

BIG DATA EN ENTREPRISE : QUELLES STRATEGIES D'IMPLANTATION ?

De

Mohammed ZOUHRI

**Professeur à la FSJES de Meknès, Vice Doyen de la FSJES de Meknès,
Département Economie, Université Moulay Ismail –Meknès.**

Résumé

Ces dernières années, nous avons assisté à une croissance exponentielle dans le volume et la disponibilité des données. Ce fait résulte principalement de l'émergence d'une multitude de sources (ordinateurs, appareils mobiles, capteurs ou réseaux sociaux) qui produisent de manière continue des données structurées, semi-structurées ou non structurées. C'est l'ère du Big Data.

Pour les entreprises, le Big Data constitue un nouveau défi puisqu'il implique de nouvelles exigences en termes de stockage de données, de traitement et de visualisation. L'analyse de ces données massives est d'une grande opportunité puisqu'elle permet de découvrir au sein de ces données des modèles et des corrélations originales.

Cet article vise à apporter quelques éléments de réponse quant à la stratégie d'implantation du Big Data dans l'entreprise. Pour traiter cette question, nous rappelons l'évolution du processus d'informatisation des entreprises depuis les fameux mainframes jusqu'au Cloud Computing d'aujourd'hui. Ensuite, nous levons le voile sur le nouveau paradigme émergent du Big Data en termes de définitions, enjeux et défis. Enfin nous montrons que plusieurs stratégies d'adoption du Big Data sont possibles suivant l'existant en solutions analytiques et les besoins de l'entreprise.

Mots clés

Système d'information, Business Intelligence, Big Data, Stratégie.

Abstract

In recent years we have witnessed an exponential growth in the volume and data availability. This is mainly due to the emergence of a multitude of sources (computers, mobile devices, social networks or sensors) that produce continuously structured data, semi-structured or unstructured. This is the era of Big Data.

For businesses, Big Data is a new challenge as it involves new requirements in terms of data storage, processing and display. The analysis of massive data is a great opportunity because it allows to discover in these data models and original correlations.

This article aims to provide some answers about the Big Data implementation strategy in the company. To address this issue, we recall the evolution of corporate computerization process from the famous mainframe to cloud computing today. Then we lift the veil on the new emerging paradigm of Big Data in terms of definitions, issues and challenges. Finally we show that several

big data adoption of strategies are possible depending on the existing analytical solutions and business needs.

Keywords

Information System, Business Intelligence, Big Data Strategy.

Introduction

Aujourd'hui, les entreprises opèrent dans un environnement turbulent, changeant dynamiquement et qui a un impact important sur leurs processus de prise de décision. Cette prise de décision est accompagnée d'une masse volumineuse de données et de logiciels pour les analyser. Les entreprises ne peuvent plus se contenter d'utiliser des données structurées collectées et stockées par leurs systèmes transactionnels, mais aussi des données semi-structurées ou non structurées se trouvant à l'extérieur de leur carcan.

Le paradigme du Big Data connaît un fort engouement aussi bien dans le milieu académique que dans l'industrie des technologies de l'information et de la communication. Dans le monde numérique, l'information est produite et collectée à une vitesse fulgurante. Actuellement, plus de 3 milliards de personnes dans le monde sont connectés à l'Internet, et plus de 5 milliards d'individus possèdent un téléphone portable. D'après une étude de l'IDC, les données numériques créées dans le monde seraient passées de 1,2 zettaoctets (1 Zo = 10^{21} octets) en 2010 à 2,8 Zo en 2012 et s'élèveront à 40 Zo en 2020. Le nombre d'objets connectés est passé de 600 millions en 2003 à 12 milliards en 2010 et est estimé à 50 milliards en 2020.¹

Selon le McKinsey Global Institute, le 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. Les ensembles de données sont de plus en plus volumineux en raison des recherches sur Internet, des réseaux sociaux et de la montée en puissance des systèmes informatiques des entreprises. Le fait de travailler avec des ensembles de données plus volumineux présente de nombreux avantages : les entreprises peuvent notamment détecter des tendances et des problèmes évitables.

Pour rester compétitif, les décideurs ne peuvent plus passer à côté de ce « nouveau pétrole » qu'est le Big Data. Le flair et l'expérience ne suffisent plus pour prendre de bonnes décisions. Les entreprises disposent, avec le Big Data, d'une mine d'informations prédictives permettant d'anticiper et de faire mieux que la concurrence.

Une étude académique portant sur le lien entre la prise de décision fondée sur les données et les performances de l'entreprise a montré que, dans un contexte équivalent, les entreprises qui basent leurs décisions sur l'analyse de données ont une rentabilité de 5% à 6% supérieure à celles qui s'appuient essentiellement sur l'intuition et l'expérience.² Une autre étude de Xerox auprès de 330 cadres d'entreprises européennes concluait que trois entreprises sur cinq (61%) estiment qu'en 2015, les prises de décision s'appuieront vraisemblablement davantage sur les données que sur d'autres critères tels que l'instinct, l'opinion ou l'expérience.³

Le Big Data irrigue non seulement tous les domaines fonctionnels de l'entreprise (marketing et vente, production et logistique, risques et fraudes, ressources humaines) mais tous les secteurs d'activités (télécommunications, énergie, banque, assurance, commerce, santé).⁴

¹ERICSSON, More than 50 Billion Connected Devices, White Paper, 2011

² The deciding Factor : Big Data and decision making, Economist Intelligence Unit, Capgemini, 2012

³ Enquête du cabinet Forrester Consulting pour le compte de Xerox, janvier 2015

⁴ J.-C. Cointot, Y. Eychenne, La révolution Big Data, les données au cœur de la transformation de l'entreprise, Dunod, 2014.

