

# Master data management: Exploiter le Big Data



Nombreux sont ceux qui considèrent le Big Data comme la nouvelle révolution du système d'information de l'entreprise. Toutefois, ses quatre fondements rendent difficile l'exploitation de tout son potentiel :

1. **Volume** (quantité) : selon une étude de Digital Universe, 800.000 pétaoctets d'informations ont été enregistrés dans le monde en 2009. D'ici 2020, ce chiffre devrait atteindre 35 zettaoctets.<sup>1</sup> Un zettaoctet équivaut à un millier de milliards de gigaoctets.
2. **Variété** : le Big Data est souvent déstructuré. Exemple : commentaires textuels, images ou vidéos.
3. **Vélocité** : le débit du Big Data généré par les nombreux capteurs et autres appareils de mesure peut être 1000 fois supérieur au débit généralement accepté par l'informatique.<sup>2</sup>
4. **Véracité** : le Big Data peut s'avérer inexact et difficile à vérifier. Par exemple, vous pouvez savoir quels clients ont déposé sur Facebook des commentaires négatifs concernant votre restaurant, mais leurs réclamations peuvent être fausses ou exagérées.

Face à ces quatre caractéristiques de base, il faut disposer d'outils et de méthodes spécialisés pour gérer et exploiter le Big Data. Les données maîtres, le sous-ensemble le plus fiable de toutes les données, peuvent aider les entreprises à donner un sens à l'énorme volume de données disponible. Les données maîtres sont la base d'un référentiel d'informations partagé utilisé dans toute l'entreprise. Elles relèvent d'une entité de l'entreprise, des dossiers clients aux produits, et ont été triées et vérifiées pour constituer un "golden record" de cette entité. Le Master data management (MDM) englobe les outils et la discipline de gouvernance et d'exploitation des données maîtres.

En combinant le Big Data à des données maîtres vérifiées et fiables, l'entreprise peut obtenir de nouvelles connaissances sur ses clients et produits, ou en déduire immédiatement des informations pouvant découler sur des actions (voir Figure 1).

