

# Les entreprises data-driven

Par Mark Schwartz, Stratège d'entreprise chez AWS





## Introduction

À l'heure actuelle, nous entendons beaucoup parler des entreprises guidées par les données ou, data-driven, et de la nécessité d'en devenir une. Mais que signifie exactement être « data-driven », et pourquoi est-ce si important dans l'environnement numérique actuel ? Quelles mesures concrètes peuvent prendre les entreprises pour mettre en place un état d'esprit et des pratiques basés sur les données ? Et quel est le lien entre les données et l'agilité opérationnelle et technique, cette dernière étant l'autre priorité de l'ère numérique ? Dans cet e-Book, nous allons voir ce que signifie être orienté sur les données et aborder quelques exemples d'entreprises qui utilisent les données pour gérer leurs activités. Nous allons également faire le lien entre l'orientation sur les données et l'agilité, la transformation numérique et l'innovation continue.

Les entreprises orientées sur les données cherchent à **baser leurs décisions métier stratégiques** sur les preuves fournies par les données (ce qui requiert une certaine rigueur et, en même temps, une capacité à innover reposant sur l'identification, au sein des données, d'opportunités pouvant conduire à de nouveaux produits ou marchés). Ces entreprises exploitent également les données comme un actif, pour améliorer leurs interactions avec les clients et augmenter leur efficacité. En d'autres termes, elles **analysent** les données pour éclairer la prise de décision et les **utilisent** pour servir leurs clients. Les données peuvent constituer la base, par exemple, de la personnalisation, de la tarification dynamique, de l'expansion du marché, de l'innovation produit ou de l'optimisation de la chaîne d'approvisionnement.

Cependant, jusqu'à récemment, les entreprises ne parvenaient pas à utiliser les données de cette façon, car elles les envisageaient uniquement dans le cadre des transactions. Résultat : elles les enfermaient dans des bases de données cloisonnées qui convenaient parfaitement au traitement des transactions, mais qui étaient moins adaptées à l'analyse ouverte. Notre modèle mental était celui du formulaire de facture ou de bon de commande : « veuillez me donner 20 widgets au prix de 100 USD par widget ». Ou : « veuillez me payer 20 widgets au prix de 100 USD par widget ». Les données étaient performatives et impératives, autrement dit, un stimulant ou un artefact pour effectuer une transaction. Aujourd'hui, la valeur des données dépasse largement le rôle transactionnel.

Comment envisager cette valeur en termes financiers et comment l'optimiser ?

## La valeur métier des données

Chaque donnée peut être utilisée dans autant d'analyses engendrant des résultats métier que nécessaire. Ces données deviennent ainsi précieuses car elles permettent d'obtenir ces résultats. Par exemple, si l'entreprise analyse l'historique de ses transactions, trouve comment optimiser sa chaîne d'approvisionnement et réduit ainsi ses coûts, les données auront donc participé à cette réduction de coûts. Par conséquent, les données ont une valeur commerciale qui découle de leur utilisation potentielle dans l'augmentation des bénéfices ou dans l'atteinte des objectifs de missions.

Les exemples d'utilisation de données pour leur valeur non transactionnelle ne manquent pas. L'entreprise Johnson & Johnson, par exemple, utilise les données transactionnelles stockées dans le cloud pour améliorer le respect de conformité réglementaire des médecins, optimiser sa chaîne d'approvisionnement et découvrir de nouveaux médicaments. Nike collecte des données sur les performances des clients pour améliorer leur expérience numérique sur NikePlus. Lyft collecte et stocke les coordonnées GPS de tous ses trajets, qui après analyse, ont révélé que 90 % des trajets coïncidaient partiellement à d'autres trajets à proximité. Cette information a permis la création de Lyft Line, un service permettant aux passagers de partager une voiture et de recevoir des remises allant jusqu'à 50 %.<sup>1</sup>

Ces utilisations pouvant donner lieu ultérieurement à des bénéfices (même si ces bénéfices n'ont pas encore été réalisés), nous pouvons envisager les données comme un actif financier (bien que non-GAAP dans la plupart des cas). C'est donc sans surprise que les données cumulées par une entreprise peuvent jouer un rôle dans la valeur d'acquisition de l'entreprise ou lui permettre de nouer des partenariats avec d'autres entreprises. C'est le cas, par exemple, de l'acquisition par Microsoft de LinkedIn, avec ses données sur 433 millions de clients, pour 26,2 milliards d'USD, ou encore de la procédure de faillite de Caesars Entertainment Operating Corp. Inc. entre 2015 et 2017, lorsque les créanciers soutenaient que les données concernant les 45 millions de clients de son programme de fidélité client Total Rewards valaient 1 milliard d'USD et représentaient son actif le plus précieux.<sup>2</sup>

Il peut être utile de voir les données comme une valeur commerciale qui prend la forme d'option d'achat financier (c'est-à-dire qu'elles nous permettent d'apporter des changements à la chaîne d'approvisionnement ou de lancer un nouveau produit, sans toutefois nous obliger à le faire). Nous pouvons choisir d'exercer l'option ou non, en fonction de la valeur que les données prévoient pour cette option. C'est là que la valorisation des actifs de données pose des difficultés : évaluer une option d'achat est bien plus compliqué que de calculer le ROI d'un flux prévisionnel de mouvements de trésorerie. C'est la raison pour laquelle les entreprises négligent souvent cette valeur, mais comme je l'explique dans mon livre *War and Peace and IT*<sup>3</sup>, de nombreuses techniques de prestation informatique agile conduisent à ce type de valeur d'option.

<sup>1</sup> Études de cas AWS. Voir <https://www.youtube.com/watch?v=6A1tOFqvgek>, <https://aws.amazon.com/products/databases/> et <https://aws.amazon.com/solutions/case-studies/lyft/>.

<sup>2</sup> Ces deux exemples sont issus de <https://sloanreview.mit.edu/article/whats-your-data-worth/>. Une analyse détaillée de la faillite de Caesars est disponible sur <https://turnaround.org/sites/default/files/11.%20Paper%20-Caesars.pdf>. La faillite était extrêmement complexe et la valeur de Total Rewards était intégrée à d'autres actifs, tant et si bien qu'il est impossible de connaître avec précision la valeur qui lui a finalement été attribuée.

<sup>3</sup> Mark Schwartz, *War and Peace and IT : Business Leadership, Technology, and Success in the Digital Age* (Portland, OR : IT Revolution Press, 2019).



