

# BIG DATA

2013 / 2014

## GUIDE DU BIG DATA

L'ANNUAIRE DE RÉFÉRENCE  
À DESTINATION DES UTILISATEURS



GUIDE DU BIG DATA

2013 / 2014

# Big Data at the Speed of Business

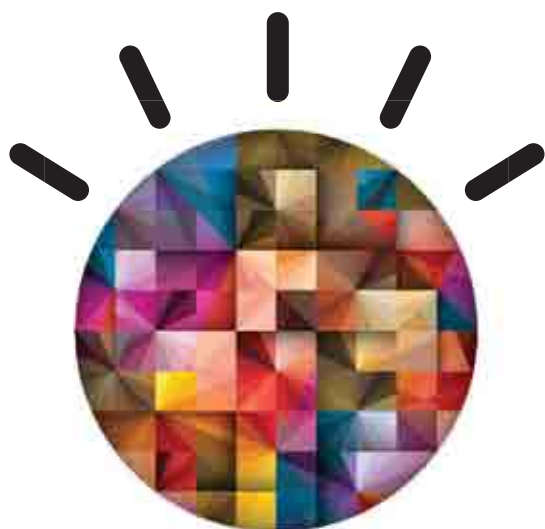
Big data n'est pas qu'une question de Volume.

Il s'agit pour les organisations de saisir les opportunités qu'offrent

l'exploitation de toutes les natures de données qu'elles soient internes ou externes.

IBM Software

# IBM big data



IBM a développé le portefeuille le plus étoffé pour gérer la donnée, la transformer en information puis générer des actions :

**Système pour Hadoop** : traiter et analyser tout type de données sur des clusters de serveurs.

**Stream Computing** : analyser en temps réel (flux) des volumes massifs de données.

**Data Warehousing** : effectuer des analyses "in-database" poussées.

**Data Visualisation** : chercher, découvrir et visualiser des données quelles que soient leur source ou leur format.

**Business Intelligence** : accéder et analyser les informations nécessaires pour améliorer la prise de décisions.

**Analyse Prédictive** : découvrir des tendances non visibles et anticiper avec précision les résultats de l'entreprise.

Avec une croissance annuelle du volume des données de 56%, Passez au Big Data avec IBM :

<https://ibm.biz/BdxkwP>

Apparu il y a seulement trois ans dans la sphère décisionnelle, le Big Data s'est déjà imposé comme l'innovation business numéro 1 de ce début de décennie, à la frontière entre technologie et management. Dans tous les secteurs où la gestion des données représente un actif indispensable de l'entreprise, l'engouement a été immédiat avec à la clé la formation d'un véritable écosystème du Big Data : des fournisseurs IT, qui ont développé une offre variée autour des technologies Hadoop, aux utilisateurs finaux (grandes entreprises, start-ups), qui cherchent à construire un avantage concurrentiel sur l'exploitation de leurs données, en passant par des entreprises spécialisées « data » (éditeurs business intelligence, consultants Big Data...), **le paysage français s'est étoffé et structuré en l'espace de deux ans.**

La ministre chargée de l'Economie Numérique, Fleur Pellerin, a d'ailleurs annoncé, au cours du premier semestre 2013, un plan de soutien à la filière du Big Data : 11,5 millions d'euros issus des Investissements d'Avenir seront ainsi alloués à des projets Big Data et un programme de formation et de financement sera mis en place.

Pourquoi un tel engouement ? Parce que les promesses de croissance autour de l'exploitation des grands volumes de données sont immenses : réduction des coûts, augmentation des ventes, relance de l'innovation, aide à la décision...

**les espoirs liés à la « data-driven company » attirent de nombreux professionnels, dans tous les secteurs** (télécom, finance, grande distribution, industrie... ).

Mais ce nouveau modèle impose de réfléchir aux conditions de mise en œuvre : le choix de la technologie d'abord, l'organisation et les compétences ensuite.

**Ce Guide a pour vocation d'aider les professionnels désireux d'exploiter leurs données à appréhender l'environnement du Big Data et à comprendre les fondamentaux d'une démarche Big Data en entreprise.** Quelles technologies

choisir ? Qui sont les acteurs du marché ? Quels métiers mettre en avant ? Et quel est le cadre juridique ?

Nous avons recherché pour vous les informations les plus récentes sur le phénomène Big Data, en faisant un état des lieux de l'offre et de la demande sur l'année 2013 et en photographiant le panorama actuel des fournisseurs, dans la deuxième partie du Guide. De nombreux experts ont également accepté de répondre à nos questions pour témoigner de leurs expériences et apporter leurs conseils.

Vous trouverez donc dans ce guide les réponses aux questions les plus usuelles posées par les curieux du Big Data. Il vous permettra de mesurer l'impact et l'efficacité de cette démarche pour votre entreprise et d'envisager, peut-être, à terme, un basculement vers ces technologies et la maîtrise instantanée de l'information.



**Blandine LAFFARGUE**

Directrice de publication

+33 (0)1 45 63 43 58 / blaffargue@corp.events.fr



*Le Guide du Big Data est réalisé  
par la société Corp Events*

<b>LE BIG DATA, UNE REVOLUTION DANS LE TRAITEMENT DES DONNEES ?</b> .....	<b>05</b>
• <b>Comment définir le Big Data ?</b> .....	<b>05</b>
• <b>Une matière première, les données</b> .....	<b>06</b>
• <b>Une Révolution technologique</b> .....	<b>07</b>
• <b>Les usages du Big Data</b> .....	<b>08</b>
<i>Interview de Mathias Herberts, Disruptive Engineer &amp; CTO, Cityzen Data</i> .....	<b>09</b>
<b>UN MARCHÉ MONDIAL – BILAN ET PERSPECTIVES</b> .....	<b>10</b>
• <b>Le marché Big Data en Chiffres</b> .....	<b>10</b>
• <b>Les acteurs du Big Data</b> .....	<b>11</b>
<i>Interview de Matteo Pacca, Partner at McKinsey&amp;Company</i> .....	<b>12</b>
<b>LES TECHNOLOGIES BIG DATA</b> .....	<b>14</b>
• <b>Petit Lexique des Technologies Big Data</b> .....	<b>14</b>
• <b>Tour d’horizon des briques fonctionnelles en lien avec le Big Data</b> .....	<b>14</b>
• <b>Les technologies Big Data face aux besoins de la DSI</b> .....	<b>15</b>
<i>Interview Isabel Drost, membre de l’Apache Software Foundation, co-Fondatrice et contributrice pour Apache Mahout</i> .....	<b>16</b>
<b>EXEMPLES D’APPLICATIONS EN ENTREPRISES</b> .....	<b>18</b>
• <b>Projet Big Data de customisation marketing: Monoprix et dunnhumby</b> .....	<b>18</b>
• <b>Projet Big Data d’optimisation du traitement industriel : La Poste</b> .....	<b>20</b>
• <b>Projet Big Data de Recherche : Projet DEUS de l’Observatoire de Paris</b> .....	<b>22</b>
• <b>Projet Big Data de Localisation : Le géomarketing de SFR</b> .....	<b>24</b>
• <b>Projet Big Data d’Agrégation de Données : Le projet Ellicible de Coface Services</b> .....	<b>26</b>
<b>LES METIERS DU BIG DATA</b> .....	<b>28</b>
<b>Quelles sont les nouvelles compétences liées au Big Data ?</b> .....	<b>28</b>
<b>Les nouveaux métiers du Big Data</b> .....	<b>28</b>
<b>Les offres de formation en France</b> .....	<b>29</b>
<b>L’impact du Big data sur l’organisation de l’entreprise</b> .....	<b>29</b>
<i>Interview de Jacques Froissant, président d’Altaïde</i> .....	<b>31</b>
<b>LES POUVOIRS PUBLICS FACE AU BIG DATA</b> .....	<b>32</b>
<b>L’Open Data</b> .....	<b>32</b>
<b>Soutenir le décollage du marché : les investissements publics dans le domaine du Big Data</b> .....	<b>32</b>
<b>La problématique de confidentialité des données</b> .....	<b>33</b>
<i>Interview de Thierry Dor, associé au cabinet Gide Loyrette Nouel</i> .....	<b>35</b>

## GUIDE PROFILS DE SOCIÉTÉ

<b>ACCENTURE</b> .....	<b>36</b>	<b>PARSTREAM</b> .....	<b>68</b>
<b>ACTUATE</b> .....	<b>38</b>	<b>PENTAHO</b> .....	<b>70</b>
<b>AFFINI-TECH</b> .....	<b>40</b>	<b>QLIKTECH</b> .....	<b>72</b>
<b>BIME</b> .....	<b>42</b>	<b>QUARTET FS</b> .....	<b>74</b>
<b>CGI</b> .....	<b>44</b>	<b>SAS INSTITUTE</b> .....	<b>76</b>
<b>CLUB DÉCISION DSI</b> .....	<b>46</b>	<b>SEMSOFT</b> .....	<b>78</b>
<b>DATA &amp; DATA CONSULTING</b> .....	<b>48</b>	<b>SENTELIS</b> .....	<b>80</b>
<b>EXALEAD</b> .....	<b>50</b>	<b>SINEQUA</b> .....	<b>82</b>
<b>FASTCONNECT</b> .....	<b>52</b>	<b>SOGETI HIGH TECH</b> .....	<b>84</b>
<b>HEWLETT-PACKARD</b> .....	<b>54</b>	<b>SPLUNK INC.</b> .....	<b>86</b>
<b>IBM</b> .....	<b>56</b>	<b>SRD CONSEIL</b> .....	<b>88</b>
<b>INTERSYSTEMS</b> .....	<b>58</b>	<b>SYNCSORT</b> .....	<b>90</b>
<b>JASPERSOFT</b> .....	<b>60</b>	<b>TABLEAU SOFTWARE</b> .....	<b>92</b>
<b>KEYRUS</b> .....	<b>62</b>	<b>WORLDLINE</b> .....	<b>94</b>
<b>MAPR TECHNOLOGIES</b> .....	<b>64</b>	<b>YSANCE</b> .....	<b>96</b>
<b>NEO TECHNOLOGY</b> .....	<b>66</b>		



# Le Big Data, une révolution dans le traitement des données ?

## Comment définir le Big Data ?

*Si le Big Data est devenu un terme à la mode dans les milieux décisionnels, il est parfois difficile de s'accorder sur une définition unique : on parle tantôt de technologies de traitement, tantôt de modèle économique, voire de mode d'organisation spécifique de l'entreprise.*

### La définition technologique

La définition initiale donnée par le cabinet McKinsey and Company en 2011 s'orientait d'abord vers la question technologique, avec la célèbre **règle des 3V** : un grand Volume de données, une importante Variété de ces mêmes données et une Vitesse de traitement s'apparentant parfois à du temps réel. Ces technologies étaient censées répondre à l'explosion des données dans le paysage numérique (le « data deluge »). Puis, ces qualificatifs ont évolué, avec une vision davantage économique portée par le 4ème V de la définition, celui de Valeur, et une notion qualitative véhiculée par le 5e V, celui de Véracité des données (disposer de données fiables pour le traitement). Ces cinq éléments ont servi pendant longtemps de boîte à outils pour comprendre les fondements du Big Data, à savoir **l'apparition de technologies innovantes capables de traiter en un temps limité de grands volumes de données afin de valoriser l'information non exploitée de l'entreprise.**

### L'aspect organisationnel et stratégique

Aujourd'hui, cette définition s'est amplifiée : **on parle davantage de « démarche Big Data » pour signifier l'importance prise par les données dans l'univers de l'entreprise.** Celles-ci deviennent un mode de décision, un actif stratégique pour se différencier de la concurrence : c'est grâce à ces données rapidement transformées

en informations que les managers sont capables de prendre les bonnes décisions et de créer de la valeur.

Du coup, le Big Data sous-entend un renversement total de paradigme dans le mode d'organisation de l'entreprise, celle-ci devenant « data-centric », c'est-à-dire guidée par les données. Les métiers sont amenés à évoluer et à travailler davantage en synergie, tandis que les temps de décisions sont considérablement raccourcis, en raison des actualisations permanentes de l'information. Mettre en place le Big Data, c'est donc lancer une « démarche », plutôt qu'installer un outil.

### Une Révolution économique ?

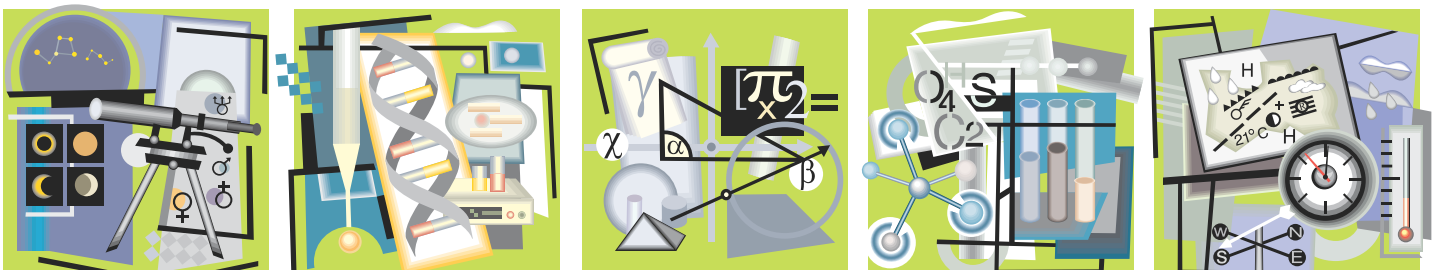
En cela, **certaines considèrent que le Big Data s'apparente à une véritable Révolution Industrielle parce qu'on se trouve en présence d'un vrai bouleversement qui se généralise à tous les secteurs économiques** : la donnée joue le rôle de matière première, les technologies jouent celles d'outils de production, et le résultat est un changement de paradigme dans la manière d'organiser les entreprises et de générer de la valeur dans tous les domaines de l'économie. Une comparaison osée et peut-être excessive, mais qui traduit bien l'impact global du Big Data et sa nature autant économique que technologique.

## Les 5V du Big Data

Volume  
Vitesse  
Variété  
Valeur  
Véracité

On le voit, les définitions du concept sont en pleine évolution, en lien certainement avec le degré de mise en pratique du Big Data dans les entreprises. Si les premières expérimentations sont encore timides en France en 2013, l'intérêt des managers est bien réel et le Big Data pourrait dès lors très rapidement aborder son tournant économique.

« Du coup, le Big Data sous-entend un renversement total de paradigme dans le mode d'organisation de l'entreprise, celle-ci devenant « data-centric », c'est-à-dire guidée par les données »



# Une matière première, les données

*A la base du Big Data, il y a donc ces fameuses données... Un « data deluge » comme le dit le terme consacré, car la production de données numériques aujourd'hui se mesure en exaoctets, soit  $10^{18}$  octets. Ce qui interpelle, c'est essentiellement la croissance exponentielle de ces données : on estime que tous les deux ans, il se crée autant de données que depuis le début de l'humanité, ce qui amène les prévisionnistes à parler de 40 zettaoctets ( $10^{21}$ ) pour 2020 !*

## Le data deluge

D'où proviennent ces données et comment expliquer cette explosion soudaine ? Il y a bien sûr la mise à disposition généralisée d'outils numériques de plus en plus performants et connectés, tels que les ordinateurs et les smartphones reliés à Internet en permanence. L'accès à l'information instantanée s'est démocratisé par les réseaux sociaux et, dans un nombre important de pays, les administrations publiques lèvent le voile sur leurs données à travers le mouvement d'Open Data.

Mais au-delà, on envisage de plus en plus la multiplication de données issues d'objets interconnectés, de type capteurs ou puces mobiles : ce qui avait été approché avec le RFID (par exemple au sein des cartes de transport ou par code-barre dans les chaînes logistiques) pourrait être généralisé dans les années à venir à travers l'Internet des Objets. Concrètement, il est aujourd'hui technologiquement possible de connecter entre eux les objets de son quotidien, par exemple son réveil avec son calendrier Outlook, ou bien son véhicule avec les applications de stationnement, ou encore son réfrigérateur avec son panier de courses en e-commerce. Des perspectives technologiques qui restent encore expérimentales mais pourraient vite exploser : Ericsson a ainsi estimé le nombre d'objets connectés dans le monde à 50 milliards en 2020 (12 milliards aujourd'hui).

## Une variété de données

### Les données multistructurées

Cette mutation dans l'univers numérique n'est pas seulement quantitative : au-delà de leur nombre, les données mises à disposition sont également de plus

« Loi de Moore : Largement répandue dans la sphère scientifique, bien qu'elle n'ait pas été prouvée, la loi de Moore affirme que la capacité des ordinateurs (et notamment le volume des données) double tous les 18 mois. »

## octets

1 Mégaoctet =  $10^6$  octets

1 Gigaoctet =  $10^9$  octets

1 Téraoctet =  $10^{12}$  octets

1 Pétaoctet =  $10^{15}$  octets

1 Exaoctet =  $10^{18}$  octets

1 Zettaoctet =  $10^{21}$  octets

# 40

zettaoctets  
de données  
en 2020

# 50

milliards  
d'objets connectés  
à la même date

en plus variées et de plus en plus distribuées (c'est-à-dire issues de sources disparates).

Les données textuelles, les données de logs, les données mobiles, les vidéos, etc. sont ainsi venues perturber les technologies de traitement traditionnelles, car ce ne sont pas des données structurées telles qu'on pouvait en trouver auparavant (comme par exemple les données chiffrées ou transactionnelles...). On parle alors de données non-structurées ou bien multistructurées.

Tout l'enjeu technologique des années 2000 a donc résidé dans la captation et le traitement de ces nouveaux formats de données.

## La question de la qualité des données

Mais, passée cette interrogation, une autre problématique s'est installée : comment savoir si ces données sont intrinsèquement fiables et propres à l'utilisation ? Les données erronées, les doublons, les informations périmées sont en effet légion au cœur du data déluge ce qui signifie que, traitées de la même façon que les autres, elles peuvent fausser des analyses et des décisions. Quels outils mettre en place dès lors pour empêcher ces altérations ?

Ce risque lié à la qualité des données a fait dire à de nombreux experts que nous étions entrés dans l'ère de la « smart data » davantage que dans celle du Big Data : plutôt que de traiter en bloc tout l'univers des données disponibles, il s'agit de trouver la bonne donnée, fiable et utile, qui permettra d'obtenir de l'information et créer de la valeur.



# Y2K

Les années 2000 ont permis la captation et le traitement de nouveaux formats de données

## Une révolution technologique

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

### L'augmentation des capacités de stockage

Le stockage, d'abord : avec la croissance exponentielle des données, il est vite devenu indispensable de faire évoluer les baies de stockage traditionnelles pour les mettre en capacité d'absorber au fur et à mesure la multiplication de ces données. Une problématique d'autant plus réaliste que les coûts de stockage étaient en chute libre depuis le début des années 2000, à moins de 1 dollar le gigaoctet...

Technologiquement, ce besoin s'est traduit par l'apparition de la « scalabilité », c'est-à-dire l'ajustement progressif et continu de l'outil de stockage à la masse des données collectées. Une innovation majeure permettant aux DSI de stocker toutes sortes d'informations très rapidement, sans arbitrage sur leur utilisation future, et surtout sans la nécessité d'accroître (et donc changer) l'infrastructure existante. Aujourd'hui, de nombreux fournisseurs développent ce type de technologies au sein d'appliances (c'est-à-dire d'outils combinés de stockage et de traitement).

### L'influence du Cloud

Mais c'est particulièrement avec l'apparition du Cloud que le Big Data a connu un développement accéléré : en mutualisant les données dans le nuage et en démocratisant ainsi l'accès des entreprises au stockage et au traitement de tous types de données, le Cloud a créé les conditions pour une généralisation du Big Data. Petites et grandes entreprises ont pu souscrire à des offres « as a service » pour explorer leurs données internes mais également d'autres données accessibles depuis le Cloud. Il est d'ailleurs très fréquent de voir des offres Big Data appuyées sur l'informatique en nuage.



### L'apparition de technologies de traitement adaptées

Enfin, le dernier facteur technologique qui a permis au Big Data de décoller, probablement le plus célèbre, concerne

la révolution des outils de traitement. Conjointement à l'explosion des données numériques, les années 2000 ont assisté à l'apparition de nouvelles bases de données, adaptées aux données non structurées (Hadoop), ainsi qu'à la mise au point de modes de calcul à haute performance (MapReduce) : c'est la révolution Hadoop/MapReduce (cf partie IV, Les Technologies Big Data).

Ces deux innovations, portées par Google et Yahoo, ont posé les bases du traitement Big Data actuel : il est devenu possible, grâce à elles, de traiter d'importants volumes de données en un temps limité (réduit de près de 50 fois par rapport aux précédentes technologies), pour tous types de données, structurées ou non. Le fait qu'Hadoop soit mis à disposition en Open Source a probablement contribué à la généralisation rapide de l'outil et à l'explosion ultérieure du Big Data.

Les conditions d'un essor du Big Data étaient donc réunies dès la fin des années 2000 : la mise à disposition de technologies a accompagné l'explosion des données. Plus qu'un effet cause-conséquence, c'est davantage une évolution conjointe qu'il faut mettre en avant, à la manière d'un processus auto-entretenu. En revanche, peut-on parler de « révolution » dans les usages du Big Data ?

« La mise à disposition de technologies a accompagné l'explosion des données. »



# Les usages du Big Data

*La question agite en effet de nombreux experts : le Big Data est-il si révolutionnaire que le laissent entendre les journalistes et les acteurs du secteur ? L'utilisation de données à des fins de traitement et d'analyse est en effet inhérente à tout pilotage d'entreprise, et des outils anciens existent pour cela (depuis l'antique livre de comptes jusqu'aux tableaux de Business Intelligence, en passant par les traditionnelles bases Excel et Access).*

Beaucoup estiment donc que le Big Data n'est qu'un passage à l'échelle des traitements traditionnels : ce que l'on traitait auparavant en petite quantité, on le traitera en grande quantité avec une fiabilité renforcée. Les approches statistiques et la nature des résultats trouvés ne seraient en rien modifiés...

C'est ici que les désaccords naissent car le Big Data, en traitant une multiplicité de données issues de sources variées et selon un quasi-temps réel, a ouvert la voie à d'autres utilisations de la donnée et à de nouvelles méthodes d'analyse :

- **En marketing**, c'est tout le secteur qui se trouve renouvelé : le Big Data permet en effet aux professionnels du secteur de connaître leur client « à 360° », c'est-à-dire à la fois par son parcours internet mais également par ses achats en magasin ou ses préférences affichées sur les réseaux sociaux. Anticiper les besoins de celui-ci et cibler des offres personnalisées est devenu le credo du marketing « data-driven », qui met en avant des techniques inédites : le Real-Time Bidding pour l'achat d'espaces en temps réel, le Retargeting pour le ciblage personnalisé ou encore l'analyse de sentiment pour la détection de comportements sur les réseaux sociaux. Le marketing se fait de plus en plus prédictif avec le Big Data, et l'on assiste à une éclosion de nouveaux modèles statistiques davantage inductifs.



- **Dans le domaine du pilotage de l'entreprise**, les usages sont également nombreux et porteurs d'innovation : en assurant une circulation immédiate et généralisée de l'information sur l'activité, le Big Data laisse entrevoir une optimisation complète des processus et des ressources métiers. Il réduit de facto le temps de réaction face à des erreurs ou des pannes et permet d'ajuster en permanence les équilibres offre-demande et temps-ressource. C'est une promesse importante dans des secteurs comme ceux de l'énergie ou des transports qui sont constamment portés par la logique de flux ; outre une réduction importante des coûts, le Big Data permet ici d'identifier au plus près les moteurs de l'activité, ce qui n'était pas possible avec les indicateurs traditionnels, soumis à des délais de latence bien plus importants.



- **Pour la Recherche**, domaine d'application originel du Big Data, l'apport de celui-ci est assez évident : en autorisant le traitement de multitudes de données, le Big Data permet à la science de réaliser des avancées importantes, lorsqu'il s'agit d'explorer l'infiniment petit (ex : exploration géologique), de croiser des données complexes (ex : imagerie) ou d'effectuer des simulations (ex : domaine spatial). C'est d'ailleurs en génétique que le Big Data a fait ses premières armes car ce secteur réclamait une approche à la fois quantitative et qualitative avancée.



- **Dans le domaine de l'Information**, tout simplement, qu'elle soit corporate ou grand public, le traitement des Big Data a profondément modifié la donne : pour une requête donnée, il est désormais possible d'accéder à un croisement d'informations très disparates, issues de sources jusque-là négligées. L'instantanéité des réseaux sociaux est à ce titre une innovation de taille : l'analyse des tweets est devenue une source de renseignements courante pour comprendre les comportements ou les goûts de populations segmentées. De plus, au-delà de la compréhension de phénomènes, la data s'avère un outil de communication efficace pour faire passer des messages factuels aux publics : c'est tout le rôle du data journalisme, qui vise à mettre en lumière des événements complexes au travers de la représentation de données multiples.

On le voit, le Big Data a donc eu un effet profondément novateur sur de nombreuses activités : il a ouvert de nouveaux champs d'exploration, avec à la clé des informations et fonctions inédites. **Mais, au-delà, le Big Data a également permis de créer intégralement de nouveaux services, de nouvelles filières fondées sur des données inexplorées jusqu'alors.**



• **Les données mobiles**, par exemple, dont le traitement temps réel est devenu un enjeu stratégique pour de nombreux secteurs, sont un important gisement d'innovation : en effet, la possibilité de géolocaliser des utilisateurs a ouvert de nombreuses perspectives aux annonceurs qui y voient une opportunité unique de placer leurs publicités à proximité des lieux d'achat. De la même façon, les opérateurs telecom travaillent avec les autorités publiques pour comprendre les flux de populations grâce à ces données et, dans le secteur du tourisme, des services inédits de localisation sont mis en place.

Les usages du Big Data sont donc variés et très innovants, ce qui peut expliquer l'engouement suscité auprès des entreprises et administrations publiques. Le débat sur l'aspect révolutionnaire du Big Data reste cependant ouvert, car certains considèrent que, si les données ont changé, les algorithmes restent quant à eux assez similaires. L'évolution de ceux-ci reste probablement l'un des chantiers de ces prochaines années.



Mathias Herberts  
(à droite) reçoit le prix de  
l'innovation Big Data



# Interview

## Mathias Herberts

Disruptive Engineer & CTO, *Cityzen Data*

### « L'élément de nouveauté dans le Big Data, ce n'est pas tant la technologie que le pilotage »

#### 1) Vous êtes un des pionniers du Big Data en France... Comment (et pourquoi) avez-vous eu l'idée de mettre en place ce type d'outils ?

Tout est parti d'une expérience chez Google, où je suis allé travailler quelque temps en 2008 et où j'ai découvert une toute autre façon de prendre des décisions. Google avait théorisé une philosophie de travail régie par deux principes : 1) ne pas arriver en réunion en disant « je pense que » mais « les données prouvent que », 2) ce n'est pas l'opinion de la personne la mieux payée qui compte. Cela faisait écho à un livre que je lisais à l'époque, qui s'appelait « Super Crunchers » et qui prônait le raisonnement par les chiffres et les données.

Partant de là, je suis revenu au Crédit Mutuel avec la certitude que les données étaient fondamentales et qu'il était nécessaire de modifier nos processus de décision : pour éviter d'avoir une vision erronée de nos clients, de nos partenaires et de notre fonctionnement, il fallait remettre à plat notre socle technologique.

#### 2) Pour vous, le Big Data correspond-il à une rupture technologique majeure ? Quels sont les éléments de nouveauté ? Sont-ils uniquement technologiques ?

Pour moi, l'élément de nouveauté dans le Big Data, ce n'est pas tant la technologie que le pilotage : désormais, on a réduit les boucles de décision, tout est plus court dans l'information et la réaction. C'est là qu'il y a un tournant.

Bien sûr, cela ne se ferait pas sans la mise à disposition d'outils particulièrement opportuns à ce moment-là. Ce sont des outils qui ont accéléré le changement de comportement vis-à-vis des données : désormais, il est possible de stocker sans savoir exactement l'utilisation qu'on fera des données, simplement parce qu'on considère qu'elles ont une valeur intrinsèque.

#### 3) Quand on parle de Big Data, on parle d'Hadoop... pour vous, le lien est-il forcément évident ? Quels sont les apports de cet outil et, à l'inverse, que pouvez-vous lui reprocher ?

Même s'il existe d'autres outils, Hadoop reste la référence du traitement Big Data. Certes, c'est un outil qui réclame de s'investir dans la programmation mais il permet vraiment de faire des choses nouvelles à des coûts restreints et sur un grand volume de données.

Le principal reproche qu'on peut lui adresser, c'est d'avoir une courbe d'apprentissage assez longue et complexe ; mais une fois qu'on a compris comment fonctionnait la communauté Hadoop, et qu'on a pris le temps de rentrer dans les subtilités du code, c'est un outil très efficace.

#### 4) Quels sont pour vous les principaux défis à relever au sein des organisations utilisatrices du Big Data ?

Pour moi, le principal défi est organisationnel : la démarche Big Data est forcément transverse donc elle implique

de casser les cloisons entre services, ce qui n'est pas toujours évident. En terme SI, cela signifie qu'il faut aller récupérer les données dans tous les recoins, ce qui a de grandes implications au niveau de l'urbanisation du système d'information. Autre défi : aujourd'hui, on a tendance à penser qu'il faut passer par des offres packagées plutôt que par le code open source Hadoop. Cela peut être un gain de temps mais ça n'apporte pas les solutions en interne. Pour moi, il faut avant tout avoir des équipes dédiées qui se plongent résolument, de façon personnalisée, dans l'intégration de l'open source au système d'information, et qui soient capables de comprendre les besoins propres de l'entreprise.

Enfin, dernier challenge : l'utilisation de la donnée risque de remettre en cause pas mal de convictions et de stratégies. Il y aura donc forcément des réticences face au pouvoir de la donnée. Aux équipes informatiques et business de montrer la voie en pointant la valorisation et le potentiel des données !

« La démarche Big Data est forcément transverse donc elle implique de casser les cloisons entre services »



# Un marché mondial Bilan et perspectives

*Quelles sont les perspectives économiques du Big Data pour les fournisseurs et les utilisateurs ? Qui sont les acteurs qui dynamisent le marché ? Et dans quels secteurs les trouve-t-on ? Voici un petit panorama du marché du Big Data autour du monde pour l'année 2013.*

## Les chiffres du Big Data

→ Un marché de  
24 milliards de \$  
en 2016

→ Taux de croissance  
annuel : 31,7%

→ Pourcentage  
d'entreprises lançant  
des projets Big Data  
ou envisageant de  
le faire

- France : 10%
- Allemagne : 18%
- UK : 33%

→ Le Big Data fait  
partie des 5  
secteurs clé de  
la croissance  
américaine

→ En France, le  
Big Data pourrait  
générer 2,8 milliards  
d'euros et 10 000  
emplois directs  
(prévisions de l'AFDEL)

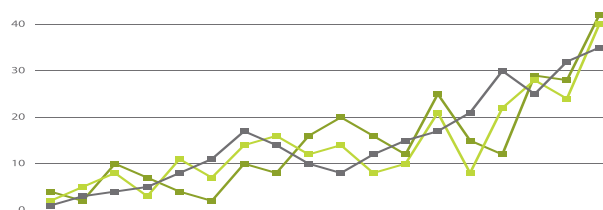
## Le marché Big Data en Chiffres

Depuis son décollage en 2011, le Big Data n'a cessé d'afficher des scores de croissance importants côté utilisateurs mais surtout côté fournisseurs : à ses débuts, on avait d'ailleurs parfois qualifié ce marché de « marché d'offre » pour souligner le fait que la demande était encore à la traîne. Qu'en est-il en 2013 ?

### Le marché des fournisseurs

Plusieurs études menées par des cabinets de conseil soulignent la forte croissance du marché du Big Data :

- Selon IDC, le marché du Big Data s'établirait à **24 milliards de dollars en 2016** (contre 3,3 milliards en 2010) avec un taux de croissance annuel de 31,7%.
- Ce marché est constitué des produits et services liés au Big Data, avec le découpage suivant :
  - Serveurs : 27,3%
  - Logiciels : 34,2%
  - Stockage : 61,4%
- D'après une étude IDC pour EMC de juin/juillet 2012, **les trois technologies privilégiées** dans le Big Data seraient les suivantes :
  - Les bases de données objets ou graphiques : 47%
  - L'indexation de contenus (moteurs de recherche avancés) : 38%
  - Les technologies en mémoire (in-memory database) : 37%



### Le marché des utilisateurs

Côté utilisateurs, la situation est plus contrastée en fonction des zones géographiques :

- Début 2013, le cabinet Forrester établissait sa **prévision de croissance des budgets IT à 6% au cours de l'année en Europe**, alors qu'elle est de 7,5% aux Etats-Unis
- IDC estime que seulement 10% des entreprises françaises utilisent déjà des solutions Big Data, tandis que 70% d'entre elles n'ont engagé ni projets ni réflexions en ce sens (chiffres de septembre 2012)
- Ainsi **58% des entreprises françaises estiment être en retard par rapport à leurs concurrents** (et 47% estiment être très en retard)
- En comparaison, 33% des entreprises britanniques et 18% des entreprises allemandes ont lancé une démarche Big Data
- Parmi les freins évoqués, la problématique du coût est souvent avancée, tout comme celle du manque de ressources et d'organisation pour mettre en place le Big Data (**64% des équipes IT estiment que la collaboration entre l'IT et les métiers devrait être renforcée**). Enfin, la question de la qualité des données (41%) ou de leur utilité pour les analyses prévues (37%) est un autre point soulevé par The Economist.

## Pourtant, les promesses sont nombreuses en matière de création de valeur :

- En juillet 2013, Mc Kinsey estime ainsi que le **PIB des Etats-Unis pourrait croître de 1,7% d'ici 2020 en s'appuyant sur le Big Data**. Celui-ci ferait partie des 5 secteurs phare de la croissance américaine.
- Les économies réalisées dans la santé et l'administration américaines pourraient atteindre 300 milliards de dollars !
- Dans le domaine de la grande distribution, le chiffre d'affaires pourrait croître de 40% en 5 ans, et de 30% sur la même période pour les télécoms !
- En France, le constat fait par l'AFDEL (association des éditeurs de logiciels français) est que **le Big Data pourrait générer 2,8 milliards d'euros et 10 000 emplois directs si les conditions de son déploiement étaient réunies** (c'est-à-dire avec un financement de 300 millions d'euros et la mise en place d'un écosystème). C'est cette estimation qui a d'ailleurs conduit le ministre à mettre en place un programme de soutien à la filière.
- Les 3 principales motivations des dirigeants pour mettre en place des solutions Big Data sont évoquées par Mc Kinsey : en Europe, c'est l'aspect « détection des tendances » qui occupe la première position, suivi par « un meilleur engagement client » puis « une prise de décision plus rapide ». Dans les marchés émergents en revanche, cet aspect « prise de décision » est placé en première position.

