analytics for 2020 ». <https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020/>

d'apprentissage automatique sont également plus aptes que les anciens systèmes à répondre à d'autres défis comme la cybersécurité, la modernisation informatique et l'accès aux données<sup>2</sup>.

Figure 1 : Solutions/Plateformes de science des données et d'apprentissage automatique utilisées par les adopteurs précoces



Source : Appen (2020). *The State of AI and Machine Learning*. <https://resources.appen.com/wp-content/uploads/2020/06/Whitepaper-State-of-Ai-2020-Final.pdf>

Gartner a récemment évalué 20 plateformes d'apprentissage automatique qui permettent de collecter des données, de créer des modèles et d'opérationnaliser l'apprentissage automatique. Ces plateformes en ligne permettent aux organisations d'exécuter toutes les opérations nécessaires et d'adapter les services en fonction de leurs besoins. Il est ainsi possible aux organisations de stocker, de préparer, d'annoter et d'étiqueter des données brutes, puis de sélectionner des algorithmes adaptés, notamment grâce à l'intelligence artificielle explicable. Ces plateformes offrent aux organisations l'énorme capacité de traitement requise pour former les modèles d'apprentissage automatique à partir de leurs données. Elles peuvent même leur donner accès à des ensembles de données supplémentaires et à du

<sup>2</sup> Karthik Ramachandran et David Linthicum, Deloitte (2020). « Why organizations are moving to the cloud ». <https://www2.deloitte.com/us/en/insights/industry/technology/why-organizations-are-moving-to-the-cloud.html>

soutien pour la mise en œuvre, la surveillance et la maintenance des modèles<sup>3</sup>. Comme le montre le graphique ci-dessous, les chefs de file du domaine comprennent SAS, IBM, Dataiku, MathWorks, TIBCO Software et Databricks.

Figure 2 : « Quadrant magique » des plateformes de science des données et d'apprentissage automatique



Source : Gartner (mars 2021). « Magic Quadrant for Data Science and Machine Learning Platforms ».

<sup>3</sup> Gartner (2021). « Magic Quadrant for Data Science and Machine Learning Platforms ». [https://content.dataiku.com/nurturing-why-ai-platforms-lob/gartner-mq-21?utm\\_campaign=CONTENT+Nurturing+Workflows+June+2020&utm\\_medium=email&\\_hsmi=90911789&\\_hsenc=p2ANqtz-8PKZwucx5qQZoOZvnCh8uiFqEL3U-\\_IqFJiH\\_m8Icw-M7V2x1dT50Ujgzl-ixPZKQ7SzaCj7vYLS2u4\\_fK8iBXMJJT3A&utm\\_content=90911789&utm\\_source=hs\\_automation](https://content.dataiku.com/nurturing-why-ai-platforms-lob/gartner-mq-21?utm_campaign=CONTENT+Nurturing+Workflows+June+2020&utm_medium=email&_hsmi=90911789&_hsenc=p2ANqtz-8PKZwucx5qQZoOZvnCh8uiFqEL3U-_IqFJiH_m8Icw-M7V2x1dT50Ujgzl-ixPZKQ7SzaCj7vYLS2u4_fK8iBXMJJT3A&utm_content=90911789&utm_source=hs_automation)

Fait intéressant, les organisations ne mettent pas toutes leurs données dans le même panier. La plupart d'entre elles disent plutôt faire appel à plusieurs fournisseurs lorsqu'elles étendent le déploiement des solutions d'IA et d'apprentissage automatique à différentes unités fonctionnelles. Les clients veulent demeurer indépendants et éviter d'être « pris » avec un seul prestataire, préférant avoir la liberté de choisir parmi tous les grands fournisseurs d'infonuagique<sup>4</sup>. Les organisations se servent ensuite de logiciels exclusifs pour faire le pont entre les ensembles de données, les applications infonuagiques, les applications gérées par des tiers spécialisés et les centres de traitement de l'information privés<sup>5</sup>.