## Données et agilité

La valeur n'est pas seulement créée par les données en soi, mais également par les outils et les processus mis en place pour les analyser et produire des résultats opérationnels. Dans le monde numérique d'aujourd'hui, caractérisé par le changement rapide, l'incertitude et la complexité (voire par le bouleversement), il est nécessaire d'utiliser les données pour favoriser l'agilité opérationnelle et répondre rapidement et avec flexibilité aux situations changeantes. L'agilité est ce qui permet aux entreprises de transformer le changement rapide en opportunité et d'éviter tout bouleversement en réagissant promptement aux menaces concurrentielles. À l'ère du numérique, les entreprises ont compris qu'il leur fallait des versions anticipées de produits à commercialiser rapidement et à faire évoluer en fonction de la réaction continue du marché.<sup>4</sup>

Les dernières années ont apporté des techniques de construction d'agilité dans le **processus** de développement produit, y compris, par exemple, le développement de logiciels en modes Agile, DevOps et Lean. Le cloud est utilisé pour accélérer la livraison des capacités informatiques, à la fois en matière de logiciel et de matériel. Les structures organisationnelles basées sur les équipes ont permis la mobilisation de ressources pour répondre aux besoins changeants. Tous ces développements ont permis aux entreprises de rendre leurs processus plus agiles.

Cependant, les processus agiles ne sont qu'une partie du problème : les données de l'entreprise doivent elles aussi être agiles. Elles doivent être facilement disponibles pour des utilisations inattendues et en constante évolution. Elles doivent être accessibles et utiles. Il faut mettre des outils entre les mains des employés afin qu'ils puissent travailler sur les données, ainsi que des compétences pour le faire. La capacité à utiliser les données de manière flexible (pour rendre possibles de nouvelles utilisations encore inexplorées) est le maillon manquant pour obtenir une agilité de l'entreprise et démarquer l'entreprise agile d'une autre ayant seulement adopté les cadres et les attributs des modèles agiles. L'agilité opérationnelle nécessite une agilité des données. Les entreprises orientées sur les données maîtrisent les deux.

Cet accent sur l'agilité conférée aux données est une nouveauté. Lorsque les données n'étaient que transactionnelles, nous pouvions les enfermer dans des bases de données hautement structurées afin de refléter leur utilisation pour les transactions. Nous utilisions des systèmes de bases de données relationnelles comme outils, par exemple Oracle ou SQL Server, dont les atouts résident dans le traitement transactionnel. Nous utilisons les données pour effectuer les transactions et pour produire des rapports opérationnels et ainsi appuyer les transactions.

Dans la mesure où nous prêtons attention à la confidentialité, nous les appliquons en limitant strictement leur accès plutôt qu'en cherchant des moyens de les rendre disponibles dans les limites du respect de la vie privée. Nous pratiquons une sorte de « confidentialité par l'anonymat » plutôt qu'une « protection de la vie privée dès la conception ».

Certes, il y a eu des tentatives de libération des données pour une analyse ad hoc avec des systèmes d'« informatique décisionnelle » (BI). Cependant, les outils que nous avons aujourd'hui sont bien plus avancés que les systèmes de BI : nous avons maintenant le machine learning, une série de bases de données sur mesure visant à gérer différents types de données, des algorithmes pour des traitements hautement parallèles, de gros volumes de données non structurées comme les vidéos et les discours, des périphériques IoT fournissant des flux de données grâce à des capteurs, et... simplement de très gros volumes de données. Avec ces outils, nous pouvons libérer nos données de leur contexte transactionnel et opérationnel.

Et surtout, nous avons compris que se concentrer sur les données ne représente des difficultés non seulement techniques mais également organisationnelles. **Pour être orientée sur les données, une entreprise doit repenser sa prise de décisions et ses interactions avec les clients. Il s'agit d'un engagement à l'égard de la valeur des données, une sorte d'humilité organisationnelle selon laquelle « les données savent mieux que nous ».**

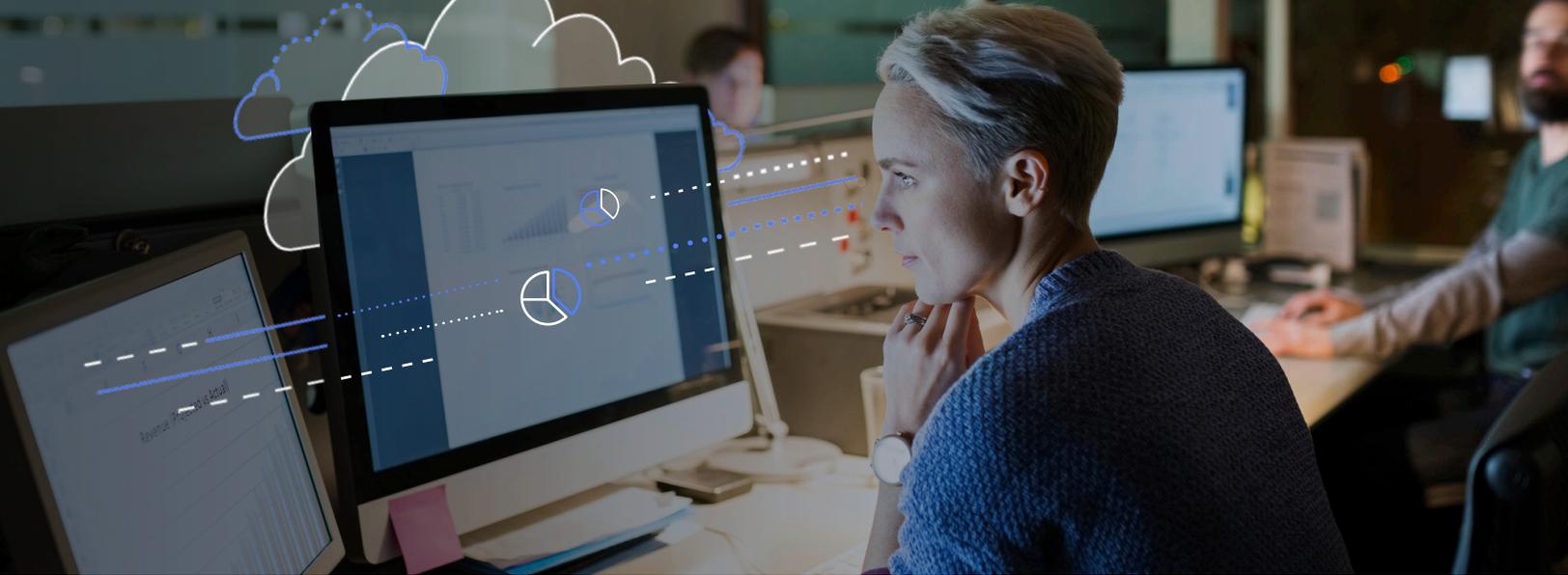
Comment rendre nos données disponibles pour qu'elles soient utilisées de manière inattendue ? Autrement dit, comment les utiliser pour rendre notre entreprise agile ? Comment faire pour apporter rigueur et créativité à la prise de décision opérationnelle ? Comment changer la culture d'entreprise de manière à bénéficier de cette nouvelle flexibilité ?

