

2014 / 2015

GUIDE DU BIG DATA

L'ANNUAIRE DE RÉFÉRENCE
À DESTINATION DES UTILISATEURS



BIG DATA

by **Corp.**
in Corporations we Trust

ASK BIGGER QUESTIONS



cloudera®

PARTIE ÉDITORIALE MARCHÉ, ENJEUX ET APPLICATIONS (PAGES 3 À 41)

EDITO	page 3
CONTEXTE	page 5
<ul style="list-style-type: none"> • BIG DATA : nouvelles définitions et concept approfondi • Le marché et les enjeux : chiffres, croissance, secteurs d'application <ul style="list-style-type: none"> - Dans le monde <ul style="list-style-type: none"> • Les Etats-Unis, précurseur du Big Data • Les pays émergents • L'adoption du Big Data en entreprise - En France/Europe <ul style="list-style-type: none"> • Le cas français : offre et demande, une incompatibilité ? • Données personnelles et cyber-sécurité : enjeux technologiques et réglementaires • Vers une cohésion européenne • L'open data et la création de valeur • Focus Big Data et formation : les programmes, les profils, la création d'emplois <ul style="list-style-type: none"> - Mise en place du Big Data : le rôle des métiers, la place de la DSI - Les formations - La création d'emploi 	
PORTRAITS	page 15
<ul style="list-style-type: none"> • Axelle LEMAIRE • Henri VERDIER • Anne LAUVERGEON • Gilles BABINET • Pashu DEWAILLY CHRISTENSEN • Doug CUTTING 	
HADOOP : enfin une solution d'entreprise ? Par José DIZ	page 24
<ul style="list-style-type: none"> • Hadoop 1 : la nouvelle star de l'analytique est née <ul style="list-style-type: none"> - 10 ans : des origines de MapReduce à Hadoop 2.x - Le centre névralgique HDFS et MapReduce - Quelques briques de l'édifice Hadoop - Des manques importants fragilisant Hadoop 1 • La revolution Hadoop 2 <ul style="list-style-type: none"> - Naissance de Yarn et réduction de MapReduce - Une gestion des fichiers amplement améliorée - Petit panorama de quelques modules Hadoop 2 - Sans oublier : sécurité, administration et gouvernance 	
PROJETS	page 29
<ul style="list-style-type: none"> • Un projet humanitaire : le développement économique et le Big Data • Un projet culturel : le data journalisme pour les flux migratoires en Europe • Un projet de lutte contre la fraude : la détection de la contrefaçon en ligne • Un projet dans le secteur automobile : l'assurance évolutive • Un projet dans le secteur des loisirs : les communautés virtuelles • Un projet de recherche et développement dans l'industrie : l'energy-monitoring • Un projet de recherche : l'Institut de Génomique du CEA • Un projet dans les transports : la sécurité aérienne • Un projet en climatologie : l'évolution climatique au niveau mondial 	

PARTIE FOURNISSEURS ANNUAIRE DES ANNONCEURS (À PARTIR DE LA PAGE 42)

37 fournisseurs : une double page par annonceur (1 page profil entreprise + 1 page PROJET)

ACTUATE / A.I.D. / AMAZON WEB SERVICES / BIME ANALYTICS / BITTLE / BLUESTONE / CAPGEMINI / CENTRALE PARIS Executive Education / CLUDERA / Club Decision DSI / CSC / DATASTAX / EDIS CONSULTING / EKIMETRICS / EXALEAD / GFI INFORMATIQUE / Hewlett-Packard / IBM France / KEYRUS / MICROPOLE / Neo Technology / PENTAHO / PIVOTAL / QLIK / SAP France / SAS / SCINETIK / SEMDEE / SENTELIS / SINEQUA / SPLUNK / SYNCSORT / TABLEAU SOFTWARE / TIBCO JASPERSOFT / TALEND / TERADATA France / VISEO.

CONTEXTE

BIG DATA : NOUVELLES DÉFINITIONS ET CONCEPT APPROFONDI

C'est officiel, on l'appellera désormais « megadonnées ». Le Big Data est un terme si fréquemment utilisé que la Commission générale de terminologie et de néologie s'en est emparé, révélant le 22 Août dernier sa traduction officielle en français. La définition associée est la suivante : « données structurées ou non dont le très grand volume requiert des outils d'analyse adaptés ». La commission précise que l'on trouve également l'expression « données massives », d'ailleurs souvent privilégiée ces dernières années.

L'effervescence autour de ce phénomène apparu il y a quelques années, génère une certaine confusion de définition, si tant est que certains détracteurs y trouvent un argument de lassitude.

S'il est vrai que le mot Big Data semble parfois utilisé à outrance, n'oublions pas que le phénomène est bien réel. Rappelons ici quelques notions, pour les initiés et ceux qui le sont moins. De façon assez communément admise, le Big Data est défini par les 3V, voire les 4 ou 5V. C'est la façon la plus schématique et synthétique possible pour expliquer ce que renferme cette notion.

Les 3V expliquent en quoi le Big Data représente l'arrivée de technologies permettant une toute nouvelle approche de la donnée. Le Big Data, c'est traiter des Volumes de données conséquemment supérieurs à ceux traités auparavant, à une Vitesse incomparable, le tout en intégrant une Variété de données largement plus riche. Chacun de ces V mérite quelques précisions.

La notion de volume peut s'exprimer en chiffres : aujourd'hui on parle de stocker et traiter des exaoctets (10¹⁸) voire zettaoctets (10²¹) alors qu'il y a à peine 10 ans on parlait de mégaoctets (10⁶), stockés sur des disquettes. L'information est cependant peu intelligible si elle n'est pas mise en relief : il est estimé que 90 % des données récoltées depuis le début de l'humanité ont été générées durant les 2 dernières années. Le plus impressionnant réside dans le fait que la création de données est exponentielle. Le graphique ci-dessous donne quelques exemples de données générées, leurs volumes, et leur impact. (source : étude EMC).

La notion de Vitesse nécessite l'illustration par l'exemple : quand une entreprise traitait ses données, en mégaoctets auparavant, en plusieurs jours parfois, elle peut désormais réaliser la même étude, sur des volumes Big Data, en quelques heures voire minutes. L'impact économique et le gain d'efficacité sont alors évidents.

"Il est estimé que **90 %** des données récoltées depuis le début de l'humanité ont été générées durant les 2 dernières années"

La notion de Variété est un peu plus technique : les données préalablement utilisées étaient hautement formatées, renseignées selon des critères communs qui eux seuls garantissaient la capacité de comparaison et de traitement de l'information. Ce que le Big Data apporte, c'est la possibilité de traiter tout type de données, dans sa forme originelle, en intégrant les nouveaux modes d'expression, de mesure et d'interactions. On peut alors traiter images, sons, vidéo, commentaires de blogs, logs...

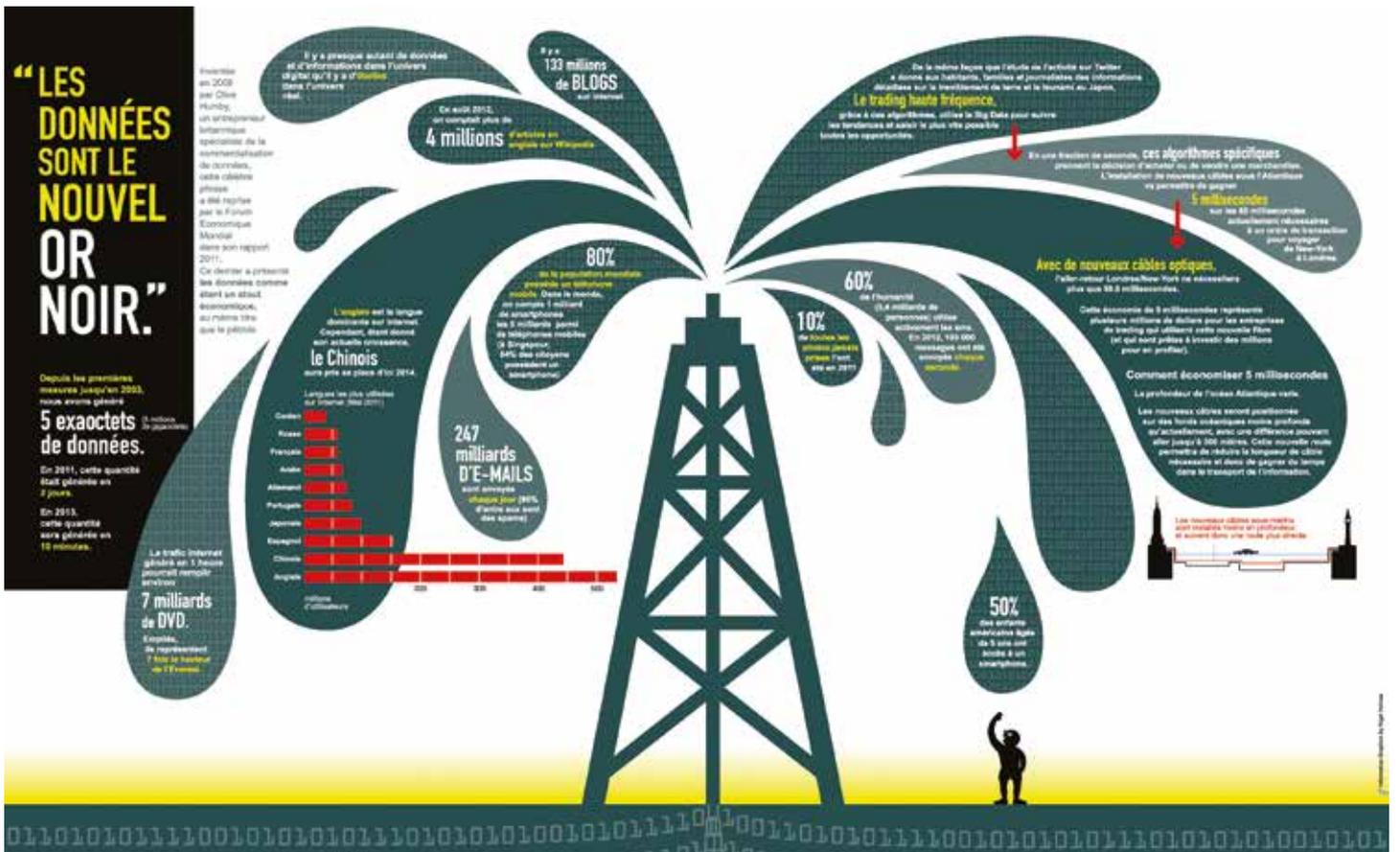
À ces 3V, on ajoute désormais le V de Valeur et celui de Véracité, qui expriment le besoin de disposer de données fiables, pertinentes et significatives pour donner suffisamment de sens et d'intérêt économique des analyses menées.

La réelle idée derrière le terme Big Data, celle qui justifie qu'on parle bien d'une révolution et non pas d'une simple amélioration du traitement de la donnée, c'est le fait que ces 3V changent entièrement la façon dont on aborde l'information. La donnée est au centre de cette transformation.

Dans un premier temps, ce sont les entreprises et institutions qui en exploitent ces nouvelles capacités de traitement et analyse de la donnée. En effet, 70 % des données créées le sont par des individus, mais ce sont les entreprises qui stockent et gèrent 80 % d'entre elles.

Parler du Big Data implique donc que l'on aborde les conséquences sur l'économie, les entreprises et organisations qui l'utilisent, et la façon dont il y est mis en place.

Ce que les premiers projets montrent, c'est que le Big Data décuple les possibilités d'analyse dans tous les secteurs et ouvre de nouveaux horizons à la prise de décision.



Infographie : "les données sont le nouvel or noir" par EMC

LE MARCHÉ ET LES ENJEUX : CHIFFRES, CROISSANCE, SECTEURS D'APPLICATION

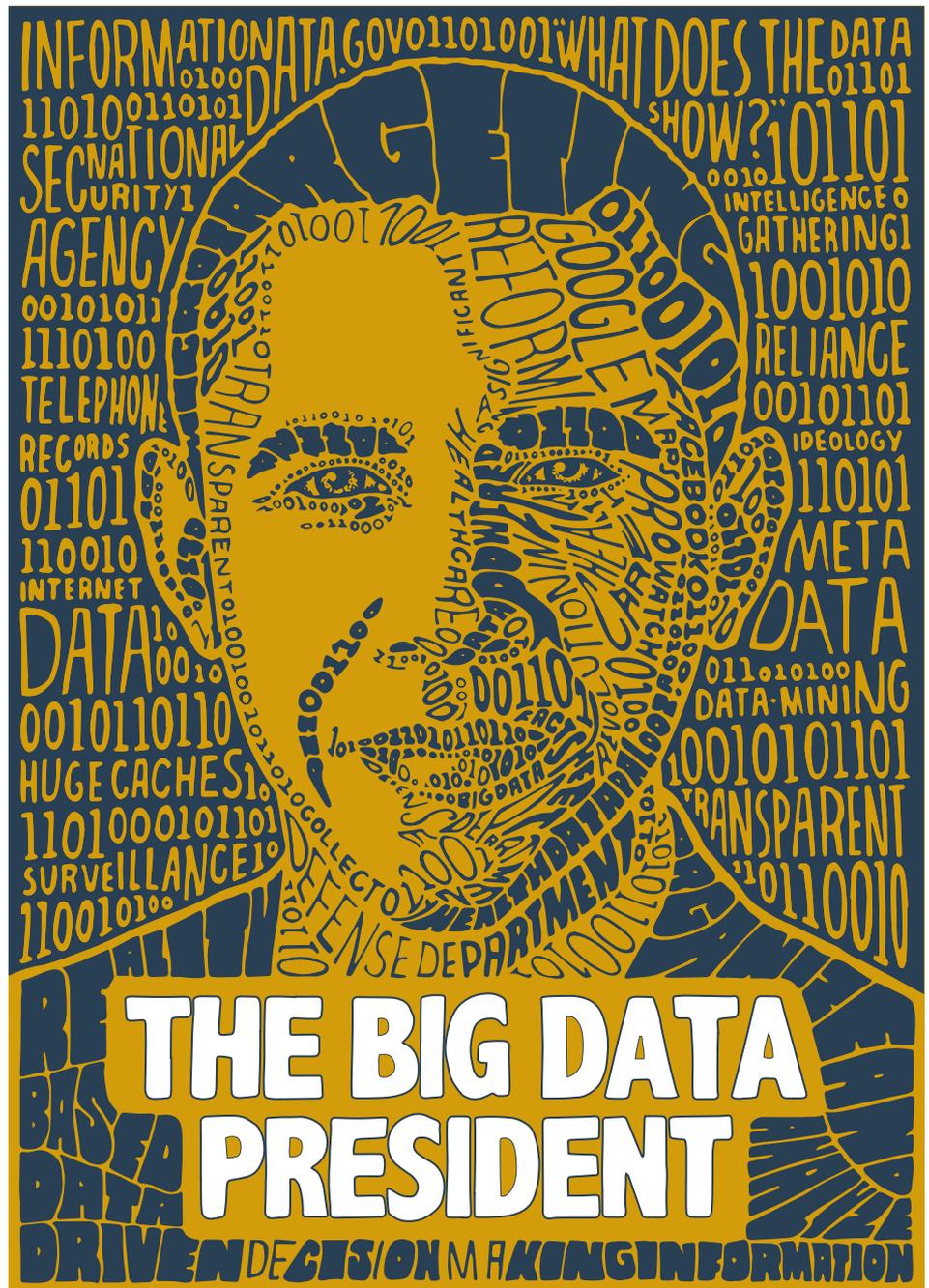
DANS LE MONDE

Les États-Unis, précurseur du Big Data

Google, Yahoo, Apache. Ces 3 noms sont indéniablement liés aux origines du Big Data. Une histoire qui trouve ses débuts dans les années 2000, alors que Google asseyait son leadership en tant que moteur de recherche, et que les spécialistes s'interrogeaient sur la technologie qui permettait au géant en devenir de proposer un service incomparable à ses concurrents. En 2003, Google publie un premier papier sur le Google File System, et révèle ainsi les premiers secrets de son succès. **En 2004 on découvre le fonctionnement de MapReduce**, et l'année suivante, Doug Cutting et Michael Cafarella, à l'époque employés chez Yahoo et inspirés par les travaux de Google, créent Nutch Search Engine, qui deviendra Hadoop. Le Big Data est né. En 2006, Yahoo lègue le projet à Apache, qui reste depuis le cœur névralgique d'Hadoop. Cette petite histoire montre bien en quoi le Big Data est né, s'est construit et a élu domicile outre Atlantique. De façon assez logique, les premiers projets de grande ampleur ont eu lieu en Amérique du Nord, comme celui de la société de retail TESCO, qui reste aujourd'hui l'un des cas d'école du Big Data.

Cependant il est important de noter que les principaux clients seraient, encore à ce jour, principalement les gouvernements nord-américains, qui demeurent les principaux investisseurs et les « early-adopteurs » du Big Data. Le gouvernement Obama n'y est pas pour rien. **En 2012, l'État américain annonçait la mise à disposition de 200 millions de dollars pour un fond de recherche sur la thématique du Big Data.** Dès sa campagne, Obama avait pressenti l'utilité du Big Data et les possibilités d'application multisectorielles. En 2012 on apprend l'existence d'études Big Data menées grâce à des informaticiens venus de Google et Facebook. Il est révélé en quoi elles ont permis de cibler les potentiels nouveaux électeurs pour Obama, et en quoi elles sont allées contre toutes les prédictions de suffrages en annonçant, et en permettant, sa réélection. **Le terme « Big Data President » faisait alors son apparition.** C'est donc tout naturellement que le second mandat du président Obama a été porteur de lourds investissements dans le domaine du Big Data. L'organisme MeriTalk, chargé d'encourager la collaboration dans le domaine de l'IT au niveau national, a publié en juin 2013 une étude estimant que le Big Data permettrait à l'État américain de réaliser 14 % d'économies, soit 500 milliards de dollars.

VOIR L'INFOGRAPHIE « SMARTER UNCLE SAM:
THE BIG DATA FORECAST » DE MERITALK



© Sarah A. King

Formation, recherche, sécurité nationale, santé, services publics... Des programmes aux noms les plus intrigants, tels ADAMS, Mind's Eye ou encore TCGA ont pour objectif de décupler les capacités et l'utilisation du Big Data dans les diverses administrations.

Le département de la défense – qui à lui seul regroupe plus de 10 projets Big Data et bénéficie de plus de 250 millions de dollars d'investissement annuel – mène notamment le programme ADAMS. Ce dernier a pour objectif de repérer des comportements anormaux, des changements d'attitude inquiétants chez un soldat ou un citoyen américain. Mind's Eye est également un projet de la DARPA (Defense Advanced Research Projects Agency). Il vise à améliorer les performances de reconnaissance vidéo et d'analyse automatisée. Le projet TCGA, porté par les archives nationales du cancer (The Cancer Imaging Archive), permettra d'analyser plusieurs péta-octets de données de séquences génétiques issues de patients atteints. Les exemples sont nombreux, et révèlent la

“le Big Data permettrait à l'État américain de réaliser **14 % d'économies**, soit **500 milliards de dollars**.”

motivation du gouvernement américain pour le Big Data.

Si l'État américain est capable d'estimer rapidement l'immensité des gains économiques que lui permettrait le Big Data, certaines entreprises privées ont fait le même calcul. **En termes d'investissement, suivraient donc les compagnies d'assurances, les banques, le secteur de la santé, le retail...** Début 2014, la très médiatique opération

“Le département de la défense à lui seul regroupe plus de 10 projets Big Data.”

d'Amazon a marqué un nouveau pas dans la façon d'aborder la relation commerciale et le processus de vente. Le leader mondial de la vente en ligne a en effet annoncé qu'il était désormais en mesure de prédire le prochain acte d'achat de chacun de ses clients, et donc de le préparer à l'expédition avant même qu'il ne soit commandé.

Le buzzword du Big Data se cache derrière cette opération. Une bonne campagne marketing a fait le reste. Car ce qu'effectue le géant du e-commerce, c'est une analyse Big Data prédictive et presque classique. La réelle innovation, c'est de transformer toute la chaîne logistique en adéquation avec les conclusions des analyses menées. En d'autres termes, ce qui différencie Amazon réside dans le fait qu'ils investissent et se fient aux analyses Big Data, et ne craignent pas de métamorphoser leur organisation qui pourtant ne montrait pas de failles majeures. Amazon joue là son rôle de leader, en prenant les devants sur une tendance qui ne devrait que se développer dans les années à venir.

Les pays émergents

L'Amérique du Nord est indéniablement en avance sur le sujet du Big Data. L'Europe suit la tendance, tout comme la plupart des pays développés, où le tissu économique constitué de nombreuses startups et de grands groupes nécessite et facilite l'adoption du Big Data.

Les regards se portent donc souvent sur le géant américain, alors que des investissements colossaux ont lieu sur le continent africain par exemple. Une étude menée par *MarketsandMarkets* indique que les pays émergents sont voués à grappiller leur retard dans les années à venir et représenteront une part importante de la croissance du marché du Big Data.

Quelques chiffres rendent compte du potentiel que représentent les pays émergents, ne serait-ce qu'au niveau des télécommunications. Selon la Banque Mondiale, il existe 6 milliards d'appareils mobiles sur Terre, dont 5 milliards se trouvent dans les pays émergents. Les applications marketing sont assez évidentes,

“Il existe 6 milliards d'appareils mobiles sur Terre, dont 5 milliards se trouvent dans les pays émergents.”

mais les pays émergents ne se limitent pas à ça. Services, industries, finance, programmes de développement économique : le Big Data stimule toutes ces économies, qui ont l'avantage d'être jeunes, agiles et innovantes.

IBM a mené une étude auprès de décideurs et d'influenceurs IT au Nigeria et au Kenya. 64 % d'entre eux déclarent prévoir adopter le Big data dans les prochaines années et le perçoivent comme un outil pour asseoir leur leadership dans leur zone. L'Afrique, en plein développement économique, n'est pas si en retard qu'on pourrait le croire. **L'Afrique du Sud est un réel pionnier en la matière : avec plus de 10 datacenters sur le territoire et une pénétration du Big Data dans 60 % des entreprises**, le pays s'affiche comme un précurseur sur le continent, et au niveau mondial.

Le Brésil, fréquemment associé au Big Data lors de la coupe du monde, voit se développer des projets dignes des plus grandes opérations américaines.

L'adoption du Big Data en entreprise

8,9 milliards de dollars, c'est le chiffre d'affaires que devrait générer le marché du Big data en 2014, selon une étude du cabinet Transparency Market Research. **Avec une croissance de près de 40 % par an, le marché représenterait déjà plus de 24.6 milliards en 2016.**

Malgré toutes ces promesses et prévisions, l'adoption du Big Data dans les entreprises françaises et européennes reste encore très limitée. Selon une étude menée par EMC fin 2013, 74 % des entreprises en France sont convaincues de l'intérêt du Big Data pour leur activité, mais **41 % d'entre elles n'ont encore engagé aucunes dépenses sur la question.**

En cause, la faible prévisibilité du retour économique de ces investissements (**35 % des décideurs s'interrogent encore sur le retour sur investissement**, selon un sondage EMC). Rassurons-nous, les entreprises françaises ne sont pas les seules. Contrairement à l'idée d'une Amérique fortement convaincue et engagée sur le sujet, une étude du cabinet Gartner aux États-Unis révèle que 56 % des entreprises ne savent pas comment tirer profit du Big Data. Jusqu'à présent, les projets impliquant un traitement massif de données n'étaient menés que par quelques leaders du marché. **Une étude publiée par DELL en avril 2014 affirme cependant que 41 % des entreprises de « middle market » - les challengers - ont désormais entamé un ou plusieurs projets Big Data.** Parmi eux, 89 % ont dès à présent identifié des retours positifs de leurs projets. Encourageant donc.

La source de ces interrogations se situe principalement sur le retour sur investissement. A priori, en matière de technologies, hardware et software, les outils se démocratisent, notamment grâce à l'apparition du Cloud. Cependant, le Big Data renferme d'autres coûts, moins mesurables et relatifs à l'approche culturelle de l'entreprise et à son organisation interne.

Si au niveau mondial il faut admettre que le secteur privé reste prudent sur les dépenses liées au Big Data, on peut s'interroger sur son adoption sectorielle. Y a-t-il des domaines d'activité où le Big Data est plus reconnu, adopté et accepté? De façon assez unanime, les spécialistes s'accordent à dire que les acteurs du e-commerce et les fonctions marketing ont été les premières à se lancer sur le sujet. Par nature ancrées dans l'innovation et en lien di-

rect avec les consommateurs, ces fonctions ont rapidement perçu les apports du Big Data pour leur activité. Rentabilité, efficacité et personnalisation sont les maîtres mots. Les concepts de RTB (real time bidding), la géolocalisation, les cookies, le suivi du parcours client, l'individualisation et l'optimisation du CRM ont alors révolutionné la fonction marketing. Des acteurs comme Criteo, notre champion français désormais expatrié aux États-Unis, ont apporté une nouvelle approche de la vente. Une approche fondée sur la « philosophie » du Big Data. Le numéro un mondial du e-commerce Amazon n'a pas investi massivement sur le Big data sans raison. Si en 2014 l'entreprise a initié la refonte de son processus de livraison, c'est que le management avait déjà fait la preuve de la rentabilité d'une bonne analyse de statistiques commerciales. En effet, **Amazon réaliserait environ 30 % de ses revenus grâce aux ventes croisées**, soit les suggestions d'achats effectuées grâce aux paniers et profils de ses clients.

“8,9 milliards de dollars, c'est le chiffre d'affaires que devrait générer le marché du Big data en 2014.”

74%

DES ENTREPRISES EN FRANCE SONT CONVAINCUES DE L'INTÉRÊT DU BIG DATA

Au sein de l'opinion publique, le Big Data est d'ailleurs souvent associé au marketing et à la vente, parfois avec une connotation négative. L'idée d'un client épié, disséqué et harcelé fait souvent peur. Or un courant nouveau assure qu'en tant que consommateur nous ne pourrions pas échapper à cette tendance et ferions mieux d'en savourer les bénéfices : une meilleure connaissance de nos goûts, des promotions personnalisées, des recommandations ciblées. Tout le monde y trouverait alors son compte. Reste la question de la protection des données personnelles, que nous aborderons plus tard.

Considérer le Big Data comme un serviteur du marketing serait cependant extrêmement réducteur. On connaît davantage les projets liés au marketing et à la vente parce que ce sont ceux qui sont les plus facilement intelligibles par le grand public et qui l'affectent directement. Les plus beaux projets Big Data se trouvent pourtant dans des secteurs que l'on observe moins, mais qui nous concernent tout autant. Logistique, maintenance prédictive, recherche, santé, énergie, culture et humanitaire même.

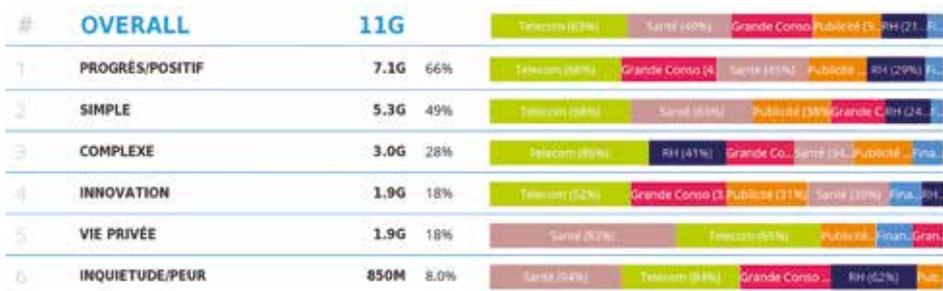
Les domaines d'applications sont vastes. La partie projet de ce guide présente quelques-uns de ces projets dans des domaines variés.

Conduire plus prudemment grâce au laboratoire d'accidentologie PSA-Renault, optimiser les consommations énergétiques d'un bâtiment avec Cofely, détecter une épidémie avec des outils comme OpenHealth ou encore endiguer la pauvreté avec les Nations Unies : le Big Data est bien une révolution de la société dans son intégralité.

Chaque secteur doit donc développer sa propre connaissance du Big Data, son intérêt stratégique, financier et les moyens à mettre en œuvre.

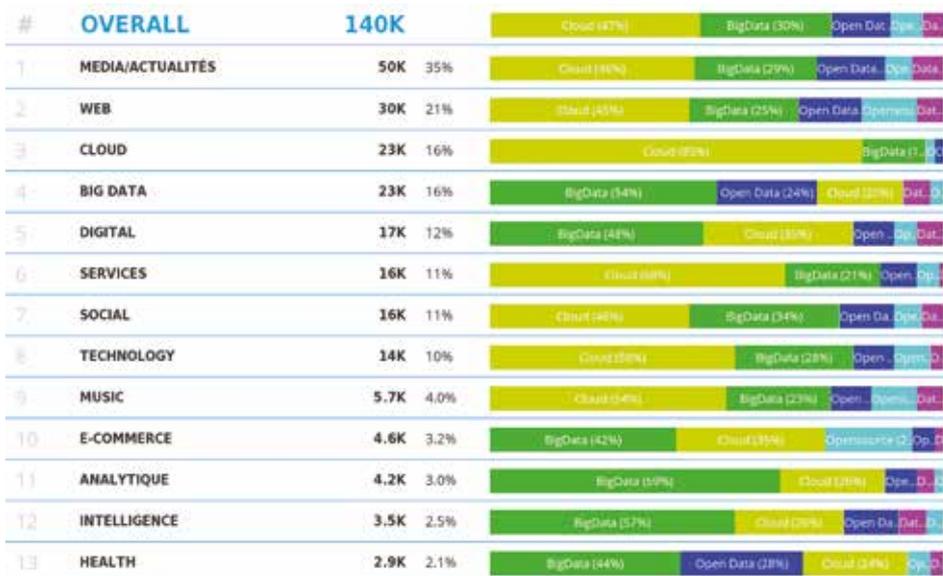
En termes d'image, la perception selon le domaine d'application est souvent connotée, positivement ou négativement.

À l'aide d'une écoute menée sur plus de 150K messages, sur la période du 27/10/2013 au 04/09/2014, avec Focusmatic, nous avons tenté de comprendre quels termes et sentiments sont le plus souvent associés au Big Data, et cela par secteur d'activité. Les Télécoms représentent la majorité des messages connotés, quel que soit l'aspect évoqué. En dehors des Télécoms, la santé, la grande consommation et la publicité sont les secteurs le plus souvent associés à des inquiétudes et à la vie privée. Dans les domaines d'activités des RH et de la grande consommation, le Big Data est encore un terme et procédé perçu comme complexe. (% des messages associés à, en audience).



La connaissance du Big Data en elle-même est discutable en France. Qui parle réellement du sujet ? Est-il si répandu qu'on le dit ? Ou sommes-nous encore aux prémices de ce phénomène ?

Sur la même période, nous avons observé le profil des auteurs qui parlent le plus du Big Data. Il s'agit ici d'analyser l'activité des auteurs ayant renseigné leur profil twitter sur le réseau utilisé. On remarque alors que pour les auteurs se réclamant être des spécialistes du Big Data, les thématiques principalement abordées sont dans l'ordre le Big Data, l'open data, le cloud, la dataviz puis l'opensource. Les auteurs spécialistes du « cloud » ont une activité aussi importante que les spécialistes du Big Data, bien que moins variée en termes de thématiques. Fait intéressant, ce sont les spécialistes de la santé qui dédient proportionnellement le plus de leurs messages à l'open data. (% des messages associés à, en activité. Du 27/10/2013 au 02/09/2014)



EN FRANCE / EUROPE

Le cas français : offre et demande, une incompatibilité ?

On parlait plus haut des impressionnants programmes d'investissement américains dans le Big Data pour ses administrations. Le cas de la France est bien différent, pour l'instant du moins.

Le plan Big Data pour la Nouvelle France industrielle, porté par François Bourdoncle et Paul Hermelin, et dont la feuille de route a été validée en juillet 2014, est à ce jour la première réelle impulsion économique émanant de l'État visant à développer l'écosystème Big Data en France.

Une question d'approche et de stratégie se pose alors, et on remarque, comme c'est souvent le cas, que la démarche américaine est bien différente de celle menée par l'État français.

Les deux gouvernements sont certes, désormais, tous deux persuadés que le Big Data est un secteur d'avenir, multisectoriel, plein de promesses économiques, sociales et environnementales. Mais les deux gouvernements n'ont pas la même façon de stimuler le secteur. Outre Atlantique, la prise de position du gouvernement Obama passe par le développement de projets financés par les administrations et dédié à leur propre utilisation. Ce sont ces grands projets d'envergure qui sont censés servir les objectifs de l'État mais aussi montrer la voie. Si l'État est réellement persuadé de l'intérêt de l'utilisation du Big Data, pourquoi ne commencerait-il pas par l'utiliser lui-même ? Le raisonnement est pragmatique, mais efficace.

On ne procède pas exactement pareil en France. Tout d'abord, nous avons pris du retard par rapport à nos voisins américains. Nous connaissons nos forces — de très bonnes compétences en mathématiques et statistiques — mais nous devons développer nos potentiels. Pour ce faire, le choix du gouvernement est de construire les conditions de l'épanouissement d'une filière Big Data en France. Le discours, les structures d'accueil, et les enveloppes budgétaires, sont là pour atteindre cet objectif. Quelques projets seront également menés au sein des administrations, et on ne peut pas ignorer l'important engagement sur le sujet de l'Open Data, notamment via Etalab. Cependant, il est évident que l'approche n'est pas la même.

Si la France souffre encore fortement de la crise économique et peine à défendre son statut à l'international, elle dispose encore de grands groupes industriels puissants et agiles. C'est là que l'action du gouvernement sur le sujet du Big Data trouve un appui non négligeable. Les besoins de financements, les ressources humaines, la capacité d'innovation et les structures de R & D sont disponibles au sein du CAC40. Le plan de la nouvelle France industrielle, et sa nouvelle feuille de route, puise sa force exactement dans cet environnement favorable. **Orange, La Poste, GDF Suez, Alstom, AXA, Société Générale, Covéa (le groupe MMA, GMF et MAAF)**... Ces grandes entreprises ont été choisies pour participer aux grands travaux du plan initié par Arnaud Montebourg, et désormais chapeauté par François Bourdoncle et Paul Hermelin.

Quatorze initiatives qui couvrent la **protection des données personnelles, la fraude, le développement des objets connectés, la ville intelligente** et d'autres applications sectorielles.

objets connectés **fraude**
données **personnelles**
ville intelligente



François Bourdoncle et Paul Hermelin à l'ouverture de la conférence BIG DATA Paris 2014

Objectif affiché : contrer l'impressionnante suprématie et capacité d'innovation des groupes américains comme Google, Apple, Amazon ou Facebook.

En définitive, l'approche française est bien différente de l'approche américaine. Elle prend cependant en compte ses forces, tente de contrer ses faiblesses, et s'autorise de grandes ambitions. Une dynamique positive donc. Reste à savoir quelle sera réellement l'enveloppe allouée à ces projets — via la Bpi notamment — l'implication des diverses parties prenantes et le résultat opérationnel. **On parle de dizaines de millions d'euros**, vraisemblablement puisés dans les 215 millions débloqués par Fleur Pellerin en 2013.

Reste à noter quelques absents non-abordés dans cette feuille de route, et pourtant abordés dans la construction du plan Big Data : la formation, l'évangélisation, et la réglementation. Sur le point de vue de la formation, nous en reparlerons plus loin, mais les initiatives sont nombreuses et les besoins en termes de recrutement génèrent plus ou moins naturellement les formations supérieures adaptées. L'enseignement primaire et secondaire reste à étudier. **« On enseigne le latin au collège, je ne vois pas pourquoi on n'y enseignerait pas l'informatique »** déclarait Fred Potter, CEO et fondateur de Netatmo, lors du dernier Hack4France. Sur le plan de l'évangélisation, c'est-à-dire le discours de vulgarisation auprès des potentiels utilisateurs du Big Data en France, on peut également se demander si le marché agira par lui-même. Le meilleur moyen de promouvoir le Big Data dans les entreprises est de montrer des cas d'applications concrets, à succès et reproductibles. Si les quatorze initiatives du plan Big Data affichent un bilan positif, l'État tiendra alors les meilleurs ambassadeurs possibles. Le reste suivra, a priori.

La question de la réglementation est bien différente. L'action de la CNIL, les discussions au niveau européen, les disparités au niveau mondial : les éléments de réglementation sont complexes, pris en étau entre intérêts économiques, protection des données personnelles et sécurité de la propriété intellectuelle. Le débat mérite plus de détails.

Données personnelles et cyber-sécurité : enjeux technologiques et réglementaires

Le Big Data ne peut désormais plus être dissocié des problématiques de sécurité. Le terme est vaste et renferme en réalité plusieurs notions bien distinctes.

Il renvoie dans un premier temps, et c'est là que se situe l'essentiel du débat public, à la protection des données personnelles. **Une donnée personnelle** telle qu'elle est définie par la CNIL (Commission Nationale de l'Informatique et des Libertés) est **« toute donnée permettant d'identifier directement ou indirectement une personne physique »**. On parle donc bien ici d'individus et non pas d'organismes, entreprises ou associations. Sont exclues les données utilisées dans le cadre d'une activité personnelle — comme un répertoire téléphonique par exemple. Leur utilisation est soumise à la loi Informatique et Libertés et à la directive du 24 octobre 1995 sur la protection des données personnelles. L'histoire de la loi Informatique et Libertés remonte bien avant l'apparition du Big Data. Elle a été depuis adaptée à de nombreuses reprises, mais à l'origine la loi est née d'un scandale digne de la très médiatisée affaire PRISM. Le gouvernement français avait alors pour projet la création d'un fichier administratif national identifiant chaque citoyen

par un numéro et interconnectant les fichiers. SAFARI, c'était son nom, a alors été vivement décrié, entraînant **la création d'une commission informatique et libertés qui proposa la loi en 1978**. L'opinion publique était alors sensibilisée à l'utilisation de ses données.

