

## Une mise en perspective du phénomène Big Data

Par : **Mohammed ZOUHRI**<sup>89</sup>

**Saïd AMALI**<sup>90</sup>

### Résumé

*Aujourd'hui, nous assistons à un véritable déluge de données produites par les particuliers, les entreprises, les scientifiques et les acteurs publics. Cet accroissement de données, couplé au développement d'outils informatiques innovants, offre d'innombrables perspectives d'analyses et de possibilités pour les états et les entreprises. Tout le monde s'accorde à dire que le traitement de ces masses de données, ou Big Data, est un enjeu majeur dans la société de demain, car il trouve des applications dans des domaines variés : les sciences, le marketing, les services clients, les transports, la santé, le développement durable et l'éducation.*

*Dans cet article, on revient sur ce nouveau concept de Big Data et on en donne les dernières caractérisations produites par la littérature. Ensuite, on présente les principales technologies liées à cette notion en termes d'outils et méthodes. Puis on s'attache à détailler quelques applications concrètes du phénomène relatées par les spécialistes. Enfin, on discutera des conditions indispensables au développement de l'analyse des Big Data par les entreprises.*

---

<sup>89</sup> Professeur à la Faculté des Sciences Juridiques Economique et Sociales, *Equipe OMEGA, Laboratoire LERES, FSJES – Meknès- Maroc*

<sup>90</sup> Professeur à la Faculté des Sciences Juridiques Economique et Sociales, *Equipe OMEGA, Laboratoire LERES, FSJES – Meknès- Maroc.*

## Introduction

Avec la démocratisation des technologies de l'information et de la communication (essor des réseaux Internet et Wi-Fi, des ordinateurs, des smartphones, des objets connectés et des réseaux sociaux), les utilisateurs sont conduits à produire une masse de données de formes variées qui reflètent en partie leurs profils en termes de préférences et de pratiques de consommation. Cette grande quantité de données constitue une source d'information importante permettant aux organisations, entreprises et états de mieux prédire les comportements et les habitudes de ces utilisateurs (clients, consommateurs, citoyens).

Cette profusion de traces numériques laissées par les utilisateurs technophiles est à l'origine d'un nouveau phénomène très en vogue depuis quelques années, il s'agit du Big Data. Ce phénomène impacte en particulier les entreprises qui sont amenés à manipuler des ensembles de données beaucoup plus volumineux qu'il y a dix ans et qui nécessitent une infrastructure particulière pour leur création, leur stockage, leur traitement, leur analyse et leur récupération.

Le Big Data est considéré par certains comme le nouvel « or noir » de croissance et la valeur de son marché, en pleine expansion, est évaluée à 125 milliards de dollars dans le monde (soit environ 111 milliards d'euros) pour la seule année 2015.<sup>91</sup>

La Big Data représente beaucoup pour le monde de l'entreprise. Au-delà d'une simple mode, il s'agit d'une véritable révolution qui va profondément et réellement changer la face du monde du business. Certes, le marketing reste le principal impacté, mais il ne faut pas sous-estimer les conséquences de la maîtrise de ce sujet sur le monde des études, de la gestion, en passant par celui de la production, de la gestion de la chaîne logistique et de la Recherche et Développement.

Mais, c'est quoi au juste, ce Big Data ? Protéiforme, le phénomène Big Data ne se laisse pas aisément circonscrire. La conception « Big Data » comme son nom l'indique est d'origine anglophone et est souvent traduite par les expressions « grosses données », « données massives » ou dernièrement « mégadonnées »<sup>92</sup>. Le concept a pris un tel essor après la diffusion des trois principaux rapports écrits sur le sujet par l'institut McKinsey ou McKinsey Global Institute (MGI): *Clouds, Big Data, and smart assets : Ten tech-enabled business trends to watch* (Manyika et al, 2010), *Are you ready for the era of « Big Data »* (Brown,

---

<sup>91</sup> Le numérique, une chance à saisir pour la France, Livre Blanc, CGI, 2015.

<sup>92</sup> Journal officiel de la république française, 22 août 2014.

Chui et Manyika, 2011) et le classique **Big Data : The next frontier for innovation, competition and productivity**(Manyika et al, 2011).

En effet, pour McKinsey Global Institute, le terme de « Big Data » décrit des jeux de données dont la taille ou la nature ne permet pas une capture, un stockage, une gestion et un traitement par des outils de gestion de bases de données classiques. Pour autant, le cabinet ne définit pas de limite de taille pour le concept de Big Data et note que le volume de données considéré comme étant « Big Data » peut varier d'un secteur d'activité à un autre (d'une douzaine de Téraoctets à plusieurs Pétaoctets<sup>93</sup>). Une chose est certaine, précise toutefois, McKinsey, les données de type « Big Data » progressent à un rythme soutenu.

Cet article est une mise en perspective de ce nouveau phénomène non seulement technologique, mais qui présente de grands enjeux aussi bien économique qu'organisationnel et stratégique. Nous en donnons une définition suite à plusieurs lectures faites aussi bien dans la littérature académique que professionnelle. Nous reviendrons sur le fondement et la matière brute de ce phénomène, à savoir les données.