Et comment contrôler correctement les données pour protéger la vie privée, tout en permettant leur utilisation flexible et rapide ?



Nous devons nous poser deux questions :

1. Comment rendre nos données plus agiles ?
2. Comment utiliser les données pour appliquer la méthode agile dans notre entreprise ?



## L'agilité des données

### Comment rendre nos données plus agiles ?

Pour atteindre une agilité opérationnelle, il faut se préparer à réagir aux changements inattendus dans des environnements commerciaux et concurrentiels et créer des innovations réellement inédites. Nous devons donc être en mesure d'exploiter nos données de façons auxquelles nous n'avons pas forcément pensé au moment de les recueillir.

#### Nos difficultés :

- Nos données sont probablement enfermées dans des bases de données transactionnelles et relationnelles, probablement cloisonnées, ce qui les rend inaccessibles à différents services de l'entreprise.
- Nous n'avons peut-être pas les outils d'analyse appropriés, ou ceux-ci ne sont peut-être pas accessibles aux bonnes personnes et aux bons moments.
- Nos modèles de sécurité et de confidentialité sont ad hoc, car nous n'avons peut-être jamais pensé à utiliser les données à des fins d'exploration. Le plus souvent, nous favorisons la confidentialité en rendant tout simplement les données aussi inaccessibles que possible.

#### Nos objectifs :

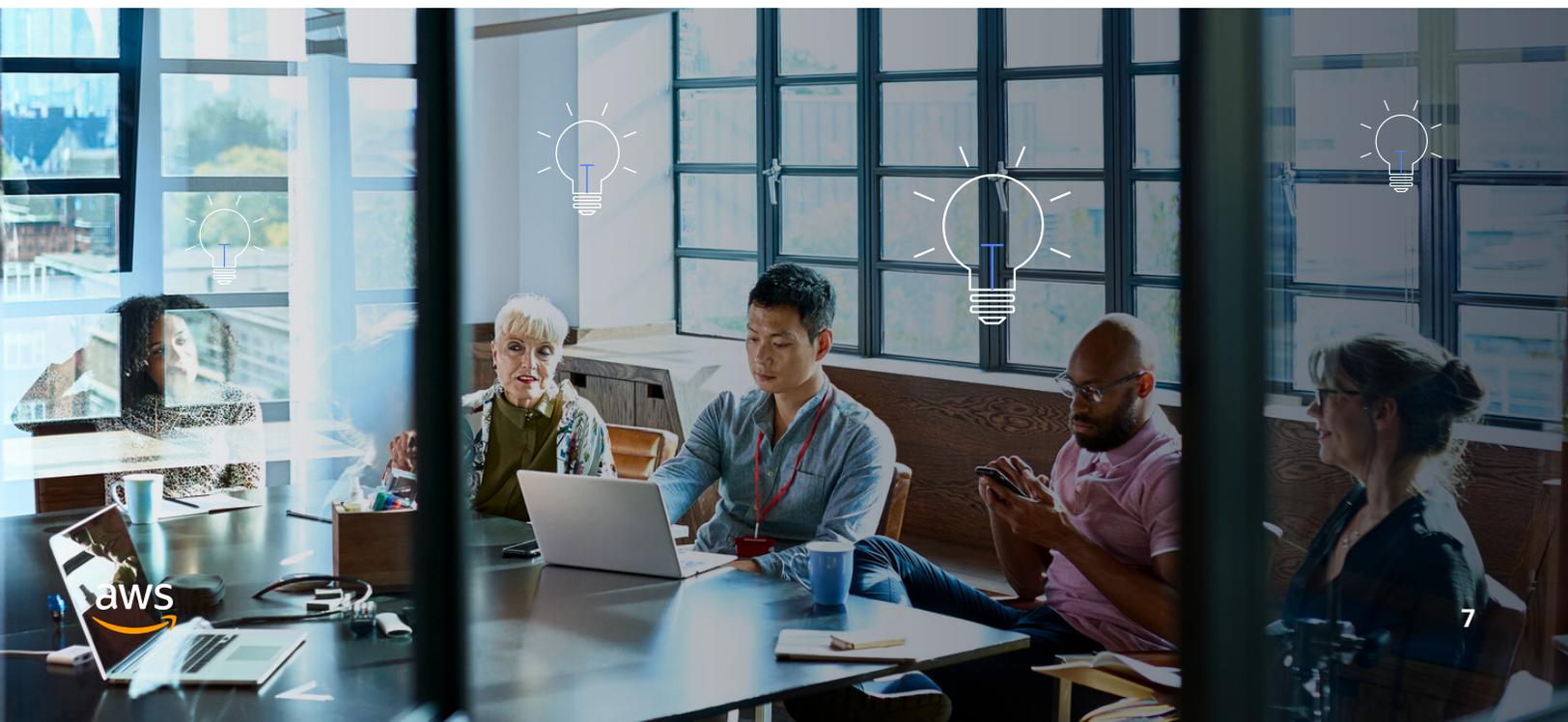
- Rendre les données aussi disponibles que possible, tout en garantissant leur confidentialité.
- Encourager la transparence dans l'entreprise en éliminant les silos d'informations.
- Fournir aux employés les outils appropriés pour explorer les données avec inventivité et en exploitant les derniers progrès analytiques.
- S'assurer que les employés ont l'expertise nécessaire pour interpréter les données de manière rigoureuse et créative.

