

Introduction et enjeux du Big Data

Panorama en juin 2012

Kim Estivalet

Selon une enquête réalisée par le cabinet spécialisé en sécurité, Échelon One (commanditée par LogLogic), « Plus de sept décideurs sur dix ne comprennent pas (38 %) ou pas vraiment (35 %) ce qu'est réellement le phénomène du Big Data » [28].

Fait préoccupant au vu du Gartner Group qui, pour 2012, a cité le Big Data parmi les technologies stratégiques et désigné l'analytics dont le Big Data comme la première préoccupation technique des DSI [16].

Des données au « Big Data »

De nouvelles sources d'information

La problématique des données et de leur traitement n'est pourtant pas nouvelle pour les entreprises [11].

Il y avait déjà les données de connexion et de navigation Internet, les données d'habitude des sites marchands, l'utilisation des publicités, les pages vues...

La consomérisation, concept de BYOD (Bring Your Own Device), en 2011, a, sans nul doute, contribué à changer la donne en propulsant « à l'intérieur des environnements professionnels des données provenant d'environnements particuliers » [2].

Il faut maintenant faire face à un afflux soudain de nouvelles informations : de nouveaux canaux de communication où « l'utilisateur se fait directement producteur d'information » [11] : téléphones mobiles et sites Web, données de géolocalisation, télémétrie, Facebook, Twitter, médias sociaux, RFID, sondes et autres capteurs.

Cette mondialisation pressentie avec l'avènement du moteur de recherche Google a bouleversé les usages et « Ce qui était jusque-là connu d'une personne ou d'un petit groupe, un déplacement, un choix de livre, une conversation, gravite à présent quelque part dans la galaxie numérique. » [10]. Sans oublier l'Open Data [39] [40] [41] [45] et ses nouveaux usages [46] [58].

Cette production en temps réel peut sembler déconcertante et inutile.

Elle reste cependant « stratégique pour peu que l'on sache en extraire l'essentiel. » [11].

Roxane Edjlali, directeur de recherche au cabinet Gartner, confirme qu'il s'agit maintenant d'une réalité : « On peut choisir de l'ignorer ou de s'en servir. Chez Gartner, nous pensons qu'utiliser ces données de façon cohérente sera plutôt un avantage concurrentiel pour les entreprises » [27].

Premier enjeu : ces données sont-elles valorisables ?

Des systèmes dépassés

L'impact de ces nouveaux usages a immédiatement été ressenti sur les systèmes et la volumétrie.

Si à l'origine était l'octet qu'il a fallu compter en kilo-octets, méga-octets et Giga, il faut maintenant utiliser le Téra (= 10^{12} soit 1 000 000 000 000 pour les non-matheux), le Péta (= 10^{15}), l'Exa (= 10^{18}), et même le Zetta (= 10^{21}) !

Chaque minute :

- Youtube affirme recueillir un flux représentant 24 heures d'images [10].

Chaque jour :

- Zynga, éditeur de jeux pour les réseaux sociaux, traite 1 pétaoctet de contenu pour l'ensemble des joueurs [9] ;
- plus de 2 milliards de vidéos sont visionnées sur Youtube ;
- l'humanité, produit « la moitié des données qu'elle a produites depuis ses origines jusqu'à 2003 » [8].

Et chaque mois [9] :

- un adolescent envoie en moyenne 4 762 messages textuels ;
- 600 millions d'utilisateurs ont ajouté à Facebook plus de 30 milliards de données (messages, photos) ;

32 milliards de recherches ont été effectuées le mois dernier sur Twitter.

Deuxième enjeu : Peut-on conserver les mêmes systèmes pour demain ?

De nouvelles problématiques

Pourtant pas si nouvelles que cela.

Déjà en 2001, Doug Laney, analyste du Cabinet Gartner, définissait les défis autour de la croissance des données ou « Big Data » [17].

Cette évolution impliquant toutes les catégories de population [8] est naturelle et, aujourd'hui, ces problématiques sont réelles.

Force également est de constater que les architectures traditionnelles atteignent leurs limites :

- Les systèmes traditionnels comme les SGBD ne peuvent gérer rapidement ces flux massifs d'informations non structurées. Car ces données nouvelles « ne remplissent pas les conditions très formalisées des bases de données » [2].
- La multiplication des volumes ne peut impliquer la multiplication parallèle des supports de stockage et rester rentable.
- L'analyse en temps réel de ces multiples sources d'informations est également devenue un enjeu économique majeur.

Les nouvelles architectures doivent aussi répondre aux besoins de : « l'extraction des données, la découverte, le nettoyage pour passer des volumes à la pertinence, l'intégration au sein d'un système de référencement, l'analyse pour dégager de la valeur des données ainsi traitées... » [8].

Sans compter que l'entreprise doit maintenant faire face à la « non-connaissance a priori du contenu de l'information à analyser » [36].

En conséquence, « L'institut d'étude Forrester estime que moins de 5 % des données disponibles dans les organisations sont utilisées avec efficacité. En cause, plus leur coût que les technologies.

Sans offre spécifique, le coût d'exploitation de 95 % des données disponibles et sous exploitées représente un montant trop important à supporter » [18].

Troisième enjeu : quelles solutions pour quel budget ?

Ces forts enjeux permettent-ils aux acteurs de s'accorder autour d'un consensus ?

Seulement un concept ?

En première approche, Laurence Dubrovin (CXP) résume simplement la situation : « Le Big Data désigne l'accumulation de tout type de donnée (structurée ou non) de toute provenance atteignant une taille et un niveau de granularité difficile à exploiter avec les outils classiques » [35].

Ce qui peut sembler suffisant, car définir le Big Data semble difficile, car plutôt vague.

Comme le résume le LCIA [32] : Big Data n'est pas un terme précis, mais plutôt une caractérisation de l'accumulation sans fin de tous les types de données, la plupart non structurées.

Il décrit des ensembles de données qui sont en croissance exponentielle et qui sont trop volumineux, trop bruts ou trop déstructurés pour être analysés en utilisant les techniques de base de données relationnelles.

Plus pratique, silicon.fr cite « La gestion et l'analyse des données non structurées qui envahissent les systèmes d'information, réclament l'usage de nouveaux outils et services, de matériels et logiciels, de frameworks, réunis sous le vocable de Big Data » [15].

Voilà là un élément commun de toutes les approches, également vu par IDC : « la capacité à rechercher, à extraire et à analyser les informations non structurées qui représentent la très grande majorité de ces données » [15].

Nous trouvons maintenant les premiers éléments de synthèse : de nouveaux outils et techniques, une explosion de la masse d'informations, capacité à pouvoir analyser, traiter, extraire et prendre des décisions pertinentes. Ce qui, comme le rappelle GFI, lie, en fait, trois concepts :

- « la capacité à adresser et stocker de fortes volumétries de données, qu'elles soient structurées ou non structurées ;
- l'interaction avec les nombreuses sources d'informations extérieures à l'entreprise et non maîtrisées par elle ;
- l'analyse de l'ensemble de ce nouveau capital informationnel. » [36].