L'explosion des données numérique, en tant que telle, n'aurait eu aucune valeur sans accompagnement technique : toute l'innovation liée au Big Data a d'abord été une innovation technologique, portée par la mutation des outils de stockage et de traitement. Nous présenterons quelques principales technologies liées au Big Data.

Aujourd'hui le « buzz » Big Data agite de nombreux experts : ce phénomène est-il vraiment si révolutionnaire que le laissent entendre les acteurs du secteur et les médias ? L'utilisation des données à des fins de traitement et d'analyse est en effet inhérente à tout pilotage d'entreprise, et des outils anciens existent pour cela (depuis l'antique livre de comptes jusqu'aux tableaux de Business Intelligence, en passant par les traditionnels tableurs). Nous nous attacherons donc à détailler quelques usages et possibilités offertes par les analyses de masses de données et leurs applications concrètes.

Les entreprises ont tout intérêt à anticiper l'impact du Big Data sur leur modèle économique car ce dernier pourra être remis en cause, voire menacé, par certaines entreprises « pure players » (nouveaux concurrents potentiels) qui, disposant de toutes les données, pourront investir des nouveaux métiers demain (Google, Amazon,...) (Cigref, 2013). Nous

---

<sup>93</sup> Un téraoctet (To) correspond à mille milliards d'octets et un pétaoctet (Po) vaut mille téraoctets.

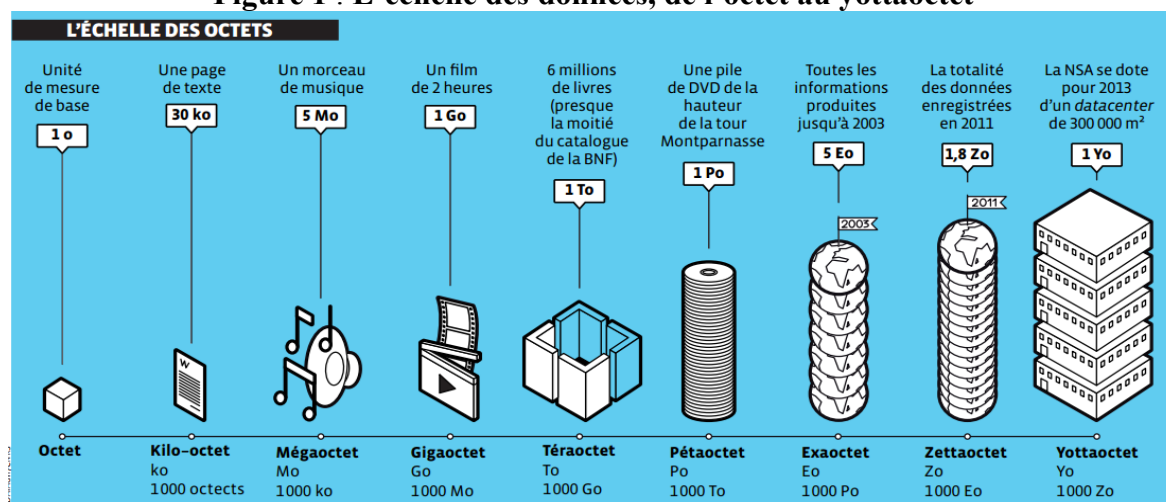
discuterons enfin des conditions indispensables au développement d'une stratégie Big Data au sein des entreprises.

## 1. Big Data : éléments de définition

Si le Big Data est devenu un terme récurrent dans divers milieux : scientifiques, cabinets de conseil et média spécialisé, il est quand même difficile de s'accorder sur une définition unique du concept. Cependant, une caractérisation récurrente du phénomène semble s'orienter vers la question technologique, avec la célèbre formule des 3V (Volume, Variété et Vitesse ou Vélocité) proposée par le cabinet McKinsey (Manyika et al, 2011).

En effet, le volume des données numériques produites est en augmentation exponentielle. On estime que 90% de l'ensemble des données disponibles aujourd'hui ont été créées ces deux dernières années (Brasseur, 2013). On ne parle plus de gigaoctets ( $10^9$  octets), mais plutôt de téraoctets ( $10^{12}$  octets), de pétaoctets ( $10^{15}$  octets), d'exaoctets ( $10^{18}$  octets), voire même de zettaoctets ( $10^{21}$  octets).

Figure 1 : L'échelle des données, de l'octet au yottaoctet



Source : « La déferlante des octets » Le journal du CNRS, n°269, novembre-décembre 2012.

Cette explosion de données s'explique principalement par la mise à disposition généralisée d'outils numériques de plus en plus performants et connectés, tels que les ordinateurs (PC, portables, tablettes) et les Smartphones reliés à Internet en permanence.

Mais au-delà, une multiplication de données issues d'objets interconnectés, de type capteurs ou puces mobiles est envisagée. Ericsson prédit ainsi qu'il y aura 50 milliards d'objets connectés dans le monde d'ici 2020, contre environ 12 milliards aujourd'hui (Ericsson, 2011).

La deuxième caractéristique est la variété. Les données analysées ne sont plus forcément structurées mais peuvent être de format diverse (texte, image, photo, vidéo, son,

log technique,...). Par exemple, on peut aujourd'hui étudier l'état de l'opinion via les tweets (social medias analysis) ou encore proposer une aide au diagnostic en exploitant les bases de données médicales. La variété porte également sur les usages possibles associés à une donnée brute, par exemple un même fichier son généré sur un plateau téléphonique pourra servir à créer un fichier texte ou à échantillonner la voix en vue d'une reconnaissance vocale ultérieure.