Dans « [Analytics without Limits: FINRA's Scalable and Secure Big Data Architecture](#) », John Brady, responsable la sécurité de l'information (RSSI) de la Financial Industry Regulatory Authority (FINRA), formule ces objectifs avec élégance en indiquant qu'il souhaite **réduire le coût de la curiosité**. Il se réfère ainsi au coût au sens le plus large, incluant le temps nécessaire pour tirer des conclusions des données et le risque que représente le fait de les rendre disponibles. L'objectif de la FINRA est d'analyser les quelque 37 milliards de transactions qui ont lieu chaque jour sur les marchés financiers afin de détecter tout modèle de fraude. Étant donné qu'il est parfois impossible de savoir à l'avance à quoi ressemble un modèle de fraude, la FINRA doit compter sur l'expertise de ses analystes pour repérer des comportements suspects. Cette tâche se résume à la curiosité : lorsqu'ils examinent les données, les analystes doivent être assez curieux pour se demander pourquoi tel ou tel modèle apparaît. La mission du service informatique consiste à réduire le coût de cette curiosité ainsi que les efforts déployés par les analystes pour analyser une intuition.

L'idée de John Brady s'applique à tous les secteurs et à tous les postes. Les vendeurs peuvent-ils facilement explorer des données pour trouver des modèles inattendus dans les achats des clients ? Les employés d'un service des opérations peuvent-ils explorer des données pour identifier les optimisations de performances ou pour diagnostiquer des problèmes dans les processus d'exploitation ? Les employés d'un service des finances peuvent-ils trouver de nouvelles manières de stimuler les performances ou de décortiquer les données pour étayer la prise de décision des dirigeants ? Les responsables informatiques peuvent-ils vérifier leurs hypothèses sur l'optimisation des dépenses de cloud avec rigueur et créativité ?

La curiosité entraîne l'innovation et l'amélioration. Les données agiles permettent aux employés d'explorer des idées, des intuitions, des hypothèses et des conjectures librement et à la vitesse de l'éclair ainsi que de promouvoir de nouvelles idées soutenues par les données.

Pour rendre les données agiles, une entreprise doit déterminer comment obtenir des données et quelles données obtenir, comment les préserver, comment et sous quelles conditions les rendre disponibles et les outils et compétences dont elle dispose pour les exploiter.



# 1

## Obtenir des données



Pour utiliser des données de manière agile, il faut d'abord avoir des données. Et étant donné que nous ne connaissons pas encore toutes les utilisations que nous allons en faire, nous devons recueillir *davantage* de données que celles que nous savons déjà comment utiliser. C'est, en somme, ce que nous appelons le « Big Data ». Heureusement, avec le cloud, le coût de stockage des données est faible et en baisse. Par conséquent, nous pouvons équiper nos processus métier d'instruments qui permettent de produire des données, beaucoup de données, et de les rendre disponibles pour analyse. Par exemple, les applications de l'internet des objets (IoT) incluent souvent des capteurs qui dispersent un flux de points de données dans le cloud que l'entreprise peut analyser immédiatement ou stocker pour une future analyse. Les entreprises peuvent maintenant travailler avec une gamme bien plus vaste de types de données : vidéos, textes et discours, par exemple. Les perspectives d'utilisation innovante et intéressante de ces informations sont considérables.

GE Oil and Gas, par exemple, opère un périphérique type scanner appelé un « pig » (cochon) à travers ses pipelines pour colliger plus de 750 To d'informations ce qui l'aide à détecter des problèmes potentiels dans l'infrastructure pipeline. Hudl a collecté environ 10 Po de vidéo et d'autres données que les entraîneurs peuvent revoir avec leurs joueurs. Peloton réunit des données de ses cycles d'exercices et les analyse pour fournir des informations à ses clients. Airbnb accumule environ 50 Go de données par jour pour une analyse rapide dans le cloud à l'aide d'Amazon Elastic MapReduce (EMR), un outil qui permet d'analyser rapidement et en parallèle de grands volumes de données.<sup>5</sup>

# 2

## Stocker les données



Après avoir acquis des données, il faut pouvoir les stocker afin qu'elles soient disponibles pour les analyses. Avant, nous stockions les données dans un format structuré correspondant à l'utilisation transactionnelle que nous allions en faire. Par exemple, nous avions un champ « quantité commandée » et un champ « prix à l'unité » dans une base de données. Nous collections les données pour remplir ces champs et les classions en remplissant les cases vides correspondantes dans la base de données, en sachant que nous pourrions toujours les multiplier par des valeurs données pour calculer un prix total. Ainsi utilisées, les données étaient utiles pour les transactions, mais nous perdions des informations qui auraient pu être utiles pour l'analyse. Il s'agissait d'un modèle de bases de données relationnelles.

Les dernières décennies ont été dominées par l'utilisation de ces bases de données relationnelles, qui sont tout à fait adaptées au traitement efficace et bien connu (« multiplication du prix à l'unité par la quantité commandée ») des volumes d'antan de données transactionnelles. Mais si vous travaillez avec des données non transactionnelles, que vous effectuez des transactions à d'immenses échelles sur Internet ou que vous gérez des données qui ne rentrent pas aussi facilement dans des « champs de données » prédéfinis, il existe désormais de bien meilleures solutions spécialement conçues pour le cloud.

Par exemple, Amazon Timestream est une base de données spécifiquement conçue pour gérer des données de séries chronologiques (comme les données produites au fil du temps par un capteur industriel ou par le suivi chronologique de l'activité du marché) ; Amazon Quantum Ledger Database est destiné au type de données utilisé dans Blockchain (données dont l'historique doit être vérifiable à l'aide de techniques comme le chiffrement) ; et Amazon Neptune est conçu pour représenter des connexions et des relations complexes, comme les réseaux sociaux. Les entreprises ne sont plus limitées aux données adaptées aux modèles relationnels.

Mieux encore (pour l'agilité), les données qui seront utilisées pour des analyses encore indéterminées, peuvent être stockées dans un référentiel flexible appelé « lac de données », où chaque donnée est simplement stockée sous la forme sous laquelle elle a été reçue. Le pouvoir des lacs de données réside dans les outils qui peuvent être utilisés pour les analyser : des outils qui vous permettent de combiner des informations hétérogènes mixant des données structurées et non structurées, des données provenant de différents silos organisationnels, ainsi que des données en grandes quantités. Les outils d'aujourd'hui peuvent appliquer des algorithmes de machine learning et des analyses statistiques, et peuvent travailler avec du texte, des vidéos et des discours en langage naturel.