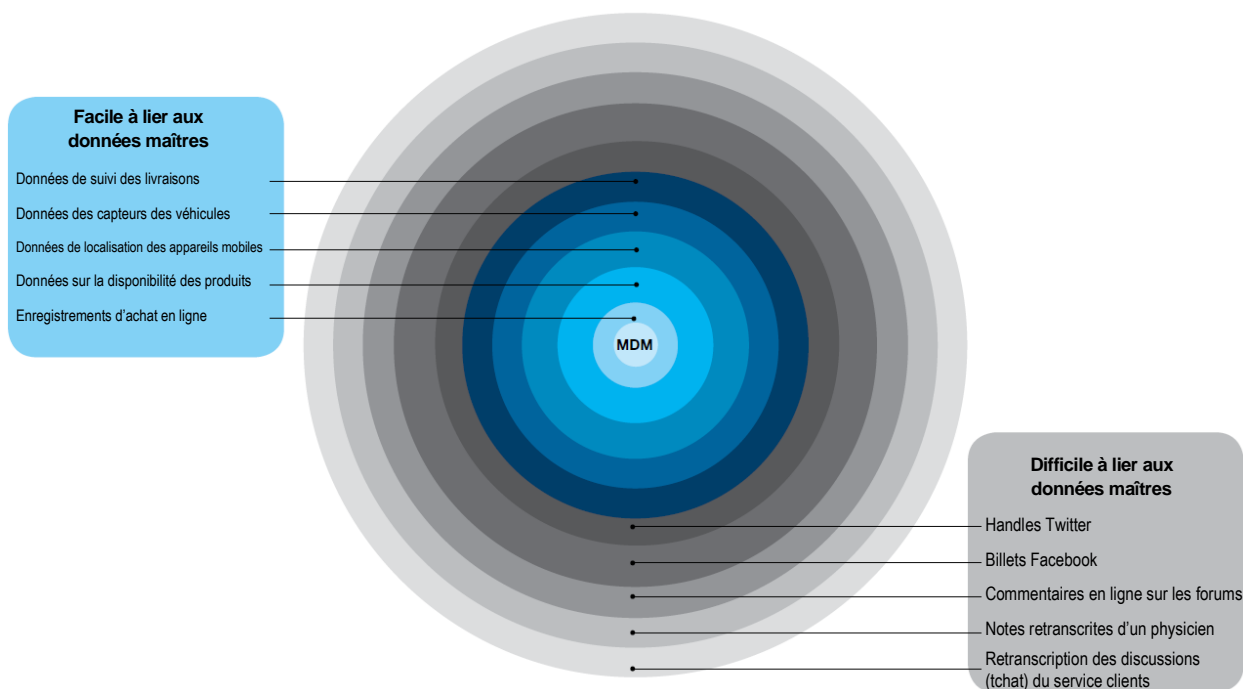


Figure 1: Le MDM aide à transformer le Big Data en données fiables grâce à la gouvernance et à la vérification.

## Gouvernance du Big Data : faire la liaison avec le MDM

Le Big Data peut faire l'objet d'une gouvernance, mais dans un espace-temps. Peu d'organisations, voire aucune, traiteront et analyseront tout le Big Data qu'elles accumulent. Celui-ci continuera d'augmenter et de se diversifier plus vite que les tentatives visant à le mapper et à l'analyser. Toutefois, en identifiant, classant et utilisant le plus possible de données, les organisations peuvent obtenir de précieuses connaissances dont elles n'auraient pas bénéficié sans le Big Data.

Les DSI qui s'efforcent d'utiliser efficacement le Big Data rencontreront probablement un problème de MDM, car les informations sur les clients enfouies dans le Big Data doivent être mises en relation avec les informations déjà capturées dans les applications structurées classiques. À moins de les lier aux données maîtres, les informations sur les nouveaux clients ou produits extraites du Big Data ne sont qu'un autre silo de données de plus.

### Augmenter la véracité du Big Data en le liant aux données maîtres

Une approche bien définie du MDM augmente la valeur du Big Data en lui donnant de la structure et du sens. Les outils de MDM peuvent établir des correspondances entre des produits et des commentaires laissés sur le Web et des messages de réseaux sociaux. Le MDM apporte également du contexte aux données obtenues par des machines, telles que les données fournies par des appareils mobiles sur les heures et lieux d'achat de certains produits.

Le MDM aide à améliorer la vue à 360 degrés du client en complétant les données maîtres de contenus non structurés. Par exemple, des messages Facebook peuvent indiquer à une chaîne d'hôtels qu'une forte proportion de ses clients professionnels ont trop d'enfants pour les chambres standard proposées à un tarif avantageux. L'hôtel peut alors proposer à ses clients prestigieux des avantages sur les suites plus grandes afin de renforcer leur fidélisation.

Le MDM peut aussi aider à mieux comprendre les clients grâce à des données structurées capturées à partir de flux de données très rapides. C'est notamment le cas des compagnies d'assurance qui évaluent les habitudes de conduites, bonnes et mauvaises, grâce aux informations données par des capteurs posés sur les véhicules, et les utilisent pour ajuster le montant de leurs polices d'assurance. Le MDM aide en outre les organisations à identifier des relations qui améliorent leur connaissance de leurs clients et leur apportent de nouvelles opportunités commerciales ou de faire des économies. En se rendant compte que deux clients distincts appartiennent au même foyer, l'entreprise peut par exemple éviter d'envoyer plusieurs courriers identiques à la même adresse.

Le Big Data est une ressource qui peut aider les organisations à identifier, comprendre et agir, en vue de résoudre des problèmes de grande envergure tels que la fourniture de services centrés sur le client, la fraude, l'efficacité des soins de santé et des offres promotionnelles individualisées en temps réel aux utilisateurs mobiles. En outre, en combinant avec exactitude le Big Data aux enregistrements maîtres d'un système de MDM, les organisations ont une plus grande confiance dans les informations qu'elles utilisent pour prendre des décisions stratégiques.

