

Data Management

qualité des données et compétitivité

Christophe Brasseur

الطبعة الأولى 2010

Hermes
Science
—publications—

COLLECTIONS SOUS LA DIRECTION DE NICOLAS MANSON

Collection Management et Informatique

Collection Etudes et Logiciels Informatiques

Collection Nouvelles Technologies Informatiques

Collection Synthèses Informatiques CNAM

La liste des titres de chaque collection se trouve en fin d'ouvrage.

TABLE DES MATIÈRES

Introduction	13
PREMIÈRE PARTIE. SYSTÈMES D'INFORMATION ET DONNÉES	19
Chapitre 1. L'information, matière première vitale de l'entreprise	21
1.1. L'évolution des systèmes d'information	21
1.2. Information et données	22
1.3. Les données : pour faire quoi et pour qui ?	23
1.3.1. Les contraintes légales	24
1.3.1.1. La publication des comptes	24
1.3.1.2. Le système d'information comptable	26
1.3.2. Les besoins opérationnels	27
1.3.3. Les besoins d'analyse et d'aide à la décision.	28
1.4. Différents types de données	29
1.4.1. Le texte	29
1.4.2. Les dessins vectoriels	30
1.4.3. L'image matricielle	31
1.4.4. La vidéo numérique	32
1.4.5. Le son	33
1.4.5.1. La numérisation du son.	33
1.4.5.2. La compression du son	36
1.5. La structuration des données.	36

1.5.1. Evolution de la technologie et du stockage . . .	37
1.5.2. Prolifération des sources de données	38
1.5.3. Les bases de données	39
1.5.3.1. Structuration de l'information	39
1.5.3.2. Stockage centralisé et non-redondance des informations	42
1.5.3.3. Indépendance des données et des traitements	42
1.5.3.4. Sécurité des données	43
1.5.4. Le langage XML (<i>eXtensible Markup Language</i>)	44
1.6. Conclusion	46

Chapitre 2. Tour d'horizon des principales tendances informatiques 47

2.1. Les progiciels de gestion intégrés (ERP)	47
2.1.1. Les raisons de l'essor des ERP	48
2.1.2. Les facteurs-clés de succès d'un projet ERP . .	49
2.2. Le commerce électronique (<i>e-business</i>)	50
2.2.1. L'expansion du commerce électronique	51
2.2.2. Les raisons du succès	52
2.3. Les systèmes d'aide au pilotage (<i>Business Intelligence</i>)	53
2.3.1. La définition des indicateurs	53
2.3.2. Collecte, stockage et restitution des informations	54
2.3.3. Principaux domaines d'application	55
2.4. Les systèmes de gestion de la relation client (CRM)	56
2.4.1. La connaissance des clients	56
2.4.2. L'optimisation de la relation client	57
2.5. Quelques mots sur les réseaux informatiques.	58
2.5.1. Différents types de réseaux	59
2.5.2. Le routage.	60
2.5.3. La commutation.	60
2.5.4. Les réseaux sans fil.	61
2.6. Conclusion	62

DEUXIÈME PARTIE. POURQUOI CHERCHER À OBTENIR DES DONNÉES DE QUALITÉ ?	63
Chapitre 3. La qualité appliquée aux données	65
3.1. L'évolution de la qualité	65
3.2. La qualité appliquée aux données	66
3.2.1. Qualité du contenu	68
3.2.1.1. Justesse de l'information.	68
3.2.1.2. Adéquation aux besoins	68
3.2.1.3. Facilité d'interprétation.	68
3.2.2. Accessibilité	69
3.2.2.1. Disponibilité	69
3.2.2.2. Facilité d'accès	69
3.2.3. Flexibilité	70
3.2.3.1. Evolutivité	70
3.2.3.2. Cohérence avec d'autres sources	70
3.2.3.3. Possibilités de traduction.	71
3.2.4. Sécurité	71
3.3. Conclusion	71
Chapitre 4. Causes et conséquences de la non-qualité des données.	73
4.1. Les conséquences de la non-qualité des données	73
4.2. Quelques exemples de défauts de qualité des données	75
4.2.1. Doublons	75
4.2.2. Coordonnées des partenaires	76
4.2.3. Montants facturés.	76
4.2.4. Disponibilité des produits	77
4.2.5. Stocks	78
4.2.6. Indicateurs d'activité.	78
4.2.7. Ergonomie	78
4.3. Sources de non-qualité des données	79
4.3.1. Sous-estimation de l'enjeu des données	80