En d'autres termes, le lac de données répond aux besoins de l'entreprise, car il permet de stocker des données alors que les différentes utilisations qui pourront en être faites ne sont pas encore toutes connues. Nous pouvons ajouter les données au lac à partir de différents silos opérationnels et les analyser ensemble. Nous pouvons rapidement trouver un moyen d'ajouter au lac les données d'une entreprise récemment acquise et ainsi mieux cerner ses opérations, et nous pouvons intégrer ses données aux nôtres. Les ingrédients qui permettent tout cela sont : (1) le faible coût du stockage, (2) la disponibilité des outils qui fonctionnent avec des données hétérogènes et peu structurées, et (3) la disponibilité des services, qui vous permet d'ajouter vos données au lac de données avec un débit élevé et de manière asynchrone (envoyez simplement les données vers le lac de données lorsque vous les recevez, et elles y seront aussi vite que possible, sans que vous n'ayez besoin d'attendre, un peu comme pour les e-mails).

# 3

## Rendre les données disponibles



L'étape suivante pour que les données soient agiles consiste à les rendre disponibles, quand et où elles sont utiles. (Remarquez que je n'ai pas indiqué quand et où elles sont **nécessaires**. Nous parlons bien d'agilité et d'innovation ici.) Le modèle qui est souvent utilisé aujourd'hui est un modèle de mise en libre-service. Lorsqu'un analyste est curieux, il peut utiliser un ensemble d'outils et un sous-ensemble des données à analyser sans avoir à demander à quelqu'un de les fournir ni à attendre de les avoir. La liberté qui en découle permet à l'analyste de poursuivre son raisonnement sous forme de « flux », plutôt que de procéder par intermittence, ce qui nuit à sa créativité (et augmente le coût de la curiosité). Le cloud est un instrument décisif à cet égard, car il permet de mettre en service et d'utiliser de nouveaux environnements de travail, puis de s'en débarrasser lorsqu'ils ne sont plus nécessaires. Il facilite également la mise en place de garde-fous pour protéger la vie privée (plus d'informations à ce sujet à la suite).

# 4

## Fournir des outils



Une entreprise qui s'appuie sur les données met facilement et rapidement à la disposition de ses employés les outils d'analyse appropriés, souvent par le biais d'un modèle d'auto-allocation, tel que décrit ci-dessous. Il existe une grande diversité de logiciels et de services : si vous souhaitez exécuter des requêtes structurées d'une manière traditionnelle sur des données, vous pouvez, par exemple, configurer un entrepôt de données basé sur les données du lac de données ou fournir un outil qui vous permet d'exécuter de vieilles requête de type SQL directement sur le lac de données.

Mais aujourd'hui, bien d'autres possibilités s'offrent à vous. Vous pouvez, par exemple, visualiser vos données grâce à des outils de modélisation, et aussi élaborer des scénarios et déterminer leurs conséquences. La révolution analytique actuelle repose sur l'intelligence artificielle et le machine learning, qui ouvrent de nouvelles perspectives en matière d'exploitation de nos données : prévision des résultats, repérage des anomalies, catégorisation des données, analyse des sentiments, identification de modèles, guidage des robots, et bien plus encore...

Citons à titre d'exemple Capital One, qui utilise le machine learning pour déceler les fraudes tout en maintenant des niveaux élevés de service client. À des fins d'amélioration de son service client, T-Mobile a recours au machine learning pour prédire les articles qui seront le plus utiles au client, ce qui permet de les faire parvenir rapidement aux agents du service clientèle. Au moment de la diffusion du mariage royal britannique, Sky News a fait usage du machine learning pour reconnaître le visage des célébrités dans la foule et les identifier pour les téléspectateurs. Et enfin, la Formule 1, la Ligue majeure de baseball et la National Football League utilisent toutes le machine learning pour améliorer l'expérience du spectateur dans leurs sports respectifs.<sup>6</sup>

Pour appliquer le machine learning, vous entraînez un modèle basé sur des jeux de données antérieurs et vous l'appliquez ensuite aux nouvelles données au fur et à mesure des observations. Dans le cadre d'AWS, trois approches générales du machine learning se profilent : (1) utilisation d'un modèle préentraîné comme Amazon Rekognition, qui a déjà été entraîné à la reconnaissance d'objets dans des images, ou Amazon Lex, qui a été entraîné à comprendre des intentions exprimées en langage naturel, (2) entraînement et utilisation de votre propre modèle basé sur l'un des algorithmes couramment utilisés pour le machine learning à l'aide d'Amazon SageMaker, ou (3) recours à vos propres algorithmes et approches d'entraînement, si vous disposez d'employés compétents en machine learning, en travaillant directement avec l'infrastructure Amazon optimisée pour le machine learning.

Grâce à des outils comme ceux-là, les entreprises peuvent donner libre cours à la créativité de leurs employés et trouver de nouvelles façons d'utiliser pleinement les données.

# 5

## Perfectionnement



Le prochain élément essentiel et sous-jacent à la valorisation de vos données consiste à vous assurer que vos employés possèdent les bonnes compétences... en plus du sens de la curiosité. C'est pour cette raison que les scientifiques des données sont tant convoités à l'heure actuelle. Oui, il existe de nombreux outils, même pour les personnes ayant peu de compétence et d'expérience en matière de statistiques. Mais pour exploiter au mieux les données, toujours rigoureusement, il est important d'être entouré de personnes capables de tirer des déductions à partir des données.

Un exemple simple : ceux d'entre nous qui ont moins d'expérience dans le domaine des statistiques ont tendance à se fier excessivement aux moyennes, même si l'examen d'une distribution entière de valeurs peut souvent mener à des éclairages importants. Je me souviens d'un cas précis, lorsque j'étais CIO chez USCIS, où nous cherchions à réduire le temps nécessaire au traitement de certains types d'applications. Nous avons créé des tableaux de bord pour suivre le temps moyen de traitement, mais chaque tentative de changement ne semblait avoir qu'un faible impact sur la métrique. Nous avons manqué un élément, le petit nombre d'applications qui soulevaient des problèmes de sécurité nationale ou de fraude et qui prenaient beaucoup plus de temps à traiter, ce qui faussait la moyenne. Nous n'avions aucun moyen de contrôler le temps que cela prenait. Bien que nos améliorations soient applicables à la grande majorité des cas, en raison de la moyenne extrêmement faussée, nous ne pouvions pas vraiment visualiser leur impact. Lorsque nous avons pris conscience du problème et que nous avons commencé à surveiller, par exemple, le 85e percentile du délai de traitement, nous avons pu identifier l'impact significatif de nos changements sur la grande majorité des cas. Nous disposions des données, des outils et de l'accès ...nous ne disposions tout simplement pas des compétences nécessaires pour en tirer les bonnes conclusions.