## Les acteurs du Big Data

### Les fournisseurs

De nombreux acteurs se sont positionnés rapidement sur la filière du Big Data, dans plusieurs secteurs :

- Dans le secteur IT, on trouve ainsi :
  - **Les fournisseurs historiques de solutions IT** (ex : IBM, SAP, Oracle, HP...)
  - **Les acteurs du Web** (ex : Facebook, Google...)
  - **Les spécialistes de solutions data et Big Data** (ex : Teradata, MapR, Hortonworks, EMC...)
  - **Les intégrateurs** (ex : Atos, Sopra Group, Accenture, Cap Gemini...)



→ Dans le secteur de l'analytique, de nombreux acteurs sont également présents :

- **Les éditeurs BI** (ex : SAS, Microstrategy, Qliktech...)
- **Des fournisseurs spécialisés dans l'analytique Big Data** (ex : Datameer, Zettaset...)

D'autre part, de nombreuses PME spécialisées dans le Big Data ont vu le jour, sur toute la chaîne de valeur du secteur. En France, les pionniers ont ainsi été :

- Pour la fourniture d'équipements et logiciels Big Data : Hurence, Dataiku
- Pour l'analyse de données : Criteo, Squid, Captain Dash, Tiny Clues
- Pour le conseil : Ysance

Ces fournisseurs établissent souvent des partenariats entre eux afin de fournir au client des solutions packagées intégrant à la fois l'infrastructure et les applications.

### Les utilisateurs

La mise en place du Big Data est vécue différemment selon les secteurs : **on a coutume de dire que les principaux secteurs d'application du Big Data sont les telecoms, la grande distribution (online et offline) et la finance**. Ce sont en effet les premières activités qui se sont penchées sur la problématique du Big Data car la captation de grands volumes de données pouvait rapidement leur donner un avantage comparatif.

Pour les opérateurs télécoms, c'était même une opportunité de création de

nouveaux services grâce au traitement des données mobiles géolocalisées.

Mais au-delà de ces secteurs habituellement réputés pour leur propension au Big Data, on voit germer des initiatives Big Data dans d'autres environnements : l'industrie française commence en effet à s'intéresser au phénomène, notamment sur ses qualités d'optimisation logistique et financière.

Les médias sont également de plus en plus friands de traitement de données, d'abord parce qu'ils peuvent obtenir une information temps réel et aussi parce qu'ils peuvent mesurer les préférences de leurs audiences.

**Mais le secteur qui fonde les espoirs les plus nombreux sur le Big Data est probablement celui de la santé**, où les perspectives dans la recherche mais aussi dans le ciblage des médicaments sont importantes. C'est d'ailleurs le terrain d'exploration originel du Big Data puisque c'est dans le domaine de la génétique que les premières expérimentations avaient eu lieu.

Actuellement, **c'est le mouvement du Quantified Self qui focalise les attentions** : mouvement orienté grand public, il vise à permettre aux individus de surveiller en permanence leurs indicateurs de bien-être (ex : poids, nutrition, fatigue) à l'aide de capteurs et de traitement de données temps réel. Encore au stade de l'expérimentation, cette activité pourrait être amenée à se développer rapidement à la faveur de l'essor des appareils mobiles.



# Interview

**Matteo Pacca**

Partner, *McKinsey&Company*



## « Le big data est un levier de performance sur lequel les entreprises françaises doivent accélérer »

### 1) Où en est le big data aujourd'hui en France et partout ailleurs dans le monde ? Quels sont les secteurs les plus moteurs ? Observez-vous une évolution significative ?

Il y a aujourd'hui un consensus autour de l'immense valeur potentielle que présentent les masses de données que nous générons chaque jour dans le monde. Selon les études de McKinsey, à l'échelle mondiale, le marché du Big Data devrait dépasser les 30 milliards de dollars US en 2013. Si le big data présente un potentiel pour tous les secteurs d'activité, trois secteurs sont particulièrement bien positionnés pour en tirer le meilleur parti : la distribution, du fait de grandes bases de clients et d'historiques d'achats, la santé (pharmacie et assurance santé), pour l'analyse de la

performance des médicaments (avant leur mise sur le marché, en phase de test, et après, pour mesurer leur efficacité), et enfin la banque-assurance, pour mieux cibler les produits financiers et surtout l'analyse du risque. Selon nos projections, **le marché du big data devrait croître de près de 15 % par an, pour répondre aux besoins engendrés par une croissance estimée des données pouvant aller jusqu'à 40 % par an.** Les dépenses des entreprises dans ce domaine devraient doubler d'ici cinq ans.

Pour les entreprises, investir dans le big data représente une opportunité majeure d'améliorer leur performance ; des travaux récents de McKinsey montrent d'ailleurs

### 2) Le Big Data véhicule de nombreuses promesses de développement pour les entreprises utilisatrices... Comment évaluez-vous le potentiel du Big Data pour les utilisateurs ? Pouvez-vous chiffrer ce développement ?

Le McKinsey Global Institute, l'institut de recherche de McKinsey, a étudié le potentiel de transformation du big data dans cinq secteurs : la santé aux Etats-Unis, l'administration du secteur public dans l'Union Européenne, la distribution aux Etats-Unis, l'industrie mondiale, et la géolocalisation\*. Ensemble, ces secteurs représentaient près de 40 % du PIB mondial en 2010 ; dans ces cinq domaines, nous avons identifié des leviers qui devraient permettre aux entreprises d'améliorer leur efficacité, grâce à une meilleure productivité et en accroissant la valeur ajoutée des produits et services qu'elles proposent. Le big data dans le domaine de la santé aux Etats-Unis présente par exemple un potentiel de création de valeur supplémentaire de près de 230 milliards d'euros par an, les deux tiers provenant d'une diminution des dépenses de santé nationales, soit près de 8 % des dépenses estimées en 2010. Dans le secteur manufacturier, le big data pourrait permettre de réduire de moitié les coûts d'assemblage et de développement de produits. Le secteur public européen pourrait voir ses coûts réduits de 15 à 20 %, diminuer la fraude et améliorer la collecte des impôts pour un total de 250 milliards d'euros par an. Pour les groupes de la grande distribution, le recours au croisement des données pourrait se traduire par un potentiel d'amélioration de 60% de leur rentabilité opérationnelle. Plus généralement, une meilleure utilisation des services de géolocalisation pourrait rapporter 100 milliards d'euros aux différents fournisseurs de services et 700 milliards d'euros aux consommateurs, qui bénéficieraient par exemple de la possibilité d'acheter un produit moins cher et plus près de chez eux.

« Le marché du big data devrait croître de près de 15 % par an, pour répondre aux besoins engendrés par une croissance estimée des données pouvant aller jusqu'à 40 % par an »

que les entreprises innovantes croissent de 10 % par an, contre - 3 % pour celles qui n'innovent pas. Pourtant, les entreprises - notamment françaises - sont encore trop timides et n'ont pas mobilisé les ressources adéquates en regard de leurs ambitions, que ce soit en termes de moyens financiers ou de talents : elles se situent aujourd'hui derrière les entreprises américaines, britanniques, scandinaves et allemandes. Parmi les freins au développement du big data - au-delà des aspects techniques, des questions de sécurité et du débat public - se pose une question organisationnelle : **afin de tirer le meilleur parti du big data, les entreprises vont devoir former et recruter les talents qualifiés**, mais aussi repenser leur structure organisationnelle et leur management. En outre, pour en saisir tout le potentiel, le big data doit être défini comme priorité stratégique de l'entreprise et nécessite une allocation de moyens financiers et humains adéquats. Les autorités ont d'ailleurs pris la mesure de l'importance du sujet et la France s'est engagée pour développer le big data avec un plan comportant plusieurs volets, comme la formation de « data scientists », la mise à disposition de certaines technologies, des financements... Le développement du big data en France passera par la mise en place d'un véritable écosystème.

\*McKinsey Global Institute, *Big Data : The next frontier for innovation, competition, and productivity*, mai 2011

# Les promesses du Big Data par secteur

- Santé :  
surplus de  
230 milliards € / an
- Industrie :  
réduction de 50%  
des coûts
- Secteur public en  
Europe : création  
de valeur de  
250 milliards d'€
- Grande distribution :  
amélioration de la  
rentabilité de 60%
- Valeur de la  
géolocalisation : gain  
de 700 milliards d'€  
pour les consom-  
mateurs



### 3) Quelles sont les principales applications du Big Data dans les entreprises ? A l'heure actuelle, quelles sont leurs proportions respectives ? Y en a-t-il certaines qui n'ont pas encore été explorées ?

Nous avons identifié cinq grandes applications du big data. Le big data permet d'abord **la transparence et la circulation des données** : dans le secteur public par exemple, une meilleure accessibilité des données entre différentes administrations ou services peut réduire les temps de recherche et de traitement des dossiers. Le big data permet également **la facilitation des diagnostics** : les entreprises peuvent ainsi identifier des variations de performance et mieux en comprendre les causes racines afin d'y remédier. Le big data permet encore **la segmentation des clients** afin de les cibler de manière plus fine : des distributeurs et des sociétés de services commencent à mettre en œuvre des techniques sophistiquées de big data telles que la micro segmentation en temps réel afin de mieux cibler les promotions et la publicité. Le big data permet aussi **d'améliorer certaines décisions humaines** par des algorithmes qui minimisent le risque et mettent en évidence des informations clés : les distributeurs peuvent ainsi optimiser certains processus, comme les inventaires ou le pricing. Enfin, le big data permet de concevoir de **nouveaux produits et services**, d'améliorer l'existant et in fine, d'inventer de nouveaux business models : par exemple, la localisation en temps réel a fait émerger de nouveaux types de services géo-localisés (navigation, assurances calculées sur la base des déplacements effectués, etc). Bien sûr, de nombreuses applications n'ont pas encore été explorées.

### 4) Que faut-il faire pour accélérer le développement du big data dans les entreprises ?

Peu d'entreprises ont développé à ce jour un véritable plan big data, incluant des discussions au sein de la direction générale, avec notamment une réflexion sur les investissements prioritaires et

des choix effectifs. Un tel plan repose sur trois étapes : d'abord, l'assemblage et l'intégration d'immenses volumes de données nouvelles afin d'identifier des informations intéressantes ; ensuite, la sélection de modèles d'analyse des données permettant d'optimiser les opérations et de prédire les conséquences des décisions business ; enfin, la création d'outils intuitifs qui transcrivent les analyses des modèles en actions réelles, et la formation des employés à l'utilisation du modèle. Une fois ce plan défini, pour le mettre en œuvre, il ne faut pas avoir peur d'expérimenter : choisir le domaine que l'on souhaite explorer, lancer un pilote et le tester, avec un esprit entrepreneurial et le courage d'une certaine prise de risque face à l'incertitude des découvertes qui seront faites.

De fait, développer le big data au sein des entreprises implique une évolution du rôle des DSI : ceux-ci doivent désormais, tout en se construisant leur propre réseau technologique, être force de proposition pour les business units afin d'améliorer leur performance. Pivot de la mise en œuvre du big data dans l'entreprise, ils doivent favoriser le dialogue entre les métiers et la technologie, au service d'une meilleure performance.

« Peu d'entreprises ont développé à ce jour un véritable plan big data »

Enfin, pour les fournisseurs de technologie, cela implique également une évolution. Ceux-ci ont encore une approche trop technique, pas assez tournée vers le client. Ils parlent d'accélération des performances techniques et ont une vision orientée vers les systèmes d'information : l'enjeu est désormais de proposer une approche plus « métiers », de montrer aux entreprises comment le big data peut véritablement avoir un impact business. Dans un contexte économique morose, les entreprises sont d'autant plus à l'écoute de nouveaux relais de croissance potentielle.



# Les technologies Big Data

## Petit Lexique des Technologies Big Data

*L'univers technologique du Big Data s'appuie sur des outils bien identifiés qui constituent la base innovante de ce mode de traitement. A eux seuls, ces outils résument le vocabulaire technologique du Big Data et en constituent la référence. Mais de quoi parle-t-on exactement lorsque l'on évoque ces innovations ? Voici un petit lexique simplifié qui permettra même aux moins technophiles de comprendre les fondements techniques du Big Data.*



- **TRAITEMENT MASSIVEMENT PARALLÈLE** (MPP = massively parallel processing) : le Big Data repose sur des systèmes de traitement de données organisés sur plusieurs nœuds ou clusters parallèles. Généralisé dans les années 2000, ce type de traitement permet de générer des calculs à très haute performance sur des infrastructures (processeurs, serveurs) réduites.
- **HADOOP** : emblème par excellence du Big Data, Hadoop est une architecture spécifique de bases de données, permettant de traiter en grand nombre tous types de données (y compris les données non structurées). On dit qu'elle est organisée sur un mode non-relationnel, c'est-à-dire que les données ne sont pas triées en fonction de leurs relations entre elles. Elle constitue en quelque sorte la librairie à partir de laquelle on pourra effectuer des calculs parallèles (via MapReduce), sachant que ces données sont distribuées (c'est-à-dire qu'Hadoop les « récupère » depuis diverses sources). Créé par la fondation Apache (voir interview d'I. Frost-Dromm), Hadoop est un framework Open Source, c'est-à-dire que son code et ses spécifications sont disponibles en libre accès pour ceux qui cherchent à l'implémenter dans leur système.
- **MAPREDUCE** : couplé à Hadoop (qui est en fait son implémentation principale), MapReduce est le mode de calcul permettant de traiter les big data. Il présente une fonction Map (distribution des données sur plusieurs clusters parallèles où les calculs intermédiaires seront effectués) et une fonction Reduce (les résultats des calculs intermédiaires distribués sont recentralisés en vue du calcul final). MapReduce est issu de la recherche Google et a été développé au même moment qu'Hadoop ce qui explique l'association traditionnelle que l'on fait des deux outils.
- **NOSQL** : on parle de NoSQL lorsque des requêtes sont effectuées dans un langage de programmation qui n'est pas limité au SQL. Ce dernier est en effet le langage le plus utilisé pour des traitements de données mais ses normes de définition, parfois trop strictes, ne sont pas adaptées aux données massives et non-structurées. Du coup, certaines requêtes sont effectuées indépendamment de ces normes et donc de ce langage propre, ce qui fait dire à certains que l'on bascule sur du NoSQL (littéralement : « au-delà du SQL »). Pour autant il est quasiment impossible de définir le NoSQL selon un corpus de règles puisqu'il représente précisément l'affranchissement par rapport à un ordre établi.
- **IN-MEMORY** : on parle de traitement in-memory pour évoquer les traitements qui sont effectués dans la mémoire vive de l'équipement informatique, plutôt que sur des serveurs externes. L'avantage du traitement in-memory est celui de la vitesse puisque les données sont immédiatement accessibles. En revanche, ces données ne sont pas stockées sur le long terme, ce qui peut poser des problèmes d'historisation.

## Tour d'horizon des briques fonctionnelles en lien avec le Big Data

Au-delà de ces outils de base, le Big Data se doit d'être applicatif et de proposer des services en lien avec les besoins directs de l'entreprise. Il est donc possible et recommandé de rajouter des briques fonctionnelles aux spécifications initiales d'Hadoop/MapReduce : c'est tout l'objet de l'écosystème Hadoop qui s'est développé en parallèle de ces outils.

Dans cet écosystème, on retrouve ainsi parmi les plus connues :

- **PIG** : également développé par Apache, il vise à augmenter les performances d'Hadoop/MapReduce en proposant un langage de programmation permettant un traitement plus rapide.
- **HIVE** : issu de la recherche Facebook, Hive permet d'effectuer des requêtes sur un cluster Hadoop directement en langage SQL.

A côté de ces briques complémentaires qui rendent Hadoop plus facile d'accès et plus rapide, on retrouve des structures de bases de données corollaires à Hadoop :

- **CASSANDRA** : également développé par Facebook, Cassandra est une

base de données orientée colonnes de type NoSQL. Elle supporte le traitement MapReduce et est particulièrement reconnue pour sa capacité à faciliter l'accessibilité des données, quel que soit le volume géré. Des compatibilités avec Hadoop sont aujourd'hui possibles.

- **HBASE** : de la même façon, HBase, développé par Apache, est un système de gestion de base de données distribué. Il est complémentaire d'Hadoop et particulièrement performant pour le traitement des très grandes tables (milliards de rangs X millions de colonnes).



Bien souvent ces différentes applications fonctionnent de manière packagée avec Hadoop : on parle donc davantage de « plateforme » Hadoop pour évoquer la structure constituée des différentes briques.

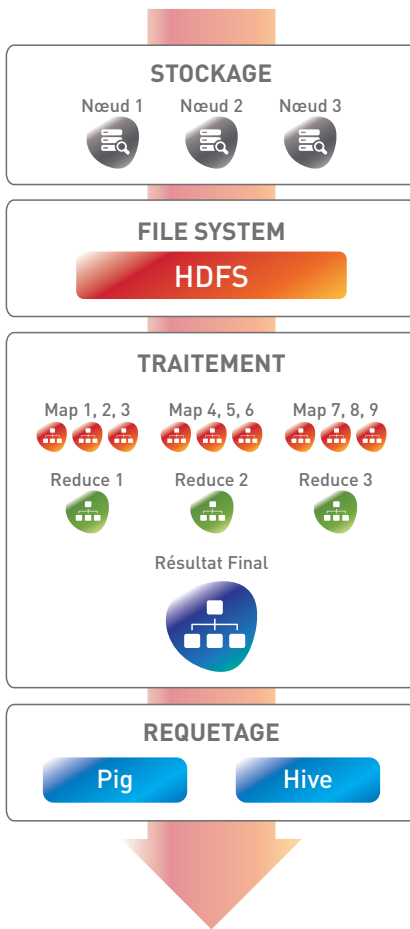


Schéma simplifié d'une architecture Hadoop/MapReduce

## Les technologies Big Data face aux besoins de la DSI

### Comment choisir les outils adéquats parmi les nombreuses offres disponibles ?

On le voit, les bases du Big Data sont assez simples du fait qu'elles soient limitées à un nombre restreint d'outils. Pourtant les offres sur le marché sont très nombreuses : les fournisseurs de technologies proposent en effet des offres variées appuyées sur Hadoop. L'avantage de ces solutions packagées et applicatives tient souvent dans la facilité d'utilisation : comme Hadoop est un framework complexe qui réclame des efforts de programmation pour son intégration dans l'architecture de l'entreprise, les fournisseurs proposent des solutions directement utilisables qui épargnent ces efforts à l'utilisateur. Une étude réalisée pour Infochimps a ainsi

montré que 61% des CIO préféreraient s'équiper de solutions extérieures plutôt que de développer leur propre infrastructure, et cela en raison du trop grand investissement en temps et en personnel représenté.

Le problème est qu'il est parfois difficile de reconnaître les atouts d'une solution par rapport à une autre : comment choisir entre une offre mettant en avant la rapidité de traitement par rapport à une autre qui se focalise sur les gros

« 61% des CIO préfèrent s'équiper de solutions extérieures plutôt que de développer leur propre infrastructure »

volumes de données ? Et quelle est la capacité d'adaptation de l'offre par rapport aux réels besoins de l'entreprise ? Cette problématique, évoquée très souvent par les DSI fait écho à la structuration très rapide du marché de l'offre par rapport à celui de la demande : tous les fournisseurs historiques de technologies se sont en effet lancés très tôt sur ce marché prometteur et ont ainsi proposé leur offre propre, ce qui multiplie d'autant le nombre de propositions commerciales faites aux utilisateurs.

### Quelle architecture mettre en place ?

Cette question renvoie aux modèles d'organisation privilégiés par l'entreprise : faut-il centraliser les données ou bien disposer d'architectures réparties au sein des directions métiers ? (cf *Les métiers du Big Data*). Ce qui est certain, c'est que, quel que soit le choix d'organisation retenu par l'entreprise, **l'intégration des outils Hadoop/MapReduce ne relève pas seulement d'un arbitrage technologique mais plus d'une stratégie liée à l'utilisation des données.** En fonction de cela, l'urbanisation des systèmes d'information peut profondément évoluer : accès aux données, gestion des entrées et des sorties, formats de données et métadonnées, etc.

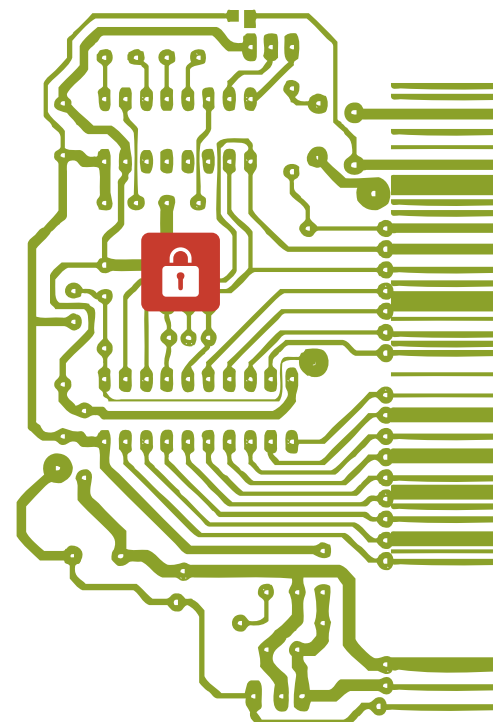
D'autre part, une des interrogations des DSI concerne l'acquisition progressive ou non des technologies Big Data : faut-il faire un « Big Bang » du Big Data par rapport à l'architecture existante, ou bien peut-on construire à partir de cette dernière ? Et faut-il nécessairement tout installer dès le premier jour ou peut-on rajouter progressivement des

applications ? Pour répondre à ces questions, **les fournisseurs ont eu à cœur de mettre en place des infrastructures plutôt légères qui s'adaptent à l'existant** ; la manifestation la plus évidente de ce phénomène étant la croissance du marché des « appliances », ces outils mixtes « hardware/software » qui se branchent sur le système existant pour une utilisation immédiate.

La question du stockage est également récurrente. Que choisir parmi les serveurs traditionnels, les appliances et les offres cloud ? **Certaines entreprises privilégient le cloud pour rendre l'infrastructure plus flexible et mutualiser les données dans le nuage.** Cependant, la migration n'est pas forcément adaptée à des gros volumes de données et les questions de sécurité restent prégnantes.

### La question de la sécurité

Evoquée concernant le Cloud, la question de la sécurité est devenue un leitmotiv avec l'explosion du Big Data : en effet, **la multiplication des données disponibles et leur valeur théorique sur le marché rendent encore plus présente la menace de cyberattaques.**



Avec le Cloud et l'apparition du BYOD, les possibilités d'infraction externes dans le système d'information de l'entreprise sont de plus en plus nombreuses.

Voilà pourquoi la segmentation et la restriction des accès aux données deviennent des éléments de régulation importants. Mais au-delà, les fournisseurs se penchent de plus en plus sur le perfectionnement d'outils de sécurité en lien avec le Big Data : et l'ironie de la manœuvre est que, pour cela, ils utilisent précisément... le Big Data ! En effet, l'intérêt du Big Data est de pouvoir détecter des anomalies, même les plus infimes, parmi les événements remontés. Dans les journaux de sécurité, le Big Data est donc un atout de taille pour prévenir les attaques et guérir les failles analysées ! Une étude réalisée par Varonis a ainsi montré que le Big Data permettait de détecter des données sensibles, d'identifier des activités malignes et de déterminer les utilisateurs dotés de droits d'accès excessifs... mais que seulement 54% des DSI le savaient !

La sécurité a été l'un des thèmes phare de 2013, au même titre que la qualité des données, l'objectif étant pour les deux de maîtriser le data deluge et de garantir l'intégrité de l'utilisation.

Les questions liées à l'implémentation du Big Data au sein de la DSI restent donc nombreuses car le phénomène est encore jeune, sans retours d'expérience de long terme, et les compétences sont encore balbutiantes en interne pour apprivoiser rapidement le phénomène. Pourtant, **81% des CIOs déclarent que le Big Data fait partie de leurs cinq principales priorités en 2013**, ce qui montre la volonté palpable des entreprises à s'équiper rapidement.

Du coup, les cabinets de conseil et d'intégration se sont positionnés rapidement pour apporter une expertise opérationnelle sur la mise en place de ces infrastructures. Des études montrent ainsi que les recrutements en SSII sur la thématique Big Data ont considérablement augmenté au cours de l'année écoulée.

# Interview

## Isabel Drost-Fromm

Co-Fondatrice / Contributrice, Apache Mahout. Membre

### « L'Open Source implique un grand nombre de défis »

#### 1) La fondation Apache est l'organisation qui a vu naître Hadoop... Pouvez-vous résumer pour nous les origines d'Hadoop ?

Si l'on se plonge dans les origines lointaines d'Hadoop, tout a commencé au début des années 2000 avec un projet appelé Apache Lucene : il s'agissait de mettre en place une bibliothèque numérique en open source, qui serait dédiée à la création d'index pour la recherche textuelle (exactement comme l'index que l'on trouve à la fin des livres et qui indique le numéro de la page contenant tel ou tel terme...sauf que Lucene indiquerait le document contenant ce terme).

Alors que Lucene avait simplement pour objet de répertorier des textes afin de faciliter les recherches, l'idée s'imposa peu à peu qu'il fallait créer un moteur de recherche de grande échelle, sur l'ensemble de la Toile. Pour cela, il fallait récupérer deux éléments : un outil d'indexation pour les nouvelles pages découvertes sur Internet en suivant des liens connus (web crawling), et un outil d'extraction de contenu (pour les pdf, le contenu html, les documents Microsoft Word...). La réunion de tous ces éléments a donné naissance au projet « nutch ».

Nutch s'est construit progressivement entre 2002 et 2004 ; comme il visait à fournir un moteur de recherche pour l'ensemble d'Internet, on comprit vite qu'il fallait recourir au traitement distribué sur plusieurs machines et non plus une seule. Heureusement, c'est à peu près au même moment que Google publiait son article sur le Google File System ainsi que sur le traitement MapReduce...

De tout cela: Lucene (l'index), Nutch (le "collecteur de données") et MapReduce (le système de traitement), toutes les conditions pour le Big Data étaient réunies!

En 2005, des premières expérimentations pour un système de fichiers distribués (« nutch distributed file system ») furent lancées au sein de nutch (celles qui mèneraient ensuite au HDFS, le système de fichiers à la base d'Hadoop). Lorsqu'Hadoop fut converti en un sous-projet de Lucene, indépendant de nutch, Yahoo! décida de l'adopter pour accroître ses capacités de cartographie du web et de contribuer au projet. C'est finalement en 2008 que le projet acquit un statut de priorité (« top-level ») menant à la création d'une communauté autonome pour le gérer !

De tout cela: Lucene (l'index),  
Nutch (le "collecteur de données")  
et MapReduce (le système de  
traitement), toutes les conditions  
pour le Big Data étaient réunies!

Capture d'écran de la page Issues d'Apache Hadoop

\*<https://issues.apache.org/jira/browse/HADOOP#selectedTab=com.atlassian.jira.plugin.system.project%3Aissues-panel>



## 2) Selon vous, quels ont été les bénéfices d'une approche open source pour la gestation d'Hadoop ? Et ses limites ?

Le grand atout de l'Open Source a été d'aider le projet à atteindre sa maturité : plus le projet attirait de contributeurs (comme Yahoo!, Facebook, LinkedIn et autres), plus il s'améliorait.

Mais bien sûr, il s'agit d'être prudent avec ce type d'approches car l'Open Source implique un grand nombre de défis : comme chacun apporte sa pierre à l'édifice, vous devez faire attention à ce que l'ensemble reste cohérent, à ce que le but initial soit toujours poursuivi et à ce que les contributeurs comprennent qu'ils travaillent sous l'enseigne Apache, et non pas pour leurs entreprises d'origine, dont les stratégies pourraient se trouver en conflit avec les développements du projet. Et également que les acteurs sont assez formés pour maintenir la bonne qualité de la structure. C'est un défi bien sûr, mais les résultats et les améliorations durables parlent d'elles-mêmes, je crois !

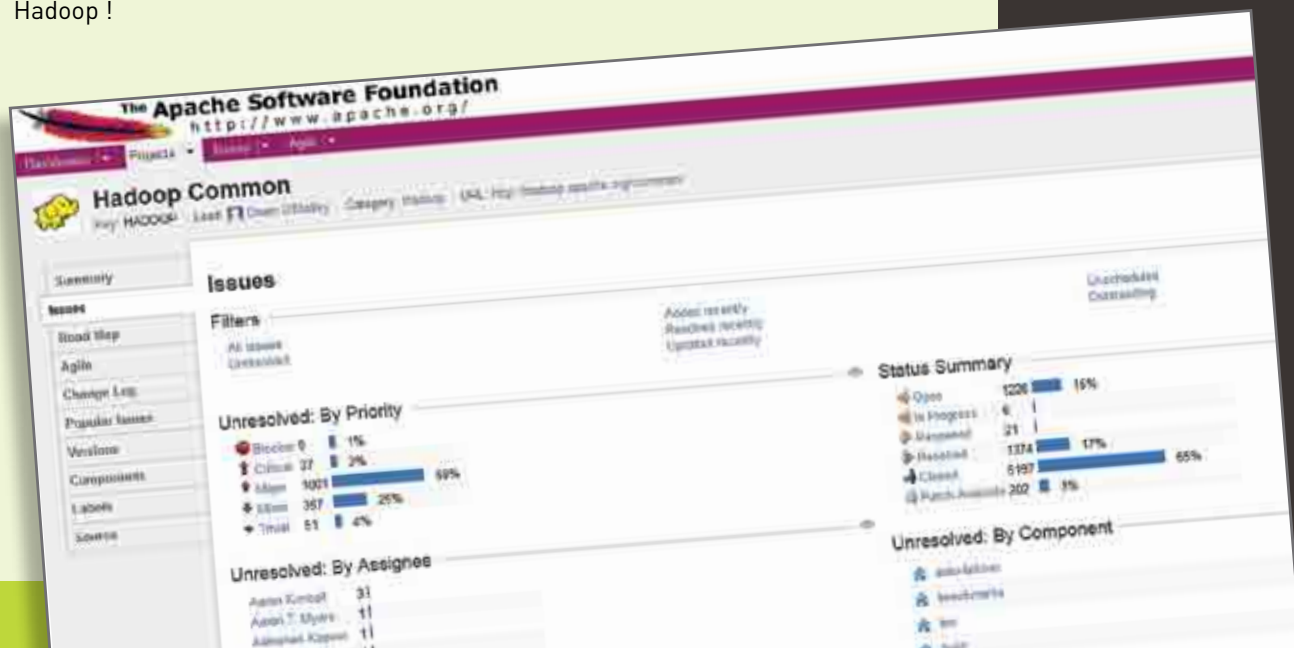
Si vous avez le temps, rendez-vous sur la page « Issues » d'Apache Hadoop\*, vous trouverez les chantiers actuels sur lesquels travaillent les contributeurs pour améliorer l'outil ; cela vous donnera une idée du dynamisme de la communauté Hadoop !

## 3) Justement, qu'est-ce qui pourrait être amélioré sur Apache Hadoop ? Quels sont les terrains de recherche actuels ?

Là encore, la meilleure réponse concernant les développements actuels et les terrains d'exploration se trouve sur cette page Issues. Comme le développement est complètement ouvert, les forces, les faiblesses et les bugs peuvent être découverts facilement, à la fois par les utilisateurs et les contributeurs.

Les principaux défis relevés sur d'autres projets de l'écosystème Hadoop sont ceux qui visent à rendre les requêtes plus rapides et plus faciles à réaliser. Peu de personnes les formulent en pur Java Map/Reduce ; la plupart utilisent d'autres langages de requêtes plus sophistiqués comme Pig, Hive, Cascading, Cascalog, Scalding, etc.

Concernant Hadoop en lui-même, **une voie intéressante d'exploration vise à rendre les clusters Hadoop plus économes en énergie**, en orientant les données les moins utilisées vers des disques qui peuvent être ralentis, permettant ainsi la réduction des coûts d'énergie... mais avec le risque de rallonger le temps d'analyse.



# Customisation marketing : Monoprix et dunnhumby

Monoprix et leur partenaire dunnhumby ont présenté un projet original d'utilisation marketing du Big Data : il s'agit d'une **expérience de personnalisation de mailings pour chaque client, avec des offres de discount adaptées** aux préférences de consommation de ce client.

Cette opération, qui est une déclinaison du programme de fidélité de Monoprix, vise à mettre en adéquation les offres de couponing proposées par les marques avec les profils de consommation clients : pour chaque client, un mailing unique est créé grâce à un algorithme d'allocations qui génère des combinaisons inédites d'offres promotionnelles. Les clients qui reçoivent ce courrier, mail ou papier, ont la possibilité de se rendre en magasin avec l'offre imprimée pour bénéficier de ces promotions personnalisées.

Au total, ce sont **33 millions de transactions de clients** Monoprix qui ont été analysées, ce qui représente **350 millions d'articles achetés**. Ces données sont donc essentiellement des données transactionnelles mais dunnhumby intègre également les données « préférence de média », à savoir le taux de retour des coupons papier pour établir une préférence média.

**L'allocation coupons/clients dure moins d'une heure.**

Lancée en février 2012 pour Monoprix, cette initiative (déjà expérimentée auprès de Tesco en Grande Bretagne et Kroger aux Etats-Unis) a fait la preuve de son efficacité : le taux de retour des coupons a été doublé par rapport à des coupons non personnalisés.

# Interview

## Hannah Whittall

Directrice Coms & Media, *dunnhumby France*

« Il faut que les équipes data et les équipes créatives travaillent main dans la main »

### 1) Pouvez-vous nous parler de la genèse de ce projet?

La communication personnalisée de Monoprix est fondée sur le même modèle que les expérimentations menées, entre autres, chez Kroger (aux Etats-Unis) et Tesco (en Angleterre), pour récompenser les clients de façon régulière avec des offres personnalisées calculées à partir de leurs achats. Nous avons démontré le gain généré par cette approche et avons été capables d'automatiser tout cela grâce à un outil propriétaire qui permet la personnalisation des communications à partir de millions de transactions traitées en quelques minutes. Monoprix est le premier distributeur français à bénéficier de cet outil et de cette approche, déjà mis en place chez d'autres distributeurs de par le monde, avec un succès certain à chaque fois.

### 2) Comment choisissez-vous les clients qui recevront ces offres personnalisées?

Nous contactons uniquement les clients qui sont « opt-in » (ceux qui choisissent de recevoir nos communications) puis nous effectuons une sélection à partir d'autres critères en lien avec les objectifs du programme. Une fois que les clients sont choisis, ils reçoivent donc leur mailing en lien avec leurs préférences. Nous suivons et analysons alors leur réponse à ce courrier, afin de pouvoir améliorer et optimiser en permanence le ciblage (qui reçoit la communication ?) et le canal (par quel canal devons-nous atteindre le client ?).







### 3) Quels sont les plans futurs pour ce programme? Avez-vous des améliorations à apporter?

Il y a des améliorations à apporter sur le canal de distribution et également sur le processus de retrait de l'offre (dématérialisation de l'offre sur mobile ou directement par mail, au lieu de l'impression papier). Nous travaillons là-dessus en lien étroit avec nos clients du retail en fonction des technologies dont ils disposent. Nous pourrions également songer à intégrer dans le traitement d'autres types de données clients : les historiques de navigation, les données sociales, la géolocalisation également afin de fournir du contenu et des offres plus précises, en temps réel, lorsque le client en manifeste le besoin ou les recherche activement.

### 4) Craignez-vous que ce nouveau type d'approche transforme le métier du marketer?... En termes de créativité par exemple?