Les données marketing que nous évoquons plus haut sont fortement concernées. Le concept de l'opt-in est à ce jour la principale mesure permettant de contrôler l'information personnelle que nous divulguons, mais les outils numériques sous-entendent un minimum d'ouverture de l'information.

Il semble que suivre les tendances des nouveaux modes de consommation et communication et parfaitement protéger ses données personnelles soit incompatible. S'en suit une certaine schizophrénie de la société, hésitant entre un désir de modernité et la volonté de préserver son intimité et ses libertés.

[VOIR LE SITE DE LA CNIL](#)



“Le volume de données produites par les internautes doublerait tous les **18 à 24 mois.**”

Les objets connectés sont au cœur du débat. Ils véhiculent inévitablement une masse importante d'informations sur nos modes de vie, notre façon de consommer, nos habitudes, préférences ou sur notre profil. D'après les estimations, **en 2018, chacun d'entre nous possédera en moyenne 8 objets connectés à titre personnel. En 2020, nous en aurons déjà 10.** La tendance ne risque pas de s'atténuer, étant donné que, selon un sondage publié par Havas Media France en janvier 2014, 60 % des internautes envisagent la généralisation des objets connectés d'ici 5 ans, et les perçoivent comme source de progrès (75 %) qui facilitent la vie (71 %). Le « quantified self », en français le « soi augmenté », représente l'humain du futur, un futur très proche.

Courant 2014, un ingénieur a fait une étrange découverte. Visionnant des publicités qui lui paraissaient de plus en plus ciblées, il a compris que son téléviseur LG espionnait son comportement à son insu. Le fabricant avait activé par défaut une fonction de collecte de données. Après avoir déconnecté cette dernière, l'ingénieur a identifié que le téléviseur continuait à transmettre des informations sur ses habitudes télévisées et même ses visionnages de fichiers privés. L'affaire a été révélée au grand jour et LG a dû présenter publiquement ses excuses tout en s'engageant à supprimer toute collecte de données de ses téléviseurs. La méfiance s'est alors généralisée à d'autres objets potentiellement connectés à l'insu de leurs utilisateurs. Le parallèle avec George Orwell semble alors évident, nourrissant une certaine psychose. L'équilibre est nécessaire et c'est là que se joue le rôle de la CNIL et des instances européennes, toutes conscientes de l'urgence d'une réglementation assurant et rassurant les citoyens.

Les objets connectés seront donc part intégrante de notre quotidien. Là où ils soulèvent encore davantage d'inquiétudes, c'est lorsqu'ils concernent la santé connectée. Bracelets, podomètres, balances, tensiomètres, appareils sportifs, assistants de confort, aide aux personnes âgées, contrôles médicamenteux... Les données de santé sont considérées comme particulièrement sensibles. Elles permettent de caractériser un individu sur des éléments extrêmement précis et privés. Ces données sont traitées à part dans les actions de la CNIL. Les acteurs du monde de la santé sont tenus de s'engager sur l'anonymisation totale des données sensibles, données de santé comprises. La recherche médicale en est parfois pénalisée, mais l'importance de la sécurisation de ces données est désormais acceptée. Quelques scandales ont participé à cette prise de conscience, notamment lorsque certains patients ont retrouvé leur dossier médical en accès libre sur internet après un passage à l'hôpital. Comme l'explique Pierre-Yves Lastic, Chief Privacy Officer chez SANOFI, l'utilisation des données de santé doit être particulièrement précautionneuse. S'assurer que les informations sont parfaitement anonymisées est un prérequis, quel que soit le pays où se déroule leur traitement et la réglementation en vigueur à cet endroit. Un projet de recherche, même à des fins tout à fait honorables, ne doit pas justifier l'utilisation de données aussi sensibles. Le monde de la santé joue avec des informations d'une haute valeur aux yeux de la société. Il se doit de les protéger.

Les objets connectés cependant ne touchent pas uniquement les données personnelles. La machine-to-machine par exemple, ou M2M, vise à recueillir un maximum d'informations sur des installations diverses afin d'optimiser leur fonctionnement. Selon l'Idate, **plus de 80 milliards de produits seront connectés à Internet d'ici 2020.**

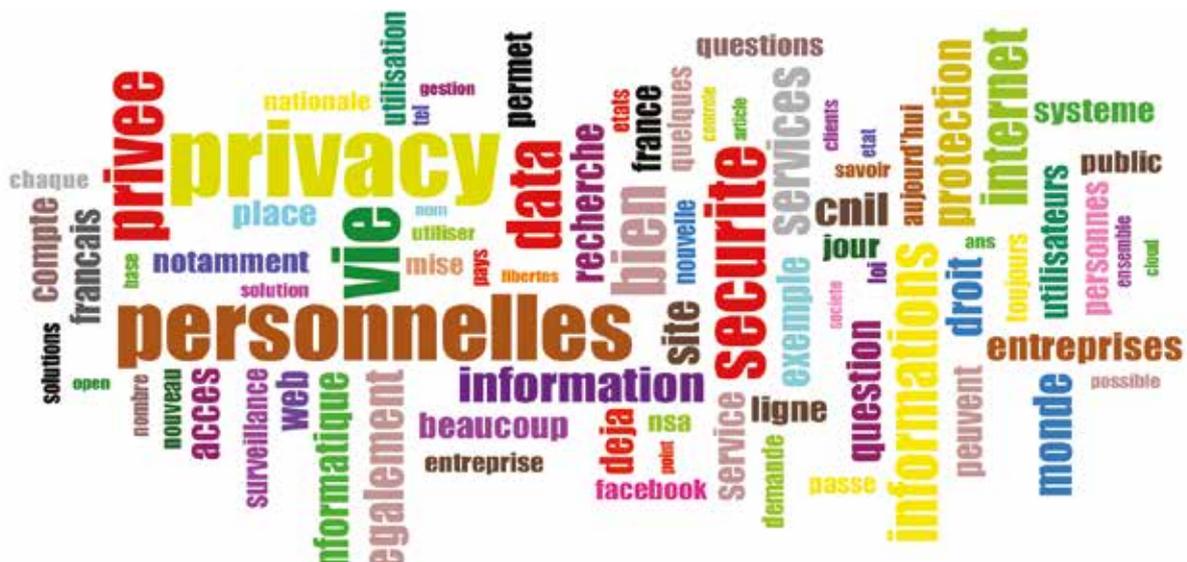
Dans l'industrie, des sociétés comme Airbus ou Total utilisent déjà massivement les capteurs pour leurs projets Big Data. Début 2014, Google investissait 2,3 milliards de dollars dans le rachat de la société Nest Labs, confirmant que le marché des objets connectés est voué à une croissance impressionnante.

Au-delà de ce type de données, les entreprises dans leur ensemble détiennent une masse d'information considérable. Le volume de données



Parlement Européen

produites par les internautes doublerait tous les 18 à 24 mois. C'est une source de valeur souvent majeure pour un acteur privé comme public, et également un point de faiblesse non négligeable. L'affaire PRISM, révélée par Edward Snowden en juin 2013 a mis le doigt sur la criticité et la valeur des données. Le Big Data est censé porter une révolution de la donnée. Cette dernière devient alors une arme puissante sur le plan économique comme géopolitique. Dans la sphère privée, on parle principalement de viol de l'intimité ou d'espionnage industriel, mais dans le domaine public il s'agit bien de jeux de pouvoirs. Le Patriot Act caché derrière l'affaire PRISM révèle comment le gouvernement américain a mis en place les conditions d'une récolte d'informa-





tions globale et continue. La structure s'appuie sur l'omniprésence d'acteurs comme Google, continuellement aux prises avec les législateurs européens, mais devenu presque incontournable pour leurs internautes. Cependant, quelques chiffres rassurants ont été publiés par la CNIL dans son rapport annuel de 2014 : **la commission a constaté une baisse des plaintes relatives à la protection des données et une augmentation de 17% des demandes de droit d'accès indirect.** L'utilisateur va donc peut-être apprendre lui-même à gérer ses données, si on lui donne les outils nécessaires.

Avec Focusmatic, nous avons observé quels termes sont les plus présents lorsque l'on parle de big data et de vie privée sur le web et les réseaux sociaux (% des messages associés à, en activité. Du 27/10/2013 au 02/09/2014). On remarque que certains acteurs ou organismes y apparaissent, comme en premier la CNIL mais aussi Facebook, la NSA ou encore l'État.

Vers une cohésion européenne

Dès lors qu'on considère que les données représentent un élément géopolitique et économique majeur, on est en droit d'estimer que les États et plus largement l'Union Européenne est concernée. L'affaire PRISM a brisé une forme de confiance tacite entre l'Europe et les États-Unis, et a poussé divers acteurs européens à s'interroger sur la façon de protéger le patrimoine et les intérêts de l'Union. La question de l'hébergement des données est particulièrement prégnante. Si la majorité des données produites par la zone européenne est stockée au sein d'entreprises américaines,

elles-mêmes soumises au Patriot Act, il est légitime de s'inquiéter de leur utilisation à des fins néfastes. Quelques échanges politiques ont bien eu lieu, visant à rassurer les divers États de la bonne foi de chacune des parties. En parallèle, la volonté de créer un « cloud souverain » s'est renforcée, avec des acteurs capables d'héberger les données nationales et de les protéger (Numergy, Cloudwatt). Là, diverses opinions s'opposent : les uns prônent des actions rapides afin de sécuriser les données, les autres sont convaincus de la nécessité de préserver le libre-échange. La surprotection n'a pas toujours, dans l'histoire, porté ses fruits. Et si l'Europe risquait plus à tenter de s'enfermer dans un système de vase clos qu'à se laisser scruter ?

De plus, une volonté d'assurer une relative harmonie sur la question de la protection des données personnelles en Europe se fait ressentir. À ce jour, les discussions font encore vivement débat. **La dernière avancée majeure sur ce sujet date du 21 Octobre 2013**, lorsque la commission des Libertés publiques (LIBE), alors largement portée par Viviane Reding, vota le compromis sur la réforme de la protection des données dans l'UE. Le site de la CNIL qualifie ce vote de « *signal politique puissant qui exprime une identité politique forte de l'Union européenne sur un sujet essentiel, tant sur le plan des valeurs que sur celui des enjeux économiques.* » Il poursuit en déclarant que « *concernant les suites de l'affaire PRISM, et comme la CNIL l'avait expressément demandé dès début 2013, le texte introduit un contrôle des autorités de protection sur les demandes d'autorités administratives et judiciaires de pays tiers d'accéder aux données relatives à des citoyens européens. Cette réponse, certes partielle, mais politiquement*

importante est une première étape vers la construction d'un cadre juridique protecteur à l'encontre des activités de renseignement de pays tiers. » On y découvre alors notamment la notion de droit à l'oubli. A priori les débats vont donc dans la bonne direction, mais certaines zones d'ombre persistent. D'ici fin 2014, plusieurs déclarations sont attendues, à l'image du projet de loi sur la géolocalisation dans le cadre d'enquêtes discuté en Février dernier.

Au niveau européen, Neelie Kroes, la vice-présidente de la Commission européenne en charge du numérique, s'est désormais construit une image forte et multiplie les déclarations. Elle porte la mouvance pro-numérique au sein des instances européennes. Critiquée pour sa position jugée faible face aux lobbies des opérateurs télécom, elle est reconnue pour ses actions en termes de cybercriminalité. Si Angela Merkel a fait beaucoup de bruit suite à la révélation de la mise sur écoute de son téléphone par la NSA (National Security Agency), il semble encore discutable de mettre en place des actions de contrôle et d'hébergement strictement européen des données. D'un point de vue pratique, stratégique et économique, les avis divergent, même si le débat devra statuer rapidement. La pression exercée par les désormais surnommés GAFA (Google, Amazon, Facebook, Apple) joue un rôle central dans cette relation de force.

L'open data et la création de valeur

L'open data c'est, littéralement, l'ouverture des données. Données publiques, données privées, données anonymisées, données environnementales... Dès lors qu'une donnée n'est pas personnelle ou stratégique pour son détenteur, elle peut être partagée. L'idée derrière cette approche, c'est que **l'ouverture des données permet la création de valeur.** Une information détenue par un acteur d'un certain secteur, qu'il n'utilise pas particulièrement, peut, si elle est partagée, permettre à un autre acteur de développer une innovation, un service, une analyse.

Sans pénaliser les acteurs qui ouvrent leurs données, l'open data est profondément ancré dans la vision participative de l'économie. Ouvrir ses données, c'est aussi contribuer à un processus qui permettra de générer de la valeur dans un écosystème, et s'inscrire dans une démarche d'émulation qui sert les intérêts de tous.

Selon un rapport du cabinet McKinsey, **l'open data pourrait contribuer à générer chaque année une valeur ajoutée de plus de 3 000 milliards de dollars.** Cette étude a mis en alerte les plus sceptiques, et désormais l'ensemble de la communauté Big Data voue une importance indéniable à l'open data, observant les divers projets qui naissent à travers le monde.

L'OPEN DATA EN INFOGRAPHIE



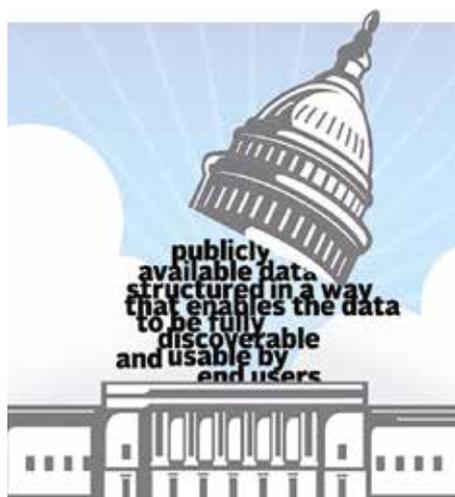
**“13 259 jeux de données
sont disponibles à ce
jour sur la plate-forme
gouvernementale
dédiée à l’Open data.”**

En juillet 2014, la ville de Chicago a initié un projet de collecte de données sur 8 lampadaires installés à travers la ville. Le Centre urbain de calcul et de données (UCCD) aura ainsi à sa disposition des données météorologiques, de trafic urbain, de pollution, d’environnement sonore et olfactif, et de fréquentation grâce aux signaux émis par les smartphones. Les données seront ensuite ouvertes.

La question de la protection des données est évidemment engagée, mais la ville assure que l’usage ne sera qu’à des fins de recherche, développement de services, et restera anonyme. L’objectif de ce projet s’inscrit parfaitement dans la dynamique d’Open Data à des fins de création de valeur et de génération de start-ups et de projets innovants. Les premiers résultats permettront de décider si l’expérience doit être étendue dès fin 2014, et si d’autres villes peuvent s’en inspirer.

L’investissement est ici limité puisqu’il représente moins de 300 euros par boîtier installé, et s’appuie sur les connaissances et ressources de l’UCCD.

La France est loin d’être en retard sur le sujet de l’open Data. L’action d’Etalab, l’engagement de l’État et la collaboration de nombreux grands groupes ont permis d’ouvrir un grand nombre de données et de créer diverses start-ups et services innovants. 13 259 jeux de données sont disponibles à ce jour sur la plate-forme gouvernementale dédiée à l’Open data (data.gouv.fr). **Cela représente plus de 350 000 fichiers contenant des données publiques.** Et les actions sont vouées à se multiplier. Porté par Henri Verdier, dont vous pourrez retrouver une interview dans la partie portraits, Etalab se positionne comme une communauté favorisant le partage d’information, l’émulation et la création de valeur.



Source : U.S. General Services Administration



**FOCUS
BIG DATA & FORMATION :
LES PROGRAMMES, LES
PROFILS, LA CRÉATION
D’EMPLOIS**

**MISE EN PLACE DU BIG DATA : LE RÔLE
DES MÉTIERS, LA PLACE DE LA DSI**

Comme le montrent les chiffres, l’adoption du Big Data reste relativement modérée dans les entreprises françaises. Plusieurs points de vue peuvent expliquer ce relatif retard. D’après Philippe Nieuwbourg, les Français réfléchissent trop à l’architecture de leurs projets Big Data, fixent trop d’objectifs et ne laissent pas la place à l’expérimentation. À l’image des POC (Proof Of Concept), des projets pilotes, largement adoptés en Amérique du Nord, l’analyste estime que **la France devrait insuffler plus de souplesse dans son approche.** La « philosophie » Big Data n’est-elle pas justement fondée sur l’expérimentation ?

À l’inverse, Fabrice Benaut, ancien DSI du groupe GFK, pense que la prudence et **la rigueur appliquée aux projets français constitueront notre force dans les années à venir.**

Quoi qu’il en soit, à ce jour, les projets Big Data semblent effrayer un grand nombre d’entreprises. Trop risqué, trop cher, trop compliqué à mettre en place, trop loin de la culture de l’entreprise... Les arguments sont nombreux pour repousser l’échéance en scrutant les premiers

retours des projets les plus médiatisés. Au-delà des considérations économiques et de l’intérêt que présente le Big Data, le facteur humain est souvent évoqué. Qui mènera le projet ? Quelles sont les qualifications nécessaires ? Comment impliquer les diverses équipes concernées ? Comment coordonner les projets ?

Ces dernières années, le terme de datascientist est apparu. Un mouton à cinq pattes pour certains, capable de comprendre les problématiques métier tout en étant force de proposition sur les outils et infrastructures. Ce datascientist est-il issu de la DSI ? Peut-être, mais il doit davantage se positionner comme un électron libre au sein de l’entreprise, naviguant entre les divers services et fédérant les équipes autour du même objectif. Un manager en somme ? Pas tout à fait. Un leader plutôt. Un homme ou une femme, qui saura appréhender les besoins de ses collaborateurs, connaître les solutions Big Data qui y répondront et les mettre en place rapidement. Un datascientist doit également faire preuve de qualités relationnelles : en tant qu’électron libre, il doit pouvoir impliquer ses collaborateurs, rendre intelligibles les résultats et leur communiquer l’avancement. C’est celui qui vulgarisera le Big Data dans l’entreprise, pour le rendre plus accessible aux divers métiers.

Une étude menée par IT Research réalisée en partenariat avec le Club Décision DSI et le JDN révèle que 35 % des DSI s’orientent vers le Big Data en France. Ils réfléchissent aux infrastructures et aux finalités des projets, mais aussi au nouveau rôle qu’ils incarneront dans leur entreprise avec le Big Data. L’image des DSI en France n’est cependant pas toujours

35%

DES DSI S'ORIENTENT VERS
LE BIG DATA EN FRANCE

Soraya, Mehdi et Denis, diplômés du Mastère Spécialisé Big Data Télécom ParisTech la première promotion BigData en France

très positive. Souvent considérés comme des départements renfermés sur eux-mêmes, les DSI qui souhaiteront prendre en main le Big Data devront adopter le profil du datascientist.

Les formations

C'est donc bien d'un nouveau profil dont on parle. Un profil qui requiert une formation mathématique, statistique, informatique mais aussi managériale.

C'est la raison pour laquelle de nombreuses formations ont ouvert leurs portes récemment. **Telecom Paristech a ouvert la voie en diplômant la première promotion Big Data de France en Mai dernier.** Une promotion dont la majeure partie des étudiants avait déjà signé son futur contrat d'embauche avant de recevoir son diplôme. Grenoble EM et l'EMSI ont ouvert en septembre leur propre formation, s'appuyant sur leur complémentarité. Les grandes écoles parisiennes d'ingénieur ou de commerce suivent elles aussi la tendance. Où recruter son futur datascientist ? Y a-t-il des spécialisations plus ciblées ? Difficile de s'y retrouver désormais dans la multiplicité de l'offre. Certaines écoles sont identifiables rapidement de par leur réputation, mais nous avons tenté d'observer les principales formations et de voir à quels domaines de spécialisation elles sont le plus souvent associées.

Avec notre partenaire Focusmatic, nous avons sélectionné quelques écoles qui ont initié des formations liées au Big Data et tenté d'identifier les secteurs qui sont attachés à ces formations (nombre de messages associés à, en activité. Du 27/10/2013 au 02/09/2014). On remarque ainsi que la formation grenobloise est associée aux sujets RH et Telecom, tout comme celles d'HEC ou de l'ENSAI. La thématique RH est souvent mentionnée de par la pénurie de datascientist à ce jour en France. Ce qu'on remarque surtout, c'est que certaines écoles ont un écho dans une variété élevée de secteurs alors que d'autres sont évoquées pour des enjeux bien précis. L'ESSEC par exemple, est principalement associée au secteur de la publicité. HEC est l'école la plus associée au secteur de la finance, tandis que Telecom Paristech et les Mines le sont à la santé.

De façon globale, **HEC, l'ESSEC, la formation grenobloise (Grenoble EM & Ensimag) et Telecom Paristech sont les écoles qui font le plus parler d'elles et du Big Data** (nombre de messages évoquant le Big Data et l'école- période du 27/10/2013 au 04/09/2014).

Retrouvez
l'interview de

Soraya



Retrouvez
l'interview de

Mehdi



Retrouvez
l'interview de

Denis

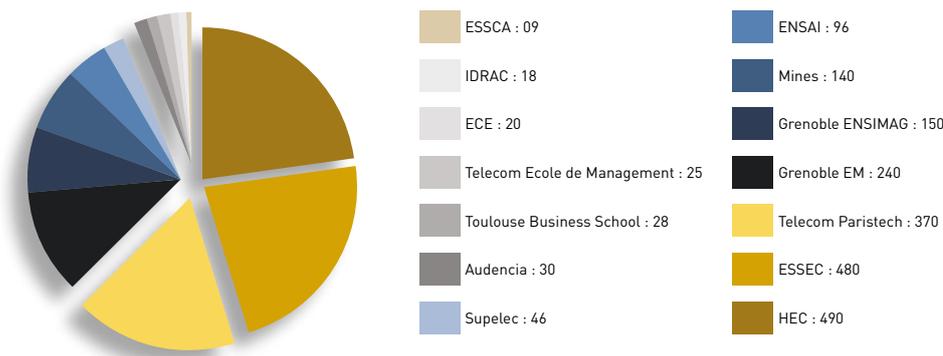


Secteurs associés, par Ecole



Données issues d'une étude menée avec Focusmatic du 27/10/2013 au 02/09/2014. Nombre de messages associés à, en activité

le Big Data et les Ecoles



Données issues d'une étude menée avec Focusmatic du 27/10/2013 au 04/09/2014. Nombre de messages évoquant le Big Data et l'école

“L'Afdel estime que
1 000 emplois directs
seront créés en France
d'ici 2018.”

LA CRÉATION D'EMPLOI

La croissance annoncée et amorcée du marché de la Big Data a une conséquence sur le marché de l'emploi. Comme nous l'avons vu, les projets Big Data nécessitent une réelle nouvelle fonction, celle de datascientist. Elle stimule également toute la sphère économique des SSII, cabinets de conseils, intégrateurs, fournisseurs hardware... Dans l'ensemble, l'Afdel estime que 1 000 emplois directs seront créés en France d'ici 2018. Au niveau mondial, **le cabinet Gartner a suscité l'effervescence en annonçant en 2013 qu'il évaluait la création d'emploi à 4.4 millions d'ici 2015.** Près de la moitié le serait sur le territoire américain. On comprend donc le besoin de formation et la prolifération des programmes académiques. Certains vont même jusqu'à penser que l'éducation nationale devrait intégrer des modules de familiarisation avec le codage et le traitement de la donnée, afin de susciter des vocations et d'éveiller la curiosité dès le plus jeune âge. Les générations à venir seront certainement plus familiarisées avec les nouvelles technologies, mais les datascientist demanderont quoi qu'il arrive une formation spécifique.

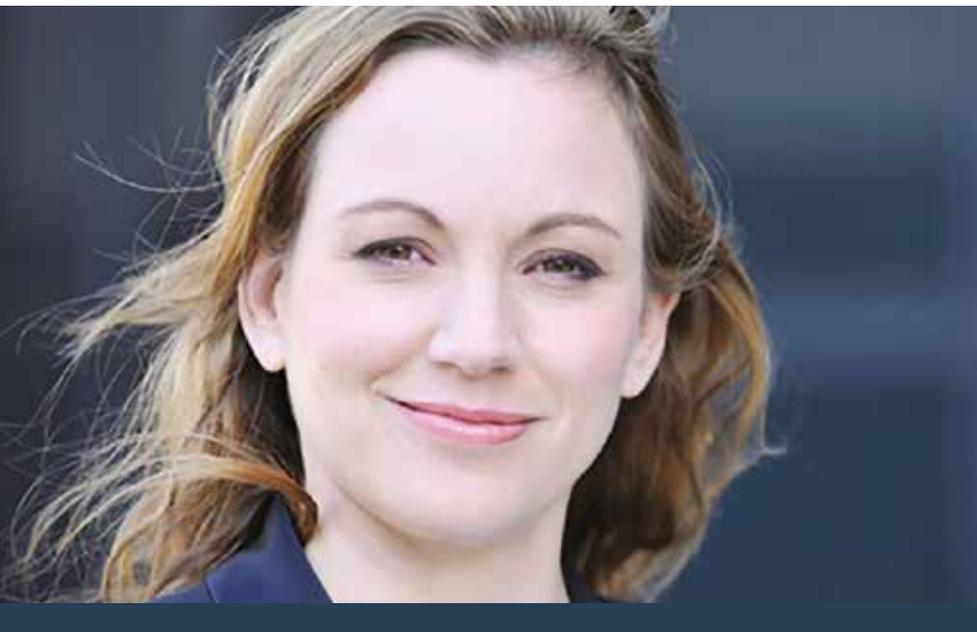
Dans le contexte d'un marché de l'emploi plutôt morose en Europe, la perspective d'une nouvelle fonction fortement demandée dans les plus grandes entreprises attire de nombreux étudiants, qui remplissent déjà les bancs des meilleures écoles. Argument non négligeable : **le salaire moyen d'un datascientist aux États-Unis serait de 89 000\$ annuels. En France, on parle de 40 000 à 80 000 € par an pour les plus expérimentés.**

4.4

MILLIONS D'EMPLOIS BIG DATA SERONT
CRÉÉS DANS LE MONDE D'ICI 2015

AXELLE LEMAIRE

SECRÉTAIRE D'ÉTAT
CHARGÉE DU NUMÉRIQUE
(Portrait)



Plus vraiment besoin de présenter la Secrétaire d'État chargée du Numérique. Sa prise de poste le 9 Avril 2014 a fait beaucoup de bruit dans la communauté numérique française.

Non pas que les compétences de la jeune femme aient été mises en doute, c'est surtout la forte popularité de celle qui fut son prédécesseur, Fleur Pellerin, qui anima les débats. En effet, l'ex ministre déléguée à l'économie numérique avait mis quelque temps à faire ses preuves, mais elle avait su gagner le respect et la confiance de la majorité de la sphère IT française. Imprégnée du sujet et réellement concernée par les enjeux du secteur, elle s'était positionnée en tant que moteur politique et économique du rayonnement numérique français à l'international. À l'apogée de son succès, le hashtag #keepfleur a traduit le soutien des acteurs du numérique lors du changement de fonction de Fleur Pellerin.

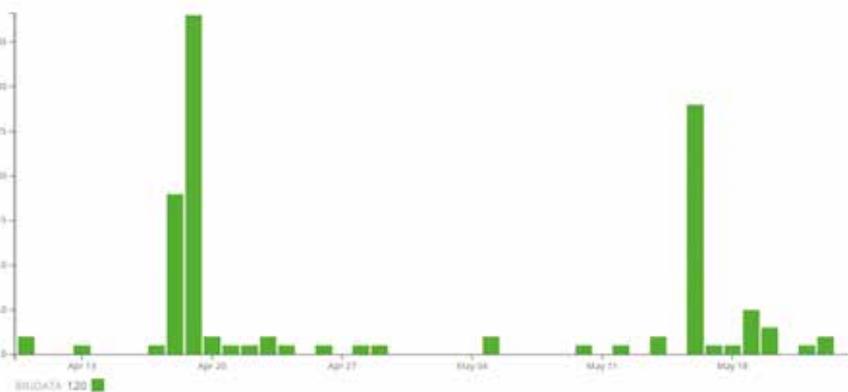
Une prise de poste compliquée pour Axelle Lemaire donc. Prendre ses marques et s'emparer du sujet tout en évitant la comparaison et les rancœurs des supporters de Fleur se présentait comme un équilibre difficile à trouver. Elle-même admettra quelques mois plus tard, lors d'une interview accordée au journal Le Monde, avoir « eu un peu peur en voyant le hashtag #keepfleur sur Twitter ». Cependant, il ne faut pas oublier que la jeune femme a déjà une carrière politique étoffée et ne découvre pas tout à fait la question du numérique. Au cours de ces dernières années, elle a développé un intérêt pour le sujet. Dans un premier temps attirée par le numérique dans le cadre de sa vie privée, elle a rapidement compris les intérêts économiques et les enjeux qui se cachent derrière ce marché en pleine expansion. Alors membre de la commission des Affaires européennes, elle rédige en 2013 un rapport sur la stratégie numérique de l'Union Européenne. De profil international — elle est née à Ottawa et a été députée des Français établis hors de France — Axelle Lemaire présente l'avantage de pouvoir naviguer facilement dans l'univers anglo-saxon, prédominant dans le numérique.

Si la suite de Fleur Pellerin a certes été difficile, on ne peut s'empêcher de noter une petite nuance qui renferme tout de même quelques réalités. Le 8 Avril, le poste occupé par Fleur Pellerin était celui de ministre délégué, le 9

« L'Europe n'est pas les États-Unis, nous ne voulons pas d'un internet fractionné. [...] Si on parvient à s'accorder au niveau européen, ce sera un grand pas en avant »

Avril, celui endossé par Axelle Lemaire porte le titre de secrétaire d'État. On a beaucoup vu Fleur Pellerin en première ligne sur la thématique du Big Data notamment. Pour Axelle Lemaire, la fonction se présente différemment. Officiellement, le poste concerne « les droits et libertés fondamentaux dans le monde numérique et la sécurité des échanges, des réseaux et des systèmes d'information ». En d'autres termes, elle est principalement en charge de la gouvernance Internet, la sécurisation des don-

LA TOP ACTU D'AXELLE LEMAIRE



(Données Focusmatic - messages associant « Big Data » et « Axelle Lemaire » - messages issus du web et des réseaux sociaux)

Afin de mesurer la présence numérique de la secrétaire d'État et ses interventions les plus remarquées lors de sa prise de poste, nous avons observé ses quelques actus Big Data qui ont fait le buzz entre le 09/04/2014 et le 24/05/2014

SES 2 PICS D'ACTUALITÉS SE TROUVENT

TOP #1 : LE 18/04/2014 GRÂCE À CET ARTICLE :

EXCLUSIF. Axelle Lemaire veut « rétablir un Internet qui garantit les libertés fondamentales » - Le Point.fr



TOP #2 : LE 19/05/2014 GRÂCE À CETTE INTERVIEW :

Axelle Lemaire : « L'Europe n'est pas les États-Unis, nous ne voulons pas d'un Internet fractionné ». - 20 Minutes



nées et la question de l'open data, le tout sous l'égide du ministre de l'économie, de l'industrie et du numérique. Sous l'égide ou dans l'ombre ? C'est la principale critique qu'elle a reçue lors de sa prise de poste, lorsqu'elle travaillait avec Arnaud Montebourg. Trop en retrait sur les dossiers clefs, elle aurait pu perdre en crédibilité et ne pas profiter de son capital sympathie, pourtant largement reconnu. A priori, cette période d'observation lui a au contraire permis de démontrer sa tempérance et son esprit d'analyse, qualités qui ont justement parfois fait défaut à son supérieur. Désormais dans l'équipe d'Emmanuel Macron, qui remplace Arnaud Montebourg depuis le 26 Août dernier, Axelle Lemaire poursuit sa mission.

Des prises de positions fortes, des actions concrètes et beaucoup de terrain, c'est ce que la sphère numérique attend d'elle.

A ce jour, ce qui ressort, et c'est certainement dû à son parcours, est sa vision européenne. Comme elle le déclarait dans une interview dans 20Minutes, « l'Europe n'est pas les États-Unis, nous ne voulons pas d'un internet fractionné. [...] Si on parvient à s'accorder au niveau européen, ce sera un grand pas en avant ». Un axe pour son mandat ? L'échéance se profile-t-elle pour fin 2014 ou la tâche demandera-t-elle encore de nombreux allers-retours ? La protection des données personnelles est une priorité qui elle le sait se jouera aussi au niveau européen. Fin Juin 2014, les révélations sur les expérimentations de Facebook menées en 2012, ont piqué au vif la communauté et fortement interpellé la secrétaire d'État. Que sur le plan légal, aucune disposition ne permette de contrôler et empêcher ce type de manipulations représente une faille majeure dans le rôle d'un Etat qui affiche sa volonté de protéger la vie numérique de ses citoyens.

L'open data pourrait être son autre champ de bataille. Passionnée par le sujet, convaincue que « **toute donnée publique doit être ouverte par défaut**. Et s'il y a fermeture, il faut qu'elle soit expliquée, justifiée et réversible » - Axelle Lemaire affiche l'open data comme une des priorités du numérique en France. Créateur de valeur, il devrait permettre l'émulation nécessaire à une effervescence de l'économie numérique nationale. Le Big Data ne serait pas en reste, puisqu'il bénéficierait largement de cette ouverture des données.

HENRI VERDIER

DIRECTEUR D'ETALAB (Interview)

Normalien de formation, Henri Verdier est un entrepreneur dans l'âme. En 1996, il crée sa première entreprise, une web agency, qui marque le début de 20 ans de création d'entreprises dans le secteur du numérique.

En parallèle de ses expériences entrepreneuriales, Henri Verdier a mené des missions de conseil et a fait partager son expertise professionnelle. Il a également occupé le poste de directeur en charge de l'innovation chez Lagardère Active.

Engagé dans l'écosystème numérique français, Henri Verdier prend la présidence de Cap Digital en 2007. Au sein de cette association de plus de 800 start-ups, 25 grands groupes tels Thales, Alcatel ou Dassault système, et 250 laboratoires de recherches publique, dont l'INRIA par exemple, il se familiarise avec les problématiques de la donnée. Au sein de Cap Digital, les projets naissent à toute vitesse. Le réseau soutient plus de 300 projets de recherche par an, les dotant de plus de 70 millions d'euros d'aides. **C'est là qu'Henri Verdier réalise qu'une révolution de la donnée est en marche.** L'effervescence qui a alors lieu dans la Silicon Valley semble confirmer cette tendance. Avec la démocratisation du coût de traitement de la donnée, il estime alors que « **la bataille commence tout juste** » !

En 2010, il crée, avec Pierre-Louis Lions et Jean-Michel Lasry, MFG Labs, engagée dans le domaine du Big Data. Le projet rencontre le succès et est acquis par le groupe Havas quelques années après sa création. Cette expérience conforte Henri Verdier dans sa conviction de l'importance à laquelle est voué le Big Data.

En 2012, il coécrit « *L'âge de la multitude* », qui traite de la transformation numérique à l'heure où la valeur économique se déplace de plus en plus vers les particuliers, et leurs créations conscientes (contribution) ou inconscientes (traces et données). On y découvre alors sa conviction que « la puissance publique peut s'inspirer des stratégies des grandes plateformes Internet ».

Rien de plus naturel qu'en 2013 il accepte le poste de directeur d'Etalab. À la croisée entre entrepreneuriat, action publique et univers de la donnée, Etalab sert des objectifs clairs. **Henri Verdier voit l'action d'Etalab comme la « rencontre de la Déclaration des droits de l'Homme et du web 2.0 ».** Le raccourci peut paraître étonnant, mais on comprend mieux lorsqu'il l'explique : « depuis 200 ans, l'État construit patiemment une certaine transparence. C'était le sens des rapports publics de la Cour des comptes, ou de la création de l'INSEE. Aujourd'hui, l'open data, c'est la mise à disposition des informations détenues par l'État, mais dans les formats les plus bruts possibles pour favoriser le maximum de réutilisations et la co-production avec les citoyens. »

Car c'est bien le collaboratif qui prime. Les données seules, sans idée, sans travail, sans émulation, ne sont pas créatrices de valeur. Le web 2.0 permet justement de mobiliser une communauté de plus en plus étendue. C'est ainsi qu'**Etalab a choisi de soutenir les 15.000**

« la puissance publique
peut s'inspirer des
stratégies des grandes
plateformes Internet »

contributeurs bénévoles d'OpenStreetMap pour recenser la géolocalisation de l'intégralité des adresses en France.

La plateforme www.data.gouv.fr mise en place par Etalab connaît déjà un grand succès, ayant même reçu les honneurs de TechCrunch. Les données qui y sont partagées sont extrêmement variées, nombreuses et proviennent de différentes sources, comme les administrations et les établissements publics, mais aussi l'UFC Que





Exemples de quelques jeux de données : Registre Parcellaire

Choisir, la Croix Rouge ou bien OpenStreetMap. Mais « pour qu'un portail vive, il faut lui donner du sens, alimenter les interactions et animer la communauté ». La construction de la plateforme a donc demandé une longue réflexion, une concertation avec divers acteurs, l'aide de spécialistes et la présence dans de nombreux événements participatifs tels les hackatons, pour en comprendre le fonctionnement et les enjeux. **L'outil fonctionne désormais comme un réseau social.**

Cette plateforme est un bon exemple de l'agilité dont peuvent faire preuve les instances publiques. Pour Henri Verdier, **on a tort d'opposer les secteurs privé et public** : la véritable différence se situe entre les innovateurs et les conservateurs, qui se situent également dans les deux univers. L'État peut tout à fait intégrer les méthodologies agiles de la culture startup, ou les règles de l'open source, avec les avantages de l'efficacité, des budgets réduits et du développement rapide des projets.