---

*“L'ère du Big Data donne la chance à tous les acteurs du monde technologique de décider [s'ils travaillent pour un salaire ou pour changer le monde] ...cette ère offrira aux entreprises et individus les meilleures opportunités depuis l'avènement d'Internet.”*

— Rob Thomas, vice-président, IBM Software Group<sup>3</sup>

---

### Augmenter la valeur des données non structurées

Sans approche stratégique du Big Data à l'échelle de l'entreprise, la croissance des données peut rapidement devenir incontrôlable. Le MDM aide à contrôler l'augmentation des données entrantes, et les outils qui appliquent le MDM au Big Data élargissent la portée des faits vérifiables concernant les clients, les fournisseurs et les produits.

Toute stratégie de Big Data devrait permettre avant tout à l'entreprise de gérer sa clientèle de façon plus rentable. Le MDM peut aider une chaîne hôtelière telle que TripAdvisor à faire la relation entre les avis sur ses hôtels (données non structurées) et certains profils de clients (données structurées). Cette chaîne d'hôtels peut alors savoir presque avec certitude que certains de ses clients haut de gamme ont un grand nombre de suiveurs sur Twitter, FourSquare or Facebook. Elle peut alors contacter ces clients par téléphone pour les remercier de leur visite et leurs proposer certains avantages soigneusement sélectionnés.

### Offre MDM de IBM WebSphere

Le logiciel IBM® InfoSphere® Master Data Management (InfoSphere MDM) est une solution de MDM complète, puissante et opérationnelle. Elle propose à la fois des fonctions de MDM orientée transaction et des capacités collaboratives de création et de workflow, pour répondre aux besoins de tous les domaines, styles d'implémentation et cas d'utilisation de différents secteurs. Un ensemble d'outils applicatifs basés sur le MDM apporte de la valeur métier en permettant de créer des applications de gouvernance grâce à des projets prédéfinis et à des widgets intégrables aux applications en place. Les modèles de données prédéfinis sont adaptables à tous les domaines et optimisés pour le MDM afin de produire rapidement des résultats. Les organisations peuvent aussi importer des modèles de données ou en concevoir de nouveaux. De plus, InfoSphere MDM est conçu pour gérer de grandes quantités de transactions ainsi que des environnements opérationnels à haute disponibilité.

---

### Meilleures pratiques pour augmenter la fiabilité des données non structurées avec la MDM.

Mettez en place des processus augmentant la valeur des grands ensembles de données, par exemple en motivant vos clients pour qu'ils mettent en correspondance leurs données non structurées et leurs données maîtres. Les outils logiciels risquent de rater la correspondance entre les ID Twitter et les dossiers clients, mais le client peut vous procurer une correspondance parfaitement exacte bien plus rapidement. Le MDM peut jouer un rôle important dans la gouvernance des données. Une approche réaliste de gouvernance du Big Data implique de :

- Faire concorder les éléments qui peuvent être mis en correspondance de façon fiable avec les données maîtres
  - Classer les données non structurées le plus exactement possible
  - Utiliser les métadonnées communes entre le Big Data et les initiatives de MDM
  - Déterminer les niveaux de qualité et de véracité acceptables pour votre organisation, et centrer les efforts sur le Big Data conforme à ces niveaux
  - Appliquer les mêmes règles de confidentialité et de sécurité des données maîtres au Big Data
-

## Combiner la puissance d'InfoSphere MDM et les capacités du Big Data

Plusieurs solutions IBM sont spécialisées dans l'analyse du Big Data pour aider les organisations à obtenir des connaissances jusqu'alors inaccessibles et à les utiliser pour les prises de décision dans toute l'entreprise (voir Figure 2). L'association de la puissance d'InfoSphere MDM et de ces solutions de Big Data IBM crée une relation précieuse : le Big Data peut apporter des connaissances au MDM, et celui-ci peut fournir des définitions de données maîtres au Big Data.

### Analyse du Big Data au repos

En analysant de grandes quantités de données conservées, les organisations peuvent identifier des schémas et perspectives qui leur permettent d'optimiser les processus et la rentabilité. Une mutuelle de santé a par exemple analysé plus de 400 millions de demandes de remboursement pour identifier des interactions médicamenteuses potentiellement mortelles.

