

Chapitre

Importance du Big Data pour vous et votre entreprise

C'est indéniable, dans le Big Data (ou grandes données), les volumes d'informations manipulées sont colossaux. Voilà pourtant un nom mal choisi. Cette expression fourre-tout désigne des données dont la taille ne permet pas de les caser dans les conteneurs habituels. Elle fait référence à des informations qui sont trop volumineuses pour tenir sur un seul serveur, qui sont dépourvues d'une structure leur permettant d'être stockées dans des bases de données orientées lignes et colonnes, ou qui arrivent sous la forme d'un flux permanent empêchant leur conservation dans un entrepôt de données statique. L'accent est souvent mis sur le volume, mais la complexité des grandes données réside en réalité dans le manque de structure.

En général, les ouvrages sur le Big Data indiquent dès les premières lignes la quantité de données disponibles dans le monde. Vous avez certainement déjà lu ces chiffres et connaissez les comparaisons habituelles. Une entreprise de taille normale dispose d'un volume d'informations égal à 427 fois celui

de la bibliothèque du Congrès¹. Le nombre de données photographiques détenues par Facebook est supérieur au nombre de pixels jamais traités par Kodak. Chaque jour produit plus de vidéos que les cinquante premières années de la télévision. Si ces allégations sur les volumes et les types des données sont imaginaires, elles ne sont probablement pas très loin de la vérité.

Ce livre ne procède pas à une telle énumération car le plus important n'est pas là. C'est vrai, les volumes de données disponibles sont gigantesques. Si l'on en croit les résultats d'une étude [1], le monde a manipulé en 2012 plus de 2,8 zettaoctets (2,8 milliards de gigaoctets, une quantité difficile à appréhender). C'est plus que ce que l'on n'a jamais connu, et la tendance est à l'augmentation. Vous ferez certes impression dans une soirée en annonçant la quantité de données disponibles, mais pour les entreprises qui doivent gérer des données massives et les exploiter, la question cruciale n'est pas le volume total. Pour reprendre un cliché, dans un contexte différent, « la taille ne compte pas ».

Le volume des données ne doit pas être le point de focalisation. L'important reste leur analyse, c'est-à-dire leur transformation en connaissances, en innovations et en valeurs commerciales. L'étude précédente suggère que seulement 0,5 % des 2,8 zettaoctets sont analysés d'une manière ou d'une autre. Le principal obstacle à cette analyse reste la nécessité de structurer les informations, car la majorité d'entre elles ne vient pas sous forme de lignes et de colonnes. La tâche qui nous attend est immense – les données doivent être structurées, analysées et exploitées. L'intégralité des informations ne

1. N.d.T. : fondée en 1800, la bibliothèque du Congrès américain est la bibliothèque nationale des États-Unis et la plus importante au monde avec 32 millions de documents imprimés, 3 millions d'enregistrements sonores, 12 millions de photographies et 5,3 millions de cartes (source : Bibliothèque nationale de France).

sera pas forcément utile – l'étude estime que seuls 25 % présentent une valeur potentielle. Mais quel que soit le pourcentage réel, l'étendue des possibilités est vaste.

Dépasser le cadre du battage médiatique

Vous êtes sans doute sceptique vis-à-vis du Big Data et de tout le battage médiatique qui l'entoure. Je l'étais aussi, mais c'était avant mes recherches sur le sujet. J'ai travaillé auprès de nombreuses entreprises sur l'utilisation de l'analytique à des fins d'avantage concurrentiel, et j'ai rédigé de nombreux ouvrages sur le sujet, notamment *L'Analytique, nouvel outil stratégique* (avec Jeanne Harris) et *Analytics at Work* (avec Jeanne Harris et Bob Morison). Le Big Data semblait alors revenir à présenter du vieux vin dans une nouvelle bouteille. Cette appellation a commencé à se répandre fin 2010, mais très peu en dehors de la Silicon Valley; autrement dit, un nouvel exemple de battage médiatique orchestré par un fournisseur, un consultant ou un analyste en technologie. J'ai même un instant envisagé de reprendre mes ouvrages, de procéder à un remplacement global avec « Big Data » et de publier ces nouveautés (je plaisante!).

