CPA COMPTABLES PROFESSIONNELS AGRÉÉS CANADA

Mégadonnées et veille stratégique

TENDANCE TECHNOLOGIQUE

« La révolution des mégadonnées ne réside pas dans les données elles-mêmes. Mais dans l'analyse que nous pouvons en faire, et dans l'interprétation que nous devons en faire. »

— Gary King, professeur et directeur de l'Institute for Quantitative Social Science, Université Harvard

Afin de prendre des décisions d'affaires éclairées, les entreprises doivent rassembler et analyser des données. Avec l'aide d'outils et de méthodes de veille stratégique, elles peuvent maintenant procéder à l'analyse rapide et à moindre coût de grands volumes de données.

Pour relever les défis que représente une telle analyse, les entreprises doivent déterminer le type de données qu'elles détiennent et la façon dont ces données peuvent être efficacement stockées et consultées ultérieurement. D'ailleurs, en raison du recours croissant aux données dans la prise de décisions d'affaires, l'intégrité des données devient un aspect de plus en plus important à considérer. Il faut tenir compte de certaines méthodes, comme les plans de classification des données, les taxonomies et l'utilisation de métadonnées. Plus la quantité de données augmente, plus les besoins de stockage et les coûts afférents augmentent eux aussi.

Description

Mégadonnées

Le terme « mégadonnées » ne décrit pas seulement un « grand nombre de données ». Les mégadonnées englobent les outils (comme Hadoop et Cassandra) qui permettent le traitement à haute vélocité d'un volume élevé d'informations d'une grande variété, de manière à améliorer la compréhension et à éclairer la prise de décisions. Selon IBM, les mégadonnées se caractérisent par quatre éléments : le volume, la vélocité, la variété et la véracité.

Les entreprises sont confrontées au paradoxe de détenir une grande quantité de données, mais peu d'informations. Elles doivent donc trouver des moyens plus efficaces de dégager la valeur informative de ces données, de les stocker, de les archiver, de les gérer et de les récupérer. Puisque la quantité de données recueillies croît continuellement, les organisations veulent extraire plus efficacement et plus rapidement les informations leur permettant de prendre des décisions toujours plus complexes.

Veille stratégique

La veille stratégique (VS) désigne les processus, les outils et les techniques conçus pour tirer de l'information de grands volumes de données structurées au sein d'une organisation. Aujourd'hui, la VS profite de la convivialité grandissante de ses technologies et du fait que l'analyse migre hors du service central des TI vers les secteurs d'activités générateurs de revenus. Les détaillants peuvent utiliser les outils de VS pour évaluer comment le placement de produits fera augmenter les ventes. Par exemple, si l'épicier place la salsa tout près des tortillas plutôt qu'avec les autres sauces, le nombre de ventes pourrait croître.

Importance

Pour bénéficier pleinement des nouvelles sources d'information, il faut mettre au point de nouvelles techniques. La croissance de l'information non structurée (vidéos, billets de blogues, gazouillis, relevés de capteurs d'objets connectés à Internet, etc.) surpasse déjà celle des sources traditionnelles de données transactionnelles. Les mégadonnées s'inscrivent dans la tendance plus générale consistant à fonder la prise de décisions sur les données. Bien que les organisations soient très au fait de la nécessité d'avoir recours à *un minimum* d'analyse, beaucoup d'entre elles rencontrent des défis, comme le manque d'expertise interne. Par conséquent, dans le cadre de leur rôle, les CPA doivent considérer l'applicabilité de l'analyse et la valeur potentielle d'investir dans celle-ci. Comme mentionné précédemment, la proposition de valeur ultime (c'est-à-dire une meilleure connaissance des clients, des produits, des services et autres) est cohérente, peu importe qu'il s'agisse d'analyses de mégadonnées ou de veille stratégique.

Avantages et considérations pour les entreprises

Les mégadonnées peuvent fournir une immense valeur et présenter d'énormes avantages pour les organisations. Voici certains de ces avantages :

- Efficacité accrue des promotions et des initiatives d'affaires;
- Meilleure compréhension du comportement des clients et de la situation du marché;
- Compréhension plus rapide;
- Économies de coûts.