L'objectif de cet article est de répondre à la question suivante : quelles approches faut-il adopter pour utiliser les technologies Big Data dans l'entreprise ?

Pour ce faire nous rappelons dans un premier temps l'évolution des systèmes d'information des entreprises depuis l'informatisation de la comptabilité et la gestion de la paie du personnel aux architectures d'aujourd'hui. Ensuite, une mise en lumière du phénomène Big Data est indispensable pour comprendre les enjeux et les défis de cette technologie innovante. Enfin et suite à une revue de littérature spécialisée et professionnelle, nous proposons trois stratégies d'adoption des systèmes Big Data en entreprise : une approche disruptive, une méthode évolutive et une stratégie hybride.

1- L'informatique en entreprise

Avant de se pencher sur ce qui, au-delà de l'effet de curiosité, fait l'originalité du phénomène Big Data, nous allons revenir sur l'informatisation des firmes et la place qu'occupe la technologie aujourd'hui dans les entreprises.

Historiquement, les systèmes d'information des entreprises se sont développés les uns après les autres dans trois grands domaines (l'opérationnel, le décisionnel et le communicationnel) et continuent d'évoluer de manière significative.

Les premières informatisations ont porté sur l'automatisation des tâches opérationnelles. L'objectif était alors de supprimer ou d'alléger des activités administratives répétitives, comme la comptabilité ou le suivi des commandes. Ensuite, on s'est focalisé sur la manière de réorganiser au mieux les processus à partir de la refonte des systèmes d'information internes des entreprises en tenant compte des opportunités offertes par les technologies. La plupart des firmes ont ainsi opté pour les progiciels de gestion intégrés ou ERP (Enterprise Planning Systems) pour harmoniser, optimiser et fluidifier leurs processus.

Le deuxième champ de développement des systèmes d'information est celui d'aide à la décision. Des premiers infocentres des années quatre-vingt aux systèmes décisionnels ou de business intelligence d'aujourd'hui, l'informatique décisionnelle a fait un saut technologique important et des progrès considérables, en particulier dans les domaines du marketing (aide à l'analyse et la segmentation client) et du contrôle de gestion (mise en place de tableaux de bord permettant de suivre la performance des activités).

Enfin le troisième domaine est celui de la communication. Avec l'avènement de l'Internet, les systèmes d'information de communication ont connu une évolution en profondeur : réseaux d'échange avec les clients et les fournisseurs, développement des services sur Internet et exploitation des réseaux sociaux pour plus de business.

Aujourd'hui, de nouveaux défis se posent aux systèmes d'information des entreprises. Il ne s'agit non seulement de stocker et de traiter des données métiers structurées mais surtout de collecter des données semi-structurées ou non structurées en volume grandissant et en flux continu. Or, l'infrastructure technologique classique des systèmes d'information a montré ses limites face à l'augmentation spectaculaire du volume des données qui sont de plus en plus collectées par des appareils mobiles qui les détectent, des caméras, des microphones, des lecteurs d'identification par radiofréquence et des réseaux de capteurs sans fil. Il s'agit du phénomène Big Data que nous allons tenter d'élucider dans la section suivante.

2- Big Data : enjeux et défis

Éléments de définition

Le Big Data est aujourd'hui un terme à la mode ou « buzzword » que nous entendons partout. Ce phénomène émergent ces dernières années ne cesse de capturer l'intérêt de nombreux chercheurs dans de nombreuses disciplines (informatique, management, économie, droit, médecine, etc.), mais aussi des professionnels de toute activité et des décideurs politiques. Malgré cet engouement soutenu pour ce phénomène, une définition claire des Big Data n'est pas pour autant tranchée.

Si on se réfère au Journal officiel de la République française du 22 août 2014, le terme Big Data est traduit en « mégadonnées » que la commission générale de terminologie et de néologie définit comme « des données structurées ou non, dont le très grand volume requiert des outils d'analyse adaptés ».⁵

Certains auteurs réfèrent le concept Big Data à de très grands ensembles de données dont la collecte, le stockage et l'analyse est impossible avec les outils et méthodes de gestion de bases de données traditionnels (SGBDR, SQL⁶). Les données sont de l'ordre au minimum du Téraoctet, du Pétaoctet... et surtout ce volume ne cesse de croître à grande vitesse.

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 fait, ce slogan « 3V » est issu d'un rapport du Meta Group datant de 2001, qui ne mentionnait pas le vocable « Big Data » mais constatait simplement une croissance des formats et du volume des données et recommandait de mieux en formaliser la gestion.⁷ Cette définition populaire s'est enrichie de deux autres « V », une notion qualitative véhiculée par le 4^{ème} V, celui de la Véracité des données, et une vision davantage économique portée par le 5^{ème} V, celui de la Valeur (Figure 1). Ces cinq dimensions du Big Data méritent précision.

La Volumétrie : le monde est submergé de volumes de données croissants de tous types. On projette que le volume de données numériques créées dans le monde sera multiplié par 10 entre 2013 (4 Zo) et 2020 (40 Zo). La génération de cette quantité de données est directement induite par la densité d'appareils en interaction sur internet (Internet of Things / Everything). Cependant, l'aspect volume pur ne suffit pas pour caractériser une approche Big Data.

La Vélocité : cette dimension réfère aux délais d'actualisation et d'analyse des données numériques. Les données sont traitées en temps réel ou quasi réel. Dans certains cas, les données même massives doivent être traitées en temps réel pour être utiles (la détection de fraudes).

La Variété : les données massives se présentent sous une forme structurée ou non (textes, chiffres, sons, images, vidéos, tweets, blogs, mails, sms, etc.). D'autres nouvelles données et à forte valeur ajoutée sont générées en temps réel par des capteurs et des senseurs vont au fur et à mesure enrichir le «conglomérat» Big Data.

