

BIG DATA

Produit Marketing ou Nouvel Eldorado ?

Thomas Morisse & Laurent Hanaud

thomas.morisse@adeli.org

laurent.hanaud@adeli.org

Résumé :

Synthèse de la conférence sur le Big Data animée par Laurent Hanaud et Thomas Morisse.

Mots-clés :

Big Data



Cet article est une synthèse de la conférence animée par Laurent Hanaud & Thomas Morisse sur le Big Data lors de l'Assemblée Générale d'ADELI le 18 janvier 2016 et de la soirée « événements-conférence » de l'ATEP de juin 2015.

La vidéo de la conférence est disponible en cliquant sur le lien suivant :

<http://www.adeli.org/contenu/big-data-lenregistrement-video-conference>

LE BIG DATA, C'EST QUOI ?

Pour faire simple, le Big Data signifie littéralement « Gros volume de données ».



Il s'agit donc de traiter des gros volumes de données structurées, ou non, mais uniquement de données numériques, c'est-à-dire une information codifiée, figée et transmissible.

Le Big Data s'intéresse à tous les formats de données, c'est-à-dire des images, des textes, des sons...

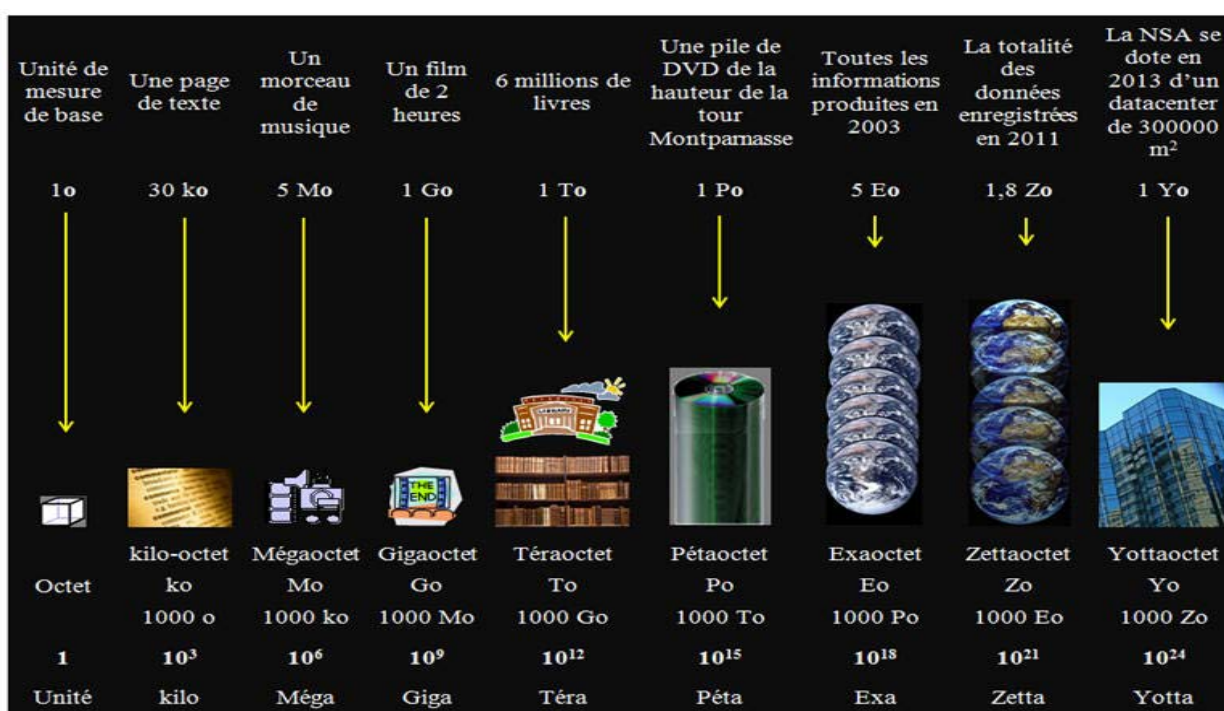
Ces données sont d'autant plus intéressantes que, lorsqu'elles sont captées ou mesurées, elles le sont avec leurs méta-informations (appelées les « métadonnées »). Les métadonnées pour une photo sont : les coordonnées de prise de vue, l'appareil utilisé, la date et l'heure de la capture...