Les décisions fondées sur des données peuvent également être mal fondées lorsque les données sont présentées (même involontairement) de manière trompeuse. Dans son ouvrage *The Visual Display of Quantitative Information*, Edward Tufte démontre comment les données peuvent être dénaturées ou brouillées par la façon dont elles sont présentées.<sup>7</sup> Là encore, une entreprise qui se dit vouloir être rigoureuse dans son utilisation des données doit s'assurer qu'elle dispose des compétences adéquates en matière d'analyse et de présentation, autant que dans le traitement des données.

# 6

## Fournir des garde-fous



Avant de pouvoir mettre des données à disposition pour des utilisations nouvelles (pour satisfaire la curiosité ou pour ainsi dire), nous devons mettre en place des garde-fous dédiés à la protection de la vie privée et de la confidentialité. Les entreprises axées sur les données pratiquent le « respect de la vie privée dès la conception », en établissant délibérément des garanties fondées sur la planification et la proactivité. Elles gagnent en rapidité et en souplesse en s'assurant qu'elles ont déjà réfléchi à ce qui doit être protégé et qu'elles ont mis en place des moyens automatisés pour le faire. En fait, le récent Règlement général de la Communauté européenne sur la protection des données (RGPD) exige le respect de la vie privée dès la conception.

Le cloud fournit de nombreux outils pour la mise en place de contrôles d'accès automatisés et concrétise cela à un niveau granulaire qui vous permet de donner aux employés l'accès aux données auxquelles ils devraient précisément avoir accès. Il est possible de suivre la provenance et la validité des données, de les chiffrer ou de les brouiller, et de restreindre l'accès champ par champ ou enregistrement par enregistrement. En d'autres termes, vous pouvez préciser à quelles données de clients un employé a accès et quelles données associées à ces clients l'employé peut consulter. Amazon Macie utilise même le machine learning pour identifier quelles données de votre lac de données sont des informations personnellement identifiables (IPI) et permet de suivre leur utilisation. Vous pouvez également choisir de gérer les données uniquement à un niveau agrégé ou avec des informations masquées ou anonymisées. Toute la flexibilité réside à ce stade ; chaque entreprise data-driven doit prendre des décisions responsables en matière de protection de la vie privée, compte tenu du type de données qu'elle gère.

L'utilisation des gros volumes de données dont dispose l'entreprise pose de nombreux autres problèmes. Il est souvent difficile d'associer avec précision les données issues de différents systèmes informatiques concernant une même personne, surtout dans des pays comme les États-Unis qui ne disposent pas d'un système d'identification national unique. Les données peuvent être inexactes non seulement en raison d'erreurs commises lors de la saisie des données, mais aussi à cause des limitations des systèmes informatiques qui collectent les données. Par exemple, il existe des systèmes informatiques qui ne permettent d'enregistrer qu'un nom de famille et un prénom, ce qui impose une imprécision pour les personnes qui ont plus de deux noms.<sup>8</sup>

Quoi qu'il en soit, l'objectif d'une entreprise data-driven revient à mettre à disposition des données permettant de prendre des décisions rigoureuses et précises et d'innover en permanence. Cela implique de recueillir et de stocker les données en vue d'une utilisation ultérieure flexible, de les mettre, avec les bons outils, à la disposition de ceux qui les utiliseront sans aucune friction, de garantir le respect de la vie privée et la confidentialité dès la conception, de cultiver les compétences nécessaires pour faire des déductions valables et de résoudre les problèmes d'hygiène des données qui peuvent mener à des décisions mal éclairées. C'est cela, rendre les données « agiles ».

# Comment utiliser les données pour appliquer la méthode agile dans notre entreprise ?

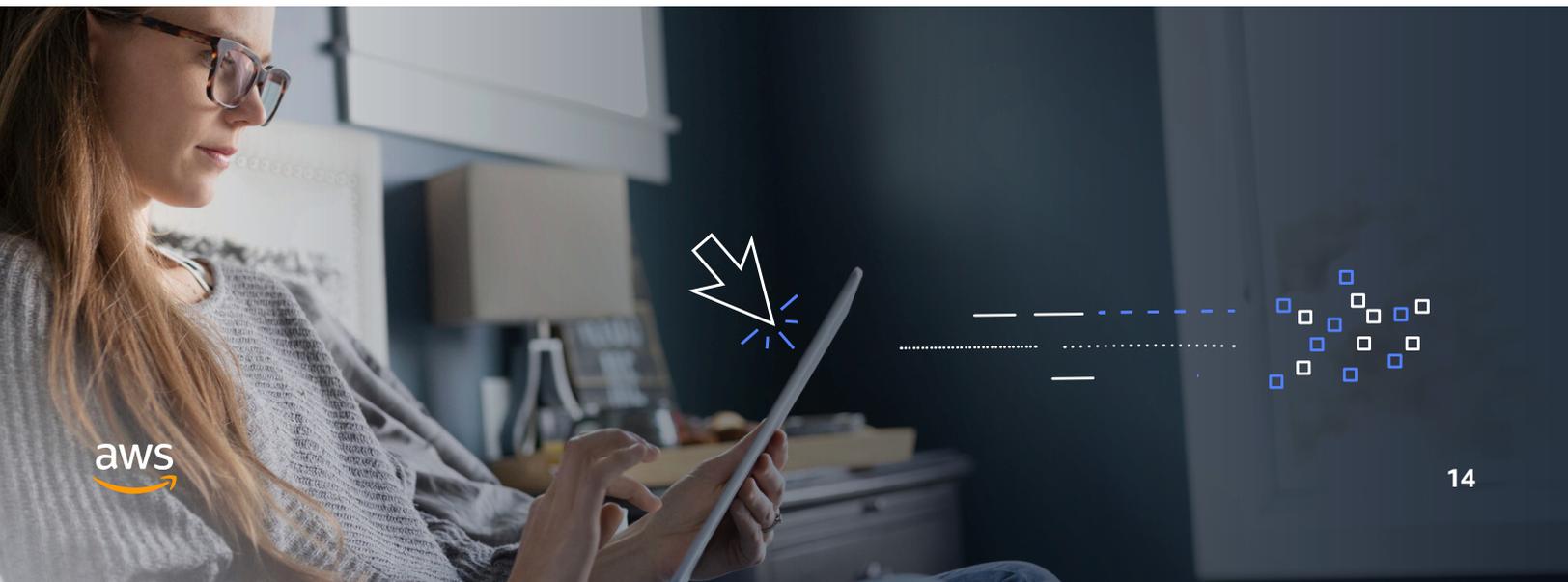
À l'ère du numérique, une entreprise agile procède en testant une idée, en obtenant un retour d'information (feedback), puis en ajustant son cap, et ce, de manière répétée. Cette approche de feedback rapide permet à l'entreprise d'innover (à faible risque, à grande vitesse et à faible coût) et de réduire le risque d'investissement en testant les idées avant de se lancer. Il en résulte une bonne adéquation entre les produits de l'entreprise et les marchés auxquels ils sont destinés et la garantie que l'entreprise résout le bon problème de la bonne manière et au bon moment.