Bien sûr, cela a un impact sur le travail et les compétences du marketer mais cela n'atténue en rien l'importance de la créativité. Le plus de données et d'informations nous réunissons à propos du client, le mieux nous pouvons informer et briefer les équipes créatives et contenu. Sur ce projet, nous traitons principalement d'offres commerciales, mais nous pouvons également personnaliser les contenus et les messages délivrés afin qu'ils soient encore plus pertinents. Comme nous sommes de plus en plus capables de le faire en temps réel via les médias digitaux, de nombreuses opportunités s'offrent à nous. On s'accorde de plus en plus sur le fait qu'il faut que les équipes data et les équipes créatives travaillent main dans la main. Une expérience 100% créative, même si elle est très bien réalisée, sera toujours difficile à quantifier dans ses résultats, mais elle vise toujours à comprendre qui est le client ; c'est là où un savant mélange d'art et de science peut ajouter énormément de valeur !

Plus que par ses technologies

innovantes ou son marché,

le Big Data suscite surtout la

curiosité pour ses nombreuses

applications en entreprise.

Celles-ci restent encore limitées

à des retours d'expérience

ponctuels mais elles sont en

pleine phase de généralisation.

Vous trouverez ici quelques

exemples inspirants, dans

plusieurs domaines d'activité.







### 3) Votre projet se poursuit et s'amplifie... Combien de personnes seront à terme impliquées dans ce nouveau process ? Comptez-vous apporter des évolutions qualitatives à la plateforme ?

Nous pensons doubler le nombre d'utilisateurs, mais en particulier nous nous appuyons de plus en plus sur ces données pour générer des rapports et des interactions dans d'autres outils et d'autres process : à travers la mise en place de web-services directement sur la base Orest, celui-ci devient le support de reportings et de nouveaux services en dehors du monde industriel et aussi pour nos clients.

DOE037C	2003/2013-10-24 16:10	TFP-NEG 1 (0934)	PF	ECO	29410-830-0-27-FF	EN COURS	42000
DOE037B	2003/2013-10-28 15:45	TFP-NEG 1 (0934)	PF	ECO	29410-830-0-27-FF	TERMINE	43770
DOE037C	2003/2013-10-24 16:10	TFP-NEG 1 (0934)	PF	ECO	29410-830-0-27-FF	TERMINE	43000Z/088

**+ Enveloppe**

Identifiant : DOE037C  
 Format : PF  
 Priorité : Non urgent  
 Etat : CREE

**+ Analyse adresse**

**- Adresse d'origine**  
 Origine : OCR  
 Code Adresse : 10117P00003102K/VOIE  
 Code Acte (Distr) : 43000Z/088  
 ID RAO : 4300022478  
 Type ID RAO : voie  
 Etat : FINALISE

**+ Traitements**

**- Traitements**  
 Date de traitement : 2003/2013-10-24 16:10  
 PIC : LEMPCES CLERIGHT AUVERGNE PIC  
 Machine de tri : TFP-NEG 1 (0934)  
 Phase : Départ  
 Programme de tri : 29410-830-0-27-FF  
 Réceptacle : T57

**Analyse de tri : filtre sur PF (Petit Format) / ECO**

Code	Description	Quantité	Pourcentage
001	...	1000	100%
002	...	500	50%
003	...	250	25%
004	...	125	12.5%
005	...	62.5	6.25%

# Recherche : Projet DEUS

Mené par l'Observatoire de Paris, avec l'aide de GENCI (Grand Equipement National de Calcul Intensif), le projet Deus lancé en 2011 visait à **modéliser pour la première fois la formation des structures cosmiques dans tout l'Univers observable, depuis le Big Bang jusqu'à nos jours.**

Pour cela, trois hypothèses avaient été retenues, correspondant à trois modèles d'énergie noire (composante énergétique de l'univers qui constitue un moteur de son expansion) : il fallait comprendre ainsi quel était l'impact de cette énergie sur l'univers et donc en déduire sa nature.

Pour cela, **un grand nombre de données représentant 550 milliards de particules de l'univers** (issues du télescope spatial WMAP) ont été injectées dans le supercalculateur CURIE de GENCI, installé au TGCC du CEA, qui présente une capacité de **traitement de 2 pétaflops par seconde (2 millions de milliards d'opérations par seconde)**. Il s'agissait d'opérer les traitements pour chacune des trois simulations en utilisant toutes les capacités de calculs de la machine (76000 cœurs de calcul sur 80000, 360 To de mémoire). **Au total, ce sont 150 pétaoctets de données brutes qui ont été générées**, ce qui posait une problématique de stockage : la solution a donc consisté à développer un outil de post-traitement qui, en parallèle du calcul, a permis de réduire à la volée ce volume de données générées, sans pénaliser les temps de simulation.

**Au final, ces traitements ont permis de réduire l'étude à 1,5 pétaoctets de données raffinées, grâce à un débit de stockage des données très important (50 Go/s) !**

Le résultat, disponible pour la communauté scientifique internationale, est une modélisation de la distribution de matière dans tout l'Univers observable, semblable à celle qui pourrait être observée par les télescopes : c'est donc un support exceptionnel d'étude pour les agences spatiales ESA et NASA, et ces modélisations permettront d'apporter des interprétations substantielles aux observations du satellite européen Planck.

**Outre la performance en cosmologie, cette expérience, unique à l'échelle mondiale, est également une prouesse dans l'univers du calcul haute performance** : elle ouvre le champ à d'autres applications du calcul intensif (HPC), comme par exemple pour le projet ITER ou les simulations climatiques, selon le même principe d'optimisation des entrées et sorties et de post-traitement à la volée des résultats.

# Interview

## Stéphane Requena

Responsable technique, *GENCI*

### « Big Data et HPC sont amenés à converger »

#### 1) Cette expérience est un modèle de performance technologique et d'optimisation des machines... au-delà de cet aspect « automatisé », quels sont les travaux d'analyse que vous avez réalisés ?

Il reste aux chercheurs de l'Observatoire de Paris un gros travail d'analyse des résultats à effectuer : en effet, même si les données raffinées ont un volume réduit par rapport aux données de sortie brutes, elles restent volumineuses et complexes. Aujourd'hui, nous avons un film global sur l'évolution de l'Univers mais si nous voulons analyser plus précisément la formation de certaines structures, il faut identifier les données pertinentes. Il y a donc un effort d'analyse à faire pour corréliser ces données de sortie, corroborer les résultats et effectuer de nouveaux calculs. Forcément, les équipes de recherche doivent être dimensionnées pour absorber ce volume massif de données, et c'est là un nouvel enjeu : du coup, on cherche à développer l'analyse automatique Big Data, c'est-à-dire utiliser des outils automatisés qui identifieront uniquement les données pertinentes pour les chercheurs.

#### 2) Suite à cette expérience, quels sont les prochains projets de ce type que l'on peut attendre ?

Il y a de nombreux domaines dans lesquels on peut envisager un couplage du HPC avec du Big Data, par exemple pour la combustion (dans les secteurs automobiles, aéronautiques et dans l'énergie) ; en effet, les calculs modélisant la combustion peuvent représenter des dizaines de téraoctets de données, qu'il est intéressant de traiter pour garantir une combustion homogène et stable. En effet, sans homogénéité et stabilité, les anomalies de combustion peuvent conduire à un mauvais rendement des moteurs avec à la clé des problèmes de surconsommation, de pollution et même d'explosion...d'où la nécessité de faire des simulations.

Autre secteur d'application : la climatologie où l'on souhaite modéliser l'évolution du climat avec des maillages très fins de la Terre, ce qui implique plusieurs modèles complexes de simulation que l'on lie entre eux. A terme, la communauté de la climatologie envisage de traiter des exabytes de données !

#### 3) Pour vous, quel est le lien entre HPC et Big Data? Parle-t-on de la même chose ? En quoi diffèrent-ils ?

Quand on parle de Big Data, on parle du traitement de grands volumes de données, c'est-à-dire une fouille dans des données et une interprétation des résultats qui peut, dans le cas de données complexes, recourir à du calcul. Le HPC quant à lui correspond à du calcul scientifique, que l'on applique à plusieurs



Le supercalculateur CURIE de GENCI installé au TGCC du CEA - © CADAM





domaines scientifiques ; c'est donc davantage l'utilisation de modèles mathématiques sur des superordinateurs. Forcément, ces deux domaines sont amenés à converger : si l'on prend l'exemple des données livrées par les nouveaux télescopes, nous avons besoin à la fois de calcul haute performance pour accélérer le traitement, et d'interprétation des données pour tirer de véritables informations. Le rythme de traitement est important lorsque l'on gère des gros volumes : si l'on optimise un calcul mais que le traitement des données est lent, l'analyse au global ne sert à rien.

#### 4) La France a été à la manœuvre sur ce projet... pouvez-vous nous expliquer les origines ?

La France dispose du supercalculateur CURIE qui était à l'époque une des trois machines capables dans le Monde de réaliser cette simulation DEUS (les autres se trouvant en Chine et aux USA). CURIE présente de grosses capacités à la fois pour le calcul et le stockage, en mémoire et sur les disques ; en 2012, elle est mise à disposition par GENCI à destination de la communauté scientifique européenne via l'infrastructure de recherche PRACE. Le projet DEUS émanait de l'Observatoire de Paris qui cherchait une machine pour pouvoir effectuer ses simulations donc nous avons pu la leur mettre à disposition.

Depuis d'autres machines ont été rendues également accessibles dans PRACE : en Allemagne, des chercheurs anglais sont actuellement en train de finaliser des simulations massives en climatologie, notamment sur la modélisation des cyclones dans le golfe du Mexique. Si elles aboutissent, leurs recherches auront permis grâce à PRACE d'avoir 3 ans d'avance sur les travaux en cours dans le domaine !



*Exemple de modélisation de l'univers*



*Stéphane Requena (à droite) et Jean-Michel Alimi de l'Observatoire de Paris, reçoivent le prix spécial du jury lors des Trophées de l'Innovation Big Data*

# Localisation : Géomarketing selon SFR

Menés depuis deux ans par une équipe dédiée au sein de l'entreprise, **les travaux de géomarketing de SFR ont visé à collecter et analyser les données de géolocalisation des utilisateurs réseaux**, afin d'en tirer des informations sur la fréquentation et les flux de populations sur certaines zones géographiques. Ainsi SFR a pu établir que le Festival de Cannes avait réuni 389 500 visiteurs dont 16 % d'étrangers !

Comment fonctionne cette infrastructure ? En réalité, SFR a mis en place un serveur de collecte développé par Intersec qui récupère, via les antennes relais, toutes les données relatives à des événements de consommation (ex : allumer son téléphone, appeler, etc.), ce qui correspond à **un total d'un milliard d'événements de mobilité par jour !**

Puis **ces données anonymisées sont croisées avec des données externes (de type Open Data)** au sein d'un outil logiciel développé en interne sur deux bases, une NoSQL et une autre Hadoop. Les données d'événements sont également corrigées statistiquement pour correspondre aux parts de marché locales de SFR.

**L'intérêt d'une telle démarche est de fournir une information inédite aux collectivités locales et au secteur du tourisme** : les analyses menées évitent en effet l'aspect déclaratif des questionnaires et les données sont disponibles en quasi temps réel.

Pour l'opérateur, c'est une opportunité unique de valoriser ses données en vendant des études aux entités intéressées. Mais au-delà, ces données ont également une utilisation interne puisqu'elles permettent à l'opérateur de suivre la consommation de ses abonnés et d'intervenir plus rapidement sur la fidélisation, la gestion des fraudes et le service client.

# Interview

## Mathieu Gras

Head of LBS & Big Data Business Development Manager,

**« Aujourd'hui, on voit beaucoup d'opérateurs internationaux qui se lancent dans l'aventure alors que nous étions les seuls en Europe à l'époque »**

### **1) Pouvez-vous revenir sur les origines de ce projet ? Comment s'est déroulé le développement de cette nouvelle activité ?**

C'est un projet qui a connu une phase de business development assez longue. En fait, nous avons rapidement constitué une équipe transversale et autonome, et nous souhaitions répondre à la fois à un besoin externe (connaître les flux de population) et à un besoin interne (avoir des indicateurs marketing temps réel). Nous sommes d'abord partis sur la réponse au besoin externe parce qu'il nous semblait le plus valorisable mais nous nous sommes rendus compte assez vite que nous étions trop en avance sur le marché et que, même si nos projets étaient bien reçus, ils étaient trop en rupture.

Du coup, nous nous sommes concentrés sur la partie interne et nous avons constaté que les équipes adhéraient rapidement à nos projets car la donnée était immédiatement valorisée !

C'est par la suite que les activités externes se sont vraiment développées, car le marché était prêt et l'engouement autour du Big Data jouait sûrement en notre faveur ! Aujourd'hui, on voit beaucoup d'opérateurs internationaux qui se lancent à leur tour dans l'aventure alors que nous étions les seuls en Europe à l'époque.

### **2) Pour vous, y'a-t-il eu un basculement technologique ou cette activité aurait-elle pu être créée il y a cinq ans ?**

Très honnêtement, je pense que l'on avait les ressources technologiques : l'opérateur fait du temps réel depuis des années. Mais le traitement aurait forcément nécessité plus de ressources, plus de temps et donc plus d'investissement : le prix d'un serveur était bien supérieur il y a cinq ans !

De plus, d'autres sources de données sont à présent disponibles, ce qui rend les analyses plus complexes et donc plus intéressantes.

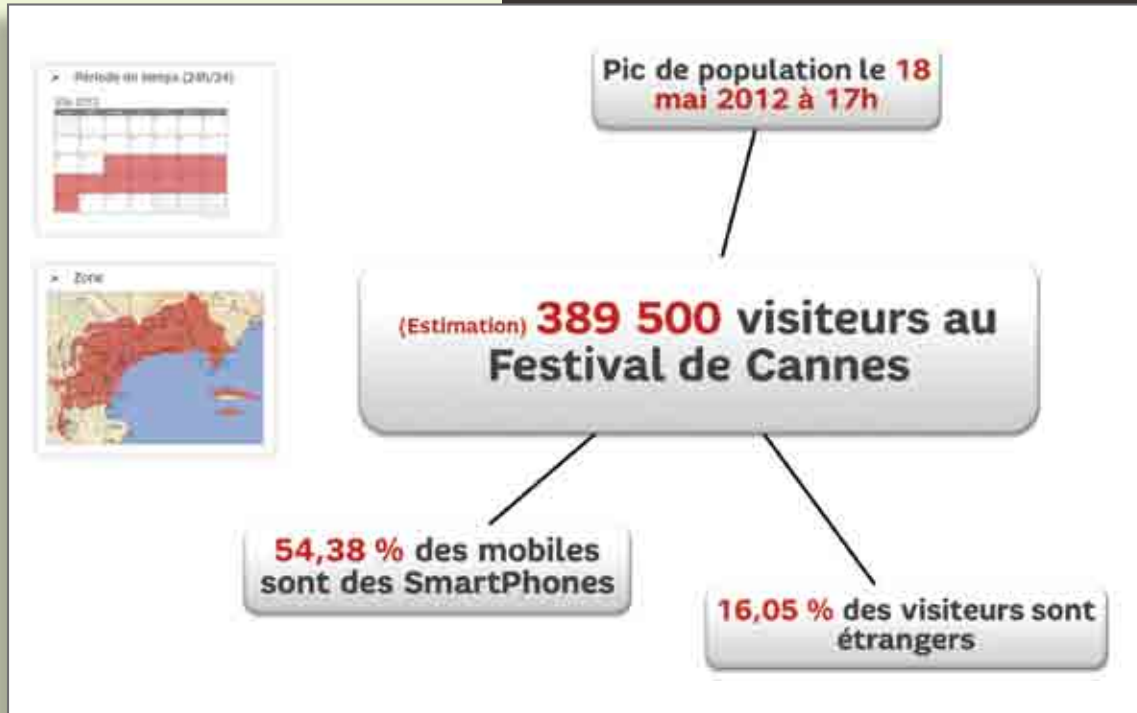
Donc, oui technologiquement, nous aurions pu le faire, mais cela n'aurait pas eu le même intérêt économique et le marché n'aurait certainement pas répondu de la même façon !

### **3) Y'a-t-il certains projets récents que vous avez réalisés avec cette plateforme de géomarketing ?**

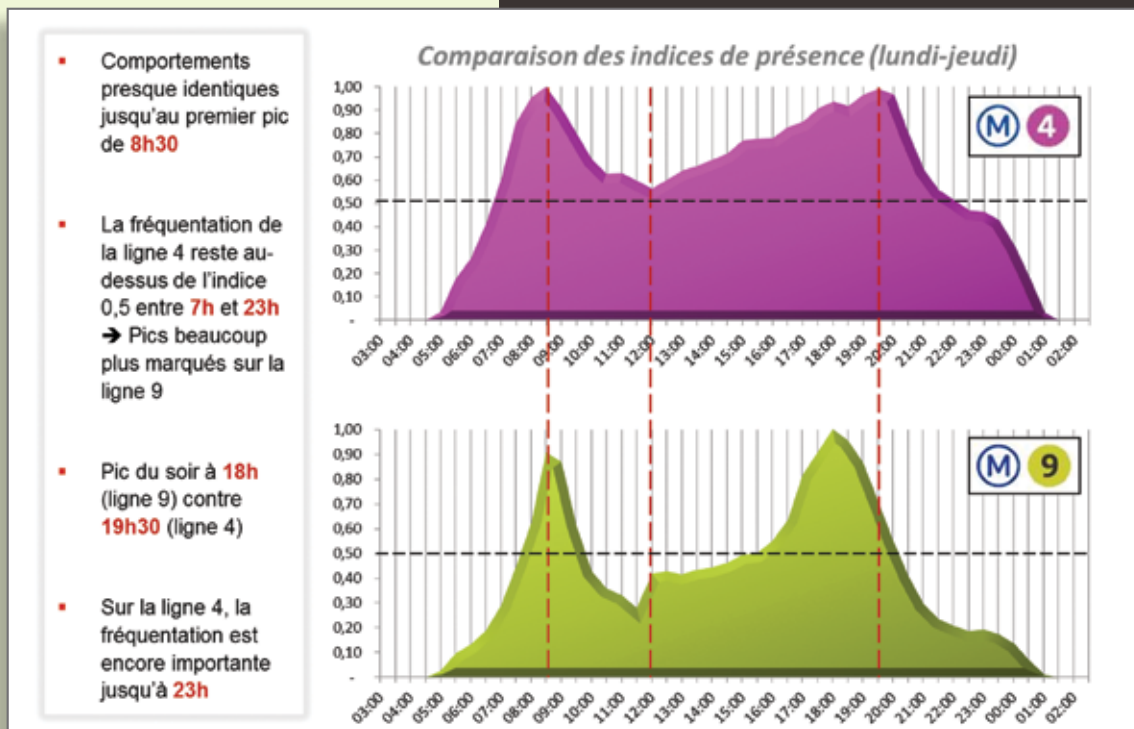
Oui, nous avons travaillé en juillet dernier sur le Tour de France : l'objectif était d'observer l'affluence des populations sur le Tour et de déterminer l'impact de celui-ci sur le tourisme local (restauration, nombre de nuits d'hôtels, etc.). En effet, les municipalités investissent de fortes sommes pour obtenir le passage du Tour sur leur territoire et le retour sur investissement en terme touristique est forcément une information qui les intéresse !



SFR



Etude de fréquentation du festival de Cannes



Etude de fréquentation du Métro parisien

# Agrégation de Données : Projet Ellicible de Coface Services

Coface Services (Ellisphere au 1<sup>er</sup> janvier 2014) est l'entité de référence concernant l'information d'entreprises en France : elle propose en effet une base de données unique sur les entreprises françaises, qui repose sur de nombreuses sources publiques et privées.

Afin de développer son offre de marketing direct et afin de la mettre en adéquation avec les besoins de « real-time marketing », Coface Services a choisi d'évoluer vers une plateforme plus réactive qui permet au client de bénéficier d'un service haut de gamme combinant de multiples critères pour un résultat immédiat : **ce service est adapté aux besoins du client dans la mesure où l'organisation de l'information (ciblage, segmentation, surveillance de groupes etc.)** lui revient de façon autonome, contrairement aux segmentations fonctionnelles observées dans les bases de données traditionnelles. En effet, un effort important en matière de finesse, de granularité et de résolution a été réalisé au niveau de la segmentation.

La plateforme Ellicible a ainsi vu le jour le 20 mars 2013 : **elle est appuyée sur des technologies New SQL (et non Hadoop/NoSQL) intégrant le traitement massivement parallèle** et développées par la société Parstream. Ainsi des milliers de requêtes à très fortes volumétries (plusieurs milliards de données) peuvent être exécutées en même temps, avec un très faible temps de traitement (quelques millisecondes). La grande avancée de cette plateforme tient dans le **raccourcissement des temps d'affichage des résultats** : ils sont cette fois-ci disponibles en temps réel.

Enfin cette plateforme est évolutive car elle intègre de nouvelles sources partenaires de grand volume, en France et à l'international.

Pour les clients, Ellicible crée ainsi de nouveaux usages dans leur recherche de prospects et d'informations sur les entreprises : ciblage facilité, segmentation et organisation de l'information personnalisées, obtention immédiate des résultats.

# Interview

## Jean-Daniel Ruegger

Directeur de domaine au sein de la DSI, *Coface Services*

« Nous observons que l'usage de l'application parcourt l'ensemble de ses fonctionnalités »

**1) Cette plateforme combine volume et performance de traitement, mais elle agglomère surtout de multiples sources de données : pouvez-vous nous parler du processus d'agrégation de ces données distribuées ?**

En effet, hormis le volume et la vitesse qui sont nos contraintes techniques, nous avons également la contrainte fonctionnelle de la variété des informations que nous assemblons. Aujourd'hui nous collectons nos données d'une vingtaine de sources différentes et nous produisons aussi nos propres données en interne. Les sources sont différentes mais leurs cycles de renouvellement aussi. Nous avons donc un processus ETL relativement classique pour chaque filière, l'originalité résidant dans l'assemblage final et sa mise à disposition.

**2) Vous avez opté pour une technologie SQL et non pour des plateformes de type Hadoop. Pour quelle raison ? Pensez-vous évoluer vers ce type de technologie à l'avenir ?**

Coface Services a une longue histoire et toute une série de « premières » dans le traitement de l'information. Sans remonter jusqu'aux racines de l'entreprise avec la création de Bottin en 1796, l'activité « Solutions Data » a innové avec la télématique dès le début des années 80, et les traitements massivement parallèles au début des années 90. Nous savions que nous voulions une technologie non-intrusive tant matériellement que du point de vue des compétences nécessaires, donc nous cherchions une solution New SQL à traitement parallèle. Par ailleurs Hadoop – orienté batch – n'était pas envisageable.

**3) Quels sont les retours que vous avez eus sur cette plateforme ? Les clients se sont-ils appropriés ces nouveaux usages ?**

Les retours que nous avons eus sont nombreux, en interne aussi car nous utilisons la technologie pour des opérations de segmentations clé-en-main, vraiment « as a service ». Nous mesurons le degré d'appropriation par deux signes forts : d'une part, nous observons que l'usage de l'application parcourt l'ensemble de ses fonctionnalités, avec des retours très positifs ; et d'autre part, nos clients (qui vont de la PME à l'enseigne nationale ou internationale) nous demandent et suggèrent des extensions très innovantes, ce qui montre leur intérêt !



# Les métiers du Big Data

*Que change le Big Data aux compétences en place ? Comment s'y adapter ? Et quels sont les modes d'organisation à privilégier pour capter toute la valeur des données ? Voici un petit tour d'horizon de l'entreprise en format Big Data.*

## Quelles sont les nouvelles compétences liées au Big Data ?

En modifiant peu à peu les paradigmes de l'analyse de données, le Big Data a créé de nouveaux besoins : en effet, les compétences informatiques sont mises à l'épreuve par la complexité des supports de base que sont Hadoop, MapReduce, NoSQL, etc. Même en faisant appel à des fournisseurs et des sociétés de service, la nécessité de disposer en interne de compétences techniques assez poussées se fait de plus en plus prégnante : si le Big Data doit être déployé de façon généralisée dans plusieurs services de l'entreprise, il est indispensable d'avoir à disposition des équipes pouvant connecter rapidement ces nouvelles technologies avec les besoins business.

**Autre impératif réclamé par le Big Data : la possibilité d'analyser les données traitées avec une vision métier.** En effet, c'est dans la partie analyse que le Big Data révèle bien souvent ses qualités stratégiques, ce qui nécessite forcément d'exploiter les données selon des objectifs business propres à chaque département. En marketing par exemple, cette approche est de plus en plus observée : les marketers s'orientent de plus en plus vers l'intégration de compétences « data » dans leurs équipes, ou bien ils se forment eux-mêmes à ce nouvel impératif.

Enfin, le Big Data porte en lui de nombreuses promesses liées à l'évolution des algorithmes et des statistiques : pour que les données puissent être vraiment exploitées à des fins prospectives, les modèles statistiques sont amenés à évoluer, ce qui pousse nécessairement les entreprises à envisager **l'embauche de statisticiens et diplômés en mathématiques.**

« Informatique, statistique, business » est donc le triptyque des compétences

# Informatique Statistique

## Business

liées au Big Data. On pourrait également rajouter celles liées à la communication car la prise de décision appuyée sur les données repose bien souvent sur la capacité de représentation et de hiérarchisation de l'information, telle qu'on peut l'observer dans la datavisualisation. Dès lors, ces compétences réunies sont-elles à disposition pour l'entreprise ?

## Les nouveaux métiers du Big Data

La question des compétences est une problématique souvent évoquée lorsqu'on parle de la mise en place du Big Data en entreprise : en effet, selon Gartner, seules 31% des entreprises estiment pouvoir lancer des projets Big Data en faisant appel à leur personnel existant.

Elles recherchent en effet des profils capables d'agglomérer les trois qualités

: informatique (capacités de programmation sur les nouveaux outils), statistiques (capacité d'innovation et de modélisation) et business (capacité à interpréter les indicateurs et à les transformer en langage opérationnel). Ces profils polyvalents correspondent à un nouveau type de métier appelé Data Scientist, pour lequel les postulants sont encore rares :

**McKinsey estimait ainsi que la demande de Data Scientists pourrait atteindre 1,5 millions de postes dans le monde en cinq ans !**

Par conséquent, les offres d'emploi liées à ce métier se multiplient depuis 2010 (*cf interview de Jacques Froissant*) et les rémunérations se montrent attractives mais les efforts de formation restent encore le nerf de la guerre et les profils expérimentés manquent à l'appel.

En attendant de trouver la perle rare, les entreprises patientent en faisant monter en compétences leurs propres équipes,

« La nécessité de disposer en interne de compétences techniques assez poussées se fait de plus en plus prégnante »





déclaraient avoir augmenté et enrichi leur offre de formation en business intelligence.

Au-delà, ce sont les entreprises des fournisseurs qui prennent le relais en mettant en place des séminaires de formation et des initiatives de type « université en ligne ».

## L'impact du Big data sur l'organisation de l'entreprise

Au-delà de la mise à disposition d'outils et de ressources humaines et financières, le Big Data soulève logiquement une problématique d'organisation au sein de la structure qui choisit de l'implémenter : en effet, puisque le Big Data est censé prendre en compte des données réparties dans plusieurs services de l'entreprise et mettre en place des processus décisionnels transverses, il implique forcément une approche globale mettant la donnée au centre des attentions.

notamment sur des métiers comme le data mining qui présentent déjà certaines caractéristiques proches de la data science.

## Les offres de formation en France

Afin de pallier le déficit de profils « data scientist » disponibles, les organismes de formation et les établissements universitaires ont lancé des programmes intégralement dédiés au Big Data, ou envisagent de le faire : c'est le cas de Télécom ParisTech ou l'ENSAI mais également d'organismes privés qui font évoluer leurs offres de formation sur le segment décisionnel. Ces dernières sont bien souvent sectorielles ou orientées Hadoop mais elles ont l'avantage de permettre aux professionnels en poste de se mettre à niveau.

Quant aux parcours universitaires, ils répondent généralement à la dichotomie « informatique vs statistique » : il est très rare en effet de trouver des masters réunissant ces deux domaines de formation (seul le Master spécialisé Big Data de Télécom ParisTech, 1er du genre en Europe, le propose actuellement). Du coup, ce sont plutôt les écoles et DUT de statistiques qui sont à la manœuvre : on recense ainsi 70 spécialités de master en statistiques à l'université mais aucun n'est pour l'instant dédié au Big Data.

Selon l'étude State of Business Intelligence Survey réalisée en 2013 pour Teradata, 41% des professeurs d'université interrogés

« Selon Gartner, seules 31% des entreprises estiment pouvoir lancer des projets Big Data en faisant appel à leur personnel existant »

### La nécessaire communication entre les métiers

Plusieurs départements sont ainsi concernés : les métiers IT, les métiers de la statistique et de la BI, et bien sûr les métiers opérationnels, ceux qui disposent des données et qui cherchent à les mettre à profit.

Pour toutes ces entités, le Big Data bouleverse l'ordre établi dans la mesure où il invite à davantage de synergies entre les métiers afin de faciliter le traitement et la communication des données : les services supports doivent en effet travailler main dans la main avec les départements business pour éviter un aller-retour chronophage de l'information ainsi que des lourdeurs technologiques. Les équipes business sont donc de plus en plus impliquées dans le traitement et l'analyse de leurs propres données (d'où l'importance de leur montée en compétences) ce qui modifie également le rapport à la décision : en effet, par leur rapport direct et immédiat aux données, les agents opérationnels disposent d'un avantage comparatif indispensable qui leur confère une position particulière dans la « data-centric organization ». Contrairement à une approche top-to-bottom, l'entreprise en mode Big Data fait ainsi de la remontée des informations une clé de sa croissance.

Quant aux départements IT, leur rôle sera particulièrement mis en avant lors de l'implémentation des outils et lors des premières expérimentations avec les équipes métiers ; au fur et à mesure de la courbe d'apprentissage, ils pourront peu à peu se recentrer sur une mission de support et de développement, laissant aux équipes d'analystes le soin de récupérer les données.

Enfin les profils BI et statistiques pourront être soumis à plusieurs types d'organisa-

## Panorama des enseignements dispensés au sein du Master «Gestion et Analyse des Données Massives» de Telecom Paristech

- **L'écosystème Big Data**
- **Bases de Données et Bases de Données Avancées**
- **Données personnelles**
- **SI Décisionnels et Gestion de la Relation Client**
- **Ingénierie du web**
- **Transport et routage dans l'internet**
- **Machine-Learning**
- **Visualisation**
- **Apprentissage Statistique « en ligne » et distribué**
- **Systèmes Répartis**
- **Sécurité des systèmes d'information**

(source : site internet du Master)

tion en fonction de leur proximité avec les métiers : dans le cas d'une centralisation de leurs compétences, ils seront amenés à récupérer les données issues des services opérationnels afin de les traiter en lien avec ceux-ci. Mais on peut également imaginer une intégration de ces profils directement au sein des équipes métiers afin de raccourcir le cycle de décision.

### Plusieurs types de « data-driven organization »

En écho à cette pluralité d'organisations pour les métiers décisionnels, on remarque que les entreprises pionnières ont adopté plusieurs schémas mettant en avant leurs données.

- **L'option « décentralisée »**, à savoir la mise en place du Big Data « métier par métier » ou « service par service » (souvent au fur et à mesure, en fonction de l'importance relative des données pour les différents départements) : dans ce schéma, le département métier (ex : le marketing) intègre les outils en son sein avec l'appui de la DSI. Les analystes métiers sont mis à profit au cœur de la division pour l'exploitation des données, ce qui réclame une montée en compétences (*cf + haut*). Dans le cas où l'entreprise dispose d'un département BI dédié, celui-ci est amené à servir de support pour les projets menés, ou peut être éclaté dans les différentes directions opérationnelles.

- **L'option « centralisée »**, à savoir la constitution d'une direction dédiée au Big Data au cœur de l'entreprise, afin de centraliser toutes les données et d'effectuer le traitement et l'exploitation pour le compte des entités opérationnelles. Bien sûr, ce type de structure fonctionne également en synergie avec d'autres services (les métiers bien sûr, mais également l'IT) mais elle présente le grand avantage de réunir en autonomie des professionnels du Big Data qui pourront ainsi impulser le mouvement « data-centric » et proposer des projets transversaux (ex : mutualiser les données de plusieurs services).

- Enfin, l'option la plus utilisée actuellement est certainement **l'option de la sous-traitance** : confier à une entreprise spécialisée dans le Big Data la gestion des données et les analyses conjointes. Cette option qui a le mérite de la souplesse et qui permet d'enrichir l'exploitation des données par une expertise extérieure (ainsi que par des données externes) pose toutefois la

« La question de l'organisation-type n'a pas encore été tranchée »

question de la rapidité du cycle de décision, les allers-retours avec les équipes opérationnelles étant moins intégrés. On le voit, la question de l'organisation-type n'a pas encore été tranchée, essentiellement parce que les entreprises n'ont pas toutes les mêmes besoins et n'utilisent pas les données de la même façon... C'est la raison pour laquelle **une réflexion stratégique doit précéder tout achat d'outils Big Data** ou toute mise en place hâtive car les entreprises doivent avant tout faire l'inventaire de leurs besoins et de leurs ressources. C'est d'ailleurs une des principales préconisations des cabinets de conseil qui agissent en ce sens auprès des entreprises pour dessiner des organisations censées réagir efficacement et de façon opérationnelle à la problématique du Big Data.

### Le Chief Data Officer

La réorganisation de l'entreprise autour des données pose nécessairement la question de **la structuration de la prise de décision** et des remontées hiérarchiques auprès des instances dirigeantes : en effet, l'utilisation de la donnée, pour être efficace, doit s'inviter jusqu'au Comité de Direction des entreprises !

Mais pour cela, il manque encore actuellement des porte-parole identifiables dont la fonction serait d'orienter les décisions globales de l'entreprise en fonction des données traitées. Ce déficit de personne ou plutôt de fonction a vite été pallié par la création d'un nouveau personnage au cœur de l'entreprise : **celui de Chief Data Officer**.

Au même titre que le **CTO** ou le **CIO**, le **CDO serait donc un responsable exécutif pour les données de l'entreprise, qui s'efforceraient d'appuyer les décisions du CODIR sur les données**. Il chapeauterait donc toute l'organisation dédiée aux données, qu'elle soit centralisée ou non. Pour l'instant, ce rôle est resté à l'état

d'intention en France : on ne recense pas encore de CDO français, et, actuellement, ce sont plutôt les DSI qui assument le relais auprès de la direction. Mais, aux Etats-Unis, les exemples se multiplient, essentiellement dans l'administration (villes de Chicago, San Francisco, NYC, US Army, Federal Reserve, etc.), ce qui a conduit un cabinet de conseil, Russell Reynolds Associates, à prédire **la création d'un tel poste pour la moitié des firmes du classement Fortune 500 d'ici à 2015...**



Mark Headd,  
CDO de  
la ville de  
Philadelphie

« L'utilisation de la donnée, pour être efficace, doit s'inviter jusqu'au Comité de Direction des entreprises »

# Interview

Jacques Froissant

Président, *Altaïde*



## « Les offres d'emploi liées au Big Data ont doublé dans les six derniers mois »

### 1) On parle beaucoup de la fonction de « data scientist » en lien avec l'émergence du Big Data. Comment définiriez-vous ce nouveau poste ?