Le Big Data se caractérise par les 3V : Volume, Vitesse et Variété.

Pour illustrer les **3V**, utilisons l'analogie avec le remplissage d'une feuille de calcul Excel :

- le **Volume** correspond aux lignes, représentant la quantité de données différentes qui sont traitées ;
- la **Variété** correspond au nombre de colonnes différentes à traiter : c'est-à-dire à la quantité d'éléments différents autour d'une même donnée ;
- la **Vélocité** correspond à la vitesse à laquelle se remplissent les lignes et les colonnes.

La quantité de données traitées est l'élément le plus important du BIG DATA, voici donc les ordres de grandeur et quelques chiffres :



Source : http://sciencesphysiques04.esy.es/documents/octets.htm#_Toc381598517

Par minute, nous échangeons dans le monde :

- 350 000 tweets ;
- 15 millions de SMS ;
- 200 millions de courriels.

Par jour cela fait :

- 24 Po traités par Google ;
- 10 To générés sur Facebook (300 millions de photos...) ;
- 7 To sur Twitter.



Quatre-vingt dix pourcent des données numériques existantes aujourd'hui ont été créées depuis 2014 et seulement 5 % sont réellement utilisées (Source : les échos - décembre 2015).

Selon le cabinet Transparency Market Research, les acteurs du Big Data se sont partagé un marché de 12,4 milliards de dollars en 2015 ; il sera de 23, 8 milliards en 2016... et pas loin de 50 milliards en 2018.

LE PROCESSUS DE VALORISATION

Les données brutes n'ont pas de valeur tant qu'elles ne sont pas utilisées pour prendre des décisions. Elles doivent suivre un processus de valorisation pour être exploitées.

Le processus est relativement simple : il s'agit tout d'abord de collecter les données, puis de les stocker et enfin de les analyser.



La Collecte

La collecte s'organise autour de différentes sources qui sont :

- les « capteurs » : vidéos, compteurs autonomes, sondes météo, satellite... ;
- les sites Internet ;
- les données d'entreprises ;
- les données personnelles : blogs, réseaux sociaux... ;
- l'OpenData (libération des données publiques – encadré par la loi – data.gouv.fr) ;
- les objets connectés.

La capture peut être volontaire (don de la personne, échange contre un service dit « gratuit ») ou bien involontaire, par exemple dans le cadre des vidéos sur la voie publique.

Le stockage

La fonction de stockage est assurée par les data centers (centre de données). Leurs nombres nécessitent de croître. Il y en a plus de 10 000 déclarés à ce jour à travers le monde, alors que l'on en comptait un peu moins de 2 300 en 2012.



Nombre et répartition des datacenters dans le monde en 2012

Source : <http://www.usinenouvelle.com/article/ou-sont-vos-donnees.N169826>

Ces machines à stocker sont organisées comme des usines où les données sont conservées, attendant d'être un jour utilisées.

On estime que la consommation des datacenters représente 2% de la production électrique mondiale et 1,5 % de la consommation mondiale des installations de refroidissement.

Corrélation et construction des schémas de décision

Les données sont corrélées suivant des modèles mathématiques afin d'en obtenir la « quintessence ». Les outils d'analyse sont de plus en plus nombreux et sont de plus en plus capables de fournir des résultats précis et fiables, pour permettre aux donneurs d'ordre de prendre des décisions.

LES ENJEUX

Risques pour les uns, opportunités pour les autres, le Big Data pose un certain nombre de questions que l'on qualifie d'enjeux.

Ces enjeux sont de plusieurs ordres.

Environnementaux

Un data center de 10 000 m² consomme autant qu'une ville de 50 000 habitants et seulement 10% de l'énergie consommée est utilisée pour traiter les données.

Sécuritaires

- l'hyper surveillance ;
- la fuite de données ;
- le piratage.

Enjeux sur les libertés individuelles

- vie privée & données personnelles ;
- droits d'accès (CGU) ;
- anonymisation, conservation (archivage) et droit à l'oubli ;
- illusion du tout gratuit (récupération de données personnelles contre services gratuits).