⁵Journal officiel de la république française, 22 août 2014.

⁶ Système de Gestion de Bases de Données Relationnelles, Structured Query Language

⁷<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> [consulté le 6 avril 2016]

La Véracité : les données recueillies sont souvent bruitées et imprécises. Leur exploitation ne peut être envisagée que si on en a confiance. Elles doivent donc être traitées pour en extraire l'information utile.

La Valeur : Les acteurs du secteur Big Data mentionnent la valeur qu'il est possible de tirer de cette avalanche de données et les usages qu'elles produisent. Dans un contexte d'info-obésité, les organisations doivent être capables de se concentrer sur les données ayant une réelle valeur et étant actionnables.

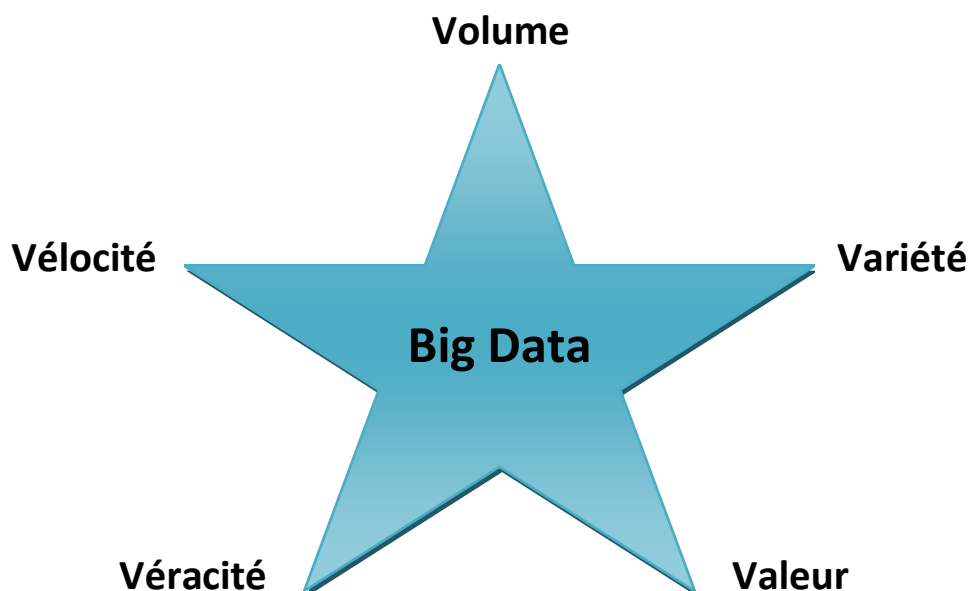


Figure 1 : les 5 V du Big Data

Cette série de V semble ne pas s'arrêter au chiffre 5 puisque dernièrement, un expert Big Data en a proposé une dizaine : en plus des 5 V précédents, il cite la Validité, la Variabilité, la «Venue», le Vocabulaire et l'imprécision (Vagueness en anglais).⁸

D'autres auteurs ont procédé à la classification des Big Data en différentes catégories afin de mieux comprendre leurs caractéristiques. La classification est faite selon cinq critères : les sources de données (web, capteurs, réseaux sociaux, objets connectés, etc.), le format de contenu (données structurées, semi-structurées ou non structurées), le stockage des données (bases orientées colonnes, documents, graphes, clé/valeur), la mise en scène des données (identification des données incomplètes ou incohérentes, transformation des données sous une forme appropriée à l'analyse et minimisation de la redondance en normalisant la base de données) et le traitement des données (par lot, en temps réel) (Figure 2).

⁸<https://www.mapr.com/blog/top-10-big-data-challenges-%E2%80%93-serious-look-10-big-data-v%E2%80%99s>
[consulté le 25/5/2016]