Même si les fournisseurs proposent des tâches et des processus préétablis en lien avec l'accès aux données et le contrôle des données (comme des modèles de données, de gouvernance et de collaboration), il revient aux organisations d'élaborer et de mettre en œuvre des politiques en la matière, puis d'en assurer le suivi.

Selon Tari Dwiek, directeur des alliances technologiques à Snowflake, une société spécialisée dans le domaine de la science des données et de l'IA, les organisations désirent avoir un accès aux données sûr et encadré. « Comme l'infonuagique nous donne la possibilité de gérer des données à grande échelle, les clients sont conscients à la fois de la nécessité de mettre en place une gouvernance des données de grande ampleur et de l'occasion que cela représente<sup>6</sup>. »

Du côté des organisations qui font un usage de plus en plus poussé de l'IA et de l'apprentissage automatique, la tendance est à la généralisation de l'utilisation des données. Par conséquent, un grand nombre d'analystes et de spécialistes ont soudainement besoin d'accéder aux données des organisations. On voit donc l'apparition de nouveaux services de gestion des données faisant appel à l'architecture de lac de données – un dépôt de stockage dans lequel il est possible de conserver des données brutes n'ayant subi aucune préparation. Une telle infrastructure permet le stockage des données opérationnelles, de même que leur utilisation et leur analyse en temps réel, sans qu'il soit nécessaire de les copier, de les étiqueter ou de les annoter<sup>7</sup>.

---

4 Dataiku (2021). *Getting the Most of AI in 2021*, p. 14. [Traduction libre]

5 Dataiku (2020). « 2021 Trends: Where Enterprise AI is Headed Next ». <https://www.dataiku.com/stories/2021-trends-where-enterprise-ai-is-headed-next/>

6 Dataiku (2021). *Getting the Most of AI in 2021*, p. 14. [Traduction libre]

7 Dremio (2021). *The Next-Generation Cloud Data Lake: An Open, No-Copy Data Architecture*. <https://hello.dremio.com/wp-the-next-generation-cloud-data-lake.html>

Le partage de données se fait aussi de plus en plus entre organisations. Les protocoles de partage des données et les exigences en matière de protection doivent alors aussi s'appliquer aux fournisseurs. Une révision des contrats en vigueur s'impose donc pour veiller à ce que leurs conditions reflètent adéquatement les [politiques de gestion des données](#), les procédures et les protocoles de l'organisation. Les questions de propriété des données, de propriété intellectuelle, de droit d'auteur, de conservation des données et de respect des règles de protection de la vie privée et d'éthique applicables doivent aussi être prises en considération.

# La gestion des nouveaux contrôles de partage des données : le rôle des CPA

Comme certaines organisations adoptent une approche hybride qui fait appel à plusieurs fournisseurs, sources de données, modèles et formats de sortie, la gestion de l'accès aux données et la présentation d'information sur la conformité sont devenues des questions cruciales. Afin d'établir le climat de confiance nécessaire pour inciter les diverses unités fonctionnelles à partager des données de qualité qui pourront être réutilisées, les organisations doivent mettre en place un cadre redditionnel crédible. Dans le contexte d'un virage numérique, c'est le contrôleur des données qui est responsable de la garde des données partagées aux fins de réutilisation et qui doit accroître leur valeur grâce à leur protection, à leur curation et à leur bonne utilisation.

Habituellement, les contrôleurs protègent les ressources et s'assurent que seules les personnes disposant des droits d'accès appropriés (en fonction des besoins et des permissions) sont autorisées à les utiliser. Traditionnellement, les professionnels comptables assument des rôles de garde à l'égard des ressources financières et physiques. À ce titre, ils veillent à ce que les ressources financières de l'organisation soient protégées, à ce que les lois et les exigences connexes soient respectées et à ce que les activités entreprises par l'organisation soient stratégiquement cohérentes.