Pour Henri Verdier, Etalab a, au moins, 3 raisons d'être. La première, la plus évidente, et déjà évoquées ci-dessus, répond à l'objectif de transparence, de concertation et de coproduction. Un outil comme OpenFisca en est un bon exemple. Son modèle dynamique permet de manipuler plus de 400 règles fiscales afin de mesurer et anticiper le montant de ses impôts mais aussi de modéliser des réformes fiscales. La deuxième raison d'être d'Etalab est l'innovation. Indéniablement, la donnée est une ressource essentielle de l'économie moderne. Elle permet de créer des applications, des services, d'améliorer des processus ou de mieux connaître notre entourage. Aux États-Unis, l'open data des données GPS a fait naître un secteur industriel entier. L'idée est de diffuser cette réussite à tous les secteurs : énergie, transports, médecine personnalisée... Les axes sont vastes. Les questions qui restent alors en suspens sont l'anonymisation des données et le risque de captation de cette valeur. Mais pour Henri Verdier, « **il ne faut pas s'arrêter par peur des Gafa** (*Goggle, Apple, Facebook, Amazon-N.D.L.R.). S'ils trouvent des modèles de valeur, ils créeront eux-mêmes les données dont ils ont besoin. Il vaut mieux dépenser notre énergie à faire naître un tissu industriel fort en France, et nous positionner au centre des innovations. »

Le troisième et dernier objectif d'Etalab concerne l'État lui-même. Si se moderniser et donner accès à davantage de données est important pour l'ensemble des citoyens et de

l'écosystème numérique, ça l'est également pour les administrations elles-mêmes. Peu familières encore de l'utilisation du Big Data, les institutions publiques ont pourtant un grand intérêt à se positionner rapidement sur le sujet.

C'est là qu'Henri Verdier agira, non plus au titre de directeur d'Etalab, mais dans le cadre de sa nouvelle fonction d'administrateur général des données. Il a en effet été nommé « CDO » de l'État en Août dernier et se trouve dès à présent face à un vaste mais passionnant chantier : faire circuler l'information entre administrations, utiliser la donnée même quand elle est considérée comme sensible et « chasse gardée » d'un ministère, jouer le rôle de datascientist de l'État en somme. Par exemple, communiquer les prix des logements en temps réel au ministère du logement permettrait de prendre des décisions appuyées par la réalité immédiate du marché. À l'image de quelques administrations américaines, **l'État français se met donc au Big Data.** On y croit, on y va.

Car pour Henri Verdier, « **le Big Data est une réelle révolution** ». Les cabinets tels McKinsey ou Gartner nous l'avait annoncée, on le vérifie aujourd'hui en assistant à la fulgurante ascension de startups qui n'auraient jamais eu les moyens de leurs ambitions si le Big Data et le cloud ne leur avaient pas permis de rivaliser avec les ultra-financés départements R & D des géants américains. C'est une démocratisation, l'ouverture du jeu à de nombreux nouveaux acteurs.

Comme le dit Lev Manovich dans son ouvrage « *Le langage des nouveaux médias* », « **une société de données n'est pas statistique** », elle ne suit pas la discipline mathématique des moyennes, mais met en place de nouvelles règles, sans hiérarchie. Le Big Data change la donne, il implique précision et personnalisation. Il entraîne de nouvelles représentations de la société. Etalab veut s'inscrire dans cette démarche de démocratie plus mature, qui vit avec son temps. Le Big Data est également une révolution philosophique. Il permet d'adresser les problèmes différemment. En biologie par exemple, le domaine d'expertise initial d'Henri Verdier, la méthode exigeait qu'on émette des hypothèses pour ensuite les vérifier. Chaque expérimentation était coûteuse, l'hypothèse devait donc être prometteuse. Aujourd'hui, les données sont nombreuses et gratuites grâce à des acteurs comme Etalab. Les temps de traitement n'ont plus rien à voir avec les méthodes tradi-



Exemples de quelques jeux de données : transparence de la vie publique

tionnelles. La recherche est entre guillemets libérée. Ce n'est même plus une question de volume de données, d'espace de stockage ou de capacités de traitement. **C'est la mentalité qu'a apporté le Big Data qui décuple les capacités**, qu'on parle de small data, smart data ou autre.

En tant que CDO de l'État, Henri Verdier s'attachera donc à exploiter cette approche. Peut-être qu'il s'agira essentiellement d'une question d'adoption de la culture Big Data avant même ses outils. L'important sera d'intégrer l'ensemble des acteurs, privés également, qui pourront participer à cette modernisation.

EXEMPLES DE QUELQUES JEUX DE DONNÉES DISPONIBLES SUR LA PLATEFORME DATA.GOUV.FR

REGISTRE PARCELLAIRE
GRAPHIQUE 2012: CONTOURS
DES ÎLOTS CULTURAUX ET LEUR
GROUPE DE CULTURES
MAJORITAIRE DES EXPLOITATIONS



TRANSPARENCE DE LA VIE
PUBLIQUE



ANNE LAUVERGEON

EX-PRÉSIDENTE D'AREVA, RESPONSABLE DE LA COMMISSION INNOVATION 2030
ET PRÉSIDENTE DU CONSEIL DE SIGFOX (Portrait)

Photo : ©Bruno Levy pour Challenges

« Arrêtons l'auto french bashing ! »

[...]

«Ténacité, vision stratégique long terme et cohésion européenne doivent primer».

Anne Lauvergeon est principalement connue pour sa carrière de femme d'affaires à la forte personnalité. Ses années chez Areva ont forgé son personnage. **Surnommée « Atomic Anne »**, l'ancienne sherpa du président Mitterrand a parfois divisé mais ses compétences ont souvent été saluées.

Elle a fait son entrée dans la sphère numérique en avril 2013, lorsque Jean-Marc Ayrault, alors premier ministre, la désigne à la tête de la commission innovation 2030. Sa mission ? **Définir, avec une équipe dédiée, les « sept ambitions stratégiques » pour la France à l'horizon 2030.** Dans son rapport rendu en octobre 2013, le Big Data figure à part entière comme secteur clef pour l'avenir économique de la France. Elle devient alors porte-drapeau d'une sphère qui milite pour un développement d'un écosystème puissant du Big Data en France. Elle crée même la surprise en acceptant le poste de présidente du conseil de la startup SIGFOX en avril 2014. Le numérique, elle y croit, et elle y jouera désormais un rôle actif.

Concernant le Big Data, la commission 2030 préconise 5 leviers d'actions : l'open data comme accélérateur d'innovation, la valorisation des données publiques (au sein des administrations), la mise à disposition de ressources technologiques au sein d'un centre à destination de startups, l'aide à l'export et le « **droit à l'expérimentation** ».

Ce dernier axe peut paraître surprenant quand on connaît les discussions au sein de la CNIL ou même de l'Europe. Le droit à l'expérimentation viserait à permettre aux acteurs du nu-

mérique de suivre le rythme effréné de leur secteur, sans être freiné par des instances et contraintes réglementaires. La commission précise évidemment qu'il faudra trouver un équilibre entre libre innovation et protection des données personnelles. C'est un idéal. À voir quelles idées pour l'atteindre émergeront.

Avec Anne Lauvergeon à sa tête, on se doit au moins de prêter attention aux recommandations livrées par la commission. La langue de bois ne fait pas particulièrement partie des habitudes de la femme d'affaires, et le pragmatisme est souvent maître mot. Lors d'une conférence en petit comité organisée par l'ESCP, Anne Lauvergeon expliquait justement certains points sur sa vision de l'entreprise, de l'État, sur le travail de la commission et le rôle du numérique dans l'économie nationale.

Dans un premier temps, et c'est intéressant au regard de la mission de la commission, elle définit le rôle des États. Un État, selon elle, doit avoir une vision stratégique. Areva par exemple, est le fruit de 50 années de politique continue, menée au niveau européen, avec un objectif clair. Cependant elle le reconnaît facilement, la formule ne s'applique qu'aux grands secteurs stratégiques. C'est alors qu'intervient **le second rôle de l'État : favoriser des écosystèmes, créer des « aquariums »** comme elle le dit. Plus précisément, cela signifie que lorsque l'État souhaite accélérer ou accompagner un secteur dit, il se doit de lui procurer les conditions nécessaires à son épanouissement. Ses outils ? La fiscalité, la réglementation, l'éducation. Sur ce dernier point elle insiste fortement, persuadée que **la France doit éduquer**

ses nouvelles générations à l'innovation, ou du moins ne pas les inciter à un raisonnement « anti prise de risque ». Sans vouloir renier notre culture et nos particularités, il nous faut évoluer avec notre entourage. Elle prend alors l'exemple des tablettes numériques, parfaitement développées dans deux pôles d'attraction qui sont pourtant deux pays très différents : les États-Unis et la Corée du Sud. Il n'existe donc pas de recette universelle, mais une capacité à l'adaptation. Chaque pays conserve ses particularités tout en favorisant les conditions d'épanouissement de l'innovation.

Enfin, cela peut sembler paradoxal, elle prône **« l'État zéro »**. L'État en retrait, qui laisse son économie se développer et les innovations émerger. L'État qui ne cherche pas à lutter contre la modernité mais plutôt à l'accompagner. En définitive, le rôle de l'État pour favoriser l'innovation consiste à reconnaître les secteurs clefs où il mènera quelques grands projets et créera un environnement favorable pour les entreprises, leur laissant la charge d'en faire bon usage.

Un point qui ne manque pas d'échapper à Anne Lauvergeon est le manque de confiance dont nous souffrons en France. « Arrêtons l'auto-french bashing » clame-t-elle ! Nous devons avoir confiance en nos capacités.

Identifier les capacités de la France, c'était justement l'objet de la commission 2030. Trois questions se sont posées à ses membres : où sommes-nous bons en France, que pouvons-nous décliner en Europe et où y a-t-il une forte demande mondiale ?



Quelques lauréats : big data pour le sport professionnel

Avant tout, la commission s'est projetée dans le monde tel qu'il pourrait être en 2030. Plus peuplé, c'est certain, avec plus de classes moyennes, le numérique encore bien plus présent dans nos vies, d'importants changements climatiques mais aussi des changements de mode de pensée et de relation à l'argent, au partage, au collaboratif. Dans quelle mesure la France aura-t-elle un rôle à jouer sur la place économique mondiale ?

Concrètement, la commission en elle-même est un point de départ, un fil conducteur. Le CGI (Commissariat Général à l'Investissement) a débloqué 300 millions d'euros, gérés par la Bpi. **Les sept ambitions de la commission 2030 ont**

donné naissance à sept concours mondiaux de l'innovation. Mondiaux ? C'était important pour la commission de laisser les candidatures ouvertes à tous, à travers le monde, avec pour seule contrepartie de s'engager à investir en France. Les premiers lauréats ont déjà été désignés.

Dans l'ensemble, la commission a reçu plus de 1200 dossiers de candidature depuis son lancement en décembre 2013 jusqu'à début juillet 2014. L'exercice est même voué à être répété tous les deux ans, selon une déclaration de François Hollande.

Anne Lauvergeon nous rappelle alors une particularité de la France, qui influe fortement sur ses démarches d'innovation et par conséquent sur des initiatives telles le concours mondial de l'innovation. Au monde, c'est le seul pays à appliquer le principe de précaution, après l'avoir inscrit dans sa constitution (voir la loi Barnier). D'autres pays l'ont reconnu, comme le Brésil ou l'Allemagne, mais la France s'attache fortement à son respect. Souvent désigné comme frein à l'innovation, **le principe de précaution requiert dans son application la recherche d'un équilibre avec le principe d'innovation.** Les deux ne doivent pas s'opposer mais se tempérer. On rejoint alors le concept du « droit à l'expérimentation » stipulé dans le rapport de la commission, qui doit s'équilibrer avec la protection de la vie privée. Comme le dit Anne Lauvergeon, il faut faire le « ying et le yang entre les deux notions ». Tout serait donc une question d'équilibre.

Selon la femme d'affaires, l'innovation ne doit pas se voir comme une finalité ni se limiter à un seul usage. **Il faut « la partager, la faire vivre,**

pour que l'innovation puisse être exploitée ailleurs que dans un domaine unique ». Dans les faits, cela demande la création de « réseaux sociaux d'acheteurs des innovations », où les idées se transmettent, se valorisent, s'exportent et s'enrichissent.

Une belle vision de l'innovation en somme, qui ne serait non plus un outil pour son générateur, mais bien un moteur pour l'économie tout entière.

Elle conclut sur ces quelques mots, qui traduisent bien la personnalité de cette femme de caractère : « ténacité, vision stratégique long terme et cohésion européenne doivent primer ».



Quelques lauréats : modélisation prédictive pour les villes

QUELQUES LAURÉATS DE LA CATÉGORIE BIG DATA DU PREMIER CONCOURS MONDIAL DE L'INNOVATION LANCÉ PAR LA COMMISSION.

PROJET : BIG DATA POUR LE SPORT PROFESSIONNEL

Mac-Lloyd propose des technologies de rupture dans le domaine du sport de haut de niveau : capteurs de mouvements, mesure par analyse vidéo, et traitement intelligent de données massives par machine learning.

PROJET : SNIPS

Snips est une startup spécialisée en modélisation prédictive pour les villes. En s'appuyant sur des données de géolocalisation ainsi que des données de contexte, nous pouvons mesurer, comprendre et anticiper les comportements dans les villes, contribuant ainsi à personnaliser et améliorer le quotidien des citoyens.



GILLES BABINET

FONDATEUR DE CAPTAINDASH
ET DIGITAL CHAMPION
AUPRÈS DE LA COMMISSION
EUROPÉENNE
(Interview)



Votre parcours :

Au-delà de la fiche Wikipedia qui parle « d'un passage difficile au collège et au lycée »... Quel parcours avez-vous suivi ? Comment avez-vous commencé à fonder vos premières entreprises et pourquoi vous êtes-vous lancé dans cette aventure ?

Tout cela est un peu singulier mais en deux mots ce qu'il faut en retenir c'est que j'étais effectivement un cancre à l'école et je ne m'y adaptais pas. J'avais une certaine habileté manuelle et j'ai commencé à travailler dans le bâtiment. C'est là que j'ai eu l'idée de ma première entreprise, que j'ai donc développée dans le monde de l'alpinisme et du bâtiment. Puis les choses se sont enchaînées et j'en suis venu à travailler dans l'univers numérique, qui m'a toujours passionné. Ce que j'en retiens c'est que **notre système éducatif n'est pas adapté pour certaines personnes comme moi**. Je ne cesse de penser que nombreux sont les enfants qui voient leurs vies gâchées ou en tout cas sérieusement perturbées parce qu'ils n'arrivent pas à se familiariser avec la rigidité de notre système français.

Vous et le numérique :

D'où vient votre intérêt pour le secteur du numérique ? À quel moment vous êtes-vous dit que le Big Data représentait un secteur d'avenir, de nouvelles opportunités, une révolution ? Et comment est né CaptainDash ?

Je ne suis pas à proprement parler un programmeur même si je me suis beaucoup intéressé à l'électronique lorsque j'étais jeune. J'ai toujours voulu introduire l'informatique, les réseaux dans mes entreprises. **Je pense avoir fait partie de la première vague des fondateurs d'agence web**, dès 1994 et donc tout ceci représente finalement un certain continuum. Il y a sept ou huit ans, j'ai rencontré Bruno Walther, qui était alors CEO de l'agence internet Ogilvy et nous avons eu des discussions passionnantes sur le constat que nous faisons que le monde de la donnée allait représenter la prochaine révolution ; de fil en aiguille nous avons créé une société.

L'État et le numérique :

Président du CNN, Digital Champion auprès de la Commission Européenne... Que pensez-vous du rôle de l'État dans la construction d'un écosystème du numérique ? Open data, financement, éducation, avantages fiscaux, incubateurs, projets publics : dans quelle mesure doit s'engager l'État ?

Le rôle de l'État est assez particulier, car il doit avoir une pratique à la fois très libérale, pro-entrepreneur en favorisant la prise de



risque, pro-business en luttant contre les professions réglementées et tout à la fois s'assurer que la qualité du système éducatif est à la hauteur d'une société de la connaissance, que les infrastructures numériques sont financées et disponibles, que les libertés ne sont pas altérées, etc. À mon avis, rares sont les États qui parviennent à mettre en œuvre cette nouvelle alchimie. **La France progresse mais elle a un long chemin à faire.**

Focus réglementation :

On vous connaît pour vos propos assez virulents envers la CNIL. Quel est selon vous l'attitude à adopter sur la protection des données personnelles, la question du cloud souverain, l'harmonisation européenne ? Comment contrer les dérives, rassurer les utilisateurs, et défendre la place de la France et de l'Europe sur le plan international ?

Tout cela est si rupturiste qu'il me faudrait un livre — que je suis en train d'écrire — pour d'écrire convenablement ce qu'il faut faire. Je ne peux ici que recommander que l'on cesse

d'être en permanence dans le principe de précaution, à avoir peur du futur et de notre ombre. Les attitudes réactionnaires ne réussissent jamais aux nations. Idéalement, il conviendrait de favoriser l'éclosion d'une nouvelle forme de Droit, ce qui ne manquera pas d'arriver. À chaque révolution industrielle son droit. La première a vu l'apparition du Code Civil, la seconde les droits sociaux (code du Travail), nous assisterons un jour à l'émergence du code de l'individu et de la Donnée.

« Je ne peux ici que recommander que l'on cesse d'être en permanence dans le principe de précaution ».

PASHU DEWAILLY CHRISTENSEN

MARKETING AND
PARTNERSHIP MANAGER
CHEZ THE HIVE, PALO ALTO
(Interview)

Pashu a un parcours atypique. Après un master en économie du développement suivi à Dauphine, elle part en stage à l'ambassade de France en Équateur. Elle revient en France pour effectuer quelques missions en conseil, avant de prendre en charge la communication d'Aide Médicale en Indonésie. Elle intègre Orange à San Francisco alors que le sujet du Big Data commence à émerger. Pashu construit alors son intérêt pour la thématique et découvre l'étendue des domaines d'application du Big Data, allant de la finance à la médecine, servant le commerce et l'industrie, sans même oublier le développement durable. Comme elle le dit, c'est là qu'elle a pris conscience que « **dans l'économie, dans nos vies quotidiennes, grâce aux objets connectés: le Big Data va tout chambouler** »!

C'est donc tout naturellement qu'elle accepte de rejoindre, en Octobre 2012, alors qu'il vient d'être fondé, l'incubateur The Hive. Travailler dans une structure qui se focalise sur la création de startups dans le domaine du Big Data a tout de suite attiré Pashu. Autre avantage notable, elle précise qu'elle a la chance de travailler avec « des gens brillants, qui ont aidé à construire l'infrastructure — chez Yahoo notamment — et qui reconnaissent les tendances de demain ».

En effet, The Hive, basé à Palo Alto, le cœur névralgique de l'écosystème numérique, a été fondé par T.M Ravi et Sumant Mandal, deux icônes de la Silicon Valley.

Avec une équipe d'environ 10 personnes, The Hive, que Pashu définit comme un « studio », finance, incube et lance des data-driven startups qui proposent des applications B to B. Les secteurs concernés sont principalement le marketing, la sécurité et la santé. La philosophie de The Hive: le « low volume, high touch ». C'est-à-dire que l'incubateur investit dans 5 à 10 startups par an, triées sur le volet, et sélectionnées au « seed-stage » de leur développement, à la naissance de l'idée. À la différence des investisseurs (capital risqueurs) ou des incubateurs traditionnels, The Hive est un studio qui inves-



« The Hive est là pour permettre aux startups de se concentrer sur l'essentiel ».

tit entre \$1,5 million à \$3 millions dans chacune des startups, avec une équipe (CTO, Data scientist, Architect, Marketer) qui travaille quotidiennement avec les startups. L'objectif est donc d'accompagner de A à Z des projets auxquels l'incubateur croit, du concept au produit final. Au-delà de l'investissement financier, l'appui humain est au cœur du fonctionnement de The Hive. À voir l'organigramme de The Hive, on comprend que **le studio doit souvent être en effervescence, avec cette « équipe de rockstars ! »** comme le dit Pashu.

En quelques années, l'incubateur connaît déjà un succès certain, qui l'a poussé à se développer au-delà de ses frontières, en Inde dans un premiers temps — le bureau de Bangalore a été inauguré il y a quelques mois. Comment fonctionne la structure? Les investisseurs sont variés: entreprises privées, individuels ou institutionnels ont apporté 8 millions de dollars à la création. Aujourd'hui, **The Hive bénéficie d'une enveloppe de plus de 30 millions.** Des conseillers extérieurs sont égale-

ment mis à disposition, partageant la connaissance d'organismes tels Google, LinkedIn, EMC ou l'université de Stanford.

En 2 années, on peut déjà mesurer les retours et voir quelques beaux projets émerger. Pashu insiste sur le fait que The Hive est là pour permettre aux startups de se concentrer sur l'essentiel: des bureaux sont mis à disposition, des équipes sont présentes en continu, l'infrastructure est fournie et les financements permettent d'avancer rapidement. Une startup ne demeure généralement pas plus d'un an dans le studio. Rien d'étonnant alors à ce que de petites pépites en soient déjà sorties, à l'image de DeepForestMedia ou E8 Security.

Deep Forest Media fournit un système intégré pour les annonceurs mobiles pour optimiser la valeur de leurs dépenses de publicité sur mobile. E8 Security détecte les attaques avancées et les activités d'initiés malveillants qui ont contourné les mesures préventives de sécurité d'entreprise. Sa mission est d'aider les entreprises à se défendre contre la croissance continue et la sophistication des cybercriminels. La technologie de pointe de E8 Security analyse automatiquement les comportements des différents acteurs de l'entreprise et peut identifier les activités suspectes sans nécessité de règles ou de signatures.

Si The Hive a pour mission première de lancer des startups innovantes, ce n'est pas sa seule activité. L'appui et le réseau nécessaire à son activité exigent du studio un tissu relationnel étendu. Pashu est donc responsable depuis décembre 2012 de **The Hive Think-Tank**. « **C'est un peu mon bébé** » explique-t-elle affectueusement. On comprend vite pourquoi: le think-tank vise à réunir à raison d'une fois par semaine la sphère data de la région, autour de thématiques pointues et avec des speakers de haute renommée. Avec plus de 5 000 membres aujourd'hui, 200 participants à chaque édition et des financements des plus grandes entreprises telles IBM ou Microsoft, The Hive Think-Tank est une structure à part entière. Privacy, environnement, évolutions technologiques: Pashu s'attache à assurer une variété des thématiques abordées. Les conférences et networking du think-tank sont devenus des rendez-vous incontournables dans la région. C'est aussi là que le studio puise ses axes de développement, affine son appréhension des opportunités et besoins du marché et étend son réseau. Connaître les attentes du secteur pour mieux savoir quels projets soutenir est essentiel pour The Hive. C'est dans cette expertise que le studio puise sa valeur.

Pour Pashu, The Hive est comme une petite famille, accueillante, pleine de vie et bien ancrée dans la ville de Palo Alto. Installée à Oakland, qu'elle décrit comme le nouveau San Francisco, plus accessible et offrant davantage d'opportunités, Pashu a la tête pleine de projets et d'envies.





DOUG CUTTING

CO-FONDATEUR D'HADOOP
(Portrait)

Doug Cutting. Un grand nom dans la sphère Big Data. Incontournable même. Un homme qui inspire respect et admiration pour avoir été l'un des fondateurs de la fabuleuse épopée du Big Data.

Un homme simple pourtant. Incroyablement accessible, Doug aime se fondre dans l'effervescence de la communauté Big Data. Passionné toujours, il s'intéresse, observe et écoute. Sans aucune prétention, il explique volontiers comment il s'est retrouvé à l'origine d'Hadoop.

Diplômé de l'université de Stanford, Doug débute sa carrière dans quelques prestigieuses entreprises, telles Excite, Apple ou Xerox. Il travaille ensuite sur les projets Lucene, Avro ou Nutch, en open source. C'est aussi à cette période que Google gagne sa place de leader dans les moteurs de recherche. Il est beaucoup plus pertinent et rapide que tous ses concurrents. Il utilise une technologie que personne d'autre ne connaît. La communauté scientifique et toute la Silicon Valley sont intriguées par la recette du succès du géant internet. En 2003, Google publie un premier papier où est expliqué le fonctionnement du Google File System. Doug et ses collègues, dont Michael Cafarella, découvrent les premières bribes de la recette de Google et prennent la mesure de la révolution qui s'annonce. Un an plus tard, MapReduce, qui se cache derrière le fonctionnement du moteur de recherche, est enfin révélé au grand public. Doug comprend alors qu'il se trouve face à la technologie qui lui permettra d'élever Lucene au stade supérieur. Le puissant algorithme développé pour MapReduce permet de traiter des volumes conséquemment plus importants de données, les analysant en parallèle de façon simple et rapide. Lucene peut passer à l'échelle

du web, alors en pleine expansion et faire face à l'immensité des informations disponibles sur la toile. Doug Cutting prend alors la tête du projet qui se nommera par la suite Hadoop. Toujours dans la dynamique open-source et open-innovation qui est depuis perpétuée dans la Silicon Valley, Yahoo lègue Hadoop à la fondation Apache en 2006. Les meilleurs ingénieurs de Google, Yahoo et des autres géants du centre névralgique de l'innovation numérique mondiale se retrouvent régulièrement pour faire avancer ces projets colossaux, conscients de la portée de leurs travaux de recherche. C'est peut-être aussi de là que vient cette certaine humilité de Doug. Il semble convaincu qu'un projet d'une telle ampleur, porteur de tant d'innovation, est le fruit d'heureuses collaborations, et rarement d'un seul homme. En 2009, Doug rejoint la société Cloudera, où il travaille toujours actuellement en tant que Chief Software Architect. L'année suivante, Doug est nommé à la tête de la fondation Apache. Emblématique du Big Data aujourd'hui, l'éléphant d'Hadoop est lui aussi issu de l'imagination de Doug, qui adapta le jouet préféré de son fils pour en faire le logo de la technologie qui révolutionnera la donnée.

Lors du dernier congrès Big Data Paris, Doug nous fit l'honneur de venir exposer ses dernières réflexions sur l'avenir de la donnée. S'il admet que nul ne peut prédire le futur, Doug Cutting constate néanmoins que le rapprochement de faits conduit souvent à d'assez bonnes anticipations. Un premier fait indiscutable réside dans l'augmentation constante des performances de la technologie, dont le coût décroît de façon tout aussi régulière, conformément à la loi de Moore. **Nous serons en mesure, à**

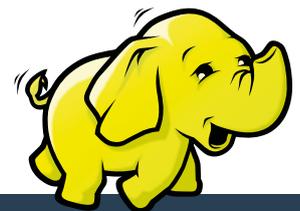
l'avenir, de stocker des volumes de données encore plus importants qu'aujourd'hui et ce n'est pas anodin car ces données sont précieuses. Les secteurs qui réussissent sont ceux qui utilisent les données de la manière la plus intelligente et ce principe semble aujourd'hui s'appliquer à tous les secteurs d'activité.

Au cours de la dernière décennie a émergé un nouveau type de logiciels, en Open Source, ce qui a fait de Linux le système d'exploitation le plus répandu dans le monde. Il ne fait pas de doute que l'Open Source devient la norme et que les développements majeurs sont appelés à utiliser ce type de plate-forme.

On remarque d'ailleurs que sur les trois principales sociétés qui distribuent Hadoop aujourd'hui, que sont Cloudera, Hortonworks et MapR, deux sont à 100 % open source [N.D.L.R.].

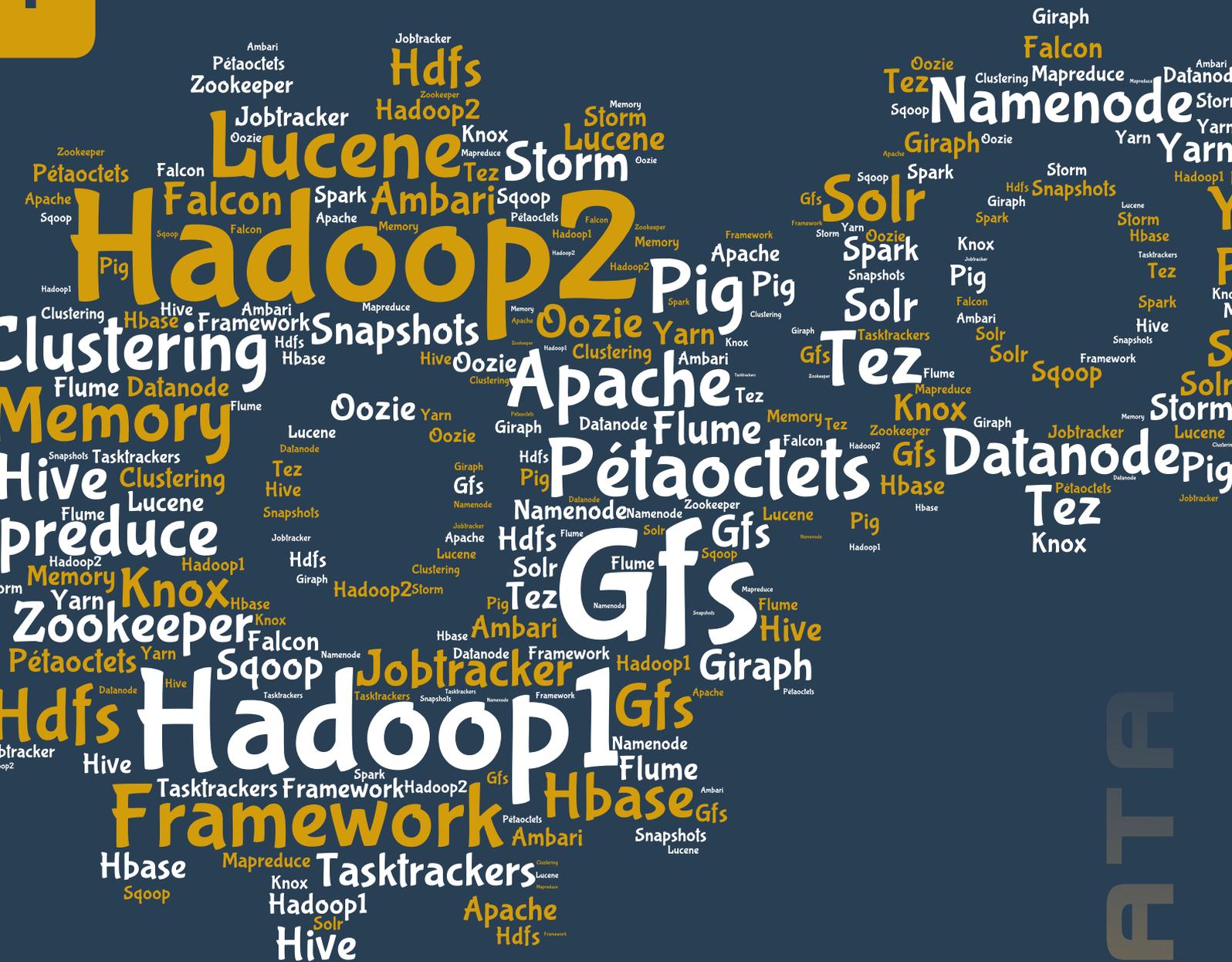
Doug explique comment il s'est modestement lancé, il y a quelques années, dans le projet Hadoop, en utilisant cinq machines simultanément, puis des méthodes de répartition. Les promoteurs du projet ne savaient pas comment changer d'échelle pour faire fonctionner des milliers d'ordinateurs en même temps. Yahoo, intéressé par le projet, y a investi de façon importante. La jeune entreprise a alors pu continuer d'améliorer l'outil en changeant d'échelle. Hadoop permet ainsi de stocker et traiter des volumes considérables de données réservés jusqu'alors aux traitements en « batch ». Le module PIG, qui est un langage de flux de données, a aussi été ajouté à la plateforme afin de faciliter l'utilisation du programme et la définition de spécifications pour des séries de données.

Hadoop a finalement été adopté par un grand nombre d'opérateurs qui en ont fait la plateforme de référence en matière de Big Data. Cette évolution ne paraît pas devoir connaître de limite et rien ne semble s'opposer à la poursuite de cette logique de partage. Même les transactions sont désormais possibles sur Hadoop. Ce conglomérat de projets qu'est devenu Hadoop a fait de celui-ci un « hub de données d'entreprise » (Enterprise Data Hub) permettant à toutes ces données d'être traitées en un seul lieu, avec l'ensemble des applications souhaitées. **Nous ne sommes encore qu'aux prémices de l'adoption du Big Data** par l'industrie et il ne fait pas de doute que chacun voudra copier les « gagnants ».



« Les secteurs qui réussissent sont ceux qui utilisent les données de la manière la plus intelligente ».

Techno



Techno

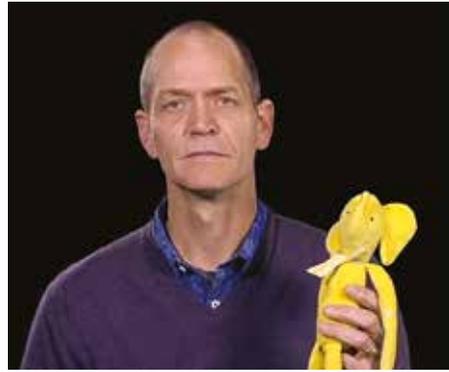
BIG DATA

HADOOP ENFIN UNE SOLUTION D'ENTREPRISE ?

PAR JOSÉ DIZ

Par Big Data on entend l'application de traitements analytiques à de grands volumes de données, structurées ou non. Depuis longtemps, des technologies de streaming analysent en "temps réel" des informations sur des bus de données. Ces solutions souvent onéreuses (reposant sur des environnements critiques) analysent par exemple en continu des flux financiers pour détecter des fraudes comme du blanchiment d'argent.

Les systèmes de gestion de données traditionnels peinent ou se montrent inefficaces pour traiter les énormes volumes d'information provenant du web (plus de 2 milliards d'internautes), des smartphones (7,3 milliards en 2014), des étiquettes RFID, tweets (7 To par jour), de Facebook (500 To par jour), des caméras, des capteurs et autres senseurs. Et plus de 75 % de ces informations ne sont pas structurées ou seulement semi-structurées.



Doug Cutting, père d'Hadoop (Source: Cloudera)

HADOOP 1 LA NOUVELLE STAR DE L'ANALYTIQUE EST NÉE

Depuis quatre à cinq ans, le framework open source Hadoop est devenu synonyme de Big Data. Même si au départ, il n'effectuait que des traitements par lots (en batch) et donc en différé. Les raisons de ce succès ? Première raison : la possibilité d'utiliser à moindre coût des algorithmes open source conçus pour fonctionner sur des serveurs x86 standard, et capables de répondre très rapidement à des questions en analysant jusqu'à plusieurs pétaoctets d'informations. Et surtout une simplicité (relative) d'utilisation et de déploiement inédite. Enfin, Hadoop combine l'utilisation de technologies de pointe très appréciées ou très innovantes : parallélisation des traitements, clustering, traitement et données sur un même nœud dans une infrastructure distribuée, fiabilité par réplication des informations, etc. Et avec la version 2 : In-Memory, temps réel, haute disponibilité améliorée... Des technologies enfin à portée budgétaire des PME. Enfin, outre les problèmes de volumes difficiles à analyser avec les technologies classiques, Hadoop permet d'envisager la collecte et le traitement de données en volume comme l'Internet des objets et les multiples capteurs et senseurs, le RFID, les tweets, etc.

10 ans : des origines de MapReduce à Hadoop 2.x

Pour appréhender ces technologies, un petit historique s'impose. Fin 2004, Jeffrey Dean et Sanjay Ghemawat employés chez Google pu-

blent l'article « MapReduce : Simplified Data Processing on Large Clusters ».

En bref, ils expliquent concrètement comment cet algorithme distribué utilise des serveurs x86 standard en cluster pour paralléliser des traitements dans le but d'analyser en un temps record un très grand volume de données de tout type. À l'œuvre chez un moteur de recherche déjà plus que prometteur à l'époque. Bien entendu, MapReduce repose sur le système de gestion de fichiers en clusters GFS (Google File System).

À cette époque, Doug Cutting (ex-employé du moteur de recherche Excite, d'Apple et de Xerox Park) mène le projet d'indexation Apache Lucene (bibliothèque open source Java pour indexer et rechercher du texte), et son utilisation dans son moteur de recherche Nutch. Il voit dans cet article la possibilité de résoudre les problèmes de masses de données de Nutch. Jusqu'en 2006, il intègre MapReduce et un système de fichier distribué DFS à son moteur. Il entre alors chez Yahoo et crée un framework regroupant MapReduce et DFS qu'il baptise du nom que son fils donne à son éléphant en peluche jaune : Hadoop. Le logo sera créé par un ami graphiste. En juin 2009, Yahoo rend le code source d'Hadoop public via un projet de la fondation Apache.

Au-delà de MapReduce et HDFS (Hadoop distributed file system), de nombreux modules viendront compléter Hadoop, dont les principaux sont devenus des projets Apache à part entière (top-level), parmi lesquels : HBase (mai 2010), Hive et Pig (septembre 2010), ou encore Zookeeper.