4.3.2. Défauts de conception	81
4.3.3. Anomalies du logiciel	81
4.3.4. Initialisation des données.	83
4.3.4.1. L'initialisation des données au démarrage d'une application	84
4.3.4.2. La saisie des données au fil de l'eau	85
4.4. Avantages compétitifs d'une politique qualité de l'information et des données.	87
4.4.1. Réactivité	87
4.4.2. Connaissance des clients	88
4.4.3. Gestion opérationnelle optimisée	88
4.4.4. Pertinence des décisions	89
4.4.5. Maîtrise des coûts.	90
4.5. Conclusion	91
TROISIÈME PARTIE. COMMENT AMÉLIORER LA QUALITÉ DES DONNÉES ?	93
Chapitre 5. L'évaluation de la qualité des données	95
5.1. Mesure du taux de données erronées	95
5.1.1. Identifier une activité	96
5.1.2. Identifier un processus métier.	97
5.1.3. Sélectionner les données	98
5.1.4. Définir les critères à mesurer	98
5.1.5. Sélectionner un échantillon représentatif	98
5.1.6. Contrôler l'échantillon	99
5.1.7. Etablir une synthèse des résultats	100
5.1.8. Définir un plan d'actions	101
5.1.9. Mesurer l'impact des actions	101
5.2. Détection d'erreurs au cours d'un processus	101
5.3. Contrôle statistique	103
5.4. Conclusion	105
Chapitre 6. L'élimination des défaillances	107
6.1. L'approche globale	107
6.2. Identifier les causes d'anomalies	109

6.3. Définir une solution.	109
6.4. Lancer un projet d'amélioration.	110
6.5. Construire l'équipe projet.	111
6.6. Détailler et mettre en œuvre la solution.	112
6.7. Mesurer le niveau de qualité résultant.	112
6.8. Conclusion	113
Chapitre 7. Les bonnes pratiques de Data Management	115
7.1. La compréhension des besoins	115
7.2. La codification des données	117
7.3. La documentation des données	119
7.4. L'administration des données	120
7.4.1. L'architecture logique	121
7.4.2. L'organisation de la gestion des données	122
7.5. Le nettoyage des bases de données (<i>clean-up</i>)	125
7.5.1. Définition	125
7.5.2. Dans quels cas faut-il nettoyer les données ?	126
7.5.3. L'alternative au <i>clean-up</i>	128
7.5.4. Le cas des doublons (ou <i>n-uplets</i>)	130
7.5.4.1. Détection des doublons dans la base de données.	130
7.5.4.2. Contrôle des doublons lors de la saisie d'un nouvel enregistrement	131
7.6. La surveillance des données	132
7.7. La sécurité des données	132
7.7.1. Sécurité de la saisie des données.	132
7.7.2. Sécurité du traitement des données	134
7.7.3. Sécurité des accès à l'application	134
7.8. L'auditabilité des données	135
7.8.1. Les documents à conserver pendant trente ans.	136
7.8.2. Les documents à conserver pendant dix ans	136
7.8.3. Les documents à conserver pendant cinq ans	137
7.8.4. Les documents à conserver pendant trois ans	137
7.9. Conclusion	138

QUATRIÈME PARTIE. LA REPRISE DES DONNÉES	139
Chapitre 8. Stratégie de reprise des données	141
8.1. Définition de la reprise des données.	141
8.2. Différents types de données et mode de reprise . . .	143
8.3. Méthodologie de reprise des données.	144
8.3.1. La préparation de la reprise	145
8.3.1.1. Nettoyage des données	145
8.3.1.2. Spécification des règles de gestion	145
8.3.1.3. Définition des règles de conversion (<i>mapping</i>)	146
8.3.1.4. Détermination du mode de reprise (automatique/manuel)	146
8.3.1.5. Définition de la séquence de chargement (ou de saisie) dans le nouveau système	147
8.3.1.6. Développement des outils de conversion . .	147
8.3.1.7. Définition de la structure des fichiers de reprise.	147
8.3.2. Planification et réalisation de la reprise	148
8.4. Les outils de conversion de données.	150
8.4.1. L'extraction des données	151
8.4.2. La transformation des données	152
8.4.3. Le chargement des données	152
8.4.4. Quelques recommandations	152
8.5. Conclusion	153
Chapitre 9. Facteurs-clés de succès de la reprise des données.	155
9.1. Ressources et management.	155
9.2. Formation	157
9.3. Planification et suivi des tâches préparatoires à la bascule	158

9.4. Tests	158
9.5. Validation des données	159
9.6. Planification et suivi détaillé de la bascule	159
9.7. Conclusion	160
Conclusion	161
Bibliographie	163
Index	165