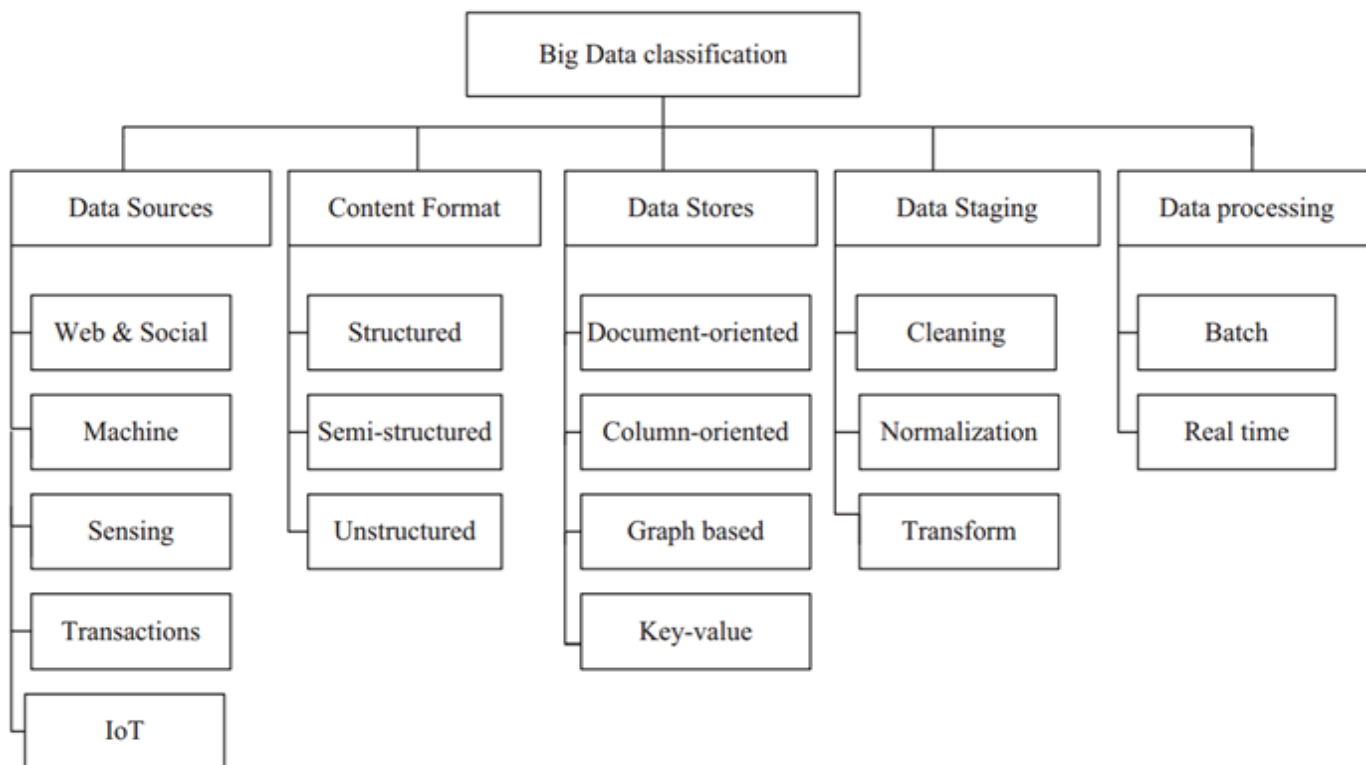


Figure 2 : classification des Big Data⁹

Facteurs d’émérgence

L’originalité de l’approche Big Data réside en ce qu’elle ne s’appuie pas sur les schémas préexistants à la collecte de données comme ça s’est fait jusqu’au là dans les systèmes aussi bien opérationnels que décisionnels, mais entend découvrir au sein de ces données des modèles. En effet, partir d’un modèle, c’est adopter un raisonnement déductif, par contre partir des données, comme en Big Data, c’est adopter plutôt un raisonnement inductif, avec une incertitude à apprécier. Il y en a même qui ont poussé le raisonnement un peu trop en véhiculant l’idée qu’on pourrait désormais se passer de toute théorie, que l’analyse statistique des données suffirait à elle seule comme source de connaissance.¹⁰ Avec les Big Data on est en présence d’un changement de paradigme : l’intérêt glisse du pourquoi vers le quoi, autrement dit de la causalité vers la corrélation. Mais comment on est-on arrivé là ? Plusieurs facteurs sont à l’origine de l’essor fulgurant du Biga Data. Tout d’abord les facteurs technologiques qui constituent la base innovante de l’univers des Big Data.

Dans le monde de l’informatique, trois lois résument ces facteurs. La loi de Moore qui prévoit que la capacité de calcul des ordinateurs double tous les dix-huit mois, celle de Kryder qui prévoit le doublement de densité des disques durs tous les treize mois et celle de Nielson qui estime un doublement des capacités des réseaux publics tous les vingt-et-un mois.¹¹ Ces lois technologiques mettent en lumière notre capacité croissante à pouvoir capter, stocker, transformer et transmettre des données massives, condition nécessaire au Big Data.

⁹ I.A.T Hashem & al, The rise of “Big Data” on cloud computing : Review and open research issues, Information Systems, 47, 2015, pp. 98-115

¹⁰ Chris Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, Wired, juin 2008

¹¹Pierre DELORT, Le Big Data, Que Sais-je?, 2015

Un autre facteur clé qui se trouve en filigrane du développement des Big Data est le Cloud Computing ou « l'informatique en nuage ». Pour pouvoir exploiter les mines d'informations et les flux de données du Big Data, d'importantes capacités de calcul sont nécessaires et qui sont souvent uniquement disponibles dans de grands centres de données (data centers). Le Cloud Computing permet aux chercheurs, industriels et entreprises de « louer » une puissance de calcul et un espace de stockage adaptés pour un traitement Big Data. En effet, seuls les grands acteurs de l'Internet (les GAFA¹²) sont en mesure d'effectuer ce traitement avec leurs propres infrastructures, au vu des équipements informatiques nécessaires. Le Cloud Computing va donc démocratiser le Big Data et le mettre à la portée des petites et moyennes entreprises (PME) et des acteurs non experts du traitement des données.

Les autres facteurs sont en relation avec l'évolution des modes d'interaction développés ces derniers temps, en particulier les réseaux sociaux, la recherche frénétique de compétitivité dans les affaires et d'avantage concurrentiel et la multiplication des nouveaux formats de données, en lien notamment avec l'Open Data.¹³

Opportunités décuplées

Le Big Data est non seulement une nouvelle technologie mais plutôt un phénomène qui ne cesse de prendre de l'ampleur en transformant le quotidien des individus et en dotant les entreprises d'applications innovantes.

Dans le domaine du marketing par exemple, le Big Data permet d'évoluer d'un service réactif à un service proactif et ainsi d'avoir une connaissance « à 360 degrés » du client. Cette vision peut être obtenue à partir du système d'information interne (ERP, CRM) et des données extérieures, notamment de l'activité du client sur les réseaux sociaux. Le Big Data a permis ainsi à l'opérateur américain de télécommunications T-mobile de réduire de 50% le taux de résiliation d'abonnement (churn) en un trimestre. Mais le marketing n'est pas l'unique domaine de prédilection des Big Data.

La gestion des données massives est à l'origine d'autres opportunités comme l'optimisation des processus et l'analyse prédictive des anomalies (réduction des déraillements des trains et des émissions de CO₂ dans les zones urbaines), la sécurité publique en recourant au suivi des réseaux sociaux et à l'intégration des données spatiales et celles des capteurs (l'analyse des tweets permet de repérer et de prévenir les mouvements des foules aux Pays-Bas par exemple).¹⁴

Défis à relever

Malgré les grandes opportunités d'évolution offertes par le Big Data, de nombreux défis au niveau de chaque maillon (collecte, stockage, analyse et visualisation) de la chaîne de traitement des données massives restent à relever.