Pour moi, il n'y a pas une mais deux fonctions autour de la data : 1) une fonction technique, scientifique, appelée « data scientist » qui correspond à une culture ingénieur statistique et informatique, et qui s'occupe de récupérer les données, de les structurer et de les produire, 2) puis une fonction davantage business appelée « data analyst » qui exploite ces données afin d'aboutir à des préconisations business. Généralement, ce deuxième profil provient d'école de management ou de formations en marketing. On le retrouve beaucoup dans le secteur du e-commerce où les budgets et préconisations en matière d'acquisition client ont une importance stratégique.

### 2) Les deux fonctions sont-elles complémentaires ? Faut-il embaucher les deux ?

Cela dépend beaucoup de la taille de l'entreprise : souvent les petites entreprises vont d'abord se doter d'un data analyst, car la problématique web correspond à des enjeux business immédiats. Mais dès que l'on monte en taille, on rencontre de plus en plus de data scientist. Dans les grandes entreprises, dans le retail ou en banque, ils sont de plus en plus présents. La demande a beaucoup augmenté pour ces postes-là : si vous regardez les offres d'emploi liées au Big Data sur les moteurs de recherche, vous constaterez qu'elles ont doublé dans les 6 derniers mois.

### 3) Justement, pensez-vous qu'il y ait un décalage entre la demande et l'offre ? Y'a-t-il un manque de profils disponibles pour répondre aux besoins des entreprises ? Et comment y remédier ?

Il est effectivement difficile de trouver les deux types de profils, car la demande pour ces deux fonctions a explosé très

rapidement. C'est sûrement pour le data scientist que le décalage est le plus fort... ils sont très sollicités.

Cependant, en France, on recense 17 formations professionnelles et universitaires qui proposent des programmes en lien avec ces fonctions. J'ai confiance dans le fait que ces efforts porteront leurs fruits rapidement : la France est plutôt bien armée pour combler le fossé entre offre et demande sur ces compétences-là.

### 4) Autre sujet actuel : on parle de plus en plus d'application Big Data dans les RH (recrutement, gestion de la paie...). En connaissez-vous quelques-unes ?

Effectivement, il y a de plus en plus d'initiatives RH liées au Big Data. Par exemple, sur le recrutement, l'entreprise américaine Evolv a décidé d'exploiter les données d'une entreprise de plus de 20 000 employés pour tenter de mettre en avant des facteurs de performance. Ils se sont ainsi rendus compte que : 1) il n'y avait aucune corrélation entre le CV et la performance, 2) les personnes actives sur les réseaux sociaux étaient souvent les plus performantes et les plus fidèles à l'entreprise... assez surprenant non ?

Linkedin est également assez en avance sur ces questions-là : ils travaillent en permanence sur les données de leurs membres et permettent aux entreprises de cibler beaucoup plus finement les candidats et prospects qui leur semblent correspondre à leur activité.

### 5) Pensez-vous que les entreprises seront effectivement amenées à généraliser ce type d'approche dans les années à venir ?

Oui, je pense qu'on y viendra, notamment pour le management... On peut très bien imaginer qu'un DRH analyse les données issues des échanges de ses employés (sur les réseaux sociaux, sur internet ou en interne) pour mesurer l'ambiance, comprendre les attentes, les inquiétudes et être capable d'intervenir en amont. Ce sera beaucoup plus efficace que des enquêtes annuelles de satisfaction !

Mais les RH restent encore méfiants par rapport à ce type de technologies : c'est d'ailleurs souvent la dernière fonction à acquérir et intégrer les outils les plus récents. Après, on peut comprendre les inquiétudes si l'on se réfère à certaines applications : le grand risque du Big Data appliqué à la RH, c'est qu'un profil type de performance soit défini à partir de ces données, et que l'entreprise n'embauche plus qu'en fonction de ce modèle. Ce serait catastrophique pour la diversité et la capacité d'innovation de l'entreprise !

« La France est plutôt bien armée pour combler le fossé entre offre et demande sur ces compétences-là. »



Capture d'écran du site d'Evolv

# Les pouvoirs publics face au Big Data

*Par son ampleur et par ses nombreuses promesses, le Big Data a rapidement attiré l'attention des pouvoirs publics : ils y ont vu un potentiel de croissance important pour la filière numérique en France mais également des menaces éthiques fortes sur le citoyen, concernant la confidentialité des données notamment.*

*Cependant, bien avant ces problématiques, l'univers public était présent dans l'écosystème Big Data à travers un phénomène antérieur qui connaît aujourd'hui un développement corollaire au Big Data : l'Open Data.*

## L'Open Data

L'Open Data est un mouvement amorcé au milieu des années 2000 visant à transmettre au grand public les données numériques détenues par les administrations publiques, dans un but de transparence et d'efficacité de l'action publique.

En effet, de nombreuses informations démographiques, économiques, sociales pouvaient revêtir un caractère stratégique pour les organisations privées et le grand public (ex : régularité des transports publics, temps d'attente dans les services publics, niveau de rémunération dans une zone géographique donnée, affluence des zones touristiques, etc.) et les rendre accessibles, libres de droits, constituait donc **une avancée économique autant que citoyenne** (l'institution publique se trouvant ainsi engagée vis-à-vis de ses administrés à corriger les imperfections soulevées par ces données). Développé dans les administrations européennes et nord-américaines, **ce mouvement a désormais gagné les organisations privées** qui peuvent choisir de mettre à disposition du public et de partenaires les données qu'elles tirent de leur activité et qui concernent la collectivité.

A l'heure du Big Data où les données prennent une importance centrale dans le décisionnel de l'entreprise, ces informations véhiculées par l'Open Data et croisées avec les données internes, constituent donc un nouvel outil précieux pour nourrir la vision stratégique de

l'entreprise. En 2006, un rapport financé par la Commission Européenne établissait ainsi que le marché lié à l'utilisation de ces données Open Data représentait 27 milliards d'euros. Alliées aux technologies mobiles, ces données sont également un vecteur d'innovation important, permettant la **création de nouveaux services** (ex : conseils géolocalisés dans le secteur du tourisme, renseignements voyageurs dans le domaine des transports, etc.).

L'Open data s'est donc imposé en quelques années comme un secteur en croissance, attirant de nombreuses start-ups dont la mission est de traiter, trier et analyser ces données à destination des entreprises et des collectivités. En France, c'est le portail [data.gouv.fr](http://data.gouv.fr), chapeauté par Etalab, qui met à disposition du public les données du gouvernement sous l'autorité du Premier Ministre.



## Soutenir le décollage du marché : les investissements publics dans le domaine du Big Data

Le dynamisme porté par le Big Data et l'éclosion de nombreuses start ups en lien avec ce sujet ont rapidement inté-

**27 milliards d'€**

→ C'est le potentiel de marché de l'Open Data

(chiffre de 2006)

# Les chiffres du financement public pour le Big Data

- USA  
200 millions de \$
- Union Européenne  
81 millions d'€
- France  
11,5 millions d'€

ressé les pouvoirs publics qui ont proposé des **mesures de financement et d'accompagnement pour soutenir le décollage du marché**. Aux Etats-Unis, le programme « **The Big Data Research and Development initiative** » avait ainsi été lancé au printemps 2012 pour soutenir les programmes de recherche sur les technologies de traitement de grande échelle : 200 millions de dollars ont été mis à disposition des grandes agences fédérales dans la Santé, l'Energie ou la Défense pour trouver des applications sectorielles immédiates. En Europe, la Commission Européenne s'est également emparée du sujet en lançant des appels à projets : le premier début 2012 concernant le « **Management de l'information intelligente** » (50 millions d'euros), le second en 2013 intitulé « **Scalable data analytics** » (31 millions d'euros). Elle fait du Big Data un de ses principaux sujets d'étude : un programme de travail sur le Big Data et la Smart Data est ainsi prévu pour 2014 et 2015.

La France, quant à elle, a commencé à s'intéresser au sujet en mars 2012 avec le lancement d'un premier appel à projets et l'inscription du **Big Data dans le programme des Investissements d'Avenir** au sein de la section consacrée à



l'économie numérique. Cet appel à projets a permis de sélectionner 7 projets pour un total de 11,5 millions d'euros attribués au Big Data en avril 2013, et 4 projets restent encore à l'étude. Pour l'heure, la ministre chargée de l'Economie Numérique, Fleur Pellerin, a annoncé en janvier 2013 la création d'une filière Big Data, capable d'attirer les investissements étrangers, suivie de l'annonce d'un plan de soutien en juin 2013. Ce plan visera à construire un écosystème du Big Data, à fournir la formation nécessaire aux jeunes data scientists et à promouvoir les start ups prometteuses de la filière. Ce plan, salué par l'AFDEL (l'association française des éditeurs de logiciels), s'inscrit dans le projet plus global de création d'une Silicon Valley à la française, appelé Paris Capitale Numérique.

« Fleur Pellerin a annoncé en janvier 2013 la création d'une filière Big Data capable d'attirer les investissements étrangers »



La ministre chargée de l'économie numérique, Fleur Pellerin

## La problématique de confidentialité des données

L'une des questions centrales soulevées par le Big Data est probablement celle de la protection des données personnelles : en effet, l'utilisation massive de données à des fins d'optimisation ou de ciblage marketing renforce la crainte déjà prégnante d'une surveillance et d'une manipulation automatisées du citoyen : géolocalisation, suivi des historiques sur Internet, personnalisation non consentie des offres, etc. sont autant d'incarnations du phénomène Big Data qui peuvent être perçues comme des menaces pour la vie privée des consommateurs finaux. Cette inquiétude, dont la récente affaire sur le programme américain PRISM s'est fait l'écho, pourrait également s'avérer néfaste économiquement puisqu'une rupture de confiance du consommateur vis-à-vis du fournisseur entraînerait automatiquement un retour en arrière dans le partage des données et donc, à terme, un recul du Big Data et de ses promesses.

Pour toutes ces raisons, les autorités de régulation se sont penchées sur la question, reprenant leurs législations en matière de protection des données personnelles afin de les adapter au Big Data : en France, c'est la loi Informatique et Libertés du 6 janvier 1978, actualisée plusieurs fois dans les années 2000, qui fait office de socle législatif. Dans le droit actuel français, la protection d'une donnée personnelle\* est garantie par 5 critères :

- **la finalité de l'utilisation** : on garde une donnée pour une utilisation déterminée à l'avance.
- **la pertinence** : les données exploitées doivent être pertinentes au regard de l'objectif avancé.
- **la conservation limitée (eq « Droit à l'Oubli »)** : les données ne peuvent être conservées indéfiniment dans un serveur. La durée de conservation est établie en cohérence avec l'objectif d'utilisation.
- **la sécurité et la confidentialité** : le responsable des données doit assurer la sécurité de celles-ci.
- **le respect des droits des personnes (loyauté et transparence)** : elles doivent avoir accès aux données, être informées de leur utilisation et pouvoir s'opposer.

SMILE ... BIG BROTHER IS WATCHING YOU!

### \*Selon la loi

« constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres »

La progression du Big Data interroge la pertinence de ces critères car **l'exploitation automatique de données met à mal le principe de finalité**, tout comme la conservation dans le cloud conduit à des incertitudes en matière de sécurité. **L'exercice du droit des personnes est, quant à lui, très largement bafoué**, les internautes n'étant souvent pas informés de la récupération de leurs données de navigation.

La Commission Européenne s'est donc emparée du sujet, par l'entremise de la Commissaire Viviane Reding en charge de la Justice et des Droits Fondamentaux, qui proposait en janvier 2012 **un projet de règlement unifiant les régimes de protection des données européens et renforçant les droits des personnes ainsi que la responsabilité des entreprises**. Des outils liés au consentement préalable des internautes (sur les cookies par exemple) ont été mis en place, mais les discussions sur le projet de règlement restent vives, certains craignant qu'un encadrement trop strict freine l'innovation et creuse un décalage avec les Etats-Unis, beaucoup moins contraignants sur l'exploitation des données. **Après un an et demi de débats, le projet de règlement a finalement été retoqué en juin 2013** par les ministres de la Justice des Etats Membres, mais Viviane Reding a d'ores et déjà annoncé qu'elle comptait obtenir **un accord sur le projet en 2014**. Une réunion entre ministres de la Justice s'est tenue à Vilnius en juillet 2013, ouvrant la porte à un accord en 2014.

En France, la ministre Fleur Pellerin a renouvelé en 2013 son souhait de présenter un projet de loi relatif à la protection des données personnelles au premier semestre 2014 mais l'ampleur du projet reste floue.

Le débat reste donc encore largement ouvert sur les mesures à adopter pour garantir à la fois le dynamisme économique et la protection des citoyens. **Beaucoup d'analystes s'inquiètent du retard que pourrait prendre l'Europe par rapport aux Etats-Unis et à l'Asie** craignant une fuite des données vers ces zones géographiques et donc, à terme, une perte de souveraineté (les données constituant désormais une matière première stratégique).

A ce titre les récents événements sur le programme PRISM ont ravivé les débats sur ce dossier, rendant la question de la protection des données encore plus stratégique dans le contexte actuel d'explosion de l'univers numérique et d'industrialisation du Big Data.

*La Commissaire Viviane Reding annonce son projet de règlement sur la protection des données en janvier 2012*



« Une rupture de confiance du consommateur vis-à-vis du fournisseur entraînerait automatiquement un retour en arrière »

## 2012



Appel à projets «Management de l'Information Intelligente»



Proposition de règlement européen sur les données personnelles



Lancement du programme «The Big Data Research and Development Initiative»



Inscription du Big Data dans les Investissements d'Avenir

## 2013



Lancement de l'appel à projets «Scalable data analytics»



Annnonce de la création d'une filière Big Data par la ministre Fleur Pellerin



Sélection de 7 projets Big Data au titre des Investissements d'Avenir



Annnonce d'un plan de soutien à la filière du Big Data



Le projet de règlement sur les données personnelles est retoqué



Réunion de Vilnius ouvrant le champ à un accord sur le projet de règlement en 2014

# Interview

**Thierry Dor**

Associé au cabinet *Gide Loyrette Nouel*



## « La France a toujours su faire entendre sa voix sur la scène européenne en matière de protection des données personnelles »

### 1) Pour vous, en quoi le Big Data a-t-il changé la donne dans la protection des données personnelles ? Pourquoi remet-il en cause l'arsenal législatif actuel ?

En matière de protection des données personnelles, au-delà des questions classiques qui sont exacerbées par le volume et la variété des données traitées par le Big Data, ce qui change vraiment, c'est l'application du principe de finalité. En effet, l'une des particularités du Big Data est la remise en cause de l'approche scientifique classique consistant à devoir poser une hypothèse de travail pour ensuite collecter et traiter des données destinées à valider ou invalider cette hypothèse. Avec le Big Data, il est désormais possible, compte tenu des capacités de traitement, d'opérer sans hypothèse de travail préalable. En d'autres termes, mener des recherches sans savoir à l'avance ce que l'on va pouvoir trouver.

Dans un tel contexte, il est difficile de respecter le principe central posé par la loi du 6 janvier 1978 selon lequel « (les données) sont collectées pour des finalités déterminées, explicites et légitimes et ne sont pas traitées ultérieurement de manière incompatible avec ces finalités ».

De même, il est difficile de respecter les autres dispositions de cette loi qui font directement référence au principe de finalité : - l'obligation de ne collecter que des données adéquates, pertinentes et non-excessives au regard des finalités, - ou encore, en ce qui concerne la durée de conservation des données, l'obligation de ne pas excéder la seule période nécessaire aux finalités pour lesquelles elles sont collectées.

### 2) La France a-t-elle une position particulière sur la question par rapport à ses collègues européens ? Comment définiriez-vous cette position ? Et, plus globalement, quels sont les principales différences législatives entre les Etats-Unis et l'Europe concernant la protection des données ?

La France a toujours su faire entendre sa voix sur la scène européenne en matière de protection des données personnelles. Elle est particulièrement active dans le cadre des négociations qui devraient mener à l'adoption du règlement européen en matière de données personnelles : la CNIL, la Ministre de la Justice, le Sénat et l'Assemblée Nationale contribuent au débat public. Toutefois, la France ne fait pas bande à part parmi les Etats Membres et elle partage l'opinion majoritaire.

Concernant les principales différences entre les Etats-Unis et l'Europe, il faut tout d'abord rappeler qu'à la différence de l'Union Européenne, il n'existe pas de loi générale en matière de protection des données personnelles aux Etats-Unis, même si des réglementations sectorielles très strictes peuvent exister dans le domaine des informations financières ou de la santé. Par ailleurs, l'affaire PRISM nous rappelle que des textes américains comme le «USA Patriot Act» (2001), le «Protect America Act» (2007) ou le «Foreign Surveillance Intelligence Act» (2008) permettent une très grande intrusion dans la sphère privée par rapport à ce qui est autorisé en Europe. Les autorités américaines peuvent notamment obtenir les données traitées par toute entreprise présente aux Etats Unis, y compris les données qui concerneraient des ressortissants étrangers, avec une facilité qui déconcerte et effraie en Europe. Il était question, en 2012, d'un texte permettant d'étendre l'accès des autorités américaines aux données si-

tuées en dehors du territoire américain, dans le but de lutter contre la cyber-criminalité et le terrorisme. Il s'agissait du Cyber Intelligence Sharing and Protection Act. Bien que ce texte n'ait finalement pas été adopté, un nouveau projet, très similaire serait en préparation.

### 3) Comment voyez-vous l'évolution des discussions sur le projet de règlement européen relatif aux données personnelles ? Quelles sont les orientations qui se dessinent pour ce texte ? Le rejet de ce texte en juin dernier a-t-il mis en péril les discussions ?

L'instruction du projet de règlement européen, présenté par Viviane Reding début 2012, est toujours en cours et peut encore réserver des surprises, néanmoins les négociateurs se montrent persévérants dans la recherche d'un consensus.

La rencontre informelle de Vilnius, qui a réuni les ministres européens de la Justice le 19 juillet dernier, affiche un bilan positif. L'adoption du règlement n'est pas envisagée avant début 2014, mais le dossier a bien avancé. Les ministres français et allemand, notamment, ont manifesté leur soutien à ce projet.

Par ailleurs, il faut prendre en considération les impacts de l'affaire PRISM sur l'avancée des discussions. Les débats sont aujourd'hui animés par les révélations de M. Edward Snowden, qui avait dévoilé l'existence de ce programme américain collectant des données personnelles d'individus hors des Etats-Unis. Selon Viviane Reding « PRISM a réveillé les Européens sur la nécessité d'avoir des règles fortes et strictes ». Voilà de quoi motiver ceux qui prônent l'adoption du règlement dans les meilleurs délais.



Moins de suppositions.  
Plus de certitudes.  
Analytics. Delivered.\*\*

Les données sont des éléments clés.  
À condition de savoir les interpréter.  
Nous aidons nos clients à utiliser  
l'analyse pour transformer les données  
en idée, l'idée en action et l'action en  
résultats concrets. High performance.  
Delivered.\*

High performance. Delivered.\*

  
accenture

conseil | technologies | externalisation



## COORDONNÉES ACCENTURE

118 avenue de France  
75636 Paris Cedex 13

Tél : +33 (0) 1 53 23 55 55

Fax : +33 (0) 1 53 23 53 23

Mail : xavier.cimino@accenture.com

## CONTACT

**Xavier  
CIMINO**

Responsable Accenture Interactive  
Digital & Analytics

## Les BIG DATA, une opportunité particulièrement lucrative, facteur de différenciation compétitive

Dans un monde en perpétuelle évolution et accélération, des données de plus en plus volumineuses, riches et variées – les fameuses BIG DATA – représentent un atout majeur pour les entreprises, à condition de savoir les capter, les raffiner, les combiner et les analyser. Ce formidable vivier de nouvelles informations permet non seulement de mieux répondre aux attentes des clients mais surtout d'identifier de nouveaux axes stratégiques ainsi que de capter d'autres marchés.

Grâce à la réduction des coûts de stockage et aux développements de nouvelles techniques de gestion de la donnée, ce monde de l'exploration et de l'interprétation, autrefois plus réservé à des spécialistes, s'ouvre à présent à toute entreprise brassant un grand volume de données.

De la stratégie à la mise en œuvre, Accenture vous accompagne dans le développement de capacités analytiques et prédictives nécessaires à la prise de décision et au pilotage de votre activité.

Accenture vous aide également à adresser les nouveaux défis technologiques des BIG DATA, notamment :

- gérer des plates-formes de données non structurées de grande capacité ;
- faire face aux nouvelles demandes pesant sur les infrastructures réseau ;
- accueillir un grand nombre de serveurs standards au sein des centres de données ;
- déployer des capacités de stockage de plusieurs pétaoctets ;
- exploiter ces précieuses données ;
- développer et mettre en place une gouvernance adaptée à la complexité et l'hétérogénéité des données ;
- intégrer le traitement de données non structurées dans les organisations et les processus existants.

## Notre offre de services Accenture Analytics, vous permet de disposer :

- **Des meilleurs professionnels.** Notre réseau mondial de professionnels offre un accès à une combinaison de compétences analytiques approfondies, de connaissances métier et d'expérience des processus.
- **D'une offre éprouvée.** Grâce à notre longue expérience, notre offre de services couvre plus de 100 secteurs d'activité individuels ou croisés et de services fonctionnels, plus de 200 pratiques et normes sectorielles clés, des outils de développement automatisés et des composantes structurelles réutilisables.
- **Des services de recherche et développement inégalés.** Les outils et les services analytiques d'Accenture constituent l'aboutissement d'un long historique d'innovation dans la gestion des data.
- **D'alliances stratégiques.** Bien que nous ne privilégions aucun prestataire spécifique, nous avons formé un réseau de partenaires stratégiques avec lesquels nous nous associons éventuellement lors de la mise en œuvre de nos projets, notamment SAS, SAP, Oracle et Microsoft.
- **D'une envergure mondiale.** Grâce à notre réseau mondial de centres dédiés à l'analytique, nous modélisons automatiquement nos processus analytiques dans un souci de constante amélioration. Nous générons ainsi une source conséquente de connaissances tirées de l'analyse des données.

En vous appuyant sur Accenture, vous aurez l'assurance de maximiser l'utilisation de vos données et ainsi créer plus rapidement davantage de valeur.

## A propos d'Accenture

Accenture est une entreprise internationale de conseil en management, technologies et externalisation. Combinant son expérience et ses capacités de recherche et d'innovation développées et mises en œuvre auprès des plus grandes organisations du monde sur l'ensemble des métiers et secteurs d'activité, Accenture aide ses clients - entreprises et administrations - à renforcer leur performance. Avec plus de 266 000 employés intervenant dans plus de 120 pays, Accenture a généré un chiffre d'affaires de 27,9 milliards de dollars au cours de l'année fiscale clôturée le 31 août 2012.



(Responsable technologique des offres Business Intelligence, Big Data & Analytics : Laurent Gaultier)

# Débloquez le potentiel de vos Big Data !

Trouvez enfin un moyen efficace et facile d'analyser l'ensemble de vos données

Les informations disponibles sur votre activité et votre clientèle prennent des formes de plus en plus variées avec l'évolution des technologies de communication et de stockage des données. Elles offrent de nouvelles possibilités et l'idée de les exploiter est séduisante. Mais comment faire pour que cela reste facile et accessible ?

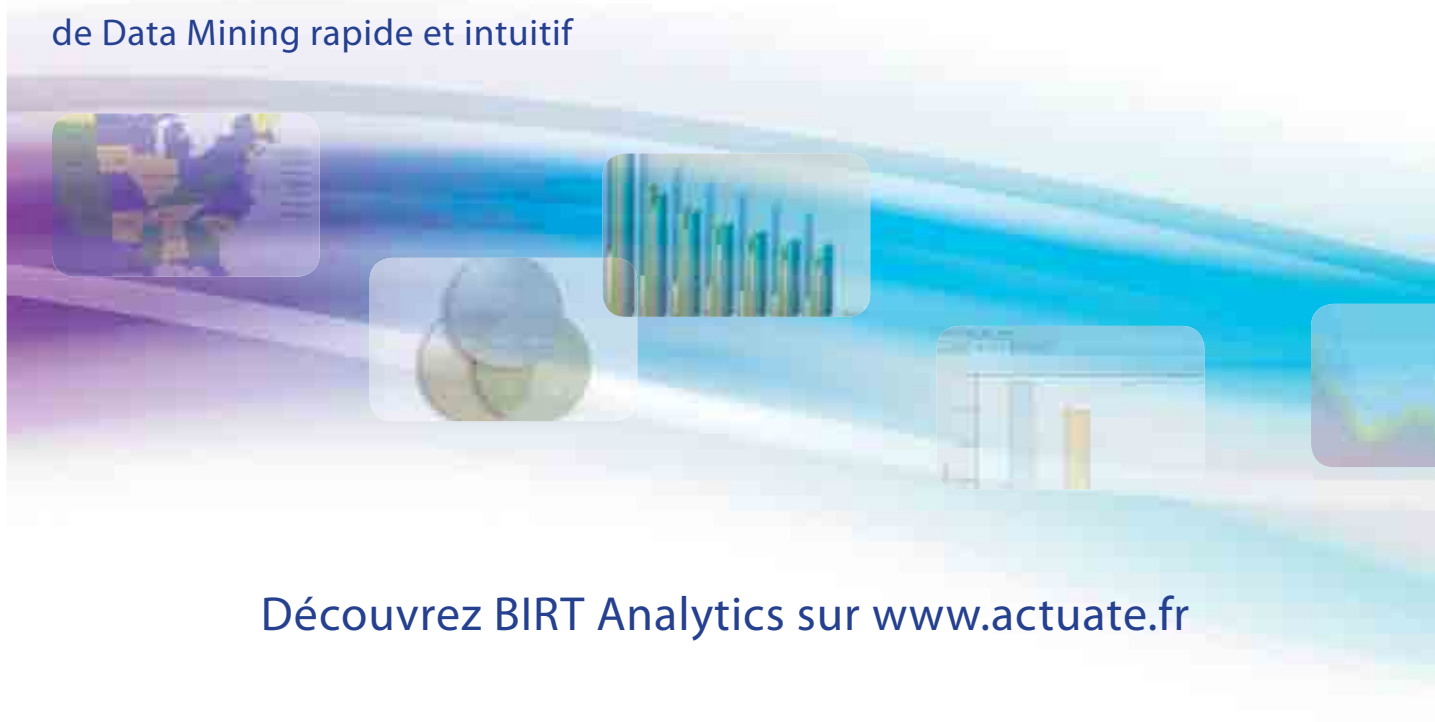
**Il existe désormais des solutions simples et rentables**

Avec son module de Data Mining intuitif, Actuate – the BIRT Company – donne aux experts marché et métier le pouvoir d'analyser directement leurs Big Data sans l'aide d'un Data Scientist. Plus réactive et avec une meilleure connaissance de son activité, votre entreprise en devient plus compétitive. Nos clients utilisent BIRT Analytics pour :

- Détecter des opportunités croisées
- Découvrir des modèles de comportement
- Prédire des évolutions
- Grouper et croiser des populations

**Le Data Mining n'est plus dévolu qu'aux Data Scientists**

**Analysez des données brutes à la volée grâce à notre module de Data Mining rapide et intuitif**



Découvrez BIRT Analytics sur [www.actuate.fr](http://www.actuate.fr)

## COORDONNÉES

### ACTUATE INTERNATIONAL CORPORATION

57 rue de Châteaudun  
75009 Paris

Tél : +33 (0) 1 76 21 56 50

Fax : +33 (0) 1 76 21 56 56

Mail : fr\_contact@actuate.com

## CONTACT

**Magali  
CARIO**

*Marketing Campaigns Manager  
Southern Europe*

### Actuate - The BIRT Company™

Actuate a fondé et codirige le projet Open Source BIRT, qui est utilisé par plus de 2,5 millions de développeurs à travers le monde et sert de socle à la suite commerciale proposée par Actuate. Pouvant être facilement déployées auprès de millions d'utilisateurs, les applications construites avec BIRT offrent une meilleure vision métier ainsi qu'une compréhension approfondie des comportements des clients. Actuate aide ainsi les entreprises à tirer pleinement parti des Big Data et de l'essor des appareils tactiles.

Les solutions BIRT d'Actuate facilitent le développement d'applications personnalisées de business analytics et de communication client basées sur BIRT. Les applications construites avec le format unique de design BIRT peuvent accéder et intégrer tous types de données, y compris les données non structurées. Elles assurent une expérience utilisateur ergonomique et homogène, quel que soit le niveau de connaissances techniques, et sont supportées pour tous les déploiements, qu'ils soient sur site, Cloud, hybrides, web ou mobiles.

Actuate, cotée au NASDAQ sous le symbole BIRT, compte plus de 5 000 clients à travers le monde dans divers secteurs d'activité, dont les services financiers, les technologies et le secteur public. Son siège social est basé dans la Silicon Valley.

### BIRT Analytics, le Data Mining et l'Analyse prédictive intuitifs

BIRT Analytics™, la solution Actuate permettant l'exploration de données graphique et l'analyse prédictive, permet de mettre en place une analyse rapide à partir de n'importe quelle source de données sans besoin de connaissances techniques et statistiques spécifiques.

BIRT Analytics procure aux analystes et aux opérationnels des fonctionnalités avancées pour prendre des décisions plus avisées et opportunes à l'ère des Big Data. La solution allie la facilité d'utilisation et la flexibilité des outils d'exploration de données à la puissance statistique et algorithmique des outils d'analyse avancée, généralement réservés aux experts IT.



BIRT Analytics donne la possibilité aux analystes métiers de découvrir par eux-mêmes des corrélations et de prévoir des comportements et des résultats par le biais de techniques telles que la segmentation client, la 'prochaine meilleure offre', le risque fournisseur ou la fidélisation client.

Au-delà de l'analyse des données, BIRT Analytics offre également la possibilité de concevoir des campagnes et d'en prédire les retombées pour une amélioration constante des résultats opérationnels.

« Avec BIRT Analytics, nous proposons le meilleur de l'univers analytique », déclare Nobby Akiha, senior vice-président du Marketing chez Actuate. « Les analystes métier n'ont plus à se contenter d'analyses prédictives 'sommaires' et profitent de la puissance analytique des outils statistiques traditionnels dans un produit visuel, facile à utiliser, qui analyse tous les types et toutes les tailles de données. Et même le Big Data » ajoute-t-il. Grâce à BIRT Analytics, les analystes métiers qui doivent traiter des milliards de données disposent d'une solution puissante et simple d'utilisation pour effectuer des prédictions sans avoir à coder.

« Combiné aux autres modules de la suite de produits basés sur BIRT, BIRT Analytics fait le lien entre la gestion métier

orientée données interne et externe, et vous permet de remplir votre mission d'aujourd'hui, c'est-à-dire gérer et décider en vous fondant sur les données pour gagner un avantage concurrentiel » conclut Nobby Akiha.



# DISPOSEZ DES MÊMES ATOUTS QUE LES LEADERS DU WEB

Affini-Tech met en oeuvre pour vous des briques technologiques packagées permettant de réaliser très rapidement des retours sur investissements importants. Ces briques vous permettent d'élargir le paysage des données que vous utilisez dans vos analyses et de bénéficier de toute la performance des algorithmes de "Machine Learning". L'idée maitresse de ces outils est de transposer à toute activité (online ou traditionnelle) les recettes qui ont fait le succès des leaders de l'économie numérique.

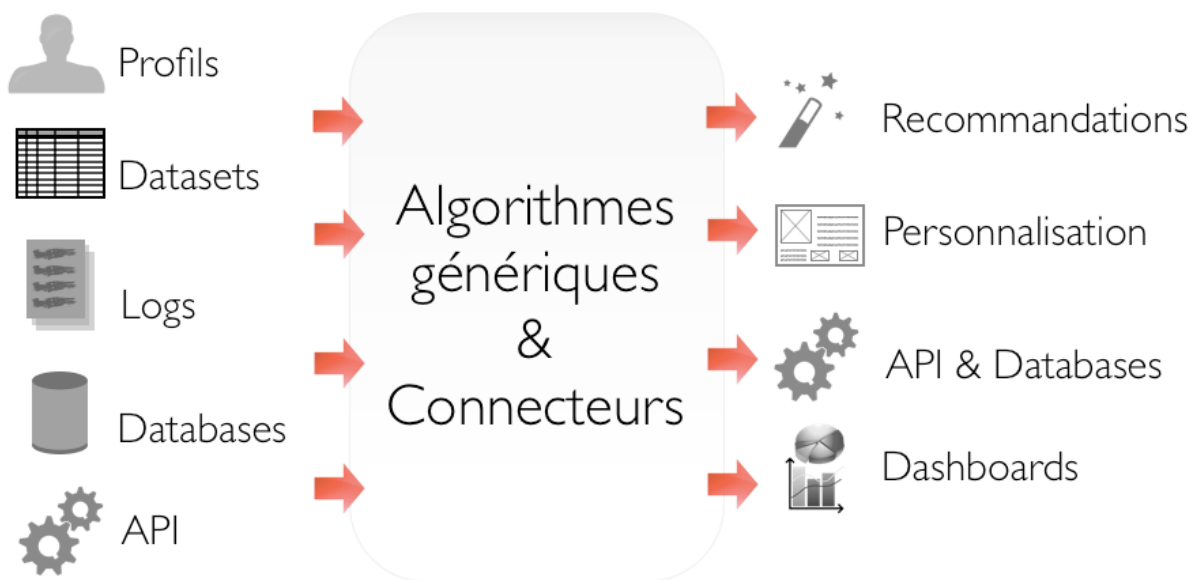
## Hybridation de données

Les données de votre organisation sont à 80% liées à des dates ou des lieux géographiques. Accédez à de potentielles liaisons fortes de vos informations à des données complémentaires, pour par exemple cibler plus finement vos clients.

## Moteur de Recommandation

Qui n'a jamais été frappé par les grands sites eCommerce vous proposant d'acheter les produits que d'autres consommateurs semblables à vous ont acheté ? Ce levier d'action peut être mis au profit de votre business simplement et rapidement.

Avec les technologies de moteur de recommandations d'Affini-Tech, tirez parti des informations que vous possédez sur vos clients pour mettre en oeuvre des stratégies de recommandations et de personnalisation de vos relations, et ainsi améliorer la performance de vos business.