## Feedback rapide

Le feedback, en ce sens, ne consiste pas à demander aux clients s'ils apprécient une nouvelle fonctionnalité ou un nouveau produit. Le plus souvent, les entreprises axées sur les données ont recours à un feedback quantitatif (le type de feedback recueilli en observant la manière dont les clients *agissent réellement*) ou à la surveillance des changements de comportement sur le marché ou encore à d'autres métriques.

Par exemple, les entreprises améliorent souvent la navigabilité de leurs sites web par des tests A/B, c'est-à-dire en essayant deux variantes sur un élément de conception (généralement, une variante est la version actuelle, celle du statu quo, et l'autre est un nouvel élément de conception qu'elles envisagent d'introduire). Elles montrent la version A et la version B à certains clients. Elles recueillent des données sur l'activité des clients et les analysent par rapport aux résultats qui les intéressent. Si elles veulent décider de rendre un bouton vert ou rouge afin d'optimiser le nombre de fois où l'internaute cliquera dessus, elles peuvent montrer à certains utilisateurs une version verte et à d'autres une version rouge, et voir laquelle récolte le plus de clics. Expedia et Netflix sont des exemples d'entreprises qui effectuent régulièrement des tests A/B, en utilisant de gros volumes de données provenant d'un entrepôt de données dans le cloud.<sup>9</sup>

<sup>9</sup> <https://www.youtube.com/watch?v=k8PTetgYzLA>.





L'approche puissante de l'apprentissage et de l'ajustement par le biais du feedback va bien au-delà du simple test de l'interface utilisateur A/B. De nouvelles idées de produits, par exemple, peuvent être testées en créant un « produit minimum viable », la version la plus petite et la plus simple du produit que l'entreprise peut utiliser pour recueillir des informations sur le succès du produit ou sur ce qu'il faut changer pour qu'il le soit. Les stratégies de marketing, les promotions, les alternatives technologiques, tout cela peut être testé par des versions d'essai et des métriques afin de réduire l'incertitude. Et le fait de rassembler des données et de les mettre à disposition pour analyse constitue la clé du succès.

La technique consistant à utiliser le concept du produit minimum viable et un feedback rapide est décrite dans l'ouvrage d'Éric Ries, *Lean Startup. Adoptez l'innovation continue*.<sup>10</sup> Selon M. Ries, à tout moment, une start-up émet deux hypothèses : une hypothèse de valeur, sur la façon dont le produit proposé créera de la valeur pour les clients, et une hypothèse de croissance, sur la façon dont l'entreprise pourra développer son marché, c'est-à-dire amener les clients à utiliser le produit. Le produit minimum viable est le plus petit produit qui donnera à la start-up des informations permettant de confirmer ou de réfuter ces hypothèses, après quoi elle pourra apporter des modifications et les tester à nouveau en les confrontant au marché.

Cet ensemble de pratiques ne s'applique pas uniquement aux start-up ou au développement de nouveaux produits. Cet aspect est devenu un élément central de la manière dont les organisations, y compris les grandes entreprises, arrivent à appliquer l'agilité métier en changeant de cap sur la base des enseignements tirés. Si une entreprise envisage de développer un nouveau système informatique à l'usage de ses propres employés, elle a probablement une hypothèse sur la manière dont ce système informatique produira les résultats commerciaux proposés dans son analyse de rentabilité. Cette hypothèse doit être testée, et des changements doivent être apportés en fonction de ce montrent les données.

Par conséquent, une pratique agile nécessite des données : pour apprendre et s'adapter, l'entreprise doit recueillir des données sur l'impact de ses nouvelles initiatives et les utiliser pour informer ces dernières. L'agilité exige en outre que l'entreprise perçoive les changements dans son environnement commercial, afin qu'elle puisse y répondre de manière appropriée pour optimiser ses résultats commerciaux. Une entreprise data-driven ne met pas seulement de l'agilité dans ses données, mais utilise également les données pour soutenir son agilité.

## Changer la culture et les processus

En ce sens, s'axer sur les données exige une manière très différente de prendre des décisions ; il s'agit d'un changement culturel profond pour de nombreuses organisations. Par le passé, nous aurions pu prendre des décisions en élaborant des plans détaillés, en analysant les options avec les données disponibles et en choisissant l'option qui (en ne considérant que les données disponibles) semble générer les meilleurs résultats. À l'ère du numérique, nous refusons de n'accepter que les données qui sont disponibles au moment de la création du plan. Nous concevons plutôt des expériences pour obtenir des données supplémentaires, puis nous intégrons ces données dans notre processus décisionnel. Nous faisons face à l'incertitude en générant de nouvelles données.

La technique de gouvernance des technologies de l'information que nous avons conçue à l'USCIS en est un exemple. Au lieu de rédiger un gros document d'exigences et de le remettre aux technologues pour la mise en œuvre, nous avons simplement transmis un objectif commercial. Dans une étude de cas, par exemple, nous avons remarqué qu'un gestionnaire de cas qualifié (un « vérificateur de statut ») pouvait traiter environ 70 cas par jour, et notre objectif commercial consistait à faire en sorte que ce chiffre soit beaucoup plus élevé. Dans une autre étude de cas, nous avons constaté qu'un certain nombre de dossiers papier se perdaient au cours du transfert lors des déplacements d'un lieu de traitement à un autre, et nous voulions éliminer ces pertes.