Les appels téléphoniques des patients contiennent des références complexes et mentionnent souvent des médicaments obtenus d'autres sources ou sans ordonnance.

Un médicament courant tel que le Diflucan peut aussi être appelé Flucanor, Batacan ou traitement antifongique dans une transcription. Cette analyse a pris plus de 100 heures au système interne de l'entreprise.

Toutefois, avec IBM InfoSphere BigInsights™ pour la mise en correspondance des retranscriptions avec les enregistrements de données maîtres des traitements, l'analyse n'a pris que 10 heures pour plus de 6 To de données.<sup>4</sup> InfoSphere BigInsights excelle également dans l'extraction des données structurées à partir de l'analyse de textes et dans leur recouplement avec les données maîtres. Ces capacités ouvrent la voie à un nouveau type d'applications dédiées à l'analyse du ressenti du client ou des médias sociaux, et à la détection des fraudes.

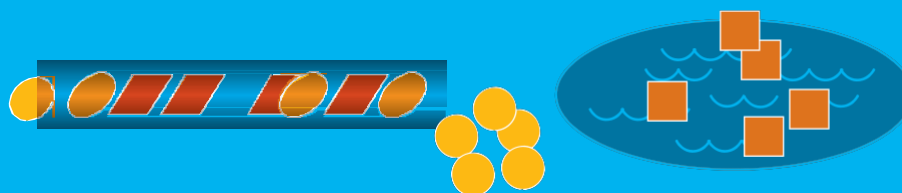
### InfoSphere BigInsights

Aide à trouver des pépites jusqu'ici cachées dans l'immensité des données, et les organise par taille, couleur et tout autre aspect, puis les répartit dans des seaux en fonction de leur type. Les seaux peuvent être déplacés vers différents lieux de stockage pour garantir leur protection.



### InfoSphere Streams

Examine les matériaux lorsqu'ils arrivent, et prélève les pépites pendant qu'elles circulent à haut débit et en grande quantité : elles sont ainsi capturées plus tôt dans le processus.



### InfoSphere Data Explorer

Identifie les types de pépites dont vous disposez et tous les endroits possibles. Il passe au tamis tous ces endroits pour attraper exactement les pépites voulues.

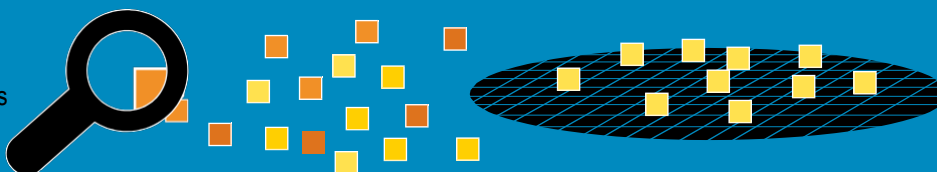


Figure 2: Comme les chercheurs d'or, vous devez souvent passer au crible des matériaux sans intérêt pour trouver de précieuses pépites, et des outils adaptés peuvent rendre ce processus plus efficace. La plate-forme IBM pour le Big Data propose plusieurs méthodes adaptées à vos besoins pour trouver des informations qui valent de l'or dans l'immensité du Big Data.

### Analyse du Big Data en mouvement

Pour certains types de données, le temps manque pour les enregistrer et il faut pouvoir agir avant de les stocker. Ces données sont souvent (mais pas toujours) générées par un capteur ou un autre appareil. La vitesse a son importance dans des cas comme la détection de fraudes, les nouveaux soins de santé et la sécurité publique, où il peut être critique d'améliorer ses connaissances des données en temps réel, et de déterminer s'il convient d'agir en accord avec des règles prédéfinies.