Le troisième critère retenu est la vitesse ou la vélocité. En effet, la vitesse réfère aux délais d'actualisation et d'analyse des données numériques. Les données sont traitées en temps réel ou quasi réel. Par exemple, en analysant en temps réel les données sur l'état de santé de bébés prématurés, des médecins canadiens ont pu détecter des infections vingt-quatre heures avant la manifestation de symptômes visibles (Mayer-Schönberger&Cukier, 2013).

Notons au passage qu'il existe deux grandes familles de projets Big Data. Celle qui traite les données en temps réel et celle qui travaille sans cette contrainte. Ces deux familles de projets induisent des approches différentes, des architectures techniques distinctes, des outils et des données différentes.

Ensuite cette caractérisation classique des 3V s'est enrichie de deux autres « V », une vision davantage économique portée par le 4<sup>ème</sup> V de la définition, celui de Valeur, et une notion qualitative véhiculée par le 5<sup>ème</sup> V, celui de la Véracité des données.

Les acteurs du secteur Big Data mentionnent donc la valeur qu'il est possible de tirer de cette avalanche de données et les usages qu'elles produisent mais aussi la véracité ou la qualité des données. Par exemple, comment s'assurer que les informations de réseaux sociaux comme *Facebook* ne sont pas des rumeurs ou des diffusions malveillantes ? Pour l'anecdote, une information d'un faux compte *twitter* a dégradé le cours de l'action d'une société cotée au NASDAQ. Le *tweet* entraîné l'échange de 300 000 actions en deux minutes, et une baisse de 25% de la valeur de l'action<sup>94</sup>.

Au-delà de cette approche très descriptive des 3V et qui a été étendue à 5V, on peut apporter un éclairage supplémentaire en termes de destination des Big Data au travers des 3P : Prévision, Personnalisation et Prévention, qui illustrent le rôle joué par les Big Data dans certains cas d'usage particulièrement pertinents. Ces 3P méritent précision.

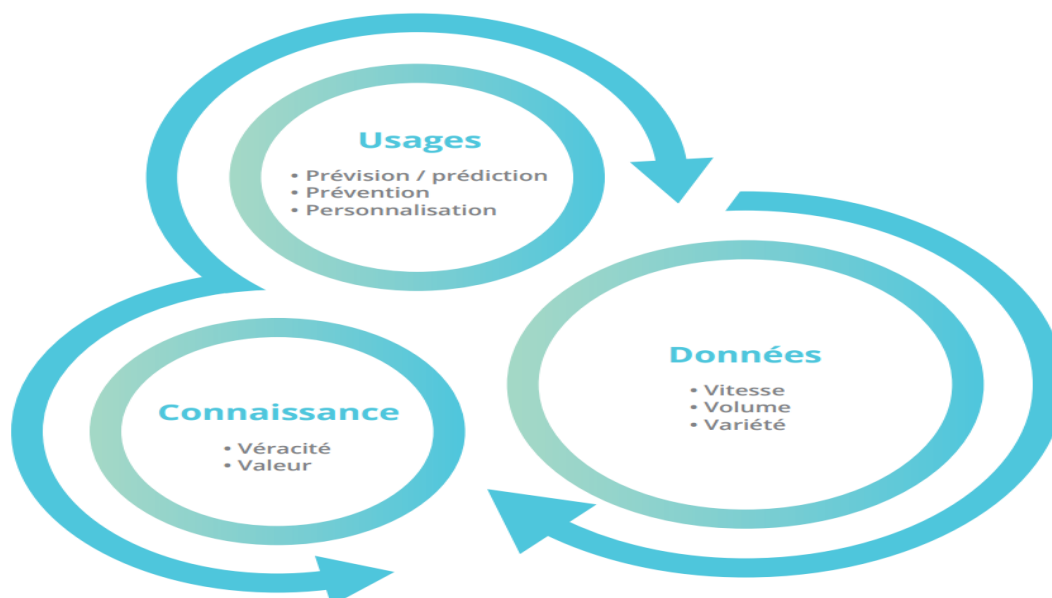
---

<sup>94</sup><https://gautierlandivier.wordpress.com/2013/02/26/un-faux-tweet-plombe-le-cours-de-bourse-dune-entreprise-le-web-et-la-folie-des-rumeurs/> [consulté le 24/12/2015]

Une première famille d'usage porte sur la notion de prévision. Comment exploiter les données pour mieux anticiper ? Comment accumuler suffisamment de connaissances pour prévoir la demande, les problèmes, les comportements et les goûts ?

Une deuxième famille d'applications porte sur la capacité de personnaliser au niveau le plus fin l'interface que propose la solution Big Data. Il s'agit ici plutôt d'une connaissance approfondie d'un environnement qui permet de configurer tout le système spécifiquement pour un groupe de personnes voire un individu.

Enfin, une troisième famille d'usages porte sur la prévention. Ici on entend se servir du Big Data pour identifier un risque, un danger et si possible, le prévenir. L'objectif ici est de définir ce qu'est le risque ou ce qui représente un danger potentiel.