Ces concepts sont mieux détaillés par les « 3 V ».

Les 3 V du Big Data

En 2001, Doug Laney, analyste du cabinet Gartner, décrivait le défi du Big Data « comme tridimensionnel : une augmentation en volume, vitesse et variété » [17].

Nous pourrions également les retrouver sous le terme : Volume, Vélocité, Variété décrits dans cette autre présentation du Gartner [30] :

- **Le premier V pour Volume** : La multiplication des formats « de toutes les directions, sous toutes les formes : voix, texte, images, impulsions électroniques, traduisant des variations de température ou de cours de Bourse [...] » [10] implique, selon une récente étude d'IDC, que les volumes de données enregistreraient une progression de... 400 % par an. L'enjeu pour les entreprises est de ne plus se contenter de simples échantillonnages. « De cette façon, les entreprises utilisent vraiment la Voix du Client et comprennent les causes réelles de mécontentement ou de satisfaction » [17].

- **Le second V pour Vélocité** : Plus qu'un gadget, c'est une obligation économique contrainte par la nécessité de pouvoir prendre des décisions pertinentes de plus en plus rapidement afin de garder ses avantages concurrentiels.

La valeur réside dans la fidélité de l'image obtenue qui sera d'autant plus réaliste qu'elle sera proche du moment. « Ces informations doivent être analysées, partagées en temps réel et ceci très rapidement afin d'appliquer les changements au plus tôt : un redressement rapide vis-à-vis du client ou bien une formation pour un agent peu performant » [17].

L'article du Gartner : « Business requirement : Moving to real Time Information » [30] complète bien cette analyse.

- **Le troisième V pour Variété de sources** : aujourd'hui multiples de par les appareils et les sources de saisie comme vu dans le chapitre précédent, mais également de par la mondialisation et la géographie des entreprises internationales.

Cette multiplicité engendrant essentiellement des données dites « non structurées ».

L'enjeu est ici l'analyse multidimensionnelle afin d'avoir « une vue complète du cycle de vie du client pour comprendre avec exactitude son parcours depuis le point de vente jusqu'au renouvellement du contrat ou de son prochain contact » [17].

Cette approche plus complète répond donc aux problématiques métiers majeures « comme mieux connaître son client, améliorer son offre, agir avant ses concurrents » [8].

Les 4 V

S'il y avait besoin de simplifier ces approches, certains analystes comme Gartner poussent le concept aux « 4 V » : « Velocity, Volume, Complexity, Variety » [30].

On distinguera ici la Valeur des données où l'enjeu réside dans la capacité à explorer et extraire les données pertinentes « parmi tant d'autres données insignifiantes » [35].

Et au-delà...

Les enjeux et usages étant multiples et encore mal identifiés, chacun pourra y apporter son expertise :

- Un premier exemple serait un cinquième V « Validité » [49].

- Un deuxième exemple serait fourni par Roxane Edjlali : le phénomène Big Data « va affecter votre infrastructure sur douze axes [...]

Aux quatre V déjà cités (volume, vitesse, variété, complexité), elle ajoute la périsseabilité des informations : Combien de temps va-t-on les garder ?

Les données doivent aussi être corrélées, reliées à d'autres qui les enrichissent. Cela fait partie de la qualification de l'information.

Si on ne peut pas relier l'information à d'autres, est-elle bien utile ?, pointe Roxane Edjlali ; elle cite encore tous les aspects juridiques sur la rétention de l'information » [27].

Pour cadrer le sujet, faisons donc un premier tour des usages.

Des usages multiples

Un des premiers à s'engouffrer dans la brèche fut notamment Google : « Google is a Big Data company. The future of search is going to be informed by the field of Big Data » [50].

Ce premier cas est donc le plus simple, mais les cas d'utilisation suivants issus de sources de l'industrie expliquent combien les types de situation dans lesquels les Big Data sont utilisés peuvent être multiples et complexes : moteur de recommandation, analyse de la campagne marketing, fidélisation de la clientèle et analyse de désabonnement, graphe d'analyse sociale, analyse de marchés, analyse prédictive, gestion des risques, détection d'opérations malhonnêtes, détection de fraude, banque de détail, surveillance de réseau, recherche et développement, archivage (cf. Livre blanc Talend [34]).

Pour dégager quelques tendances, il faudrait citer :

- en tant qu'entreprises, les acteurs de la grande distribution, des télécoms, les acteurs internet et les instituts de sondage ;
- et attendus par les équipes métiers de l'entreprise :
 - « Marketing : améliorer le cross-selling, développer le géo-marketing, entreprendre du marketing comportemental (lié à l'analyse des comportements), permettre la micro-segmentation client,
 - Merchandising : optimisation des assortiments, optimisation des prix, optimisation des emplacements
 - Opérations : optimisation des usages réseaux, détection des fraudes, prévention des incidents et amélioration de la qualité de service,
 - Logistique : gestion des inventaires, optimisation de la logistique, optimisation des négociations fournisseurs » [36].

Les thèmes dont nous pouvons détailler des utilisations probantes sont décrits ci-après.

Données clients et données sociales

Comme le rappelle Aaron Zornes (directeur de la recherche au MDM Institute aux États-Unis), « Les entreprises estiment avoir une vision à 360° de leur clientèle rien qu'en achetant des données clients auprès de Dun & Bradstreet. Cette vision n'est en fait que de 75°. Si vous souhaitez avoir une vue d'ensemble, vous devez également trier la masse d'informations, présente en ligne, sur Facebook, Twitter et LinkedIn. » [12].

Si le mailing reste un usage cible des responsables marketing, l'utilisation du Big Data leur permet de mieux « acquérir, qualifier, traiter et analyser les données » [8].

On pourra citer cet exemple :

- American Express a aidé les PME à augmenter leurs revenus d'années en années de 28 % pendant le week-end suivant Thanksgiving avec « Small Business Saturday », une campagne impliquant 1.5 millions de fans Facebook » [8].
- Un autre intérêt réside dans l'utilisation de l'effet Benjamin Franklin : « En demandant à un opposant politique de lui prêter un livre, Franklin a trouvé qu'il était traité avec davantage de respect. Juste en rendant service à Franklin, cet opposant a résolu son attitude dissonante en adoucissant son point de vue sur Benjamin Franklin. La loyauté du client peut être aussi améliorée en utilisant l'effet Benjamin Franklin en demandant simplement un retour d'avis. Le fait de donner son avis à une entreprise est similaire à lui faire une faveur et en retour le client aura un avis plus favorable sur elle » [17].
- Des outils comme Facebook et Twitter sont effectivement largement intégrés dans les sites Web, permettant aux entreprises d'interagir avec les avis clients, que ces avis soient positifs ou négatifs [1]. Ce qui offre également « une gamme d'analyse très variée » [17].

