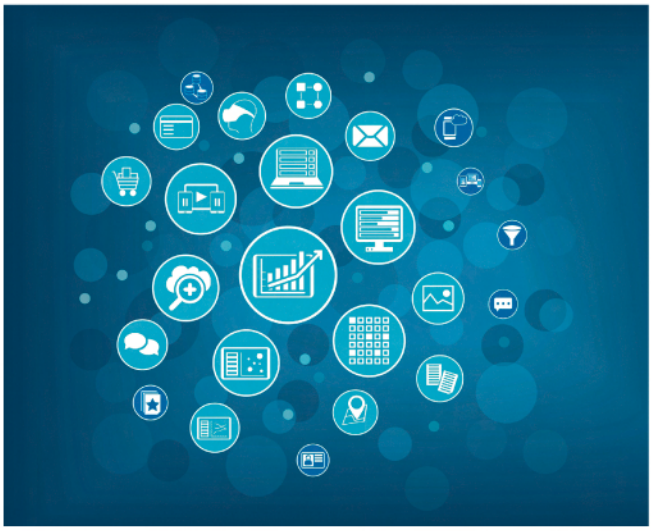
INFORMATION NUMÉRIQUE
Traitement, interprétation, communication

Christophe Brasseur

Enjeux et usages du big data

2e édition



Lavoisier
Hermes

Direction éditoriale : Emmanuel Leclerc
Édition : Céline Poiteaux
Fabrication : Estelle Perez
Composition : Nord Compo, Villeneuve-d'Ascq
Image de couverture : iconimage – Fotolia.com

ISBN : 978-2-7462-4758-1

Collection
Information numérique

Traitement, interprétation, communication

dirigée par Olivier Rioul

Professeur, Télécom ParisTech,
Université Paris-Saclay, Paris

Pour ma femme Christine et mes enfants Arthur et Capucine.

Table des matières

Chapitre 2
Enjeux du big data et du digital

Chapitre 3
Exemples d'usage du big data

Chapitre 5
Méthodes et techniques d'analyse du big data

Partie 3 — Comment tirer parti du big data ?

Chapitre 6
Compétences et ressources humaines

Introduction

Les mégadonnées au cœur de la révolution digitale

L'utilité des systèmes d'information dans les entreprises n'est plus à démontrer. Pratiquement aucune entreprise ne peut aujourd'hui s'en passer. L'activité quotidienne des banques, des assurances, des sociétés industrielles, des entreprises de distribution ou des services publics dépend du bon fonctionnement du système d'information. Celui-ci facilite et optimise les processus métier de l'entreprise, et pratiquement toutes les fonctions (comptabilité, contrôle de gestion, marketing, vente, production, achats, ressources humaines, qualité, maintenance, recherche) sont concernées.

On oublie parfois que la matière première du SI est constituée de données devenues indispensables à la vie et au développement des organisations. Il est clair que les données sont un facteur clé de succès et les entreprises qui en font bon usage gagnent en compétitivité. Des données fiables permettent notamment de prendre de bonnes décisions.

Depuis quelques années, on assiste à une prise de conscience de plus en plus importante de l'avantage compétitif que l'on peut obtenir avec des données de qualité totalement maîtrisées. Il reste encore du chemin à parcourir, mais ce mouvement vers la reconnaissance de la valeur des *data* est une bonne nouvelle. Les initiatives des entreprises pour améliorer la qualité des données notamment *via* des référentiels et des solutions MDM (*master data management*) sont des signes positifs de cette évolution. Et il était temps !

Car l'histoire des données est loin d'être terminée. Un autre phénomène relativement récent est venu se greffer aux problématiques de qualité de données : il s'agit de l'explosion du volume de données appelée plus communément « big data », terme anglais traduit depuis peu en français par celui de « mégadonnées ». Le développement

spectaculaire d'internet, et en particulier celui du web social (réseaux sociaux, blogs), la généralisation des intranets, l'essor de la mobilité avec les téléphones mobiles et les tablettes, la multiplication des capteurs et des objets connectés provoquent une avalanche de données à laquelle les organisations doivent faire face.