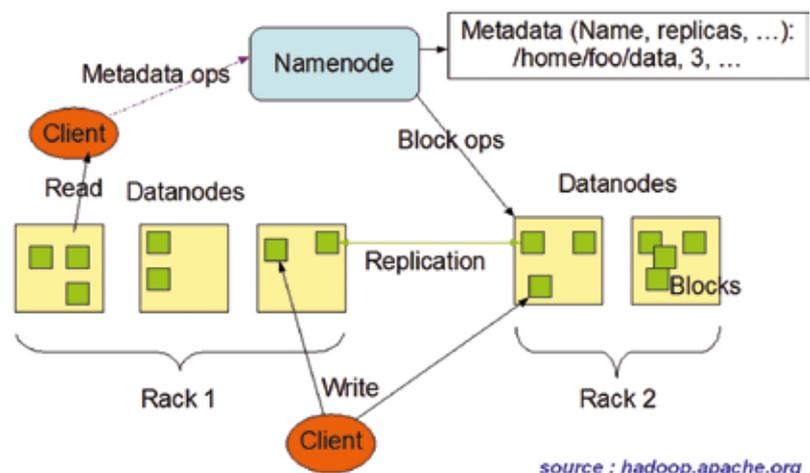
La version 1.0.0 du framework est disponible fin décembre 2011.

En octobre 2013, Hadoop connaît une évolution majeure de son infrastructure avec la version 2.2.0 qui ouvre à ce framework les portes du temps réel, de l'extensibilité fonctionnelle, du In-Memory, de l'environnement Windows, etc.

Le centre névralgique HDFS et MapReduce

Le cœur d'Hadoop 1 est composé de deux modules majeurs.

Inspiré de Google File System et développé en Java, **HDFS (Hadoop Distributed File System)** s'exécute au-dessus du système de gestion de fichiers de chaque nœud d'un cluster. Un nœud où sont stockées les données (et exécutés les traitements qui les concernent) s'appelle un **datanode**. HDFS stocke l'information sous forme de blocs



source : hadoop.apache.org

Le Namenode : mémoire et cerveau ordonnateur d'HDFS



Hadoop (64 Mo par défaut, 128 recommandés). En tant que fichier du serveur, un bloc Hadoop occupe physiquement plusieurs blocs. Si un fichier ou une partie de fichier est plus petit qu'un bloc Hadoop, sa taille s'ajuste. Chaque fichier est découpé en blocs Hadoop répartis sur des datanodes différents, et une ou plusieurs copies de chaque bloc sont enregistrées sur des Datanodes différents. Une capacité qui assure une **tolérance aux pannes** appréciable. Un serveur appelé **Namenode** gère toutes les métadonnées des blocs Hadoop et sait donc les retrouver. Certains éditeurs traditionnels ou pionniers Big Data remplacent HDFS par leur propre système de gestion de fichiers distribué.

MapReduce 1 assume plusieurs rôles. Il gère et alloue aux applications les ressources du cluster, et exécute les traitements appliqués aux données. Lorsqu'une requête est adressée à Hadoop, elle est prise en main par un **JobTracker** qui coordonne les traitements entre Map et Reduce et assure le suivi des tâches. Il distribue les processus parallélisés aux **Task Trackers** sur les nœuds du cluster Hadoop en optimisant les échanges. La fonction **Map** divise la demande initiale en séquences (clé, valeur) auxquelles vont être appliqués le ou les traitements en parallèle (vitesse optimale). Chaque tâche Map renvoie un résultat (clé-valeur). Puis un traitement (Shuffle & sort) remanie les résultats pour regrouper ceux qui ont la même clé. La fonction **Reduce** prend ces résultats et les

"réduit" en effectuant une opération sur les valeurs associées à chaque clé (montant total, nombre d'occurrences, etc.).

Quelques briques de l'édifice Hadoop

HBase est un système distribué de gestion de bases de données NoSQL en colonnes. Projet Apache, il est né suite aux publications de Google sur Big Table en 2006. Installé sur HDFS. Il fonctionne en mode cluster, est horizontalement évolutif et tolérant aux pannes. Le mode colonne réduit les accès à des index et le nombre d'accès disque. Donc performant pour l'analytique.

Pig a été conçu pour simplifier l'utilisation des API MapReduce exigeant de bonnes connaissances Java. Yahoo a créé la plateforme de traitement de données Pig et son langage Pig Latin (projet Apache). Efficace et simple, le langage Pig Latin (proche du scripting) permet d'écrire des applications exécutées sur Hadoop (runtime Pig) sans passer par MapReduce. Le développeur charge les données, y compris d'une source externe vers le cluster Hadoop, et les manipule.

Hive - Conçu chez Facebook, ce projet Apache ouvre Hadoop au SQL afin de simplifier la manipulation des données avec une syntaxe très proche du SQL (HiveQL). Hive offre des outils d'entrepôt de données (datawarehouse) : extraction, transformation et chargement d'informations provenant de diverses sources. Il stocke

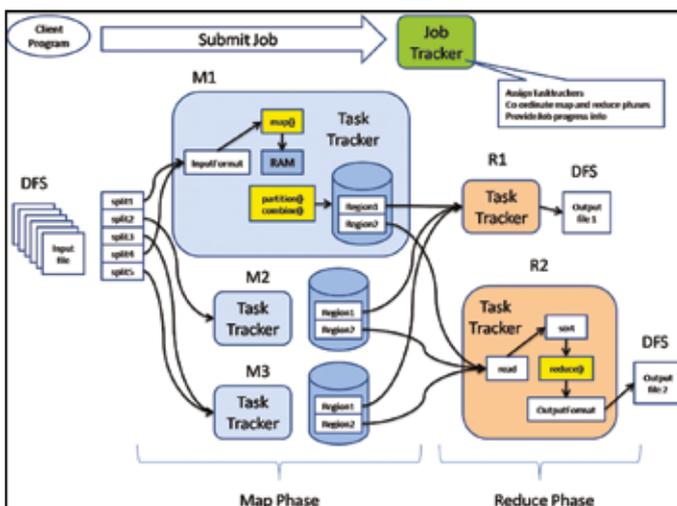
le tout dans Hadoop « à la manière d'une base de données classique », permettant sélections, jointures, agrégats... Hive compile les requêtes en jobs MapReduce qui les exécute sur Hadoop.

Zookeeper a été conçu sur la base du logiciel Chubby de Google. Il propose une gestion centralisée de configurations pour grands systèmes distribués aussi bien des machines physiques que des services applicatifs Hadoop. Il permet de suivre et de maintenir l'état des services distribués (comme MapReduce ou Hbase) pour les rendre consistants.

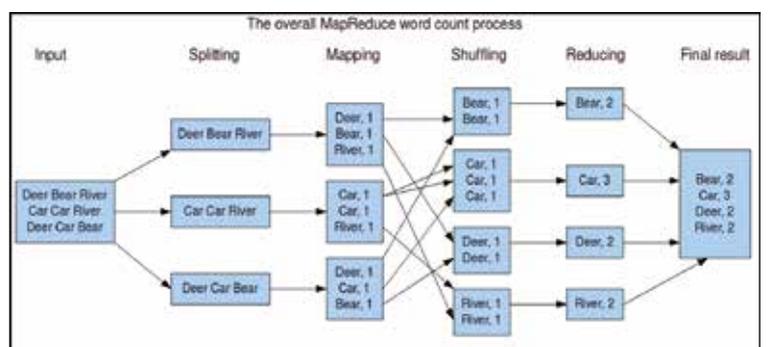
Des manques importants fragilisant Hadoop 1

- Malgré l'emballage général, Hadoop 1 souffrait de multiples carences**, qui ont fait hésiter grand nombre d'entreprises. Parmi lesquelles :
- le serveur Namenode n'était pas prévu pour être répliqué, le transformant en talon d'Achille de la plateforme ;
 - pas de partage entre les "jobs" (MapReduce et autres) ;
 - évolutivité limitée: un seul JobTracker par nœud et un maximum de 4000 nœuds par cluster ;
 - en cas de panne, toutes les tâches étaient supprimées (en cours ou en attente) ;
 - le temps réel et le streaming n'étaient pas possibles ;
 - les mises à jour d'informations s'avéraient complexes, risquées, voire impossibles ;
 - la supervision était encore primaire, et la gouvernance quasi inexistante.

Toutefois, les diverses communautés mobilisées sur et autour d'Hadoop se montrent très actives et l'innovation s'accélère. Comme le prouve la version de référence Hadoop 2.2.0 publiée en octobre 2013 par la fondation Apache.



MapReduce, JobTracker et HDFS dans le bateau Hadoop



LA REVOLUTION HADOOP 2

Plus qu'une autre version corrigeant des manques, Hadoop 2 modifie son architecture pour plus de sécurité et d'ouverture : haute disponibilité, temps réel, streaming ou encore au In-Memory.

Naissance de Yarn et réduction de MapReduce

Évolution majeure: les deux fonctions de MapReduce sont découplées. Désormais, la gestion des ressources du cluster est assurée par la nouvelle couche **Yarn (Yet Another Resource Negotiator)**. MapReduce se contente de gérer les travaux de types batch. Une amélioration qui n'oblige plus à passer par MapReduce pour accès aux données en mode parallèle.

Yarn propose une nouvelle architecture. Depuis un nœud maître, le **Resource Manager (RM)** orchestre l'ensemble des ressources du cluster et arbitre entre les applications nécessitant des ressources. Sur chaque nœud esclave, un **Node Manager** communique avec le RM.

À chaque demande, Yarn génère un container (mémoire+CPU) sur un nœud esclave, géré par un **Application Master**. En cas de besoin, d'autres containers peuvent être créés pour cette application, mais l'Application Master gère seul tous les containers de l'application. Et le Resource Manager supporte Kerberos.

Avec Yarn, il devient possible d'ajouter des modules applicatifs autres que MapReduce pour accéder aux données. La plateforme devient donc extensible.

Et surtout, **plusieurs moteurs applicatifs de différente nature** (batch, interactif, temps réel, streaming...) peuvent s'exécuter simultanément sur un même jeu de données.



Une gestion des fichiers amplement améliorée

HDFS 2 bénéficie aussi d'améliorations, même si certaines existaient déjà dans des distributions Hadoop :

- possibilité de **plusieurs Namenodes** (donc namespaces) non liés sur le stockage, avec journalisation globale partagée pour de la **haute disponibilité**,
- création de **snapshots** en lecture seule ou en lecture/écriture,
- HDFS peut devenir un disque partagé "banalisé" sur le réseau via **NFS v4**,
- compatibilité HDFS 1,
- fichiers de steaming géré par HDFS, etc. Enfin, **Hadoop2** supporte désormais **Windows Server et Windows Azure**, donc ouverture à plus de compétences.

Petit panorama de quelques modules Hadoop 2

Le projet **Apache Tez** offre un framework pour concevoir sous Yarn des applications batch ou interactives.

Pig conserve sa fonction de scripting de flux de données, mais n'a plus besoin de **MapReduce** (en passant par Tez). De même Hive -via Tez- reste la référence pour le SQL sous Hadoop. Tez apporte éventuellement

une dimension interactive.

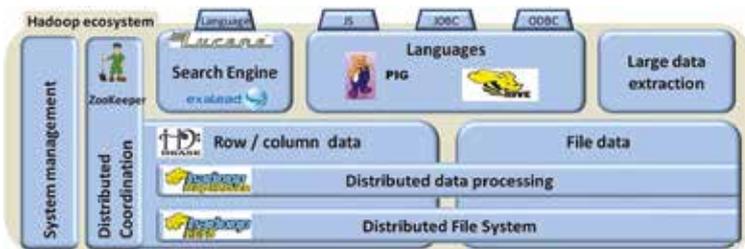
Storm est un système informatique temps réel distribué, résolvant en grande partie les problèmes de latence sur Hadoop. Avec tout langage, le développeur utilise ce framework rapide, évolutif, tolérant aux pannes (Storm relance les tâches arrêtées), simple à utiliser et déployable en moins d'un jour. Il s'installe sur Yarn, "temps réel" oblige !

Avec **Apache Spark**, Hadoop peut bénéficier du In-Memory. Directement posé sur Yarn, ce service de type MapReduce accélérerait jusqu'à 100 fois les traitements. Ses fonctions permettent de charger de jeux de données en mémoire dans le cluster et d'y effectuer des traitements répétitifs, par exemple. Il est également compatible avec l'API de stockage d'Hadoop.

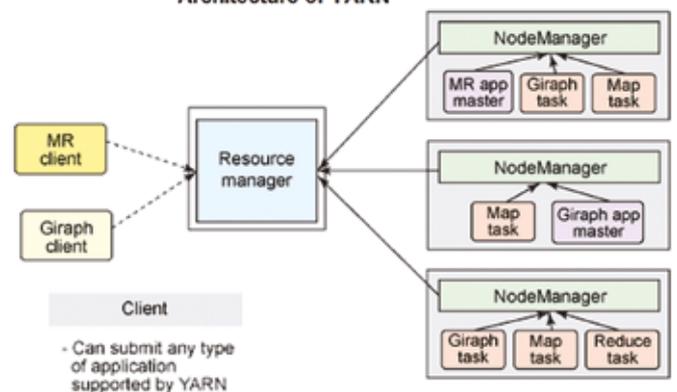
Solr amène la fonction de recherche plain-text et plus évoluée aux données Hadoop. Cette plateforme repose sur le moteur Lucene (avec lequel elle a fusionné en 2011). CNet Networks l'a offerte à la fondation Apache.

Fiable, tolérante aux pannes et évolutive, Solr propose de nombreuses innovations : indexation quasi-temps réel, API ouvertes XML, JSON et HTTP, statistiques serveur (avec exposition JMX), index autorépliqué, reprise sur incident... Directement interfacé avec Yarn.

"Hadoop 2 modifie son architecture pour plus de **sécurité** et d'**ouverture** : haute disponibilité, temps réel, streaming ou encore au In-Memory."



Architecture of YARN



Apache **Giraph** est un projet pour réaliser des opérations sur les graphes (sommets et liens), comme sur les réseaux sociaux (contacts, niveaux de proximité, etc.) de type Facebook, Twitter, LinkedIn ou Viadeo. Giraph est parti d'un article publié par Google sur son système de Graphes appelé Pregel.

Sans oublier : sécurité, administration et gouvernance

Sélection de quelques modules apportant ces indispensables dimensions pour une solution d'entreprise.

La passerelle de **sécurité Apache Knox** (API Rest) fournit un point d'accès unique pour toutes interactions Rest/Hadoop. Il propose la gestion sur toute la plateforme Hadoop de l'authentification (LDAP et AD), de la fédération SSO, des autorisations aux services, de l'audit, etc. Il gère les autorisations en workflow et vérifie les jetons d'authentification (règles). Complète les mesures de sécurité existantes dans Hive, HDFS, etc.

Ambari joue un rôle majeur dans l'exploitation et la supervision Hadoop. Une palette déjà impressionnante : gestion de la configuration fine des services (et déploiement en cluster), supervision des machines du cluster (possibilité de décommissionnement, etc.), tableau de bord global (état de services et exécutions), gestion des habilitations (compatible LDAP), gestions des jobs avec vues agrégées et timelines (+ tracking utilisateur), audit, etc.

Apache **Oozie** est un outil de workflow pour planifier et coordonner les jobs sous Hadoop. Cette application Web Java peut combiner séquentiellement plusieurs jobs sur une seule unité logique. Intégré à la stack Hadoop, il sup-

porte entre autres MapReduce, Pig ou encore Sqoop, mais aussi des applications Java ou des scripts shell.

Falcon (incubation Apache) vise à simplifier la gestion des données et de leur cycle de vie, à découvrir les schémas de données et les traitements associés, offrir la réutilisation de ces règles réalisées sans codage... Bref, une gouvernance des données sur cluster Hadoop.

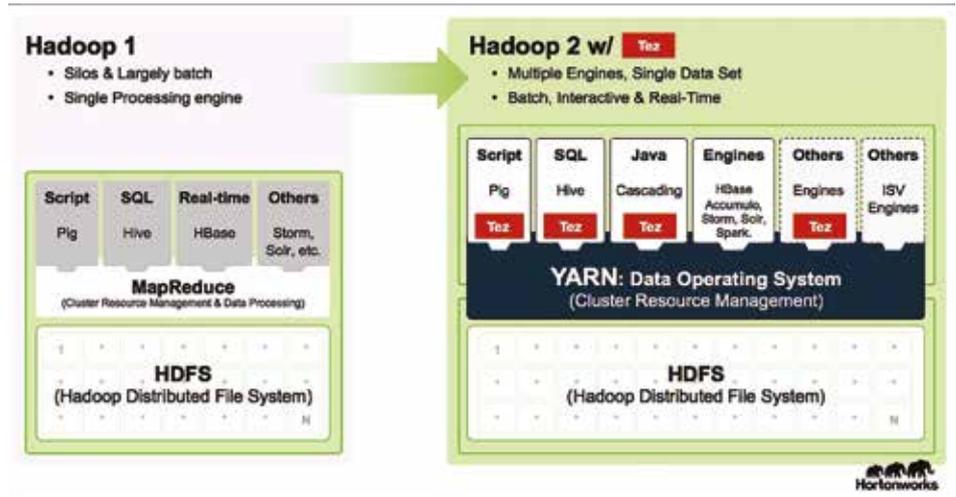
Projet Apache initié par Cloudera, **Flume** collecte, agrège et transfère sur Hadoop de gros volumes de données (logs, événements, etc.). L'écriture de ces informations sur HDFS est réalisée au fil de l'eau.

Apache **Sqoop** permet de récupérer des données structurées dans des bases de données ou des datawarehouses pour les déplacer sur un cluster Hadoop. Ensuite, il peut effectuer des opérations en les corrélant avec des informations semi ou non structurées sur Hadoop. Enfin, il renvoie un résultat structuré au format des SGBD source.

Ce panorama de quelques outils illustre les grands progrès réalisés par Hadoop en trois ou quatre ans pour transformer cette plateforme en réelle solution d'entreprise : robustesse, évolutivité, administration, sécurité, gouvernance, disponibilité, stockage... De plus en plus d'experts parlent (plus ou moins ouvertement) d'Hadoop comme stockage de référence des données, y compris pour les traitements transactionnels. Et la notion de Data Lake reprise par nombre de spécialistes semble bien aller dans ce sens.

D'ailleurs, le ralliement à de nombreux projets Hadoop de tous les géants de l'informatique (voire leur financement des pionniers Hadoop en dizaines de millions de dollars) n'est-il pas révélateur ?

Projets



UN PROJET HUMANITAIRE : LE DÉVELOPPEMENT ÉCONOMIQUE ET LE BIG DATA

UNITED NATIONS Global Pulse

Nous parlions au début de ce guide des nombreux projets menés par le gouvernement américain et de leur ampleur.

On évoquait aussi les ambitions de l'administration française.

Voici ici la preuve qu'une instance, quelles que soit sa taille et sa complexité, peut initier des projets Big Data puissants et efficaces.

Au sein des Nations Unies est né il y a quelques années un beau projet qui s'appuie entièrement sur des technologies Big Data.

Global Pulse, initié en 2009, vise à promouvoir l'adoption du Big Data à des fins d'aide au développement et de projets humanitaires. Le projet permet notamment de fournir une analyse en temps réel de l'avancement des programmes d'aide au développement dans le monde, de leur efficacité et de leurs retombées. Il est né de la volonté du secrétariat général des nations unies de faire connaître l'analyse des données massives comme un outil puissant pour le développement économique. C'est aussi une nouvelle approche pour le secteur humanitaire, peu habitué à des moyens supposés onéreux et rarement associé aux nouvelles technologies. S'engager dans des programmes humanitaires, viser à améliorer le niveau de vie de populations en situation de crise ou appuyer le développement d'un pays en retard est une lourde tâche qui gagnerait énormément à connaître ses points d'échec et de réussite. Si on peut mesurer, monitorer en temps réel et même prédire le chiffre d'affaires d'une société, on doit tout aussi bien pouvoir évaluer le degré de réussite d'une action humanitaire. Non pas à des fins économiques. C'est l'efficacité qui est visée. Un même budget idéalement réparti apportera un meilleur appui au projet mené. Les Nations Unies se positionnent là en plateforme d'innovation en quelque sorte. L'objectif est de permettre le partage d'information et d'accélérer l'utilisation du Big Data dans les programmes qu'ils soutiennent.

Concrètement, Global Pulse fonctionne comme un réseau, où est conçue et coordonnée l'innovation. Diverses entités des Nations Unies sont sollicitées pour y participer, ainsi que les gouvernements, universités et les entreprises privées souhaitant apporter leur expertise en recherche et développement. Le concept s'appelle BD4D, Big Data for Development.

“Le projet est né de la volonté du **secrétariat général des nations unies** de faire connaître l'analyse des données massives comme un outil puissant pour le développement.”

L'idée part d'un constat assez simple : l'évolution du niveau de pauvreté, les conditions de vie et les crises économiques sont analysées généralement a posteriori. Avec un certain nombre d'indicateurs, on déclare par exemple que le niveau de pauvreté s'est aggravé dans une région donnée. Or de nos jours, et ce même dans les zones les moins développées, les nouvelles technologies ont envahi le quotidien, l'économie et les administrations. **Chaque jour, en continu, les individus et institutions produisent des « signaux digitaux » sur les modes de vie.** Les réseaux sociaux par exemple, peuvent refléter les états d'esprits d'un pays ou d'une communauté. Le niveau d'épargne et son évolution traduisent la santé économique des ménages mais aussi leur confiance en l'avenir. Les mots-clés entrés sur les moteurs de recherche peuvent révéler des tendances, des mouvements sociaux voire même des épidémies. On se rappelle de l'exemple en France de l'épidémie de grippe début 2014, prédite par le gouvernement mais contredite par la société Celtipharm.

Grâce au Big Data, Global Pulse encourage l'analyse de ces signaux, qui, observés en temps réel, fournissent un réel baromètre du développement au niveau mondial. Chaque action humanitaire ou d'aide au développement peut bénéficier d'un feedback immédiat. Les projets peuvent alors être adaptés, améliorés, recadrés.

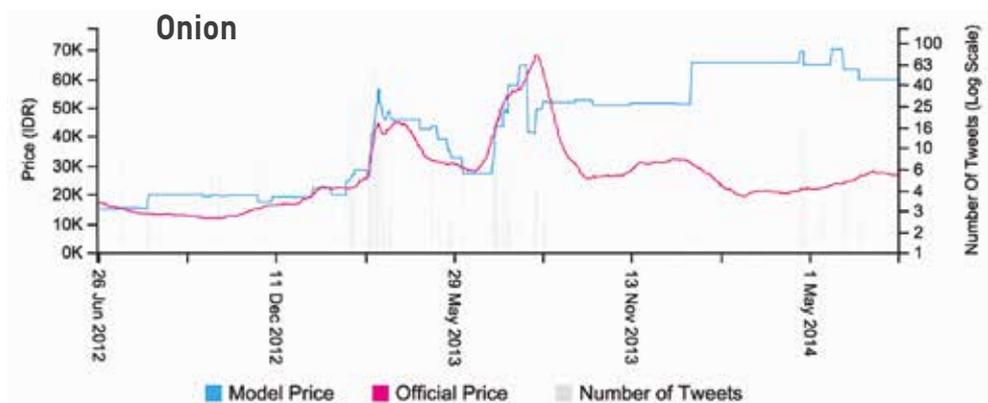
Robert Kirkpatrick, le directeur de Global Pulse, explique que l'initiative est née suite aux débuts de la crise financière mondiale. La propagation à grande vitesse des effets de la crise a alors prouvé au monde entier l'importante interconnexion de nos économies et de

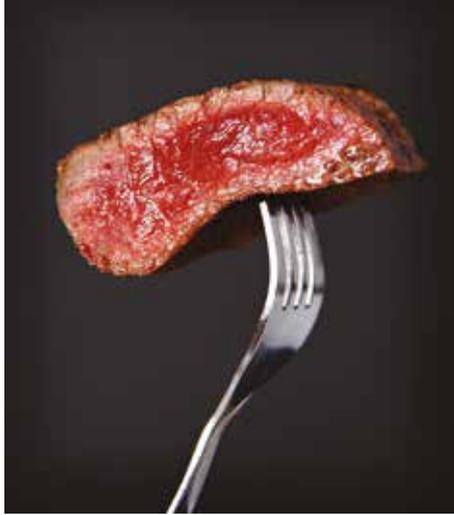
nos systèmes d'information. Paradoxalement, la majorité des études statistiques de l'époque utilisaient des chiffres datant de deux ou trois années. Même le secteur privé, alors qu'il est le principal gestionnaire de ces données, est encore difficilement capable de les traiter en temps réel. Global Pulse a donc voulu répondre à cette réalité du Big Data et encourager le secteur humanitaire à s'emparer de cet outil extrêmement puissant.

De nombreux projets, partenariats et analyses diverses sont alors nés de Global Pulse. Par exemple, les équipes se sont récemment demandé si Twitter pouvait être un bon indicateur de l'évolution du prix de biens de consommation. Ayant choisi de se concentrer sur l'Indonésie, un pays qui se classe parmi les cinq premiers foyers d'utilisateurs Twitter au monde, Global Pulse a testé cette théorie.

Avec l'aide du World Food Program et de Bappenas (Ministry of National Development Planning, Republic of Indonesia), Global Pulse a ainsi initié le projet Pulse Lab Jakarta. Trois produits de consommation parmi les plus populaires ont été sélectionnés : l'oignon, le bœuf et le poulet. Les équipes ont alors recherché sur Twitter les allusions au prix de ces trois aliments, observant leur évolution et vérifiant l'hypothèse d'une influence et corrélation entre le prix de ces produits dans le temps. Le prix du poulet au jour J influencerait donc son prix au jour J+1.

Pour chacun de ces produits, Pulse Lab Jakarta construit un modèle et fournit un graphique superposant : le prix tel que le modèle le détermine (avec l'aide du prix relaté sur Twitter), le prix tel que l'État déclare qu'il est, et le nombre de Tweets. La conclusion ? On peut, a priori, prévoir le prix d'un bien de consommation grâce à Twitter. Les exemples du bœuf et du poulet ont surtout prouvé qu'un pic d'activité sur Twitter correspondait à une hausse de prix.





UN PROJET CULTUREL : LE DATA JOURNALISME POUR LES FLUX MIGRATOIRES EN EUROPE

THE MIGRANT FILES - Entretien avec Nicolas KAYSER-BRIL, CEO and Founder, Journalism++

Après avoir dirigé le data journalisme chez OWNI, Nicolas KAYSER-BRIL a cofondé Journalism++, avec comme optique de faire parler les données à des fins journalistiques.

Le projet The Migrant Files, publié le 31 Mars dernier, montre la réalité des accidents, décès et disparitions survenues à l'occasion de migrations aux frontières de l'Europe.

Retrouvez la vidéo animée de

Global Pulse



Voir l'intégralité de l'interview de

Robert Kirkpatrick



Découvrez en visualisation interactive les résultats du projet

Pulse Lab Jakarta



Les données utilisées proviennent de sources diverses, issues pour les deux tiers d'ONG, d'articles de presse récupérés via United ou Fortress et de plus de 15 journalistes relais. Le tiers manquant provient de sources ajoutées par les équipes de Journalism++. Ces données, non structurées pour la grande majorité, sont agrégées, classées puis traitées par un programme d'extraction spécifique. Alors que jusqu'à présent de nombreuses données restaient inutilisées de par leur diversité, complexité et multiplicité, The Migrant Files permet de les réunir et d'en tirer des conclusions les plus exhaustives possibles.

Frontex, l'organisme européen issu du traité d'Amsterdam de 1997, est chargé de mutualiser les informations et d'améliorer les procédures de gestion des frontières, flux de biens, personnes, services et capitaux. Il agrège entre autres la détection des tentatives d'immigration, les décès, échecs, reconduites à la frontière, demandes de droits d'asile... Le Frontex a ainsi défini 8 zones géographiques autour de l'Europe, aux caractéristiques semblables. The Migrant Files s'est appuyé sur ces données et a observé en détail l'activité de ces 8 zones frontalières. La qualité des données subit une marge d'erreur de 10 % environ. The Migrant

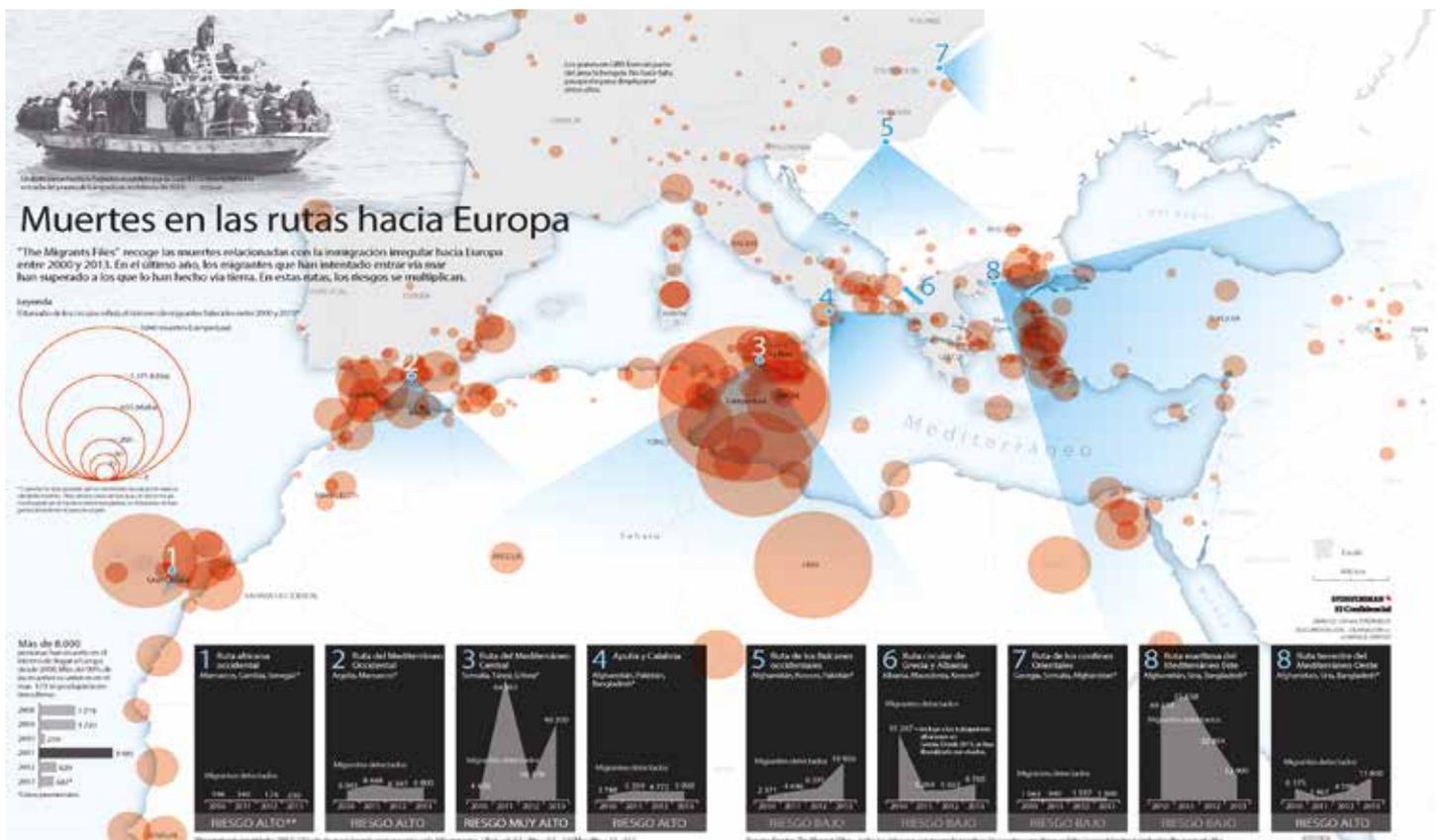
Files considère les « événements » survenus aux frontières, chaque événement représentant une disparition ou un mort. Un unique événement est caractérisé par plus de 20 points de données, toutes issues des sources précédemment énumérées.

D'après les premières conclusions de l'étude, on peut estimer le nombre de migrants décédés aux frontières européennes entre 30 000 et 50 000 depuis l'an 2 000.

Le chiffre est impressionnant et révélateur de nombreux dysfonctionnements et contradictions des Etats-Membres de l'Union Européenne. Si The Migrant Files a été relayé dans la plupart des pays de l'UE, via des médias tels Le Monde Diplomatique, L'Espresso, El Confidencial, et plus de 50 articles, les conclusions en sont à ce jour relativement peu utilisées dans le cadre géopolitique ou juridique.

Certains politiciens s'y sont intéressés, autant au niveau national qu'à l'échelle de l'Union, mais aucun n'a l'obligation de mesurer ou prendre en compte ces chiffres.

Ce que révèle également The Migrant Files, c'est la dangerosité de certaines zones par rapport à d'autres. La zone de Lampedusa par exemple, est tristement célèbre pour la mortalité des migrants qui tentent d'y passer. Le





programme Mare Nostrum initié à la suite de plusieurs naufrages meurtriers survenus en Octobre 2013 est l'une des plus grandes actions humanitaire et militaire sur le sujet et est mise en place par l'Italie. Les données analysées par The Migrant Files confirment la dangerosité de cette zone, mais ce qui donne un nouveau visage à l'immigration en Europe, c'est la possibilité de pouvoir comparer tous les points d'entrée, sur la base de données similaires. La carte montre ainsi que Gibraltar et les îles Canaries par exemple sont elles aussi des zones à fort risque.

L'autre force de The Migrant Files, qui est à mettre en relief avec une vision journalistique, est la possibilité d'effectuer une analyse temporelle. Car si analyser les flux au temps-T a du sens, observer les évolutions dans le temps apporte un regard plus fin sur l'impact des

politiques migratoires au cours des années. L'exemple de la frontière turco-grecque est significatif: alors que les politiques nationales avaient, il y a quelques années, eu pour objectif de réduire les flux par le Bosphore, The Migrant Files révèle que les migrants ont ouvert une nouvelle route dans les années qui ont suivi, privilégiant les îles de la mer Égée. Ce trajet de substitution est en réalité plus meurtrier que le précédent. Les décisions de politiques de gestion des frontières pourraient donc être analysées par le biais de la donnée, ouverte, disponible pour tous et entrant en compte dans l'équilibre recherché par l'Union Européenne.

CLANDESTINS, LES MIGRANTS ?
Remarquons qu'à aucun moment The Migrant Files ne parle « d'immigration clandestine ».

Nicolas Kayser-Bril nous explique pourquoi : les migrations ne sont pas clandestines. On devient migrant en situation irrégulière uniquement lorsque la demande d'asile a été rejetée. Aucun migrant qui traverse par voie de terre ou voie de mer n'est clandestin, ils sont tous dans le cadre du droit international fixé par la convention de Genève de 1951, ratifiée par tous les états membres de l'Union.

Visitez la plateforme

«The Migrant Files»



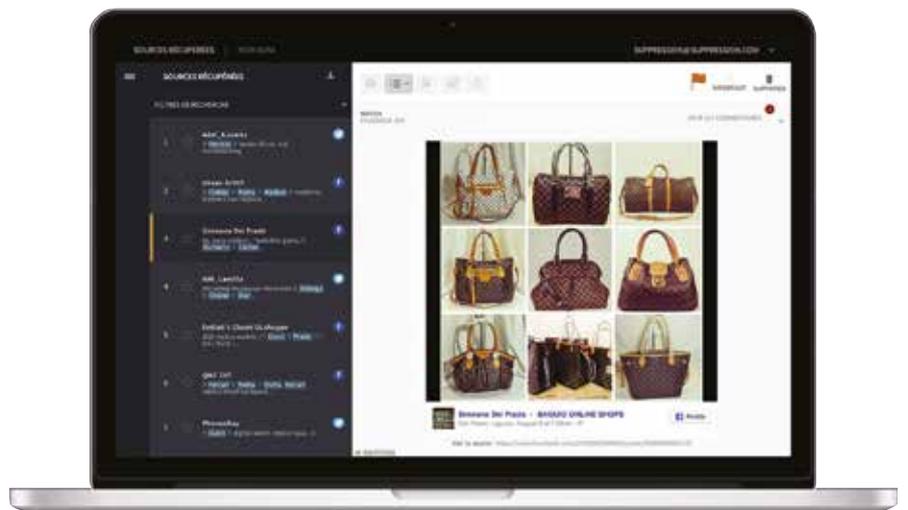
UN PROJET DE LUTTE CONTRE LA FRAUDE : LA DÉTECTION DE LA CONTREFAÇON EN LIGNE

BRAND WATCHDOG - Entretien avec Zouheir GUEDRI, CEO, Data & Data

Anciennement directeur du conseil chez PWC, Zouheir Guedri est un passionné de la data. C'est donc avec entrain qu'il a conçu et participé à la mise en fonctionnement, depuis avril 2014, d'un outil de détection de la contrefaçon.