Le Big Data devrait donc « faciliter la fidélisation client » [47].

Sciences et recherches

Trois faits significatifs :

- 200 M\$ ont été débloqués aux États-Unis pour développer les technologies de Big Data dans le domaine de la recherche scientifique [29] ;
- La France a fait de même avec 25 M€ [29] ;
- Un téraoctet de données est généré par chaque expérimentation sur l'ADN [8].

Dans les autres secteurs

Des exemples déjà probants :

- **Géologie** : « un séquenceur pour calcul sismique génère 2 téraoctets par run et des millions de fichiers associés qu'il faut prendre en compte, traiter et interpréter » [8].
- **Compagnies pétrolières** : « prendre en compte les millions d'informations collectées par l'ensemble des capteurs et appliquer les patterns de reconnaissance sur le flux de données – Stream » [8].
- **Organismes de crédit** : « coupler un système Hadoop avec un Datawarehouse et un système de détection de fraude pour assurer le suivi en temps réel des transactions » [8].
- **Constructeurs automobiles** : « qui disposent de plus en plus de possibilités de suivre le vieillissement de la flotte des véhicules qu'ils ont vendus et les habitudes d'utilisation de leurs clients, le Big Data accompagne la métamorphose de ces industries lourdes en entreprises de service » [10].
- **La municipalité de Memphis Tennessee** : « revendique une baisse de 30 % de la criminalité depuis la mise en œuvre de nouvelles méthodes de gestion de ses rondes de police appuyées sur l'utilisation d'informations compilées en grand nombre (citation IBM) » [10].
- **Equifax** (Bureau de Crédit) utilise le Big Data pour créer de nouveaux produits à partir de 800 milliards de données clients et business à travers le monde [54].
- **Le gouvernement fédéral américain** : « a mis en place un système permettant la collecte de quelques 370 000 données brutes et géolocalisées à partir de 172 de ses agences ou sous – agences. Celles-ci sont désormais disponibles à partir de 230 applications disponibles sur un portail dédié permettant un accès public à ces informations non classifiées impossible auparavant » [18] [32].

On pourra aussi citer : la Logistique et le fret maritime [37], les jeux olympiques [56] et d'autres secteurs bientôt dans la course :

- Les assureurs, les gestionnaires de risques, les banquiers : « réfléchissent tout autant aux nouvelles possibilités que leur offre le Big Data » [10].
- Gouvernements : « qu'il s'agisse de mieux faire entrer l'impôt, d'améliorer la qualité des services publics, de prévenir les fraudes, les nouvelles capacités de compilation et d'analyse des données leur offrent un immense potentiel » [10]. Mais là, nous ne sommes pas pressés.

Sémantique et Analytique

À propos de la sémantique : le dictionnaire ne suffit plus. Pour qu'un outil comme Google puisse identifier qu'un groupe de personnes parle de téléphone portable et non de pâtisseries, même si certains mots utilisés peuvent être communs « orange » ou « orange ». Il faut qu'il puisse interpréter le contexte avec les éléments complémentaires [10].

Pour les entreprises souhaitant coller à l'actualité d'un domaine particulier en mutation permanente, c'est un défi nécessaire et peu évident. « Il faut savoir ce que l'on peut en faire » [10].

En capacités d'analyses, nous devrions également voir l'essor de :

- l'analyse prédictive avec quelques réserves car si de nombreuses entreprises se tournent vers cet outil, l'analyse automatique de données ne peut cependant « se substituer entièrement à l'expertise humaine. [...] les analyses prédictives sont une science, qui vous permettra d'aller du point A au point B, mais deviennent un art quand vous faites appel à l'expertise professionnelle de la personne qui prendra, en bout de course, la décision et qui devra tenir compte de l'importance du parasitage » [21] ;
- l'analyse de sentiments qui « s'imisce désormais dans les outils analytiques afin de répondre aux problématiques Big Data des entreprises » [55].

Réglementations et sécurité

Là aussi l'entreprise doit jongler entre l'imposé et le nécessaire. Les contraintes et besoins à ce niveau sont maintenant multiples. Entre autre :

- Réglementation dont « la gestion des risques ou de la fraude pour les banques ou la LCEN (Loi pour la confiance dans l'économie numérique) dans les télécoms » [36] ;
- Sécurité et conformité : « Si le Big Data est géré correctement, il fournit des données qui permettent d'améliorer la sécurité et de répondre aux problématiques liées à la conformité » considère Mandeep Khera, Directeur Marketing de LogLogic [28].

Des solutions en maturation

Alors « Big Data, architecture ou logiciels ? ». [6]
La réponse devient, bien évidemment, les deux !

Infrastructures et stockages

Les acteurs comme Hitachi [4] se sont déjà adaptés et proposent de nouvelles infrastructures de stockage.

On notera que l'indice Big Data de Jaspersoft, « englobe un très vaste ensemble d'environnements NoSQL et Big Data, dont Hadoop Hive, Hadoop HBase, MongoDB, Cassandra, Riak, Infinispan, Neo4J, Redis, CouchDB, VoltDB, etc. » [16].

Il ne semble donc pas y avoir de difficultés de ce côté, tous les acteurs historiques proposent leurs solutions.

Hadoop et Map Reduce

Hors stockage, il faut s'outiller de solutions analytiques permettant de chercher, retrouver, exploiter les données.

Les solutions actuelles ne sont plus adaptées à la gestion des volumes en jeu. Les Frameworks de type Map Reduce tentent de résoudre ce besoin. Ils permettent d'explorer, révéler les données chaudes, brutes et potentiellement non structurées. Limitant les besoins CPU et réduisant l'inertie des traitements.

Le projet applicatif Hadoop s'est pour l'instant imposé dans ce contexte au-dessus des couches matérielles. « Même s'il est encore loin d'être stable dans ses développements, Hadoop présente un avantage et un inconvénient : porté par une communauté open source, il emporte un message bourré d'opportunisme économique ; en revanche, il affiche une rare complexité qui est loin de le rendre accessible au commun des DSI » [24].

Au plus simple, on peut le définir ainsi : selon Wikipedia Hadoop « est un Framework Java libre destiné aux applications distribuées et à la gestion intensive des données. Il permet aux applications de travailler avec des milliers de nœuds et pétaoctets de données. Hadoop a été inspiré par les publications MapReduce, GoogleFS et Big Table de Google » [6]. Mais pour approfondir ce sujet, on pourra se reporter à l'intéressant article de Romain Chaumais [22].

L'impact sur les acteurs de l'IT a été non négligeable : Microsoft a même fini par troquer sa technologie propriétaire Windows HPC Server (abandonnée en novembre dernier) contre une solution « Hadoop native » sur Windows Azure, tandis que la technologie Map Reduce développée par Aster Data, rachetée (trop tôt ?) par Teradata, n'a pas eu le temps de percer. À l'inverse, Amazon Web Services (AWS) a bâti son offre de Big Data sur le moteur Hadoop proposé en mode On Demand tandis qu'Oracle installe Cloudera au cœur de sa Big Data Appliance [22].

