

Gauthier Vasseur

Devenez  
un **data**  
pionnier !

Comprendre et exploiter  
les données en entreprise

MARDAGA



## Adapté des classes de l'auteur

« Become a Data CEO » du programme Continuing Studies Program  
de Stanford University

« Seven keys to Master Data » de L'Alliance for Inclusive AI  
à L'Université de Berkeley en Californie

© 2021 Éditions Mardaga  
B-1160 Bruxelles (Belgique)  
[www.editionsmardaga.com](http://www.editionsmardaga.com)

Coordination éditoriale : Véronique Dupas et Laura Vanham  
Design de couverture : Emmanuel Bonaffini  
Maquette intérieure : Carine Thurion – CW Design

Dépôt légal : 2021/0024/85  
ISBN papier : 9782804720056  
ISBN numérique : 9782804720339

Toute reproduction ou représentation intégrale ou partielle du présent ouvrage,  
par quelque procédé que ce soit, est strictement interdite.

Achévé d'imprimer en France en août 2021  
sur les presses de l'imprimerie Dupliprint à Domont (95).



À Marie



# SOMMAIRE

Avant-propos	9
Introduction	13
<b>CHAPITRE 1</b>	
Les racines de la transformation digitale	17
<b>CHAPITRE 2</b>	
Commencer sur de bonnes bases	29
<b>CHAPITRE 3</b>	
Comprendre la donnée dans l'entreprise	53
<b>CHAPITRE 4</b>	
Mettre les systèmes au service du data pionnier	65
<b>CHAPITRE 5</b>	
Maximiser le potentiel de la donnée	125
<b>CHAPITRE 6</b>	
Relever le défi humain de la data	225
<b>CHAPITRE 7</b>	
Construire des processus efficaces, légers et durables	249
Conclusion	279
<b>ANNEXES</b>	
Témoignages de data pionniers et paroles d'experts	281
Remerciements	321
Notes	323
À propos de l'auteur	325
Table des matières	329



# TABLE DES MATIÈRES

SOMMAIRE	7
AVANT-PROPOS	9
INTRODUCTION	13
<b>CHAPITRE 1</b>	
<b>LES RACINES DE LA TRANSFORMATION DIGITALE</b>	17
Les défis d'un monde en constante évolution	18
Le darwinisme digital	19
Les clés de la survie	20
La data est-elle vraiment le nouvel or noir ?	22
Trois illustrations non conventionnelles des bénéfices de la maîtrise de la data	25
Protéger ses dirigeants de la prison ( <i>Keep your execs out of jail</i> )	25
Préparer sa survie aux cygnes noirs	26
Faire de la data un non-sujet	26
Comprendre la data dans sa globalité	27
Sommes-nous prêts ?	27
<b>CHAPITRE 2</b>	
<b>COMMENCER SUR DE BONNES BASES</b>	29
Pourquoi est-ce capital de poser une question pour commencer ?	29
Pourquoi est-ce difficile de poser une question ?	30
L'obstacle des biais cognitifs	30
> <i>L'aversion à l'ambiguïté ou à l'incertitude</i>	31
> <i>L'ancrage</i>	31
> <i>L'argumentum ad novitatem ou ad antiquitatem</i>	31
La perception des données disponibles	38
La compréhension des mécaniques analytiques	39
La pression de l'habitude	40

## Devenez un data pionnier!

Le courage (ou son manque)	41
La tentation du voisin et de la bonne pratique	41
Comment poser la bonne question	43
Mettre la question au service de l'action	50
Pour conclure	51

