

2

Big Data : ne vous trompez pas de révolution

Arnaud Laroche (1994) – Président de Bluestone
 Bertrand Fisch (1998) – *Practice Manager* chez Bluestone

La valorisation des données de l'entreprise ne saurait se réduire au simple enjeu technologique du Big Data. Pour l'éviter, elle doit être confiée à des Data Scientists capables d'appréhender la problématique dans toutes ses dimensions : métier, informatique, statistique et mathématique.



« Les données des entreprises constituent un actif stratégique pouvant devenir un vecteur puissant de différenciation et de performance. »

Le Big Data s'impose comme le concept incontournable du moment, dont chacun s'accorde à dire qu'il va révolutionner la manière de travailler de beaucoup d'entreprises dans de nombreux secteurs d'activité, et dans des domaines aussi variés que la conception produit, le marketing, la détection de la fraude, la maintenance industrielle...

Nous assistons dans le même temps à une mise en lumière de la « valorisation des données » encore jamais atteinte par le passé : plus un jour ne passe en effet sans que de nouveaux articles vantant les mérites d'une utilisation poussée de l'information ne soient publiés (y compris dans les médias généralistes de grande diffusion), ou que de nouvelles offres technologiques ne soient mises sur le marché par leurs éditeurs (que ceux-ci fassent partie du cercle restreint des géants de l'industrie informatique traditionnelle, ou du cercle plus large des *start-up* innovantes).

Certes, le monde des affaires a l'habitude de ces révolutions annoncées – et relayées en premier lieu par les éditeurs de logiciels. Ainsi, au cours des 15 dernières années, et dans le seul domaine du traitement analytique de l'information, a-t-on vu le *business intelligence*, le *datamining* ou encore le *Customer Relationship Management*

Figure 1 : Les applications possibles du Big Data

Ventes & Marketing	 Connaissance clients Segmentation clients – Scoring – Détection du churn – Analyse de la valeur clients	 Marketing produit Pricing – Yield Management – Optimisation du Mix Marketing – Cross-selling & Up-selling	 Marketing relationnel Ciblage de campagnes – Personnalisation d'offres – Mesure de la satisfaction	 Image de marque Analyse de sentiments – Identification de leaders d'opinion – Brand Equity
Risque industriel & financier	 Pilotage du risque financier Mathématiques financières – Systèmes experts de calcul du risque de crédit – Calculs du risque de sinistralité	 Détection de fraude Identification de comportements & fonctionnements anormaux – Remontée d'alertes	 Smart alarming Détection de signaux faibles de dysfonctionnement – Remontée d'alertes – Aide au diagnostic	 Maintenance prédictive Approche prévisionnelle de la défaillance – Optimisation des tournées de maintenance
Production & Exploitation	 Ressources humaines Analyse sémantique de curriculum vitae – Aide à la constitution d'équipes collaboratives	 Conduite de réseaux Mesures temps réel et prévisions d'utilisation des infrastructures – Aide au diagnostic et à la décision	 Supply chain management Approche prévisionnelle de la demande – Optimisation des stocks – Tracking des approvisionnements	 Ressource allocation Prévision de demande et d'activité – Optimisation sous contraintes de l'allocation des ressources

(CRM) véhiculer leurs lots de promesses... et de déceptions.

Et si certains décideurs, échaudés par ces expériences passées, regardent avec circonspection le déferlement médiatique du Big Data, la perspective de manquer une opportunité réputée incontournable risque fort de l'emporter – quitte à répéter les erreurs du passé.

Dès lors, comment récolter les fruits du Big Data en minimisant les risques ? Si aucun plan ne peut garantir à coup sûr le succès dans cette voie, certaines bonnes pratiques permettent d'éviter bien des difficultés :

- Pensez Data Science avant de penser Big Data,
- Embauchez une équipe de Data Scientists,
- Placez-les sous la responsabilité d'une autorité légitime.