Tesco, un détaillant britannique de marchandises générales et d'épicerie, lie la gestion de sa chaîne d'approvisionnement à la météo, à la date et à des données géographiques de manière à cerner et à prévoir les tendances. Par exemple, grâce à l'analyse de mégadonnées, l'entreprise a pu établir qu'« un samedi ensoleillé de la fin d'avril où il fait 16 degrés se traduit par un pic de ventes. Toutefois, si les mêmes conditions se reproduisent quelques semaines plus tard, elles ne produiront pas le même effet, puisque les gens auront déjà eu leur premier barbecue de la saison. » Sur le plan des avantages globaux pour l'entreprise, celle-ci a constaté que « les projets sur les mégadonnées procurent d'énormes avantages à Tesco, car ils permettent notamment d'améliorer les promotions pour s'assurer qu'il y a 30 % de moins de produits manquants sur les étagères, d'établir des prévisions sur le comportement des consommateurs selon la météo de façon à réduire le gaspillage d'aliments, chiffré à 6 M£ au cours de l'été, de réduire les stocks en entrepôt de 50 M£ et, ce faisant, d'optimiser les activités des magasins pour aboutir à une diminution du gaspillage de 30 M£[1]. »

Bien que les mégadonnées et la VS peuvent apporter leur lot d'avantages à une organisation, il y a certains secteurs de risque importants à prendre en considération :

Secteurs de risque	Stratégies d'atténuation des risques
Les gestionnaires de risques et autres contrôleurs au sein de l'organisation peuvent choisir de ne pas utiliser les mégadonnées en raison des incertitudes et de la crainte générale qu'inspirent les technologies, ou d'autres types de doutes.	Lorsqu'on tente de cerner les nouveaux risques nets associés aux mégadonnées, il est important de comprendre la façon dont les technologies sous-jacentes diffèrent des processus de contrôle habituels propres aux systèmes transactionnels. Un bon point de départ en la matière : prendre connaissance des dix plus grands défis établis par la Cloud Security Alliance en ce qui a trait à la sécurité des données et à la protection des renseignements personnels[2]. Par exemple, on y voit comment la technologie déconstruit un ensemble de données pour ensuite en traiter chaque « bloc » individuellement. On y souligne aussi que, si le processus de lecture du bloc de données n'est pas autorisé, il y a un risque que l'analyse produite soit incorrecte. Ceux qui connaissent les contrôles généraux des TI[3] remarqueront qu'il s'agit d'un contrôle du développement d'applications pour lequel le code sous-jacent doit être testé et autorisé.

1 http://cloudofdata.com/2012/10/tesco-uses-data-for-more-than-just-loyalty-cards/

2 https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Top_Ten_v1.pdf

3 Le terme équivalent en anglais, *IT General Controls*, est celui employé par le PCAOB dans les tests relatifs à la Loi Sarbanes-Oxley.

Secteurs de risque	Stratégies d'atténuation des risques
Les contrôles à l'égard de la protection des renseignements personnels qui s'appliquaient aux technologies de bases de données traditionnelles pourraient être mal adaptés à un environnement de mégadonnées.	Pour se conformer aux règles concernant la protection des renseignements personnels dans un environnement de mégadonnées, il faut comprendre en quoi le profil de risque a changé par rapport aux données qui sont maintenant utilisées. Par exemple, les analyses de mégadonnées peuvent comprendre des photos. Même si la technologie de reconnaissance faciale n'en est encore qu'à ses débuts, certaines entreprises, comme iOmniscient, font savoir depuis quelques années que leur technologie peut être combinée aux analyses de mégadonnées[4]. Dans un tel contexte, les organisations doivent évaluer ce qui constitue un préavis et un consentement suffisants pour se conformer aux règles de protection des renseignements personnels, comme la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE). Par exemple, des experts en protection des renseignements personnels doivent déterminer ce qui constitue un préavis suffisant pour un enregistrement vidéo. Dans pareil cas, une pancarte informant les clients qu'ils sont filmés et que l'enregistrement peut être utilisé à des fins d'analyse pourrait ou non être un « préavis suffisant ». Puisque Google offre déjà des services de recherche par image, il n'est pas trop exagéré d'imaginer que des caméras en magasin puissent être connectées au moteur de recherche de Google Images afin d'identifier les clients sur place, puis de fournir cette information aux vendeurs. La Federal Trade Commission (FTC) a rédigé un document de pratiques exemplaires qui explore la question de la protection des renseignements personnels dans le recours à la reconnaissance faciale. Dans ce document, l'on cite le cas de Kraft Foods qui compte utiliser cette technologie dans les supermarchés[5]. D'autre part, les organisations doivent aussi évaluer la sensibilité de leur clientèle au sujet de la protection des renseignements personnels. Par exemple, Benetton a dû essuyer la colère de ses clients pour avoir *envisagé* de recourir à des étiquettes IRF pour faire le suivi de ses stocks. Certains d'entre eux ont craint que cette technologie porte atteinte à leur vie privée[6].