Mes recherches sur le sujet en 2011 ont montré que j'avais tort d'être sceptique. Elles ont impliqué plusieurs études systémiques, une sur les scientifiques des données et le facteur humain dans le Big Data, une sur le Big Data dans les grandes entreprises, une sur le Big Data dans l'industrie du tourisme et une sur le processus de découverte des données dans le Big Data [2]. Une centaine d'entretiens sur de tels projets ont été menés auprès de startup, de sociétés en ligne établies et de grands groupes du monde industriel classique. Dans les grandes entreprises, les dirigeants partageaient souvent ce scepticisme et cette méfiance vis-à-vis de la publicité faite au Big Data. Ils pensaient gérer des grandes données depuis des

années (si l'on s'en tient au volume) et que le sujet n'avait donc rien de nouveau. Toutefois, au cours des entretiens, la plupart admettaient que l'absence de structure dans ces données modernes soulevait de nouveaux défis et apportait de nouvelles opportunités d'affaires.

En conclusion, il existe de réelles différences entre l'analytique classique et le Big Data (voir le Tableau 1.1), même si cela ne transparaît pas toujours dans les autres livres ou articles publiés, où les distinctions sont souvent très floues. Plus loin dans ce chapitre, ainsi que dans la suite de cet ouvrage, je reviendrai sur ces différences et sur quelques similitudes essentielles. J'affirmerai aussi que le Big Data, malgré mes réserves sur le nom choisi, est là pour durer et qu'il revêt une grande importance pour de nombreuses entreprises. Vous vous rallierez probablement à cette opinion, mais pour le moment gardez votre scepticisme.

Tableau 1.1 Comparaison du Big Data et de l'analytique classique

	Big Data	Analytique classique
Type de données	Formats non structurés	En lignes et colonnes
Volume de données	100 téraoctets à plusieurs pétaoctets	Dizaines de téraoctets ou moins
Disponibilité des données	Flux constant	Pool statique
Méthode d'analyse	Apprentissage automatique	À base d'hypothèses
Objectif premier	Produits orientés données	Support aux décisions internes et services

Le Big Data est important pour vous et votre entreprise ? Alors vous allez devoir agir. Vous devrez en identifier les aspects qui s'appliquent à votre métier, et les travailler. Vous aurez à embaucher des personnes, à louer les services d'une société ou à former des gens qui rendront le Big Data opérationnel. Vous aurez également à changer votre architecture technologique. Si vous vous lancez dans toutes ces actions, je saurai que j'ai réussi. Mais attention, elles doivent avoir un

sens pour vous, et dans tous les cas vous devrez d'abord être convaincu !

Cet ouvrage va essayer non pas de vous vendre le Big Data mais de vous aider à prendre les bonnes décisions. Ses points forts seront soulignés, sans oublier de préciser ceux qui sont surfaits. Les industries et les entreprises transformées par le Big Data seront décrites, sans oublier celles qui seront peu affectées, tout au moins à court terme. Si vous décidez d'aller plus loin dans le Big Data, vous découvrirez les chemins les plus économiques et les plus abordables.

Vous avez des contraintes de temps et des priorités, mais vous devez être prêt à mobiliser des ressources autour du Big Data. Commencez par un projet exploratoire afin d'entrevoir les nouvelles possibilités. La plupart des entreprises devraient se lancer sur cette voie, tout au moins discuter au plus haut niveau de la place que peut prendre le Big Data.

La section suivante expose mes hypothèses sur les raisons qui vous poussent à lire cet ouvrage. Viennent ensuite certaines de mes réserves quant au Big Data, notamment en ce qui concerne son nom et l'engouement qu'il suscite. Le reste de ce chapitre précisera pourquoi le Big Data est digne d'intérêt.