**Figure 2 :** le Big Data ne se résume pas aux 3V, très descriptifs. On peut leur adjoindre 2V supplémentaires qui permettent de qualifier les données et enfin et surtout, les 3P qui décrivent la destination du Big Data. (Du Big Data au big business, Livre blanc, Business&Décision, 2013)

## 2. Nouveaux enjeux technologiques du Big Data

L'univers technologique du Big Data s'appuie sur des outils bien identifiés qui constituent la base innovante de ce mode de traitement. À eux seuls, ces outils résument le vocabulaire technologique du Big Data et en constituent la référence. Nous allons décliner les spécificités de ces innovations technologiques en termes d'architectures matérielle et logicielle ainsi que les bases de données Big Data.

### a- Les spécifications d'architecture matérielle en Big Data

Le traitement des données au volume important n'aurait pu être possible sans des changements majeurs opérés dans les composants assurant les fonctions transmettre, stocker

et transformer des ordinateurs et traduits par des « lois » : la loi de Moore, la loi de Kryder et la loi de Nielsen qui sont toujours vérifiées à nos jours. La loi de Moore prévoit que la capacité de calcul des ordinateurs double tous les dix-huit mois, celle de Kryder prévoit le doublement de densité des disques durs tous les treize mois et Nielsen estime un doublement des capacités des réseaux publics tous les vingt-et-un mois.<sup>95</sup>

Les traitements Big Data sont portés entre autre par l'évolution fulgurante de processeurs ces 20 dernières années, faisant passer leur architecture de 16 bits à 32 bits, ensuite de 32 bits à 64 bits. Rappelons que l'architecture d'un processeur correspond au nombre de registres en mémoire vive qu'il peut gérer directement, car aller au-delà de ce nombre nécessite des accès aux disques durs. Ainsi, un processeur 32 bits qui gère un peu plus de 4 giga-octet de mémoire vive est largement suffisant pour le fonctionnement du système d'exploitation et des programmes, mais n'autorise que peu de données en mémoire vive. Par contre, en architecture 64 bits et en théorie, on peut stocker des exaoctets ( $10^{18}$  octets) de données en mémoire vive. C'est le « In-memory ». L'avantage du traitement « in-memory » est celui de la vitesse puisque les données sont immédiatement accessibles. En revanche, ces données ne sont pas stockées sur le long terme, ce qui peut poser des problèmes d'historisation.

Un autre choix technologique ayant permis l'essor du Big Data est le « massivement parallèle (MPP) ». Les technologies du massivement parallèle sont à l'ordre du jour depuis plus de 20 ans. Les informaticiens ont adopté différentes façons de traiter leurs problèmes techniques : au travers des machines vectorielles, avec des mainframes, avec des serveurs, puis en ajoutant de plus en plus de processeurs, et enfin aujourd'hui, avec des processeurs qui ont de plus en plus de cœurs, ou encore des machines regroupées en clusters.

Une troisième alternative concerne les SSD (*Solid State Drive*) qui sont des unités de stockage uniquement constituées de mémoire flash qui présentent des avantages considérables par rapport aux unités DRAM (*DynamicRandom Access Memory*) constituant principal des mémoires vives d'aujourd'hui.

Le traitement massif de données peut prendre plusieurs voies, chacune avec ses avantages et ses inconvénients et les trois technologies (*in-memory*, *MPP*, *SDD*) vont sans doute s'installer dans les projets Big Data.

---

<sup>95</sup> Pierre DELORT, Le Big Data, Que Sais-je?, 2015

Mais c'est particulièrement l'avènement du « Cloud Computing » qui est à l'origine du développement accéléré du Big Data. En mutualisant les données dans le nuage et en démocratisant ainsi l'accès des entreprises au stockage et au traitement de tous types de données, le « Cloud Computing » a créé les conditions pour une généralisation du Big Data. Des petites, moyennes ou grandes entreprises peuvent souscrire à des offres « as a service » pour explorer leurs données internes mais également d'autres données accessibles depuis le nuage.<sup>96</sup>

#### **b- Les spécifications d'architecture logicielle en BigData**

D'un point de vue logiciel, les spécificités des Big Data découlent de ce qui précède. A une époque, les calculs parallèles étaient réservés à une communauté réduite de scientifiques, de développeurs et de spécialistes très pointus. Aujourd'hui, le calcul parallèle s'est démocratisé et les choses ont complètement changé : le Cloud Computing, puis les outils comme *MapReduce* de Google, *Hadoop* de la fondation OpenSource Apache et *Cassandra* de Facebook, tous massivement parallèles, obligent les informaticiens à se confronter à ce qu'est un développement de calcul parallèle, voire distribué.