Dans un premier temps destinée aux plus grands acteurs du secteur du luxe, la solution est vouée dès septembre 2014 à servir le secteur pharmaceutique, et à se développer dans d'autres sphères par la suite. Le principe? Screener internet et les réseaux sociaux, à la recherche d'articles contrefaits. « C'est comme un Google inversé » explique Zouheir Guedri. L'outil nécessite la construction d'algorithmes spécifiques, adaptés à chaque secteur et typologie de produits. Pour le luxe, le premier constat était que les techniques traditionnelles qui consistent à comparer un certain nombre de points ne suffisaient pas à identifier efficacement les sources frauduleuses. La photo n'est pas un indicateur fiable, puisque de nombreux sites de contrefaçon affichent la photo issue du site de la vraie marque, le prix de vente peut être un indice, mais ce n'est pas systématique, alors que l'analyse de la source du site web, son service client, sa licence, etc, est souvent révélateur. En moyenne et pour chaque source identifiée, **plus de 300 points de mesures sont**

“Selon l'**OCDE**, le marché de la **contrefaçon** en ligne connaît **plus de 25 %** de croissance annuelle depuis 2010.”





analysés, à partir d'une douzaine de médias sociaux les plus fréquentés et de l'ensemble du web. Le résultat est donc relativement exhaustif, en plus de bénéficier de traitements en temps réel.

Avant le Big Data et les possibilités offertes par le cloud, seules des structures telles Google ou Yahoo pouvaient faire ce type de traitements. L'achat d'espace de stockage adapté aux besoins, en toute flexibilité, permet à des start-ups d'implémenter des outils de type Big Data sans se heurter à la barrière financière.

L'idée ici est d'identifier les réseaux, de tracer la cartographie et de voir les nœuds avec les sites cachés derrière une filière de distribution d'articles contrefaits. L'action à suivre est définie au cas par cas, selon la stratégie de la marque, qui peut soit souhaiter dénoncer l'hébergeur, avertir les autorités, ou simplement notifier le pro-

priétaire du site. Cette stratégie est modélisée sous forme de règles, revues en moyenne une fois par mois au début, puis systématisée par la suite. L'objectif étant de se laisser le temps de connaître les retombées, analyser les sources, la façon dont est utilisé le réseau, pour pouvoir définir une stratégie adaptée.

La question qui vient alors est celle de la rentabilité et de l'intérêt réel de cette pratique: la contrefaçon en ligne représente-t-elle réellement une part importante du parasitage du chiffre d'affaires? Selon l'OCDE, le marché de la contrefaçon en ligne connaît plus de 25 % de croissance annuelle depuis 2010. Les médias sociaux représenteraient à eux seuls jusqu'à 90 % de ce trafic. Mais là encore tout dépend de la marque et de l'article concerné, qui peut parfois être davantage représenté sur des sites

merchants plutôt que sur les réseaux sociaux. Pour le secteur du luxe, les équipes ont pu constater qu'un seul et même réseau de sites frauduleux pouvait effectuer jusqu'à 10 Millions de dollars de chiffre d'affaires sur un mois. Dans certains cas, il se révèle plus intéressant de laisser libre court aux activités illicites d'un site afin de pouvoir observer, mesurer et ensuite justifier auprès des autorités le montant du préjudice. **En à peine un mois, la plateforme a ainsi pu évaluer plus de 140 Millions d'euros de préjudices.** La suite est donc prometteuse...

Cependant, certaines marques ne souhaitent pas mener d'action juridique, et ont comme unique objectif de stopper toutes les sources identifiées. Là encore, tout est une question de stratégie.

En définitive, les 2 éléments clés de ce projet sont la possibilité de mener une surveillance et d'enclencher une riposte systématique, et la prise en compte des médias sociaux.

Si à ce jour l'utilisation vise les secteurs du luxe, des cosmétiques et de la pharmacie, il paraît assez évident que de nombreux acteurs pourraient vouloir dupliquer l'idée. Et le secteur privé n'est pas le seul concerné: les services des douanes envisagent l'utilisation de ce type d'outil hyperpuissant pour l'optimisation de leur fonctionnement, qui doit désormais prendre en compte l'importance de la contrefaçon virtuelle et construire les passerelles entre la détection des sources frauduleuses et la confiscation des marchandises aux frontières, aujourd'hui de plus en plus virtuelles elles aussi.

Dans le domaine industriel, l'exemple, devenu quasi historique aujourd'hui, de la pièce contrefaite à l'origine de l'accident du Concorde, ouvre de nombreux champs pour la qualité et la sécurité dans des secteurs à forts enjeux.

À une échelle beaucoup plus large et s'inscrivant dans la dynamique open data, on peut également s'imaginer à rêver une plateforme en accès libre, gratuite, fonctionnant tel un moteur de recherche, qui permettrait à l'internaute de vérifier l'indice de fiabilité d'un site marchand qu'il visite.

UN PROJET DANS LE SECTEUR AUTOMOBILE : L'ASSURANCE ÉVOLUTIVE

PROGRESSIVE

L'assurance est un secteur hautement compétitif et crucial pour ses consommateurs. Choisir, contracter, adapter ou résilier son assurance peut parfois relever du parcours du combattant. Encore plus critique, l'assurance automobile est souvent considérée comme chère voire inaccessible. Le système des malus et la prise en compte des antécédents du conducteur ajoutent une barrière à l'entrée, au-delà de tarifs parfois prohibitifs. Le métro Parisien est d'ailleurs depuis plusieurs années fréquemment orné de publicités attirant l'œil d'automobilistes a priori délaissés par le système classique des assurances. Un créneau existe, une nouvelle approche est attendue.



Quel rôle joue ici le Big Data? Alors que les capteurs, systèmes de géolocalisation et autres technologies mobiles se multiplient, l'assurance automobile voit son nouveau modèle économique se profiler. Jusqu'à présent, le montant d'une prime était déterminé par quelques facteurs basiques, facilement identifiables et prédéterminés. Puissance du moteur, nombre de kilomètres de la voiture, modèle, âge du conducteur, nombre d'années de permis...

“Grâce au Big Data, les compagnies d’assurance peuvent désormais obtenir, en temps réel, des informations réelles sur le type de conduite de leurs assurés.”

Grâce au Big Data, les compagnies d’assurances peuvent désormais obtenir, en temps réel, des informations réelles sur le type de conduite de leurs assurés. Cette personne roule-t-elle souvent au-dessus de la limitation de vitesse ? Celle-ci grille-t-elle souvent les feux ? Combien de temps conduit-elle en moyenne par jour ? Y a-t-il eu une évolution dans son type de conduite récemment ? Vient-elle de se localiser sur Facebook dans un bar — et donc a-t-elle des risques de conduire alcoolisée ?

Tant d’éléments qui fournissent une analyse beaucoup plus fine du profil auquel la compagnie est confrontée.

En 2011, la société Progressive a créé la surprise en lançant son service d’assurance auto « Pay As You Drive ». Littéralement, cela signifie que l’automobiliste voit le tarif de son

assurance s’adapter à son type de conduite. Le prérequis ? Accepter l’installation de capteurs et systèmes de tracking sur sa voiture — Snapshot en l’occurrence. Évidemment, on peut trouver le principe extrêmement intrusif. Comment garantir que la compagnie d’assurances n’alertera pas la police si l’automobiliste dépasse la vitesse limitée ? Ce ne sont pas ses prérogatives.

Du côté de l’assuré, on peut voir les choses avec enthousiasme. Un jeune conducteur, selon les critères classiques relativement mal accueilli par les compagnies d’assurances, peut se révéler être un conducteur beaucoup plus prudent qu’un automobiliste expérimenté n’ayant apparemment jamais eu d’accident majeur. Chez Progressive, ce jeune conducteur présentant une attitude dite idéale, verra la compagnie lui offrir des réductions sur le tarif qu’il paye. À l’inverse, l’automobiliste à risque devra s’acquitter de sanctions tarifaires. L’intérêt pour Progressive ? Le jeune conducteur est, statistiquement, un automobiliste qui ne lui coûtera pas cher en remboursements et couvertures. L’autre est plus enclin à engendrer des frais importants pour la société. Et si l’automobiliste taxé se trouve être mécontent du traitement qui lui est réservé, il peut toujours changer d’assurance. Pas de perte majeure pour Progressive, qui conserve son assuré « rentable » et clôt le compte d’un client à risque.

Au-delà de l’aspect monétaire, le monitoring de la conduite permet aussi à la société de mettre à disposition de ses assurés un outil de suivi. Chacun peut, presque en temps réel et avec une traçabilité, observer l’évolution de son type de conduite. Un automobiliste bien intentionné disposera alors d’un moyen efficace de mesurer si ses efforts sont efficaces, où il doit s’améliorer et comment.

Progressive n’est pas la seule compagnie à proposer ce type d’assurance, qui se développe désormais largement en Amérique du nord. Certaines critiques et témoignages d’anciens assurés dénonçant les pratiques de Progressive ont quelque peu entaché l’image du pionnier de l’assurance « Pay As You Drive ». Cette approche traduit cependant l’avenir de l’assurance automobile, en fortes probabilités.

LE MODÈLE DE PROGRESSIVE, EXPLIQUÉ EN 3 ÉTAPES À SES CLIENTS :

PLUG - DAY 1

The Snapshot device fits right into a little port in your car, generally located below the steering wheel.



DRIVE - DAYS 1-30

We look at your driving habits to see if you could be saving more. You can track your projected savings online.



SAVE - DAY 31

Now your rate is fully personalized; any discount you earned is in full effect. Simple!



UN PROJET DANS LE SECTEUR DES LOISIRS : LES COMMUNAUTÉS VIRTUELLES

X-FIELD PAINTBALL - Entretien avec Sébastien MUNIER, Directeur Général, PCB TEAM

Sébastien Munier n’est pas un expert du Big Data. L’idée de sa société est venue avec son associé, Fabrice Tron, il y a environ 2 ans. Fabrice et Sébastien sont des entrepreneurs du numérique, tous deux ayant eu de précédentes expériences en agences web. Ils ont donc pendant longtemps répondu aux problématiques numériques de leurs clients.

En 2007, Fabrice travaille sur Second Life, un univers entièrement virtuel, recréant une société entière, avec ses individus, associations, hommes politiques...le tout en 3D. C’est le premier à être labélisé en Europe par Linden Lab, l’éditeur Californien de Second Life pour l’insertion de marques au sein de la plateforme. Ils travaillent alors avec de nombreuses entreprises du CAC40, qui perçoivent rapidement leur intérêt à être présent sur ce type de support virtuel. Les marques profitent de ce nouveau lieu de rencontre pour faire connaissance avec leurs clients, de façon virtuelle mais ciblée, par affinités communautaires. La puissance du modèle est d’autant plus forte que de nombreux utilisateurs se sont réellement pris



Finale du championnat Européen [Millennium Series] à Chantilly, monde réel.



Action identique dans le jeu.



Un des terrains les plus réputés au monde : SC Village, Los Angeles, Californie.



Et sa représentation dans le jeu (positionnement exact des obstacles via Google Earth)

au jeu, recréant leur univers réel jusqu'à répliquer leur mobilier à l'identique. Les marques présentes sur la plateforme avaient donc des informations précieuses sur les goûts de leurs potentiels clients.

Suite à cette expérience, Fabrice et Sébastien ont eu l'idée d'exploiter cet univers des communautés. Savoir reconnaître des groupes de personnes avec une affinité commune, les capter, développer la communauté et la monétiser auprès de marques liées deviendrait leur nouveau business model.

Concrètement, PCB Team a actuellement trois projets en cours, chacun centré sur un centre d'intérêt précis : le paintball, la musique électronique et le camping.

X-Field Paintball est le premier projet qu'ils ont développé, et le plus abouti. Si le secteur peut paraître surprenant, il faut tout d'abord noter que **la communauté des passionnés de paintball représente plus de 20 millions de pratiquants à travers le monde**. C'est la troisième communauté de sport extrême, derrière le BMX et le snowboard aux USA. Ce qui est d'autant plus intéressant, c'est que la pratique du paintball représente des coups similaires à la pratique du golf. Un passionné est donc susceptible de dédier des budgets importants pour son activité, et les marques de paintball ont un fort intérêt à le capter (**marché mondial de 200 millions d'euros d'équipements et 1,3 milliard d'euros de consommables – les billes**).

Fabrice et Sébastien ont donc développé une application mobile, testée sur le marché au préalable, qui atteint rapidement des centaines de milliers de téléchargements.

Leur communauté au sens large comprend les pratiquants du paintball, mais aussi les industriels (marques d'équipements), les exploitants de terrains de jeu, les ligues (des championnats sont régulièrement organisés) et la presse spécialisée. En un peu plus de 6 mois, PCB a amorcé ce nouveau média, l'objectif étant

d'agrèger le maximum d'information qualifiée au niveau mondial, le tout depuis leurs petits bureaux toulousains.

Une levée de fond et 3 millions de téléchargements plus tard, les plus grandes marques du secteur s'associent toutes à X-Field Paintball. C'est eux qui sont alors prescripteurs auprès des pratiquants tout en participant au financement de la plateforme. Leur intérêt ? Bénéficier de l'accès à leur communauté, fédérée de façon unique au monde. C'est en d'autres termes une immense base de données, la plus grande, hyperqualifiée, sur la communauté du paintball. Comment les joueurs sont-ils attirés sur la plateforme ? Grâce à une précision sans égal. Chaque terrain est reproduit, les actions réelles rendues virtuelles, les « marqueurs » (nom des lanceurs de billes), les décors, les mouvements sont transcrits à la lettre. Le joueur peut donc retrouver le plaisir du jeu réel dans un univers virtuel, avec des interactions au niveau mondial. X-Field est même allé jusqu'à reproduire les terrains des championnats, en les mettant à disposition quelques jours à l'avance afin que les participants puissent s'y entraîner. Les plus gros sponsors, comme lors des championnats réels, diffusent leurs bannières de publicité pour figurer en meilleure position sur le terrain. C'est un marché de niche, avec huit à dix acteurs majeurs dans le monde, mais qui représente des enjeux forts pour ces quelques marques. La valeur ajoutée de X-Field est de détenir une information exclusive.

En quoi ce projet est-il réellement Big Data ? Le

“Marché mondial de 200 millions d'euros d'équipements.”

volume d'informations récoltées — 3 millions d'utilisateurs ultra-qualifiés — en un temps record, incluant la géolocalisation et permettant la détection de lead, requiert l'utilisation de solutions Big Data.

Un exploitant de terrain par exemple, pourra bientôt via une plateforme de géolocalisation repérer les pratiquants répondant à un certain profil dans son entourage au temps-T, les cibler, les adresser via un système de RTB et les attirer vers son terrain. Le business est alors classique : la marque ou le terrain transmet un push proposant une offre promotionnelle à ses leads.

D'après Sébastien, deux types de pratiquants sont à cibler : les fans du jeu dans le réel et les fans de jeux vidéo au sens large, qui pourront prendre goût à X-Field et finalement aux terrains de paintball réels. L'objectif est là d'élargir la communauté autant que possible, à moindre coût et le plus rapidement possible.

Le deuxième projet le plus abouti de PCB Team est celui concernant la communauté du camping. En France, **l'hôtellerie en plein air représente 38 millions de réservations par an**. La promesse de la société est de permettre à la communauté des campeurs de personnaliser et vivre ses vacances plus intensément.

Le dernier projet en cours chez PCB est celui des communautés de musique électronique (Pink City Beats). Il est encore à un stade embryonnaire, à l'étape de captation de sa communauté (600 artistes DJ internationaux déjà), mais présente comme objectif sous-jacent la volonté pour les équipes de PCB de démontrer le caractère duplicable de leur approche.

À ce jour, X-Field en est à sa seconde levée de fonds. L'objectif immédiat est de passer de 3 à 6 millions de téléchargements de la plateforme. Le projet autour du camping vient d'effectuer sa première levée de fonds et d'entrer en commercialisation.

À long terme, la société PCB et ses deux fondateurs souhaitent répliquer leur fonctionnement dans d'autres secteurs de niche, essentiellement dans des domaines sportifs. Les trois premiers projets seront là pour montrer la puissance de ces bases de données hautement qualifiées exploitées avec des outils Big Data. L'idée, dans les rêves les plus fous de Sébastien, sera ensuite de croiser les différentes communautés. Un scénario type : cet individu, en vacances dans ce camping, joue souvent aux jeux vidéo, dont parfois au paintball, et comme il n'aime pas les musées, qu'il a une voiture et qu'il est venu sans enfants, je peux lui envoyer une offre promotionnelle pour ce terrain de paintball à moins de 30 minutes de son lieu de vacances.

La seule question : la protection des données personnelles. Là, la réponse de Sébastien s'appuie sur un principe d'opt-in et une anonymisation maximale des informations. L'identité nominative importe peu en fin de compte. C'est le profil qui prime. Si la société est capable de reconnaître un profil et de lui communiquer une offre, elle n'a pas besoin de connaître le nom de la personne concernée. Quoi qu'il en soit, les équipes garderont un œil averti sur les recommandations de la CNIL et veilleront constamment à protéger les données.

UN PROJET DANS L'ÉNERGIE : COFELY SERVICES

DEEPMI - Entretien avec Vincent BRYANT, Direction d'activités Efficacité Énergétique, Cofely Services, Groupe GDF SUEZ

Cofely Services, Groupe GDF SUEZ, est principalement connue pour ses services en efficacité énergétique. Rattaché au DG de la société, Vincent Bryant est en charge de développer les nouvelles activités ayant pour vocation d'améliorer l'efficacité énergétique.

Il n'est pas DSI, il ne fait pas partie du département R & D, mais il est à l'interface de ces métiers. Son rôle est d'utiliser les algorithmes de data analytics pour créer de nouveaux business.

Cofely Services a développé il y a 2 ans l'offre VERTUOZ, qui permet à ses clients (**plus de 41 000 sites renseignés à ce jour**) de monitorer leurs consommations en temps réel et d'amener la « Business Intelligence » jusque chez ses clients.

Mais la data pour l'efficacité énergétique, ce n'est pas que de la mesure.

Comme l'explique Vincent Bryant, **il existe 3 niveaux d'utilisation de la donnée**. Le premier correspond à la mise en place d'un système d'energy monitoring pur : connaître ses consommations réelles au temps-T, en garder une trace. Le second niveau consiste à intégrer des fonctionnalités BI (Business Intelligence) pour mieux comprendre et exploiter ces données. Le troisième niveau, qui nous intéresse ici, s'attaque au Big Data en intégrant des données externes, des volumes de données importants et un traitement en quasi-temps réel.

Dans ce contexte, le Big Data peut servir au Benchmark : connaître combien je consomme est intéressant, mais cette donnée prend tout son sens lorsque je peux la comparer de ma-

nière fine avec les consommations de structures similaires à la mienne.

Pour Cofely Services, le Big Data est également un moyen de mieux connaître ses clients, et donc de mieux les conseiller. L'optimisation de la vente croisée permet à Cofely Services de fournir à ses clients des services parfaitement adaptés à leurs besoins. En complément des données CRM déjà présentes, les algorithmes de data-analytics viennent tirer plus de valeur des données volumineuses et complexes existantes utilisées par les équipes marketing. À titre d'exemple, un seul client peut représenter plus de 1 000 sites, chaque site ayant ses propres caractéristiques de consommation, d'environnement, d'activité...**Le Big Data chez Cofely Services est donc passé par la construction d'un outil de Data Discovery puissant** mis à disposition des équipes. « Je connais mon client pour mieux lui vendre mes services, et mon client bénéficie ainsi d'une réponse adaptée au plus proche de ses besoins ». L'outil a également engendré l'identifi-

cation de patterns et typologies de clients, pour que Cofely Services comprenne de manière beaucoup plus fine les consommations de ses clients.

Mais là où l'entreprise va plus loin encore, c'est en lançant une spin-off, DEEPMI. Comme l'explique Vincent Bryant, au 19e siècle, le monde a cédé à la ruée vers le pétrole aux États-Unis. De nombreux acteurs se sont précipités au Texas, ont acheté une parcelle de terrain et quelques pelles et pioches pour creuser. Certains ont trouvé du pétrole, d'autres non et beaucoup de pioches ont été vendues. Puis des entreprises d'ingénieurs ont proposé d'analyser les sols pour identifier les gisements et les meilleurs outils pour les exploiter, en orientant ainsi les efforts là où il était probable qu'il y ait du pétrole. Cette cartographie des gisements de pétrole a apporté une grande valeur aux chercheurs de pétrole, aux vendeurs d'outils et aux propriétaires fonciers. « Aujourd'hui chez Cofely Services, nous voulons faire pareil » souligne alors Vincent Bryant. **La cartographie développée par DEEPMI vise à détecter les gisements d'économies d'énergie grâce aux données existantes du secteur privé, des ministères, collectivités, mais aussi à l'aide de données ouvertes**. Ces informations sont structurées, segmentées en classes de bâtiments à l'aide d'algorithmes, puis traitées avec un moteur d'inférence.

Les équipes obtiennent des cartographies de gisements d'économies d'énergie complétées par des diagnostics ou des systèmes de mesures. La démarche permet donc d'exclure les cas particuliers, et d'affiner autant que possible les études. Dans l'énergie, c'est une petite révolution : il ne s'agit plus uniquement de mener des études physiques, mais bien de créer une cartographie préalablement. Ces puissants algorithmes utilisent des outils statistiques bien connus des sociétés de promotion publicitaire sur le web. Le principe est relativement simple : alors qu'on aurait ten-





dance à vouloir comprendre les mécaniques d'achat d'un consommateur (par exemple, si j'achète une table, j'achèterai vraisemblablement les sièges qui vont avec), mieux vaut en observer statistiquement les comportements réels. Si l'analyse nous dit que la majorité des clients similaires ont acheté une lampe après avoir ajouté cette table, mieux vaut mettre en avant des lampes pour susciter l'achat. Le raisonnement est le même pour DEEPI: si au sein d'un parc de boutiques, on constate que les sites qui consomment le plus sont ceux qui ont la fréquentation la plus élevée, alors pourquoi continuer à benchmarker les consommations d'énergie des sites en fonction de leur date de construction? La recommandation la plus adaptée sera de s'intéresser aux boutiques ayant une fréquentation importante. **Déterminer ces « facteurs d'influence » est un enjeu crucial pour dynamiser l'efficacité énergétique.** Reste alors à savoir quelles données traiter. Car si l'on considère que des données a priori sans aucun lien de causalité avec l'événement peuvent être considérables, il faut alors élargir autant que possible le spectre des données prises en compte dans l'étude. Elles peuvent prendre la forme toujours dans notre exemple des boutiques, du trafic piétonnier alentour, des horaires d'ouverture, du type d'ouverture, de la fréquentation, du chiffre d'affaires... La question vise ensuite à déterminer quelle part prend chacun de ces facteurs dans la consommation de la structure.

Afin de sélectionner les jeux de données à intégrer à chaque étude, les équipes de DEEPI s'appuient sur deux éléments: la disponibilité de la donnée (puis-je avoir accès à cette information?) et la connaissance métier (quels sont les facteurs d'influence selon les exploitants du site?).

La première étape est donc un diagnostic technique, qui amène à la constitution d'une Data Room regroupant toutes les données disponibles et utilisables.

C'est ainsi que des constats contre-intuitifs peuvent émerger: **grâce à l'analyse réalisée sur des données récupérées sur plus de 400 boutiques en France, des segments de boutiques homogènes d'un point de vue des consommations d'énergie ont été identifiés:** le constat qui en a résulté a amené à la conclusion que les boutiques les plus consommatrices n'étaient ni les plus anciennes, ni les plus grandes.

Voilà de quoi donner des idées à de nombreux gestionnaires d'actifs sur l'analyse des données qu'ils possèdent déjà.

“plus de 41 000 sites
renseignés à ce jour.”

UN PROJET DE RECHERCHE : LA GÉNOMIQUE

France GENOMIQUE - ENTRETEN AVEC PIERRE LE BER ET PIERRE LECA

Pierre Le Ber est polytechnicien de formation. Il a complété son cursus par un doctorat en pharmacologie moléculaire, et s'est rapidement orienté vers les sciences du vivant, les biotechnologies et l'industrie pharmaceutique. Pierre intègre, en 1995, la société Genset, un des pionniers de la génomique. Puis, après avoir travaillé à Singapour, en Chine, en Suisse, il occupe depuis 4 ans le poste d'adjoint au directeur de l'Institut Génomique (IG) du CEA, ainsi que de Responsable Scientifique et Technique de l'infrastructure France Génomique.

Pierre Leca pour sa part a davantage un profil de mathématicien et d'informaticien. Docteur en mathématiques appliquées, il se plaît à résoudre des problèmes complexes à l'aide d'ordinateurs. Il a débuté sa carrière dans le secteur aéronautique et a rejoint le CEA en 2001 pour prendre la direction du Département des sciences de la simulation et de l'information.

Ce Département est en particulier responsable du complexe de calcul du CEA et mène des projets de R & D avec de nombreux industriels.



Crédits photos : © P. Stroppa/CEA

Les deux hommes sont donc résolument complémentaires, l'Institut de Génomique du CEA auquel Pierre Le Ber appartient travaillant sur la partie R & D en génomique, Pierre Leca sur la R & D en calcul et traitement de données.

LE SÉQUENÇAGE DE L'ADN

Le projet qu'ils nous présentent, dans le cadre de l'infrastructure nationale France Génomique, traite du séquençage de l'ADN, et vise à fournir des capacités de stockage et de traitement de données qui permettent de nombreuses études et analyses. Quelques notions s'imposent: le séquençage consiste à déchiffrer le code génétique. Pour le génome humain, cela représente 3 milliards de bases réparties sur 23 chromosomes. Le génome du blé, par exemple, renferme 17 milliards de bases, alors qu'une bactérie n'en contient que quelques millions. On prend alors vite la mesure des volumes et traitements nécessaires. Depuis plus

de 30 ans, les scientifiques ont développé diverses méthodes de séquençage de l'ADN, dont la plus connue est la méthode de Sanger. Dans les années 90, un programme international a été lancé dans le but de séquencer l'intégralité du génome humain. La tâche a été achevée en 2003 seulement, mais c'était une première mondiale. Plusieurs années de travail donc, mais aussi plus de 3 milliards de dollars investis pour réaliser cet exploit. **Aujourd'hui, un génome humain se lit en quelques heures pour quelques milliers de dollars!** Cette rupture technologique, scientifique et financière ouvre la voie à des applications de séquençage massif, dont les besoins en moyens de stockage et de calcul sont un challenge à relever.

Le Genoscope, qui fait maintenant partie de l'IG, avait largement participé au vaste projet du premier séquençage, étant en charge de l'intégralité du chromosome 14. Le travail, soulignent les deux collègues, était « fastidieux et



coûteux ». Depuis, le coût et la vitesse du séquençage ont évolué de façon spectaculaire. À la fin des années 90, on pouvait lire quelques millions de bases par jour. Aujourd'hui, dans un centre comme l'IG, on peut lire plusieurs centaines de milliards de bases (Gbases) par jour.

TRAITER LES VOLUMES DE DONNÉES ISSUES DES SÉQUENCEURS

En termes de volumes de données, l'équivalence est simple : une base, c'est un octet. **L'infrastructure France Génomique** (dont l'Institut de génomique du CEA, deuxième centre de séquençage européen, est le principal partenaire), **peut produire de 50 à 75 téraoctets de données par mois, soit plusieurs centaines de téraoctets par an.**

Pour Pierre Le Ber et Pierre Leca, si les données à traiter sont structurées par les appareils qui les produisent, la bio-informatique rejoint le « Big Data » par la volumétrie et la vitesse de production et de traitement des données. En outre, les données sont caractérisées par un modèle d'erreur propre à la technologie, et un taux d'erreur variable dû aux biais de la biologie moléculaire. C'est là que se situe le point critique de leurs travaux. Afin de traiter cette quantité colossale de données issues des nouvelles générations de séquenceurs d'ADN, une infrastructure centralisée, capable de mettre en œuvre des moyens de stockage multi-téraoctets accessibles avec un haut niveau de performance (débit, latence) par plusieurs milliers de cœurs de calcul, s'imposait. Cette e-infrastructure de France Génomique a donc été mise en place au CCRT (Centre de Calcul Recherche et Technologie) au sein du Très Grand centre de Calcul du CEA (TGCC). Elle a été financée dans le cadre du programme national d'investissements d'avenir, mis en œuvre par le Commissariat Général à l'Investissement.

Il y a 10 ans, pour un projet de génomique, on dépensait 10 % de la somme totale dans l'informatique. Dans le cas de France Génomique, pour 1 euro dépensé dans le laboratoire, 1 euro est investi en informatique et bio-informatique. Le besoin de stockage est expliqué par les volumes de données générés mais aussi par la nécessité, pour certains projets de grande ampleur, de les conserver accessibles en ligne sur de longues durées, afin de pouvoir comparer les nouvelles données aux plus anciennes. De plus, la communauté est très active, et les techniques de séquençage et d'analyse évoluent très rapidement. Les données sont récupérées par le CCRT et les traitements primaires sont réalisés au travers de pipelines d'analyse pratiquement automatisés. Les analyses secondaires sont ensuite effectuées sur l'e-infrastructure par les équipes projet.

DES APPLICATIONS TRÈS VARIÉES

Dans le cas du génome humain, en comparant deux génomes de deux personnes différentes, on observe quelques millions de différences, dont une minorité a des conséquences biologiques. L'analyse primaire consiste à détecter ces différences, l'analyse secondaire vise à les interpréter, à l'aide d'études statistiques, fonctionnelles, médicales... Prenons l'exemple du cancer : séquencer le génome de cellules tumorales par comparaison au génome « normal » du patient permet d'en déterminer le profil génétique, et dans certains cas de prédire quel traitement a le plus de chances d'être efficace pour un patient donné. C'est de la mé-

decine personnalisée, une grande avancée apportée par la génomique et qui s'appuie sur le Big Data.

Un autre exemple est celui du projet TARA Océans, un vaste projet d'étude de la biodiversité planctonique marine qui a amené dès 2009 les équipes à sillonner les océans pendant plus de 3 ans afin de récolter des échantillons de tous les coins du globe. La finalité : séquencer et analyser les gènes des organismes présents à divers endroits, en parallèle avec d'autres approches telles l'imagerie à très haut débit. Il s'agit le plus souvent d'organismes que l'on n'est pas capable d'isoler mais qu'il est possible de séquencer en tant qu'écosystèmes planctoniques complets (approche dite « métagénomique »). À date, environ le quart des prélèvements effectués ont été séquencés. La découverte impressionnante réside dans le fait que la très grande majorité des organismes séquencés nous étaient jusqu'alors inconnus. Ce projet révèle l'immensité de la biodiversité des océans. C'est un inventaire, en quelque sorte, une photo prise au temps-T, qui nous permet de mieux connaître les étendues aquatiques. Au-delà de l'intérêt purement informatif, **TARA Océans permettra par exemple d'étudier les effets du réchauffement climatique sur les écosystèmes marins** : renouvelé dans quelques années, le prélèvement de nouveaux échantillons permettra de comparer les organismes présents. À date, on réalise un ca-

“le génome humain représente **3 milliards** de bases réparties sur **23 chromosomes.**”

talogue de référence qui servira à observer l'évolution de la biodiversité. Ce projet, unique par son ampleur, est un des plus vastes dans lesquels s'investit aujourd'hui le Genoscope au sein de l'IG. Financé par les grands organismes de recherche publique, mais aussi par les Investissements d'Avenir (dont France Génomique) ou des mécènes privés tels Agnès B., Tara Océan prévoit d'ouvrir ses données. C'est aussi ce type de projets qui a rendu inéluctable le changement d'échelle pour les capacités de stockage et de traitement des données.

France Génomique continue à investir dans ses infrastructures. Ne serait-ce que pour mettre à jour la partie hardware, **la structure doit investir 2 à 3 millions d'euros tous les 3 ou 4 ans.** Le besoin se fait grandissant d'autant plus que de nouvelles applications se font jour : la médecine personnalisée notamment est un domaine dans lequel l'expertise que France Génomique aura accumulée dans le domaine des Big Data sera précieuse, avec néanmoins de nouvelles problématiques comme la question de la protection et la sécurité des données. Recherche et diagnostic seront concernés, la volonté étant de valider les protocoles de séquençage à très haut débit, et plus encore de gestion des données pour la médecine personnalisée.



UN PROJET DANS LES TRANSPORTS : LA SÉCURITÉ AÉRIENNE

SAFETYLINE - Entretien avec Pierre JOUNIAUX, co-fondateur, SAFETYLINE

L'aéronautique est un secteur reconnu pour ses capacités à innover et son exigence sécuritaire. Pour Pierre Jouniaux, ingénieur aéronautique de formation, le Big Data représente une réelle opportunité.

Au début de sa carrière, Pierre Jouniaux a intégré le Ministère des Transports. Il a alors passé 12 ans au bureau des enquêtes accidents, à traiter les données des boîtes noires pour comprendre les causes des accidents survenus. Il devient ensuite pilote de ligne au Vietnam, et en parallèle il participe à la mise en place d'un système de traitement des données.

C'est là qu'il constate que les données enregistrées en vol sont sous-exploitées. En effet, les compagnies sont tenues de récupérer et lire les données issues des boîtes noires et les copies dupliquées, mais elles ne mènent les analyses que vol par vol. En 2008, bien avant qu'on commence à parler de Big Data, Pierre Jouniaux fonde donc SafetyLine. L'idée à l'origine du projet est que cette masse de données récupérées par les compagnies aériennes représente de nouveaux champs d'analyse si les données sont confrontées, sur tous les vols, en temps réel. Développée avec l'aide d'un laboratoire de recherche — le laboratoire de statistiques théoriques et appliquées de Paris — la solution a donc été construite sur un socle Big Data. Comme l'explique Pierre Jouniaux, « un vol,

c'est dynamique ». Pour expliquer un accident, il ne suffit pas d'observer que quelques paramètres. C'est la combinaison de multiples facteurs complexes qui peut engendrer une anomalie. Prenons l'exemple concret des sorties de pistes, qui ne sont pas les accidents les plus dangereux mais sont les plus fréquents. Lorsqu'il travaillait en tant qu'enquêteur, Pierre Jouniaux a étudié le cas de la sortie de piste de Toronto et confirmé qu'un tel événement est dû à une multitude de données.

Le projet de Pierre Jouniaux consiste à déterminer le principe des marges et observer comment elles sont absorbées. Pour ce faire, il est donc nécessaire de faire un apprentissage sur ces marges, ces limites. L'idée est alors de connaître les valeurs des données pour une si-

tuation dite normale, la marge d'écart acceptée, et la façon dont la compagnie utilise ces marges. Sans a priori, on regarde toutes les données, essentiellement récupérées par les capteurs disposés sur les avions, mais aussi les données météorologiques par exemple. Un avion enregistre environ 1 000 paramètres par seconde, donc sur un vol entier les volumes atteignent des niveaux nécessitant une analyse experte. Pour Pierre Jouniaux, l'avantage d'avoir été pilote, tout en ayant les connaissances en traitement de la donnée, lui permet de savoir quelles informations sont pertinentes et dans quel cas de figure.

Empiriquement, on peut considérer que 60 % de marge sont généralement pris en compte pour englober tous les paramètres. Dans la grande majorité des cas, c'est largement suffisant, mais sur des vols atypiques ou pour certains aéroports par exemple, la marge de sécurité restante est faible.

Revenons à l'exemple de l'atterrissage long. L'algorithme de SafetyLine met en évidence ce qui différencie les vols normaux des vols atypiques. **Profil du vol, configurations diverses, type de pilotage...L'outil prend toutes les données de tous les vols disponibles** et en fait un apprentissage, afin de construire un modèle. On peut alors faire une prédiction, en pourcentage, des risques sur un modèle donné. Safetyline

“Un avion enregistre environ 1 000 paramètres par seconde.”

observe les signaux faibles et les paramètres inattendus. Par exemple, pour tel type d'avion, sur tel type de trajet, dans telle compagnie, la marge est davantage utilisée, et le vol est donc plus risqué. On donne ces indicateurs aux compagnies aériennes qui ont choisi le service de SafetyLine. Chaque modèle est individuel. Chaque compagnie est propriétaire de ses données de vol, qui sont confidentielles. SafetyLine s'engage à les protéger, à ne pas les héberger sur le cloud, à conserver les données isolées et à assurer un maximum de sécurité. L'enjeu est crucial tellement les données sont sensibles. Les données sont donc stockées, et les calculs effectués en in-memory. Les temps de calcul sont relativement importants — plusieurs heures — mais pour Pierre Jouniaux, ce n'est pas un problème du fait qu'il fonctionne en asynchrone. Pour des analyses prédictives, c'est largement suffisant.

Pierre Jouniaux travaille par exemple avec Air France, compagnie avec laquelle il a mis en place une démonstration grandeur nature de la sécurité des vols. Chaque modèle construit est unique, avec son propre risque, ses marges et ses méthodes d'analyse propre.

Air France, c'est 75 000 vols moyens courriers en quelques mois uniquement. C'est le spectre de la collaboration entre SafetyLine et la compagnie, et c'est déjà largement suffisant pour en tirer des conclusions pertinentes. **L'outil détecte les signaux faibles, avant même les accidents ou dérives qui pourraient survenir.** L'idée est de savoir comment la compagnie utilise ses marges. Lorsqu'on constate qu'elles sont fortement rognées, on peut estimer l'augmentation du risque engendré. Identifier la source d'un problème de façon individuelle est relativement aisé, la question des marges est plus complexe. Safetyline observe donc la corrélation des paramètres qui peut augmenter le risque, sans que l'œil humain puisse le percevoir ni qu'aucun problème ne survienne. C'est aussi en construisant ces modèles que l'on peut faire du prédictif et du préventif. Et chaque vol, chaque aéroport, chaque avion a ses par-

ticularités. Par exemple, un A318 est un petit modèle d'avion, moins sensible au problème de distance d'atterrissage qu'un autre type d'avion, mais qui nécessite un pilotage adapté à l'approche.

L'avantage des données de vol est qu'elles sont très formatées. Airbus, Boeing... tous les constructeurs ont standardisé au niveau international les capteurs et les unités mesurées.