### Affini-Tech SARL

11 avenue Henri IV, 92190 Meudon

Tél : +33 (0) 6 61 88 76 71

Fax : +33 (0) 9 58 12 97 98

Mail : [contact@affini-tech.com](mailto:contact@affini-tech.com)

Web : [www.affini-tech.com](http://www.affini-tech.com)



## COORDONNÉES

### AFFINI-TECH SARL

11 avenue Henri IV  
92190 Meudon

Tél : **+33 (0) 6 61 88 76 71**

Fax : +33 (0) 9 58 12 97 98

Mail : vhe@affini-tech.com

## CONTACT

**Vincent  
HEUSCHLING**

*Fondateur*

**«Les outils, les pratiques du Big Data s'adressent à toutes les organisations, quelles que soient leurs tailles».**

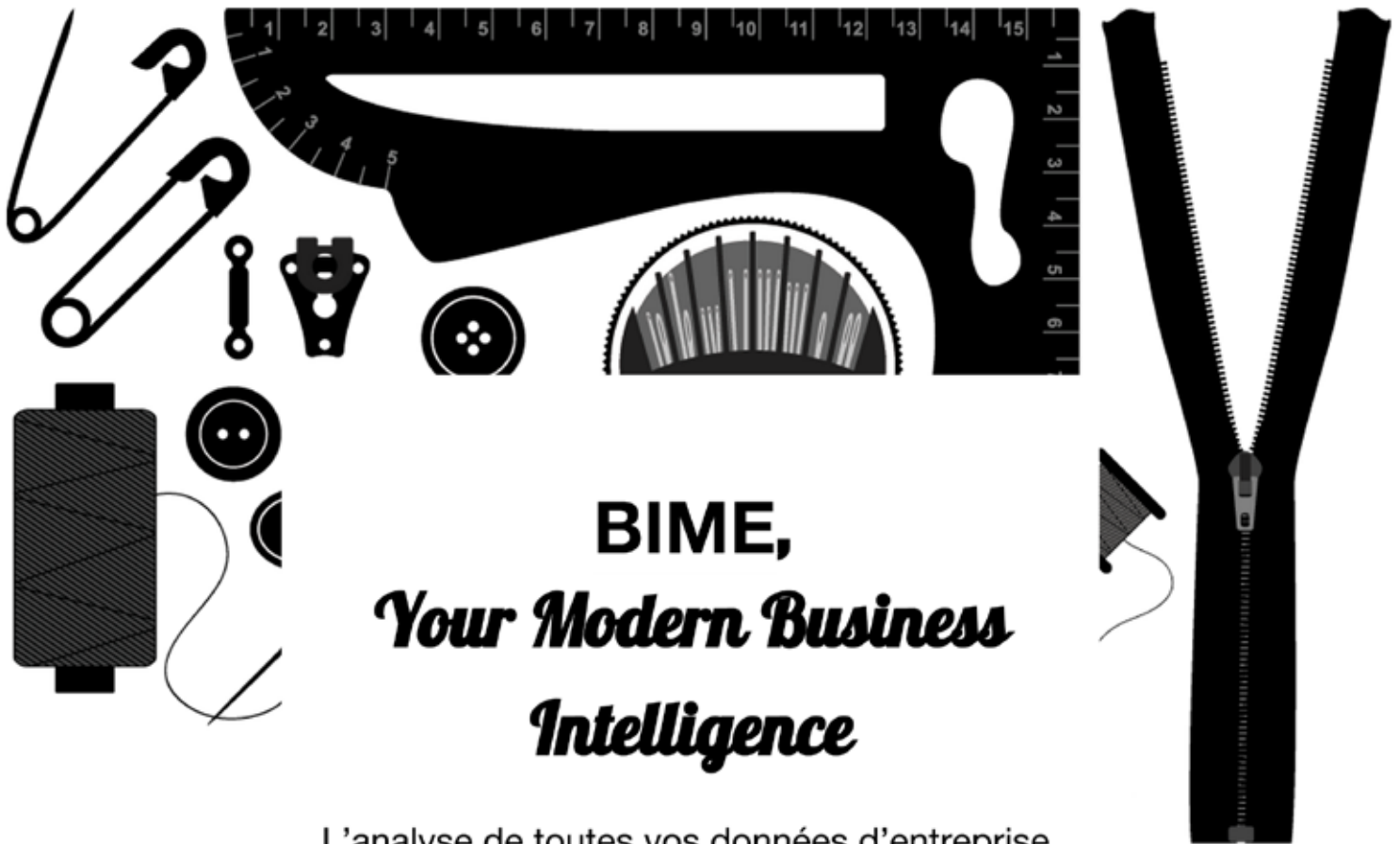
Affini-Tech, acteur français du Big Data, le démontre tous les jours chez des clients de tous secteurs d'activité. Si le Big Data est né et a grandi chez les géants du Net, c'est aujourd'hui dans les industries traditionnelles que les plus grands bénéfices sont générés.

Le positionnement affirmé et le savoir-faire d'Affini-Tech est de mettre à la portée de toute entreprise, sur ses volumes de données, les apports du Big Data : Analyse de transactions, Prédiction, Recommandation, Classification, Audit...

Par une démarche à la conjonction entre les métiers et l'IT, et par la démonstration, l'entreprise découvre de nouvelles opportunités pour des coûts radicalement plus faibles. Le retour sur investissement est obtenu en quelques semaines, et les coûts sont maîtrisés. L'utilisation de méthodes agiles permet de conserver à tout moment des objectifs atteignables. Vous disposez tant du savoir faire d'Affini-tech que de ses briques technologiques prêtes à l'emploi, développées pour répondre à des besoins précis. Affini-Tech vous permet d'en comprendre les ressorts et d'être en possession d'une RoadMap Big Data pour vos acteurs clés internes.

Affini-tech ne vous accompagne pas uniquement par la technologie. Une démarche auprès des métiers vous permet de bien plus grandes créations de valeur et de spectaculaires diminutions de vos coûts. C'est aussi grâce à la nouvelle capacité offerte par les plateformes Big Data à traiter massivement de gros volumes de données qu'il devient possible, à moindres coûts, de mettre en place des innovations. Le Big Data est abordable.





# BIME, *Your Modern Business Intelligence*

L'analyse de toutes vos données d'entreprise,  
A un seul et même endroit,  
Peu importe leur volume

Texte *Rachel Delacour, co-fondatrice*

**BIME** est une solution qui permet de croiser des données hébergées dans votre entreprise - grandes bases de données propriétaires comme Oracle par exemple - et des données issues de solutions SaaS telles que Google Analytics, Salesforce, Facebook, puis de donner accès à toute la structure de l'analyse pour créer vos propres tableaux de bord dynamiques.

Suivez vos indicateurs de performances sous forme de visuels explicites avec lesquels vous pouvez interagir. Collaborez avec vos équipes, clients ou partenaires en partageant vos analyses selon le mode de diffusion qui vous convient le mieux : envoi automatique par email, accès à un portail, intégration dans votre intranet.

BIME permet ainsi à ses clients de construire dynamiquement de l'information à partir de données nombreuses, complexes et multicanales et de faciliter son partage pour une prise de décision plus efficace. C'est la force du **big data** et de la Business Intelligence au service des entreprises et des collectivités.

## Les forces de BIME

**1** Une solution pensée en mode SaaS, opérée par le Cloud, accessible depuis votre navigateur, rendant possible le traitement rapide d'énormes quantités de données. **2** Un interfaçage rapide avec les bases de données, d'entreprises et publiques, et une grande scalabilité. **3** Une qualité visuelle et une simplicité d'utilisation réelle qui favorise une analyse plus rapide au premier coup d'œil, en ligne.



## COORDONNÉES

### BIME (WE ARE CLOUD SAS)

1 place Francis Ponge  
34 000 Montpellier

Tél : **+33 (0) 4 67 41 60 64**

Fax : N/A

Mail : [benoit@bimeanalytics.com](mailto:benoit@bimeanalytics.com)

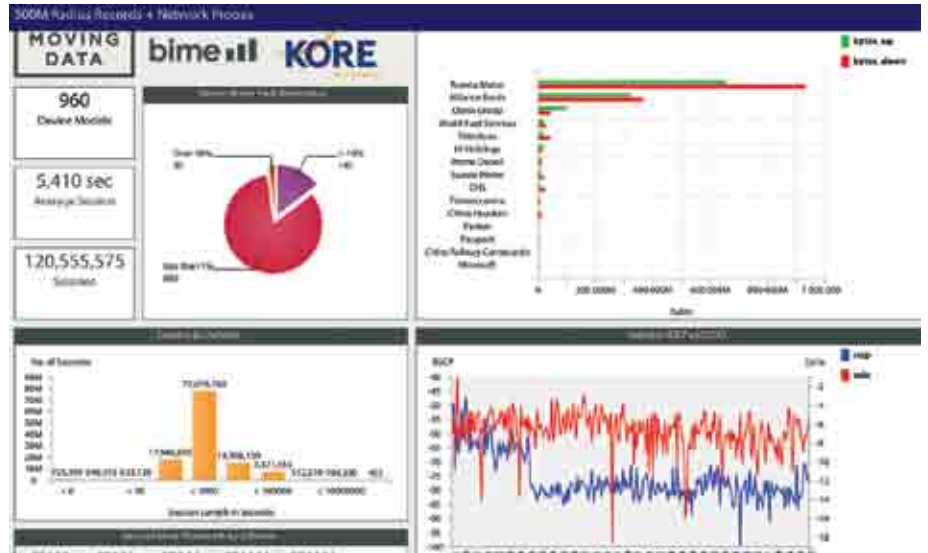
## CONTACT

**Benoit  
PARRA**

*Responsable Commercial France*

## Leur Histoire Big Data avec BIME

Puiser au cœur de l'internet des objets dans le désert Australien, c'est ce que fait la société MovingData. Afin d'améliorer la performance mobile de ses clients Telco nationaux, elle assure la qualité et la disponibilité des réseaux de téléphonie en dehors des grandes villes australiennes grâce à des milliers de capteurs répartis géographiquement.



500 Radius Records + Network Probes



Craig Morton  
Ceo of MovingData



## Pour comprendre plusieurs millions d'évènements générés sur ses réseaux :

MovingData devait répondre à 2 besoins particuliers que seul BIME pouvait satisfaire : disposer d'outils permettant une analyse interactive et visuelle rapide des incidences réseaux et pouvoir communiquer ces analyses à leurs clients Telco grâce à des tableaux de bord collaboratifs, disponibles en ligne, sur des volumétries de données très importantes stockées dans Google BigQuery.

Ce projet BigData a été réalisé en moins de trois mois. Trois partenaires Cloud, distants géographiquement, ont sans jamais se rencontrer pu l'orchestrer : l'Américain Google pour la base de données, le Français BIME pour l'analyse et les tableaux de bord, l'Australien MovingData pour les indicateurs métiers et les process d'alimentation de données.

Pensez vos projets BigData avec l'agilité et la scalabilité du Cloud.





# Une décision repose-t-elle **sur une intuition ou des données ?**



CGI Business Consulting, cabinet de conseil en innovation et transformation, fait partie du Groupe CGI inc. Ses consultants associent expertises sectorielles, fonctionnelles et technologiques pour accompagner les plus grandes entreprises et organisations. Parce que chaque client est unique, CGI Business Consulting a créé des méthodes de travail spécifiques permettant à chacun de prendre part au management de sa transformation et garantissant une amélioration durable de ses performances.

#### **Coordonnées**

CGI Business Consulting  
17 place des Reflets  
92 097 Paris La Défense cedex

#### **Contact**

Frederic Brajon, Associé  
T : +33 1 57 87 50 55  
Email : frederic.brajon@cgi.com

#### **Site web**

[www.cgi.com](http://www.cgi.com)



## COORDONNÉES

### CGI BUSINESS CONSULTING

17 place des Reflets  
92 097 Paris La Défense cedex

Tél : **+33 (0) 1 57 87 50 55**

Fax : N/A

Mail : frederic.brajon@cgi.com

## CONTACT

**Frederic  
BRAJON**

Associé

### Big Data & Data Science

Mieux connaître ses clients, anticiper les changements structurels de son marché, détecter les intentions d'achat, de résiliation, de fraude... Telles sont quelques-unes des promesses du big data.

L'ouverture croissante de l'information pose un véritable défi de compétitivité aux grandes entreprises : rendre leur offre lisible dans un catalogue chaque jour plus complexe et plus confus.

Les entreprises qui sauront non seulement triompher de cette complexité, mais en tirer profit en se positionnant comme détenteurs d'une information intelligible et consolidée, seront au centre de cette nouvelle économie.

Quelle est la véritable valeur de l'information que mon entreprise détient ? Quelle information représente un avantage compétitif potentiel ? Laquelle pourrait devenir un produit ou un service potentiel ? Quelles synergies pourrions-nous trouver dans des gisements d'information externes ?

A l'heure où une masse d'information sans précédent devient disponible, et où apparaissent des dizaines d'outils et de méthodes pour en tirer toute la valeur, une question essentielle demeure :

« Quels gains pouvons-nous dès aujourd'hui tirer de cette révolution encore naissante ? »

### Réussir un projet de Big Data nécessite un ensemble de compétences permettant d'accompagner les entreprises sur différents axes :

- Stratégique : Réconcilier les décideurs avec la donnée et faire de la maîtrise de l'information un enjeu stratégique (anticiper les bouleversements des écosystèmes, analyser les mouvements du marché, réduire les coûts, valoriser le patrimoine immatériel, ...)
- Métier : Analyser les leviers de performances, optimiser les processus (production, ventes, fraude, positionnement),
- Organisationnel : repenser la gouvernance, faire accoster un modèle multi-informatif (données, références, indicateurs, données réglementaire, données externes), repenser les accès et les stratégies d'archivages, redistribuer les rôles, adapter les politiques RH
- Création de valeur : combiner les expertises (informatique, mathématique, business) au travers d'une approche de Data Science éprouvée pour révéler le potentiel de l'information

Pour répondre à ces enjeux, CGI Business Consulting, cabinet de conseil du groupe CGI a créé une entité dédiée au traitement et à la valorisation de l'information : « Digital Information Strategy »

Ses consultants expérimentés sont à votre disposition pour vous accompagner dans vos projets de transformation





# Devenez Membre du **Club Décision DSI**

Le 1<sup>er</sup> Club Européen indépendant de Décideurs informatiques

*Décideurs informatique,  
élargissez votre cercle de  
relation, identifiez les solutions  
innovantes et tendances de  
demain, consolidez vos  
compétences*



**Club Decision  DSI**

16, Place Vendôme • 75001 Paris  
Tél. : 33 (0)1 53 45 28 65

[www.decisionsi.com](http://www.decisionsi.com)

## COORDONNÉES

### CLUB DÉCISION DSI

16 Place Vendôme  
75001 Paris

Tél : **+33 (0) 1 53 45 28 65**

Fax : N/A

Mail : [club@decisiondsi.com](mailto:club@decisiondsi.com)

## CONTACT

**Véronique  
DAVAL**

*Présidente du Club*

## Le Club Décision DSI, premier club français dédié aux directeurs informatiques

Aujourd'hui le Club Décision DSI réunit 1 100 membres «directeurs Informatique & Télécoms» de sociétés du secteur privé et du secteur public dont le parc informatique est supérieur à 200 pc. Il rassemble des hommes et des femmes d'expérience, animés par la volonté de mener des actions qui soutiennent d'autres directeurs des systèmes d'information (DSI) dans la réalisation de leurs projets professionnels.

Le club est administré par Véronique Daval, la présidente, le vice-président, ainsi que par un conseil central composé de neuf directeurs informatiques.



## Chaque année le club organise :

- 20 manifestations thématiques par an.
- 7 événements accessibles aux partenaires (éditeurs de logiciels, constructeurs, opérateurs télécoms, SSII)
- Les Rencontres de l'innovation technologique
- Le consortium du DSI
- Les DSI Meeting (table rondes, débats)
- Le dîner de Gala avec son village solutions (réservé aux membres et aux partenaires du club)
- Le dîner de gala annuel

## HELP DSI , le 1<sup>er</sup> Réseau francophone de 1100 DSI connectés

HELP DSI ce sont des DSI, des groupes de DSI et des communautés de DSI qui échangent en associant leurs idées, leur expérience et leurs connaissances Afin d'optimiser leurs choix technologiques.

Le club dispose de son propre réseau social (RSE) sur lequel de nombreuses thématiques sont abordées qui permettent aux membres de capitaliser sur l'expérience des uns et des autres. Les ambassadeurs de Club Décision DSI ont adopté cette solution afin d'améliorer la communication entre les membres, de créer des groupes de réflexion sur des thématiques variées, et de diffuser les compétences à travers le réseau des membres au niveau national et européen.

Ce moteur d'expérience permet aux sociétés adhérentes de se servir des retours d'expérience des projets IT mis en place par les membres.

Chaque année les ambassadeurs identifient des thématiques auprès des membres et définissent un programme annuel consultable sur le site internet du Club :

[www.club-decisiondsi.com](http://www.club-decisiondsi.com)





# Présentation de Data&Data Consulting

Data&Data Consulting est une startup spécialisée dans le Big Data, le conseil en stratégie et en organisation sur les problématiques liées à l'analyse, la gouvernance et la qualité des données.



• La Big Data va au delà des données de votre organisation et de ses systèmes opérationnels. Afin de réellement tirer avantage de cette opportunité vous devez comprendre et définir la valeur des données véhiculées en interne et en externe.

Notre équipe vous assiste dans la définition et le déploiement de la stratégie Big Data. Ainsi, vous pourrez réellement tirer un avantage économique de l'analyse des données internes et externes à votre entreprise.

**Nos domaines d'intervention sont les suivants :**

- Construction de la feuille de route Big Data pour votre organisation ;
- Identification et préconisation des solutions techniques ;
- Evaluation de l'actif informationnel de votre organisation et de son niveau de qualité ;
- Déploiement de la stratégie Big Data et gestion de projet ;
- Renforcement de vos équipes par des ressources expertes en Big Data: visualisation des données, « Data Scientists » et consultants expérimentés ;
- Formation et accompagnement au changement.

## Notre approche

Notre approche se caractérise par un conseil de proximité, qui favorise des solutions pragmatiques et sur-mesure.

Nous disposons d'un centre de R&D dédié aux problématiques de Big Data et de visualisation des données, et nous travaillons avec des chercheurs d'universités de premier plan, Berkley, Copenhagen Business School et Stanford University, sur la conception et le développement de solutions innovantes pour nos clients.

**Notre modèle repose sur trois grands principes :**

- 1) l'indépendance, pour garantir l'objectivité de nos recommandations ;
- 2) l'innovation au travers de notre centre de "R&D"
- 3) un large réseau de partenaires pour être en mesure de répondre à l'ensemble des problématiques de nos clients.

## Notre équipe

Notre équipe constituée de consultants, statisticiens et ingénieurs spécialisés dans la data, est reconnue par ses clients pour son implication, son professionnalisme, son expertise sectorielle, et la recherche permanente de résultats concrets.

**Nous mettons au profit de nos clients un large spectre de compétences :**

Big Data, Business intelligence, Gouvernance et qualité des données, master data management, analytics et techniques d'analyse de données pour l'audit, les investigations et les recherches de fraude.

[www.data-and-data.com](http://www.data-and-data.com)

Contact : Zouheir Guédir ASSOCIÉ • 33 6 22 35 14 20 • [zouheir.guedir@data-and-data.com](mailto:zouheir.guedir@data-and-data.com)



## COORDONNÉES

### DATA & DATA CONSULTING

31 rue Saint-Lazare  
75 009 Paris

Tél : +33 (0) 6 22 35 14 20

Fax : N/A

Mail : [zouheir.guedri@data-and-data.com](mailto:zouheir.guedri@data-and-data.com)

## CONTACT

**Zouheir  
GUEDRI**

Associé



### Zouheir Guedri, Directeur de Data&Data Consulting

#### 1) Pouvez-vous nous présenter Data&Data Consulting, ses origines et sa spécificité sur le marché ?

Data&Data Consulting est une start up spécialisée sur le Big Data, qui conseille les sociétés dans la valorisation de leurs données : nous apportons à la fois des recommandations en terme de pistes de création de valeur, de choix de technologies, de procédures et de stratégies à mettre en place.

Data&Data Consulting est une structure constituée de consultants expérimentés, spécialisés dans la data avec un background d'ingénieurs et de statisticiens.

#### 2) Quelles prestations proposez-vous à vos clients et quelles sont vos cibles principales ?

Nous travaillons essentiellement avec des grands comptes mais également avec quelques start ups ayant des problématiques Big Data ou small data. Leur besoin principal est de savoir comment valoriser leurs données : nous leur proposons donc des études de marché, des conseils en matière d'orientations stratégiques et d'organisation, ainsi que des recommandations technologiques en fonction de leur activité. Pour cela, nous disposons de compétences de développement ainsi que d'un réseau de partenaires pour pouvoir lier l'implémentation opérationnelle aux recommandations stratégiques.

Nous disposons également d'un centre de R&D dédié aux problématiques de Big Data et de visualisation des données, et nous travaillons avec des chercheurs d'universités de premier plan, Berkeley, Copenhagen Business School et Stanford University, sur la conception, le développement et l'expérimentation de solutions innovantes pour nos clients.

#### 3) Plus globalement, comment voyez-vous l'évolution du marché du Big Data en France ? A-t-il atteint sa phase de maturité ? Y a-t-il quelques retards ?

L'année 2013 a été une année d'expérimentation pour un grand nombre d'entreprises : on retrouve des projets avancés dans tous les secteurs, pas seulement les plus gros détenteurs de données, et ce sont pour la plupart de purs projets Big Data (et non pas un recyclage de projets anciens). De plus, la réflexion autour du Big Data est largement avancée, la phase de méfiance étant dépassée : au-delà d'un effet de mode, on reconnaît un véritable intérêt au Big Data et les réticences, bien que toujours présentes (notamment sur la question de la confidentialité des données) commencent à s'estomper. Cependant, nous nous trouvons pour l'heure dans la phase d'expérimentation et les premiers résultats en termes de ROI devraient plutôt être observés en 2014.

Le véritable retard en France se situe davantage du côté de la mise en place d'une organisation décisionnelle : si les technologies et les objectifs changent, les processus de prise de décisions restent les mêmes, ce qui conduit à un manque de réactivité par rapport à l'information véhiculée par les données. Pour moi, il faudrait mettre en place des boucles de décisions automatisées, comme aux Etats-Unis.

#### 4) Vous évoquez la question de la protection des données personnelles... que pensez-vous du débat actuel et des actions des pouvoirs publics sur le sujet ?

Le Big Data, comme toute avancée technologique, peut comporter des risques, qu'il ne faut surtout pas ignorer : en effet, le Big Data repose sur la confiance du consommateur et toute rupture dans cette confiance entraînerait automatiquement un retour en arrière. Ce qui s'est produit avec le programme PRISM est un exemple d'excès qui peut effrayer les citoyens et consommateurs. De la même façon, on craint que le Cloud ne soit pas assez protecteur, alors qu'il est à la base de l'essor du Big Data. Il est donc urgent de maîtriser ces risques pour garder la confiance des consommateurs ; en cela, je crois que les projets amorcés en Europe sur la protection des données vont dans le bon sens. Par contre, je juge essentiel de ne pas opposer le modèle européen au modèle américain, ce serait contre-productif et ne permettrait pas de garantir l'essor du Big Data.

De la même façon, le plan du gouvernement pour encourager la filière Big Data est un bon signe envoyé au marché ; il faudra simplement l'orienter davantage vers les start ups pour les années à venir.



**3S EXALEAD** | *CloudView*

**3S EXALEAD** | *OneCall*

**3S EXALEAD** | *OnePart*

**Ne vous laissez  
pas distancer !**



D'ici à 2015, Gartner prévoit que sur les 500 entreprises du classement Fortune, plus de 85 % échoueront à exploiter efficacement les Big Data pour rester compétitives.

Avec **EXALEAD**, aucune raison que vous soyez l'une d'entre elles !

**3S DASSAULT  
SYSTEMES**

[www.3ds.com/products/exalead/](http://www.3ds.com/products/exalead/)

## COORDONNÉES

### EXALEAD

10 place de la Madeleine  
75008 PARIS

Tél : +33 (0) 1 55 35 26 26

Fax : +33 (0) 1 55 35 26 27

Mail : [contact@exalead.com](mailto:contact@exalead.com)

## CONTACT

**Axel  
BRANGER**

*Directeur Commercial*

## EXALEAD est un fournisseur de logiciels de recherche et d'accès à l'information en entreprise et sur le Web.

EXALEAD propose EXALEAD CloudView™, une technologie logicielle industrielle qui offre agilité, fonctionnalités sémantiques avancées et performances uniques. CloudView permet de concevoir des applications d'exploration et d'analyse de l'information pour transformer les données de l'entreprise en richesse exploitable et pertinente. Ces applications reposent sur un index et réconcilient données structurées et non structurées, internes ou externes.

### Search Based Applications et Big Data

EXALEAD offre des applications d'exploration et de l'analyse de l'information pour libérer, capitaliser, valoriser votre Big Data.

Les applications construites autour d'EXALEAD CloudView permettent aux entreprises de réellement exploiter leur Big Data sans remise en question du SI existant, avec un déploiement rapide et un ROI complet en quelques mois.

Que vous cherchiez à valoriser les pétaoctets de vos entrepôts de données, ou à explorer la valeur cachée dans l'immensité du contenu non structuré comme les fichiers de logs, les emails ou des pages Web, EXALEAD CloudView propose une approche pragmatique.

### Collecte et fusion (agrégation de données)

- Contenu Web
- Contenu d'entreprise
- Systèmes existants / spécifiques
- Fusion de données

### Data mining et analyse de contenu

- Analyse de données textuelles
- Analyse de données Web
- Analyse multimédia
- Analyse de sentiments

### Recherche et navigation facilitées

- Recherche plein texte et navigation à facettes
- Accès multimédia
- Recommandation de contenu

### Des applications métiers à forte valeur ajoutée

- Déployer rapidement des applications orientées processus, enrichies des informations pertinentes extraites des Big Data

EXALEAD est une marque du groupe Dassault Systèmes.





# WITH THE POWER OF BIG DATA... ENLARGE YOUR BUSINESS!



## FastConnect a développé une expertise pour vous conseiller et vous accompagner sur vos projets Big Data :

**Data Science** : une équipe d'experts du Data-Mining, des statistiques, du Machine-Learning et de la Data Viz vous guide afin d'extraire la valeur de vos données.

**Hadoop** : nos experts maîtrisent la plateforme Hadoop et son riche écosystème. Pour vous faire profiter pleinement de cette expertise, nos consultants formateurs sont là pour vous préparer à ce nouveau standard.

**Agilité** : l'agilité est au cœur des projets Big Data. La flexibilité de nos outils et approches vous permet de démarrer sur un périmètre restreint et de développer sans limite. Pour vous aider à réaliser vos projets FastConnect vous accompagne avec une méthodologie Agile éprouvée sur de nombreux projets.

**Industrialisation** : depuis 9 ans FastConnect a développé une expertise dans la mise en place de projets complexes et exigeants en environnements distribués. FastConnect a étendu son expertise autour des technologies et outils des plateformes Big Data, Batch ou temps réel.



## COORDONNÉES

### FASTCONNECT

11, rue des Peupliers  
92130 Issy-les-Moulineaux

Tél : +33 (0) 1 45 07 81 65

Fax : +33 (0) 1 45 07 90 87

Mail : mathias.kluba@fastconnect.fr

## CONTACT

**Mathias  
KLUBA**

*Responsable Big Data / NoSQL*



## Le Big Data est une réponse aux nouveaux défis d'aujourd'hui

- Comment faire émerger la valeur d'un volume d'informations en constante croissance, dans des délais et coûts maîtrisés ?
- Comment exploiter l'information jusqu'alors inexploitée, due à sa nature non-structurée ?
- Pour prendre des décisions et des stratégies gagnantes, comment se différencier de ses concurrents par des analyses toujours plus pointues et toujours plus proches du métier ?

Les solutions techniques Big Data / NoSQL sont les outils pour atteindre ces objectifs, mais nécessitent des compétences spécifiques :



Le rôle du Data Scientist est de vous accompagner dans la conception et la présentation d'analyses élaborées pour tirer toujours plus de valeur des données.



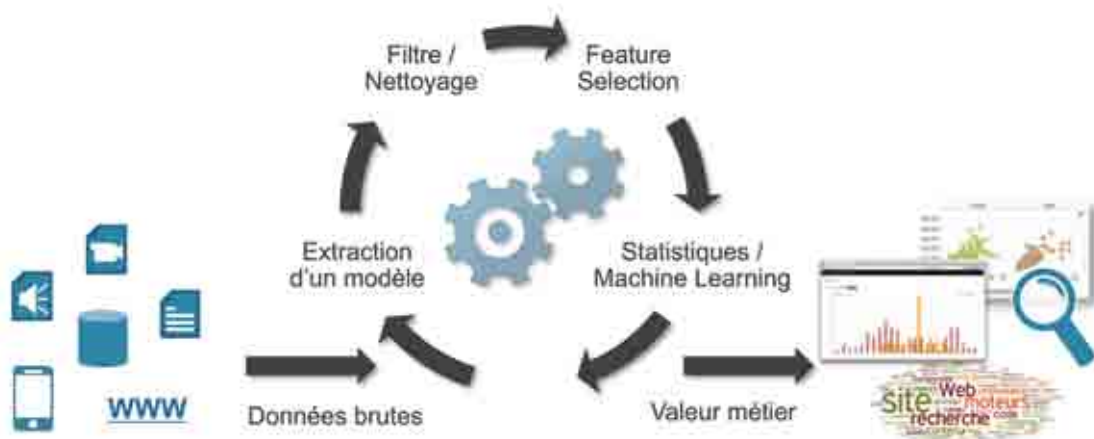
Le rôle du développeur - ingénieur est de savoir exploiter la puissance des outils Big Data et maîtriser les environnements distribués.



Le rôle de l'Ops est de vous aider dans la mise en place d'une plateforme Big Data robuste et performante, dans votre DataCenter ou sur le Cloud.

Grâce à une méthode de projet Agile, l'équipe Big Data réalise vos projets de manière transparente et maîtrisée.

FastConnect vous accompagne dans tout le cycle de votre projet :








Il existe différentes solutions Big Data pour répondre à différents besoins et l'écosystème Hadoop est devenu une plateforme incontournable bientôt omniprésente dans les SI comme le sont devenus les bases de données, les serveurs d'applications et les plateformes décisionnelles.

Nos experts sont là pour vous aider dans la mise en place de votre solution Big Data, ainsi que dans la formation de vos équipes.



# Plus de données. Meilleurs résultats.

Vertica, la plateforme analytique temps-réel, conçue pour le Big Data.

-  **Real-time analytics**  
50x-1000x faster than traditional databases
-  **Open architecture**  
support for Hadoop, R and wide variety of BI and ETL tools
-  **Optimized storage**  
10x-30x more data per server
-  **Massive scalability**  
effectively unlimited number of nodes, petabyte (and beyond) scale
-  **Flexibility**  
deployment via industry standard hardware, optimized appliances, or private/public cloud

Découvrez pourquoi des milliers de clients utilisent Vertica.

**Enregistrez-vous pour un test gratuit de 30 jours sur [vertica.com/evaluate](http://vertica.com/evaluate)**



## COORDONNÉES

### HEWLETT-PACKARD FRANCE

20 Quai du Point du Jour  
92100 Boulogne-Billancourt

Tél : **0820 211 211**

Fax : N/A

Mail : [vertica-France@hp.com](mailto:vertica-France@hp.com)

## CONTACT

**Frédéric  
LEMEILLE**

*Sales Manager Vertica*

## HP Vertica Analytics Platform

### La plate-forme d'analyse de nouvelle génération spécialement conçue pour tirer le maximum des Big Data

Issue des travaux de recherche de Michael Stonebraker, le créateur d'Ingres et de PostgreSQL, HP Vertica Analytics Platform est une base de données analytique en temps réel capable des plus hautes performances à ce jour.

Elle intègre des fonctions analytiques « in-database », une intégration étroite avec Hadoop et R, un déploiement rapide et une administration minimale de la plate-forme.

Vertica est bâti sur une architecture orientée colonnes tout en supportant le langage SQL, tant pour le chargement que pour les requêtes. Ainsi, il n'est pas nécessaire d'apprendre un langage spécifique et tous les outils amont et aval (ETL, décisionnels et Front-end) sont utilisables.

Avec Vertica, nos clients obtiennent désormais des réponses de 10 à 1000 fois plus rapides qu'avec des bases traditionnelles. Chez Kansys, un leader US du « Big Data as a Service », une requête qui prenait précédemment 18 heures s'exécute maintenant en 10-15 secondes. Chez Guess?, la marque de mode, les temps de requêtes complexes sont de 60 à 80 fois plus rapides qu'avant Vertica.

Le chargement des données est aussi extrêmement rapide et les données chargées sont accessibles de suite pour les traitements analytiques, sans nécessité d'indexation, 24h sur 24, 7 jours sur 7. Zynga, un leader mondial des jeux en ligne, capture les événements issus des interactions des joueurs et injecte ainsi dans Vertica 60 milliards de lignes, représentant 10 Téraoctets de données semi-structurées, par jour.

Le taux de compression peut atteindre 90%, réduisant considérablement les coûts de stockage, en utilisant des architectures de stockage standards. KDDI, opérateur mobile japonais, charge dans Vertica 8 millions d'enregistrements d'appels (CDR) par seconde.

Même si HP propose des appliances configurées pour Vertica, la solution repose sur l'architecture x86 et les unités de stockage standards.

**Quelques clients Vertica :** AOL, AT&T, Groupon, Mozilla, Twitter, Guess, Comcast, Level3, Sprint, Verizon, First American, The Washington Post...

**Pour en savoir plus :** Testez et déployez Vertica Community Edition, une version gratuite et fonctionnellement complète de Vertica Enterprise Edition, seulement limitée à un téraoctet de données et trois noeuds.

Rendez-vous sur <http://vertica.com/community/eDM> ou taggez ce code pour obtenir Vertica Community Edition ! »

# HP Vertica 6.1 "BULLDOZER"

Faster Analytics. Enhanced Hadoop Integration.  
Simplified Cloud Deployments



# Interview Corinne Baragoin et Béatrice Cochard, IBM

## 1) Les outils Big Data deviennent un complément indispensable des entrepôts de données au sein des entreprises : qu'apportent-ils en plus ? Qu'est ce qui a changé par rapport à ces datawarehouses traditionnels ?

Pour moi, les datawarehouses traditionnels ne permettent pas de répondre aux nouveaux besoins métier. Ils présentent des limitations majeures et se montrent contraignants pour l'utilisateur.

Tout d'abord, le gros défaut du datawarehouse est le fait qu'il n'exploite pas l'ensemble des données disponibles dans et hors de l'entreprise : ni les données non structurées (ex : texte libre, et au-delà audio et vidéo) ni les données semi-structurées provenant de logs ou de capteurs... c'est donc une perte de valeur importante que les outils de Big Data ont permis de pallier en élargissant le périmètre par l'exploitation de ces nouvelles informations.

Ensuite, concernant leur aspect contraignant et monolithique... Impossible de prendre en compte de nouvelles demandes métier sans bouleverser tout l'entrepôt et en revoir le modèle ! Cela implique une latence forte pour les utilisateurs, le temps nécessaire à l'IT pour prendre en compte la modification du modèle et son alimentation.

Le Big Data a profondément bousculé cela puisqu'on peut s'appuyer sur une vision purement exploratoire pour tirer parti des données... C'est donc un vrai changement de paradigme !

## 2) Justement, vous parlez de la variété des données... Est-ce un sujet nouveau chez IBM ?

Pour tout dire, cela fait longtemps que nous travaillons sur ce sujet, depuis le début des années 2000. Il est apparu évident qu'il faudrait analyser assez rapidement des vidéos, les textes, etc. qui se développaient sur les différents supports numériques, avec un haut niveau de granularité.

Aujourd'hui, on parle beaucoup de « données non-structurées » : je préfère parler d'informations « multi-structurées ». L'avantage du Big Data est justement de pouvoir exploiter tout type de donnée et de les analyser pour les transformer en informations pertinentes pour l'entreprise, et ce quelle qu'en soit la structure.