Des données très rapides et en grande quantité exigent de procéder à l'analyse dans l'action. Les données en cours de diffusion peuvent arriver déjà liées à un identifiant de données maîtres (tel qu'un numéro de téléphone). Dans une unité de soins intensifs pour prématurés, les données des capteurs peuvent être corrélées avec le nourrisson correspondant. Des transactions financières sont associées sans ambiguïté à un numéro de carte bancaire ou de sécurité sociale. Pour autant, toutes les données en circulation ne présentent pas un intérêt : elles doivent être analysées par un outil comme IBM InfoSphere Streams avant d'être associées aux données maîtres.

InfoSphere Streams analyse de grandes quantités de données avec une latence très faible. Au lieu de commencer par accumuler et stocker les données, Streams les analyse dès leur arrivée à la recherche de conditions qui déclenchent des alertes (par exemple des transactions aberrantes identifiées par une banque comme potentiellement frauduleuses lors d'une procédure d'autorisation de carte bancaire). Les données sont alors extraites du flux et recoupées avec les données maîtres afin d'améliorer les résultats. InfoSphere Streams produit un récapitulatif des connaissances obtenues grâce à l'analyse du flux et les met en correspondance avec des informations fiables telles que les dossiers des clients, pour compléter les données maîtres.

### Analyse des données mobiles en mouvement

Les données maîtres ajoutent une valeur stratégique au Big Data généré par les communications mobiles. Globe Telecom, un grand opérateur téléphonique philippin, utilise InfoSphere Streams pour le traitement analytique en temps réel (RTAP) des données sur les appels, les recharges et le l'attrition de la clientèle. L'opérateur utilise ces informations pour identifier de meilleurs modèles promotionnels pour chacun de ses clients. C'est ainsi que lancement d'une offre de recharge au moment opportun a eu un impact important sur l'activité.

L'établissement d'une relation entre le Big Data mobile et le MDM débouche également sur diverses opportunités de ventes croisées et incitatives, grâce à l'identification d'une liaison entre les préférences personnelles et les données maîtres, et à la définition de priorités en fonction du jour, de l'heure et du lieu.

---

### Comprendre le client : l'importance du contenu pour les offres promotionnelles

Un samedi à 18h, John Sanchez gare sa voiture près de la plage et paie le parking à l'aide de son téléphone. En croisant les informations récupérées dans des messages Facebook et à partir de ses propres données maîtres, l'opérateur téléphonique a déterminé que John est un père de famille de 40 ans. Il envoie un SMS à John dans lequel il lui propose un code de remise pour une pizza au format familial dans un restaurant situé à quelques pas de l'endroit où il est garé. Si le profil de John avait révélé qu'il avait 21 ans et était célibataire, il lui aurait plutôt proposé une remise pour une boîte de nuit ou un salon de tatouage des alentours.

Le Big Data des appareils mobiles permet souvent d'établir une corrélation solide avec un individu. Les connexions de tablettes dans les cafés et les aéroports génèrent d'immenses quantités de données de géolocalisation, qui, utilisées efficacement, s'avèrent précieuses pour les vendeurs

---

### Explorer le Big Data pour identifier des schémas et de nouvelles perspectives

IBM InfoSphere Data Explorer complète InfoSphere BigInsights et InfoSphere Streams lors de l'utilisation des données maîtres pour exploiter le Big Data. Cette plate-forme permet de naviguer dans de grandes quantités de données de presque n'importe quelle forme, et de les filtrer. Elle propose un tableau de bord visuel croisant les données maîtres d'une personne avec son historique complet de transactions obtenu à partir des e-mails, médias sociaux, applications packagées, etc. (voir Figure 3). Ces données peuvent aider les vendeurs, notamment les opérateurs de télémarketing, à prendre contact avec les clients potentiels de façon plus individualisée et améliorer leurs taux de vente.

Conçu pour gérer les données sociales et informations générées par des capteurs, InfoSphere Data Explorer permet, une fois un "axe" précisé, de rechercher les données correspondantes susceptibles d'être fusionnées aux données maîtres. En s'appuyant sur les données maîtres, InfoSphere Data Explorer peut aussi placer les flux d'activités en temps réel dans un contexte en vue de les analyser.