Au-delà de la notion de risque, Safetyline réfléchit à d'autres applications du traitement de la donnée de vol. **L'optimisation de la consommation de carburant par exemple est une piste particulièrement prometteuse.** En modélisant les caractéristiques du vol, les équipes SafetyLine peuvent optimiser son profil, particulièrement au moment de la montée, le point critique de la consommation. Concrètement, le principe consiste à communiquer des consignes de vol qui permettront d'améliorer cette consommation. Les principaux paramètres influant sur la consommation sont la température, la masse de l'avion et son profil aérodynamique. Chaque profil entraîne des consignes de vol différentes.

Pour l'instant, cela n'a été étudié que pour un seul type d'avion, mais ce projet est voué à se dupliquer. Dans le jargon, on parle de niveau de consommation 3R : c'est là que l'avion consomme le moins, et c'est ce que veulent atteindre les compagnies aériennes autant que possible. L'enjeu financier est considérable, la solution attire donc déjà car elle permet des gains substantiels. Sur un vol unique, l'économie n'est pas notable. Elle est de quelques dizaines de kg par vol. Mais lorsqu'on considère qu'un avion peut effectuer jusqu'à 5 vols par jour, et en mettant cette donnée en relief avec la taille de la flotte de compagnies telles Air France, on perçoit vite l'intérêt.

D'autres projets sont en développement, comme un travail avec les Aéroports de Paris qui vise à détecter les problèmes de glissance de la piste, ou encore des solutions de maintenance prédictive pour anticiper les pannes.



“Air France, c'est 75 000 vols moyens courriers en quelques mois.”



UN PROJET EN CLIMATOLOGIE : L'ÉVOLUTION CLIMATIQUE AU NIVEAU MONDIAL

Projet DADA - Entretien avec Alexis Hannart, Climate Scientist au CNRS

Le projet DADA est une collaboration scientifique internationale en climatologie, structurée autour de l'IPSL (Institut Pierre Simon Laplace, Île-de-France) et financée par l'ANR (Agence Nationale de la Recherche), impliquant des chercheurs français, européens et américains. Alexis Hannart, climatologue au CNRS, en est le responsable. Scientifique de formation (Polytechnique, ENSAE), Alexis a rejoint la recherche académique dans les années 2000 après une expérience de plusieurs années en R & D dans le secteur privé, sur des sujets constituant alors les prémices du Big Data.



DADA propose une application environnementale du Big Data. Son ambition est d'exploiter les flux d'observations climatiques, de plus en plus nombreuses et diverses, dans le but in fine de pouvoir statuer en temps réel sur l'existence d'un lien de causalité éventuel entre les phénomènes climatiques survenus, de quelque nature et échelle de temps qu'ils soient, et les émissions de gaz à effets de serre – ou autres émissions d'origine humaine (ozone, aérosols...).

Pour cela, on confronte l'observation aussi complète et exhaustive que possible d'un événement ou d'une série d'événements, à des simulations numériques : d'une part du monde « factuel » – c'est-à-dire du système climatique tel qu'on le connaît – et d'autre part du monde « contrefactuel » – c'est-à-dire du même système mais tel qu'il serait si les gaz à effets de serre n'avaient jamais été émis. Selon que l'événement apparaîtra – ou non – comme plus vraisemblable dans l'avatar numérique du monde factuel que dans son pendant contrefactuel, on pourra invoquer un lien de causalité – ou son absence.

Si le principe général est simple, la mise en œuvre est délicate : elle exige en effet de manipuler des flux de données en très grand nombre et d'origines diverses (stations météo, satellites, réseaux de capteurs divers...), des modèles numériques de très grande dimension, un interfaçage complexe entre modèles et observations, enfin des algorithmes ad hoc permettant de convertir « à la volée » à partir de cet interfaçage un très grand volume de données brutes en une série de scores synthétiques de vraisemblance et, in fine, de causalité.

Cette chaîne de calcul d'une haute technicité est fortement innovante. Même si les études de causalité sont déjà nombreuses en science du climat, la nouveauté proposée par DADA est d'abord de se placer dans la perspective qui caractérise le Big Data : celle des grands volumes de données, de la rapidité du traitement « temps réel » et des problèmes spécifiques qu'ils engendrent. **Cette approche est originale en détection et attribution climatique par comparaison aux travaux antérieurs, qui se placent plus volontiers dans un « temps différé »** – avec des délais d'analyse de l'ordre de quelques années – et en suivant une approche plus artisanale qu'industrielle qui ne répond pas entièrement aux besoins futurs.



Mais surtout, la principale nouveauté est l'idée, pour mettre en œuvre cette approche Big Data, de capitaliser au maximum sur les procédures et les infrastructures de calcul d'interfaçage modèle-observations déjà existants et entièrement opérationnels dans les services météorologiques. En effet, **la prévision du temps repose sur de gros modèles numériques de l'atmosphère qui sont réinitialisés toutes les 6h sur la base d'un flux de nouvelles observations entrantes.** On appelle cela l'assimilation de donnée et c'est de ce terme que provient l'acronyme DADA (Data Assimilation for Detection and Attribution). L'équipe DADA s'inspire de ce cadre tout en l'enrichissant et en l'adaptant.

In fine, ce dispositif technologique innovant vise à répondre à des besoins concrets d'attribution causale – en particulier liés à l'indemnisation des dommages futurs attendus dans le changement climatique en cours. Il comporte de multiples défis théoriques comme pratiques ; la preuve de sa pertinence reste donc largement à établir avant de passer à une mise en œuvre opérationnelle. C'est justement cette exploration préliminaire qui est le but de DADA. Comme le souligne d'ailleurs pour conclure Alexis, « nous n'en sommes encore qu'au stade de l'idée générale, l'avenir dira si c'en est une bonne ou pas, c'est de la recherche... Mais d'ici là, on va essayer de suivre au plus près les préceptes du Dadaïsme, ce mouvement artistique

né il y a un siècle et auquel notre acronyme est bien sûr un clin d'œil : innover sans aucune censure, en provoquant des hybridations et des rapprochements inattendus, et surtout, en s'amusant ! Maintenir cet esprit ludique sera facile, car le Big Data est un terrain de jeu fantastique et inépuisable. »

Alexis estime que « dans le cadre du Big Data, les travaux d'un climatologue peuvent intéresser d'autres profils, comme un responsable marketing par exemple ». À bon entendre...

“Dans le cadre du Big Data, les travaux d'un climatologue peuvent intéresser d'autres profils, comme un responsable marketing par exemple .”



Le Guide Big Data répertorie pour vous les acteurs référents et symboliques du marché afin de vous permettre de faire votre propre sélection.

Dans cette partie, retrouvez les acteurs de l'écosystème Big Data qui font l'actualité en France et à l'étranger. Ils ont souhaité vous faire part de leur vision, de leurs principales évolutions technologiques ou de leurs plus belles réalisations.

Au cours de ces quelques pages, apprenez à les connaître ou à les redécouvrir, car ce sont eux qui participent à la croissante maturité du Big Data en France. Que vous soyez à la recherche de partenaires et fournisseurs, ou tout simplement passionné du sujet et attentif aux évolutions de ses protagonistes, cette partie vous intéressera.

Dans un souci d'organisation et de clarté, nous vous proposons diverses catégories, représentées par des labels, qui vous permettront d'identifier immédiatement les domaines de compétences clés de chaque entreprise et pouvant donc répondre à vos besoins. Ces labels sont répartis en 6 catégories



Equipementiers Hardware Editeurs Software Editeurs BI
 Editeurs Dataviz Consultants SSII Intégrateurs Institutionnels



Labels
Annonceurs

Equipementier
Hardware

Editeur
Software

Editeur
BI

Editeur
Dataviz

Conseil / SSII
Intégrateur

Institutionnel

Pages

ACTUATE		.	.	.			44
A.I.D				.	.		46
AMAZON WEB SERVICES		.	.				48
BIME ANALYTICS				.			50
BITTLE		.	.				52
BLUESTONE					.		54
CAPGEMINI					.		56
CENTRALE FORMATION						.	58
CLOUDERA		.					60
CLUB DECISION DSI						.	62
CSC					.		64
DATASTAX		.			.		66
EDIS CONSULTING					.		68
EKIMETRICS			.	.	.		70
EXALEAD DASSAULT SYSTEMES		.		.			72
GFI INFORMATIQUE					.		74
HP	.		.				76
IBM		.	.	.			78
KEYRUS					.		80
MICROPOLE					.		82
NEO TECHNOLOGY		.	.				84
PENTAHO		.	.	.			86
PIVOTAL		.	.				88
QLIK			.	.			90
SAP		.	.				92
SAS		.	.	.			94
SCINETIK			.		.		96
SEMDEE			.				98
SENTELIS					.		100
SINEQUA		.	.	.			102
SPLUNK		.					104
SYNCSORT		.					106
TABLEAU SOFTWARE		.	.	.			108
TALEND		.	.				110
TERADATA	.	.					112
TIBCO JASPER SOFT			.				114
WISEO				.	.		116

Fournisseurs

BIGDATA



COORDONNÉES ACTUATE FRANCE

57, rue de Châteaudun
75009 Paris

Tél. **+33 (0)1 76 21 56 50**
Mail fr_contact@actuate.com

CONTACT

**Sarah
LALLAM**

Responsable marketing

Actuate aide les entreprises à exploiter les sources de données hétérogènes dont le **Big Data** pour la prise de décision et la diffusion d'informations personnalisées en proposant des rapports, des analyses et des tableaux de bord **interactifs**, accessibles à un grand nombre d'utilisateurs depuis une interface Web ou mobile.



BIRT iHub, la plateforme modulaire innovante d'Actuate, permet de développer et déployer rapidement des applications de Business Intelligence sur mesure. Elle exploite la technologie **Open Source** BIRT, créée par Actuate au sein de la Fondation Eclipse et adoptée par plus de trois millions de développeurs dans le monde.

Les applications construites avec BIRT iHub sont idéales pour la diffusion de masse d'informations personnalisées.



Restitution interactive,
Intuitive et
hautement visuelle



Vision intégrée
de données
Multi-source



Sécurité sur de grands
volumes d'utilisateurs
et de supports



Montée en charge
prouvée pour des
millions d'utilisateurs



Actuate : L'Éditeur à l'origine de BIRT

Analyses, tableaux de bords et rapports personnalisés pour plus de 200 millions de personnes.



Plateformes décisionnelle BIRT IHUB

La plateforme décisionnelle BIRT IHUB d'Actuate offre des visualisations de données interactives, intuitives et personnalisées à travers des outils analytiques et des tableaux de bord pouvant être facilement déployés auprès de millions d'utilisateurs à travers le web et le mobile.

Les applications construites avec BIRT IHUB sont idéales pour la diffusion de masse d'informations personnalisées. Actuate aide ainsi les entreprises à tirer pleinement parti du Big Data et de l'essor des appareils tactiles.

Développement

BIRT Designer Pro crée des contenus basés sur la technologie BIRT Open Source

Déploiement

BIRT IHUB gère jusqu'à des millions d'utilisateurs via des clusters de servers à haute disponibilité
BIRT Mobile visualise des contenus BIRT sur des supports mobiles tels que smartphones et tablettes

Visualisations

BIRT Viewer permet la visualisation de rapports BIRT statiques

BIRT Interactive Viewer permet aux utilisateurs de visualiser et de personnaliser leurs rapports en toute liberté

BIRT Studio donne aux utilisateurs la possibilité de développer leurs propres rapports web Ad Hoc, les données restant sous le contrôle de la DSI.

BIRT Analytics

BIRT Analytics est la solution d'analyse exploratoire et prédictive des données qui permet de réaliser des études à partir de n'importe quelles sources de données dont le Big Data sans besoin de connaissance techniques et statistiques spécifiques.

A la main des utilisateurs métier, BIRT Analytics offre également la possibilité de concevoir des campagnes et d'en prédire les retombées.



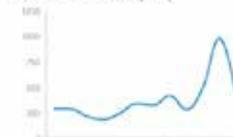
Les composants de la plateforme BIRT HUB



Synthèse Des Ventes - 2010-2012

Ventes par commerciaux	
George Chausoff	668 379
Faouf March	584 394
Umar Boufdir	368 490
Abdy Fider	362 369
Silvia Hernandez	355 475
Phan Van Thang	488 213
Hanno Spahn	457 519
Harbin Gansel	387 477
Bala Anelli	386 442
Leticia Thompson	347 538

Evolution CA sur 2012 (en €)



Evolution CA sur 3 ans





COORDONNÉES

A.I.D.

4 rue Henri Le Sidaner
78000 Versailles

Tél **01 39 23 93 00**

Mail info@aid.fr

CONTACT

**Arnaud
CONTIVAL**

Président Directeur Général

L'accélération du digital modifie profondément la façon de communiquer des entreprises avec leurs consommateurs, devenus multi-canaux, interactifs, informés et partageurs. Les entreprises qui relèvent ces défis et accompagnent les consommateurs dans ces évolutions structurelles sont celles qui **exploitent et valorisent leur capital « data »** pour parler à leurs clients **au bon moment, avec un message pertinent sur le bon canal**. Face au phénomène de volumétrie croissante des données, le leitmotiv d'A.I.D. est d'accroître en permanence l'expertise « data » **avec des méthodes innovantes d'analyse, de traitement et d'exploitation de la donnée** afin de vous permettre d'atteindre vos objectifs.

A.I.D. propose une **chaîne de valeur complète** autour de l'analyse, du traitement, de l'exploitation et de la transformation des données avec notamment :

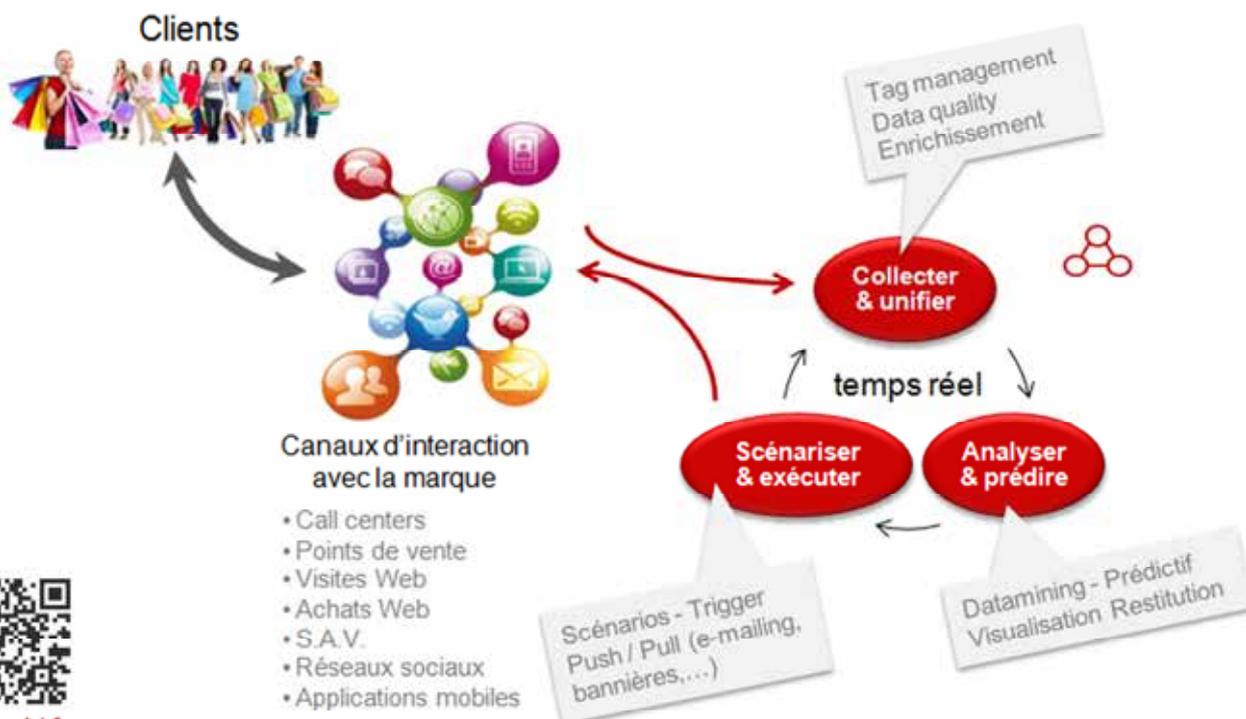
- La création, l'hébergement et l'exploitation de systèmes d'information marketing, au service d'une communication à la fois plus personnalisée et industrialisée.
- La mesure et l'optimisation des actions marketing, pour maximiser l'impact de vos opérations marketing, détecter les clients en partance ou gérer la pression commerciale par la pertinence.
- La data visualisation pour vous aidez à lire, comprendre, et analyser vos données.
- La qualité des données pour garantir une qualité optimale des données multi-source et une bonne gouvernance des données.
- L'analyse des parcours clients cross-canal afin de prédire et d'influencer la prochaine interaction client.

Nos clients tels que AUCHAN, BOUYGUES TELECOM, CRÉDIT MUTUEL ARKEA, DARTY, EDF, MOBIVIA GROUPE (NORAUTO, MIDAS...), PMU, PUY DU FOU, RENAULT, SEPHORA, SFR, SIMPLY MARKET, SYSTEME U, TRUFFAUT, **nous recommandent avec une note moyenne supérieure à 8/10**

(Enquête réalisée 2 fois par an sur la base du Net Promoter Score).

Depuis plus de 40 ans, A.I.D. innove continuellement dans le traitement et l'analyse des données. Notre R&D est la garantie que nous apportons à nos clients les solutions les plus performantes à leurs problématiques BtoB et BtoC autour du Datamining, de la Dataquality, du CRM, du CIM et de l'Hébergement.

Customer Interaction Management, la solution en mode SaaS de valorisation des parcours clients cross-canal



Customer Interaction Management

Visualisation des parcours clients cross-canal

(achats magasins, site e-commerce, call center, réseaux sociaux)

Porteurs du projet

Au sein d'un distributeur et e-commerçant français, le projet a été porté par la Direction Marketing en la personne du responsable e-CRM. Ce projet transversal a eu le soutien de l'équipe informatique et relation client (Call center).

Objectifs

Collecter toutes les interactions clients sur tous les canaux et donner une information actionnable aux collaborateurs de l'entreprise pour les aider dans la prise de décision.

- Fournir au call center lors d'un appel entrant l'ensemble des interactions du client et leur motif avec la marque (call center, navigation web, réseau sociaux).
- Analyser les comportements clients (via les interactions) afin de déterminer les situations à risques et prendre les mesures adéquates.

Présentation

Le projet s'appuie sur la solution C.I.M. (Customer Interaction Management) conçue par A.I.D. et utilisable en mode SaaS.

Les données traitées sont de sources multiples :

- E-CRM (signalétique client, activités – achats, retours d'articles, retours de campagnes marketing - ouvertures, clics, ...)
- Call-center (interactions, appels entrants, appels sortants, suivi des tickets ...)
- Site e-commerce (taggage du site avec la technologie A.I.D.)
- Réseaux sociaux (veille et capture sur l'e-réputation de la marque)

Les intervenants sur le projet sont pluridisciplinaires : un data scientist, un architecte Big Data (Hadoop, Hbase), des développeurs (java), un ergonomiste et un business analyste.

Le projet a été réalisé en mode agile (scrum master, product owner...).

Les processus mis en place suivent la logique : collecte, stockage, traitement (constitution des référentiels), analyse, modélisation prédictive et restitution.

Aspect novateur

La solution C.I.M. amène l'ensemble des innovations :

- collecte de données hétérogènes, en quelques clics
- stockage et analyse en temps réel dans un environnement Big Data et création à la volée de référentiels,
- interface de visualisation unique et novatrice conçue et utilisable directement par le métier,
- constitution d'un modèle statistique permettant de prédire les futures interactions clients (attrition, insatisfaction, abandon de panier...).

Valorisation observée de la donnée

La donnée collectée est stockée, transformée, référencée, analysée et modélisée afin de permettre la prise de décision en temps utile selon le canal : soit en contact direct avec le client, soit en préventif pour une interaction à venir.

Etapes à venir et démarche Big Data dans l'entreprise

Le projet a démontré l'intérêt de la démarche de collecte et de valorisation des interactions entre magasin, site e-commerce, call-center et réseaux sociaux. Les enjeux commerciaux et le ROI ont été clairement validés et le déploiement de la solution est en cours.

Customer Interaction Management

Comprendre, anticiper et actionner en temps réel les parcours cross canal de vos clients



Add intelligence to Data

Copyright, tous droits réservés, A.I.D. 2014





COORDONNÉES

AMAZON WEB SERVICES INC

410 Terry Avenue North, Seattle, WA
98109

Tél +33 (0)1 46 17 10 08

Mail lavocats@amazon.fr

CONTACT

**Stéphanie
LAVOCAT**

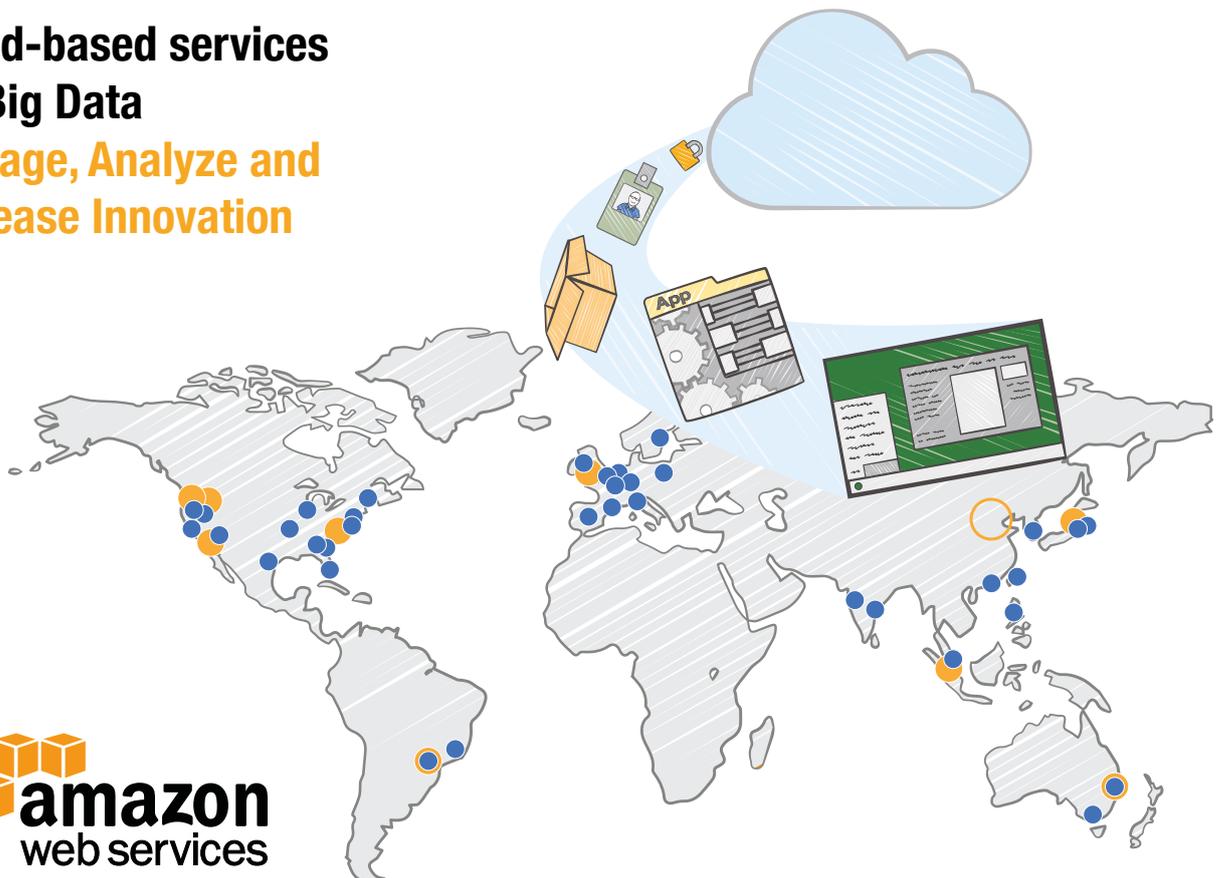
Marketing Manager, France

Amazon Web Services (AWS) est une compagnie d'Amazon.com, créée en 2006, qui offre à ses clients un accès à des services d'infrastructure dans le cloud afin de leur permettre de créer des applications sophistiquées et extensibles.

Aujourd'hui, AWS offre plus de quarante services à ses centaines de milliers de clients, grandes et petites entreprises, dans 190 pays, tout en étant un réel moteur pour l'innovation dans de nombreux domaines comme les médias, la finance, le e-commerce, ...

Amazon Web Services propose notamment aux scientifiques, développeurs et aux autres professionnels ayant des problématiques Big Data, de tirer parti du cloud pour effectuer leurs analyses et relever les défis liés à l'augmentation du volume, de la diversité et de la vitesse des informations numériques. En effet, Amazon Web Services fournit un catalogue complet de services cloud afin d'aider ses clients à gérer le Big Data pour obtenir un avantage concurrentiel et augmenter leur rythme d'innovations. Plus de détails sur les services Big Data du cloud AWS : <http://aws.amazon.com/fr/big-data/>

Cloud-based services for Big Data Manage, Analyze and Increase Innovation



Pensez Big Data, pensez Cloud

L'idée nouvelle du **Big Data** est de rassembler la totalité des données d'usages historiques disponibles (logs, données patrimoniales, réseaux sociaux...) sans pré-formatage ou pré-filtrage, puis de traiter ces données a posteriori sur des historiques de plusieurs mois voire plusieurs années.

Se lancer dans un projet de Big Data dans le cloud apporte une valeur unique par rapport à des infrastructures sur site, notamment dans les étapes de stockage et d'analyse:

- **Stockage élastique** : le principe est d'avoir une capacité illimitée dans laquelle les clients ne paient que pour le stockage réellement utilisé, sans engagement dans le temps.
- **Capacité serveurs à la demande** : les clients ont accès – à la demande et en quelques minutes – à des serveurs virtualisés pour effectuer des traitements sans limite de taille. La facturation des ressources est à l'heure d'utilisation.

C'est dans ce domaine que les technologies élastiques proposées par le cloud Amazon Web Services prennent tout leur sens. **L'architecture informatique élastique (stockage, serveurs...) proposée par AWS s'adapte aux projets Big Data ponctuels ou récurrents des entreprises de toutes tailles.**

Rendez-vous sur: aws.amazon.com/big-data pour découvrir comment les services AWS vous aident à gérer vos projets Big Data en réduisant les coûts, pour obtenir un avantage concurrentiel et augmenter le rythme de vos innovations.

Cas d'usage AWS : Botify

Botify est une start-up française qui utilise l'analyse Big Data pour faire progresser l'audience SEO et le chiffre d'affaires de leurs clients tels que Dailymotion, Viadeo, Hotels.com, BlaBlaCar, ELLE et Venere.com. En effet, **Botify gère aujourd'hui l'analyse de plus d'une centaine de millions de pages** ce qui impose à la start-up des services afin de mieux d'adapter aux besoins de ses clients en leur permettant de récupérer les informations indispensables en termes d'optimisation du référencement sur les moteurs de recherche. Lorsque Botify 'crawl' un site de 1 million de pages, ils accumulent plus de 100 Go de données. **Botify utilise donc la plateforme AWS pour sa capacité de calcul haute performance mais aussi pour son faible coût**, sa flexibilité et sa simplicité, qui permet à l'entreprise de gagner du temps en phase de lancement de projets et pour la gestion des requêtes. Plus d'information sur Botify : www.botify.com

The screenshot shows the Botify website with a dark header containing the logo and navigation links. Below the header is a white banner with the text "A NEW ERA FOR SEO" and "2 great tools for Search Engine Optimizations. Thousands of fantastic clients worldwide!". Underneath the banner are logos for clients: ELLE, BlaBlaCar, viadeo, venere.com, and hotels.com. The main content area is divided into two colored sections: a teal section for "Botify Log Analyzer" and a red section for "Botify Analytics".

1 Botify Log Analyzer
How well does Google know your site?
Optimize your crawl and organic visits performance
REQUEST A DEMO

Discover strategic links ignored by search engines
Monitor search engines' crawl errors
Eliminate toxic content discovered by search engines
Drive search bots towards quality pages
Increase the number of pages with organic traffic
Monitor your indexing progress across time
No analyst resources anymore

2 Botify Analytics
How well do you know your site?
Pinpoint issues with real room for improvement
START MY 30-DAY FREE TRIAL

Customize your dashboard with dash.3
Remove duplicate content
Report broken links
Adjust internal linking to promote key pages
Map related internal and external pages to other important
Query and export personalized data to take action
Analyze up to 5 million pages





COORDONNÉES BIME ANALYTICS

1 place Francis Ponge
34000 Montpellier

Tél +33 (0)4 67 41 60 35

Mail stephan.khattir@bimeanalytics.com

CONTACT

**Stephan
KHATTIR**

Directeur des Ventes EMEA

BIME est un service de Business Intelligence moderne. Né dans le Cloud et dimensionné pour vos Big Data, BIME a été pensé pour les entreprises devant mixer leurs données pour identifier leurs facteurs clés de réussites. BIME masque la complexité pour l'utilisateur métier : il ne nécessite aucune infrastructure complexe, ni serveurs, ni logiciel client. Un simple navigateur, uniquement.

BIME se connecte aux nouvelles sources de données en ligne telles que Salesforce, Twitter, Facebook, Youtube, Google Analytics ainsi qu'aux données installées dans l'entreprise - base de données propriétaires comme Oracle, Excel et plus. Réalisez vos analyses personnalisées et transposez les en graphiques visuels interactifs pour créer vos propres tableaux de bord. Partagez-les en temps réel via un simple lien URL : consultez et modifiez-les sur tablette ou mobile pendant vos déplacements.

BIME propose une série de connecteurs Big Data tels que Google BigQuery, Amazon Redshift, SAP HANA, HP Vertica et propose des solutions de Stockage Cloud extensibles pour analyser vos flux de données, sans limite. Depuis votre navigateur.

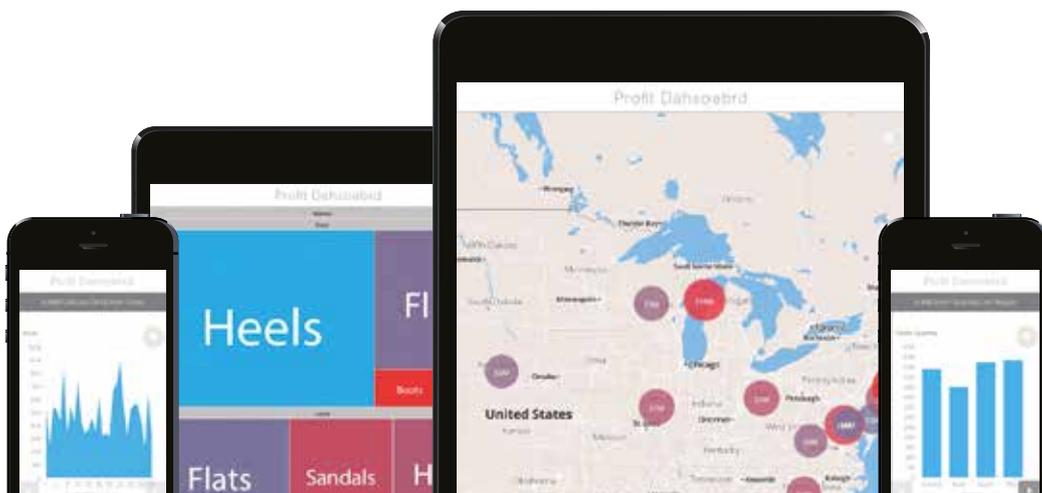
La nouvelle version BIME V6 marque une étape importante dans l'ère des applications de Business Intelligence. S'inspirant des meilleurs applications grand public, elle présente une interface utilisateur de nouvelle génération, fluide, rapide, claire et élégante. Toute la complexité habituellement associée aux outils BI disparaît au profit de l'efficacité et du plaisir de travailler sur vos jeux de données.

BIME V6 vous accompagne toute la journée pour répondre à vos questions métiers au fur et à mesure qu'elles se posent. Vous ne serez pas uniquement dans la consommation de tableaux de bord construits par vos analystes mais vous pourrez les interroger, posez vos propres questions, n'importe où, depuis votre tablette aussi facilement que lorsque vous créez une présentation. Utilisez l'intégralité du Cloud comme votre propre data warehouse, créez et partagez vos analyses à la qualité visuelle la plus moderne.

Make your company

SMARTER

ON WEB, MOBILE & TABLET



Porteur du projet

Adrien Charaix, Business Analyst / Contrôleur de gestion. adrien.charaix@weekendesk.fr

Objectifs

Weekendesk est une agence de voyage 100% en ligne, leader de la vente de courts séjours en Europe. Le site génère plus de 46M€ de Volume d'Affaires sur les 3 pays que sont la France, la Belgique et l'Espagne. Porté par le département Finance & Business Analytics, leur projet BigData s'articulait d'abord sur une logique de sortie du Reporting en silo, dû à l'hétérogénéité de leurs sources de données. Il s'agissait en priorité d'agréger des données financières et de les mixer avec celles du trafic de leurs sites web et du comportement de leurs consommateurs.

Présentation

a. Nature des données traitées : Weekendesk est un site web marchand dont les 25000 commandes par mois génèrent un important volume de données transactionnelles entre les clients et les hôtels. Weekendesk analyse aussi constamment l'évolution de son trafic et son taux de conversion clients en rapport avec son offre produit et/ou service de son catalogue. Avec plus de 2 millions de visites par mois en moyenne, Weekendesk utilise Google Analytics. Les données transactionnelles de réservation sont stockées dans des bases de données SQL et un ERP propriétaire en ligne. D'autres données sont disponibles dans le Google Drive de l'entreprise.

b. Ressources et méthodologie (compétences humaines, outils techniques, fonctionnement du traitement de la donnée) : Le Reporting était précédemment produit via Excel avec une multitude de tableaux croisés dynamiques associés et d'onglets, pour couvrir tous les indicateurs clés des différents départements métiers. Ce Reporting a rapidement atteint les limites techniques d'Excel (volumétrie croissante et problème de disparité des données), devenant difficilement exploitable. En sélectionnant BIME, Weekendesk peut désormais reproduire son schéma de Reporting initial avec comme amélioration majeure la fluidité de se connecter nativement à toutes les sources de données (SQL, Google Analytics et Drive) depuis un seul et même endroit, puis d'agréger les informations en quelques clics, pour enfin construire et partager des tableaux de bord de pilotage en ligne avec le top management et les différents départements.

c. Calendrier de déploiement : Le projet a démarré en avril 2014. La Direction analyse les tableaux de bord depuis août 2014. Weekendesk a fait appel à un intégrateur pour travailler en amont les connexions on premise aux bases de données SQL.

Aspect Novateur

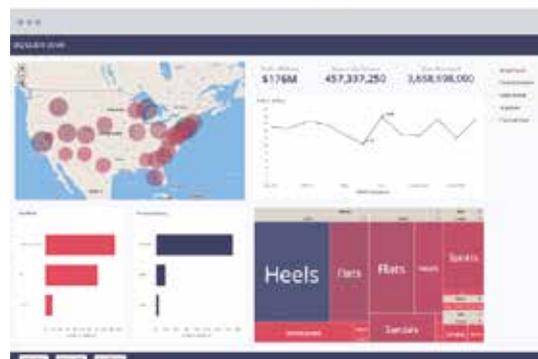
Weekendesk est une société 100% web, orientée cloud et mobilité. Chaque employé doit impérativement accéder à ses données métiers depuis n'importe où. La solution de Business Intelligence de BIME, 100% web correspond parfaitement à cet ADN. Grâce à la gestion des profils et droits utilisateurs, chaque Business Analyst rentabilise son temps de production des analyses. Un même tableau de bord, agrégeant plusieurs sources de données, peut en effet être partagé avec le Comité de Direction et les départements métiers. Le Comité accède aux informations globales de l'entreprise tandis que les métiers (vente, marketing) ne voient qu'un périmètre autorisé.

Valorisation observée de la donnée

Exploitation de la richesse du croisement des informations multi sources dans les tableaux de bord, à la volée, au même endroit et en ligne. Production simplifiée de graphiques visuels permettant une compréhension rapide des performances des canaux d'acquisition. Visibilité accrue des résultats des ventes web par région, par canal de vente (force de vente interne, service client, pur web), par source (ordinateur, tablette, smartphone), dans un contexte croissant de volume de données. Actualisation real time des indices de compétitivité des offres, disponibles en ligne pour les métiers, chacun étant capable d'exécuter des requêtes adhoc pour répondre à leur question.

Etapas à venir et démarche Big Data dans l'entreprise

Optimisation, augmentation et mix des collectes de données de ventes et de réservations pour renforcer la granularité de l'analyse au client près. Objectif : migrer les données CRM (dont Salesforce) & Marketing (dont les données d'Attribution comme C3 Metrics) dans Google BigQuery pour produire des tableaux de bord des profils clients et de leurs comportements d'achats, le plus rapidement possible. Bénéficier des technologies Cloud pour rester fidèle à l'ADN Web de Weekendesk et disposer d'une plateforme d'analyse puissante connectée en direct à toutes les sources de l'entreprise, pouvant grandir à l'infini.





COORDONNÉES

BITTLE

235 Rue Denis Papin
Domaine du Tourillon Bât C
13 857 AIX EN PROVENCE Cedex 3

Tél +33 (0)4 42 12 54 57

Mail contact@bittle-solutions.com

CONTACT

**Christophe
SUFFYS**

Président

BITTLE est le premier moteur décisionnel conçu et imaginé pour fonctionner en mode Cloud Computing. L'offre est proposée en mode web selon le principe du Software as a Service (SaaS), et dispose de capacités de traitement et de stockage quasi illimitées.