## 3) Partant de ce constat sur l'intérêt du Big Data et les limites des datawarehouses, quelles technologies avez-vous mises en place pour répondre aux défis évoqués ?

Face aux limites des entrepôts, nous sommes convaincus qu'il faut repenser les architectures de données selon 3 briques : l'une permettant l'analyse d'historiques de données multi-structurées à bas coût appelée InfoSphere BigInsights, une autre permettant l'analyse en temps réel appelée InfoSphere Streams et nous avons depuis longtemps vu l'importance d'enrichir les outils d'analyse par des outils plus fins liés à l'exploration de données, comme InfoSphere Data Explorer. Et bien évidemment à l'entrée de toutes ces briques, nous nous assurons de la qualité de l'intégration des données et de leur gouvernance, par des outils d'analyse dédiés.

## 4) Pourriez-vous nous décrire la brique InfoSphere Streams plus précisément ?

InfoSphere Streams traite des flux de données multi-structurées (données structurées, non-structurées, audio, vidéo) en temps réel. Au-delà de sa performance sur de grandes volumétries, sa très faible latence (microseconde) et ses capacités d'optimisation dynamique des flux, InfoSphere Streams a l'énorme qualité d'exécuter tout type d'analyse, qu'elles soient prédictives tirées d'outils de data mining, de moteur de règles, de fonctions de time series, etc.

## 5) Et pour BigInsights, diriez-vous qu'il est une déclinaison d'Hadoop afin de faciliter l'usage de celui-ci ?

InfoSphere BigInsights s'appuie effectivement sur le code open source de Hadoop et l'enrichit pour en faire une version prête pour l'entreprise sur les aspects performances, accélérateurs analytiques, déploiement de nouveaux services, facilités d'administration, etc... Pour cela, InfoSphere BigInsights complète la version open source par de nombreux composants : des connecteurs pour l'intégrer au SI de l'entreprise, des outils d'installation simplifiée et unifiée, des fonctions d'administration système ou d'accélération de l'analyse de tout type de données (réseaux sociaux, logs, capteurs, etc.).

## 6) Passons maintenant au challenge le plus important : l'intégration dans l'entreprise. Diriez-vous que vous avez un rôle de facilitateur du Big Data auprès des entreprises ?

Notre approche n'est pas seulement technologique. Nous partons des besoins métiers, nous demandons à nos clients quel type d'informations ils auraient besoin de mieux exploiter, et nous leur présentons des expériences clients déjà vécues. Ce qui est également important, c'est de faire un pilote avec l'IT et les métiers afin d'appréhender les défis technologiques avec une vision à 360° et d'adapter la plateforme aux besoins spécifiques du client.

Pour IBM, les technologies sont prêtes : le frein est plus souvent du côté des équipes des clients qui manquent parfois de connaissances pour les appréhender. Il faut également faire évoluer l'organisation de l'entreprise pour mettre en place ces outils, ce n'est pas toujours évident... C'est pourquoi un temps de formation est nécessaire et nous le proposons au travers de nos workshops.

## 7) Vous adressez-vous de la même façon aux petites et aux grandes entreprises, qui n'ont pas forcément les mêmes volumes de données ?

Pour IBM, la question du Big Data n'est pas uniquement celle du volume : on peut être une

petite entreprise avec des volumes de données inférieurs à des téraoctets, et avoir des problématiques de Big Data. A partir du moment où l'on décide de se démarquer de sa concurrence par les données, où l'on a compris que les informations générées à l'extérieur de l'entreprise étaient fondamentales, où l'on réalise l'apport du temps réel dans son contexte pour réduire la latence d'exploitation, alors on se trouve face à une démarche de Big Data et notre rôle est d'accompagner cette démarche.

## 8) Si l'on revient aux technologies que vous proposez, sur quoi travaillez-vous actuellement ? Quels sont les outils que vous allez mettre prochainement à disposition en terme de stockage ou d'analyse ?

En terme de stockage et de solution, notre travail se concentre essentiellement sur les appliances, en plus de solutions purement logicielles. Nous allons sortir au 2e semestre 2013 l'appliance « PureData Systems for Hadoop » : accélérateur pour le déploiement de solutions Hadoop qui intégrera InfoSphere BigInsights et permettra une installation rapide et un temps de mise en œuvre réduit.

Mais c'est essentiellement sur le domaine de l'analyse avancée de l'information pour en tirer du sens pour l'entreprise que nous progressons. Au-delà de solution de BI et d'analyse prédictive, nous travaillons au développement de systèmes à même de tirer plus de sens de l'information (analyse sémantique) et de la contextualiser, voire des systèmes d'intelligence cognitive. Par exemple, trouver un diagnostic basé sur des évidences, à partir de réponses de patients : c'est ce sur quoi nous travaillons avec Watson en « deep QA » dans le domaine de la santé !

## 9) Plus généralement, quels sont les défis auxquels vous pensez devoir faire face dans les années à venir ?

Les principaux défis sont probablement ceux liés au fait de placer l'information au centre des SI de demain et de repenser les architectures de données avec ces nouvelles technologies.

Il faudra répondre à la question de leur intégration dans l'entreprise, à leur gouvernance, mais aussi à la sécurité, la qualité et puis, bien entendu, penser à la formation des équipes ! Le Big Data c'est un mouvement de fond pour les entreprises. Elles ont à faire face au défi majeur de leur organisation : tirer avantage des informations pour être plus compétitives. Il est donc crucial d'avoir une approche globale, pas uniquement fondée sur des concepts informatiques avec des solutions qui intègrent toutes les dimensions de l'entreprise.



**Corinne Baragoin**  
Certified IT  
Specialist Big Data



**Béatrice Cochard**  
Big Data Technical  
Sales Manager





## COORDONNÉES

### IBM FRANCE

17 avenue de l'Europe  
92275 Bois Colombes

Tél : **+33 (0) 1 58 75 00 00**

Fax : N/A

Mail : [isabelle\\_lefevre@fr.ibm.com](mailto:isabelle_lefevre@fr.ibm.com)

## CONTACT

**Isabelle  
LEFEVRE**

*Marketing Manager Big Data*

IBM est un acteur technologique au service de l'innovation et de la transformation des PME, des groupes internationaux et des administrations. IBM participe à créer une planète plus intelligente pour permettre aux organisations qu'elle sert d'optimiser leur positionnement dans un monde en perpétuelle évolution.

## OFFRE IBM BIG DATA

Big Data permet de comprendre, avec plus de précision que jamais, comment fonctionne notre monde afin de produire une plus grande croissance économique et du progrès social. Il constitue une opportunité d'obtenir des connaissances sur des types de données et de contenus nouveaux, afin de rendre les entités plus agiles. Plateforme Big Data IBM – IBM seul propose une plateforme complète qui permet de résoudre les défis Big Data qui se posent aux organisations. Le principal bénéfice d'une telle plateforme est de tirer parti de composants pré-intégrés afin de réduire le coût et le temps d'implémentation.



<b>IBM DB2 10.5 with BLU Acceleration</b>	DB2 10.5 with BLU Acceleration est une base de données en mémoire orientée colonne. C'est une option de paramétrage du nouveau moteur de base de données DB2 10.5 qui permet à la fois de gérer dans une même instance de base de données des tables relationnelles, mais aussi des données stockées 'à plat' et exploitées sous forme de tableau en mémoire.
<b>IBM InfoSphere Streams</b>	IBM InfoSphere Streams est une plateforme de traitement qui permet d'analyser des données en mode flux à très haut débit. Optimisée pour le trafic intense, IBM InfoSphere Streams est capable d'absorber, de filtrer, d'analyser et de corréler des volumes massifs de flux de données structurées ou non structurées, tels que : <ul style="list-style-type: none"> <li>• les fichiers texte, fichiers de tableur, images, enregistrements vidéo et audio.</li> <li>• les Emails, contenus de dialogues en ligne, SMS, trafic web, blogs et sites de réseaux sociaux.</li> <li>• les transactions financières, données de scanner de police, rapports de systèmes et d'applications.</li> <li>• les données satellite, données GPS, enregistrements de lecteurs de badges et d'accès.</li> </ul>
<b>IBM InfoSphere BigInsights</b>	IBM InfoSphere BigInsights est un produit qui s'appuie sur la technologie Hadoop. Conçu comme une solution intégrée, il offre des fonctions de gestion, de sécurité et de fiabilité nécessaires à un déploiement à grande échelle. Ses fonctions de recherche, de visualisation et d'analyse permettent de mettre en place rapidement des fonctions analytiques puissantes.
<b>IBM InfoSphere Data Explorer</b>	IBM InfoSphere Data Explorer permet d'exploiter les nombreuses sources d'information et leur mise à jour en temps réel. Cette solution permet une vision unique des données du Data Warehouse, des applications d'entreprise, des applications de gestion de contenus et du web. Une interface Web, centrée sur l'information, permet de rechercher et de naviguer dans les Big Data à travers des sources fédérées.
<b>IBM InfoSphere Master Data Management &amp; Information Server</b>	La bonne gouvernance des données mais aussi la cohérence de ces données sont des propositions de valeur qu'IBM souhaite apporter à ses clients. IBM propose des solutions complètes, qui permettent d'identifier, de nettoyer, de transformer et distribuer vos données sans obstacle et de manière sécurisée dans l'ensemble de l'entreprise. Dans le domaine du MDM, IBM propose une plateforme disposant d'un puissant moteur de corrélation qui résout à la fois les rapprochements entre objets et les relations entre objets.
<b>PureData for Analytics</b>	Appliance décisionnelle haute performance conçue pour l'analytique sur des volumes de données sans cesse croissants avec plus de simplicité, de rapidité et une meilleure accessibilité. Réduit le temps d'exécution d'heures à la minute sur des Petabytes de données.



# INTERSYSTEMS

**BIG DATA? ALL DATA!**  
Tirez facilement le bénéfice de  
toutes les données disponibles



**InterSystems All Data: collectez, stockez, exploitez sans limite** toutes vos données: internes, externes, web, réseaux sociaux, ...

Donnez du sens aux masses de données structurées et non structurées, et dégagez rapidement de la valeur métier pour vos utilisateurs.

## COORDONNÉES

### INTERSYSTEMS FRANCE

Tour Europlaza - La Défense 4  
20 avenue André Prothin  
92400 Courbevoie

Tél : +33 (0) 1 77 49 16 00

Fax : +33 (0) 1 47 76 02 63

Mail : Robert.Bira@InterSystems.com

## CONTACT

**Robert  
BIRA**

*Market Development France*

## Relevez le défi du Big Data... Tout de suite!

La problématique fondamentale des entreprises à l'heure actuelle est d'essayer d'exploiter toutes les données à leur disposition pour en tirer un réel avantage concurrentiel dans leurs métiers.

Toutes les données, ou ce que nous appelons le 'All Data', cela veut dire principalement trois types de données :

- Les données structurées à l'intérieur de l'entreprise, a priori elles sont bien maîtrisées et ce depuis longtemps,
- Les données non-structurées à l'intérieur de l'entreprise (emails, contrats, documents, intranet...) représentent un gisement important en fonction des métiers, et sont peu ou mal exploitées,
- Les données non-structurées à l'extérieur de l'entreprise, le grand inconnu, tant en termes de volumes, que de qualité, et de pertinence pour le métier.

Cette notion de 'All Data' est évidemment une variante du Big Data tant toutes ces données peuvent représenter des volumes considérables, une variabilité de sources et de formats extrême, et enfin un besoin de traitement rapide et performant si on veut rendre tout cela exploitable. Face à ces trois sources qu'il faut réconcilier, les sociétés se retrouvent donc confrontées aux «3 V» bien exprimés par les analystes du Big Data: Volume, Variabilité, Vitesse

Ce challenge technologique, mais aussi métier, est résolu nativement par les solutions InterSystems, originellement issues du MIT, qui permettent de mettre en œuvre un projet Big data sans rien sacrifier : la performance extrême est au rendez-vous sur des technologies standard (performance de niveau base mémoire tout en restant sur des technologies disques 'classiques'), ouverte maximale sur toutes les données structurées et non structurées qui restent ou deviennent accessible une SQL, volumes traités virtuellement illimité de part la technologie de stockage même.

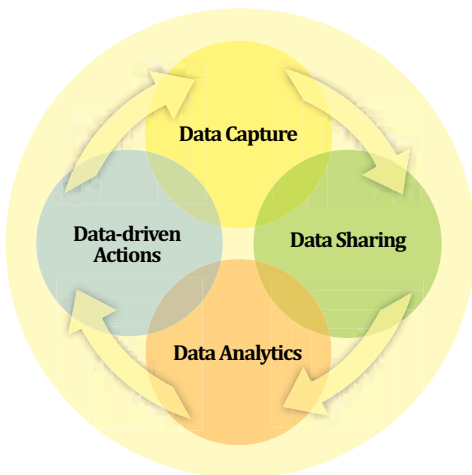
Au cœur de nos technologies: InterSystems Caché®, base de données NoSQL/objet/relationnelle ultra performante, supporte tous les formats de données: aussi bien les structures simples que complexes, les modèles relationnels, Objet ou NOSQL et est capable d'exploiter les données structurées et non structurées de manière intelligente. Grâce à InterSystems DeepSee® elle offre en outre le décisionnel intégré et en temps réel sur tous les formats de données. Le moteur d'analyse sémantique iKnow permet enfin de générer de l'information pertinente à partir de données non structurées.

Au niveau projets BIG DATA, InterSystems Caché est notamment au cœur du projet GAIA de cartographie de la Voie Lactée (plusieurs centaines de To avec forte contrainte de temps de traitement), du système de trading du Crédit Suisse (forte contrainte de performance temps réel hauts volumes), de solutions de «Smart Metering» en Europe, ainsi que de très nombreuses entreprises qui exploitent les immenses volumes de données textuelles du web pour leur enjeu métiers (marketing, presse...).

Au delà du Big Data, InterSystems propose à ses clients une démarche vers le 'All Data', ce que nous appelons «libérer les données».

InterSystems est, depuis plus de 30 ans, l'un des principaux éditeurs mondiaux de logiciels novateurs dans le domaine des bases de données, du développement rapide, et de l'intégration d'applications.

Les technologies InterSystems sont reconnues et mises en œuvre dans tous les secteurs d'activité, adoptées par plus de 2500 éditeurs, mises en œuvre par les plus grands intégrateurs et SSII, et plus de quatre millions de personnes à travers le monde utilisent chaque jour des applications stratégiques s'appuyant sur l'environnement logiciel d'InterSystems.



**Hadoop**

**NoSQL/NewSQL**

**Analytic DBMS**





## COORDONNÉES

### JASPERSOFT SARL

3 rue Colonel Moll  
75017 Paris

Tél : **+33 (0) 9 70 44 40 19**

Fax : **+353 1 686 5249**

Mail : [gcarbonnel@jaspersoft.com](mailto:gcarbonnel@jaspersoft.com)

## CONTACT

### Georges CARBONNEL

*Responsable Enterprise Accounts,  
France*



Jaspersoft permet à des millions de personnes de prendre chaque jour, de meilleures décisions plus rapidement, en leur donnant accès à des données utiles et exploitables au cœur de leurs applications et processus métier. Sa plateforme d'outils de reporting et d'analyse intégrés offre à chaque utilisateur une autonomie réelle et les réponses dont il a besoin. Elle s'adapte à toutes les architectures et à tous les budgets afin de servir l'ensemble de l'entreprise. Grâce à une communauté forte de plusieurs centaines de milliers de membres, les logiciels open source de Jaspersoft comptent des millions de téléchargements. Ses solutions décisionnelles sont intégrées au cœur de centaines de milliers d'applications et de processus métier d'où Jaspersoft – The Intelligence Inside. Jaspersoft est une société privée qui possède des bureaux dans le monde entier.

### Pourquoi Jaspersoft ?

Le décisionnel libre-service pour tous

Jaspersoft propose une suite décisionnelle complète de bout en bout, en libre-service pour :  
TOUS LES UTILISATEURS : En fournissant des rapports interactifs non pixellisés, des tableaux de bord et des analyses ad-hoc. Jaspersoft offre un décisionnel libre-service intuitif aux utilisateurs occasionnels et expérimentés.

TOUTES LES QUESTIONS : Grâce à un moteur en mémoire et une couche évoluée de méta-données, les utilisateurs reçoivent des réponses ultrarapides à toutes leurs requêtes.

TOUS LES APPAREILS : Les utilisateurs peuvent accéder à toutes les fonctionnalités décisionnelles depuis leur bureau ou en déplacement sur smartphone ou sur tablette.

### Une plateforme évolutive pour les créateurs de décisionnel

La plateforme décisionnelle Jaspersoft permet aux administrateurs du décisionnel de s'adapter facilement pour mieux prendre en charge :

TOUS LES UTILISATEURS Le faible coût de l'abonnement et la rapidité d'implémentation permettent aux entreprises de toute taille de se lancer rapidement et de passer à des milliers d'utilisateurs de manière économique.

TOUTES LES INFRASTRUCTURES Flexible et entièrement basée sur les standards Web, l'architecture de Jaspersoft peut fonctionner en version autonome ou s'intégrer facilement à n'importe quelle application sur site ou sur environnement Cloud grâce aux API étendues et aux thèmes CSS.

### Conçue pour le nouveau monde des TI

La plateforme décisionnelle moderne de Jaspersoft est conçue pour suivre l'évolution du monde des TI, en particulier :

### LE BIG DATA

Jaspersoft est la seule plateforme décisionnelle capable de créer des rapports et des analyses, directement depuis des sources Big Data tels que Hadoop, MongoDB et Cassandra ainsi que les principales bases de données analytiques comme HP Vertica, IBM Netezza, EMC Greenplum et bien d'autres. Jaspersoft a déjà travaillé sur des projets Big Data au sein de groupes comme Groupon, Ericsson, Vimeo, Jagex et SoftServe entre autres. Pour plus d'informations, consultez [www.jaspersoft.com/bigdata](http://www.jaspersoft.com/bigdata).

### LE CLOUD

La plateforme décisionnelle entièrement mutualisée de Jaspersoft offre des fonctionnalités SaaS pour les entreprises ainsi que des fonctionnalités PaaS pour les développeurs. Elle est la première à s'intégrer aux plateformes Cloud leaders tel qu'Amazon, Red Hat et VMware.

### LE MOBILE

Jaspersoft propose le seul framework open source du marché pour le décisionnel mobile sur les appareils iOS et Android en tant qu'applications natives et sur navigateur, grâce à la technologie HTML 5.

*Contacts: Anthony Pangnanouvong, Business Development, France  
[apangnanouvong@jaspersoft.com](mailto:apangnanouvong@jaspersoft.com)*





## COORDONNÉES

### KEYRUS

155 rue Anatole France  
92300 Levallois-Perret  
France

Tél : **+33 (0) 1 41 34 10 00**

Fax : **+33 (0) 1 41 34 10 10**

Mail : [bigdata-keyrus@keyrus.com](mailto:bigdata-keyrus@keyrus.com)

## CONTACT

**Reda  
GOMERY**

*Directeur des Opérations*

Fort d'une quinzaine d'années d'expérience au service de la maîtrise et de la valorisation de la donnée, Keyrus aide les entreprises à fournir à chacun de leurs collaborateurs les informations et la motivation nécessaires au développement de leurs activités et de leurs équipes.

S'inscrivant dans la continuité de cette mission claire et partagée, Keyrus propose une capacité unique de réalisation de solutions analytiques essentielles aux enjeux de performance des entreprises. Pour ce faire, Keyrus a réuni des compétences métiers – fonctionnelles et sectorielles –, et des expertises analytiques et technologiques travaillant de façon totalement intégrée.

Notre approche unique de proof-of-value vous permet ainsi de comprendre en quelques semaines le plein potentiel des nouveaux modes de valorisation de la donnée et d'entreprendre le chemin permettant de les intégrer au sein d'un portefeuille analytique au service de la découverte des leviers de la performance économique.

"Dans un monde de plus en plus dématérialisé où les clés de la compétitivité et de la croissance font appel à une connaissance plus fine des marchés, des clients et de leurs dynamiques, Keyrus se positionne comme le partenaire naturel de ses clients dans la transformation de leur actif informationnel en valeur économique durable.

S'il est en effet à présent largement admis que les données, sous ses formes et ses origines les plus diverses, constituent un actif majeur des organisations dans toutes les industries et que leur rôle de catalyseur de la capacité d'anticipation et de l'agilité opérationnelle ne fait plus aucun doute, les enjeux dessinés par leur exploitation n'en restent pas moins à la hauteur des opportunités qu'elle augure.

Conscient de l'importance du changement que les nouveaux paradigmes sous-tendent au sein d'un écosystème économique en pleine mutation et des questions bien concrètes que l'exploitation de la richesse et de l'accessibilité toujours croissante des données soulèvent, Keyrus vous accompagne dans la mise en œuvre des organisations et des solutions permettant de répondre aux enjeux économiques modernes d'un monde devenu numérique.

Fort d'une capacité unique à mobiliser conjointement l'expertise métier, analytique et technologique indispensable au succès des ambitions analytiques et Big Data de ses clients, Keyrus répond aux défis d'aujourd'hui par des modalités bien tangibles et pragmatiques, ancrées dans la réalité de l'entreprise et ciblées sur des projets concrets et accessibles, construisant ainsi les fondations d'un avantage concurrentiel dans la durée.



# WHEN YOUR JOB DEPENDS ON HADOOP



LET MAPR WORK FOR YOU  
[WWW.MAPR.COM](http://WWW.MAPR.COM)

MapR delivers on the promise of Hadoop with a proven, enterprise-grade Big Data platform that supports a broad set of mission-critical and real-time production uses.

MapR Europe du Sud & Benelux

171 bis av. Charles-de-Gaulle - 92 200 Neuilly/Seine - France

[sales-southEU@mapr.com](mailto:sales-southEU@mapr.com)



## COORDONNÉES

### MAPR TECHNOLOGIES

MapR Europe du Sud & Benelux  
171 bis avenue Charles de Gaulle  
92200 Neuilly sur Seine

Tél : +33 (0) 1 70 91 86 49

Fax : N/A

Mail : sales-southEU@mapr.com

## CONTACT

**Xavier  
GUERIN**

*VP Southern Europe and Benelux*



MapR est la distribution la plus à la pointe du secteur pour les solutions Apache Hadoop, faisant de la gestion et de l'analyse du Big Data une réalité pour un nombre croissant d'entreprises utilisatrices.

MapR contraste avec les autres distribution Hadoop, grâce à une architecture dédiée entreprise qui fournit aux clients une plate-forme fiable et très performante leur permettant non seulement d'analyser les données d'une manière nouvelle et puissante, mais aussi de créer de nouveaux produits et services pour leur clientèle.

MapR a considérablement perfectionné Hadoop en ajoutant à la plate-forme des extensions puissantes dédiées entreprise, la rendant plus simple à utiliser, plus fiable et beaucoup plus rapide que ce qui est proposé par les autres distributions Hadoop. Il existe plus d'une douzaine de fonctionnalités, disponibles uniquement chez MapR. Non seulement MapR supporte toute la gamme des API Hadoop, mais elle fournit également un accès NFS pour le partage de fichiers à travers le réseau ; ainsi, chaque application à base de fichiers peut accéder à des données du groupe sans réécritures à coder. De surcroît, la plateforme MapR fournit des améliorations des performances, celles-ci étant multipliées au moins par deux, voire même par cinq, et elle permet une fiabilité de MapReduce sans précédent. MapR englobe, entre autres, les suites communautaires open-source populaires comme HBase, Hive, Cascading et Zookeeper. En plus de ses fonctions particulièrement simples à utiliser, MapR permet aux clients de réduire le coût des matériels hardware nécessaires de moitié, autorisant ainsi un plus grand nombre d'entreprises et d'organisations de profiter de la puissance de l'analyse du Big Data et de devancer ainsi leurs concurrents.

MapR Technologies a connu une croissance rapide au niveau de la clientèle et de son expansion puisqu'elle opère désormais en Europe, au Moyen-Orient et en Asie. MapR est la seule plateforme Hadoop disponible sur Google Compute Engine. Elle est également présente dans Amazon Elastic MapReduce sur Amazon.com.

Grâce à sa technologie de pointe et sa réputation de pionnière dans l'environnement Hadoop, l'entreprise permet à des milliers de clients de mieux gérer et analyser leurs données issues du Big Data. Actuellement, la plateforme Big Data MapR est utilisée par des sociétés issues de la finance et de l'assurance, des médias, de la recherche et de la santé, de l'industrie, des administrations, des chaînes de commerce de détail, ainsi que par les principales entreprises Fortune 100 et Web 2.0, permettant à toutes ces sociétés ou organisations d'analyser des centaines de milliards d'éléments sur une base journalière, des données concernant 90 % de la population connectée à Internet sur une base mensuelle, et plus de mille milliards de dollars de transactions de commerce de détails sur une base annuelle. Le siège social de l'entreprise est situé à San Jose, Californie, USA, et elle a ses bureaux européens à Londres, Paris, Stockholm et Munich. Consultez [www.mapr.com](http://www.mapr.com) pour de plus amples renseignements concernant l'entreprise et ses produits.

Pour télécharger la version la plus récente de Apache Hadoop de MapR Distribution, veuillez vous rendre sur [www.mapr.com/products/download](http://www.mapr.com/products/download)





# Neo4j

## The World's Leading Graph Database

### La base de données de graphes Neo4j

- Neo4j vous permet de représenter les données connectées naturellement, en tant qu'objets reliés par un ensemble de relations, chacun possédant ses propres propriétés. La base de données de graphes, permet au développeur de commencer immédiatement à coder, car les données stockées dans la base font le parallèle direct avec les données elles mêmes.
- Comparé aux bases relationnelles, la base de données de graphe Neo4j peut être jusqu'à plusieurs milliers de fois plus rapide pour traiter les données associatives, tout en simplifiant considérablement les requêtes qui peuvent s'étendre plus facilement à de larges ensembles de données, car elles ne nécessitent pas de recourir aux coûteuses jointures du monde SQL. Les traversées utilisent un langage spécialement adapté pour parcourir les connections très simplement, et très rapidement. Vu qu'elles ne dépendent pas d'une modélisation rigide, Neo4j est plus propice pour la gestion de données changeantes et de schémas évoluant fréquemment.
- Les bases de données de graphes sont un outil puissant pour exécuter des requêtes de type parcours de graphe. La recherche du plus court chemin entre deux points du graphe permettant, par exemple, de résoudre facilement les requêtes sociales ainsi que de géographie et d'analyse d'impact.

Si vous vous arrachez les cheveux avec de nombreuses jointures et les procédures stockées complexes, il est fort possible que vous soyez en train de construire un graphe sans le savoir : car les graphes sont partout. Depuis les applications sociale web jusqu'à l'analyse d'impact sur un réseau Telecom en passant par la recherche en Biologie, les recommandations, la sécurité, et la détection de fraude, de nombreuses organisations ont adopté la bases de données de graphes Neo4j pour augmenter leurs capacités et leurs performances. C'est un système conçu pour gérer les données connectées que nous rencontrons de plus en plus au quotidien, qui est néanmoins fortement mûri avec plus de 10 ans en production.

Découvrez avec nous ce qu'apporte la puissance de Neo4j à une large variété de clients tels que HP, SFR, Cisco, Deutsche Telekom, Viadeo et Glowbl.

## COORDONNÉES

### NEO TECHNOLOGY FRANCE

Tél : **+33 (0) 8 05 08 03 44**

Fax : N/A

Mail : [cedric.fauvet@neotechnology.com](mailto:cedric.fauvet@neotechnology.com)

## CONTACT

**Cédric  
FAUVET**

*Business development France*

## Neo Technology

Les chercheurs de Neo Technology sont les pionniers de la base de données de graphes depuis 2000. Ils ont contribué à mettre la puissance des graphes à la disposition de nombreuses PME innovantes telles que Glowbl, Sodifrance ou Perigee, mais aussi d'organisations réputées mondialement telles que Cisco, HP, Deutsche Telekom, SFR, et Viadeo.

Ses 10 ans de service en production dans des environnements critiques pour le métier, ses partenaires Français et globaux et ses milliers de déploiements réussis ont fait de Neo4j le leader mondial de la base de données de graphe.

## Cas d'usages de Neo4j

### Analyse d'impact

Avoir la main sur son système et identifier en temps réel le client impacté par une maintenance ou une panne.

### Logistique

Calculer le meilleur chemin pour livrer un client.

Social, collaboration : Rechercher très facilement qui sont les amis de mes amis/

Recommandation : Définir en temps réel la liste des produits achetés par mes amis que je n'ai pas moi-même achetés.

### Master Data Management / Gestion de configuration

Construction d'un référentiel standardisé performant et sans redondance pour vos données critiques hiérarchisées : Hiérarchie d'entreprise et de produit.

### Droits d'accès

Gestion des groupes, utilisateurs et droits rapide et sans redondance.

### Géo-Spatial

Modélisation d'une carte routière et calculs d'itinéraires.

### Biologie, interactions moléculaires

Réduire les risques d'effets secondaires des médicaments en calculant en temps réel les interactions entre une protéine et une future molécule.,

## Bénéfices clés de Neo4j

### Des requêtes haute performance

Le modèle de données du graphe permet l'exécution des requêtes complexes avec une haute performance, inhérentes aux données connectées des applications d'aujourd'hui. En un mot le bon outil pour le bon besoin.

### Des projets livrés plus rapidement

Modéliser une base de données de graphes est aussi simple que d'écrire sur un tableau blanc. Les entreprises peuvent capturer rapidement toutes sortes de données, structurées, semi-structurées et déstructurées et ainsi les stocker dans Neo4j. Ceci résultant dans une réduction des temps de développement, une réduction de coûts de maintenance et de plus hautes performances.

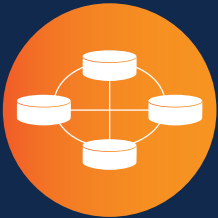
### La découverte de nouveaux cas d'usage

Très souvent les utilisateurs et clients de Neo4j démarrent un projet car ils n'ont pas pu trouver de solution à un problème bien précis avec les bases de données relationnelles. Puis ils découvrent, par la simplicité du modèle de données de graphes, de nouveaux cas d'usage très utiles dans le monde des startup ou des industries fortement concurrentielles.





# NEXT GENERATION REAL-TIME BIG DATA ANALYTICS



Highly Scalable,  
Distributed  
Database



Analyze  
Billions of  
Records



Filter  
Thousands of  
Columns



Sub-second  
Query-times  
without Cubes



Continuously  
Import  
Data



Thousands of  
Concurrent  
Queries





## COORDONNÉES

### PARSTREAM

La Grande Arche - Paroi Nord  
Paris La Défense - 92044

Tél : **+33 (0) 1 40 90 34 20**

Fax : N/A

Mail : [sales@parstream.com](mailto:sales@parstream.com)

## CONTACT

**Peter  
LIVAUDAIS**

*Sr. Director Solutions*



## ParStream est une plateforme innovante et primée de Big Data Analytics.

Créée en tant que technologie totalement nouvelle à partir de 2008, son index compressé haute-performance (HPCI) et son architecture état de l'art (share nothing, columnar) permet d'obtenir des temps de réponse en millisecondes sur des milliards d'enregistrements. Conçue dès l'origine pour tirer partie des environnement multi-cœur, ParStream passe à l'échelle linéairement et apporte l'infrastructure analytique la plus efficace en coût et en rendement énergétique.

ParStream a commencé comme défi technologique pour un groupe d'ingénieurs allemands : effectuer des opérations analytiques sur un jeu de données de 18 milliards d'enregistrements pour des centaines d'utilisateurs tout en fournissant des temps de réponse interactives. Afin d'obtenir de telles performances dans les contraintes serrées de coûts du e-commerce, des recherches importantes ont été effectuées conduisant à une percée technologique et un brevet pour l'indexation de grands ensembles de données.

ParStream se présente comme une base NewSQL permettant une prise en main aisée aussi bien par des équipes BI, des DBA et des développeurs. Elle est disponible sous forme logicielle et – ne requérant pas de matériel spécifique – s'intègre avec une empreinte économique et énergétique optimisée.

La version initiale de ParStream, montrée en 2010, a reçu le prix « One to Watch » du Emerging Companies Summit de Nvidia, ce qui lui a valu une attention immédiate et depuis soutenue d'acteurs telles que Red Herring (Top 100 Europe), Gartner (Cool Vendor) ou encore Computer Week Allemagne (Big Data Innovation). En 2013, ParStream a été classée le premier des Startup du Big Data parmi une sélection de leaders de l'industrie par CIO.com ParStream travaille en étroite relation avec le groupe « Database Systems et Information Management » de l'université de Berlin et en particulier avec le professeur Markl qui préside le conseil scientifique de la société aux côtés de Don Haderle (« le père de DB2 »), Marie-Anne Neimat (co-fondatrice de TimesTen et ex-VP du Développement d'Oracle) et Andy Palmer (co-fondateur et ex-CEO de Vertica). ParStream a ainsi accès à une ampleur et profondeur de connaissances sur le marché et la technologie analytique pour soutenir et développer son avance technologique.

ParStream est membre de plusieurs initiatives et projets du gouvernement fédéral allemand. Par exemple, ParStream est l'un des six membres du projet de recherche MIA qui est une place de marché Cloud pour l'information et l'analytique du Web Germanophone à l'initiative du BMWI (Ministre fédéral de l'Economie et de la Technologie). Egalement, Jörg Bienert le CTO de ParStream, est membre du Conseil sur la Nouvelle Economie Numérique du Ministre de l'Economie et de la Technologie et il est fréquemment membre des délégations internationales du ministre.

ParStream est soutenue par des fonds reconnus de la Silicon Valley tels que Khosla Ventures, Data Collective, Baker Capital, ... et a des clients initiaux tels que Coface Services en France. ParStream est basée à Cologne avec des bureaux à Hamburg, Paris ainsi qu'à Cupertino. ParStream est utilisée dans des applications variées qui vont de l'e-commerce à la recherche scientifique en passant par la supervision et l'optimisation d'infrastructures et de systèmes complexes.



# Bring Your Big Data to **Life**

Big Data Integration and Analytics

Optimisation d'Hadoop et plus encore.  
[Découvrir comment sur pentaho.fr](http://pentaho.fr)



## COORDONNÉES

### PENTAHO

Paris - France

Tél : 0800-915343  
Fax : 1 407 517-4575  
Mail : ocharlery@pentaho.com

## CONTACT

**Olivier  
CHARLERY**

*French Inside Sales*

Pentaho construit le futur de l'analytique. Son héritage Open Source mène son innovation continue pour une plate-forme analytique intégrée et moderne conçue pour les diverses exigences de l'analyse d'entreprise de demain, notamment celles du Big Data. L'analytique performante à moindre coût est facilement accessible grâce à la suite d'outils de Pentaho qui permet l'accès aux données, la visualisation, l'intégration et l'analyse de données ainsi que le datamining.