En conséquence, les entreprises sont submergées de données dont les volumes se chiffrent désormais en dizaines de téraoctets et parfois même en pétaoctets. Une bonne partie de ces nouvelles sources d'informations ne sont pas structurées, et les fournisseurs de bases de données et de solutions d'intégration se sont adaptés à cette nouvelle situation.

Ces dernières années, un énorme *buzz* autour du big data s'est développé à tel point qu'un certain nombre d'analystes n'y ont vu qu'un phénomène marketing de plus, visant à favoriser les ventes des fournisseurs de technologie. Les protagonistes des systèmes d'information, à commencer par les entreprises utilisatrices, s'aperçoivent à présent que le phénomène est bien réel. Et les enjeux sont de taille, car le big data n'est pas qu'une question technique de volumétrie et de stockage. Il constitue au contraire l'opportunité de comprendre le contenu de ces nouvelles sources et d'en tirer profit. La valeur des données non structurées issues des réseaux sociaux en est le parfait exemple. Quel service marketing n'a pas rêvé de connaître avec précision les sentiments des consommateurs à propos de ses produits ou de ses marques ?

Encore récemment, les moyens d'exploiter de telles volumétries de données, structurées ou non, étaient inexistants. La technologie est aujourd'hui disponible, les bases de données ont évolué et les solutions techniques dédiées à l'exploitation des données massives sont opérationnelles. La *business intelligence* laisse peu à peu place au *business analytics et aux algorithmes d'intelligence artificielle tels que le machine learning*. Ces solutions s'appuient sur des modèles mathématiques et des algorithmes pointus permettant entre autres de faire de l'analyse prédictive. Grâce à ces progrès remarquables, l'avalanche de données se transforme en valeur ajoutée, et les entreprises peuvent ainsi accroître leurs performances, être à la fois plus proactives et plus compétitives. En pratique, les entreprises sont de plus en plus nombreuses à mettre en œuvre des programmes de transformation digitale incluant un important volet big data.

Le big data ne se réduit pas à une problématique unique ; il se décline au contraire en différentes variantes selon le métier et la situation particulière de chaque entreprise. Les techniques doivent donc être choisies et intégrées en fonction de la situation. Les projets big data ont cependant des caractéristiques communes dans leur approche et leur méthode de mise en œuvre. Le présent ouvrage développe aussi cet aspect essentiel à la réussite des projets.

1. Objectifs de l'ouvrage

L'ouvrage a pour objectifs d'expliquer les enjeux du big data, de présenter différents cas d'usage et de détailler les méthodes et techniques utilisées. Il présente aussi les principes de mise en œuvre d'un projet big data avec les recommandations associées.

Cette seconde édition prend en compte les principales évolutions du big data, et approfondit les nouveaux enjeux numériques parmi lesquels figurent l'expérience client, les réseaux sociaux, les objets connectés, le *cloud* et l'analytique. Les exemples et les cas ont notamment été renouvelés pour coller aux préoccupations actuelles des entreprises. Enfin, l'ouvrage met l'accent sur les méthodes, les techniques et les ressources nécessaires pour permettre aux entreprises d'entrer avec succès dans l'ère de l'information à grande échelle.

Le livre s'adresse aux dirigeants d'entreprise, aux managers opérationnels et aux professionnels des systèmes d'information (responsables informatiques, directeurs et chefs de projets, consultants, ingénieurs). Les entreprises ne peuvent plus ignorer le phénomène du big data et il est essentiel que leurs responsables soient au fait des enjeux, des usages et des pratiques.

Le livre s'adresse aussi aux chercheurs et aux scientifiques confrontés à des volumétries de données de plus en plus importantes dans le cadre de leurs travaux. Outre le monde de l'entreprise, le big data concerne en effet des domaines divers tels que la recherche biomédicale, la météorologie, l'éducation, la criminologie et d'une façon plus générale la science, dont les applications et/ou les expérimentations nécessitent de plus en plus de données. Ainsi, l'informatique cognitive (*cognitive computing*) est un domaine prometteur s'appuyant sur des données volumineuses.

Cet ouvrage s'adresse aux étudiants des écoles de commerce, d'ingénieurs et des universités soucieux de comprendre les problématiques du big data auxquelles ils seront confrontés, de près ou de plus loin.