Qui êtes-vous ?

Qui êtes-vous pour vous intéresser aux implications commerciales et à la valeur du Big Data ? Les lecteurs de mes précédents ouvrages sur l'analytique étaient généralement des professionnels intelligents et ambitieux qui pensaient que les données et la technologie pouvaient changer les règles et les pratiques commerciales. Je soupçonne chez vous les mêmes attributs.

Les cadres les plus intéressés par le Big Data sont souvent ceux dont les fonctions impliquent de nombreuses données,

comme le marketing, la chaîne logistique, et de plus en plus la finance et les ressources humaines. Les personnes qui assurent la mise en place du Big Data, comme le service informatique, ont elles aussi envie d'approfondir leur connaissance du concept. L'intérêt est plus marqué encore dans les industries qui sont déjà fortement axées sur les données, comme les entreprises en ligne, ou celles qui ont la possibilité de se transformer – potentiellement n'importe quelle industrie, mais plus particulièrement celles qui disposent d'un grand nombre de données clients –, comme le commerce de détail, les voyages et le transport, les télécommunications, les médias et le divertissement, ainsi que les services financiers. Si vous êtes encore étudiant et prévoyez une carrière, ou un ou deux postes, dans le Big Data, félicitez-vous de votre perspicacité. Ce secteur a des chances de rester en plein essor pendant de nombreuses années.

Comme de nombreux managers et professionnels, vous connaissez certainement l'expression Big Data et savez que la production des données croît à un rythme effréné. En revanche, les différences entre le Big Data et la gestion et l'analyse traditionnelles des données sont sans doute moins claires. Plus important encore, une telle solution n'existe probablement pas dans votre entreprise. Un sondage, réalisé en 2013 auprès d'un millier de lecteurs de la *Harvard Business Review*, révèle que la plupart des personnes interrogées sont familières du concept du Big Data, mais seulement 28 % d'entre elles ont indiqué que leur entreprise « utilisait des solutions Big Data pour améliorer les décisions commerciales ou pour créer de nouvelles opportunités d'affaires ». 23 % ont répondu que leur entreprise avait une stratégie de Big Data. Seul un petit 6 % a confirmé que « l'entreprise avait pris en considération l'impact du Big Data sur les activités essentielles de son métier ». Un pourcentage encore plus faible, 3,5 %, affirmait que « l'entreprise savait comment appliquer le Big Data à son secteur ».

Voilà peut-être pourquoi vous lisez ce livre : vous voulez aider votre entreprise à répondre à de telles questions et, probablement, en faire profiter votre carrière. Vous avez choisi le bon point de départ !

Que cache l'expression Big Data ?

L'ensemble du concept de Big Data plaît à de nombreux cadres, à l'exception toutefois de son nom. Le principe est révolutionnaire et apporte des possibilités de transformation dans n'importe quelle entreprise, mais le terme lui-même pose problème, pour diverses raisons.

Tout d'abord la notion de taille, donnée par « big », ne représente que l'un des aspects distinctifs des nouvelles formes de données, et pour la plupart des entreprises, cette caractéristique n'est pas la plus importante. Une enquête effectuée en 2012 par NewVantage Partners auprès de plus de 50 dirigeants de grandes entreprises révèle que résoudre le problème d'absence de structure sur les données s'avère plus important que prendre en compte leur volume. 30 % des personnes interrogées estimaient que le principal problème du Big Data était « l'analyse de données provenant de sources variées ». 22 % étaient principalement préoccupées par « l'analyse de nouveaux types de données », tandis que 12 % se focalisaient sur « l'analyse de données arrivant en continu ». Seuls 28 % s'inquiétaient de l'analyse de jeux de données dont la taille dépassait 1 To, et presque la moitié de ces personnes (13 %) manipulaient des jeux de données avec une taille allant de 1 à 100 To – un volume relativement faible par rapport aux standards du Big Data [3].