*MapReduce* vient de Google et inverse un concept traditionnel – envoyer les données aux unités de traitement -, car il est conçu pour envoyer le traitement aux unités de stockage. *MapReduce* est un algorithme qui est principalement utilisé pour manipuler un volume important de données qui sont distribuées sur des clusters ou machines parallèles. Ce processus se déroule en deux étapes :

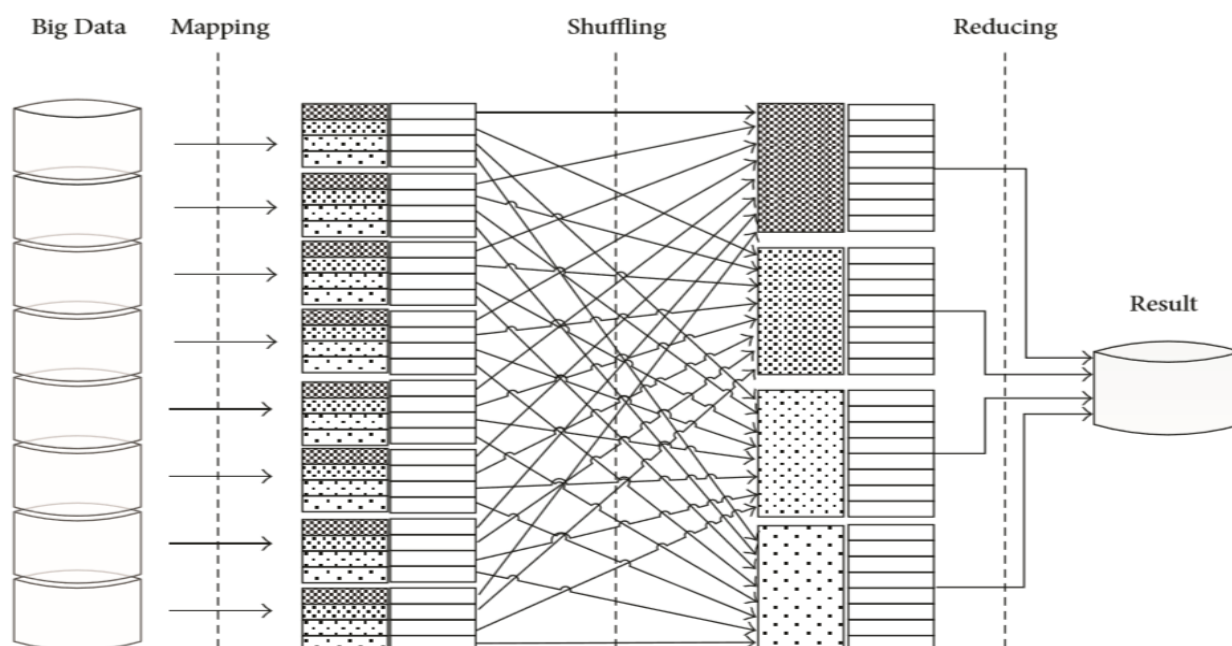
- Une première étape *Map* pour découper le problème en morceaux plus ou moins gros ; on va ensuite dédier chaque machine à un sous-problème particulier. Chaque machine va elle-même redécouper ce sous-problème, et de manière récursive, on arrive à ce que chaque machine traite une toute petite partie du problème.
- Une deuxième étape *Reduce* pour regrouper les morceaux. Chaque nœud va utiliser cette deuxième étape qui consiste à faire un calcul de manière indépendante (*Shuffle*) et à remonter vers la machine qui lui a donné l'ordre de faire le calcul, jusqu'à revenir au nœud de départ.

---

<sup>96</sup>Ibrahim AbakarTargio Hashem et al, The rise of "Big Data" on cloud computing: Review and open research issues, Information Systems 47 (2015) pp. 98-115



Avec cette méthode, on découpe les problèmes, soit pour atteindre un problème suffisamment simple pour pouvoir être traité par l'algorithme, soit pour s'adapter à la puissance de la machine.



**Figure 3 : Fonctionnement de MapReduce<sup>97</sup>**

Qui dit Big Data, dit outils de gestion de données. Jusqu'à une époque, on avait tendance à stocker toutes les données de la même façon sans réfléchir à leur pertinence, dans des bases de données dites relationnelles ou transactionnelles qui sont plutôt orientées « lignes ».

Les bases de données orientées « lignes » se caractérisent par :

- des données stockées par tables composées de colonnes et de lignes ;
- des lignes identifiées uniquement par une clé primaire, des index en accélèrent l'accès ;
- des clés primaires et étrangères pour définir des relations entre deux tables (identité entre colonnes) ;
- une utilisation multi-utilisateurs concurrents.

Ces dernières outillent des applications s'inscrivant dans des process, avec un accès aux objets de gestion (les lignes : par exemple des clients) décrits par des attributs (les colonnes : par exemple le nom, l'adresse, ...). Les attributs d'une entité seront sur le disque, contigus car manipulés ensemble et le besoin de prévisibilité des requêtes (plusieurs utilisateurs

<sup>97</sup>Nawsher Khan and al., Big Data: Survey, Technologies, opportunities, and Challenges. The Scientific World journal, 2014

concurrents) élevé (c'est le cas par exemple pour les bases de données de type Oracle, DB2, SQL, SQL Server).

Une base de données orientée « colonne », par contre, placera de manière contigüe les données d'une colonne pour les balayer rapidement, en exploratoire et fouille de données, dès lors avec un moindre besoin de prévisibilité (c'est le cas de *Hbase*, inspirée de *BigTable* de Google ou *Cassandra* de Facebook).