## CHAPITRE 3

### COMPRENDRE LA DONNÉE DANS L'ENTREPRISE 53

Qu'est-ce que la donnée pour l'analytique en entreprise? 53

Comment appréhender la data 54

La data, c'est « tout » 54

Identifier quelle data est importante 55

Collecter une donnée pour une utilisation efficace,  
précise et durable 57

Qu'est-ce qu'une base de données? 58

Les fichiers plats: le plus simple des formats FAIR 60

> *Un média quasiment sans limite en termes de volume* 62

> *Un format plus performant au chargement et  
qui se compresse mieux* 63

> *Un format d'échange universel* 63

> *Un format sur lequel on peut travailler directement* 63

Les trois règles cardinales en analytique 64

## CHAPITRE 4

### METTRE LES SYSTÈMES AU SERVICE DU DATA PIONNIER 65

Quelle est l'utilité de la technologie en data? 65

Les solutions transactionnelles 69

Une brève histoire de l'ERP 71

Les forces des ERP sont leurs faiblesses pour l'analytique 71

L'analytique transactionnelle n'est pas l'analytique décisionnelle 73

> *La naissance de monstres* 73

> *La mutation en escargot* 74

> *La création de cocktails explosifs* 75

> *Le péché cardinal de la requête sans fin* 76

Les nouveaux besoins analytiques	76
Le cycle de management	76
La naissance de la Business Intelligence	78
Comprendre les solutions BI ou décisionnelles	80
Le Middleware	81
> Récupérer et préparer la donnée	82
> Les solutions de Master Data Management	85
> Les solutions de gestion de la qualité des données	85
> Pour conclure	86
La base de données	88
> L'émergence difficile des bases vraiment décisionnelles	88
> Les qualités intrinsèques d'une bonne base décisionnelle	89
> La place du Cloud dans le traitement des données	93
Les solutions d'analyse	94
> La présentation de la donnée	94
> Les calculs finaux	95
> La création de mini-processus analytiques en local	96
Le reporting	96
> À quoi sert le reporting ?	97
> Les grandes familles de reporting	99
> Les grands formats de reporting	101
Pour conclure	119
Comment passer des solutions transactionnelles aux solutions décisionnelles ?	119
L'accès à l'équipement	119
Le frein organisationnel	121
Le frein humain	122
Pour conclure	124

## CHAPITRE 5

### MAXIMISER LE POTENTIEL DE LA DONNÉE 125

Les types de bases de données	125
Les fichiers plats, texte ou CSV	126
Les bases de données relationnelles	128
Les cubes de données	130
Les bases de données en colonnes	132
Les bases de données NoSQL (Not Only SQL)	133
Les bases de données décisionnelles virtualisées	133
Pour conclure	134

Organiser la donnée pour une meilleure analytique	136
Les fondamentaux pour comprendre la donnée	136
> Identifier les formats de la data	136
> Identifier les rôles de la data	139
> Connaître ou définir les règles d'agrégation de la data	141
> Pour conclure	144
Jauger la qualité de données initiale : profiler la donnée	146
> La revue des clés existantes	147
> L'audit des attributs	149
> Le contrôle des métriques	150
> La validation des calculs	152
> Les fonctions automatisées de profilage (profiling)	152
> Pour conclure	155
Connecter la data pour une analytique plus riche et pertinente	156
> La notion de jointure	156
> La cardinalité d'un modèle	156
> Les types de jointures	159
> Les résultats de jointures	161
> Pour conclure	165
Tirer le meilleur de la donnée avec une bonne préparation	172
Ajouter des tables de groupements ou des tables de hiérarchies	173
Créer des clés pertinentes pour de nouvelles jointures	176
> Se conformer aux clés du transactionnel	176
> Ne pas générer des clés très longues	176
> Ne pas choisir des clés numériques qui commencent par 0	176
> Éviter les clés sans format standard	177
> Choisir des clés universelles ou ISO quand c'est possible	177
Optimiser les analyses par les données calculées/préparées	178
> La préparation de calculs complexes	178
> Les agrégations et les réductions du nombre de colonnes	179
> Les corrections de format	181
> Les snapshots, les sauvegardes d'instantanés	182
Les options pour travailler la donnée : à la volée ou stockée	183
Optimiser le choix du moteur de calcul	185
Augmenter la valeur des données	187
Gérer les données maîtres: le Master Data Management (MDM)	187
> Définition	187
> Quelques exemples de problèmes avec des données maîtres mal maîtrisées	188
> Le processus de Master Data Management	197