En moins de quatre ans, Hadoop est devenu la référence sur les solutions de type Map Reduce, étouffant la concurrence. « Les éditeurs traditionnels en sont ramenés à un rôle d'assemblage de briques complémentaires, sommées d'interagir autour de ce même noyau ouvert » [22].

Deux intérêts majeurs font que cette situation devrait durer encore quelque temps :

- Une normalisation bénéfique de la solution ;
- Hadoop a été placé au cœur des offres de traitement de données en volume par les ténors de l'industrie principalement pour son coût [2].

Autres services

Comme il l'a été vu dans les usages, le Big Data doit offrir une famille complète d'outils de la source à la décision. Il faudra donc explorer tout le processus [8] : les composants d'acquisition, la qualification des données, l'analyse de flux en temps réel, des briques datawarehouse, le Master Data Management, l'analyse prédictive, le data mining, les statistiques, l'intelligence artificielle, le traitement du langage, etc.

Pour les tendances, toujours selon l'indice Big Data de Jaspersoft, on a pu observer l'année passée « une nette augmentation de la demande pour des méthodes plus rapides et plus transparentes permettant de se connecter, d'analyser et de présenter les informations extraites des Big Data.

Ces données reposent sur le nombre total de téléchargements de connecteurs sur JasperForge, le site Web communautaire Open source de Jaspersoft » [16].

Au final, chaque éditeur propose bien ses propres outils. Par exemple chez IBM [6] :

- Stream et Big insight font partie de la gamme Infosphere.
Le « IBM Entity Analytics group » a développé entre autre, InfoSphere Identity Insight Solutions pour l'analyse en continu de jeux de données de milliards de lignes en temps réel avec des ordres à la milliseconde ; [13]
- déclinaisons de services et de logiciels d'analyse prédictive, Smarter Analytics Signature Solution [26], (lutte anti-fraude, relation client et risque financier).

On pourra approfondir ces différentes solutions avec les liens fournis : « 7 top tools for taming big data » [42], IBM [8] [60], Oracle [31], Talend et autres solutions [34].

Big money ?

De nouvelles perspectives pour les utilisateurs finaux et les entreprises.

Certains n'ont pas attendu et même si la valorisation actuelle de Facebook autour des 100 milliards de dollars peut être controversée, sa valeur hors norme basée sur la base de données autoalimentée par les 900 millions d'utilisateurs est indéniable. « Les premiers à avoir su capter, stocker et commencer à utiliser cette nouvelle matière première sont déjà riches » [10].

Pour les autres, si le bénéfice net n'est pas disponible, la valeur est indiscutable :

- identifier des compétences et des savoirs : « La production par des salariés de commentaires sur Intranet, d'échanges entre collègues, de notes ou de posts informels sur les réseaux de l'entreprise peut s'avérer précieuse » [11] ;
- améliorer l'efficacité opérationnelle et générer de l'innovation : « Pour les entreprises, le Big Data offre des opportunités importantes [...], considère Dan Vesset, vice président Business Analytics Solutions d'IDC » [16] ;
- un gisement de connaissances clients et de performances : « qui, bien analysé et traité, s'avérera un puissant atout économique » [11] ;
- amélioration du ratio d'informations disponibles : « deux entreprises peuvent avoir la même efficacité dans le traitement des données mais si l'une d'entre elle arrive à exploiter 15 % des data disponibles là où la seconde n'est qu'à 5 %, la première bénéficie d'un avantage compétitif de taille » [18] ;
- Time To Market : « selon les DSI du Leadership Council, bien utilisé le Big Data peut conduire à améliorer significativement les efforts de développement des entreprises en redirigeant efficacement les flux d'information, en aidant à débusquer les blocages dans l'organisation, en fluidifiant la chaîne logistique, en optimisant la connaissance des clients et en permettant l'émergence de nouveaux produits, services ou modèles » [18] [32],
- Amazon : « affirme réaliser aujourd'hui 30 % de ses ventes grâce aux propositions ciblées qui apparaissent à l'écran en fonction des préférences exprimées par l'internaute » [10].

Quelques résultats donc, mais surtout beaucoup d'attendus.

Du côté des acteurs, les chiffres semblent plus clairs.

Un marché conséquent pour les acteurs

En 2010, le cabinet McKinsey prédisait déjà « une augmentation de 60 % de la marge d'exploitation des distributeurs qui utiliseraient pleinement ces énormes volumes de données » [59].

Ce qui n'a pas manqué de susciter des convoitises : « EMC2, Teradata, HP, Microsoft, Google, Amazon, Informatica, SAP/Sybase, Oracle, SAS, MicroStrategy, Syncsort, Talend, Jaspersoft, Pentaho, Tableau Software etc., mais aussi des experts de centre de recherche (Inria, etc.) et des acteurs innovants dans le domaine de la visualisation des données (Captain Dash, Dataveyes, etc.) » [35].

Plus précisément, selon IDC, le marché des technologies et services Big Data précédemment évalué à 3,2 milliards de dollars en 2010 devrait atteindre 16,9 milliards en 2015. [35] [15]. Soit une progression de 40 % par an jusqu'en 2015. Pour relativiser : « sept fois plus rapidement que l'ensemble des technologies de communication et d'information » [15] [11] [66].

Tous se mobilisent donc et les mouvements sont nombreux :

- Parmi les start-up et Hadoop : « Au départ Cloudera et Datameer. L'éco-système commercial d'Hadoop au début 2011 était essentiellement composé de start-up » [2] ayant pris la vague du Big Data, de l'Open Source et du traitement des données non structurées et « basé leur modèle économique sur une déclinaison commerciale de Hadoop » [2]. « À ce jour, plus de 500 millions de dollars ont été investis par le capital-risque dans la création de jeunes pousses spécialisées dans le Big Data » [16] ;
- Netezza et « Enterprise Search Software Company Vivisimo » [43] [44] pour IBM ou Greenplum pour EMC ;
- 2011 : HP avec Vertica et Autonomy ;
- sans oublier le rachat de Platform Computing par IBM ou celui d'Aster Datapar Teradata en mars dernier.

Ceci n'est pas prêt de cesser. « En 2012, l'offre Big Data devrait se consolider, et les éditeurs combler encore les manques de leur technologie » [2].

Il faut préciser pourtant que cette croissance sera inégalement répartie : « serveurs (27,3 %), logiciel (34,2 %), ou stockage (61,4 %). De même, plusieurs approches pourront être adoptées pour déployer une solution : des appliances, de l'outsourcing, voire du cloud computing » [15].

Le choix d'usage appartenant aux organisations.

Et, qui plus est, ce qui est facile à dire n'est pas forcément facile à faire.