Enjeux de Souveraineté numérique

En 2012, 50 % des datacenters étaient situés aux États-Unis. Les GAFAs (Google, Amazon, Facebook et Apple) collectent toutes les données et investissent massivement dans le Big Data : la Silicon Valley devient la Big Data Valley.

Enjeux concernant l'absence de maîtrise

- banalisation de la collecte ;
- perte de localisation des données (par quel droit est régie la donnée ?) ;
- manipulations statistiques – l'information obtenue par corrélation d'une multitude de données est-elle fiable ?

Enjeux d'accessibilité

Les données doivent être accessibles en temps réel.

Enjeux de disponibilité

Les données doivent être disponibles en temps réel (cf. IoT).

LES USAGES

Le traitement de masse des données permet d'affiner les modèles de décisions existants, de définir des modèles plus précis et plus fiables.

De plus, le Big Data permet de passer de l'analyse reporting (le passé) à l'analyse prédictive (le futur) ou en d'autres termes de « réactif » à « prédictif ».

L'utilisation des données couvre aujourd'hui tous les domaines de l'activité économique et sociale de la société. La liste paraît infinie tellement les utilisations potentielles sont nombreuses telles que pour la sécurité intérieure, les prévisions météorologiques, la régulation des flux de transports...

Concrètement, comment cela peut-il se pratiquer ? Démonstration à partir de trois exemples posant la problématique et les résultats obtenus.

Exemple n°1 : Profilage et prédiction dans le marketing

Domaine d'activité : Téléphonie

Besoin traité :

La guerre des prix faisant rage sur le marché de la téléphonie mobile, SFR cherchait à savoir s'il était possible d'anticiper un éventuel non-renouvellement d'un abonnement grâce à l'étude des sites concurrents visités par leurs clients. Et à partir des résultats obtenus, il fallait aussi déterminer les actions à mettre en œuvre pour conserver l'abonné.

Éléments utilisés :

Le projet s'appuya sur trois pôles :

- le Big Data ;
- les échanges sur les réseaux sociaux ;
- le monitoring des conversations.

Résultat obtenu :

Grâce à cette démarche :

- Plus de 81 % des profils concernés par une potentielle résiliation ont été détectés.
- Cela a permis de conserver 75 % des clients.

Exemple n°2 : Analyse des données

Domaine d'activité : Industrie pharmaceutique

Besoin traité :

Dans le développement d'un médicament, le temps de recherche est toujours un point crucial. La société Quinten¹ proposa à son client, une solution en vue de réduire la durée d'un programme d'optimisation de molécules.

¹ Fournit une expertise technique et fonctionnelle dans le domaine du Big Data à différentes entreprises, en l'occurrence dans le cas présent : un laboratoire pharmaceutique <http://www.quinten-france.com/>

Éléments utilisés :

Pour y parvenir :

- à partir du Big data ;
- elle traita 400 composés pendant les 6 premiers mois du programme ;
- identifiant dans ce laps de temps, 4 profils de composés présentant les meilleures propriétés.

Résultat obtenu :

Pour obtenir ce résultat, le temps de recherche fut divisé par 2.

Exemple n°3 : Optimisation & Régulation avec le Smartgrid

Domaine d'activité : Distribution d'électricité

Besoin traité :

Au début des années 2000, la société Enel² devait faire face à une double problématique :

- Comment lutter contre la fraude ?
- Comment optimiser la distribution d'électricité à la demande ?

Éléments utilisés :

Avec l'aide de CapGemini, elle résolut cette problématique, en déployant des infrastructures Big Data pour recueillir les données issues des équipements de distribution³. Pour résumer, cette opération reposa sur :

- du Big data ;
- le sourcing des données avec des compteurs intelligents ;
- les échanges avec les équipements de distribution hétérogènes.

Résultat obtenu :

Après avoir déployé 30 millions de compteurs, la fraude fut réduite de 75 %.