Le terme soulève plusieurs autres questions. La notion de taille est évidemment relative – ce qui est grand aujourd'hui ne le sera pas nécessairement demain et ce qui est grand pour une entreprise est petit pour une autre. S'il faut établir une

base, disons que grand correspond à un dixième de pétaoctet ou plus. Quoi qu'il en soit, s'il y a un aspect des données où « la taille compte », c'est au niveau du matériel à acheter pour les stocker et les traiter.

Le Big Data a été défini par certains en fonction de trois V (volume, variété et vitesse). D'autres V ont aussi été ajoutés (vérité, valeur – vénalité est peut-être le suivant). Cependant cette définition pose également problème. Certes, ces caractéristiques sont importantes, mais que se passe-t-il si vous n'avez qu'un ou deux V ? N'avez-vous alors qu'un tiers ou que deux cinquièmes d'un Big Data ?

Par ailleurs, l'expression Big Data est déjà employée pour faire référence à toute utilisation de l'analyse ou, dans les cas extrêmes, à la production de rapports et à l'informatique décisionnelle classiques. Les fournisseurs et les consultants ont la fâcheuse habitude de prendre tout nouveau mot à la mode et de l'appliquer à leurs offres – Big Data n'a pas échappé à la règle. Attention donc lorsque vous lisez des publicités, des livres ou des articles concernant ce sujet ! S'ils font référence à la production de rapports, à la prise de décision fondée sur des données ou à l'analyse classique, vous étudiez des idées certes utiles et précieuses, mais pas particulièrement nouvelles.

En raison de ces problèmes de définition, la durée de vie de ce terme malheureux sera relativement courte. Il a été adopté par les médias et les startup, mais au sein des grandes entreprises, en particulier celles qui manipulent des données massives depuis de nombreuses années, comme les banques et les sociétés de transport, les personnes axées sur la technologie montrent une certaine réticence à son usage. Le Chapitre 8 reviendra sur le sujet, mais sachez qu'elles considèrent que les sources et les types de données prétendument nouveaux aujourd'hui ne sont que l'une des nombreuses nouvelles générations précédentes. Bien entendu, cela ne signifie pas que « le phénomène précédemment appelé Big Data » va dis-

paraître. Pour décrire la grande diversité des nouveaux types de données apparus cette dernière décennie, il n’y a pas encore de meilleur terme que Big Data.

En raison de son manque de précision, les entreprises doivent décortiquer cette expression afin d’affiner leurs stratégies et d’indiquer aux intéressés ce qu’elles souhaitent réellement faire avec ces nouveaux types de données, en précisant ceux qui sont les plus importants. Le Tableau 1.2 montre que le menu des possibilités du Big Data est digne d’un restaurant chinois. Faites votre choix en partant de n’importe quelle colonne.

Tableau 1.2 Le menu des possibilités du Big Data

Forme des données	Source des données	Secteur concerné	Fonction affectée
Grand volume	En ligne	Services financiers	Marketing
Non structurées	Vidéo	Santé	Chaîne logistique
Flux continu	Capteur	Production	Ressources humaines
Formats multiples	Génomique	Voyage/transport	Finance

Plutôt que d’annoncer « nous nous lançons dans un projet de Big Data », il est probablement plus constructif de dire « nous allons analyser les données vidéo de nos distributeurs et de nos agences pour mieux comprendre la relation client ». Si vous travaillez dans le domaine de la santé, vous pouvez décider de « combiner les enregistrements médicaux électroniques et les données génomiques de façon à établir un schéma de traitement personnalisé pour les patients ». Outre la clarté de vos intentions et de votre stratégie, cette approche permet d’éviter des discussions sans fin sur la taille, petite ou grande, des données concernées (en réalité, peu d’entreprises annoncent travailler avec des « petites données », même si cette activité est tout à fait respectable – et un terme est utile uniquement si son contraire est également pertinent).