Enfin, on trouve aussi les bases de données orientées « documents » où on va stocker uniquement des couples clé/valeur (key/value) en vrac. Ce sont des bases de données très déstructurées, où on stocke tout en vrac et où les algorithmes de recherche de canalisation des données vont utiliser les spécificités de la base de données pour améliorer les performances (c'est le cas par exemple de *MongoDB* ou *CoucheDB*).

Il faut dire que les technologies des bases de données ne cessent d'évoluer que ce soit en mode open source ou chez les grands acteurs de l'informatique, comme l'illustre la dernière base de données *Spanner* de Google. Le choix de la base de données va surtout dépendre en premier lieu des besoins. Dans une base de données relationnelle, une recherche ou une sélection en fonction d'un champ donné par exemple, est très peu coûteuse. Par contre, c'est beaucoup moins performant si on veut faire beaucoup d'insertion avec des jointures entre deux bases de données pour faire des statistiques par exemple. Ce même type d'opérations sur une base de données orientée « colonnes » est quasi-instantané car elles sont faites pour cela.

### **3. Nouveaux usages induits par les Big Data**

Comme mentionné précédemment, l'analyse des Big Data offre d'innombrables possibilités tant pour le secteur public que privé. Les usages des technologies innovantes sont nombreux, qu'il s'agisse de rendre la gestion plus efficace, d'améliorer les services rendus ou de prévenir des phénomènes nuisibles (épidémies, criminalité, etc.)

La technologie des Big Data aide à mieux être à l'écoute des usagers, à comprendre les modes d'utilisations des services et à affiner l'offre. Par exemple, Google Analytics, service gratuit d'analyse d'audience d'un site Web ou d'applications, permet une meilleure adaptation de son site web en analysant les traces des internautes. Mais les Big Data ne sont pas l'apanage que du secteur privé, le secteur public y trouve aussi des applications innovantes.

Les modes d'enseignement (dont les MOOC – Massive Open Online Courses) peuvent être considérablement améliorés en analysant les activités des apprenants en ligne (temps

consacré, façon de suivre les programmes, arrêt-retour dans les vidéos pédagogiques, recherches internet parallèles, etc.).

Dans le registre des transports, on peut par exemple utiliser les données des transports en commun, des vélos et de la géolocalisation de personnes ou de voitures pour adapter les infrastructures et les services (horaires des trains, etc.).

Les analyses des données massives peuvent faciliter la recherche d'emploi en combinant les qualifications avec l'ensemble des offres d'emploi en ligne (provenant des sites spécialisés, des sites d'entreprises, des sites administratifs, etc.).

Un autre usage des Big Data est d'anticiper, avec une probabilité donnée, des comportements ou des besoins. On analyse ainsi un volume important de données sur les us des internautes pour établir des corrélations et ainsi prédire leurs achats.

Un autre domaine d'application des technologies Big Data est bel et bien celui de la santé. On peut grâce à l'analyse des requêtes des internautes prévenir des maladies ou des épidémies mais aussi améliorer les diagnostics ou permettre à l'industrie pharmaceutique de meilleurs ciblage thérapeutiques. Ainsi, quand la crise de la grippe H1N1 a éclaté en 2009, le système mis au point par Google s'est révélé un indicateur plus utile et plus précis que les statistiques gouvernementales. Les responsables de la santé publique ont ainsi disposé de renseignements précieux<sup>98</sup>.

La prévention des crimes est un autre type d'applications des Big Data. En effet, la police américaine de Memphis a développé un logiciel (*Blue CRUSH*) permettant d'identifier les zones et les heures où des délits et des crimes sont le plus à même d'avoir lieu, afin d'optimiser l'affectation des services. Ce programme informatique, prélève et rassemble avec l'aide de caméras et des forces de police un maximum de données sur les délits qui surviennent dans un territoire. A mesure qu'elles s'accumulent, les données sont traitées par l'ordinateur qui, sous formes de statistiques, fait émerger des tendances et des probabilités; probabilités sur lesquelles les policiers vont se baser pour régler leurs patrouilles. Cependant, recourir aux ordinateurs et aux calculs de probabilité pour prévenir la criminalité peut poser de sérieux problèmes éthiques.<sup>99</sup>

---

<sup>98</sup>Big Data : la révolution des données est en marche. Viktor Mayer-Schönberger, Kenneth Cukier, RoberLaffont, 2014.

<sup>99</sup><http://www-03.ibm.com/press/us/en/pressrelease/32169.wss> [consulté le 23/12/2015]

#### 4. Stratégie Big Data dans les entreprises

Le Big Data gagne de jour en jour en popularité, et les expressions infobésité, datamasse, mégadonnées fleurissent chez les spécialistes en innovations technologiques, business, et même en marketing et SEO<sup>100</sup>. Considéré comme le « nouveau pétrole » par le forum économique mondial<sup>101</sup>, le Big Data permet d'améliorer la prise de décision, de raccourcir les temps de commercialisation et d'augmenter les profits. Mais selon quelle stratégie ?