Pensez Data Science avant de penser Big Data

Si le Big Data est bien, à strictement parler, une révolution technologique (cf. introduction du

présent dossier), il serait particulièrement dangereux de ne le considérer que sous l'angle des solutions techniques, et de laisser les informaticiens s'emparer du sujet pour piloter la mise en œuvre d'outils qu'ils auraient eux-mêmes choisis.

Car le Big Data n'est en réalité « que » la composante informatique d'une autre révolution, plus profonde mais moins visible : l'émergence de la Data Science dans les entreprises et, avec elle, la compréhension que les données des entreprises constituent un actif stratégique pouvant devenir un vecteur puissant de différenciation et de performance.

L'idée n'est certes pas totalement nouvelle mais, bénéficiant de l'attention médiatique portée au Big Data, c'est probablement la première fois qu'elle peut être placée de façon aussi centrale et à aussi haut niveau dans le débat.

Au-delà du choix d'un outil ou d'une technologie nouvelle, la réflexion sur la place de la Data Science dans l'entreprise lui permet de questionner ses pratiques, ses données et ses méthodes :

- Quels sont les processus sur lesquels une analyse fine de l'information permettrait d'être plus efficace (conception produit,

L'émergence du Data Scientist : une chance pour le GENES ?

Gartner estime à un peu plus de 4 millions le nombre total d'emplois liés au Big Data et à la Data Science à l'horizon 2015 au niveau mondial. Et McKinsey estime à environ 200.000 le nombre de postes de Data Scientists qui resteront non pourvus à l'horizon 2018 pour les seuls Etats-Unis...

Parce qu'elles forment des ingénieurs de très haut niveau dans les domaines de la statistique et plus largement du traitement de l'information, et parce que ces ingénieurs ont les compétences et les facultés qui leur permettent de s'ouvrir à de nombreux sujets et à de nombreuses problématiques métier, l'ENSAE ParisTech, l'ENSAI, le CEPE et les différents laboratoires de recherche du Groupe peuvent acquérir une position de premier plan sur ce marché d'avenir.

Pour y parvenir toutefois, la dimension informatique de ce métier ne devrait pas être négligée dans ces enseignements, au motif qu'elle serait moins noble que les disciplines qui y sont traditionnellement enseignées. L'exemple des Etats-Unis - et le cursus des formations en Data Science qui y sont dispensées - est à ce titre révélateur (cf. <http://datascience101.wordpress.com/2012/04/09/colleges-with-data-science-degrees> pour une liste de ces formations).

marketing produit, relation client, vente en ligne, production opérationnelle, chaîne logistique, maintenance industrielle, gestion du risque financier, pilotage de la marge...)?

- Quelles sont les données disponibles dans les systèmes d'information de l'entreprise, qui permettraient de répondre plus efficacement aux questions que se posent les pilotes de ces processus, et quelle est la nature de ces données (volume, qualité, variété...)? A l'inverse, quelles sont les données non stockées dans les systèmes d'information, et qui permettraient pourtant d'enrichir fortement l'analyse métier (données de logs web, données issues de capteurs industriels...)?
- Quelles sont les méthodes de traitement de l'information (analyse statistique, analyse numérique, intelligence artificielle, machine learning...) qui permettraient effectivement de transformer ces données en des réponses concrètes et opérationnelles aux questions posées par les pilotes de processus?

Ces analyses et questionnements pourront effectivement conduire certaines entreprises à recourir

aux nouvelles solutions technologiques offertes par le Big Data, et ce notamment si les données à traiter sont volumineuses et peu structurées (texte, voix, image...), et si les processus métier qu'elles doivent éclairer imposent des exigences de traitement rapide (voire en temps réel).

Mais il est également possible que des réponses adaptées puissent être trouvées dans les technologies plus traditionnelles du *business intelligence* et du *datamining*, sous réserve toutefois que l'entreprise sache apporter une attention beaucoup plus soutenue aux approches quantitatives dans la résolution des problématiques métier auxquelles elle fait régulièrement face.