La liste pourrait être sans fin, citons dans le domaine :

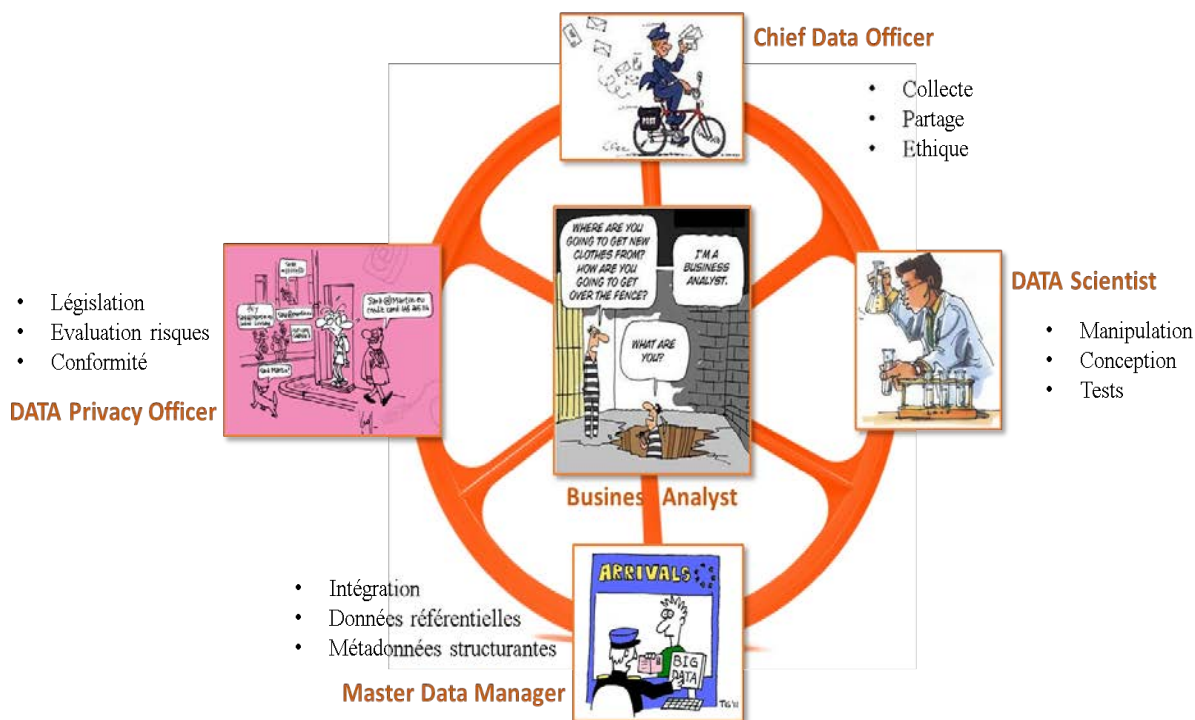
- de la sécurité intérieure : la prédictivité des événements pour anticiper les délits avec le système PREDPOL ;
- des transports : Les véhicules connectés – La gestion de trafic ;
- de l'agriculture : L'utilisation des drones.

² Opérateur italien d'énergie électrique. <http://www.enel.fr/>

³ Livrés par différents fournisseurs, tel que GE Energy, Alstom, Siemens ou encore Schneider Electric.

LES MÉTIERS DU BIG DATA

Pour traiter ces données dans l'entreprise, de nouveaux métiers sont apparus ces dernières années. Le schéma suivant synthétise la situation :



Le Big Data a fait émerger quatre métiers « type » :

- le chief Data Officer ;
- le Data Scientist ;
- le Master Data Manager ;
- le Data Privacy Officer.

Bien entendu, d'autres métiers ou d'autres appellations peuvent apparaître dans les différentes publications traitant du Big Data, rien n'est figé. Mais nous disposons là, en quelque sorte, des quatre points cardinaux du métier. Ajoutons aussi, bien qu'il ne soit pas spécifiquement lié au sujet - mais qui est important en tant que garant du besoin utilisateur final - le Business Analyst.

Le **Chief Data Officer** a un rôle triple, celui de :

- diriger la collecte des données ;
- organiser le partage de leur analyse avec les autres directions et le Business Analyst ;
- faire respecter l'éthique en matière d'usage.

Le **Data Scientist** manipule l'information. Il conceptualise cette dernière pour le métier en donnant sens aux données. Il les trie, les teste et émet des hypothèses.