Pour ce qui est de la collecte des données, il faut pallier les difficultés liées à la variété, la vitesse, la provenance et la qualité des données. Au niveau du stockage des données, les systèmes de gestion de bases de données relationnelles s'avèrent insuffisants et le recours à de nouvelles technologies

¹² Google, Apple, Facebook, Amazon

¹³ L'«Open data» ou « données ouvertes » est une démarche de publication de données numériques en ligne selon des critères garantissant leur libre accès et réutilisation par tous.

¹⁴ www.ayming.fr/typo3conf/ext/almacg/user_upload/Innovation_taxes/Guide/Cahier_Big_Data_ALMACG2015.pdf [consulté le 25/5/2016]

est nécessaire (bases NoSQL, architecture distribuée). Quant à l'analyse des données, le problème est de pouvoir chercher la donnée sur une multitude de serveurs et de procéder à l'analyse en temps réel (technologies Mapreduce, Hadoop, loadbalancing, loadshedding). Enfin le défi de visualisation concerne essentiellement le reporting dynamique en temps réel et en particulier le choix des graphiques facilement interprétables à proposer au décideur.

3- Stratégies d'implantation du Big Data en entreprise

La collecte, le stockage, le traitement et l'analyse des volumes massifs des données (surtout les données non structurées comme la vidéo, le son, les messages, etc.) pose de grands défis aux entreprises et les poussent désormais à déployer des technologies Big Data.

L'introduction du Big Data dans les firmes est synonyme de changements en profondeur, d'évolution de ses modèles économiques et de sa réactivité vis-à-vis de ses clients et de son marché. Les projets Big Data sont en général transverses et impliquent toutes les fonctions de l'entreprise. De nouvelles compétences sont nécessaires pour accompagner de tels projets et le rôle des experts dans des métiers donnés sera lui-même transformé.

Les grandes données ou mégadonnées promettent de bouleverser les systèmes d'information actuels mis en œuvre dans les entreprises. Le stockage des données et leur traitement avant analyse, ainsi que les matériels et les logiciels employés, sont transformés par les technologies du Big Data. Certaines sont véritablement nouvelles, d'autres par contre existent depuis longtemps mais avec un usage différent.

La question qui se pose alors est : comment introduire la technologie Big Data dans l'entreprise ? Quels sont les modes de sa mise en œuvre, les compétences requises et les changements organisationnels à prévoir ? En somme, quelle stratégie faut-il adopter pour implanter le Big Data dans la firme ?

Comme nous l'avons rappelé dans la première section, la plupart des entreprises, en particulier les grandes, ont mis en place des systèmes d'information et les ont fait évoluer en fonction de leurs besoins et en tirant parti à chaque fois des nouvelles tendances IT ou informatiques. Ces dernières disposent ainsi d'un système transactionnel, basé sur des serveurs uniques et des systèmes de gestion de bases de données, qui gère le quotidien mais aussi d'un système décisionnel pour aider les décideurs à avoir la bonne information au bon moment. Ces derniers utilisent des technologies qui se sont développées depuis plus de vingt ans et qui englobent toute la chaîne décisionnelle. De la collecte des données de plusieurs sources (internes, externes), en passant par la préparation et le stockage de la donnée décisionnelle dans ce que l'on nomme entrepôt de données (datawarehouse) ou magasin de données (datamart) en allant vers la restitution de l'information d'aide à la prise décision via des outils variés comme le reporting, les cubes OLAP¹⁵ ou les techniques du datamining.

Aujourd'hui, les entreprises voudraient bien profiter des technologies Big Data mais elles ont déjà des environnements et des technologies de données à gérer. Plusieurs modèles de développement de

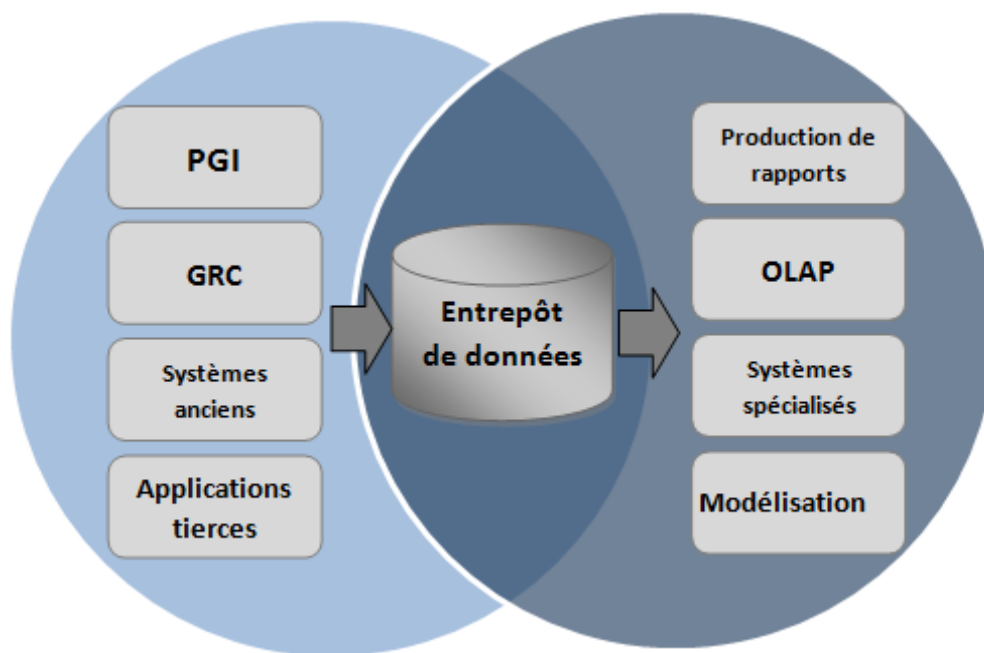
¹⁵ On-Line Analytical Processing : représentation abstraite d'informations multidimensionnelles prévue à des fins d'analyses interactives par les utilisateurs du métier que ces données sont censées représenter.