Embauchez une équipe de Data Scientists

Conduire efficacement un projet Big Data est une tâche complexe, portant sur un périmètre vaste et touchant de très nombreux acteurs; elle ne peut être confiée qu'à une figure triplement compétente : le Data Scientist.

« Data Scientist » est un terme qui cristallise une série d'évolutions dans la pratique des « professionnels de l'exploitation de données » confrontés :

- d'une part, à la progression rapide des moyens informatiques – hardware et software – mis à leur disposition,
- d'autre part, au rôle grandissant que leur reconnaissent les départements opérationnels dans les prises de décision métier.

Dans ce contexte, le Data Scientist apparaît comme une synthèse de différentes compétences essentielles pour le projet Big Data, mais difficiles à réunir dans l'entreprise :

- Le Data Scientist comme **mathématicien** : parce qu'il choisit, adapte et applique des approches issues de domaines variés de la statistique et de l'intelligence artificielle pour extraire la valeur des données qu'il manipule, le Data Scientist doit être un mathématicien à même d'évaluer et de comparer différents modèles ou méthodes de calcul, d'en anticiper les avantages et les inconvénients, et de les exploiter en connaissance de cause dans un environnement très métier.
- Le Data Scientist comme **informaticien** : qu'il s'agisse d'extraire les données parti-

nelles des systèmes d'information, de programmer les algorithmes qui lui permettront de les traiter, ou d'aider à concevoir les plateformes qui faciliteront l'exploitation rapide des résultats obtenus, le Data Scientist se doit de maîtriser les langages de programmation et les environnements technologiques (en particulier ceux du Big Data) adaptés aux différents cas qu'il pourra rencontrer.

- Le Data Scientist comme **expert métier** : parce que ses analyses doivent être menées dans une logique de recherche d'efficacité et de rentabilité de l'entreprise, le Data Scientist doit nourrir un dialogue métier avec les pilotes de processus qu'il accompagne, et doit par ailleurs être force de proposition sur les stratégies à mettre en œuvre ou tactiques à adopter, eu égard aux enseignements qu'il tire de ses analyses. A ce titre, il ne peut être un simple expert technique, mais doit garder les yeux grands ouverts sur les enjeux *business* de ses travaux.

Pour les décideurs désireux de ne pas manquer le virage du Big Data, le recrutement (ou l'identification dans les équipes en place) d'un ou plusieurs Data Scientists semble donc être une condition incontournable du succès.

Les pays anglo-saxons ne s'y sont d'ailleurs pas trompés : des formations dédiées à la Data Science se mettent progressivement en place dans les plus grandes Universités; les entreprises embauchent à prix d'or et en volume les Data Scientists capables de « faire parler leurs données » et nombreuses sont celles qui commencent à les placer aux plus hauts niveaux de responsabilité dans la hiérarchie.