Le rôle de garde du contrôleur des données ne se limite pas aux ressources financières. Il s'applique à toutes les ressources de données et chevauche dans une certaine mesure le rôle du contrôleur financier. Il est dans l'ordre des choses que les professionnels comptables élargissent leur rôle traditionnel de gérance financière afin d'y inclure toutes les données destinées à être partagées et réutilisées. Toutefois, garde

ne signifie pas propriété. La propriété des données non financières appartient souvent à des fonctions opérationnelles au sein d'une organisation. La garde est une fonction habilitante qui vise à s'assurer que les propriétaires des données les protègent, en font la curation, les partagent et les utilisent dans le respect des contraintes externes (lois, règlements, etc.) et internes (politiques).

À mesure que des marchés sont créés pour l'échange de données entre acheteurs et vendeurs, on s'attend à ce que des exigences officielles soient élaborées en vue de la certification des données vendues, partagées ou négociées. La certification nécessitera probablement la capacité de prouver l'exactitude et la source des données, autrement dit leur provenance ou leur traçabilité. Un contrôleur des données serait certainement appelé à fournir cette certification.

Les organisations ont de nombreuses occasions de mettre à profit les nouvelles technologies, ce qui peut présenter des défis considérables du point de vue juridique et éthique, car les utilisations de la technologie ne cadrent pas toutes avec les valeurs de sociétés différentes (équité, sécurité, protection des renseignements personnels, compréhensibilité et transparence). Comme il est indiqué dans *l'International Code of Ethics for Professional Accountants*, on attend des professionnels comptables que, compte tenu de leur poste et de leur autorité au sein de l'organisation, ils y encouragent et promeuvent une culture axée sur l'éthique. De ce fait, ils sont en bonne position pour aider les organisations en exerçant le rôle de contrôleur des données.

La fonction clé de ce rôle est la protection des données, ce qui implique de s'assurer que seules les personnes autorisées ont accès aux données, qu'il est possible de démontrer la provenance des données et que les données sont uniquement utilisées aux fins prévues.

Les professionnels comptables ont un rôle élargi à jouer, qui est de s'assurer que les limites de compétence sont respectées en effectuant une surveillance appropriée de l'utilisation des données. Un rôle tout aussi important est celui de certifier que les données utilisées ou vendues sont adaptées à l'usage prévu, en ce sens que la provenance et la traçabilité des données peuvent être prouvées<sup>8</sup>.

---

<sup>8</sup> CPA Canada et IFAC (2021). *Le rôle du professionnel comptable dans la gestion des données*, document de travail, 44 pages.

# Les opérations d'apprentissage automatique

Au fur et à mesure que les organisations généralisent l'utilisation des données, la gestion de la performance d'un écosystème de partage des données et la présentation d'information sur la conformité se complexifient. Une fois que les processus de transformation numérique ont été systématiquement intégrés aux activités courantes, le modèle de la petite équipe hybride peut être appelé à évoluer et à prendre de l'expansion pour permettre la gestion des opérations d'apprentissage automatique (*Machine Learning Operations* ou MLOps). Cette nouvelle discipline a vu le jour vers la fin de 2018; elle vise à aider les organisations à passer au niveau supérieur (c'est-à-dire à passer de la gestion d'un seul modèle d'apprentissage automatique à la gestion de centaines ou de milliers d'entre eux) et à gérer la complexité et les risques découlant de l'utilisation de modèles d'apprentissage automatique. Les opérations d'apprentissage automatique font appel à l'expertise de spécialistes, de scientifiques de données, d'ingénieurs de données, d'ingénieurs en logiciels, d'experts en développement et exploitation, d'architectes en apprentissage automatique, de responsables de la gestion des risques liés aux modèles et d'auditeurs. Cette approche a pour but de standardiser et de rationaliser la gestion du cycle d'apprentissage automatique.