Pour chacun de ces objectifs, nous avons commencé par créer un tableau de bord qui indiquait la métrique clé : le nombre de cas par jour ou le nombre de fichiers manquants. Au lieu de rédiger un document d'exigences, nous avons créé une équipe multidisciplinaire d'opérateurs commerciaux et de scientifiques des technologies de l'information, et nous leur avons demandé d'améliorer la métrique. Nous leur avons fourni les outils nécessaires pour apporter rapidement des modifications aux systèmes informatiques et aux processus métier, puis nous avons surveillé les tableaux de bord avec eux. Ils ont tenté de petits changements progressifs et ont vérifié les résultats chaque jour. Sur la base de ce qu'ils ont vu dans les données, ils ont pu décider de la suite à donner pour optimiser le résultat. Et la direction pourrait décider de continuer à financer l'initiative ou de diriger les fonds ailleurs. Il en est ressorti un processus de gouvernance léger, axé sur les données et à risque réduit, qui a permis d'obtenir rapidement une valeur ajoutée.

Autre point important : la responsabilisation est renforcée par la transparence. En rendant les données largement disponibles, nous avons augmenté la visibilité des progrès de l'équipe. Par conséquent, les organes de surveillance pourraient constamment revoir la décision d'investissement, soit en investissant plus ou moins, soit en redéfinissant les objectifs, ou encore en stoppant complètement les investissements. Les résultats étaient le seul critère de réussite, et les résultats pouvaient être obtenus rapidement. Mais ces résultats devaient être étayés par les données.

## Repérer des modèles

Un autre domaine dans lequel les données peuvent favoriser l'agilité est la détection des changements ou la reconnaissance de modèles au sein de l'environnement. Par exemple, le machine learning peut être utilisé pour détecter les anomalies et y répondre. Nous pouvons entraîner un modèle de machine learning avec des données historiques ou de routine afin qu'il s'habitue à ce qui est « normal » et l'appliquer ensuite pour déceler une activité qui est anormale. Cette technique peut être utilisée, par exemple, pour repérer les transactions frauduleuses ou les intrusions de pirates sur les réseaux. Ou pour repérer, sur une chaîne de production d'une usine, un équipement qui dévie de son comportement normal et qui pourrait faire l'objet d'une réparation ou d'un remplacement ; et ce, avant qu'il ne tombe effectivement en panne.

Lorsque nous collectons de gros volumes de données, force est de constater que nous pouvons identifier des relations, jusqu'alors inconnues. Les entreprises de réseaux sociaux rassemblent de vastes bases de données sur les relations entre les personnes. La Sécurité intérieure pourrait découvrir qu'un terroriste potentiel sur lequel elle enquête a vécu à la même adresse qu'une personne déjà connue pour être un terroriste, ce qui pourrait l'amener à poser des questions lors de sa prochaine rencontre avec cette personne. Un certain nombre de demandes d'immigration frauduleuses peuvent se révéler avoir été préparées par le même avocat spécialisé en immigration. Ici, nous sommes allés bien au-delà de la simple utilisation de données pour traiter des transactions : nous pouvons maintenant trouver des relations importantes et intéressantes entre ces transactions. Mais, une fois de plus, nous ne savons pas exactement quelles relations nous pourrions trouver ; l'agilité, la flexibilité et la curiosité sont les clés qui permettent de tirer profit des données.

Pour citer un autre exemple d'utilisation des données dans le but de « garder un œil sur les événements », l'existence d'un point de données peut servir à confirmer qu'une activité a eu lieu (par exemple, lorsque des journaux d'audit sont créés automatiquement). En suivant la trace des activités, les auditeurs peuvent être en mesure de valider la conformité ou d'enquêter sur une activité inappropriée. La Blockchain est souvent utilisée pour stocker des données qui confirment que des activités ont eu lieu, par exemple, un transfert d'argent entre deux parties ou l'approbation d'un contrat par les parties concernées. En utilisant des garde-fous automatisés et des données d'audit pour établir la conformité, les entreprises peuvent souvent éviter des processus de conformité lourds qui réduisent l'agilité.

Bien entendu, l'utilisation des données pour soutenir l'agilité des entreprises pose des problèmes. Comme nous l'avons indiqué plus haut, il faut jouir de certaines compétences pour tirer les conclusions appropriées des données. Les données ne nous indiquent pas toujours les mesures à prendre : nous devons les interpréter et prendre de bonnes décisions. Souvent, il s'agit de trouver un compromis entre les faux positifs et les faux négatifs. Par exemple, si nous utilisons les données pour repérer les transactions anormales et identifier une fraude potentielle, nous courons le risque de signaler trop de transactions repérées comme anormales et d'ennuyer nos clients ou d'en signaler trop peu permettant ainsi à la fraude de se faufiler. Plus l'ensemble de données est important, plus il est probable que des modèles insignifiants apparaissent ou que des modèles importants soient enfouis parmi le simple nombre de connexions potentielles. Le bruit s'accumule en même temps que le signal.

## Pour terminer

Une organisation data-driven est une organisation qui met les données au service de l'amélioration des résultats commerciaux, à la fois en utilisant les données pour mener à bien un processus de prise de décisions rigoureux, et en rendant les données disponibles pour stimuler l'innovation et apporter de la valeur aux clients. Lorsque les données sont enfermées dans un cadre rigide, sont cloisonnées ou difficiles à obtenir, elles deviennent un obstacle à l'agilité de l'entreprise, l'empêchant de répondre aux opportunités ou de mettre rapidement ses produits sur le marché. Pire encore, lorsqu'une entreprise *n'oriente pas* ses processus et ses investissements vers l'utilisation de données, elle renonce à un contact important avec le marché qu'elle essaie de servir ou laisse passer un retour d'information qui pourrait l'aider à mieux réussir dans ses initiatives. Une organisation data-driven, en revanche, utilise les données pour gagner en agilité et exploite cette agilité pour rendre ses données plus précieuses.

### À propos de l'auteur

Mark Schwartz est un stratège d'entreprise chez Amazon Web Services et auteur de *The Art of Business Value, A Seat at the Table: IT Leadership in the Age of Agility*, et *War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age*. Avant de rejoindre AWS, il a été directeur de l'informatique du Service de la citoyenneté et de l'immigration des États-Unis (section du département de la Sécurité intérieure), directeur de l'informatique chez Intrax et directeur général d'Auctiva. Il est titulaire d'un MBA de Wharton et d'un BS en sciences informatiques et d'un MA en philosophie de l'Université de Yale.



Mark Schwartz,  
Stratège d'entreprise, AWS

[En savoir plus de Mark Schwartz»](#)