Vous pouvez évidemment profiter de la publicité faite autour d'un terme. Si votre entreprise fait partie de celles qui réagissent uniquement aux nouveaux et attrayants objets de management – et si c'est encore le cas du Big Data au moment où vous lisez ces lignes –, lancez-vous dans un projet de grandes données. Autrement dit, si avoir l'étiquette Big Data permet de mobiliser de l'énergie et de l'enthousiasme au sein de votre entreprise, n'hésitez pas un instant. Vous devrez toutefois rester prêt à basculer sur le prochain terme à la mode attribué à ces activités dès qu'il apparaîtra ; IBM dispose déjà d'un centre de recherche sur les « données massives » – les données « gargantuesques » ou « gigantesques » ne sont pas loin ! Cet ouvrage conserve le terme Big Data car il n'en existe pour le moment aucun autre qui corresponde aux caractéristiques en question. Quoi qu'il en soit, extraire une valeur réelle de cette ressource mal nommée demande de s'y plonger en profondeur.

Quid de la pérennité du Big Data ?

Son nom a probablement été mal choisi, mais le Big Data va-t-il perdurer ou n'est-il qu'un phénomène de mode dans le monde de la gestion stratégique de l'information, comme ont pu l'être le hula hoop ou les Pet Rock dans d'autres contextes ? Si ce n'est qu'une lubie, les chefs d'entreprise peuvent l'oublier sans crainte. En revanche, s'il s'agit d'une composante importante et pérenne du domaine professionnel, les entreprises et les managers prennent des risques en l'ignorant.

Le Big Data comprend fatalement des éléments éphémères, notamment parce que le principal concept sous-jacent n'est pas vraiment nouveau. L'idée d'analyser des données de façon à comprendre ce qui se passe dans une entreprise a émergé depuis longtemps (au moins depuis 1954 lorsque UPS a mis

en place un groupe d'analyse). Dans ce cas, pourquoi a-t-il fallu lui trouver de nouveaux noms ? Rechercher un sens à des données a été successivement appelé *aide à la décision*, *aide à la direction*, *traitement analytique en ligne*, *informatique décisionnelle*, *analytique*, et à présent *Big Data* (voir le Tableau 1.3) [4]. Chaque changement de terminologie comporte certainement de nouveaux éléments, mais les évolutions ne semblent pas suffisantes pour mériter six générations.

Tableau 1.3 Désignation de l'utilisation et de l'analyse des données

Termes	Période	Signification
Aide à la décision	1970-1985	Utilisation de l'analyse des données pour aider à la prise de décision
Aide à la direction	1980-1990	Analyse des données pour aider à la prise de décision chez les cadres dirigeants
Traitement analytique en ligne (OLAP)	1990-2000	Logiciel pour l'analyse multidimensionnelle de tables de données
Informatique décisionnelle	1989-2005	Outils pour faciliter les décisions fondées sur des données, avec un accent mis sur les rapports
Analytique	2005-2010	Décisions fondées sur des analyses statistiques et mathématiques
Big Data	2010-aujourd'hui	Intérêt porté aux données très volumineuses, non structurées et en constant changement

Dans le cas du Big Data, le changement de terme peut s'expliquer par les nouvelles formes de données volumineuses en jeu – selon une estimation [5], 2,5 trillions d'octets (ou 2,5 suivi de 18 zéros) sont générés chaque jour dans le monde. La prise en charge d'informations peu structurées justifie plus encore le nouveau terme et la nouvelle approche. Certains types de données, comme le texte et la voix, nous accompagnent depuis très longtemps, mais leur production en masse, sur Internet et sous d'autres formes numériques, annonce une nouvelle ère, tout comme les nouvelles technologies qui en permettent l'analyse. La nouveauté des données des médias sociaux, qu'elles proviennent des pages Facebook, des pages Pinterest ou des tweets, ne fait aucun doute. Toutes

ces formes de données ne vont probablement pas perdurer, le schéma classique tendant vers une inclusion des fonctionnalités de ces outils dans des applications plus vastes. En revanche, les données qu'ils produisent et les connaissances qu'ils dévoilent sur leurs sources ne vont pas disparaître.