### **Pentaho permet d'exploiter plus rapidement les Big Data grâce à sa solution d'analytique pour Big Data la plus complète du marché. Pentaho offre :**

- Pentaho Big Data Analytics Center
- Une continuité totale depuis l'accès aux données jusqu'à la prise de décisions : une plate-forme complète analytique et d'intégration de données conçue pour n'importe quel entrepôt de données Big Data
- Un développement et une exécution plus rapides : développement visuel et exécution distribuée
- Une analyse interactive et instantanée : aucun codage ni compétences ETL requis Instaview

Pour répondre aux besoins spécifiques des analystes Big Data, Pentaho a créé Instaview, première application interactive et instantanée pour Big Data. Instaview convertit les données en analyses en trois étapes simples visant à réduire considérablement le temps de recherche, de visualisation et d'exploration de larges volumes de données complexes et diverses.

### **Pentaho facilite l'analyse des données Hadoop**

L'utilisation de Pentaho Business Analytics avec Hadoop offre une conception graphique hautement productive associée à une analyse et à une visualisation des données instantanées :

- Une plate-forme d'analyse Hadoop complète – développement visuel pour l'ingestion, la préparation, l'intégration et l'organisation des données Hadoop et visualisation, exploration et prédiction interactives
- Une meilleure productivité de l'équipe informatique grâce à la réutilisation des compétences existantes – facilement utilisables sans formation particulière par les informaticiens, les analystes d'entreprise et les scientifiques spécialistes des données
- Une protection contre les risques – fonctionne avec toutes les principales distributions Hadoop, ainsi que les bases de données NoSQL et les entrepôts de données traditionnels, en connectant Hadoop au reste de l'entreprise
- Productivité et performance : gain en productivité de développement et en performance d'exécution multiplié par 10 ou par 15 par rapport au codage manuel MapReduce, grâce à l'utilisation de l'unique moteur ETL qui s'exécute nativement comme une tâche MapReduce à travers le cluster Hadoop

### **Analyse et visualisation NoSQL**

Grâce à un support natif des bases de données NoSQL les plus utilisées, notamment Apache Cassandra/DataStax, HBase, MongoDB/10gen et HPCC Systems, Pentaho fournit :

- Une plate-forme d'analyse Hadoop complète : développement visuel pour l'ingestion, la préparation, l'intégration et l'organisation des données NoSQL ; visualisation, exploration et prédiction interactives
- Une création de rapports et une analyse par ordinateur et par utilisateur sur le nombre croissant de données générées, telles que contenu web, documents, réseaux sociaux ...
- Une meilleure productivité de l'équipe informatique grâce à la réutilisation des compétences existantes : outils facilement utilisables sans formation particulière par les informaticiens, analystes d'entreprise et scientifiques spécialistes des données
- Une protection contre les risques – fonctionne avec toutes les bases de données NoSQL de premier ordre, ainsi que les entrepôts de données traditionnels et Hadoop, en connectant les bases de données NoSQL au reste de l'entreprise
- Productivité et performance : gain en productivité de développement et en performance d'exécution multiplié par 10 ou par 15 par rapport au développement codé manuellement pour les bases de données NoSQL

Visitez le Pentaho Big Data Analytics Center : [www.pentahobigdata.com](http://www.pentahobigdata.com) pour en savoir plus sur Pentaho et Big Data, accéder à des articles techniques et vous tenir informé des dernières nouvelles et tendances du secteur.

Pour évaluer gratuitement "Pentaho Business Analytics" : [www.pentaho.fr/download/](http://www.pentaho.fr/download/)



# WHAT WILL YOU DISCOVER?



## QlikView donne tout son sens au Big Data

Le Big Data nous concerne tous. Il transforme le monde qui nous entoure. Grâce à la plate-forme QlikView de Business Discovery, vous pouvez exploiter toute sa puissance :

- **Analyser** des données issues de différentes sources depuis une seule et même application;
- **Explorer** naturellement les associations entre les données;
- **Visualiser** les données à l'aide de graphiques soignés et performants;
- **Accéder** à vos données à partir de périphériques mobiles pour les analyser;
- **Développer** la prise de décisions collaboratives en temps réel et de façon sécurisée

Découvrez comment QlikView permet à King.com, un acteur majeur des jeux en ligne (*Candy Crush Saga*) d'analyser le comportement de jeu de 40 millions de clients, grâce à l'exploitation des données stockées dans son cluster Hadoop, où deux milliards de nouvelles lignes sont créées chaque jour.

Téléchargez le livre blanc sur :  
[www.qlikview.fr/big-data](http://www.qlikview.fr/big-data)

**QlikView**



## COORDONNÉES

### QLIKTECH FRANCE

93 avenue Charles de Gaulle  
92220 Neuilly-sur-Seine

Tél : +33 (0) 1 55 62 06 90

Fax : N/A

Mail : infoFR@qlikview.com

## CONTACT

**Stéphane  
BRIFFOD**

*Director Presales*

QlikTech (code NASDAQ : QLIK) est un leader en matière de Business Discovery, la Business Intelligence (BI) conduite par l'utilisateur. La plate-forme QlikView de Business Discovery comble l'écart entre les solutions de BI traditionnelles et les applications inadaptées de feuilles de calcul. QlikTech a mis au point une technologie pionnière de recherche associative en mémoire et grâce à cet environnement de BI en libre-service, les utilisateurs peuvent explorer les informations librement, au lieu d'être cantonnés à une série prédéfinie de questions.

### **QlikView permet aux utilisateurs métier d'appréhender leur activité de façon unique :**

- Consolider l'ensemble des données issues de différentes sources dans une seule et même application;
- Explorer naturellement les associations entre les données;
- Visualiser les données à l'aide de graphiques soignés et performants;
- Accéder à vos données à partir de périphériques mobiles pour les capturer et les analyser;
- Développer la prise de décisions collaboratives en temps réel et de façon sécurisée.

Le déploiement de l'environnement d'analyse en libre-service de QlikView ne nécessite que quelques jours ou quelques semaines, et répond aux besoins des PME comme des plus grandes entreprises internationales.

Le modèle orienté applications de la plate-forme QlikView de Business Discovery est également compatible avec les solutions de BI existantes et offre une expérience collaborative basée sur l'immersion et intégrant une dimension sociale et mobile.

Plus de 29 000 sociétés à travers le monde ont déjà fait le choix de QlikView.



# ActivePivot - Enfin l'analyse multidimensionnelle tient ses promesses !

## Profondeur d'analyse

- Agrégation de grands volumes de données
- Calculs sophistiqués à partir de vos règles métier

+33 (0) 1 4

The ActivePivot logo, which consists of a red 3D pyramid shape with a white grid pattern on its faces, set against a light gray background. The text 'ActivePivot™' is written in white across the front face of the pyramid.

ActivePivot™

## Liberté d'analyse

- Pas de limite au nombre de dimensions
- Temps de réponse instantanés

## Performance

- Calculs en mémoire, en mode incrémental
- Rafraîchissement des données à la volée

Une base de données "mixed workload" au service de la prise de décision opérationnelle

[www.quartetfs.com](http://www.quartetfs.com)

PARIS • LONDRES • NEW-YORK • SINGAPOUR • HONG-KONG

## COORDONNÉES

### QUARTET FS

2 rue Jean Lantier  
75001 Paris

Tél : +33 (0) 1 40 13 91 00

Fax : +33 (0) 1 40 13 84 51

Mail : [nina.krause@quartetfs.com](mailto:nina.krause@quartetfs.com)

## CONTACT

**Nina  
KRAUSE**

*Sales Executive*

**Quartet FS répond aux besoins des entreprises qui souhaitent améliorer leur performance en mettant l'analyse des données en temps réel au service de la prise de décision opérationnelle.**

### ACTIVEPIVOT EN BREF

Puissant moteur d'agrégation et de calcul en mémoire, ActivePivot permet aux décideurs du quotidien d'analyser leur activité et d'en optimiser la performance en s'appuyant sur des indicateurs complexes générés à la volée à partir de données rafraîchies en temps réel.

ActivePivot est donc une solution innovante particulièrement adaptée aux enjeux analytiques du « Big Data ».

Combinant les traitements transactionnels et analytiques au sein d'un seul et même environnement tirant parti du traitement en mémoire, ActivePivot se distingue par la combinaison des caractéristiques suivantes:

- Agrégation de grands volumes de données hétérogènes de manière incrémentale
- Calcul à la volée d'indicateurs à partir de n'importe quel type de règle métier
- Navigation dans les données à différents niveaux d'agrégation
- Simulation de scénarii sur des données temps réel
- Choix de l'interface de visualisation: ActivePivot Live ou interfaces MDX/XMLA

### CAS D'USAGE

ActivePivot apporte de la valeur dans les processus nécessitant l'agrégation de grands volumes de données très volatiles, comme la supply chain, le risk management, le pricing dynamique ou encore la gestion de trafic. Dans ces contextes, ActivePivot permet d'adresser des cas d'usage inédits comme la gestion par exception, l'optimisation de ressources, l'analyse d'impact, la détection d'anomalies ou encore le stress testing.

#### • Transport: Visibilité de bout en bout sur la supply chain

Un enjeu Big Data dans le monde du transport est le pilotage de la chaîne logistique. Pour garantir les délais de livraison, les opérateurs logistiques ont besoin de faire la synthèse des informations qui remontent de leur supply chain, que ce soit des systèmes de planification, de TMS ou encore de gestion des entrepôts. ActivePivot fait la synthèse en continu de toutes ces données pour permettre aux opérateurs logistiques d'avoir une visibilité de A à Z sur leur supply chain. ActivePivot permet aux opérateurs de disposer d'une véritable tour de contrôle de leurs flux opérationnels.

#### • Finance: Analyse des risques en temps réel

ActivePivot est déployé dans de nombreux établissements financiers pour répondre à un besoin croissant d'analyse et de surveillance des risques en temps réel. Chez Lloyds Banking Group, les traders utilisent ActivePivot pour obtenir une visibilité en temps réel sur leurs positions en préalable à toute activité de trading. Chez Nordea, les risk managers utilisent ActivePivot pour calculer la Value at Risk en temps réel, une mesure qui nécessite l'agrégation incrémentale de grands volumes de données non linéaires.

#### • E-commerce: Pricing dynamique

Un cas d'usage typique pour ActivePivot dans le monde du e-commerce est la prise de décision en mode "stressé" ou en flux tendu, comme par exemple en période de solde. Lorsqu'il faut ajuster les prix en quasi temps réel, ActivePivot permet aux acheteurs de déterminer les références pour lesquelles une variation de prix aura l'effet de levier le plus fort sur les ventes sans pour autant détériorer la marge de l'ensemble du rayon.

### A PROPOS DE QUARTET FS

Créée en 2005 par un groupe de cinq entrepreneurs issus du monde applicatif, Quartet FS est une société à capitaux privés avec des bureaux à Paris (siège de la R&D), Londres, New-York, Hong Kong et Singapour. Avec plus de 50 projets en production dans des groupes d'envergure internationale, Quartet FS a pour clients des entreprises dans de nombreux secteurs d'activité parmi lesquels la finance, la logistique, le transport ou encore la grande distribution.





HIGH-PERFORMANCE

# ANALYTICS

Les Big Data ont trouvé leur maître.

Il existe une solution rapide et efficace pour valoriser la masse de big data que vous affrontez. SAS® High-Performance Analytics simplifie l'analyse des grandes volumétries de données pour résoudre des problèmes complexes et permet de prendre des décisions rapides et fiables grâce à la puissance d'analyse, à la capacité et la rapidité de traitement de quantités massives de données.



Découvrez tous les atouts de l'analytique haute performance sur  
[www.sas.com/france/hpa](http://www.sas.com/france/hpa)





## COORDONNÉES

### SAS INSTITUTE

Domaine de Grégy  
Grégysur-Yerres  
77257 BRIE COMTE ROBERT CEDEX

Tél : +33 (0) 1 60 62 11 11

Fax : +33 (0) 1 60 62 11 12

Mail : [comsas@fra.sas.com](mailto:comsas@fra.sas.com)

## CONTACT

**Serge  
BOULET**

*Directeur Marketing*

## SAS, votre atout performance

### Prenez de meilleures décisions plus rapidement

SAS vous aide à résoudre vos problèmes complexes, atteindre vos objectifs clés et exploiter plus efficacement vos données. Pourquoi ? Leader mondial du décisionnel et des solutions de business analytics, nous proposons des outils analytiques puissants, reconnus sur le marché, des solutions de pilotage de l'information et de business intelligence, pour vous aider à faire face aux challenges d'aujourd'hui et anticiper les opportunités de demain.

### Regardez vos objectifs prendre forme

Une des priorités de votre entreprise : savoir tirer parti des big data. L'analytique va vous aider à atteindre cet objectif, à vous permettre de percevoir qu'un détail peut faire la différence ou à prendre du recul pour détecter de plus grandes opportunités, de plus fructueuses perspectives. C'est toute la puissance apportée par l'analytique.

A cet instant même, les clients de SAS – parmi lesquels 90 des 100 premières entreprises du classement FORTUNE Global 500® de 2012, s'appuient sur nos solutions analytiques pour :

- Découvrir et saisir des opportunités de développement et de performance dans de multiples domaines.
- Prendre des décisions stratégiques, opérationnelles et tactiques en un rien de temps.
- Utiliser des données pour des décisions aux résultats tangibles.

### Avec SAS, c'est possible

Nous nous engageons sur le succès de vos projets sur le long terme. Nos équipes de consultants, formateurs, support 24/7, vous accompagnent tout au long du cycle de vie de votre projet. La communauté de nos clients et partenaires contribuent à cet accompagnement.

### Comment nous vous aidons

Vous avez besoin d'une solution spécifique adaptée à votre secteur ? Ou d'une solution fonctionnelle transverse qui va vous permettre de répondre à vos challenges ? Quelle que soit votre problématique, SAS a une réponse pour vous. Par exemple, nous aidons à combattre la fraude dans le secteur financier, à accélérer la mise sur le marché de médicaments ou identifier des opportunités de ventes croisées et additionnelles dans la distribution.

Nous aidons aussi nos clients à valoriser leur relation client, à piloter leurs risques et à optimiser leurs réseaux informatiques. Nos solutions répondent à ces problématiques complexes en s'appuyant sur un socle technologique avec trois composantes clé : la gestion des données, les outils d'analyse et la restitution de l'information.

Vous n'avez jamais pris de décisions aussi rapidement grâce à vos données. Avec SAS® High-Performance Analytics, vous n'avez plus besoin de longs processus pour analyser vos Big Data. Obtenez des résultats en quelques minutes ou secondes, et consacrez le temps ainsi gagné à votre business.

Nous avons aussi rendu la business analytics accessible et simple. Changez votre manière d'analyser vos données : avec SAS® Visual Analytics vous explorez visuellement et rapide-

ment vos données. Vous pouvez créer vos rapports et partager les résultats sur le web ou des tablettes Android® et iPad®.

Tous vos collaborateurs amenés à prendre des décisions, quel que soit leur métier ou leur niveau hiérarchique, peuvent désormais avoir accès à des analyses prédictives rapides et précises.

Votre succès est notre succès. Et c'est pourquoi nous vous aidons à déployer et utiliser les solutions SAS de la façon qui vous convient le mieux. Nous avons des offres flexibles pour répondre à vos besoins : cloud privé, hébergement, grid ou services web d'Amazon...

SAS Institute, société privée créée en 1976. Plus de 65000 sites clients répartis dans 135 pays, dont 2600 en France.



# Business Agility through Data Agility

Une **base de données virtuelle** sans modification du SI, ni centralisation des données.

- **Vue 360°** de vos **clients**, partenaires et fournisseurs.
- **Vue unifiée de systèmes d'information** hétérogènes lors de fusion ou de migration.
- Collecte de données **simple** et **agile** pour fouille et analyse.

semsoft

agility

accuracy

Lauréat  
Data Intelligence Awards  
2013



## COORDONNÉES

### SEMISOFT

80, avenue des buttes de Coesmes  
35700 Rennes

Tél : +33 (0) 2 99 12 19 86

Fax : N/A

Mail : Damien.regnault@semsoft-corp.com

## CONTACT

**Damien  
REGNAULT**

COO

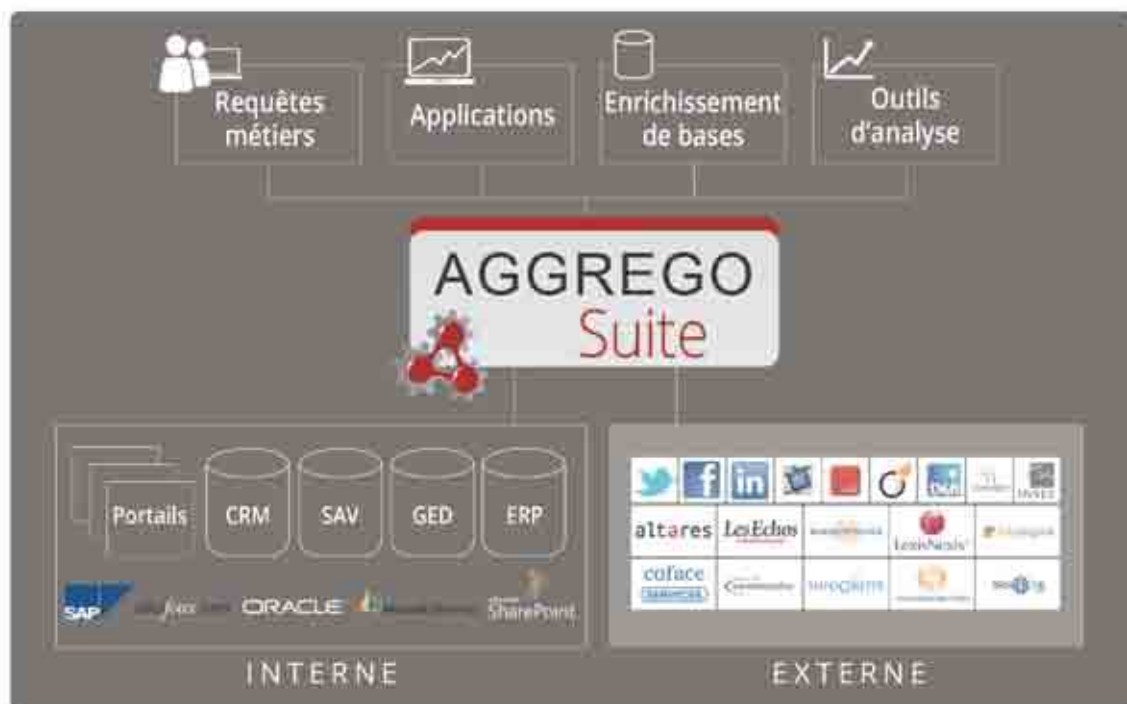
## Si votre performance dépend de l'agilité de l'accès aux données...

- Avec Aggrego Suite® de Semsoft, relevez le défi de l'intégration de données dans un contexte Big Data de manière simple et rapide tout en maîtrisant votre investissement (intégration limitée, pas de stockage, mode Cloud).

La solution Aggrego Suite© est une base de données virtuelle qui réconcilie des sources de données hétérogènes (internes / externes, structurées / non structurées) en les présentant de manière structurée et homogène.

Avec Aggrego Suite© vous valorisez avec agilité votre capital informationnel étendu sans intrusivité, ni centralisation, pour :

- Faire des requêtes métiers complexes,
  - Enrichir dynamiquement vos informations clients / produits / fournisseurs,
  - Collecter de manière simple des données pour analyse et Data mining.
- Aggrego Suite© est le fruit de plus de 10 ans de R & D conjointe avec l'INRIA, Orange Labs, l'Université d'Orsay et le CNRS autour des technologies du web sémantique et du VLDB (Very Large Database). Elle a déjà séduit un certain nombre de clients en France dont Orange, Total, Médiamétrie/EuroDataTV,...
  - La solution Aggrego Suite de Semsoft a été primée lors des Data Intelligence Awards 2013.



**ROI** Valorisation du capital informationnel étendu de l'entreprise vs. coûts d'intégration et d'exploitation faibles.

**Facilité** Mise en œuvre en quelques jours/semaines.

**Agilité** Ajout de sources et évolution du modèle instantanés, requêtes métiers complexes et contextuelles.





# NEW TO BIG DATA ? one smart ANSWER



smart foundations



## COORDONNÉES

### SENTELIS

31 Rue du pont  
92200 Neuilly-sur-Seine  
France

Tél : **+33 (0) 1 75 61 34 00**

Fax : N/A

Mail : [i.regnier@sentelis.com](mailto:i.regnier@sentelis.com)

## CONTACT

**Isabelle  
REGNIER**

*Directeur Commercial*

## smart SOLUTIONS for smart LEADERS

### Nos métiers

Stratégie / Gouvernance / Architecture

### Nos prestations

Industrialisation du SI  
Gouvernance orientée service de la DSI  
Conception d'architecture nouvelle génération  
Etude et mise en œuvre de socles SI industriels  
Etude et activation de pratiques DSI industrielles  
Marketing de la DSI

### Une vocation à accompagner les directions des systèmes d'information grâce à :

- une parfaite compréhension des tendances de fond sectorielles et des impératifs auxquels doivent répondre les décideurs
- une capacité à comprendre, conseiller et défendre des approches et solutions innovantes
- une exigence de haute qualité doublée de pragmatisme

### Un savoir-faire à forte valeur ajoutée fondé sur :

- une pratique conjuguée de la stratégie, de la gouvernance et de l'architecture des systèmes d'information
- des compétences interdépendantes en étude amont et mise en œuvre opérationnelle
- une forte expérience sur les composants et pratiques transverses (SOA, BPM, MDM, BI, BigData, Shared Services Centre, Amélioration continue...)
- une méthodologie dédiée à la mise en œuvre de ces fondations : smartfoundations™

### Une approche très personnalisée de nos clients qui repose sur :

- une compréhension approfondie de leur contexte métier et système d'information
- la recherche de solution « sur mesure »
- une équipe de talents complémentaires au sein d'un cabinet indépendant



with SENTELIS

without.



## TIREZ PROFIT DU BIG DATA DE VOTRE ENTREPRISE

### RECHERCHE ET ANALYSE EN TEMPS RÉEL POUR

- Constituer une vue à 360° d'un sujet ou d'un client
- Révéler des « réseaux sociaux implicites »

#### ANALYSE APPROFONDIE DE

Données en langues naturelles



Données de vos applications d'entreprise



INFORMATION PERTINENTE  
POUR OPTIMISER VOTRE PERFORMANCE

## COORDONNÉES

### SINEQUA

12 rue d'Athènes  
75009 Paris

Tél : +33 (0) 1 70 08 71 71

Fax : +33 (0) 1 45 26 38 92

Mail : [jeanrond@sinequa.com](mailto:jeanrond@sinequa.com)

## CONTACT

**Hans-Josef  
JEANROND**

*Directeur Marketing*

## Sinequa fournit une plateforme de Search & d'Analyse du Big Data en temps réel.

Sa solution s'adresse aux grandes entreprises et administrations publiques pour offrir à leurs utilisateurs un point d'accès unifié à l'information couvrant toutes les données textuelles et base de données. Son interface visuelle permet de découvrir de façon intuitive et conversationnelle l'information pertinente.

Big Data ne serait que « Big Noise » (grand bruit sourd) sans des capacités d'analyse puissantes pour extraire et visualiser les informations pertinentes.

Sinequa se distingue par cette capacité d'analyse approfondie pour extraire l'information pertinente à partir de toutes sortes de données : les données non structurées dans des documents, emails, sites Web et des réseaux sociaux, et les données structurées, dans les applications d'entreprise, telles que ERP, CRM, Finance, RH, etc.

L'analyse sémantique, développée pour 19 langues, permet de lever le plus grand défi du Big Data : traiter la masse des données textuelles ou « données générées par l'humain ». L'analyse sémantique permet à la fois de mieux comprendre les questions des utilisateurs, formulées en langage naturel, et le sens des textes, allant bien au-delà de la recherche de mots clés. Il contribue également à catégoriser des informations en utilisant des filtres spécifiques à un métier afin d'améliorer considérablement la découverte d'information. Sinequa offre une forte sécurité d'accès aux informations de l'entreprise avec un contrôle strict des droits d'accès de chaque utilisateur.

Plus de 120 connecteurs aux sources de données aident à faire face à la diversité inhérente au Big Data.

Nos clients utilisent notre plateforme de Search & d'Analyse du Big Data pour créer des applications spécifiques (Search Based Applications ou SBA) liées à leur cœur de métier, comme la création de vues à 360° (par exemple sur les clients) ou de révéler les réseaux sociaux implicites (réseaux d'experts sur un sujet donné), généralement avec un très fort ROI.

Les données sur les clients viennent souvent d'applications d'entreprises classiques comme le CRM, ERP ou systèmes financiers, des systèmes de gestion de documents contenant des contrats, des descriptions de produits et de services, de courriers électroniques et des échanges de centre d'appels.

Les données sur les experts et l'expertise peuvent provenir de publications internes et externes, rapports de projets, systèmes RH, RSE, etc. Sinequa ne fournit pas seulement une liste de noms comme résultat d'une recherche, mais le degré d'expertise de chacun ainsi que les documents relatifs à cette expertise et les relations entre experts. Ceci révèle les « réseaux sociaux implicites » ou la « matrice organisationnelle informelle » d'une entreprise, pas en fonction de la position hiérarchique, mais du travail accompli et de l'expertise détectée grâce à ce travail.

Dans chaque cas, de fortes capacités sémantiques sont nécessaires pour aller au-delà de la recherche simpliste par mots clés, pour découvrir les contenus traitant d'un sujet particulier, même si elles ne contiennent pas les mots clés utilisés dans une requête de recherche.

## EXTRACTING VALUE



## FROM ENTERPRISE DATA



# Quand on a **la passion...** ... l'innovation va plus loin.



**+20 ans** d'expérience  
au service de l'industrie



## **5 métiers** High Tech

- Conseil
- Ingénierie Système
- Ingénierie Physique
- Ingénierie Logiciel
- Testing



## **6 marchés** clés

- aéronautique,
- spatial,
- défense,
- énergie,
- ferroviaire,
- télécoms



**+20 000**  
collaborateurs

Sogeti dans 15 pays



## **4 orientations** R&D - innovation

- Incubateur de solutions
- Partenariats laboratoires
- Propriété intellectuelle
- R&D interne



Avec **3 000 collaborateurs** en France, et plus de **20 ans d'expérience**, Sogeti High Tech figure parmi les leaders du marché de l'Ingénierie et du Conseil en Technologies. Nous intervenons dans de nombreux projets d'innovation et d'ingénierie technologiques des plus grands groupes industriels mondiaux. Nos expertises techniques et métiers renforcent la valeur ajoutée de nos solutions au service des challenges du monde industriel du XXI<sup>e</sup> siècle.



**SOGETI**

High Tech

[www.sogeti-hightech.fr](http://www.sogeti-hightech.fr)





#### COORDONNÉES

#### SOGETI HIGH TECH

15 bis rue Ernest Renan  
92136 Issy-les-Moulineaux

Tél : **+33 (0) 1 40 93 73 00**

Fax : **+33 (0) 1 40 93 73 73**

Mail : [philippe.ravix@sogeti.com](mailto:philippe.ravix@sogeti.com)

#### CONTACT

**Philippe  
RAVIX**

*Directeur Innovation*

Analyse du risque, affinement du ciblage marketing, aiguillon pour la R&D et autres axes d'application de l'analyse prédictive sont bienvenus dans tous secteurs d'activité, en ces temps de rigueur économique, perçus comme de potentiels axes de croissance.

Donner accès, par exemple, aussi bien aux bureaux d'études qu'au service marketing, à la confrontation des données de conception et fabrication d'un produit avec celles issues de son exploitation, analyse des coûts induits incluse, afin d'optimiser le cycle de vie de ce produit et/ou alimenter la démarche d'innovation de l'entreprise. Tels sont en effet, d'un point de vue générique, les deux enjeux forts du Big Data : optimiser, dans une optique d'efficacité opérationnelle ; générer de nouvelles idées, propositions et débouchés, dans un objectif de croissance de revenus.

L'offre Big Data de Sogeti High Tech, forte des compétences de plus de 200 scientifiques et analystes spécialisés et d'un savoir-faire dans le domaine de la simulation, répond à ces besoins d'analyse, expressions des besoins d'optimisation et d'innovation des industriels. L'offre Big Data de Sogeti High Tech s'adresse aux industriels de tous secteurs selon une déclinaison de services incluant :

- le conseil avec notamment l'animation d'ateliers sur fond d'études de cas, visant à cerner l'impact potentiel d'un projet, sa nature, son périmètre, son organisation, les ressources impliquées,
- le développement de modèles d'analyse,
- la conduite de projet menant à la mise en place de systèmes Big Data, de l'infrastructure aux modèles d'analyse de données sur-mesure, et, en bout de chaîne, les modèles prédictifs ou prospectifs.

Sogeti High Tech, filiale du groupe Capgemini, est un leader du marché de l'Ingénierie et du Conseil en Technologies.



# NOTHING SPLUNKED, NOTHING GAINED.

Start with big data and Splunk® software. End with an unfair advantage.

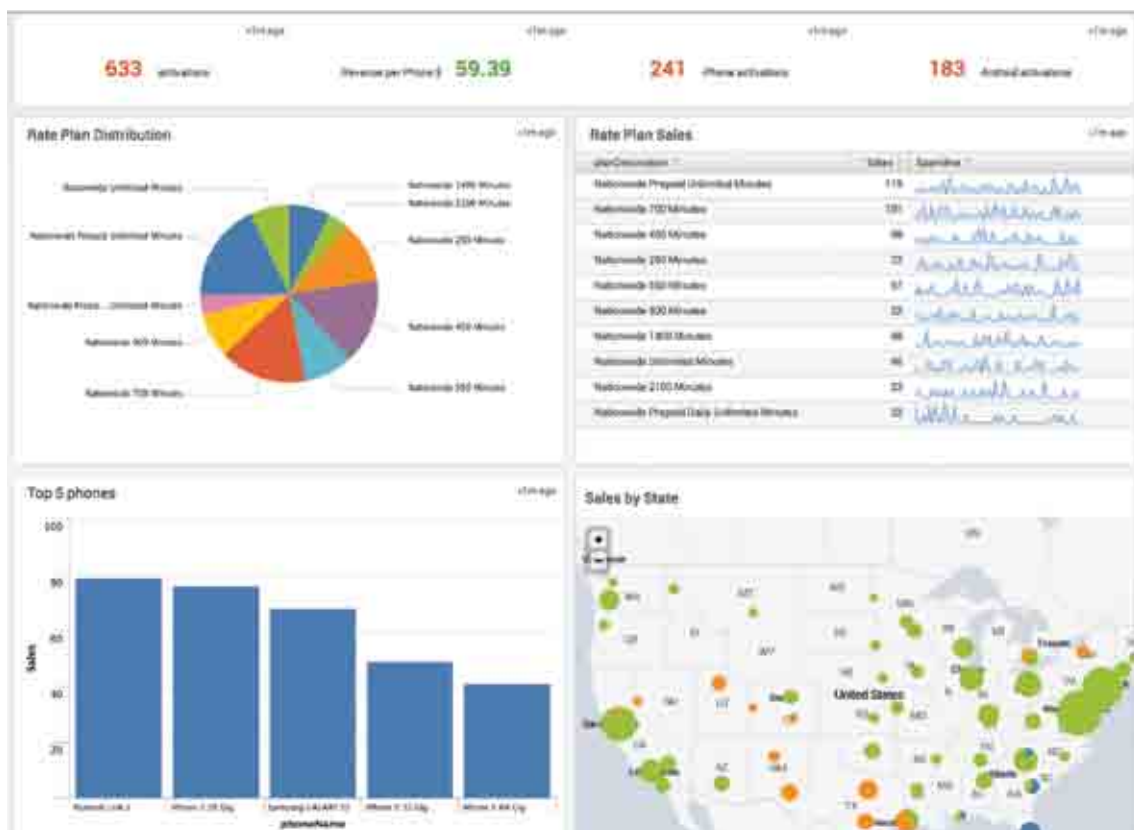
Splunk software collects, analyses and transforms machine-generated big data into real-time Operational Intelligence—valuable insight that can make your business more responsive, productive and profitable. Over half of the Fortune 100™ use Splunk software and have the business results to prove it.

Learn more at [Splunk.com/listen](http://Splunk.com/listen).

© 2013 Splunk Inc. All rights reserved.

splunk® listen to your data™

Toutes vos applications, vos systèmes et votre infrastructure génèrent des données chaque milliseconde de chaque jour. Ces données machine constituent le domaine le plus complexe, avec la plus forte croissance du « Big Data ». C'est aussi le domaine qui possède le plus de valeur car on peut y trouver tous les détails des transactions effectuées par les utilisateurs, le comportement des clients, l'activité des capteurs de toutes sortes, le comportement des machines, les menaces de sécurité, les activités frauduleuses pour ne citer que ces exemples.



## COORDONNÉES

### SPLUNK INC.

5, place de la Pyramide  
Tour Ariane La Défense 9  
92088 Paris La Défense Cedex

Tél : **+33 (0) 1 55 68 10 43**

Fax : N/A

Mail : [emea\\_sales@splunk.com](mailto:emea_sales@splunk.com)

## CONTACT

**Pierre  
GOYENEIX**

*Regional Sales Director, EMEA South*

Les données machine contiennent des informations critiques, utiles pour l'ensemble de l'Entreprise. Par exemple :

- Monitoring des transactions de bout en bout pour les business en ligne 24h/24 et 7j/7
- Comprendre l'expérience client, son comportement et l'utilisation des services en temps réel
- Respect des SLAs internes et monitoring des SLAs des Fournisseurs de Services
- Identifier les tendances et analyser les sentiments sur les réseaux sociaux
- Cartographier et visualiser les scénarii et schémas comportementaux pour améliorer sa politique sécurité

L'utilisation des données machine est un challenge. En effet, il est difficile de les traiter et de les analyser avec les outils de management de données classiques, sans parler des performances.

- Les données machine sont générées par une multitude de sources différentes : effectuer des corrélations entre des événements significatifs provenant de ces sources est extrêmement complexe
- Ces données sont non structurées et difficiles à faire rentrer dans un schéma prédéfini
- Les données machines atteignent rapidement des volumes importants et sont estampillées avec une unité de temps ; elles nécessitent une nouvelle approche pour les traiter et les analyser
- Les informations les plus précieuses à en retirer sont le plus souvent utiles en temps réel

Les solutions classiques de Business Intelligence, Data Warehouse ou IT Analytics ne sont tout simplement pas conçues pour ces types de données de très gros volume, dynamiques et non structurées. Des solutions émergentes en Open Source peuvent fournir une partie de la réponse mais nécessitent des équipes de développeurs de très bon niveau, très spécialisées et qui au final coutent très cher. De plus, quand les demandes évoluent, ces solutions fragiles n'ont très souvent pas l'agilité nécessaire pour s'adapter. Les entreprises d'aujourd'hui ne peuvent pas attendre. Les décideurs, ou qu'ils se trouvent dans l'organisation des entreprises, doivent constamment s'adapter aux changements incessants et de plus en plus rapides de leur environnement. Ils ont besoin d'une technologie qui permet les analyses temps réel, le data mining et la production de rapports à la volée : en bref, une solution qui leur apporte des réponses aussitôt qu'elles sont posées.

## Plateforme pour le Big Data

La plupart des organisations maintient en parallèle plusieurs types de données (données machine, relationnelles et d'autres formes de données non structurées). Splunk permet naturellement cette intégration, afin de vous permettre d'obtenir le maximum de vos informations. Que vous soyez développeur, professionnel des TIC, expert sécurité ou Business Analyst, la plateforme Splunk et ses « apps » rendent vos données accessibles, utilisables et pleines de valeur.