Dans un livre sur les opérations d'apprentissage automatique publié récemment par O'Reilly Media, Mark Treveil de Dataiku soutient que des processus et des contrôles de gouvernance des données rigoureux sont essentiels à une transition réussie vers le modèle d'opérations d'apprentissage automatique. À chaque projet, il convient de consigner l'ensemble des décisions et des résultats, d'en faire le suivi et de mettre en place des contrôles pour pouvoir répondre aux questions suivantes :

- D'où proviennent les données?
- Comment et selon quelles conditions d'utilisation les données initiales ont-elles été collectées?

- Les données sont-elles exactes et à jour?
- Les données contiennent-elles des renseignements personnels ou d'autres informations sensibles qui ne devraient pas être utilisées<sup>9</sup>?

Les opérations d'apprentissage automatique permettent aux organisations d'« avoir une vision d'ensemble et de savoir quelles équipes utilisent quelles données, comment elles le font et quels modèles elles appliquent ». Pour ce faire, « les organisations doivent aussi savoir que leurs données sont fiables et qu'elles sont collectées conformément à la réglementation en vigueur, en plus d'avoir une compréhension centralisée des modèles qui sont utilisés dans leurs différents processus »<sup>10</sup>. Cela dit, il n'existe actuellement aucune méthode établie de surveillance de la conformité relativement à la gouvernance des données et à la présentation d'information à ce sujet. Certaines sociétés spécialisées ont toutefois lancé de nouvelles plateformes et de nouveaux logiciels de gestion de l'accès aux données. Par exemple, Immuta et Databricks offrent maintenant des outils qui permettent d'automatiser et de contrôler l'accès aux données, de manière à soutenir les équipes d'opérations d'apprentissage automatique. Ces outils s'appuient sur des processus comme les contrôles d'accès basés sur les attributs (*Attribute-Based Access Controls* ou ABAC), les contrôles automatisés de sécurité et de protection des renseignements personnels comme l'anonymisation des données ou le repérage et l'étiquetage des données sensibles, ainsi que la vérification de l'accès aux données et la présentation d'information à ce sujet<sup>11</sup>.

En mettant en place des mécanismes de gouvernance adéquats, une organisation peut veiller au respect de ses obligations envers l'ensemble de ses parties prenantes et au respect des principes fondamentaux d'équité.

---

9 Mark Treveil, en collaboration avec l'équipe de Dataiku (2020). *Introducing MLOps: How to Scale Machine Learning in the Enterprise*, O'Reilly, 169 pages.

10 *Ibid.*, p. 10. [Traduction libre]

11 Immuta et Databricks (2020). *A Guide to Data Access Governance*. <https://www.immuta.com/downloads/a-guide-to-data-access-governance-with-immuta-databricks/>

# Que nous réserve l'avenir?

Le présent document d'information soulève la nécessité d'établir des contrôles afin d'injecter une dose de confiance qui fait actuellement défaut dans les écosystèmes de partage des données. Dans notre monde de plus en plus numérisé, il existe un besoin de confiance plus important que jamais. Lorsqu'il est question de rapports financiers, de surveillance et de contrôles, les CPA sont la référence en matière de confiance. La profession peut donc donner le ton dans la discussion sur la confiance à l'égard des chaînes de valeur des données – on parle ici de la collecte des données, de leur partage et de l'accès à celles-ci, et de la confiance envers des outils qui font appel à l'IA et à l'apprentissage automatique. Les CPA peuvent contribuer à démontrer que les nouveaux systèmes fonctionnent comme il se doit en assurant la surveillance, la vérification, l'audit et la présentation d'information, ainsi qu'en veillant à la conformité aux meilleures pratiques de partage des données, d'accès aux données et de réutilisation des données dans les activités des organisations, qui sont menées dans différentes régions et qui peuvent prendre une multitude de formes.



**CPA**

COMPTABLES  
PROFESSIONNELS  
AGRÉÉS  
CANADA

277, RUE WELLINGTON OUEST  
TORONTO (ONTARIO) M5V 3H2  
CANADA  
TÉL. : 416 977.3222 TÉLÉC. 416 977.8585  
CPACANADA.CA