De façon générale, les *données issues de capteurs* sont là pour rester. En 2011, le nombre de dispositifs en réseau a dépassé le nombre d'habitants sur Terre. Les analystes estiment que 50 milliards de capteurs seront connectés à Internet d'ici à 2025 (« l'Internet des objets »), chacun produisant sa kyrielle d'informations. Initialement, seule la connexion des appareils grand public à Internet avait été envisagée, mais le développement en ce sens est resté limité. Nos réfrigérateurs n'iront probablement pas de sitôt sur Internet pour commander du lait frais, mais les téléviseurs, les systèmes d'alarme et les thermostats sont de plus en plus reliés à des réseaux. Ces appareils génèrent des données dont le traitement peut servir à optimiser l'utilisation, la consommation et la fourniture des services.

Des capteurs sont également branchés sur des êtres vivants, animaux ou personnes. Par exemple, Roger Parks, directeur des systèmes d'information chez J.R. Simplot, une société agricole et d'élevage, emploie l'expression « vache numérique » pour faire référence à une expérimentation qui consiste à placer des capteurs dans l'estomac des vaches de façon à mesurer leur température. Si l'animal est malade, le capteur permet au vétérinaire d'être averti du problème lorsqu'il est encore temps de traiter l'affection. D'autres recherches visent à utiliser des capteurs pour détecter la présence de la bactérie *E. coli* chez ces animaux.

Les êtres humains sont également de plus en plus « équipés de capteurs », en général dans le domaine de la santé et de la forme physique. Nous sommes entrés dans l'ère de l'auto-analyse, c'est-à-dire de la collecte et de l'analyse des données de performance, de forme et de santé des personnes [6]. En 2006, une première mise en application associait les baskets

Nike+ à l'iPod. Les chaussures, connectées à l'iPod, collectaient des informations et affichaient la durée de la course, la distance parcourue, le rythme cardiaque et le nombre de calories brûlées. Nike et Apple ont ensuite proposé plusieurs solutions de capture de données lors d'un exercice physique (capteurs de rythme cardiaque, liaison à des appareils de cardio-fitness, capteurs incorporés à des vêtements, etc.). Les possesseurs de chaussures Nike+, plus de 5 millions, stockent fréquemment leurs données sur leur ordinateur et les envoient sur le site web Nike+. Ils étudient leur progression, défient des amis et bénéficient de conseils d'entraînement. Depuis 2006 et Nike+, un grand nombre d'autres systèmes d'analyse physique sont apparus, principalement dans le domaine sportif. Par exemple, Garmin Connect permet aux athlètes dans différentes disciplines de consigner leurs activités, de planifier de nouveaux parcours et de partager leurs performances. Zed9 se focalise sur la mise en forme sociale, CycleOps suit les performances des cyclistes et Concept2 gère les entraînements des rameurs.

Mais les analyses personnelles vont au-delà du cadre sportif. Elles sont aujourd'hui présentes dans le domaine de la santé, de la richesse, du travail et de la satisfaction individuelle au sens large. Withings propose des balances connectées en Wi-Fi et à Twitter; d'autres appareils de santé, comme les tensiomètres, vont suivre. Différentes sociétés, dont MyZeo, WakeMate, BodyMedia et Fitbit, proposent d'analyser votre sommeil. Me-trics, une startup du Michigan, développe un outil général de collecte et de supervision des données personnelles dans quasiment n'importe quel aspect de la vie, comme la santé, l'humeur, le budget, la forme physique, l'activité en ligne, etc. Le jour où je me suis connecté à son site, Marie faisait le suivi de ses rencontres amoureuses et Ryan, celui de sa consommation d'eau. Il ne serait pas surprenant que des capteurs adaptés à ce type de mesures apparaissent bientôt.