Fonctionnellement, BITTLE permet de suivre et piloter les indicateurs clés de l'ensemble des activités de la société, en analysant les données sources de manière simple, efficace et ludique. BITTLE s'adresse à tous types de structures qui recherchent une solution décisionnelle complète, packagée et prête à l'emploi permettant de créer et de partager facilement des tableaux de bord dans un environnement entièrement sécurisé.

Notre vision est d'apporter aux marchés visés une plateforme de Business Intelligence communautaire, compréhensible par tous (aucune compétence technique en informatique requise). Marketing, Ventes, RH, Finance, Achats, Informatique... quel que soit votre métier, Bittle vous permettra de suivre vos indicateurs et analyser vos données.

Toutes les organisations, quelque soit leur taille, privées ou publiques, sont de plus en plus demandeuses de solutions de Business Intelligence pour gérer l'ensemble de leurs activités. Les solutions présentes sur le marché ne répondent cependant pas à leurs besoins, pour différentes raisons, que ce soit la complexité de ces solutions, leur coût, ou encore les délais de mise en œuvre.

En proposant une solution BI MultiCloud, simple à mettre en place, simple à utiliser et à un prix attractif sous forme d'abonnement mensuel, BITTLE s'affranchit de l'ensemble de ces freins et s'impose comme La solution décisionnelle pour tous.

MY BIGDATA MY DECISION



Depuis sa création, Bittle a pour ligne de conduite de vouloir révolutionner la façon d'appréhender la Business Intelligence. Bittle décide maintenant d'aller encore plus loin en révolutionnant le marché du Big Data avec le même souci de démocratisation.

La plateforme BittleGoBig permet ainsi de travailler sur les données de masse qui peuvent être générées au travers de différents types d'événements.



La seule solution
intégrée de Big Data à la portée de tous!

BittleGoBig™

My BigData, My Decision



Depuis sa création, Bittle a pour ligne de conduite de vouloir révolutionner la façon d'appréhender la Business Intelligence. Bittle décide maintenant d'aller encore plus loin en révolutionnant le marché du Big Data avec le même souci de démocratisation.

Présentation

La plateforme BittleGoBig™ permet de travailler sur les données de masse qui peuvent être créées au travers de différents types d'événements. Elle permet, entre autres, à des utilisateurs de générer des données qui ne sont actuellement pas traitées informatiquement, comme par exemple des indicateurs d'utilisation, de satisfaction, de ressenti, ou encore des comportements d'utilisateurs sur des sites internet, ou sur une application mobile.

L'offre BittleGoBig™ se découpe en deux grandes parties. La première partie s'articule autour de la création d'un moteur OLAP basé sur des technologies Big Data, appelée «Bittle Foundation™». La seconde partie est la mise en place de connecteurs permettant la récupération de données simplement au travers d'une multitude de sources, appelée «Data Mixture™».

Socle technologique de BittleGoBig™, Bittle Foundation™ permet à Bittle de supporter les Big Data, sur deux aspects:

- Le stockage et l'interrogation des données permettant d'avoir des tableaux de bord toujours aussi réactifs quel que soit le volume traité.
- L'alimentation en masse, permettant le temps réel, lors de l'import de données.

Data Mixture™ est une véritable boîte à outils, permettant aux utilisateurs de collecter et mixer des données à la volée provenant de sources multiples, imaginée pour les utilisateurs ne possédant qu'un léger vernis technique.

Ce module est aussi un puissant middleware à destination du département informatique désireux de récupérer facilement les données produites par les applications existantes au sein du SI de l'entreprise via des APIs intégrées.

Aspect Novateur

Lauréat du Concours Mondial de l'Innovation 2030, Bittle innove dans sa proposition en intégrant différentes composantes permettant de couvrir tous les aspects du Big Data tout en le rendant plus accessible et plus simple.

Comment ? Les outils dédiés au Big Data nécessitent aujourd'hui des connaissances techniques, afin de comprendre les concepts utilisés et pouvoir mettre en place de tels systèmes.

Bittle ayant déjà réussi à simplifier le décisionnel standard, pour que des utilisateurs sans connaissance technique puissent facilement donner du sens à leurs données, veut mettre tout son savoir-faire et son expérience pour reproduire ce succès en démocratisant la mise en place de système Big Data.

De plus Bittle, ayant déjà développé un ETL accessible à des utilisateurs sans connaissance technique, va révolutionner la récupération et la génération de données de masse au travers de la mise en place de son module Data Mixture™ tout en gardant cette phase simple d'utilisation et extrêmement graphique.

Bittle deviendra donc par conséquent le réceptacle central des données de chaque entreprise, et le meilleur moyen de communiquer tant en interne qu'en externe ... le tout sans être informaticien.

Valorisation observée de la donnée

Les cas d'utilisation de BittleGoBig™ sont multiples. Ils vont du secteur des Facilities, pour une optimisation de la consommation énergétique des bâtiments en traitant une très forte volumétrie de données issues des compteurs mais aussi de sondes ou de capteurs, au secteur des télécommunications, pour analyser les événements sur les réseaux mobiles, en passant par le marketing et l'analyse comportementale, la géolocalisation, ou encore le web, pour suivre et analyser les actions effectuées par les internautes sur un site ou une application web.





COORDONNÉES

BLUESTONE

55, rue du Faubourg Montmartre
75009 Paris

Tél +33 1 53 25 02 10

Fax +33 1 53 25 02 11

Mail franck.bellenger@bluestone.fr

CONTACT

**Franck
BELLENGER**

Directeur Commercial

La Data Science au service de la performance

Créée en 1998 pour répondre à des enjeux d'exploitation avancée de la donnée, Bluestone a toujours considéré qu'une utilisation raisonnée des données permettait aux entreprises de :

- Créer la connaissance en analysant de façon appropriée les données et informations disponibles
 - Mobiliser la connaissance dans l'action, grâce à la mise en place d'applications « Data Driven ».
- La révolution industrielle que constitue le Big Data valide totalement ce point de vue. L'exploitation avancée de la donnée est, en effet, désormais perçue par la plupart des entreprises comme un levier d'innovation extrêmement fort à tous les niveaux.

Dans ce contexte, et avec une équipe de 130 Data Scientists, nous sommes aujourd'hui l'un des acteurs leaders dans ce domaine :

- Conseil : Accompagnement des stratégies de Data Innovation et de Management de l'Information au service des métiers
- Science : Expertise dans les domaines de la statistique, du datamining, du machine-learning, et de la recherche opérationnelle
- Technologie : Conception d'architectures Big Data industrielles, et développement d'applications Data Driven opérationnelles.

Notre offre de services est organisée autour de 3 pratiques métier :

- Sales & Marketing : Connaissance & expérience clients, tarification & yield management, marketing produit & mix modelling, pilotage & prévision des ventes
- Risque & Finance : Modélisation & gestion du risque, détection & gestion de la fraude, planification & consolidation, pilotage de la performance globale
- Opérations & Production : Ressource planning & supply chain, maintenance prédictive & smart alarming, pilotage & exploitation de réseaux complexes.

Dans chacun de ces domaines, nous disposons d'une équipe de consultants ayant développé une compétence et une expérience spécifique en la matière, qui nous permet d'assurer un relais efficace entre les besoins métier de nos clients et les réponses scientifiques et techniques que nous leur apportons.

We are data scientists

La Data est une source de **richesse**
pour votre entreprise.

C'est notre **conviction**, c'est notre **promesse**.

Développement d'une plate-forme de scoring de brevets

Porteur du projet

Caisse des Dépôts Propriété Intellectuelle

Objectifs

Le marché des brevets est à ce jour un marché encore peu transparent, sur lequel il est difficile -et coûteux- de disposer d'une valorisation objective des biens échangés.

Le projet s'inscrit dans ce contexte, et a pour objectif de développer une plate-forme permettant -sur la base des nombreuses informations publiques disponibles- de noter l'ensemble des brevets européens, et de mettre cette information à disposition du public concerné (industriels, fonds d'investissement...).

Le principe de la notation repose sur une hypothèse simple : les décisions « administratives » prises par le déposant d'un brevet à différentes étapes de son cycle de vie révèlent la qualité de l'invention que le brevet protège. Dès lors, la modélisation de ces décisions « administratives », mises en relation avec des caractéristiques observables du brevet (le texte même du brevet, le schéma de citations inter-brevets...), permettent d'élaborer des modèles prédictifs de la durée de vie des brevets, et ainsi d'approcher leur valeur.

Présentation

Nature des données traitées : Les données traitées portent sur l'ensemble des brevets européens. Il s'agit pour certaines de données structurées (classification internationale pour les domaines technologiques, citations de brevet à brevet, cycle de vie administratif du brevet...), et pour l'essentiel de données non structurées (texte des revendications et des descriptions, en français, anglais ou allemand). Ces données sont collectées à partir du référentiel de brevets de Questel (fournisseur de données spécialisé), et du site de l'Office Européen des Brevets.

Ressources et méthodologie : L'ensemble des données collectées est transformé puis stocké et historisé sur une plateforme Hadoop.

Les processus algorithmiques de calcul des notes -qui impliquent des traitements statistiques complexes, ainsi que des traitements de type textmining sur de gros volumes- sont exécutés automatiquement sous forme de batch, ou à la demande selon le paramétrage choisi par l'utilisateur (sélection d'un référentiel de calibration des modèles, modification des caractéristiques des brevets évalués).

Calendrier de déploiement : Après plusieurs expérimentations et tests en grandeur nature auprès d'utilisateurs pilotes, une première version accessible au public sera lancée au printemps 2015.

Aspect novateur

Le scoring élaboré dans le cadre de ce projet est le seul qui permette une combinaison algorithmique de nombreuses caractéristiques des brevets, dont certaines issues d'une approche textmining peu ou pas utilisée jusqu'ici. L'accent est également mis sur la possibilité de décomposer le score selon les caractéristiques les plus influentes du brevet, ce qui garantit la transparence des algorithmes de calcul utilisés. Enfin, les utilisateurs les plus avancés peuvent choisir un univers de référence pour la calibration des modèles, et obtenir des notes pour leur portefeuille et ceux de leurs concurrents aussi pertinentes que possible.

Valorisation observée de la donnée

La notation des brevets proposée permet de nombreux usages pour différents types d'acteurs de la propriété intellectuelle. Dans le cadre d'une revue de portefeuille à acquérir, elle peut par exemple aider des experts évaluateurs de brevets à identifier plus rapidement les meilleurs brevets. Elle permet de prendre en compte le portefeuille de brevets d'une société pour aider aux décisions d'acquisition ou de financement. Au sein d'un département de propriété intellectuelle, elle offre aussi un moyen de valoriser son portefeuille par rapport à celui de ses concurrents...

Etapas à venir et démarche Big Data dans l'entreprise

Les mois qui viennent vont permettre de porter l'ensemble des traitements en environnement de production.

En outre, les algorithmes de notation seront probablement encore affinés pour tenir compte des retours des premiers utilisateurs. Pour cela, il est notamment envisagé d'orienter le projet vers une logique d'open innovation grâce à la plate-forme datascience.net.

Assess Presence in Countries Weighted by Expected Duration of Renewal



Compare Patent Portfolios on Volume and Quality





COORDONNÉES CAPGEMINI

7, rue Frédéric Clavel
92287 Suresnes Cedex

Tél 01 49 00 40 00

Mail valerie.perhirin@capgemini.com

CONTACT

**Valérie
PERHIRIN**

Responsable France de l'offre Big Data

Dans un monde « constamment connecté », les sources de données se sont multipliées, des puces RFID aux transactions en ligne en passant par les analyses des codes QR et les messageries instantanées des réseaux sociaux. Non seulement le volume de données géré par les entreprises a explosé, mais analyser le passé ne suffit pas : pour pouvoir bénéficier d'un véritable avantage concurrentiel, vous devez anticiper l'avenir.

C'est là que l'expertise Big Data de Capgemini entre en jeu. Car si le marché évolue et les solutions Big Data se démocratisent rapidement, les entreprises peuvent encore avoir des interrogations sur les choix technologiques, d'infrastructure ou d'organisation. Capgemini dispose d'un solide savoir-faire avec notamment un centre d'excellence, un lab et des équipes expérimentées.

A propos de Capgemini

Fort de plus de 130 000 collaborateurs et présent dans plus de 40 pays, Capgemini est l'un des leaders mondiaux du conseil, des services informatiques et de l'infogérance. Le Groupe a réalisé en 2013 un chiffre d'affaires de 10,1 milliards d'euros.

Avec ses clients, Capgemini conçoit et met en œuvre les solutions business et technologiques qui correspondent à leurs besoins et leur apporte les résultats auxquels ils aspirent. Profondément multiculturel, Capgemini revendique un style de travail qui lui est propre, la « Collaborative Business Experience™ », et s'appuie sur un mode de production mondialisé, le « Rightshore® ».

Plus d'informations sur : www.fr.capgemini.com

Rightshore® est une marque du groupe Capgemini



Business Analytics et Big Data : les nouveaux enjeux de l'efficacité business

www.fr.capgemini.com/business-analytics-et-big-data

Télécom : optimiser l'utilisation réseau grâce au Big Data

Porteurs du Projet

La division Global Solution d'un acteur international dans le secteur des Télécoms, très présent en Europe et en Amérique du Nord

Objectifs

Notre client a missionné Capgemini pour développer une solution s'appuyant sur des outils Big Data afin de :

- Prédire les futures hausses du trafic sur des marchés et des régions spécifiques
- Identifier les zones où le réseau était surdimensionné ou sous-dimensionné
- Identifier les zones où la croissance du trafic pourrait amener des opportunités de ventes supplémentaires.

Présentation

Notre client cherchait à optimiser son réseau, à améliorer sa qualité de service et à identifier des modèles de comportements sociaux permettant d'initier de nouvelles activités génératrices de revenus. Pour y parvenir, la société a dû convertir ses données relatives à l'utilisation de son réseau en informations effectivement exploitables. La société craignait notamment que le trafic réseau, pour ses offres de gros destinées aux entreprises, ne soit en baisse en raison d'une concurrence agressive obligeant les sociétés à offrir des prix plus bas aux clients.

a. Nature des données traitées :

Capgemini a fourni au client une solution lui permettant d'analyser ses données via de nombreuses sources dont :

- le trafic réseau
- les informations sur les lignes réseau
- la qualité de service
- les informations sur les systèmes autonomes
- les données démographiques

b. Ressources et méthodologie (compétences humaines, outils techniques, fonctionnement du traitement de la donnée)

Capgemini a tiré parti des modèles de thermodynamique sociale qui permettent d'effectuer des analyses et des prédictions sur l'utilisation du réseau.

Les rapports ont été produits via la solution SAP Business Objects BI 4.1. Cloudera

Enterprise 4.5 a été utilisé comme plate-forme Big Data, tandis que les requêtes ont été traitées directement avec Cloudera Impala.

La solution incluait également :

- L'utilisation des méthodes agiles
- L'introduction d'un outil ESB/ELT pour l'intégration des données
- L'utilisation d'un processeur graphique (GPU) pour paralléliser à grande échelle les calculs thermodynamiques
- L'utilisation d'un modèle thermodynamique pour l'analyse de données

c. Calendrier de déploiement : Confidentiel

Aspect Novateur

- Utilisation d'algorithmes scientifiques permettant de mieux modéliser le trafic internet et le comportement des routeurs réseau.

- Utilisation de technologies Hadoop Cloudera dans le cloud AWS, pour une plus grande flexibilité de calcul.

Valorisation observée de la donnée

Durant la première phase du projet, les données ont été transférées et centralisées dans un Enterprise Data Hub. La solution Enterprise Data Hub de Cloudera offre une plus grande souplesse et réactivité dans le stockage et l'analyse des données, tout en permettant au client de réduire considérablement ses coûts.

Le projet a également permis au client de détecter à l'avance les zones géographiques de son réseau où il pourrait cibler ses investissements.

Autres bénéfices notables pour le client :

- Améliorations opérationnelles
- Identification de nouvelles opportunités commerciales

- Amélioration de l'appui stratégique aux activités commerciales
- Rapports d'entreprise plus flexibles, toutes les informations étant à disposition dans un référentiel unique
- Prédiction de résultats plus précises grâce à l'utilisation des modèles thermodynamiques

Etapes à venir et démarche : Confidentiel





COORDONNÉES

CENTRALESUPÉLEC EXECUTIVE EDUCATION

Grande voie des Vignes
92 295 CHATENAY MALABRY

Tél +33(0) 1 41 13 16 14

Mail info@cf.ecp.fr

CONTACT

**Rania
OUARET**

Chargée d'affaires SI

Le monde change, nous vous aidons à changer

Par la complémentarité de leur offre de formation et de leurs activités de recherche, Centrale Paris et Supélec proposent une offre de formation continue très complète qui couvre le champ des Sciences de l'Ingénieur. Nos approches pédagogiques s'attachent à faire de la formation continue un outil de valorisation et de reconnaissance pour les collaborateurs et les managers des entreprises, sans perdre de vue l'amélioration de la performance de la structure.

Conjuguons management et technologies

Parce que le monde change, parce qu'il change à grande vitesse, les formations de CentraleSupélec Executive Education sont non seulement axées sur les compétences, mais aussi sur le développement des capacités des individus et des entreprises à changer de façon continue. C'est un enjeu majeur qui nécessite autant d'agilité que de talent. CentraleSupélec Executive Education hérite de deux cultures férues d'excellence et réputées pour leurs approches avant-gardistes de la pédagogie. Car à Centrale Paris comme à Supélec, les formations dépassent largement la seule acquisition de connaissances : l'objectif est de former des personnalités conscientes des grands enjeux économiques, environnementaux et sociaux de notre époque pour transformer le quotidien.

La formation Big Data – enjeux et opportunités

A l'ère du numérique, le déluge de données crée de nouvelles opportunités économiques pour les entreprises. Toute la chaîne du décisionnel est impactée par ces « Big Data », de la collecte des données hétérogènes à leur analyse et leur visualisation, en temps contraint. Différents métiers de l'entreprise sont également impactés par ce phénomène comme le marketing ou la stratégie d'entreprise.

Objectifs de la formation

- COMPRENDRE les enjeux des Big Data
- IDENTIFIER les opportunités à travers l'open data et les cas d'usage métiers
- ACQUÉRIR le vocabulaire lié à cet écosystème complexe
- ADAPTER la stratégie d'entreprise
- MAÎTRISER les enjeux juridiques et liés à la protection des données
- ACQUÉRIR les technologies des Big Data (acquisition et stockage des données, traitement distribué, analyse de données à large échelle)
- MAÎTRISER les technologies par des études de cas concrètes
- CONCEVOIR un projet Big Data.



CentraleSupélec

Les formations continues BIG DATA de CentraleSupélec Executive Education

Stratégiques pour vous. Essentielles pour votre entreprise.

**Parce que
votre
compétence
est l'instrument
privilegié
de votre
croissance.**

EXECUTIVE CERTIFICATE BIG DATA POUR L'ENTREPRISE NUMÉRIQUE

ENJEUX, PERSPECTIVES ET PANORAMA DES BIG DATA

ASPECTS JURIDIQUES ET PROTECTION DES DONNÉES

ACQUISITION ET STOCKAGE DES DONNÉES

TRAITEMENT DISTRIBUÉ

ANALYSE DE DONNÉES & BIG DATA, CONFESS YOUR DATA

CADRAGE ET MANAGEMENT DE PROJETS BIG DATA

+ DES FORMATIONS SUR MESURE, PERSONNALISÉES, ADAPTÉES À VOS BESOINS

Documentation et catalogue
www.cf.ecp.fr
Conseil et accompagnement
01 41 13 16 14 / info@cf.ecp.fr



EX3
EXECUTIVE
EDUCATION

EXPERIENCE



Interview de Marie Aude Aufaure

4 questions à Marie Aude AUFAURE

Responsable pédagogique des formations Big Data à CentraleSupélec
Expert auprès de la commission européenne sur les appels à projets Big Data

Pourquoi se former sur les Big Data ?

Les Big Data constituent un enjeu à la fois en termes d'opportunités business et de défis scientifiques et techniques. C'est un vecteur d'innovation et l'impact socio-économique est fort. La bonne utilisation des données massives peut accroître la productivité, l'innovation et la compétitivité dans des secteurs économiques entiers. Dans de nombreux domaines, tels que l'énergie, le commerce électronique, le secteur public, les médias, l'environnement pour n'en citer que quelques-uns, les données massives s'accumulent. Ces données se caractérisent par une volumétrie importante, une forte hétérogénéité et une provenance en temps réel. Leur gestion et leur exploitation deviennent des enjeux stratégiques, en mettant en œuvre des méthodes d'analyse capable de passer à l'échelle Big Data. Le métier de Data Scientist émerge, nécessitant une solide connaissance en statistiques et méthodes avancées d'analyse de données combinée à une compréhension de la gestion de données passant à l'échelle, sur les aspects techniques et implémentation. Un autre spécialiste des Big Data, le Data Engineer, doit être capable de développer des techniques, processus et méthodes pour transformer les données en business et générer de la valeur.



Quelle est la valeur ajoutée de CentraleSupélec Executive Education sur ce sujet ?

Cette formation intègre à la fois les aspects techniques et managériaux des fonctions liées aux Big Data. En effet, il est important de comprendre leurs enjeux, les aspects légaux liés à la sécurité et à la protection des données personnelles mais aussi d'être en mesure de mettre en œuvre un projet Big Data sur le plan technique. La formation délivrée par CentraleSupélec Executive Education aborde l'ensemble des différentes facettes de la gestion et de l'exploitation des données massives, des cas d'utilisation et de la stratégie d'entreprise.

Cette formation est également modulaire, conduisant à des certifications intermédiaires (globale, management ou technique).

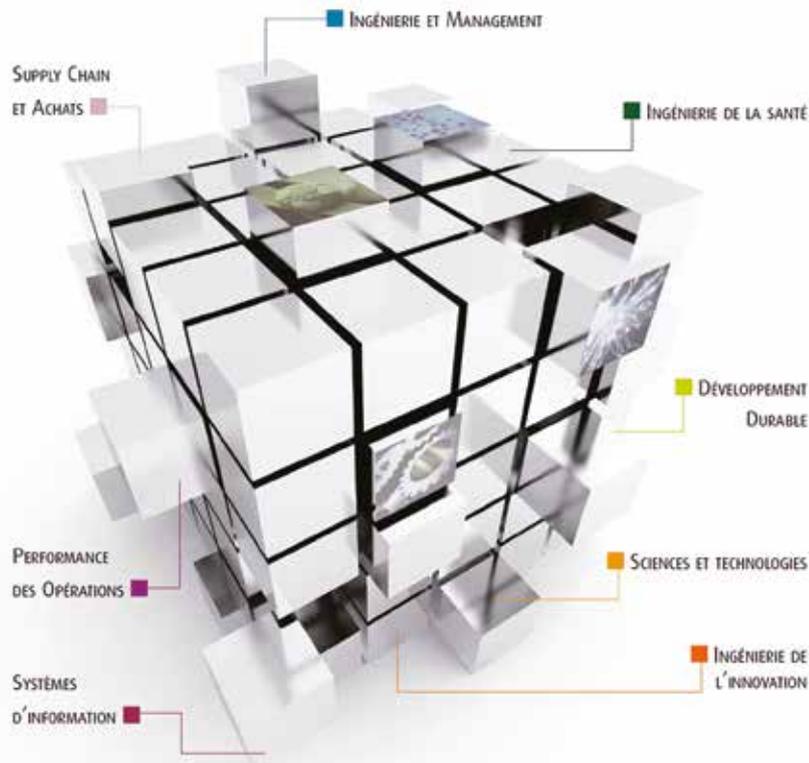
Vers qui est orientée cette formation ?

L'Executive Certificate Big Data pour l'entreprise numérique est destiné à un public de Directeurs/chefs de projet, managers des systèmes d'information, experts en business intelligence, ingénieurs de recherche et développement, consultants techniques, data miners.

Quelles en sont les modalités (durée, implication nécessaire, pré-requis, etc)

La formation dure 20 jours (140 heures) et donne lieu à une certification basée sur un test de connaissances pour chaque module et un mémoire avec soutenance en fin de cursus. La formation inclut une journée d'accompagnement à la préparation du mémoire et à la soutenance. Les méthodes pédagogiques sont basées sur un exposé de l'état de l'art, des mises en pratique via des logiciels spécifiques et des études de cas. A ceci s'ajoutent des retours d'expérience d'industriels du domaine sous forme de webinars qui pourront être suivis en ligne avec une séance de questions/ réponses, ou de manière différée.

Deux certifications intermédiaires sont également prévues, l'une d'une durée de 6,5 jours abordant les aspects enjeux, juridiques et gestion de projet et la seconde d'une durée de 16 jours abordant les aspects techniques : stockage, traitement distribué, analyse, gestion de projet ainsi que les enjeux.





COORDONNÉES

CLUDERA FRANCE SRL

Mail emea-enquiries@cloudera.com

CONTACT

**Romain
PICARD**

EMEA Regional Director



Cloudera révolutionne la gestion des données d'entreprise avec la première plateforme unifiée pour le Big Data basée sur le framework Apache Hadoop™. Cloudera fournit aux entreprises un lieu unique pour stocker, traiter et analyser leurs informations, tout en créant de nouveaux moyens de maximiser la valeur des investissements existants et de générer davantage de valeur à partir de leurs données. Pionnier de Hadoop dès 2008, Cloudera demeure le principal fournisseur et défenseur du framework Apache dans les entreprises. Cloudera propose également des logiciels de gestion d'applications critiques — stockage, accès, gestion, analyse, sécurité et recherche de données. Cloudera collabore avec plus de 1 000 fournisseurs de matériels, logiciels et services pour répondre aux objectifs ambitieux de ses clients.

Interview de Romain Picard

Romain est entré chez Cloudera en Juillet 2014, dans le cadre de la stratégie d'implantation de la société en Europe. Constatant une forte demande de solution Hadoop, Cloudera a mis en place un plan de déploiement ambitieux, incluant d'importants recrutements.

Romain en est persuadé : les données sont le nouvel « or noir », mais c'est aussi une matière à contrôler.

Dans un premier temps, Hadoop permet la démocratisation du Big Data. L'augmentation des volumes de données n'est pas réservée qu'aux grands acteurs, toute entreprise à travers le monde est concernée, PME y compris. Toute entreprise peut désormais mettre en place une infrastructure Hadoop, centralisée et accessible aux métiers en permanence. Chez Cloudera, on l'appelle l'Enterprise Data Hub. L'intérêt étant d'avoir accès à toutes les données quand on le souhaite sans limite de temps.

Toute la problématique réside alors dans leur exploitation. Comme le dit Romain, comment passe-t-on des données à l'information ? C'est là qu'intervient le datascientist, élément incontournable qui détermine quelles données seront utiles pour générer quelle information. Reste à convaincre les métiers de leur intérêt et les pousser à se poser des questions qu'ils n'auraient pas imaginées auparavant.

L'intérêt validé, demeure la question de la sécurité. Si de nombreuses problématiques métiers sont adressées au sein de l'entreprise grâce à Hadoop, il est indispensable de définir qui a accès à quelles données. Avec l'acquisition de Gazzang, Cloudera se dote d'un outil capable de gérer cette question et devient le premier fournisseur capable d'offrir une solution Hadoop entièrement sécurisée. Les données sont encodées, anonymisées, leur accès au sein de l'entreprise est limité et monitoré en temps réel, ceci pour répondre à des problématiques de compliances, audits, etc.

Dernier point important d'après Romain : la simplification IT. Plus de silot avec des systèmes transactionnels de partout mais un seul Enterprise data Hub. Un seul et même réceptacle accueille toutes les données, qui sont ensuite analysées par des outils simples et intuitifs selon les besoins de chaque métier. Cloudera apporte pour sa part l'expertise fonctionnelle et des supports tels Cloudera Manager.

Chaque secteur et chaque métier adresse ses propres problématiques. Dans le retail, la capacité de positionner le bon produit au bon moment à un client donné améliorera la compétitivité. Que ce soit augmenter le CA, limiter les coûts, améliorer la qualité de vie, garantir la sécurité, offrir des services innovants... Le Big Data répond présent.

Romain nous relate alors une anecdote de son dirigeant, Tom Reilly, qui alors qu'il se trouvait au volant de sa voiture en Californie, entendit à la radio un alerte tremblement de terre, décomptant 8 secondes avant qu'elles surviennent les secousses. 8 secondes avant ? En quoi cela est-il utile ? Si un humain a en effet peu de capacités de réaction en un temps aussi court, les machines, elles, trouvent dans ces quelques secondes un intérêt primordial. Fermer un pont, bloquer des ascenseurs... les alertes et réaction automatisées peuvent éviter de nombreux accidents.

De belles opportunités à la clé

Le Big Data offre aux entreprises de différents secteurs une formidable opportunité. En exploitant de nouveaux gisements de données de toutes sortes, elles peuvent poser des questions nouvelles et originales concernant leurs clients et leur activité. Par exemple, les entreprises utilisent ces données pour améliorer l'expérience vécue par leurs clients en vue de les fidéliser tout en générant une valeur ajoutée accrue. Dans le même temps, en disposant d'une vision pertinente et approfondie des activités commerciales, ces entreprises peuvent identifier les zones d'inefficacité qui, une fois traitées, participeront potentiellement à la réduction des coûts d'exploitation.

La plateforme Enterprise Data Hub : la clé du succès pour le Big Data

L'objectif est simple : acquérir et combiner tous volumes ou types de données en conservant fidèlement leurs caractéristiques initiales en un seul endroit et aussi longtemps que nécessaire, et offrir des informations pertinentes à tous les types d'utilisateurs, aussi vite que possible, en utilisant les investissements et les ressources avec une efficacité maximale. Place à la plateforme Cloudera Enterprise Data Hub Edition !

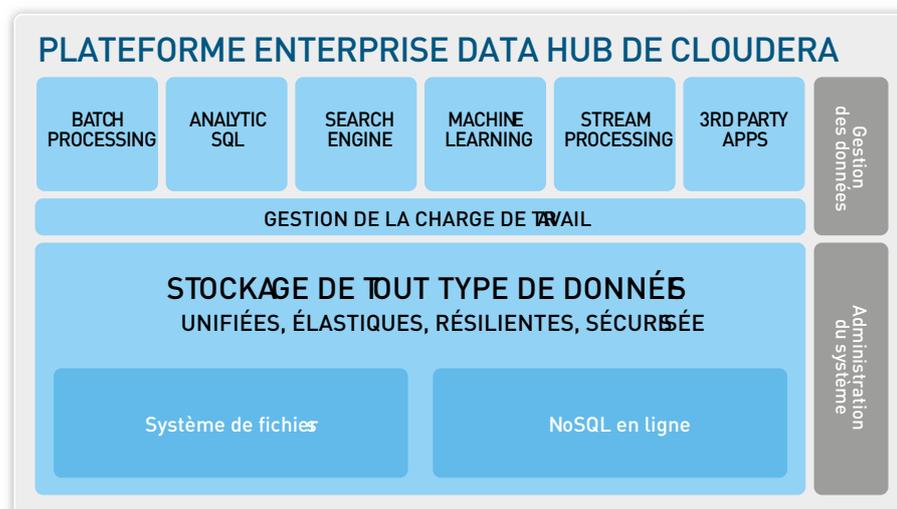
Par rapport aux systèmes existants, la plateforme Enterprise Data Hub Edition (EDH) de Cloudera présente des avantages suivants :

→ **Des archives actives** : Un endroit unique pour stocker toutes vos données, tous formats confondus, sans limite de volume, aussi longtemps que vous le souhaitez. Vous pouvez ainsi répondre aux exigences de conformité et fournir des données sur demande pour satisfaire aux exigences réglementaires, internes ou externes. Parce que ce lieu est sécurisé, vous savez qui voit quoi, et grâce aux services de gouvernance et de lignage, vous pouvez suivre l'accès à vos données, ainsi que leur évolution au fil du temps.

→ **Transformation et traitement** : Les charges de travail ETL qui devaient auparavant être exécutées sur des systèmes coûteux peuvent migrer vers la plateforme Enterprise Data Hub où elles pourront tourner à très faible coût, en parallèle et beaucoup plus rapidement. L'optimisation de l'emplacement de ces charges de travail et des données associées libère de la capacité sur les systèmes analytiques et les entrepôts de données haut de gamme. Ce sont par conséquent de solides atouts capables de se concentrer sur les applications critiques prises en charge, telles que les applications OLAP.

→ **Informatique décisionnelle exploratoire (Self-service BI)** : Les utilisateurs veulent souvent accéder à des données aux fins de reporting, d'exploration et d'analyse. Dans de nombreux cas, les systèmes d'entrepôts de données de production doivent être protégés contre un usage informel afin d'être en mesure d'exécuter les traitements analytiques et opérationnelles qu'ils prennent en charge. Une plateforme EDH permet aux utilisateurs d'explorer les données en toute sécurité, en utilisant des outils d'informatique décisionnelle (BI) interactifs traditionnels au moyen du langage SQL et de recherches par mot-clé.

→ **Fonctions analytiques avancées** : Plusieurs environnements informatiques (computing frameworks) dédiés aux fonctions analytiques, à la recherche ou au machine learning, apportent une nouvelle valeur aux sources de données, nouvelles et anciennes. Au lieu d'examiner des échantillons de données ou des instantanés correspondant à de courtes périodes de temps, toutes les données historiques peuvent être combinées dans des analyses complètes avec une parfaite fidélité. Des données tabulaires simples peuvent être panachées avec des données plus complexes et multi-structurées, ce qui était jusqu'alors impossible.





COORDONNÉES CLUB DÉCISION DSI

16 Place Vendôme
75001 Paris

Tél +33 (0) 1 53 45 28 65
Mail 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.



Véronique Daval Présidente



Julien Daval Vice-Président

LES MEMBRES DU BUREAU ET AMBASSADEURS DU CLUB



Armand ASSOULINE
DSI
GROUPE NICOTRANS



Gilles BERTHELOT
RSSI
RESEAU FERRE DE FRANCE



Claude DE SOUSA
DSI
CBRE



Christian DOGNET
DSI
CHAINE THERMALE DU SOLEIL



Damien GRIESSINGER
CTO
EPO



Trieu HUYNH-THIEN
DSI ADJOINT
CENTRE GEORGES POMPIDOU



Bertrand LEMAIN
DSI
CCI DE SEINE ET MARNE



Lionel ROBIN
DSI
GROUPE LA RESERVE



Dominique TROUVE
DSI
HOPITAL AUCENNE



Claude YAMEGGO
ARCHITECT SI
ALSTOM

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 1er 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 :



Devenez Membre du **Club Décision DSI**

Le 1^{er} 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



**COORDONNÉES
CSC**

Immeuble Balzac : 10 Place des Vosges
92072 Paris La Defense Cedex

Tél **+33 (0)1 55 70 70 70**
Mail lguiraud@csc.com

**CONTACT
Laurent
GUIRAUD**

*Responsable Offre Big Data
& Analytics France*

Si vous aviez la possibilité de l'obtenir, quelle information transformerait radicalement le mode de fonctionnement de votre entreprise ?

CSC fournit à ses clients les fondations pour la gestion de l'information et les capacités analytiques leur permettant d'obtenir un aperçu sur le passé, le présent et le futur. Et nous pouvons les aider à obtenir des résultats en moins de 30 jours, ainsi CSC aide les organisations à tirer parti de la puissance des Big Data pour orienter leur activité et faire face à la concurrence.

CSC offre des services de Consulting couvrant les aspects Métier, Data Science et Plateforme à partir d'offres packagées, des meilleures technologies, et d'une plateforme combinant des solutions traditionnelles et open source pour résoudre les cas d'usage les plus complexes de nos clients.

A propos de CSC

CSC est un des leaders mondiaux des solutions et des services informatiques de nouvelle génération. L'entreprise se donne pour mission de faire fructifier les investissements technologiques de ses clients, en associant son expertise métier et sa dimension mondiale aux meilleures solutions du marché

Grâce à sa large gamme de compétences, CSC fournit à ses clients des solutions sur mesure pour gérer la complexité de leurs projets afin de leur permettre de se concentrer sur leur cœur de métier, collaborer avec leurs partenaires et clients, et améliorer leurs performances opérationnelles.