projets Big Data en entreprise sont envisageables. Cependant, le choix de l'un d'entre eux va dépendre d'une décision fondamentale de l'entreprise :

- elle peut capitaliser sur l'existant et donc privilégier son entrepôt de données et les investissements associés disponibles;
- elle peut par contre basculer très tôt sur une solution Big Data « native », ceci en vue de stocker des volumes massifs de données (en pétaoctets et non plus en téraoctets), de les traiter rapidement (en minutes et non pas en heures) ou pour bénéficier d'une solution avec une infrastructure économique par l'utilisation de matériels à sa portée.

Etat des lieux

Généralement, dans les grandes entreprises, l'environnement d'analyse classique comprend des systèmes opérationnels qui alimentent, à partir d'une variété de sources de données, un entrepôt de données ou des magasins de données fédérés qui stockent et intègrent les données pour diverses fonctions d'analyse, ainsi qu'un ensemble d'outils relevant de l'informatique décisionnelle (business intelligence) qui permettent de prendre des décisions à partir de requêtes, de tableaux de bord et d'exploration des données (Figure 3).



© SAS Best Practices

Figure 3 : un environnement décisionnel autour d'un datawarehouse¹⁶

Cet écosystème décisionnel a nécessité de grands investissements de la part des firmes dans des plateformes matérielles, des bases de données, des logiciels d'ETL¹⁷, des outils d'analyse élaborée, des contrats de maintenance et des systèmes de stockage, notamment des entrepôts de données robustes. Ces solutions, dans les meilleurs des cas, ont permis à ceux qui les détiennent, de mieux comprendre le comportement et les habitudes d'achat de leurs clients, de rationaliser leurs processus

¹⁶ T.H. Davenport, J. Dyché, Big Data in Big Companies, SAS Institute, 2013.

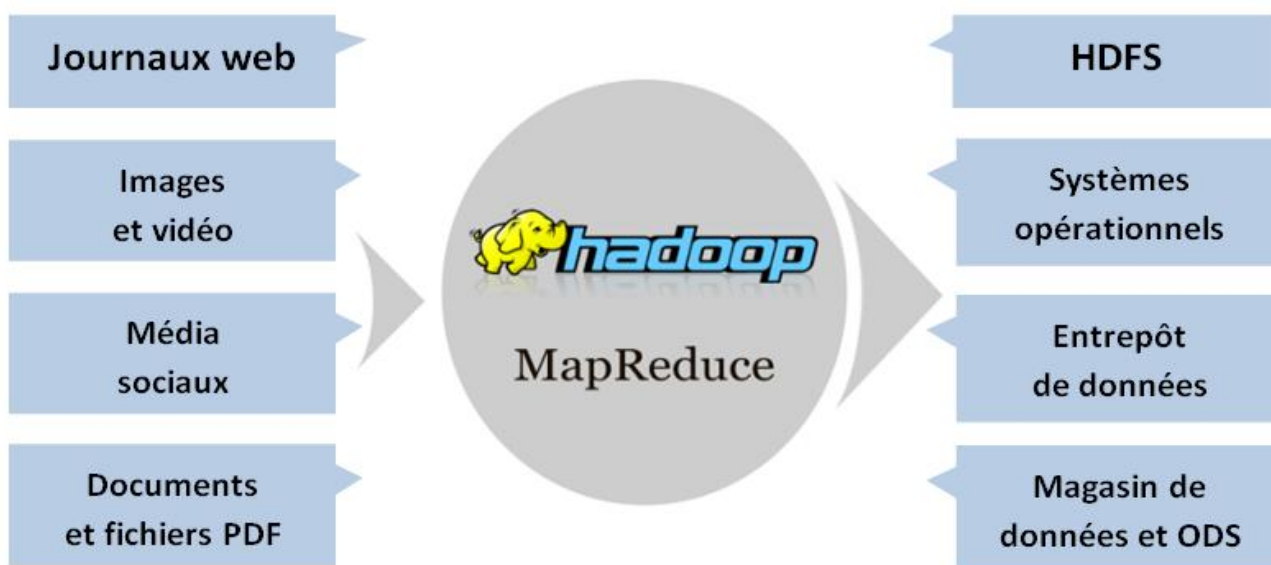
¹⁷ Extract-Transform-Load: technologie informatique permettant d'effectuer des synchronisations massives d'information d'une source de données (le plus souvent une base de données) vers une autre.

de vente et d'optimiser le prix et le conditionnement de leurs produits. Par contre et dans les cas les plus défavorables, d'autres entreprises se sont vus surinvestir dans des technologies décisionnelles tout en étant incapables de recouvrer leurs investissements dans l'analytique et par conséquent de voir leurs systèmes décisionnels comme des sources de coûts irrécupérables, avec un retour marginal en valeur commerciale.

Les firmes numériquement averties voient plutôt leurs systèmes d'information comme un ensemble de capacités métiers déployées au fil du temps, en exploitant une infrastructure commune et des données réutilisables. Le Big Data n'est qu'une nouvelle tendance technologique permettant de nouvelles opportunités que les systèmes existants sont incapables de fournir.