Si ces données sur l'activité humaine ne suffisent pas, les smartphones nous accompagnent en permanence et enregistrent des informations sur nos déplacements, nos conversations et, de plus en plus, nos achats. Les ordinateurs sont capables de suivre quasiment n'importe quel aspect de notre vie professionnelle, comme Stephen Wolfram, directeur général de Wolfram Research, l'a vérifié pour son propre poste [7]. Il peut ainsi savoir précisément à quel moment il a rédigé l'un de ses 300 000 courriers électroniques envoyés depuis 1989. Nos animaux domestiques sont également dotés de capteurs. Peut-être que nos humeurs et nos ondes cérébrales seront bientôt capturées et analysées. Toutes ces sources de données existent, mais la principale sera sans aucun doute « l'Internet industriel », c'est-à-dire le très grand nombre d'appareils connectés dans les entreprises, les systèmes de transport, les réseaux d'énergie, etc. GE (General Electric) y voit un développement révolutionnaire, tant au niveau des volumes de données produites que des bénéfices attendus. Il estime ainsi que la supervision des turbines à gaz dans la production d'énergie va générer à elle seule 588 Go par jour – sept fois le volume des informations issues quotidiennement de Twitter.

Si ces sources massives de données ne suffisent pas à motiver une entrée dans le monde du Big Data, les fournisseurs de logiciels s'en chargeront, ou peut-être qu'ils y seront poussés par leurs clients. Ils mettront parfois du vieux vin dans les nouvelles bouteilles du Big Data, mais il est incontestable que la nature même des logiciels d'entreprise fait l'objet de profonds changements. Nous passons de l'automatisation des transactions à l'analyse des données qu'elles produisent. Lorsque SAP tire plus de revenus de l'informatique décisionnelle et de l'analytique que de sa suite d'applications transactionnelles, c'est qu'une transition majeure a eu lieu. Lorsque des entreprises comme HP, EMC et Oracle procèdent à des acquisitions et à des annonces de produits majeures dans

le domaine du Big Data et de l'analytique, il est probable que des choses sont en cours. Lorsque IBM dépense près de 20 milliards de dollars dans des rachats dans le secteur de l'analytique, on peut penser que le changement est là. Même Microsoft, une société créée pour gérer les petites données sur les ordinateurs personnels, a annoncé plusieurs produits pour le Big Data. Les technologies Big Data seront examinées plus loin dans cet ouvrage, principalement au Chapitre 5, mais, à ce stade, l'arrivée de tous ces grands acteurs dans le Big Data devrait au moins vous convaincre de la pérennité de ce concept.

Plus de données, plus de technologies. Que faut-il d'autre pour être certain que le Big Data ne soit pas un simple feu de paille ? Nous le verrons tout au long de cet ouvrage, ce sont les *personnes* qui vont donner au Big Data sa véritable mesure. Le travail du *scientifique des données* (voir le Chapitre 4) constitue le principal facteur de réussite du Big Data dans une entreprise. Les données sont souvent gratuites ou bon marché, les éléments matériels et logiciels sont gratuits ou peu coûteux, mais les personnes compétentes coûtent cher et sont difficiles à trouver. Le Chapitre 4 détaillera l'origine de ces difficultés, mais sachez qu'elles vont avoir tendance à diminuer. De nombreuses universités ou écoles proposent à présent un cursus en informatique décisionnelle ou en analytique. Plusieurs ajoutent le Big Data et les connaissances associées aux formations existantes. L'arrivée sur le marché d'étudiants qualifiés ne devrait plus tarder, ce qui va faciliter la réalisation des projets de Big Data au sein des entreprises et éviter qu'une pénurie de main-d'œuvre ne fasse obstacle à son développement. Pour que ses concepts disparaissent, il faudrait que les entreprises arrêtent de vouloir gagner de l'argent, vendre plus de produits et services, et satisfaire leurs clients. Il est peu probable que cela se produise.