Le **Master Data Manager** est l'intégrateur de service. Ainsi, il est en charge :

- d'identifier les données référentielles ;
- de maintenir les catalogues métiers ;
- de faire appliquer les normes et méthodes ;
- de s'assurer de la bonne intégration des données dans le SI.

Le **Data Privacy Officer** (DPO) est en quelque sorte le « Shérif » de l'équipe, l'homme de loi. En effet il doit :

- s'assurer de la législation⁴ (conforme à la loi) ;
- informer les individus de leurs droits (données à caractère personnel) ;
- évaluer le risque tout au long de la vie des données (conformité et responsabilité).

Avant même d'évoquer le Big Data, le **Business Analyst** (BA) a toujours tenu une mission centrale pour l'entreprise :

- Il analyse les besoins métier, recherche des solutions et vérifie leur adéquation aux besoins.
- Par conséquent, son activité reste centrée sur l'analyse, la modélisation, la formalisation et la négociation.

Le Big Data produit des résultats, encore faut-il que ces derniers fournissent une valeur ajoutée à l'entreprise, qu'ils répondent au besoin. De ce fait, le rôle du Business Analyst s'en trouve renforcé et revêt une dimension supplémentaire.

CONCLUSION

Le Big Data, eldorado ou mirage ? Une réponse intéressante a pu être fournie par Gilles Babinet⁵ :

« *Le Big Data n'est pas qu'une technologie, mais bien une nouvelle structure d'information et de management. C'est donc une nouvelle façon d'interagir avec la réalité.* »

Force est de constater que ce principe commence à être appliqué dans l'analytique RH⁶ ou dans le marketing digital⁷.

⁴ Notons, qu'en cas de non-respect par l'entreprise de la réglementation relative à la protection des données personnelles, la Commission Européenne a prévu des amendes pouvant aller jusqu'à 2 ou 5 % du chiffre d'affaires mondial de l'entreprise.

⁵ Entrepreneur français, digital champion de la France auprès de la Commission européenne, auteur de « Big Data, penser l'homme et le monde autrement ».

⁶ L'analytique RH (RH pour Ressources Humaines) est un aspect du Big Data. Il est parfois utilisé sous le terme de People Analytics. L'analytique RH, permet d'optimiser la mobilité interne dans les entreprises, de prévenir le turn-over et de mieux repérer les talents. Pour en savoir plus, vous pouvez vous référer aux articles suivants :

* Le blog ManpowerGroup : « Le big data en support des RH pour gérer les talents et favoriser la mobilité interne ». <http://www.manpowergroup.fr/big-data-support-rh-gestion-talents-mobilite-interne/>

* Philippe Burger : « HR Analytics : le prochain Graal des ressources humaines ? ». <http://www.blog.deloitte.fr/technologie-et-innovation/hr-analytics-le-prochain-graal-des-ressources-humaines/>

* Andrée Laforge : « L'analytique RH, les indicateurs, le benchmarking et tout le tralala! ». <https://mesurerlecapitalhumain.wordpress.com/2016/05/25/lanalytique-rh-les-indicateurs-le-benchmarking-et-tout-le-tralala/>

* Andrée Laforge rapportant une annonce faite par Josh Bersin de Deloitte : « 10 choses à savoir sur l'analytique RH ». <https://mesurerlecapitalhumain.wordpress.com/2016/01/28/10-choses-a-savoir-sur-lanalytique-rh/>
<http://deloitte.wsj.com/cio/2015/12/03/10-things-we-know-about-people-analytics/>

⁷ Le marketing digital cherche à optimiser les ventes en établissant une relation plus personnalisée avec le client. Cette personnalisation peut être obtenue grâce aux nouveaux outils du numérique tels que les réseaux sociaux, aux référencements, aux catalogues on line, aux newsletters, etc. Et les données sur le comportement d'achat s'en trouvent enrichies, d'où sa relation avec le Big Data. Pour bien comprendre, voici deux articles de référence :

* Vincent Luciani JDN : « Le digital marketing et ses 5 mondes ». <http://www.journaldunet.com/ebusiness/expert/58592/le-digital-marketing-et-ses-5-mondes.shtml>

* Bertrand Bathelot : « La définition du marketing digital » <http://www.definitions-marketing.com/definition/marketing-digital/>