Les analystes du renseignement, des forces de l'ordre et de lutte contre la fraude doivent pouvoir tirer parti de toutes les bribes d'informations possibles pour procéder à des rapprochements et prendre instantanément des décisions bien éclairées. InfoSphere Data Explorer travaille à partir de données maîtres connues pour offrir aux agences de renseignement une vision unifiée de toutes leurs sources d'informations et apporter visibilité, transparence et connaissance.



Figure 3: InfoSphere Data Explorer obtient des informations des données maîtres pour renseigner certains domaines et les complète des informations obtenues auprès d'autres sources pour obtenir une vue complète d'une entité (par exemple d'un client).

## Big Data et MDM : associés, c'est encore mieux !

Toute entreprise qui accumule de grandes quantités de données qui circulent rapidement, qu'elles soient structurées ou non, doit mettre en place une approche bien définie pour les conserver, les analyser, les exploiter et juger de leur fiabilité.

IBM InfoSphere MDM et les outils de Big Data jouent un rôle clé dans l'exploitation optimale du Big Data. Ils aident les entreprises à analyser le Big Data, à le rapprocher des données maîtres et à en retirer des informations exploitables susceptibles de faciliter les prises de décision. InfoSphere MDM soutient les outils d'analyse et aide les organisations à fournir des données fiables pour remporter des ventes, satisfaire et fidéliser les clients, améliorer les opérations et renforcer le respect des réglementations.

### Informations complémentaires

Pour obtenir plus d'informations sur la façon dont MDM aide les entreprises à utiliser le Big Data et la plate-forme IBM InfoSphere MDM, veuillez consulter :

- [ibm.com/software/data/master-data-management](http://ibm.com/software/data/master-data-management)
- [ibm.com/bigdata](http://ibm.com/bigdata)



© Copyright IBM Corporation 2012

© Copyright IBM Corporation 2012

IBM Corporation Software Group Route 100  
Somers, NY 10589

Produit aux États-Unis d'Amérique  
Novembre 2012

IBM, le logo IBM, [ibm.com](http://ibm.com), BigInsights et InfoSphere sont des marques de International Business Machines Corp., enregistrées dans de nombreux pays du monde. Les autres noms de produits et de services peuvent être des marques de IBM ou d'autres entreprises. Une liste à jour des marques IBM est disponible sur le Web, dans la section "Copyright and trademark information" du site [ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

Ce document est à jour à sa date de publication initiale et pourra être modifié par IBM à tout moment. Toutes les offres ne sont pas disponibles dans tous les pays où IBM est présent.

Les exemples de données de performances et de clients sont cités à des fins d'illustration uniquement. Les performances réelles peuvent varier en fonction de la configuration et des conditions d'exploitation.

LES INFORMATIONS PRÉSENTES DANS CE DOCUMENT SONT FOURNIES "EN L'ÉTAT" SANS AUCUNE GARANTIE EXPRESSE OU IMPLICITE, Y COMPRIS SANS GARANTIE DE COMMERCIALISATION, D'ADÉQUATION À UN USAGE PARTICULIER NI GARANTIE OU CONDITION DE NON-CONTREFAÇON. Les produits IBM sont garantis selon les termes et conditions du contrat qui les accompagne.

Le client est responsable de la conformité aux lois et réglementations applicables. IBM ne fournit pas de conseil juridique, et n'apporte pas l'assurance que ses services ou produits garantissent le respect des lois et réglementations par le client.

<sup>1</sup>IDC Digital Universe Study, pour EMC, mai 2010.

<sup>2</sup>Hopkins, Brian. "Big Data: How enterprise architects can capitalise on future opportunities." <http://www.computerweekly.com/feature/Big-data-How-enterprise-architects-can-capitalise-on-future-opportunities>

<sup>3,4</sup>"Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data," McGraw-Hill, 2012. [ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&appname=SWGE\\_IM\\_DD\\_USEN&htmfid=IML14297USEN&attachment=IML14297USEN.PDF](http://ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&appname=SWGE_IM_DD_USEN&htmfid=IML14297USEN&attachment=IML14297USEN.PDF)



Merci de recycler