Stratégie disruptive

Cette approche consiste à mettre en place le Big Data au cœur du système décisionnel. Il s'agit naturellement du choix opté par une start-up ou d'un acteur des technologies numériques. Toute la chaîne d'acquisition des données, de stockage, d'analyse et d'extraction de rapports est construite autour des outils du Big Data : le système de stockage de données Hadoop, les outils d'analyse propres au Big Data comme Mapreduce et les techniques d'apprentissage (Machine Learning), et enfin des outils de visualisation (Figure 4).



© SAS Best Practices

Figure 4 : Un écosystème de technologies de Big Data¹⁸

Bien que cette approche pure Big Data n'utilise que les dernières technologies propres à la gestion des grandes données non structurées, elle peut aussi prendre en charge des données structurées et fournir ainsi des rapports d'analyse historique sur un domaine fonctionnel comme la relation client ou le suivi des commandes. Ceci dit, une telle approche n'est pas toujours à la portée des entreprises de taille intermédiaire.

A titre d'exemple de cette stratégie, la société Yahoo dispose d'un système Hadoop avec plus de 40 000 machines et plus de 350 pétaoctets de données pour leur moteur de recherche et le système de gestion de la publicité.¹⁹

¹⁸ T.H. Davenport, J. Dyché, Big Data in Big Companies, SAS Institute, 2013.

Méthode évolutive

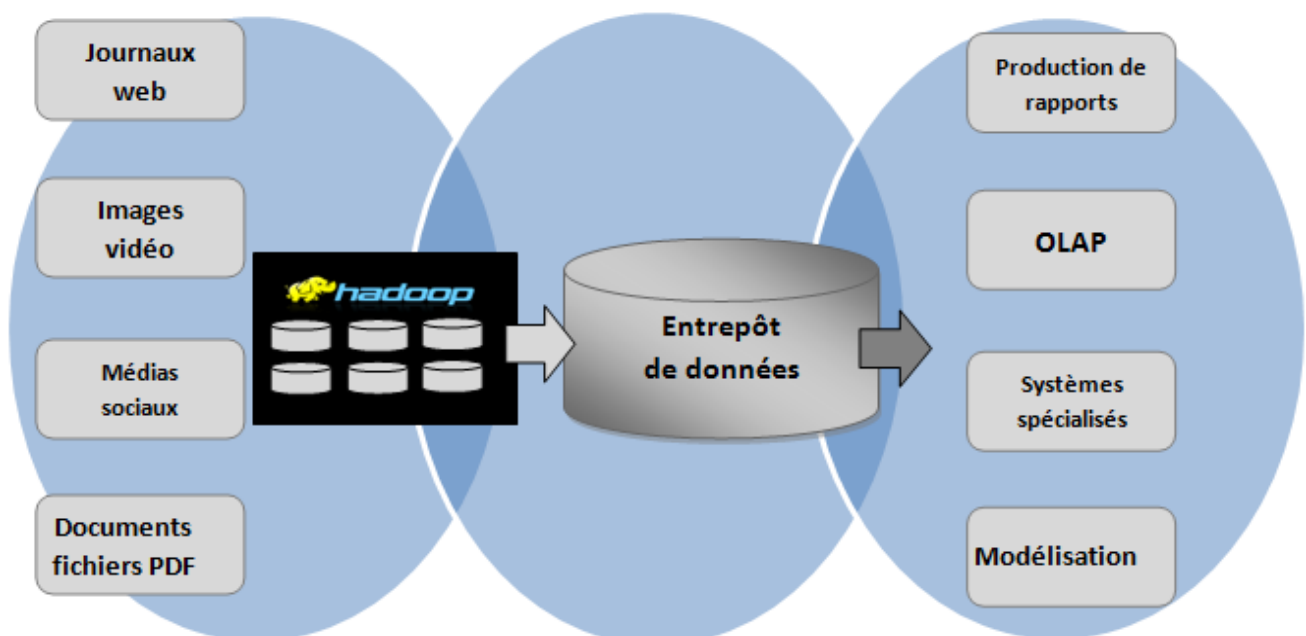
Cette approche consiste à préserver l'investissement en business intelligence déjà existant en l'enrichissant des apports des technologies Big Data dans certains domaines comme le traitement analytique des flux de données en temps réel ou l'analyse des données provenant des réseaux sociaux par exemple.

A titre d'illustration, cette méthode évolutive consiste à installer en amont du datawarehouse d'un opérateur en télécommunications, en prise direct sur les données du réseau pour analyser en temps réel des quantités massives de données et ainsi gérer au mieux les problèmes de performance de certains services ou encore des pannes de réseau.

L'explosion du volume des données produite par les appareils mobiles est à l'origine de la demande pour ce type d'approche ainsi que la contrainte que cela représente sur les réseaux mobiles ; pour collecter et traiter tous les événements instantanément, pour suivre leur performance des services, et pour comprendre l'impact sur les clients.

Approche hybride

Cette stratégie de coexistence consiste à associer les entrepôts de données et les environnements d'analyse anciens à la puissance des nouvelles solutions de Big Data, ce qui permet de tirer le meilleur parti des deux mondes technologiques (figure 5).



© SAS Best Practices

Figure 5 : la coexistence d'un entrepôt de données et du Big Data²⁰

Dans ce modèle, le datawarehouse fonctionne en coopération avec le système Big Data. L'entrepôt de données conserve sa fonction de stockage des données fournies par les systèmes opérationnels, essentiellement des données structurées, rassemblant les données commerciales et comptables par exemple. Il reste exploité pour l'informatique décisionnelle et les analyses standards, y compris les

¹⁹<http://developer.yahoo.com/blogs/hadoop> [consulté le 24/5/2016]

²⁰T.H. Davenport, J. Dyché, Big Data in Big Companies, SAS Institute, 2013.

rapports de ventes régionaux, les tableaux de bord des clients ou l'historique du risque de crédit par exemple. Les données opérationnelles peuvent aussi servir de source à l'environnement Big Data lorsqu'ils présentent un intérêt pour les calculs complexes et l'exploration des données brutes. C'est à la firme d'envoyer les données vers la plateforme appropriée.