2. Organisation de l'ouvrage

Partie 1
Origines et enjeux du big data

- Chapitre 1 — Qu'est-ce que le big data ?
- Chapitre 2 — Enjeux du big data et du digital
- Chapitre 3 — Exemples d'usage du big data

Partie 2
Technologies et méthodes du big data

- Chapitre 4 — Technologies du big data
- Chapitre 5 — Méthodes et techniques d'analyse

Partie 3
Comment tirer parti du big data ?

- Chapitre 6 — Compétences et ressources humaines
- Chapitre 7 — Gérer un projet big data

La première partie, « Origines et enjeux du big data », permet de comprendre la problématique du big data :

– le chapitre 1, « Qu'est-ce que le big data ? », rappelle le rôle stratégique des données et définit ce qu'est le big data. Les différents types de données sont présentés, de même que les différents acteurs concernés ;

– le chapitre 2 présente les enjeux du big data et du digital. La création de valeur est au cœur de ce chapitre qui s'intéresse à la transformation digitale des organisations et à l'expérience client. Le chapitre traite aussi les sujets connexes que sont la qualité des données, la confidentialité et l'ouverture des données publiques. Le chapitre se termine par l'analyse des perspectives du big data ;

– le chapitre 3, « Exemples d'usage du big data », propose plusieurs cas pratiques. Les possibilités du big data sont illustrées au travers d'exemples concrets appliqués au marketing, à la santé, à l'énergie, aux services publics, à l'assurance et au *data journalism*.

La seconde partie présente les technologies et les méthodes utilisées pour traiter et analyser les mégadonnées :

– le chapitre 4 décrit « Les technologies du big data ». La problématique technique et les différentes solutions du marché sont présentées. Un focus spécifique sur les technologies Hadoop et Spark est réalisé. Les principaux fournisseurs de solutions sont également listés dans ce chapitre ;

– le chapitre 5, « Les méthodes et techniques d'analyse du big data », dresse un panorama des principales méthodes et techniques du *business analytics*. Les techniques de visualisation sont également développées dans ce chapitre.

La troisième partie fournit les clés permettant de tirer parti du big data :

– le chapitre 6, « Compétences et ressources humaines », s'intéresse aux ressources et profils nécessaires d'une part pour réaliser un projet big data et, d'autre part, pour utiliser les modèles et outils mis en œuvre. L'organisation des ressources est également abordée dans ce chapitre ;

– le chapitre 7, « Gérer un projet big data », précise les caractéristiques propres à ce type de projet, et définit la démarche de mise en œuvre. Des recommandations et les risques à prendre en compte complètent ce chapitre pour amener les projets vers le succès.

Le développement spectaculaire d'internet, des réseaux sociaux, de la technologie mobile et des objets connectés provoque une croissance exponentielle des données à laquelle toutes les entreprises sont confrontées : c'est le phénomène du big data.

Cette nouvelle édition de *Enjeux et usages du big data* prend en compte les principales évolutions du big data et approfondit les nouveaux enjeux numériques parmi lesquels figurent l'analytique, l'expérience client, le *cloud*, les réseaux sociaux et les objets connectés. Les exemples et les cas ont été entièrement renouvelés afin d'être au plus près des préoccupations actuelles des entreprises. L'ouvrage met également l'accent sur les méthodes, les techniques et les ressources nécessaires pour permettre aux entreprises d'entrer avec succès dans l'ère de l'information à grande échelle.

Cet ouvrage concis et didactique s'adresse aux dirigeants d'entreprise, aux managers opérationnels et aux professionnels des systèmes d'information, ainsi qu'aux étudiants des écoles de commerce, d'ingénieurs et des universités, soucieux de comprendre les problématiques et les applications du big data.

Christophe Brasseur, *ingénieur diplômé de l'ESTP et de l'IAE, travaille depuis plus de 25 ans dans les systèmes d'information. Il a participé à de nombreux projets internationaux dans des postes de conseil et de management, principalement dans le secteur des services publics, de l'industrie et de l'énergie. Il est actuellement Senior Manager chez Capgemini.*

editions.lavoisier.fr

978-2-7462-4758-1