## Qu'est ce qui rend Splunk Unique ?

Splunk Enterprise est une solution temps réel, intégrée, de bout en bout, pour les données machine. Elle offre les capacités fondamentales suivantes :

- Assemblage et Indexation universelle des données machine provenant de toute source
- Langage de recherche puissant pour analyser des données temps réel ou historiques
- Monitoring de comportements et de niveaux ; alertes en temps réel quand un ensemble de conditions sont remplies
- Production de rapports et d'analyses
- Tableaux de Bord personnalisés et spécifiques pour chaque fonction de l'entreprise
- Haute disponibilité et évolutivité sur les plateformes standard du marché
- Contrôle d'accès et sécurité basée sur les fonctions
- Support flexible des déploiements multi-site et distribués
- Connectivité avec les autres outils de données comme les SGBDR et une connectivité bidirectionnelle avec Hadoop
- Une plateforme robuste et flexible idéale pour les applications Big Data

## Vous deployez Hadoop?

Hunk (Splunk Analytics on Hadoop) est une nouvelle application qui permet d'explorer, analyser et visualiser des données dans Hadoop. S'appuyant sur les années d'expérience de Splunk dans le Big Data et le déploiement auprès de milliers de clients dans le monde, Hunk permet une amélioration spectaculaire en termes de vitesse et de simplicité pour analyser et interagir avec les données situées dans Hadoop. Tout cela sans programmation, ni intégration couteuse ni migrations forcées de données.



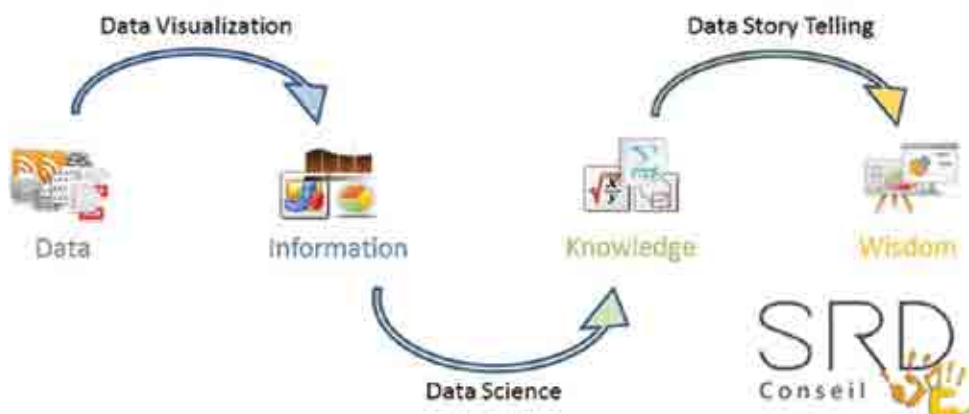
**Notre société accompagne les entreprises dans le traitement de l'information autour du CRM, de la Business Intelligence et du Big Data. Nous apportons aux responsables fonctionnelles des moyens rapides et efficaces afin qu'ils puissent prendre les meilleures décisions.**

Aujourd'hui, exploiter les données diverses (interne ou externe à l'entreprise) est possible via des technologies robustes et innovantes. Nous conseillons et mettons en œuvre des systèmes d'informations efficaces de prise de décision pour pouvoir disposer d'une meilleure connaissance des données. Nous sommes une entreprise d'experts dans les solutions CRM, Business Intelligence et Big Data.

Notre « Lab Data Scientist » est dédié à la pratique de la science sur les données. Aujourd'hui, de plus en plus d'entreprises souhaitent exploiter au mieux leurs données afin de pouvoir améliorer leur compétitivité et leur productivité. Dans cet objectif, notre entité accompagne les entreprises dans la prise en compte du besoin métier, dans la mise en place d'architecture de données, de transformation, d'acquisition de nouvelle connaissance pour ainsi transformer les données en connaissance.

**Cette entité travaille ainsi sur les volets dédiés à la Data Science à savoir :**

- Ingénierie de la donnée : architecture, stockage, modélisation, optimisation, automatisation autour des plateformes Big Data (Hadoop, Aster Data ...)
- Transformation des données : modèles statistiques (calculs de distances, calcul de corrélation, clustering, classification, arbre de décision, réseau de neurones) appliqués aussi aux documents textes, mise en place de techniques de machine learning
- Visualisation des données : graphiques adaptés aux données, mise en récit des données pour une meilleure adhésion utilisateurs
- Connaissance métiers : autour de l'intelligence sociale sur les clients et produits pour fidéliser, développer et gérer les clients





## COORDONNÉES

### SRD CONSEIL

10 rue château  
75009 PARIS

Tél : **+33 (0) 1 74 64 06 07**

Fax : N/A

Mail : [contact@srd-conseil.com](mailto:contact@srd-conseil.com)

## CONTACT

**Abed  
AJRAOU**

*Directeur Business Unit BI et Big Data*

SRD Conseil, entreprise créée en 2006, a pour objectif d'aider les entreprises dans le cœur de leurs préoccupations à savoir l'acquisition, la rétention et la fidélisation de leurs clients. SRD Conseil aide à organiser, optimiser et à mesurer l'efficacité des actions marketing et commerciales par la mise en place d'outils, de méthodes et de formations adaptés à métier aux besoins des entreprises.

### SRD Conseil intervient sous plusieurs formes de prestations à savoir :

- **Forfait** : Sur des projets cadrés, ou afin de permettre aux clients de bénéficier d'offres basées sur nos domaines d'expertise (migration, projets QuickStart, solutions standards verticalisées en mode SaaS...), SRD Conseil propose à ses clients d'intervenir au forfait.
- **Régie** : Pour un besoin de renfort technique et/ou fonctionnel, sur un périmètre et une durée indéfinie, un consultant SRD conseil peut intervenir en régie, de 0,5 à 220 jours par an.
- **TMA (tierce maintenance applicative)** : SRD Conseil propose d'intervenir sur des TMA de 20 à 1000 jours, sur engagement de 6, 12 ou 24 mois, sans nécessairement être l'intégrateur à l'origine du projet.
- **Formation** : En tant qu'organisme agréé, SRD Conseil intervient auprès de vos équipes techniques, métiers et/ou vos utilisateurs finaux avec des plans de formations adaptés.
- **Recrutement** : En cas de besoin de renfort interne (CDI), n'hésitez pas à consulter SRD Conseil: la sélection de profil fait l'objet d'une rigueur particulière (plusieurs entretiens, contrôle de références poussé, éventuels tests ...)

### Nous intervenons dans les domaines suivants :

Banque - Finance / Assurances / Mutuelles / Santé - Pharma / Energie / Transports & Sécurité / Immobilier - BTP / Hôtellerie & Restauration / Services BtoB & BtoC / Négoce - Distribution / eCommerce / FAI & Telecom / Administration / Association

Notre équipe de Data Scientist est là pour apporter leur connaissance des architectures Big data, de traitements statistiques et text-mining aux données structurées ou poly structurées, et mettre en place des data visualisations pertinentes pour mieux découvrir et comprendre ces données. L'objectif est d'accompagner nos clients dans une meilleure compréhension de leur métier et ainsi de leur permettre la mise en place d'une vraie stratégie autour de la donnée.



# Is ETL Dead in the Age of Hadoop?

By Josh Rogers

The collision between traditional approaches to Data Integration and Hadoop is one of the more interesting aspects of Big Data generally and of Hadoop specifically.

Data integration and more specifically, Extraction, Transformation and Loading (ETL), represents a natural application of Hadoop and a precedent to achieving the ultimate promise of Big Data – new insights. But perhaps most importantly at this point in the adoption curve, it represents an excellent starting point for leveraging Hadoop to tackle Big Data challenges.

## ETL Must Die

On Halloween 2012, Doug Henschen published a fascinating article around the work Sears is doing with Hadoop. Within the article, there is a section labeled, “ETL Must Die.” While I can appreciate the sentiment, let’s be clear: ETL as a capability in the information management stack is not going anywhere. Data will need to be sourced (“E”), it will need to be related and refined (“T”) and put somewhere (“L”). What Sears has done is a fundamental shift in architecture for delivering this capability. Sears needed a new architecture, and found one in Hadoop. As detailed in the article, Hadoop is providing a whole range of benefits to Sears. But to say Sears is no longer doing ETL is not accurate. In fact, the opposite would appear to be the case. Sears is sourcing more data, the scope of the transformations has increased, and the distribution of information has expanded. What Sears has done is to move existing workloads from various tools and platforms into a Hadoop environment. That has shifted the reliance on specific products but not the capabilities required to maximize the value of the data.

## Hadoop is not a Data Integration Solution

In a recent research piece entitled “Hadoop is not a Data Integration Solution,” Ted Friedman and Merv Adrian from Gartner describe the gaps between Hadoop and a proper Data Integration Solution. But, what is it about the Hadoop infrastructure that is attracting such interest despite these significant gaps?

In the era of Big Data, Hadoop’s architecture is fundamentally superior for supporting many of the most commonly deployed data integration functions. For this reason, organizations are flocking to Hadoop even if key functional capabilities must be written by hand today. Hadoop makes it easy to scale computing power horizontally with low cost components. More importantly, this architecture will also support a broader range of data integration functions.

We are seeing willingness by organizations to abandon well-understood and mature features of Data Integration products and embrace a raw, open source technology. The main problem Big Data creates is an architectural one, not a functional one. Perhaps it is fair to say that today; Hadoop is not a Data Integration solution. But, as Mr. Adrian and Mr. Friedman seem to agree, it could very well be the Big Data operating system on which next-generation data integration solutions are built. Moreover, I believe it’s a safe bet that the applications needed to perform sophisticated Data Integration in Hadoop are coming quickly.

## Rebuilding ETL and ELT in Hadoop

In his February 21st post entitled, “Big Data’s New Uses Cases: Transformation, Active Archive and Exploration,” Cloudera’s CTO, Amr Awadallah, provides a concise and accurate history of the “T” in ETL. In the late 90s and early 2000s, the conventional wisdom was to extract data from key source systems, cleanse and relate that data through transformations, then load it into the warehouse. As businesses got customized views of information combined from multiple systems, the traditional ETL approach, particularly the “T”, began to suffer under the weight of growing volumes. At the same time, databases and the hardware that supported them were evolving rapidly to support intense query load with massively parallel processing architectures.

Over time, a new approach to data integration was developed which would avoid the lack of relative scalability in traditional data integration engines. The concept was simple: Extract the tables from source systems and then load those tables directly into the data warehouse into a separate staging area. Use SQL, a common and widespread skill to combine and relate those raw tables into the tables required for reporting and analytics. This approach became known as ELT and is, in fact, the most common approach used for data integration today.

While this approach provides the scalability required to perform the “T” for integrating the growing sources and volumes of data, there are tradeoffs. Most of the logic used for data integration is expressed in SQL and it is hard to maintain. It requires constant tuning as data changes. It is difficult to test and debug. Oh, and, very often, it fails to perform well.

ELT has become the conventional wisdom in integrating data but little focus has been put on the impact these tradeoffs have from an expense and agility perspective. Current practitioners would argue that it is the cost of operating a large-scale analytics and reporting environment and that there is no other alternative that can provide the required scale to meet the insights the business demands.

By targeting the replacement of ELT or ETL in Hadoop, you can provide clear answers to many of these questions. There are massive cost savings in just offloading the data warehouse and there is a backlog of new sources that the business wants in the warehouse but has not been met because of cost and scalability..

Such an approach yields a series of side benefits to the organization that will equip the team to deliver the next Big Data project. Fundamentally, by focusing your first project on operational efficiencies, you will deliver a win for your business partners and your organization.

## ETL – the Leading Use Case for Hadoop

The adoption of Hadoop and its promise is creating hype but make no mistake – it is also creating real value today. Identifying the right starting point can create momentum and initial wins on the journey to leveraging Big Data.

Why not start with an approach that has a clear payback and builds the skills you need in order to consistently deliver results? Start by rebuilding and even augmenting existing ETL or ELT workloads. In the process, you won’t be killing ETL but rather moving this core capability to an architecture that can support the scale the business needs, at a cost it can afford. Hadoop will dramatically change the data integration landscape but it won’t kill ETL. Far from it. ETL is the killer app that will drive mainstream adoption of Hadoop, the new operating system for Big Data.



**Josh Rogers**  
President of Data  
Integration  
Syncsort.

## COORDONNÉES

### SYNCSORT FRANCE

Tour Egée - La Défense  
17 avenue de l'Arche  
92671 Courbevoie Cedex

Tél : **+33 (0) 1 46 67 46 10**

Fax : **+33 (0) 1 47 88 04 38**

Mail : [syncsortSA@syncsort.com](mailto:syncsortSA@syncsort.com)

## CONTACT

**Florence  
RABUEL**

*Regional Sales Manager*



Syncsort propose aux organisations gérant d'énormes quantités de données des solutions logicielles permettant de collecter et traiter de manière plus intelligente cette avalanche exponentielle de données.

Puisant dans son expérience de milliers de déploiements sur toutes les plateformes majeures, y compris le mainframe, Syncsort aide ses clients, partout dans le monde, à surmonter les limites architecturales des environnements Hadoop et ETL d'aujourd'hui.

Les entreprises obtiennent ainsi de meilleurs résultats en moins de temps, avec moins de ressources pour un coût de possession réduit.

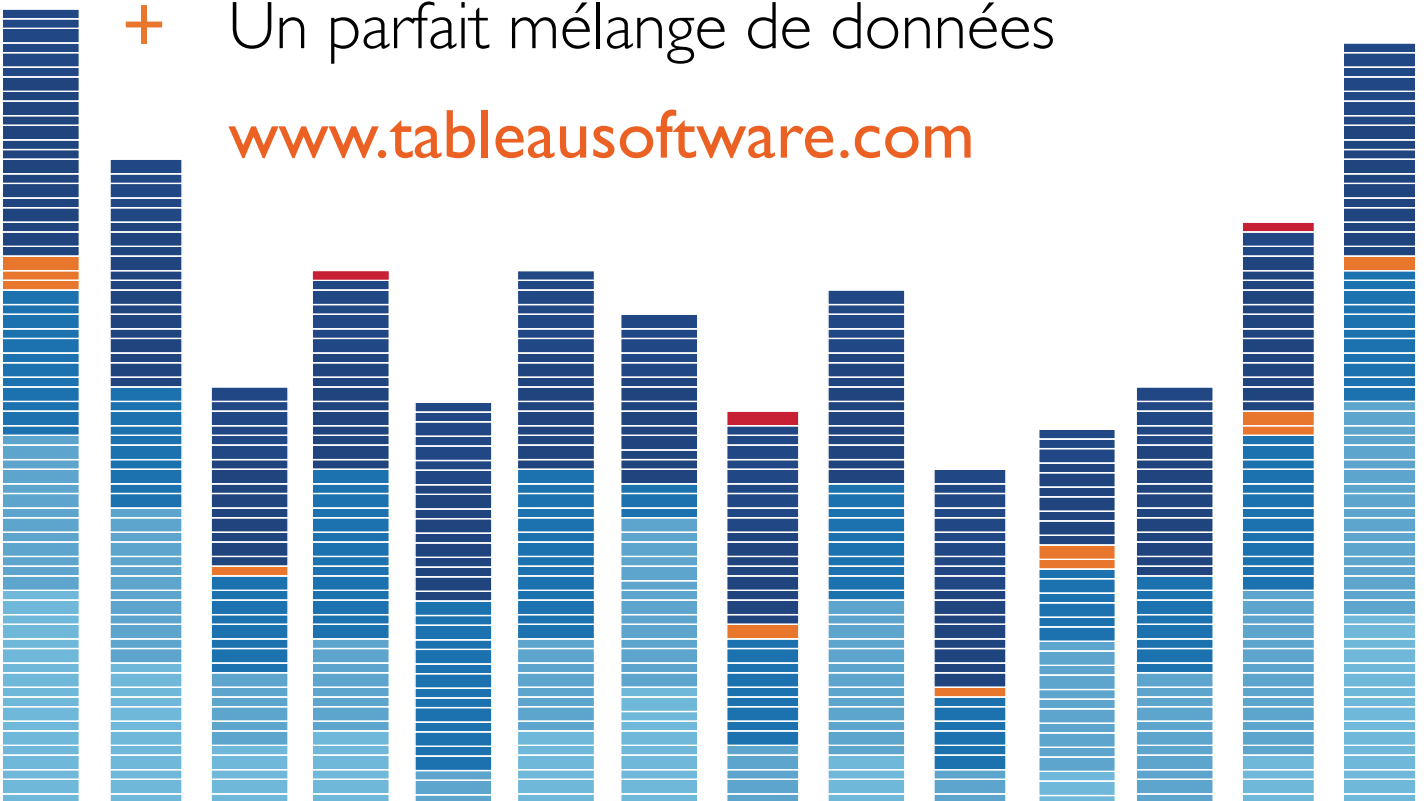
Pour en savoir plus, veuillez consulter notre site.



# Répondez aux questions immédiatement grâce à Tableau

- + 10 à 100 x plus rapide
- + Une simplicité d'utilisation à vous en donner des frissons
- + De superbes tableaux de bord intelligents et adaptés
- + Connectez-vous directement et effectuez vos tâches
- + Un parfait mélange de données

[www.tableausoftware.com](http://www.tableausoftware.com)





## COORDONNÉES

### TABLEAU SOFTWARE

Blue Fin Building  
110 Southwark Street  
London SE1 0TA - United Kingdom

Tél : +44 (0) 203 310 4500

Fax : +44 (0) 203 310 4501

Mail : europeansales@tableausoftware.com

## CONTACT

**Julien  
SIGONNEY**

Sales Area Manager

Tableau Software aide les utilisateurs à visualiser et à comprendre leurs données. Tableau permet d'analyser, de visualiser et de partager simplement des informations. La société Tableau Software propose des solutions de Business Intelligence qui aident les personnes et les organisations à travailler plus efficacement en leur permettant d'identifier plus clairement les tendances. Grâce à l'utilisation des outils de Tableau, les utilisateurs optimisent leur temps en ayant accès aux informations qui leur sont importantes de manière très rapide. Plus de 12 000 entreprises font confiance à Tableau pour obtenir rapidement des résultats, au bureau comme lors des déplacements.

Grâce à sa facilité d'utilisation, Tableau Public est un outil qui permet de traiter un gros volume de données facilement et rapidement et ensuite de les publier sur le web en un seul clic. Cet outil aide diverses typologies d'utilisateurs dont les journalistes à analyser des données complexes et extraire des informations pertinentes qui permettent de très belles visualisations de données et cela de manière dynamique.

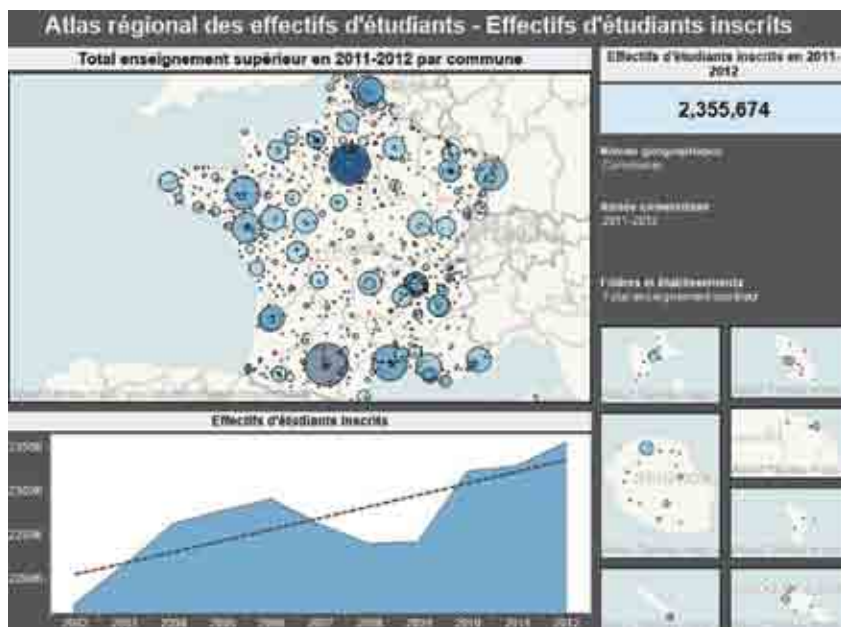
### Tableau Software a trois produits principaux pour les entreprises : Tableau Desktop, Tableau Server et Tableau Online.

#### Répondez aux questions immédiatement grâce à Tableau Desktop.

Tableau Desktop utilise une technologie innovante de l'université de Stanford qui vous permet de glisser-déposer des données pour les analyser. Vous pouvez accéder aux données en quelques clics, puis visualiser et créer facilement des tableaux de bord interactifs.

Nous avons mené des recherches pendant plusieurs années pour mettre en place un système prenant en charge la capacité naturelle des personnes à penser de manière visuelle. Basculez facilement entre les vues, suivez votre raisonnement naturel. Vous n'êtes pas coincé par des assistants d'exécution ni obligé de rédiger des scripts. Vous créez simplement des visualisations de données riches et harmonieuses.

C'est tellement simple à utiliser que c'est à la portée de n'importe quel utilisateur d'Excel. Obtenez plus de résultats en fournissant moins d'efforts. Et c'est 10 à 100 x plus rapide que les solutions existantes.



#### Tableau Server - Une incroyable alternative aux solutions décisionnelles traditionnelles.

Tableau Server est une solution décisionnelle qui fournit un outil d'analyse basé sur navigateur Web accessible à tous. C'est une alternative très rapide aux logiciels d'informatique décisionnelle traditionnels connus pour leur lenteur et leur manque de flexibilité.

Vous avez un travail à effectuer, mais les outils dont vous disposez actuellement sont lents et volumineux. Tableau Server est un nouveau type d'outil d'informatique décisionnelle 100 fois plus rapide et agréable à utiliser. Vous effectuez votre déploiement selon vos besoins. Testez-le gratuitement en ligne. Obtenez des résultats en quelques semaines seulement, et non plus après quelques années.

#### La puissance de Tableau dans le cloud.

Tableau Online est une version hébergée de Tableau Server. Elle rend l'analyse d'entreprise rapide et plus facile que jamais. Partagez des tableaux de bord avec l'ensemble de votre entreprise, et avec vos clients et partenaires en quelques minutes. Fournissez des vues de données interactives en direct qui permettront aux personnes de répondre à leurs propres questions direc-

tement dans un navigateur Web ou sur une tablette. Faites-le dans un environnement sécurisé et hébergé.

Soyez opérationnel en quelques minutes. Ajoutez de nouveaux utilisateurs en quelques clics. Vous pouvez augmenter la puissance de Tableau Online en fonction de vos besoins et le rendre accessible où que vous soyez. Aucun achat, aucune configuration et aucune gestion d'infrastructure n'est nécessaire.



.....

# Redesigning the data landscape around user experience



**worldline**  
e-payment services

..... an atos company

## COORDONNÉES WORLDLINE

River Ouest  
80 quai Voltaire  
95877 Bezons Cedex

Tél : **+33 (0) 4 78 17 69 35**

Fax : N/A

Mail : [infoWL@worldline.com](mailto:infoWL@worldline.com)

## CONTACT

**Gregoire  
WIBAUX**

*Business development & Marketing  
stratégique*

## **Worldline, filiale d'Atos, est le leader européen et un acteur mondial référent dans le secteur des paiements et des services transactionnels.**

En ajoutant à ses expertises et plates-formes transactionnelles historiques des composants Big Data industriels, Worldline construit avec ses clients et partenaires des projets exploitant pleinement la valeur intrinsèque des données. Worldline délivre ainsi à ses clients des services de bout en bout sur l'ensemble de la chaîne de valeur : captation des données, analyse, visualisation et valorisation.

- Pour la phase de captation, Worldline tire parti de ses compétences historiques en termes de collecte, intégration, stockage et nettoyage de gros volumes de données transactionnelles. Pour ces types de données, les composantes vitesse (temps réel) et variété (formats divers voire changeants) sont souvent critiques. Les données sont gérées au sein des datacenters Worldline, ce qui permet d'apporter flexibilité et fiabilité à cette phase du cycle.
- Dans chaque contexte client, Worldline apporte son expertise en termes d'analyse de données, notamment dans les domaines de la lutte contre la fraude, du scoring de transactions et de la gestion de la relation client, et met en place des modèles statistiques et des algorithmes d'intelligence artificielle sur mesure pour passer de la donnée brute à l'information à valeur ajoutée.
- La phase de visualisation est le maillon indispensable pour assurer un lien entre les données et les utilisateurs finaux. Elle est prise en compte dès la conception des services et ce quel que soit le terminal de l'utilisateur. Pour ce faire, Worldline met à profit ses réussites dans la création d'interfaces et d'applications adaptées aux multiples combinaisons contexte, support et usage.
- Enfin, pour la valorisation, Worldline dispose des expertises business nécessaires pour accompagner ses clients dans le nouveau paysage informationnel du Big Data et leur permet de tirer le meilleur parti des données, qu'elles soient internes ou externes. C'est ainsi que se concrétisent pour eux les multiples opportunités qu'elles représentent.

En associant cohérence et niveaux de qualité de service industriels de bout en bout de la chaîne, Worldline va au-delà des enjeux purement techniques pour délivrer la valeur des technologies et des nouveaux usages de la Big Data à tous ses clients. du BigData.

**40 ans**  
d'expérience  
en solutions  
de paiement

**1,1**  
milliard  
d'euros  
de chiffre  
d'affaires

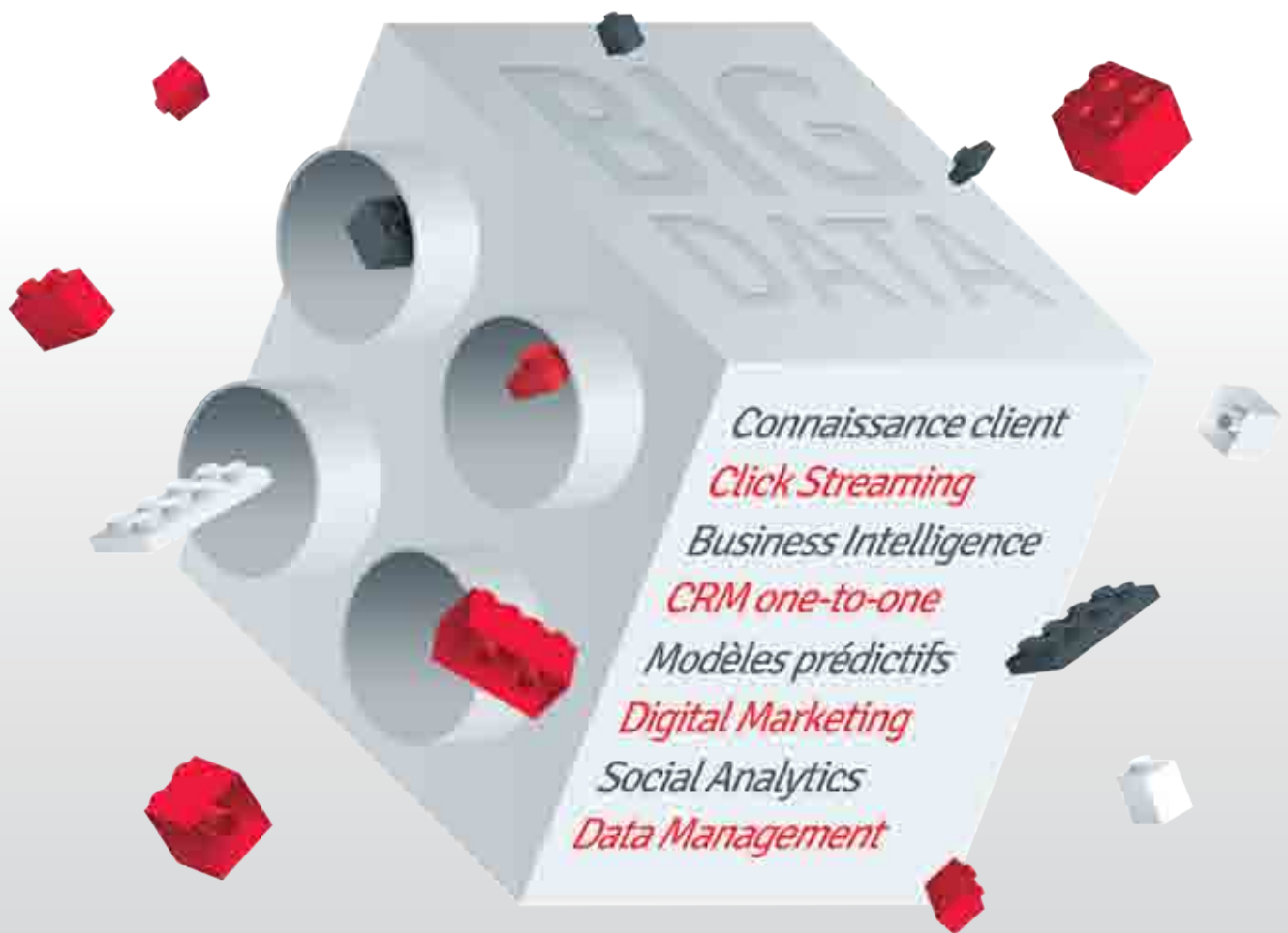
Plus de  
**7 100**  
experts  
dans le monde  
entier

Plusieurs  
milliards  
de transactions  
traitées  
chaque  
année



# Ysance

## Play with your **Big Data**



**Ysance** créateur de plateformes digitales



cloudera

Google

MAPR  
TECHNOLOGIES

QlikView

[www.ysance.com](http://www.ysance.com)



## COORDONNÉES

### YSANCE

Tour de l'Horloge  
4 place Louis Armand  
75012 Paris

Tél : +33 (0) 1 43 36 23 17

Fax : +33 (0) 1 73 72 97 26

Mail : [contact@ysance.com](mailto:contact@ysance.com)

## CONTACT

**Romain  
CHAUMAIS**

*Directeur du Développement*

**Créée en 2005 par une équipe d'experts IT, Ysance est une agence conseil en technologie digitale forte de 100 collaborateurs.**

Ysance aide les directions métier à tirer un profit des nouveaux usages du numérique en accompagnant ses clients dans la conception, la mise en œuvre et l'animation de leurs projets digitaux.

Ses activités s'articulent autour de 5 métiers : le Marketing digital, la Business Intelligence, l'e-business, le Cloud Computing et le Big Data.

Depuis 3 ans, Ysance construit des plateformes Big Data pour les mettre au service des différents métiers de l'entreprise : CRM, eCommerce, Marketing, Ventes, RH, Finances, BI, Datamining, etc.

Fort de cette expérience, Ysance a conçu une méthodologie progressive et pragmatique déclinée en 4 étapes pour rapidement accéder aux potentiels des Big Data :

### 1. Découvrir et estimer la valeur d'une démarche Big Data pour son activité

Évaluer les avantages et identifier les métiers pouvant en bénéficier  
Imaginer les services, les analyses et les opportunités rendus possibles  
Définir les scénarios Big Data et les business cases associés

### 2. Démarrer et réaliser VOS premières analyses Big Data

Définir un premier scope d'analyse métier  
Collecter les données et les charger dans la plateforme Big Data d'Ysance  
Réaliser un ensemble d'analyses et valider le modèle

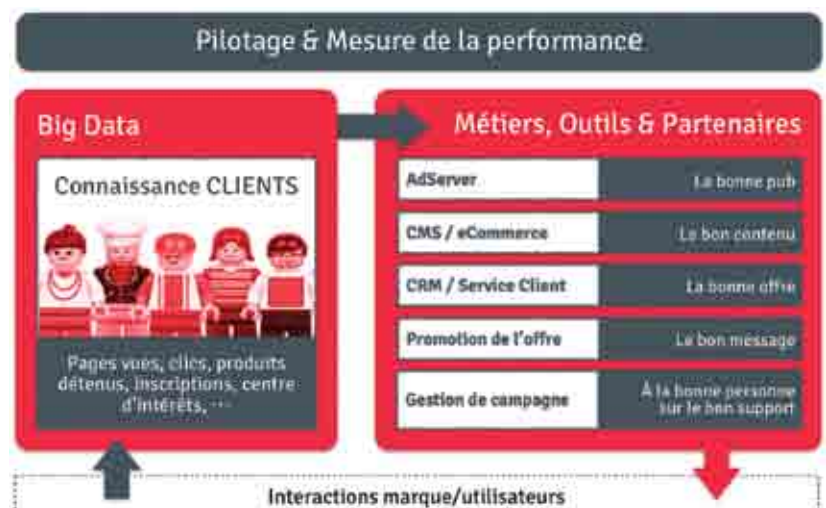
### 3. Développer et mettre en place le socle de VOTRE plateforme Big Data

Formaliser et valider les besoins techniques et fonctionnels  
Identifier les sources et les volumes de cible  
Définir l'architecture et choisir les solutions techniques  
Implémenter la plate-forme Big Data

### 4. Déployer et faire du Big Data un avantage concurrentiel

Industrialiser la collecte, le stockage et l'analyse en temps réel  
Élargir, organiser et sécuriser l'accès à la plate-forme  
Capitaliser les bonnes pratiques et les normes de développement  
Organiser la conduite du changement

Ysance dispose d'une grande expertise des technologies Hadoop, des enjeux de Data management et des environnements Cloud.



INTEGRATING BIG DATA... SMARTER  
**syncsort**

**Get on the  
Fast Track  
to Big Data!**

*SMARTER ARCHITECTURE.  
SMARTER PRICE POINT.  
JUST SMARTER.*

*COLLECT, PROCESS,  
DISTRIBUTE*

*FROM MAINFRAME TO HADOOP  
...AND EVERYTHING IN BETWEEN*

## Discover Our Solutions for:

- ✓ **Hadoop:** overcome the challenges of collecting, processing & integrating data with our Hadoop ETL, sort acceleration, & connectivity solutions!
- ✓ **ETL & Sort:** manage your data smarter with our solutions for high-performance ETL, ETL optimization, SQL migration, sort acceleration, & more.
- ✓ **Mainframe:** discover why Syncsort has led the industry. In high-performance sort technology for over four decades & see how our new solutions can lower mainframe costs!

[syncsort.com/bigdata](http://syncsort.com/bigdata)

# Put your **data** to work\*

Exploitez vos données



**DATA & DATA**  
CONSULTING

Comment créer de la valeur à partir des données au regard de mes grands enjeux actuels et futurs ? Avec quelle stratégie Big Data et comment la déployer ? Comment définir ma feuille de route et quelles solutions technologiques adopter ? Quels sont les impacts de la Big Data et comment mener à bien le changement ? Comment renforcer mes équipes avec des compétences analytiques et comment insuffler la culture « data science » au sein de mon organisation ?

Notre mission est de résoudre ces problématiques en apportant des solutions innovantes ayant un impact réel sur la performance et permettant de bâtir un avantage compétitif réel et durable.

Bienvenue chez Data&Data Consulting, votre partenaire pour mener la révolution Data.

[www.data-and-data.com](http://www.data-and-data.com)

Data&Data Consulting© 31 rue Saint-Lazare 75009 Paris [Contact@data-and-data.com](mailto:Contact@data-and-data.com)