Aaron Zornes (directeur de la recherche au MDM Institute aux États-Unis) « note que les fournisseurs de technologies ont été très lents à adresser les problématiques de la gouvernance des données notamment. Seul, un nombre réduit d'entre eux a posé ses jalons sur ce segment, et rare sont ceux qui s'y sont véritablement engouffrés » [12].

Ce qui permet à certains de douter de ces prévisions : Peter Fader (codirecteur du 'Wharton Customer Analytics Initiative à l'Université de Pennsylvanie [48] se rappelle de ce qui a pu arriver au CRM (Customer Relationship Management) depuis 15 ans, la réalité s'étant avérée bien au delà des prévisions.

Un exemple de frein pourrait effectivement être l'orientation actuelle du Big Data vers la découverte. La gestion de ces masses d'information par les outils existants peut entraîner une élimination naturelle des valeurs aberrantes. Ce qui, selon les cas, serait exactement ce qu'il fallait trouver (cf. Philip Russom, TDWI ou The Data Warehousing Institute) [53]. La solution est fournie par Hal Varian (Google's chief Economist) qui précise que même s'il faut toujours utiliser des échantillons aléatoires, les Big Data permettent de générer les ensembles de données les plus valables pour la modélisation [53].

Rien n'est donc facile et les outils devront prendre le temps de résoudre, un par un, les nombreux écueils opérationnels permettant de générer le Business tant attendu.

Confrontés à la réalité opérationnelle, il faut donc revenir à la raison qui ne peut être appréhendée que par les utilisateurs finaux et entreprises qui donneront finalement le vrai rythme. Loin du miracle business attendu et défendu par les acteurs de l'IT. La vérité est donc, comme toujours, certainement entre les deux.

Visions et enjeux du « Big Data »

Afin d'éviter toute fausse interprétation des prévisions, je vous propose cette fois-ci une simple sélection des affirmations les plus marquantes.

À court terme : 2012-2013

Marché

- Technologie stratégique ? « Le Gartner a inclus le concept Big Data dans son traditionnel Hype Cycle 2011 et l'a classé parmi les 10 technologies stratégiques pour 2012. Il est donc logique que le marché se mette en place » [2].

Matériels

- **BYOD ?** La tendance « Bring Your Own Device » (BYOD) change la donne de l'équipement informatique. Le PC de l'entreprise est de plus en plus dépassé [5].

Données

- **Nouveaux paramètres ?** « le besoin de prendre en compte des données dont on ne maîtrise pas nécessairement encore tous les paramètres » [7].
- **Temps réel ?** « le besoin de prendre en compte ces données dès maintenant » [7].
- **Web Analytics ?** « Ne plus pouvoir suivre les utilisateurs qui optent pour les cookies ce qui aura un impact extrêmement important sur la façon dont nous mesurons la qualité du site par l'intermédiaire d'analyse [1].

- **Données non structurées ?** Extension des analyses aux données volumineuses : « les données non structurées stockées sur les serveurs de fichiers, sur les périphériques NAS et sur les systèmes de messagerie électronique » [3].
- **Scientifiques des données ?** « Les scientifiques des données deviendront des postes-clés dans les entreprises » [20].
- **Open Data ?** et « Data transparency coalition » [38].

PC de l'entreprise et sécurité

- **Menaces internes :** « Les menaces internes resteront une préoccupation majeure » [3].
- **Plans de sécurité ?** L'enjeu sécuritaire n'est pas seulement une crainte mais un vrai risque économique : « La récente histoire de Sony lui a coûté 170 M\$ et potentiellement plus de 1 milliard d'opportunités perdues (Forrester : Quelles approches face à la nouvelle cybercriminalité ?). Forrester indique que les entreprises devraient commencer à élaborer maintenant, des plans de sécurité Big Data, de manière à réduire les coûts, les risques et les problèmes de déploiement » [7].
- **Analystes pour la sécurité informatique ?** Brian Fitzgerald, vice-président de RSA en charge du marketing, constate déjà un phénomène de convergence : « De nombreux responsables de la sécurité informatique ont beaucoup de données à traiter et recherchent des personnes qui connaissent l'analytique. Et même, de préférence, des analystes venus du décisionnel, qui peuvent étudier les données de sécurité sans idées préconçues, qui sont capables de trouver des schémas logiques au milieu de toutes ces données » [14].
- **Coût des violations de données ?** Les pertes ou les vols de données informatiques ont coûté en 2011, 16 % de plus que l'année précédente aux entreprises françaises, soit 2,55 M€. Ce qui s'explique par la multiplication des attaques très ciblées, l'utilisation croissante des terminaux mobiles et les nouvelles réglementations. Selon l'étude « Cost of a Data Breach » du Ponemon Institute, commanditée par Symantec, le coût par donnée compromise passe de 98 € à 122 €, le coût des violations de données a augmenté de 34 % en 2 ans [25].
- **Secure collaboration ?** « La Secure Collaboration va exploser en 2012 » [3].

Organisation

- **Entreprises non préparées ?** « Le Big Data et le Cloud se développent, les exigences de conformité sont des préoccupations majeures, mais la majorité des entreprises n'est pas préparée pour y faire face de manière adéquate » précise Bob West, Fondateur et CEO d'Échelon One [28].

- **Modèle d'entreprise et analyse des données ?** « Big Data est un sujet désormais incontournable pour les entreprises. Pour y faire face, elles doivent commencer leur transition vers un modèle d'entreprise qui place l'analyse des données au cœur des processus de prise de décision » [20].
- **Automatisation intelligente ?** « Une gouvernance des données efficace nécessite l'exploitation de la puissance des métadonnées grâce à l'automatisation intelligente. Il n'est guère surprenant de voir les experts de l'industrie affirmer que ce type d'automatisation est désormais nécessaire pour une meilleure gouvernance » [3].

À moyen terme : 2013-2014

- **Déploiement de la tech nologie ?** « IDC voit trois scénarios principaux pour le déploiement de la technologie Big Data, qui vont imposer des choix stratégiques innovants pour le stockage (scale-out) et le traitement des données (Hadoop et MapR) : — Déploiement pour des données collectées supérieures à 100 To dans un environnement In Memory (les données demeurent dans la mémoire des serveurs et non pas sur les disques); — Déploiement sur des technologies de messagerie en très grands volumes et en temps réel pour la capture et l'analyse des données (majoritairement non structurées); — Déploiement sur des métiers où les volumes de données encore réduites aujourd'hui vont grossir très rapidement » [15].
- **Tablettes et données mobiles ?** « eMarketer prédit qu'il y aura 90 millions d'utilisateurs de tablettes d'ici 2014 » [1].
- **Exploiter les données quelles qu'elles soient ?** « les entreprises doivent donc passer de l'époque où elles optimisaient leurs processus de gestion des données (à travers les ERP notamment) à un nouvel âge où elles doivent maintenant exploiter les données quelles qu'elles soient pour mieux comprendre et analyser leur écosystème et ainsi adapter leur business model. Le temps de l'optimisation des processus en vase clos est aujourd'hui dépassé et les entreprises doivent encore d'avantage tenir compte du monde extérieur pour améliorer leur compétitivité » [36].
- **Orienté Données ?** « signifie : prendre des décisions sur base de données. Les données sont un facteur clé de succès et les entreprises qui en font bon usage gagneront en compétitivité » [20].
- **Politique et gouvernance de données ?** « Selon PwC, à horizon trois ans, 30 % des grandes entreprises vont mettre en place des politiques et des cellules de gouvernance de données » [23].