4 Consultez le document http://www.iomniscient.com/Media/PR/GIT2012-Face_Recognition_in_a_Crowd.pdf. L'entreprise explique comment la reconnaissance faciale peut être mise en corrélation avec le temps écoulé pour déterminer à quelle vitesse progressent les files d'attente.

5 www.ftc.gov/opa/2012/10/facialrecognition.shtm

6 www.rfidjournal.com/articles/view?471

Secteurs de risque	Stratégies d'atténuation des risques
Les données utilisées dans les modèles d'analyse de mégadonnées ne sont pas adaptées à l'objectif ou contiennent de grosses erreurs qui mèneraient à la prise de mauvaises décisions.	Sans bons contrôles de la qualité des données, il y a risque de présence de « données sales » pouvant entraîner de piètres analyses dont les résultats comporteraient, ultimement, des « anomalies significatives ». Par exemple, une entreprise minière effectuant des analyses à partir de données de mauvaise qualité pourrait construire une plateforme pétrolière sur un puits sec plutôt que sur un puits productif, entraînant ainsi le gaspillage de millions d'euros. Par conséquent, il est important de vérifier l'intégrité des données disponibles avant de commencer à discuter de solutions liées aux mégadonnées. Si l'on souhaite veiller à l'intégrité de l'information – sujet approfondi dans la publication intitulée Cadre de contrôle de l'intégrité de l'information –, il est nécessaire d'établir une approche multidomaine en explorant les contrôles s'appliquant aux domaines du contenu (comme l'exactitude des métadonnées), du traitement (comme l'exactitude de la manipulation de l'information par la logique des programmes sous-jacents) et de l'environnement des systèmes d'information (comme l'accès logique à l'information). Un aspect important de la préparation d'un exercice lié aux mégadonnées est de veiller à ce que les données soient exemptes d'erreurs et adaptées à l'objectif. Les activités de nettoyage de base devraient permettre de s'assurer que les champs ne comportent aucune donnée qui ne devrait pas s'y trouver (comme un État ou une province invalide, des caractères alphabétiques dans un champ numérique ou un code postal invalide). Toutefois, veiller à ce que les données soient adaptées à l'objectif peut s'avérer plus ardu. Par exemple, une enquête de ProPublica a permis de révéler qu'un logiciel utilisé pour évaluer le risque de récidive criminelle comportait un biais raciste[7]. En d'autres termes, un racisme systémique avait été programmé à même le logiciel. Dans une telle situation, il faut prendre soin de n'utiliser que des données non biaisées dans le modèle prédictif basé sur les mégadonnées.

7 www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Conclusion

Étant donné la diminution des coûts de stockage et la popularité croissante des objets connectés, les données foisonnent. Mais pour pouvoir en extraire toute la valeur, les entreprises doivent nettoyer, analyser et interpréter ces données. En ayant recours aux outils de VS et en étant conscientes des secteurs de risque associés aux mégadonnées, elles peuvent commencer à appliquer les bonnes analyses à leurs données afin d'en tirer des informations précieuses.

La présente publication s'inscrit dans la **série Tendance technologique**, qui porte sur les grandes tendances du domaine touchant le milieu comptable. Les documents de cette série sont disponibles sur notre site Web.