Placez-les sous la responsabilité d'une autorité légitime

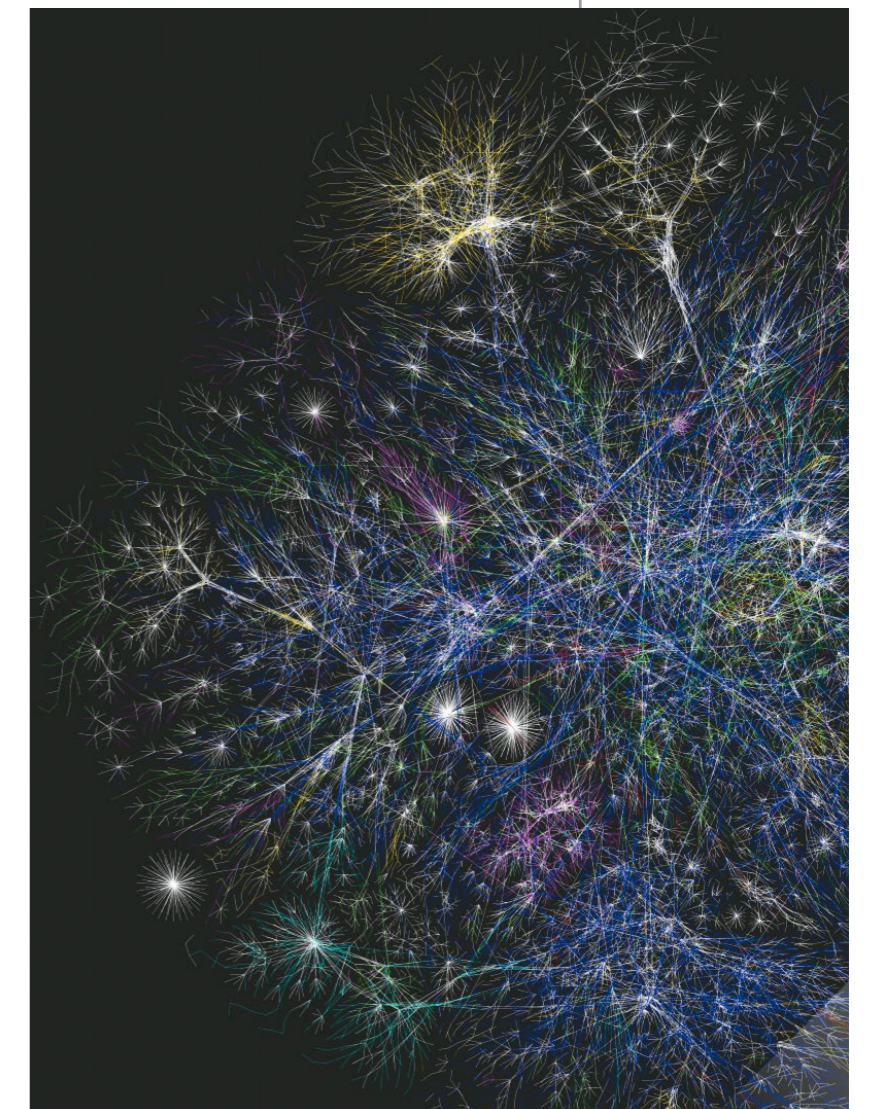
Plusieurs démarches et réflexions sur le sujet du Big Data sont probablement déjà engagées au sein de votre entreprise :

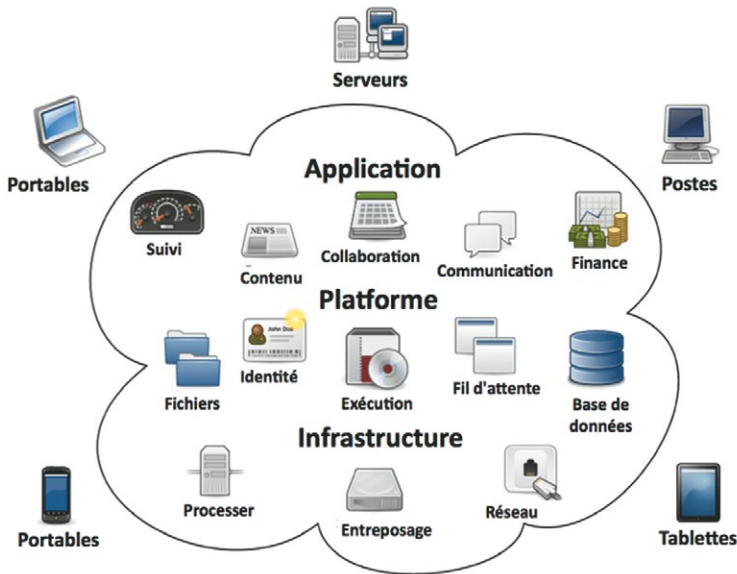
- Certaines sont initiées par la Direction des Systèmes d'Information qui se pose des questions sur l'opportunité d'investir, et sur le choix des outils à mettre en place pour prendre le relais et venir compléter les architectures décisionnelles déjà en place.
- D'autres sont engagées par la Direction Marketing ou par les Directions de Support