Cette approche contraste avec le modèle évolutif précédent où le Big Data n'est qu'une source possible de données, sans intégration forte. Ce modèle de coexistence entre les deux technologies Big Data et datawarehouse est bien adapté pour des entreprises qui ont déjà une informatique décisionnelle, un entrepôt de données et des outils d'analyse et de reporting.

Cette stratégie d'implantation du Big Data a pour avantage d'éviter des changements dans le système traditionnel déjà installé et coûteux à réaliser, pour inclure des données non structurées. Par ailleurs, si on stockait ces données dans un système décisionnel classique, l'augmentation du volume des données pour être en phase avec le système Big Data induirait des coûts d'infrastructure très élevé.

En termes d'offre commerciale, la plupart des éditeurs de solutions décisionnelles proposent des systèmes hybrides grâce à des extensions aux systèmes de business intelligence existants. Les sociétés Teradata, SAS, SAP, Microsoft ou encore IBM proposent cette approche dans leurs solutions respectives.

Mais le choix de l'un de ces modèles d'implantation ne suffit pas. Encore faut-il l'accompagner par l'acquisition des compétences nécessaires en termes de gestion des données, d'extraction de la valeur et de management. Les entreprises doivent aussi gérer les changements organisationnels induits par la nouvelle culture de Big Data. Enfin, une gestion des différentes étapes du déploiement et la mise en place d'une plateforme de Big Data s'avèrent nécessaires pour mener à terme tout projet Big Data en entreprise.

Conclusion

Avec la numérisation grandissante et l'explosion informationnelle qui en résulte, les entreprises sont en pleine métamorphose par des innovations – disruptives – rendues possibles grâce à une démocratisation des usages de l'Internet et les technologies Big Data. Ces dernières, avec le Cloud Computing, la mobilité et les réseaux sociaux, constituent l'un des quatre piliers de la transformation numérique de l'économie moderne. On désigne par Big Data, des ensembles massifs de données hétérogènes et accumulées à des vitesses sans précédent sur les individus, et même sur les organisations, à travers plusieurs canaux comme le web, les réseaux sociaux et l'Internet des objets (Internet of Things). Le Big Data est considéré comme « le pétrole » de l'ère digitale actuelle et il présage un nouvel eldorado numérique pour des économies en quête de croissance.

Dans cet article, nous nous sommes intéressés à la question d'implantation des technologies Big Data en entreprise. Aussi, nous sommes revenus d'abord sur la manière dont l'informatique s'est implantée dans les entreprises depuis les premiers EDP (Electronic Data Processing) aux entrepôts de données d'aujourd'hui. Ensuite, nous avons présenté quelques caractéristiques du concept Big Data. Enfin et pour répondre à la problématique retenue, nous avons proposé quelques stratégies permettant aux entreprises de modifier l'architecture des technologies existantes et d'acquiescer des technologies du Big Data afin de s'adapter au monde actuel des grandes données.

Au-delà de l'immensité du volume des données qui caractérise le Big Data, seule la capacité à structurer, analyser et transformer ces données en connaissances, en innovations et en valeurs commerciales pour les entreprises sera décisive.

Bibliographie

Ibrahim ABAKAR TARGIO HASHEM et al., *The rise of "Big Data" on cloud computing: Review and open research issues*, Information Systems, 47, 2015, pp. 98-115.

Brad BROWN, Michael CHUI, James MANYIKA, « *Are you Ready for the Era of «Big Data»?* », McKinsey Quarterly, octobre 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.

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

CIGREF, *Big Data : la vision des grandes entreprises françaises*, Octobre 2013.

Jean-Charles COINTOT, Yves EYCHENNE, *La révolution Big Data*, Dunod, 2014.

Thomas H. DAVENPORT, Paul BARTH, Randy BEAN, *How 'Big Data' is Different?* MIT Sloan Management Review, Vol.54, N°1, Fall 2012.

Thomas H. DAVENPORT, *Stratégie Big Data*, Pearson, 2014.

Thomas H. DAVENPORT, J. DYCHÉ, *Big Data in Big Companies*, SAS Institute, 2013.

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

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.

Myriam KAROUI, Grégoire DEVAUCHELLE, Aurélie DUDEZERT, *Big Data : Mise en perspective et enjeux pour les entreprises*, Ingénierie des Systèmes d'Information, 2014.

Myriam KAROUI, Grégoire DEVAUCHELLE, Aurélie DUDEZERT, *Systèmes d'Information et prise de décision à l'ère du «Big Data»: Le cas d'une entreprise française du bâtiment*, 18^{ème} Conférence Internationale de l'Association Information et Management, Lyon, 2013.

Nawsher KHAN et al., *Big Data: Survey, Technologies, opportunities, and Challenges*. The Scientific World Journal, 2014.

Doug LANEY, *3D data Management: Controlling Data Volume, Velocity, and Variety*. Meta Group Inc, février 2001.

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.

Viktor MAYER-SCHÖNBERGER, Kenneth CUKIER, *Big Data : la révolution des données est en marche*. Robert Laffont, 2014.

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

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

Jafar RAZA ALAM, Asma SAJID, Ramzan TALIB, Muneeb NIAZ, *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.

Bruno TEBOUL, Taoufik AMRI, *Les Machines pour le Big Data : Vers une Informatique Quantique et Cognitive*, 2014. <hal-01096689>

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

Mohammed ZOUHRI, Saïd AMALI, *une mise en perspective du phénomène Big Data*, Revue Marocaine de Gestion et d'Économie, numéro 6, 2016.