Une étude menée par le cabinet Ernst & Young auprès d'un panel d'entreprises françaises montre que deux tiers de ces dernières considèrent que le Big Data est un concept intéressant mais encore trop flou, difficile à appréhender au sein des entreprises, tant en termes de transformation organisationnelle, de stratégie, de retour sur investissement (ROI) que de gestion et de formation des compétences.<sup>102</sup>

Cette étude a révélé les éléments suivants :

- **Le nombre d'études d'opportunités Big Data demeure insuffisant** : la moitié des entreprises enquêtées n'a pas étudié les opportunités éventuelles liées au Big Data. Seules 9 % des entreprises du panel ont lancé à la fois une étude d'opportunité du Big Data et mis en place une stratégie globale de gestion de leurs données clients.
- **L'entreprise n'a pas de vision «360°» de la donnée** : la moitié des entreprises interrogées reconnaît que l'absence « d'un plan d'action clair qui constitue une feuille de route pour l'ensemble de l'entreprise » est un frein à une exploitation optimale des données clients. 57 % des entreprises non matures considèrent la perception du top management comme un frein contre 11 % pour les très matures.
- **Une « gouvernance Big Data » n'est pas encore à l'ordre du jour** : la responsabilité de la « stratégie Big Data » et son application incombe soit à la direction marketing, soit à la direction des systèmes d'information (DSI) d'impulser et/ou de mettre en place cette stratégie, mais aussi de concentrer les compétences « Big Data » de l'entreprise. Aux États-Unis par exemple, on voit émerger un nouveau profil de « Chief Data Officer » qui a la charge de l'ensemble des questions relatives à la « stratégie Big Data » de l'entreprise.

---

<sup>100</sup>Search Engine Optimization ou référencement naturel

<sup>101</sup>*Big Data, Big Impact: New Possibilities for Development*, Genève : World Economic Forum, 2012

<sup>102</sup>(Big) data : où en sont les entreprises françaises ? *Quelle maturité dans l'exploitation des données clients ?* Ernst & Young Advisory, 2014

Si chacune des approches Big Data mises en place dans les entreprises a des avantages, ce n'est pas tant l'approche en tant que telle qui fera la différence, mais la prise en compte d'un certain nombre de conditions qui constituent les facteurs clés de succès d'une stratégie Big Data. L'entreprise se doit de définir une stratégie Big Data, avec la donnée comme axe stratégique au même titre que la politique tarifaire par exemple.

Parmi les facteurs clés de succès du déploiement d'une stratégie Big Data et sans prétendre à l'exhaustivité, on peut citer :

- **Une implication forte de la direction générale**

Le développement du Big Data nécessite le soutien de la direction générale, qui doit se traduire par la nécessaire et coûteuse mise en place d'un département de « data scientists » dépendant de la direction, mais aussi par la mise à niveau des compétences de l'ensemble des salariés.

- **Une culture de la donnée dans les entreprises**

Il importe pour l'entreprise de comprendre que les données ne sont pas seulement l'affaire des spécialistes, mais celle de tous, et que les pratiques de chaque salarié s'en trouveront modifiées. Pour développer une culture de la donnée dans l'entreprise, il faut donc s'interroger sur les moyens de capter, conserver et stocker les données, et développer ces moyens en conséquence.

- **Une transversalité indispensable en amont**

L'un des premiers enjeux dans le déploiement d'une stratégie Big Data consiste à mobiliser l'ensemble des directions qui disposent souvent de points de vue divers et complémentaires en matière de connaissance d'un client par exemple. S'il intègre l'ensemble des métiers de l'entreprise le plus en amont possible, le projet stratégique Big Data permettra d'effectuer le recensement le plus exhaustif afin de révéler des corrélations et d'en tirer des enseignements inédits.

- **Un cadrage stratégique des enjeux et opportunités**

Il faut consacrer un temps significatif à la qualification des besoins et enjeux de l'entreprise et à son diagnostic maturité. Et ensuite seulement choisir les technologies les plus à même de répondre aux besoins identifiés et qualifiés.

- **Une feuille de route agile et un plan d'actions concret**

La feuille de route à moyen terme doit définir des actions prioritaires concrètes, des objectifs à court et moyen termes, et des indicateurs clefs de performance (KPI) associés à

chacune des actions. Il est important que les décideurs et utilisateurs finaux puissent utiliser eux-mêmes (et interagir avec) les résultats qui leur sont remontés (reporting).

### • La protection de la vie privée

Les entreprises doivent prendre en compte, en amont de tout projet Big Data, les enjeux de risques techniques, juridiques et réputationnels. Elles doivent garantir la sécurité et l'intégrité des données traitées, la protection des données personnelles, tout en étant transparentes sur l'utilisation de ces données.

Selon le Cigref<sup>103</sup>, le Big Data n'est ni un projet de transformation, ni un projet «Système d'information», mais une manière innovante de penser et d'appréhender la donnée. C'est une démarche transversale qui vise l'optimisation de la prise de décision aussi bien sur des questions existantes que sur des interrogations non identifiées.

En somme et quels que soient le niveau de maturité et l'approche, l'ouverture et le partage des données tant en interne qu'à l'extérieur, ainsi que la nécessaire montée en compétences induite par cette démarche demeurent des conditions sine qua non de réussite de toute stratégie dans le domaine.