aux Opérations qui s'interrogent sur ce que peut leur apporter le Big Data dans les domaines de l'amélioration de la connaissance clients, dans la mise en place d'approches relationnelles beaucoup plus individualisées, dans l'optimisation de la conception des offres et du marketing produits, dans la gestion de l'image de marque sur les réseaux sociaux.

- D'autres encore sont menées par la Direction Financière sur les problématiques de détection de fraude, de management du risque, de pilotage de la marge.

Ces approches, pour légitimes qu'elles soient, ne permettent pas d'inscrire les réflexions dans une démarche globale de valorisation des données. Mieux vaut mettre en place un « projet Data Science » transverse, réunissant l'ensemble des parties prenantes :





Le Nuage

- Les Directions Fonctionnelles (Marketing, Finances, Contrôle de Gestion, Opérations...) utilisatrices, mais aussi souvent co-auteurs des enseignements tirés d'une analyse plus fine de l'information,
- La Direction des Systèmes d'Information, en tant que garant de la pertinence des choix technologiques et responsable de la qualité et de la maîtrise des données véhiculées par les systèmes de l'entreprise,
- Les experts quantitatifs (souvent disséminés dans différentes Directions Fonctionnelles), capables de réfléchir à l'apport potentiel de méthodes quantitatives simples ou avancées pour améliorer les processus métier existants.

La mise en œuvre d'un tel projet présente bien sûr des difficultés, au premier rang desquelles la résistance naturelle des acteurs de l'entreprise confrontés à une évolution des pratiques qu'ils ne pourront pas entièrement maîtriser – d'autant que toutes les fonctions de l'entreprise sont potentiellement impactées ou *a minima* concernées par le sujet.

« Toutes les fonctions de l'entreprise sont potentiellement impactées ou, a minima, concernées par le sujet. »

- Participation à la définition du volet analytique de la trajectoire des systèmes d'information (*business intelligence, data-mining, Big Data, data visualization...*),

- Gouvernance du référentiel de données de l'entreprise et maintien en qualité de l'actif informationnel transverse.

Dès lors, un projet de ce type devrait être mis entre les mains d'une autorité explicitement reconnue par la hiérarchie de l'entreprise (par exemple, en la rattachant à la Direction Générale), maîtrisant la culture et les différents processus au cœur des métiers de l'entreprise, connaissant les principaux systèmes d'information de l'entreprise ainsi que les données qu'ils véhiculent, et aguerrie aux techniques de traitement et de valorisation scientifique de l'information (statistique, mathématique, recherche opérationnelle...).

Ainsi structurée dans une Direction autonome, l'équipe de Data Scientists pourra mener à bien sa mission de valorisation des données de l'entreprise :

- Cartographie permanente des processus métier pour lesquels une approche Data Science peut procurer un avantage compétitif,
- Qualification constante des informations nécessaires aux travaux d'analyse permettant d'améliorer les processus sélectionnés à l'aide d'approches analytiques,
- Prise en charge des travaux analytiques transverses, avec préconisations d'actions stratégiques et tactiques aux pilotes de processus métier,

Un tel dispositif - qui permet à la Direction des Systèmes d'Information de centrer ses activités sur les Systèmes en confiant celles de l'Information et de sa valorisation à une Direction dédiée - apparaît déjà aux Etats-Unis dans les entreprises les plus en pointe dans l'exploitation analytique de leurs données. La responsabilité de *Chief Data Officer* y a déjà trouvé sa place dans la fameuse « C-Suite », en tant qu'autorité en charge de la gouvernance des données de l'entreprise, et de leur exploitation au service des processus, cœur de métier de l'entreprise. ■