- **Adapter la législation ?** [49], Forrester [7].
- **Big Brother ?** Les yeux d'Equifax vous regardent- « Big Data Means Big Brother ». Equifax (Bureau de Crédit) utilise Big Data pour créer de nouveaux produits à partir de 800 milliards de données clients et business à travers le monde. Son CIO affirme « Nous en savons bien plus sur vous que ce vous voulez bien que nous sachions » [54].
- **Capacités technologiques ?** « En 2015 ; 15 % des entreprises devront adapter leurs capacités de technologiques pour les données extrêmes, le contenu socialement médiatisé et les nouveaux dispositifs connectés » [30].
- **Infrastructure unique ?** « la possibilité mentionnée par certains que Big Data soit l'infrastructure unique de production dans les années à venir au détriment des architectures aujourd'hui considérées comme traditionnelles » [8].

À long terme : 2015 et plus

Consommation des données

- **Consommation des données ? :** « En 2011, l'utilisateur moyen de Smartphone aura consommé 15 Mo de données par jour, et ce volume passera à 1 Go à l'horizon 2020 » [5].
- **Données personnelles ? :** « En 2015, environ 3 milliards de personnes seront connectées et partageront des (leurs) données. Le volume attendu est de l'ordre de 8 Zettabytes (soit 8 milliards de Tb)» [9].
- **Données non structurées ? :** Parmi les 10 tendances stratégiques pour 2012 mentionnées par le Gartner, Big Data, les schémas et l'analytique arrivent en deuxième position. La quantité de données non structurées va croître de 80 % dans les cinq ans à venir, avec à la clé un sacré challenge pour les directions informatiques » [9].
- **Quantité des données ?** Gartner prévoit une croissance de « 800 % des quantités de données à traiter d'ici cinq ans » [8].

Infrastructures et réseau

- **Trafic IP ? :** Le trafic IP international quadruplera d'ici à 2015 en partie grâce au Cloud [9].
- **Stabilité et performance ? :** « Pour rester compétitives, les entreprises devront donc chercher des solutions innovantes leur permettant d'améliorer la stabilité et les performances du réseau tout en réduisant les coûts. Celles qui ignorent cette tendance seront confrontées à un problème majeur et il y a fort à parier qu'il y aura quelques exemples significatifs » [5].

Sécurité

- **Encryptage ? :** « À l'avenir, l'encryptage va devenir le pivot de la sécurité des données et de la stratégie de confidentialité », Forrester [7].

Organisation

- **Modèles de gouvernance ?** Gartner : « Information Governance 2020 : Traditional Governance Models Are Likely to Break » [30].

Gouvernements et sociétés

- **Réorganisation des sociétés ? :** « Il y avait eu la découverte du pétrole, il y aura l'ère de la « Big Data », nouvelle matière première susceptible de provoquer des réorganisations en profondeur des sociétés en changeant la manière de faire des affaires ou de gouverner » [10].
- **La crainte du « Big Brother » ? :** « Ce ne sera pas simple à mettre en œuvre s'inquiètent les professionnels. Le Big Data a un impact sociétal et l'industrie qui promeut son développement doit anticiper une phase de négociation de nouvelles réglementations visant à en préciser les usages » [10].
- **Et toujours Big Brother ? :** Les yeux d'Equifax vous regardent - « Big Data Means Big Brother ». Equifax (Bureau de Crédit) utilise Big Data pour créer de nouveaux produits à partir de 800 milliards de données clients et business à travers le monde. Son CIO affirme « Nous en savons bien plus sur vous que ce vous voulez bien que nous sachions » [54]. ▲

kestivalet@free.fr

Pour aller plus loin

Quelques conseils :

- Feuille de route en cinq points établis par le LCIA « pour aider les organisations à progresser dans leur adoption du Big Data et à améliorer l'utilisation efficace de leurs données » [18] [32] ;
- « Five Things CIOs Should Know About Big Data » [52] ;
- « 10 Commandments of Digital Analytics » [57] ;
- « How to Avoid Big Data Spending Pitfall » [51] ;
- Le livre blanc de Fujitsu [33].

Références :

1. « Biggest Search Events of 2011 & Predictions for 2012 », <http://searchenginewatch.com/article/2133530/Biggest-Search-Events-of-2011-Predictions-for-2012>, publié le 21 décembre 2011.
2. « État du monde IT : le Big Data, une offre qui se forme », http://www.lemagit.fr/article/donnees-hadoop-big-data/10136/1/etat-monde-big-data-une-offre-qui-forme/?utm_source=essentielIT&utm_medium=email&utm_content=new&utm_campaign=20111227&xtor=ES-6, publié le 26 décembre 2011.
3. « Quelle gouvernance des données en 2012 selon Varonis », <http://www.infodsi.com/articles/126962/gouvernance-donnees-2012-selon-varonis.html?key=>, publié le 27 décembre 2011.
4. « Hitachi Data Systems Stockage et infrastructures IT, quelles tendances pour 2012 », <http://www.infodsi.com/articles/126964/hitachi-data-systems-stockage-infrastructures-it-tendances-2012-1-2.html?key=> et <http://www.infodsi.com/articles/126963/hitachi-data-systems-stockage-infrastructures-it-tendances-2012-2-2.html?key=>, publié le 27 décembre 2011.
5. « L'année 2012 selon Brocade : BYOD, Big Data et montée en puissance du cloud computing », <http://www.infodsi.com/articles/126982/annee-2012-selon-brocade-byod-big-data-montee-puissance-cloud-computing.html?key=>, publié le 28 décembre 2011.
6. « Qu'est-ce que Big Data ? », <http://www.bpmbulletin.com/2012/01/13/quest-ce-que-big-data>, publié le 13 janvier 2012.
7. « Big Data : Opportunités et problèmes de sécurité », <http://www.infodsi.com/articles/128592/etudes-forrester-big-data-opportunités-problèmes-sécurité.html?key=>, publié le 06 février 2012.
8. « Big Data : cas d'usages, exemples, intégration, retour sur la matinée 01 Big Data », <http://www.bpmbulletin.com/2012/02/14/big-data-cas-dusages-exemples-integration-retour-sur-la-matinee-01-bigdata/>, publié le 14 février 2012.
9. « Big Data et les 10 tendances stratégiques pour 2012 par le Gartner », http://www.bpmbulletin.com/2012/02/21/big-data-et-les-10-tendances-strategiques-pour-2012-par-le-gartner/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+bpmbulletin+%28BPM+Bulletin%29, publié le 21 février 2012.
10. Jacques Secondi, « Big data Le pétrole du XXIème siècle », <http://www.lenouveleconomiste.fr/big-data-13807>, publié le 22 février 2012.
11. « Big Data, le défi de l'information pléthorique », <http://www.atelier.net/trends/files/big-data-defi-de-linformation-plethorique>, publié le 27 février 2012.
12. « Dans l'ombre du Big Data, les acteurs du MDM continuent à innover », http://www.lemagit.fr/article/mdm/10604/1/dans-ombre-big-data-les-acteurs-mdm-continuent-innover/?utm_source=essentielIT&utm_medium=email&utm_content=new&utm_campaign=20120307&xtor=ES-6, " publié le 06 mars 2012.
13. « Big Data : si on se posait les bonnes questions ? », http://www.bpmbulletin.com/2012/03/09/big-data-si-on-se-posait-les-bonnes-questions/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+bpmbulletin+%28BPM+Bulletin%29, publié le 9 mars 2012.:
14. Valéry Marchive, « La sécurité se met résolument au «Big Data» », http://www.lemagit.fr/article/ids-ips-siem/10640/1/la-securite-met-resolument-big-data/?utm_source=essentielIT&utm_medium=email&utm_content=new&utm_campaign=20120312&xtor=ES-6, publié le 12 mars 2012.
15. « Big Data : un marché en pleine expansion », <http://www.silicon.fr/big-data-un-marche-en-pleine-expansion-72630.html>, publié le 9 mars 2012.