### Conclusion

Avec l'essor des réseaux Internet et Wi-Fi, la banalisation des appareils mobiles, le boom des réseaux sociaux, l'ouverture au public de certaines bases de données « Open Data » ou encore le développement de grands projets scientifiques internationaux, le phénomène du « Big Data » va en s'amplifiant. Les grands volumes de données générés quotidiennement représentent un tel enjeu économique, industriel et scientifique que les états et les entreprises investissent de plus en plus dans le Big Data. Plusieurs enquêtes au niveau mondial ont montré que les entreprises qui s'appuient sur l'analyse des données pour développer leur stratégie augmentent leurs niveaux de productivité, améliorent leurs processus de planification et de prévisions et accroissent la connaissance et la maîtrise de leur activité.

Dans cet article nous avons présenté les concepts fondamentaux du Big Data. Ces concepts incluent d'abord quelques éléments de définition du phénomène glanés dans la littérature spécialisée. Ensuite nous avons rapporté les technologies récentes développées dans le domaine. Nous avons passé en revue quelques applications directes de l'analyse des masses de données dans les différents secteurs d'activité privé et public. Enfin la question du

---

<sup>103</sup> Big Data : la vision des grandes entreprises françaises, Opportunités et enjeux, Octobre 2013

développement de projets Big Data dans les entreprises a été soulevée et nous avons cité quelques conditions pour réussir une telle stratégie.

Aujourd'hui, le Big Data est une réalité émergente au sein des états et des entreprises. Les gouvernements sont plus que jamais appelés à déployer un écosystème d'infrastructures, de compétences, de formation et de financement pour accompagner la filière Big Data.

### **Références Bibliographiques**

*Andrew McAfee, Erik Brynjolfsson, Big Data: the Management Revolution, Harvard Business Review, October 2012.*

*Brown Brad, Michael Chui et James Manyika, « Are you Ready for the Era of «Big Date»? , McKinsey Quaterly, octobre 2011.*

*Business & Decision, du Big Data au Big Business, Livre Blanc, 2014.*

*Cigref, Big Data : la vision des grandes entreprises françaises, Opportunités et enjeux, Octobre 2013.*

*Claude Brasseur, Enjeux et usages du Big Data : technologies, méthodes et mise en œuvre, Lavoisier, 2013.*

*Commission innovation 2025, Un principe et sept ambitions pour l'innovation, commission présidée par Anne Lauvergeon, Paris, La Documentation française, octobre 2013.*

*Ericsson, More than 50 Billion Connected Devices, White Paper 2011.*

*Ernst & Young Advisory, (Big) data: où en sont les entreprises françaises? Quelle maturité dans l'exploitation des données clients ?, 2014.*

*Institut Montaigne, Big data et objets connectés. Faire de la France un champion de la révolution numérique, Rapport, Avril 2015.*

*Ibrahim Abakar Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan, The rise of "Big Data" on cloud computing: Review and open research issues, Information Systems, 47, 2015, pp. 98-115.*

*Jafar Raza Alam, Asma Sajid, Ramzn Talib, MuneebNiaz, A Review on the Role of Big Data in Business, International Journal of Computer Science and Mobile Computing, Vol.3, Issue. 4, April-2014, pg. 446-453.*

*James Manyika, Jacques Bughin, Angela Hung Byers, Michael Chui, Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch, McKinsey Quarterly, August 2010.*

*James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, Big Data. The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, May 2011.*

*Khalid Benabdeslem, Christophe Biernacki, Mustapha Lebbah, Les trois défis du Big Data. Éléments de réflexion, Statistique et société, Vol. 3, N° 1, juin 2015.*

*Marcos D. Assunção, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya, Big data computing and clouds : Trends and future directions, J. Parallel Distrib. Comput., 79-80, 2015, pp. 3-15.*

*Myriam Karoui, Grégoire Devauchelle, Aurélie Duzert, Big Data : Mise en perspective et enjeux pour les entreprises, ingénierie des systèmes d'information, 2014.*

*Myriam Karoui, Grégoire Devauchelle, Aurélie Duzert, Systèmes d'Information et prise de décision à l'ère du «Big Data»: Le cas d'une entreprise française du bâtiment, 18<sup>ème</sup> Conférence Internationale de l'Association Information et Management, Lyon, 2013.*

*Nawsher Khan and al., Big Data: Survey, Technologies, opportunities, and Challenges. The Scientific World journal, 2014.*

*Pierre Delort, Le Big Data, Que Sais-je?, 2015*

*Shaokun Fan, Raymond Y.K. Lau, J. Leon Zhao, Demystifying Big data Analytics for Business Intelligence Through the Lens of Marketing Mix, Big Data Research 2, 2105, pp. 28-32.*

*Telecom ParisTech, Big Data : Nouveaux Défis, N°169, Juillet 2013.*

*Thomas H. Davenport, Paul Barth and Randy Bean, How 'Big Data' is Different? MIT Sloan Management Review, Vol.54, N°1, Fall 2012.*

*Victor Mayer-Schonberger , Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work and Think. Boston, New York, Eamon Dolan, Houghton Mifflin Harcourt, 2013.*

*Viktor Mayer-Schönberger, Kenneth Cukier, Big Data: la révolution des données est en marche. Robert Laffont, 2014.*

*World Economic Forum, Big Data, Big Impact: New Possibilities for International Development, 2012.*