Les subtilités et les risques du Master Data Management	201
> <i>Pareil ne veut pas dire le même</i>	201
> <i>Peu d'attributs ne veut pas dire anonymat</i>	201
> <i>Processus simple ne veut pas dire évident</i>	202
Établir une qualité holistique de la donnée	203
> <i>Jouer en mode Street Fighter Style</i>	203
> <i>Se concentrer sur ses besoins</i>	205
> <i>Ne jamais s'arrêter</i>	206
> <i>Se souvenir que c'est la bonne décision qui compte</i>	206
<b>Le Big Data: ses mythes, sa réalité</b>	206
Comment définir le Big Data ?	207
> <i>Pourquoi volume et vitesse ont-ils été si moteurs au départ ?</i>	208
> <i>A-t-on vraiment besoin de volume pour être pertinent ?</i>	209
> <i>L'obsession illusoire de la vitesse</i>	211
> <i>Pourquoi cette ruée vers les volumes et la vitesse ?</i>	212
> <i>La variété, le Graal du Big Data</i>	212
Les domaines du Big Data	214
> <i>Qu'est-ce que l'intelligence artificielle ?</i>	215
> <i>Qu'est-ce que le Machine Learning ?</i>	215
> <i>Qu'est-ce que le Deep Learning ?</i>	215
Trois grandes approches algorithmiques à connaître pour mieux travailler avec les data scientists	217
> <i>Les régressions</i>	217
> <i>Les classifications</i>	219
> <i>Le clustering</i>	220
Les biais en algorithme	222
Pour conclure	223
<b>CHAPITRE 6</b>	
<b>RELEVER LE DÉFI HUMAIN DE LA DATA</b>	225
Le Bad Buzz de la data	227
L'effet Dunning Kruger de l'analyste	229
La courbe de changement digital	231
L'annonce	232
La descente irrationnelle	233
> <i>Le choc</i>	233
> <i>Le déni, le refus</i>	233
> <i>La colère ou la panique</i>	234

## Devenez un data pionnier!

La transition émotionnelle	235
> <i>La dépression</i>	235
> <i>La justification</i>	236
Le retour au rationnel	236
> <i>La négociation</i>	236
> <i>L'engagement</i>	237
Le YBG-IBG (JSP-TSP en français)	238
L'obstacle du management	239
Les motivations	239
La transparence	240
Responsabilité et culpabilité	240
Les dynamiques d'équipe	241
Les visages de l'opposition	241
> <i>Geoffrey Effraie</i>	241
> <i>Claire Excel</i>	241
> <i>Larry Big Data</i>	243
> <i>Rose Royce</i>	243
> <i>Anton Python</i>	244
Nos superpouvoirs	245
La force de notre humanité et de la diversité	247
Le vrai visage du data scientist	248
Les nouveaux champs de développement	248

## CHAPITRE 7

CONSTRUIRE DES PROCESSUS EFFICACES, LÉGERS ET DURABLES	249
Le défi des processus <i>lean</i>	249
Chasser les processus <i>Creeper</i>	252
La valeur ajoutée des étapes du processus analytique	253
L'optimisation des neuf étapes d'un processus analytique	253
> <i>Neuf étapes vers le data nirvana</i>	253
> <i>Les appuis techniques et méthodologiques à chaque étape</i>	256
L'accélération par le RPA ( <i>Robot Process Automation</i> )	265
L'optimisation des processus analytiques	267
L'excellence est dans la répétition	267
> <i>Répéter pour apprendre</i>	267
> <i>Répéter pour optimiser</i>	267
> <i>Répéter pour se tromper</i>	268
L'organisation des contrôles et les check-lists	269

Extraire les trésors de l'intelligence collective	270
La dynamique de collaboration	270
Le cercle vertueux Analytiques terrain – Informatique – Data Science	275
> <i>Un passé d'initiatives déconnectées</i>	275
> <i>Le Pionnier catalyseur</i>	276
> <i>Les équipes data et informatiques libérées</i>	277
> <i>La Data Science impactante</i>	278
Vers une intelligence de l'essaim	278
CONCLUSION	279
<b>ANNEXES</b>	
TÉMOIGNAGES DE DATA PIONNIERS ET PAROLES D'EXPERTS	281
REMERCIEMENTS	321
NOTES	323
À PROPOS DE L'AUTEUR	325