16. « Le décollage du Big Data », <http://www.infodsi.com/articles/130141/decollage-big-data.html?key=>, publié le 13 mars 2012.
17. « L'impact du Big data sur la voix du client Par Arnaud Guénégan, directeur commercial, Nice Systems France », <http://www.infodsi.com/articles/130101/impact-big-data-voix-client-arnaud-guenegan-directeur-commercial-nice-systems-france.html?key=>, publié le 13 mars 2012.
18. « La révolution Big data doit être préparée pour générer de la valeur », <http://connect.emc2.fr/la-revolution-big-data-doit-etre-preparee-pour-generer-de-la-valeur/>, publié le 14 mars 2012.
19. « Virtualisation, Big Data, BYOD, Brocade explore les nouvelles tendances réseaux », <http://www.itespresso.fr/virtualisation-big-data-byod-brocade-explore-les-nouvelles-tendances-reseaux-51808.html>, publié le 19 mars 2012.
20. « Les entreprises doivent évoluer vers un modèle orienté données », <http://www.atelier.net/trends/articles/entreprises-doivent-evoluer-vers-un-modele-orientee-donnees>, publié le 19 mars 2012.
21. « L'analyse prédictive doit être le fruit d'une combinaison d'art et de science », <http://www.atelier.net/trends/articles/lanalyse-predictive-etre-fruit-dune-combinaison-dart-de-science>, publié le 20 mars 2012.
22. « Big Data : hors Hadoop, point de salut Par Romain Chaumais, co-fondateur d'Ysance », <http://www.infodsi.com/articles/130449/big-data-hors-hadoop-point-salut-romain-chaumais-co-fondateur-ysance.html?key=>, publié le 21 mars 2012.
23. « Big Data Paris : le Big Data, un accélérateur pour la qualité des données ? », http://www.lemagit.fr/article/decisionnel-analyse-big-data/10714/1/big-data-paris-big-data-accelereur-pour-qualite-des-donnees/?utm_source=essentielIT&utm_medium=email&utm_content=new&utm_campaign=20120321&xor=ES-6, publié le 21 mars 2012.
24. « Big Data : des technologies et des usages », <http://www.silicon.fr/big-data-des-technologies-et-des-usages-72999.html> publié le 21 mars 2012.
25. « Le coût des pertes de données informatiques a nettement augmenté pour les entreprises françaises en 2011 », <http://www.itchannel.info/articles/130500/cout-pertes-donnees-informatiques-nettement-augmente-entreprises-francaises-2011.html?key=862d53eea2c1d2fe>, publié le 22 mars 2012.
26. « Smarter Analytics : « IBM s'enrichit de 3 nouvelles solutions d'analyse prédictive », <http://www.itespresso.fr/smarter-analytics-ibm-s%E2%80%99enrichit-de-3-nouvelles-solutions-d%E2%80%99analyse-predictive-51918.html>, publié le 22 mars 2012.
27. « Printemps de l'USF : le Big Data expliqué en détail », http://www.cio-online.com/actualites/lire-printemps-de-l-usf-le-big-data-explique-en-detail-4224.html?utm_source=mail&utm_medium=email&utm_campaign=Newsletter publié le 22 mars 2012.
28. « Enquête Echelon One/ LogLogic Une entreprise sur quatre comprend ce qu'est le Big Data », <http://www.infodsi.com/articles/130810/enquete-echelon-one-loglogic-entreprise-quatere-comprend-est-big-data.html?key=>, publié le 29 mars 2012.
29. « Du Big Data au service de la recherche scientifique », <http://www.zdnet.fr/actualites/du-big-data-au-service-de-la-recherche-scientifique-39770206.htm#xor=EPR-100>, publié le 30 mars 2012.
30. Yvonne Genovese, Gartner, High-Tech Provider Opportunities in Big Data, 'april_10_high_tech_provider_big_data_ygenovese.pdf', publié en avril 2012.
31. Oracle, « Big Data and Extreme Analytics Summit », <http://www.oracle.com/us/dm/h2fy11/67311-emeafm11033460mpp066-oem-1590985.html>, en avril 2012.
32. LCIA, Les sources IT, « Big Data: Big Opportunities to Create Business Value, '2_Etudebigdata_LCIA-BigData-Opportunities-Value.pdf », publié en avril 2012.
33. Fujitsu, Linked-data-connecting-and-exploiting-big-data-(v1.0).pdf", publié en avril 2012.
34. Talend, « wp_en_big_data.pdf », publié en avril 2012.
35. Laurence Dubrovin, CXP, « Big Data - L'offre est là, les besoins doivent émerger », http://www.cxp.fr/flash-cxp/big-data-offre-est-besoins-doivent-emerger_1217?_S=rss, publié le 3 avril 2012.

36. Nicolas Eimery, Directeur du Pôle de compétences Performance Management & Business Intelligence, GFI Informatique, « Big Data : de la Business Intelligence à la Business Analytics », <http://www.infodsi.com/articles/131008/big-data-business-intelligence-business-analytics-nicolas-eimery-directeur-pole-competences-performance-management-business-intelligence-gfi-informatique.html?key=>, publié le 04 avril 2012.
37. « Le fret maritime aussi se met au Big Data », <http://www.atelier.net/trends/articles/fret-maritime-se-met-big-data>, publié le 16 avril 2012.
38. « Data transparency coalition forms with backing of Microsoft », <http://thehill.com/blogs/hillicon-valley/technology/221771-data-transparency-coalition-forms-with-backing-of-microsoft>, publié le 16 avril 2012.
39. « Open Data : le portail data.gouv.fr passe en V2 avec des fonctions collaboratives », <http://www.zdnet.fr/actualites/open-data-le-portail-datagouvfr-passe-en-v2-avec-des-fonctions-collaboratives-39770815.htm#xtor=EPR-100>, publié le 17 avril 2012.
40. « Data.gouv.fr : plongée dans la plate-forme technique du portail Open Data », <http://www.zdnet.fr/actualites/datagouvfr-plongee-dans-la-plate-forme-technique-du-portail-open-data-39770820.htm#xtor=EPR-100>, publié le 17 avril 2012.
41. « L'Open Data, véritable opportunité pour les collectivités », <http://www.atelier.net/trends/articles/lopen-data-veritable-opportunite-collectivites>, publié le 17 avril 2012.
42. « 7 top tools for taming big data », http://www.itworld.com/software/268940/7-top-tools-taming-big-data?source=ITWNLE_ITWNLE_nlt_bi_2012-04-25, publié le 18 avril 2012.
43. « IBM Acquires Enterprise Search Software Company Vivisimo To Boost Big Data Analytics », <http://techcrunch.com/2012/04/25/ibm-acquires-enterprise-search-software-company-vivisimo-to-boost-big-data-analytics/>, publié le 25 avril 2012.
44. « IBM s'offre le spécialiste du Big Data Vivisimo », <http://www.linformaticien.com/actualites/id/24641/ibm-s-offre-le-specialiste-du-big-data-vivisimo.aspx>, publié le 27 avril 2012.
45. « Mobile, réseaux sociaux et géolocalisation, terreau favorable à l'Open Data », <http://www.atelier.net/trends/articles/mobile-reseaux-sociaux-geolocalisation-terreau-favorable-open-data>, publié le 30 avril 2012.
46. « Une étude du cabinet Lecko L'Open Data au cœur de la nouvelle relation entre collectivités, entreprises et citoyens », <http://www.infodsi.com/articles/132007/etude-cabinet-lecko-open-data-ur-nouvelle-relation-entre-collectivites-entreprises-citoyens.html?key=>, publié le 30 avril 2012.
47. « Le Big Data doit définitivement faciliter la fidélisation client », <http://www.atelier.net/trends/articles/big-data-definitivement-faciliter-fidelisation-client>, publié le 02 mai 2012.
48. « Is There Big Money in Big Data? », <http://www.technologyreview.com/business/40320/?nlid=nlbus&nld=2012-05-04>, publié le 03 mai 2012.
49. « Vous avez dit Big Data ? », <http://www.alain-bensoussan.com/avocats/vous-avez-dit-big-data/2012/05/03>, publié le 03 mai 2012.
50. « Google, Big Data & What it Means for SEO », http://searchenginewatch.com/article/2172922/Google-Big-Data-What-it-Means-for-SEO?wt.mc_ev=click&WT.tsrc=Email&utm_term=&utm_content=Google%2C%20Big%20Data%20%26%20What%20it%20Means%20for%20SEO&utm_campaign=05%2F08%2F12%20-%20SEW%20Daily&utm_source=Search%20Engine%20Watch%20Daily&utm_medium=Email, publié le 08 mai 2012.

51. « How to Avoid Big Data Spending Pitfalls », http://www.cio.com/article/705922/How_to_Avoid_Big_Data_Spending_Pitfalls?source=CIONLE_nlt_enterprise_2012-05-09, publié le 08 mai 2012.
52. « Five Things CIOs Should Know about Big Data », http://www.cio.com/article/706290/Five_Things_CIOs_Should_Know_About_Big_Data_?source=CIONLE_nlt_datacenter_2012-05-16, publié le 11 mai 2012.
53. « Taking the measure of big data: More is more », http://www.itworld.com/big-datahadoop/277196/taking-measure-big-data-more-more?source=ITWNLE_ITWNLE_nlt_bi_2012-05-16, publié le 15 mai 2012.
54. « Equifax Eyes Are Watching You--Big Data Means Big Brother », http://www.cio.com/article/706457/Equifax_Eyes_Are_Watching_You_Big_Data_Means_Big_Brother?source=CIONLE_nlt_datacenter_2012-05-23, publié le 15 mai 2012.
55. « La BI dopée par les sentiments » http://www.lemagit.fr/article/ibm-decisionnel-bi/11128/1/la-dopee-par-les-sentiments/?utm_source=essentielIT&utm_medium=email&utm_content=new&utm_campaign=20120529&xtor=ES-6, publié le 29 mai 2012.

56. « Faster, higher, stronger: Big data will get Olympic workout », http://www.itworld.com/big-datahadoop/278832/faster-higher-stronger-big-data-will-get-olympic-workout?source=ITWNLE_ITWNLE_nlt_bi_2012-05-30, publié le 29 mai 2012.
57. « 10 Commandments of Digital Analytics », http://searchenginewatch.com/article/2180772/10-Commandments-of-Digital-Analytics?wt.mc_ev=click&WT.tsrc=Email&utm_term=&utm_content=10%20Commandments%20of%20Digital%20Analytics&utm_campaign=05%2F30%2F12%20%20SEW%20Daily&utm_source=Search%20Engine%20Watch%20Daily&utm_medium=Email, publié le 30/05/2012.
58. « Gfi Informatique planche à Nantes sur les Open Data », <http://www.decideur-public.info/article-gfi-informatique-planche-a-nantes-sur-les-open-data-106134682.html>, publié le 31 mai 2012.
59. « Les 3 V du Big Data : Volume, Vitesse et Variété » Par Stefan Schmidt, Directeur de la stratégie produits d'hybris », <http://www.infodsi.com/articles/132989/3-big-data-volume-vitesse-variete-stefan-schmidt-directeur-strategie-produits-hybris.html?key=>, publié le 04/06/2012.
60. IBM, livre blanc « Understanding Big Data », https://www14.software.ibm.com/webapp/iwm/web/signup.do?source=sw-infomgt&S_PKG=500016891&S_CMP=is_bdebook1_bdmicro, en libre téléchargement

61. Gartner, « Gartner, the top 10 strategic technology trends for 2012. »
62. Brian Hopkins, Forrester, « Big Data Will Help Shape Your Market's Next Big Winners. »
63. IDC's 2011, « Extracting Value from Chaos », <http://idcdocserv.com/1142>, publié en juin 2011
64. McKinsey Global Institute, « Big data: The next frontier for innovation, competition, and productivity », http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation, publié en mai 2011
65. IDC, « Worldwide Big Data Technology and Services 2012-2015 Forecast », www.idc.com/getdoc.jsp?containerId=233485

66. Veille IT sur <http://the-it-watcher.com>.