79 000 professionnels dans le monde travaillent chez CSC, au service de clients répartis dans plus de 70 pays. CSC a réalisé un chiffre d'affaires de 13 milliards de dollars au cours de l'année fiscale clôturée le 30 mars 2014

STORM
R
HADOOP
D3
NOSQL

ERP
EDW
CRM
HCM
PLM

AUDIO/VIDEO
LOGS WEB
FLUX DE CLICS
DONNÉES DE CAPTEURS
MÉDIAS SOCIAUX

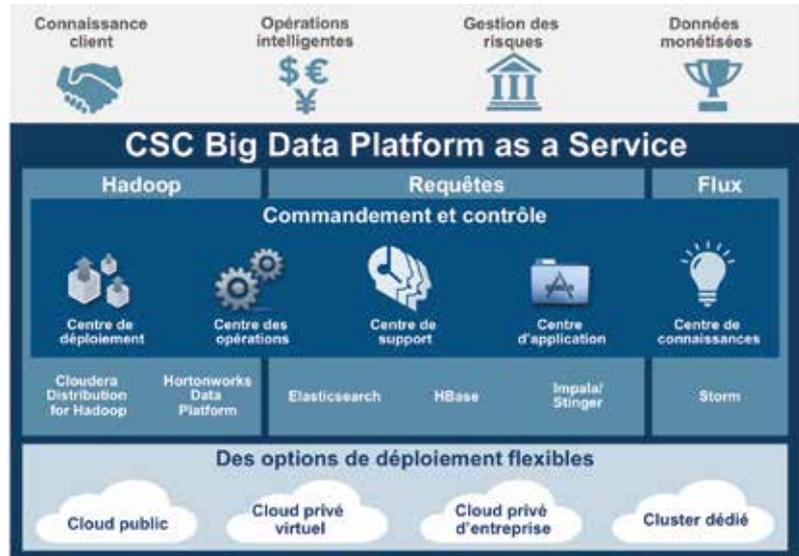


Big Data PaaS

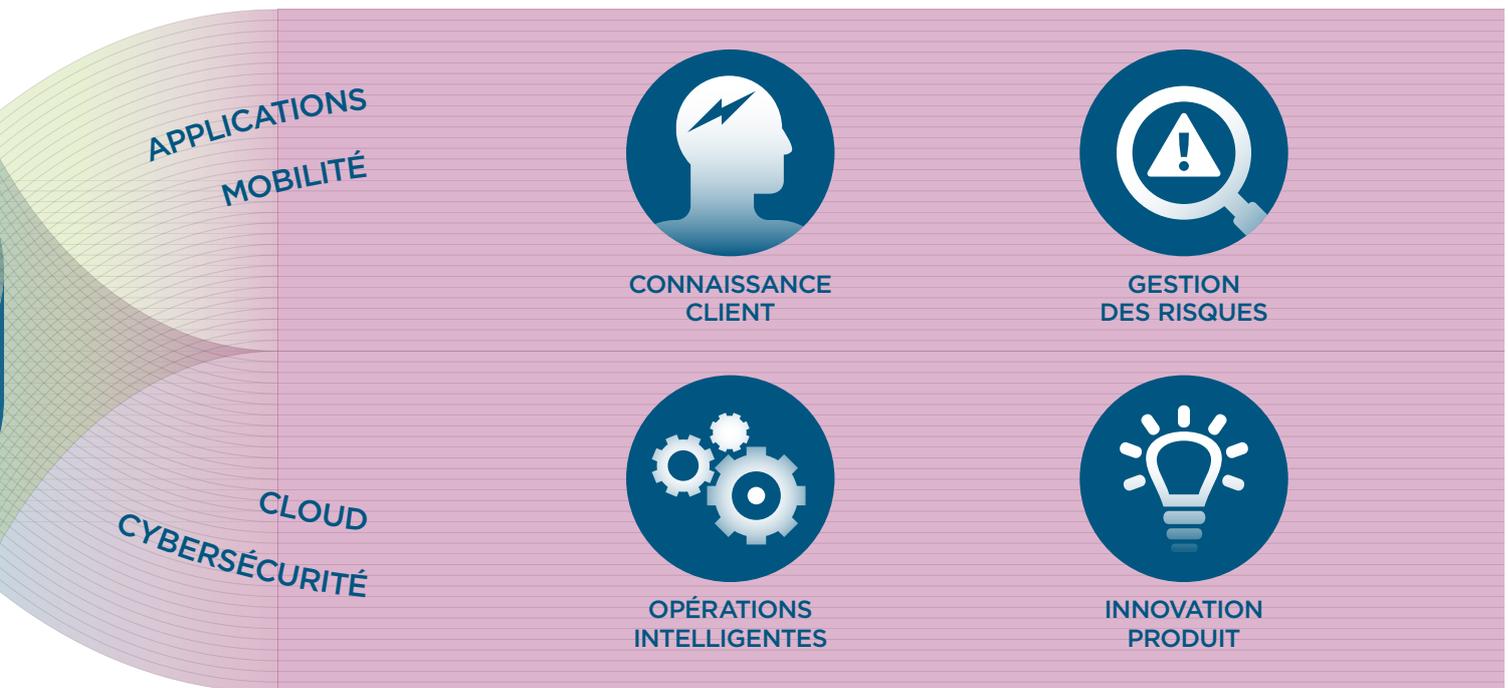
La solution **Big Data Platform-as-a-Service** de CSC est une plateforme Big Data entièrement intégrée et opérée, permettant à nos clients de développer et tirer parti d'un environnement Big Data en moins de 30 jours :

- livrée en mode "as a Service "
- utilisant des technologies web scale avancées
- s'appuyant sur des solutions open source et des modèles opérationnels utilisées par Google, Yahoo!, Facebook, Twitter & LinkedIn.
- supportant toute combinaison d'analytique ad-hoc, batch et temps réel

Avec Big Data PaaS il est beaucoup plus facile et moins coûteux de développer, mettre en œuvre et opérer des applications Big Data.



DES RÉSULTATS TANGIBLES AVEC LE **BIG DATA**





COORDONNÉES

DATASTAX FRANCE

121 avenue des Champs Elysees,
Paris, France 75008

Tél **+33 (0)1 72 71 85 85**

Fax **+33 (0)1 72 71 85 99**

Mail **svandenberg@datastax.com**

CONTACT

**Steve
VAN DEN BERG**

*DataStax Regional Director
Western Europe*

DataStax épaula les applications en ligne qui transforment l'activité de plus de 300 clients, parmi lesquels de nombreuses start-ups et 20 des Fortune 100. DataStax offre une technologie de base de données massivement extensible, flexible et constamment disponible, intégrant une version d'Apache Cassandra™ certifiée pour la production avec analyses, recherches et sécurité pour les données en temps réel, dans des centres de données multiples et dans le cloud. De grandes sociétés font confiance à DataStax pour transformer leur activité : Adobe, HealthCare Anytime, eBay, ou encore Netflix.



Base de données distribuée pour les entreprises du Web

- Commencez à n'importe quel niveau, expansif à souhait
- Architecture toujours en activité
- Cassandra certifié pour les déploiements en production
- Déploiement sur site ou dans le Cloud
- Architecture dynamique et flexible

Choisir une stratégie NoSQL

Les analystes s'attendent à ce que le marché du NoSQL progresse trois fois plus vite que celui du SQL dans les prochaines années. Les nouveaux venus dans le Big Data auront donc à cœur de comprendre pourquoi et comment ce type de technologie peut s'adapter à leur organisation. Les motivations pour passer au NoSQL sont liées aux besoins de performances, de montée en charge, de très haute disponibilité, de distribution géographique, de gestion de tout type de données et de réduction des coûts. Il existe cependant des obstacles indéniables à son adoption, aussi bien techniques que non techniques. Il est important d'étudier ceux-ci en même temps que les améliorations apportées par l'écosystème NoSQL. Par ailleurs, il est important de connaître les cas d'utilisation afin de correctement introduire cette technologie NoSQL dans l'entreprise.

Choisir une base de données NoSQL

Le site Web nosql-database.org liste actuellement plus de 150 bases de données NoSQL différentes. Avec une telle liste, comment trouver les candidats qui pourront convenir à votre scénario d'utilisation ?

Principaux critères de choix : De nombreuses caractéristiques et fonctions distinguent les différentes bases de données NoSQL. On peut cependant s'aider de critères pour mieux cerner celles qui conviendront à une utilisation définie. Pour commencer, il faut savoir quel type de données vous avez besoin de gérer. En effet, les bases de données NoSQL utilisent des modèles de données très divers (colonnes, documents, graphes...) et l'inadéquation entre le modèle de données et l'application peut être fatale au projet.

La montée en charge attendue : la question suivante concerne le potentiel de croissance de l'application et l'augmentation du volume des données. Certaines bases de données NoSQL se résument principalement au stockage en mémoire. Elles sont donc difficiles et coûteuses à faire monter en charge, à la différence de Apache Cassandra, capable d'une extension linéaire sur de très nombreuses machines.

Le modèle de distribution des données : il est important de définir le besoin de répartition des données, la prise en compte de plusieurs régions géographiques et la nécessité de récupération après sinistre. Il faut aussi se demander si les lectures et les écritures devront toutes être prises en charge sur les sites distribués. Certaines bases de données NoSQL utilisent des architectures maître-esclave (bien que leur appellation soit parfois « primaire/secondaire »), qui permettent une montée en charge des opérations de lecture seulement. D'autres architectures « Peer-to-Peer » (sans relation « maître-esclave ») comme Apache Cassandra, permettent de supporter l'augmentation massive des besoins à la fois en lecture et en écriture.

Le déploiement d'une stratégie NoSQL dans votre entreprise impose donc de bien comprendre pourquoi et comment les sociétés utilisant cette technologie ont réussi. Il faut également déterminer comment votre activité va pouvoir bénéficier des spécificités du NoSQL. Une fois que vous aurez décidé quelle technologie NoSQL est faite pour vous, il vous faudra encore comprendre les pièges à éviter, les critères de choix selon vos applications et la bonne stratégie de déploiement.

DataStax propose des logiciels, des services et des stratégies adaptés à l'entreprise pour vous aider à réussir votre conversion au NoSQL. Avec sa solution DataStax Enterprise utilisant Apache Cassandra, sécurisée et éprouvée, son assistance 24 heures sur 24, ses conseils et ses formations dispensés par des experts, DataStax peut vous aider à tirer tout le bénéfice du NoSQL tout en facilitant le changement.

Pour en savoir plus sur Apache Cassandra et DataStax et télécharger les logiciels Cassandra et DataStax Enterprise, consultez le site www.datastax.com ou envoyez un message à info@datastax.com.



World Class Training.
Cassandra Instructor-led courses
NOW AVAILABLE!





COORDONNÉES

EDIS CONSULTING - JEMS GROUP

6 bis rue des Gravieres
92 200 Neuilly-sur-Seine

Tél 01 47 89 63 10

Mail mbrancher@edis-consulting.com

CONTACT

**Matthieu
BRANCHER**

Directeur Commercial

EDIS est une structure spécialisée dans les systèmes d'aide à la décision et l'un des plus importants acteurs indépendants du marché français de la Business Intelligence.

EDIS accompagne ses clients dans la définition et le choix des technologies, l'installation de l'architecture BIG DATA ainsi que son intégration au sein de leur système d'information (SI).

Associé à JEMS Group, entreprise dédiée aux systèmes d'information, EDIS développe depuis 10 ans des offres de conseil et d'intégration de solutions décisionnelles couvrant toute la chaîne de valeur jusqu'au pilotage de la performance des processus et des activités métier.

Avec une politique de partenariat active aux côtés des principaux éditeurs de Business Intelligence historiques (SAP-BO, IBM, Oracle, MS), et des challengers aux solutions innovantes et complémentaires (Talend, MAP-R, Tableau...), toutes les équipes EDIS rassemblées en pôles d'expertise, maîtrisent leurs outils, et sont formées aux meilleures méthodes et pratiques de conception, réalisation et de management de projets.

LE pôle BIG DATA, conçoit et met à disposition des clients, de tous secteurs d'activités, des solutions innovantes permettant la collecte, le stockage, l'analyse en temps réel, la diffusion et la visualisation de données structurées ou non structurées.

EDIS propose une méthodologie de « metadata enhancement » permettant d'extraire de la pertinence et d'explorer de nouveaux axes d'analyses d'intérêts pour les métiers.

Créé également en 2002, JEMS Group est LE cabinet en systèmes d'information, fondé sur des valeurs humaines solides pour ses clients et ses consultants.

Renforcé en 2012 grâce au rapprochement avec EDIS, JEMS possède un savoir-faire et un savoir-être qui assurent la réussite de chaque projet. JEMS Group compte 260 collaborateurs et réalise 24,5 M€ de chiffre d'affaires



BIG DATA

DATAWAREHOUSE

DATAVISUALISATION

DASHBOARDING

DATAVALORISATION

HADOOP

DATALAKE

**Vous avez les données,
nous avons le savoir-faire !**

**RETROUVEZ NOTRE EXPERTISE
ET NOS RÉALISATIONS SUR
WWW.EDIS-CONSULTING.COM**

Constitution d'un Data Lake en milieu bancaire

Porteurs du Projet

L'un de nos clients, grand acteur du secteur bancaire, souhaite s'appuyer sur une innovation technologique majeure de gestion et de management de données (Hadoop) pour centraliser et partager les données internes et externes de ses différents métiers.

Objectifs

L'objectif du projet est de développer les synergies par un accès et une exploitation plus transversale du capital d'informations économiques. En favorisant l'accès, l'enrichissement, le partage et le traitement de l'information, le projet doit lui permettre de promouvoir l'échange d'information entre les différents acteurs.

Présentation

Ce projet majeur vise à regrouper dans un espace unifié et mutualisé (DATA LAKE reposant sur le framework Hadoop) des données économiques en provenance de plusieurs sources. EDIS, spécialiste du traitement de l'information et des infrastructures Big Data, accompagne cette transition technologique et organisationnelle. Nos équipes mènent de front les chantiers organisationnels, fonctionnels et technologiques.

a. Nature des données traitées : La plateforme BigData accueille un grand nombre de données hétérogènes provenant de différents producteurs internes/externes, avec des structures et formats différents et des systèmes de codification hétérogènes.

b. Ressources et méthodologie (compétences humaines, outils techniques, fonctionnement du traitement de la donnée) : EDIS, apporte son avoir faire à la méthodologie, réalise le cadrage et la définition des besoins, assure la bonne mise en œuvre du projet et participe à la définition et la mise en œuvre de l'infrastructure.

EDIS apporte également son expertise sur l'architecture Big Data :

- Collecte et stockage de gros volume de données dans Hadoop sur des bases de données NoSQL (HBase, MongoDB, Casandra...)
- Mise à disposition rapide (HDFS, NFS, YARN, MapReduce...)
- Analyse, Valorisation & Visualisation (Hive, Pig, Impala, Mahout, Logstash, ElasticSearch, Shark, Spark, R, Tableau, Kibana etc...).

Notre méthodologie pour la constitution de DATALAKE, repose sur un processus d'intégration de données qui sont ensuite enrichies en métadonnées afin de disposer d'informations contextuelles insérées dans un référentiel. Cet enrichissement s'appuie sur un dictionnaire de métadonnées et contribue à matérialiser les adhérences entre les catégories de données issues de sources différentes.

Une fois intégrées, ces données enrichies peuvent être consultées et exploitées par les utilisateurs :

- Interrogation / recherche du contenu à partir d'un langage naturel
- Utilisation du moteur de recherche et d'indexation ElasticSearch.

Aspect Novateur

Ce projet, par sa taille et ses ambitions est très certainement l'un des plus grands chantiers BIG DATA mené par un acteur Français de portée internationale. Il doit permettre à notre client sur plusieurs années de collecter, stocker, intégrer des données volumineuses, hétérogènes et variées dans le but de les analyser et les visualiser en temps réel.

Valorisation observée de la donnée

Au-delà de rassembler des données aussi variées, le projet permet de les enrichir avec des métadonnées ad-hoc (tag spécifiques, référentiels, synonymes, traductions, méthodologies..) qui facilitent l'accès et l'analyse de l'information contenue dans les données produites par les métiers depuis l'IHM posée sur le socle commun.

Etapes à venir et démarche Big Data dans l'entreprise

Le projet doit démontrer la possibilité de rassembler toutes les données de l'entreprise dans une « boîte » commune afin de faciliter leur stockage, leur accès, leur partage, leur analyse et leur visualisation conviviale adaptée à toutes les familles de métiers. Elle doit prouver la malléabilité, et la « scalabilité » d'un tel système dans le temps. Il ne s'agit pas d'une initiative locale mais bien d'une mutation complète du système d'information du client.

Autres informations

Nos clients ont les données, nous avons la méthodologie et le savoir-faire.





COORDONNÉES

EKIMETRICS

136, avenue des champs Elysées
Paris 75008

Tél +33 (0) 1 71 19 49 84

Mail fp@ekimetrics.com

CONTACT

**François
POITRINE**

Directeur Général. Responsable
de la practice BigData

Best in Class statistics au service de la stratégie marketing

Ekimetrics a été fondé en 2006 à Paris et compte 70 consultants, basés à Paris, Londres et New York. Notre approche novatrice et rigoureuse dans le traitement de la data nous a permis de devenir leader européen du ROI et du pilotage par la performance :



Ekimetrics est un des rares cabinets de conseil en stratégie spécialisé dans l'utilisation d'outils statistiques de pointe permettant l'optimisation de la stratégie Marketing et Data de grandes entreprises françaises et internationales (CAC 40 / Fortune 500)

- Grâce à notre triple expertise : stratégique, statistique et technologique ; nous sommes à même de livrer des recommandations actionnables par les CMOs et les CIOs.
- Nos 70 consultants ont tous un double profil consultant / statisticien-Data Scientist (X, Mines, HEC, ENSAE).
- Nous sommes indépendants de tout organisme de médias et d'éditeurs de logiciels, ce qui nous permet une grande agilité dans le choix des outils et des méthodes, ainsi qu'une réelle impartialité dans nos recommandations .
- Nous partons de l'usage et des besoins business de nos clients pour construire la méthodologie et l'architecture adaptée.
- Nous avons un partenariat avec la chaire d'économie des médias MINES ParisTech.
- Nous délivrons des plans d'actions au bout de 4 mois. Nos recommandations ont un impact mesurable sur la productivité de votre marketing dès la livraison de nos modèles.



Eki SocialWeb Lab est un laboratoire de R&D au sein d'Ekimetrics. Le laboratoire a pour vocation de comprendre et d'analyser les grandes tendances du SocialWeb, en s'appuyant sur l'expertise en Datascience et la compréhension métier de ses consultants. Grâce notamment à des algorithmes de classement sémantique propriétaires, les analystes du SocialWeb Lab sont à même de décrypter le bruit autour de grands événements culturels et politiques pour le compte de marques et de grands médias (Libération, LeLab Europe1, TheMediaShaker ...). Ainsi, durant l'année écoulée, notre Lab a notamment analysé la Fashion Week, les élections municipales à Paris, le festival de Cannes et la Coupe du Monde de football.

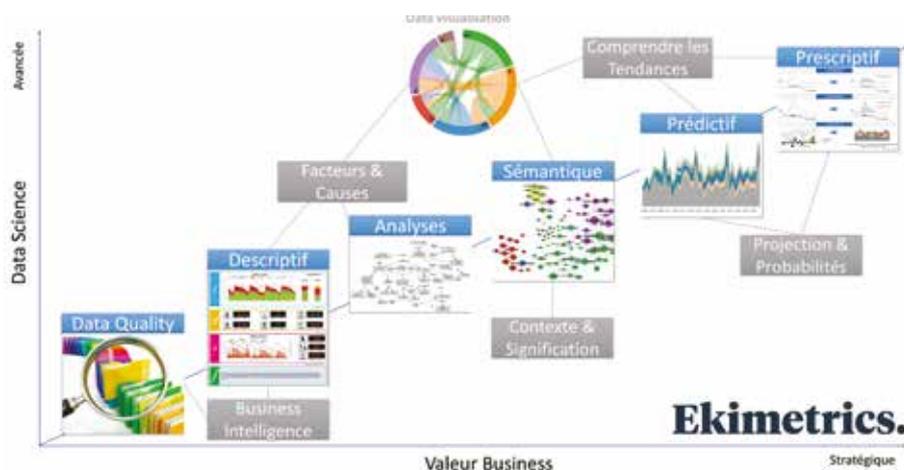
Nous mettons le Big Data au service de votre stratégie

Ekimetrics.

Notre expertise en Data Science pour vous accompagner dans vos projets Big Data

Durant les deux dernières années, Ekimetrics a accompagné des leaders de l'industrie des télécommunications, de l'automobile et de l'énergie dans l'exécution de projets BigData. Nous nous différencions par la maîtrise de l'ensemble de la chaîne de valeur de la data : en phase de conseil, en phase de « data discovery », en phase de « data factory » et surtout en phase de valorisation.

Mapping du champ d'intervention d'Ekimetrics dans le Big Data



Notre philosophie sur les projets Big Data : l'approche PoC – Proof of Concept

Dans le Big Data, nous observons un biais consistant à construire des systèmes pour ensuite chercher à les valoriser (DMP, clusters..), nous faisons le contraire :

- Nous avons une approche agile de test et d'apprentissage pour mettre à jour des initiatives réellement industrialisables.
- Nous partons de l'usage métier pour ensuite construire une architecture sur mesure.
- Nous n'avons pas de contraintes sur la technologie ni sur la méthodologie : nous choisissons la solution la plus adaptée en fonction des besoins exprimés par nos clients.

Ekimetrics : Les partners vous racontent leurs projets Big Data

Jean Baptiste Bouzige. PDG Ekimetrics

«Les APIs sont une source de données de plus en plus stratégiques et servent de socle à l'Internet des objets. Chez Ekimetrics nous suivons avec intérêt l'évolution des APIs et de l'Open Source depuis 2006. Nous avons été parmi les premiers à utiliser les outils Big Data, pour tirer profit de ces nouvelles sources d'enrichissement de données, notamment dans le cadre d'analyses du SocialWeb».

François Poitrine. Directeur Général, Responsable de la practice Big Data

«Nous accompagnons, depuis plusieurs années déjà, un leader de l'industrie automobile dans l'optimisation de sa stratégie Marketing et Data. Avec l'émergence des véhicules connectés et d'autres opportunités business liées au digital, les outils Big Data sont devenus centraux dans la valorisation des nouvelles sources de données à disposition. Nous avons pu déployer plusieurs projets de valorisation de ces données avec une approche pragmatique et cost effective de Test, Learn & Industrialize».

Quentin Michard. CEO Ekimetrics Londres

«Aujourd'hui, l'enjeu ne consiste déjà plus à valoriser sa propre donnée, mais à trouver des logiques de partenariat et de complémentarité pour avoir une offre complète.

Au sein d'Ekimetrics, nous maîtrisons les méthodologies statistiques et algorithmiques pour mener à bien de tels projets, la couche technologique apportée par le Big Data nous permet de passer de la théorie à la production».





COORDONNÉES

EXALEAD DASSAULT SYSTÈMES

10 place de la Madeleine 75008 PARIS

Tél **+33 (0) 1 55 35 26 26**

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

Mail **www.3ds.com/fr/contact**

CONTACT

**Axel
BRANGER**

*Sales Director EXALEAD
South Europe, Russia & Middle East*

EXALEAD Dassault Systèmes rompt avec 25 ans de tradition dans le développement d'applications, jusqu'ici centré sur des bases de données : sa technologie de recherche et d'indexation, couplée à des technologies Web, offre un accès unifié et un traitement haute-performance de l'information.

Les organisations ont désormais besoin de 3 niveaux d'expérience de l'information :

- Accéder agilement et rapidement à un grand volume de données dispersées et hétérogènes
- Croiser, réconcilier, contextualiser les données pour créer de nouvelles solutions métier et offrir de nouveaux services à leurs clients
- Aller plus loin dans l'analyse pour en extraire encore plus de valeur et transformer leur entreprise

EXALEAD Dassault Systèmes répond à ces défis et offre des applications personnalisées ou packagées, évolutives, parfaitement adaptées aux environnements actuels de Big Data et qui permettent de le transformer en vrais bénéfices pour l'entreprise.

Les solutions EXALEAD valorisent simplement et agilement vos données métiers pour décider mieux et agir vite. Elles adressent trois grandes problématiques :

- Interaction client : EXALEAD OneCall
- Digital assets : EXALEAD CloudView / EXALEAD OnePart
- Machine Data : EXALEAD CloudView

EXALEAD propose également la Web Mining Expérience qui s'appuie sur le moteur exalead.com et offre la puissance de la recherche Web et de l'analyse sémantique pour le développement d'applications à partir de données pur Web.



Porteur du projet

Yves-Marie Pondaven - Chief Technology Officer, Parkeon

Objectifs

Offrir aux villes et opérateurs de parkings un outil de pilotage (ventes, durées...) et faciliter la maintenance des horodateurs Parkeon.

Les défis étaient de rester performant malgré les volumes importants, d'être très simple d'utilisation et de permettre aux villes ou opérateurs de mieux comprendre les usages grâce à de nouveaux tableaux de bords intégrant de la cartographie.

Présentation

Analytics est un service développé par Parkeon grâce à la technologie EXALEAD CloudView. En indexant et analysant plus d'un milliard de tickets de stationnement émis par les horodateurs mis en place par une ville ou une communauté urbaine, Parkeon offre un outil de suivi à la fois global et détaillé du parc installé des machines pour permettre aux différents services d'agir, interagir et réagir pour améliorer le fonctionnement du stationnement dans une ville.

Les principales fonctionnalités d'Analytics

Horodateurs : Listes - Revenus des terminaux - Vue détaillée par terminal agrégeant de nombreuses informations (ventes, alarmes, états de la batterie...)

Parking : Evolution des revenus dans le temps, par zone - Evolution en fonction des jours ou des heures

Maintenance : Terminaux avec des problèmes de fonctionnements identifiés - Terminaux où des actions préventives sont nécessaires (changement de rouleau de papier...)

Transactions : Accès aux transactions par pièces ou cartes - Vues plus détaillées possibles sur les cartes pour filtrer par type de carte (Mastercard...)

Malgré les volumes de données importants (en centaines de millions pour certaines villes), l'outil devait rester aussi simple à utiliser que les sites grands publics, c'est à dire sans formation. EXALEAD apporte son savoir-faire et ses outils correspondants aux meilleures pratiques ergonomiques telles qu'un set de facettes pour filtrer les données, un champ de recherche unique sur toutes les pages configuré pour l'utilisation des requêtes en langage naturel et «trusted queries»... ce qui permet aux utilisateurs de prendre en main très facilement l'outil.

La visualisation et l'analyse sont facilitées grâce à des cartes manipulables par zones, des filtres par période et date, une création flexible de rapports, des exports Excel sous format standard... Les temps de réponse, les résultats clairs et en contexte augmentent l'efficacité des différents utilisateurs de l'application.

Aspect novateur (métier/usage)

L'application peut être utilisée par n'importe quel utilisateur de l'entreprise, quasiment sans formation. Chaque fonction peut accéder facilement aux données dont elle a besoin dans son activité quotidienne.

Analytics offre de grandes performances : une grande scalabilité, une indexation et un traitement puissant des données, un temps de réponse adapté aux challenges de la compétition...

Les gestionnaires de parcs peuvent ainsi accéder et traiter des milliards d'enregistrements à partir de tout type de sources... En quasi temps réel.

Valorisation observée de la donnée

L'exigence EXALEAD c'est de délivrer la bonne information, au bon moment, de la façon la plus intuitive possible. N'importe quelle fonction de l'entreprise, n'importe quel utilisateur peut naviguer, trouver, analyser et transmettre l'information utile. A l'instar de Parkeon, les organisations améliorent la capitalisation du savoir, l'efficacité de leurs équipes, le partage des connaissances. Parkeon l'a compris et, grâce à la technologie CloudView, a transformé des données machines basiques en application à haute valeur ajoutée offrant ainsi à ses clients un service performant pour gérer et anticiper son activité de gestion de parcs de parcètres. C'est un vrai facteur différenciant pour l'offre Parkeon.

Etapas à venir et démarche Big Data dans l'entreprise

Parkeon va continuer de faire évoluer l'outil en intégrant toujours plus d'informations et profiter des nouvelles fonctionnalités autour de la cartographie afin d'améliorer encore la solution.



COORDONNÉES

GFI INFORMATIQUE

La Porte du Parc - 145 boulevard
Victor Hugo, 93400 Saint-Ouen

Tél **+33 (0)1 44 04 50 08**

Mail alvin.ramgobeen@gfi.fr

CONTACT

**Alvin
RAMGOBEEN**

Group Practice Manager BI & Big Data

Le Big Data pour s'adapter et/ou se réinventer

L'avènement de l'ère numérique nous porte à considérer autrement le capital informationnel de tout un chacun. Notre monde évolue en permanence sur un rythme de plus en plus soutenu et porte des exigences de rentabilité exacerbées par une pression de compétition à l'échelle mondiale.

Dans tous secteurs d'activités, il nous faut nous réadapter en permanence. L'accès à la lecture de ce que nos données expriment nous donne les moyens de se redécouvrir et d'agir de façon nouvelle : prévisions de production de pièces détachées en fonction des comportements de conduite et des conditions climatiques pour un même modèle de voiture, mais aussi assurances différenciées en fonction de son attitude de conduite sont autant d'exemples parlant qui s'immiscent dans notre quotidien.

Le Big Data sert avant tout des enjeux business qui doivent être clarifiés en amont lorsque l'on traite un point dur identifié persistant malgré les investigations traditionnelles (ex. anticiper la fraude à l'export) ou en aval lorsqu'il s'agit de découvrir des 'potentiels leviers d'actions' dans le cadre d'une activité naissante (ex. des compteurs intelligents)



Gfi Informatique, architecte ensemble de vos projets Big Data

Parce qu'il n'est pas de "Mega Structure" qui se ressemble, parce que chaque projet Big Data est l'expression de votre essence et de votre culture, nous sommes à vos côtés pour relever ces défis avec une passion éclairée et pragmatique.

Gfi Informatique, acteur innovant en système d'information, propose une approche autour des thématiques Big Data à contre sens des normes du secteur. Après des travaux conjointement menés auprès d'Airbus depuis 2010 portant sur des expérimentations NoSQL dans des contextes analytiques et prédictifs, Gfi base sa stratégie Big Data en opérant un couplage fort entre le monde des startups et ses compétences propres. Seul acteur à proposer une offre de service à la pointe du savoir-faire français allié à l'approche industrielle d'un groupe polyvalent en intégration de système et en éditeur de logiciel, Gfi opère la réalisation de vos programmes Big Data en tant qu'architecte ensemble.

La posture de l'architecte ensemblier

Les demandes en Big Data sont extrêmement variées : chargement et stockage de données massives, traitement de données en temps réel, analyse de données textes ou son, élaboration de moteurs de recommandation, de corrélation, de prédiction... analyses récursives ou de réseaux (base graph), Data Visualisation spécifique, algorithmie avancée, etc. De plus, les demandes de réalisations Big Data sont souvent couplées à un existant IT et doivent également irriguer les systèmes opérationnels et les terminaux de travail qu'ils soient fixes ou mobiles. Cette diversité en besoins pointus amène Gfi, avec l'appui de son écosystème de startups, à se positionner en tant que garant de l'assemblage des compétences spécifiques pour chaque nature de besoin en Big Data.

L'offre Big Data de Gfi est multi entrée et permet de couvrir une grande majorité des attentes du marché sur les volets :

- Conseil outillé en exploration de la donnée qui permet en quelques semaines de faire émerger et de quantifier un use case afin de valider sa pertinence face à vos enjeux business
- Conseil en aide au choix de solutions et en architecture full & hybride Big Data
- Centre de service expérimenté pour la réalisation de prototypage Big Data
- Mise à disposition de Framework accélérant le déploiement en environnement de production de plateformes Hadoop sécurisées
- Réalisation de projet sur mesure Big Data, en mode agile/Dev Ops, intégrant l'exposition de services sur les terminaux digitaux

Fort du constat marché que derrière les termes de Datascientist ou d'Expert Hadoop, il est plus que difficile de trouver le profil compétent chez une seule et même personne, nous avons développé une approche en équipe pluridisciplinaire.

Cette équipe regroupe l'ensemble des profils spécialisés en Consultants Métier, Développeurs Big Data et Data Analysts, Experts BI & CRM analytique et Consultants Digital et Mobilité.

Afin de s'assurer de gagner en sécurité et performance, nous allions les compétences de nos équipes avec l'expertise pointue de startups en Datascience, Data Visualisation, Clustering Hadoop, Text mining, etc.

Illustration d'industrialisation d'un savoir-faire à la pointe

Nous co-déployons, avec notre partenaire ADALTAS, un outil simplifiant et accélérant la mise de place de clusters Hadoop multitenants sécurisés. Issu de près de 2 ans de travaux sur les plateformes d'un acteur majeur de l'énergie en Europe.

Les apports sont multiples et permettent notamment de :

- gérer plusieurs clusters et mettre à disposition des environnements de développement identiques à ceux de production
- permettre l'ajout de nouveaux composants en s'inscrivant dans un outil robuste orienté DevOps
- s'intégrer et/ou enrichir l'environnement sécurisé de l'entreprise de type Kerberos / OpenLDAP / Active Directory
- sécuriser tous les composants du cluster (Hive, HBase...)
- résumer les étapes de déploiement en une seule commande applicable à l'échelle du cluster
- assurer la haute disponibilité des composants & d'automatiser les sauvegardes

A ce jour, la distribution Hadoop supportée est l'HDP 2.1 d'Hortonworks

Les composants couverts sont HDFS et YARN, Hive, WebHCat, Oozie, Flume, Mahout, Pig, HBase, Hue, Hadoop, Sqoop et Zookeeper. D'autres composants tels que Storm, Spark et SolR mais aussi MongoDB seront prochainement intégrés.

L'offre Big Data de Gfi

Conseil outillé
en Data Exploration

Aide au choix de
solutions et Architecture
Big Data

Centre de service de
prototypage Big Data

Framework de déploiement
de production de plateformes
Hadoop sécurisées

Réalisation de projet sur
mesure Big Data avec
exposition de services digitaux





COORDONNÉES

HEWLETT-PACKARD FRANCE

20, Quai du Point du Jour
92100 Boulogne-Billancourt

Tél 0820 211 211
Mail bigdata.france@hp.com

CONTACT

**Florence
LAGET**

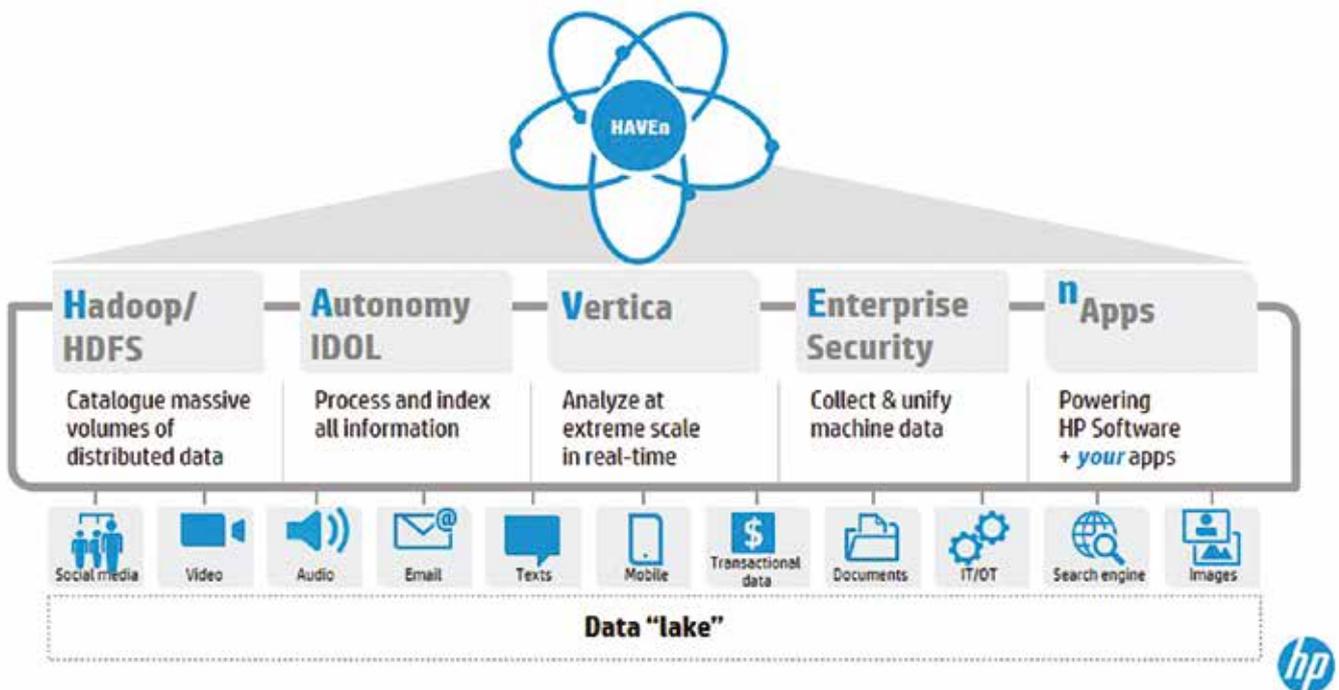
Directrice Big Data, HP France

Collecter, agréger et comprendre les informations issues des machines et des humains requiert une approche et des technologies nouvelles, pour connecter intelligemment des données de nature disparate, d'une manière comparable au fonctionnement du cerveau humain - faire travailler la technologie et non l'inverse.

HP décline sa stratégie Big Data au travers de **HAVEⁿ**, plateforme ouverte, intégrée mais modulaire, qui permet à nos clients de traiter 100% des informations à leur disposition.

- **H**adoop est le support idéal pour stocker de très grands volumes de données à faible coût.
- **A**utonomy IDOL est un moteur extrêmement puissant d'indexation et d'analyse des informations humaines : texte, mails, social media, vidéo, audio, etc.
- **V**ertica est une plateforme analytique temps réel (architecture en colonnes) capable d'ingérer les données et de traiter des requêtes complexes 50 à 1000 fois plus rapidement qu'une base de données traditionnelle
- **E**nterprise Security (Arcsight Logger) collecte et analyse en temps réel (1.000.000 d'événements par seconde) les logs des applications et des équipements (Badgeuse, Login...).
- Le « **n** » de HAVEⁿ représente les applications développées par HP et son écosystème partenaires éditeurs et intégrateurs.

L'ensemble de cette plateforme est supportée par des infrastructures innovantes et spécifiquement adaptées au Big Data. Nous sommes en mesure d'accompagner nos clients dans la conception de leurs cas d'usage, la mise en œuvre, l'exploitation ou l'externalisation de ces solutions.



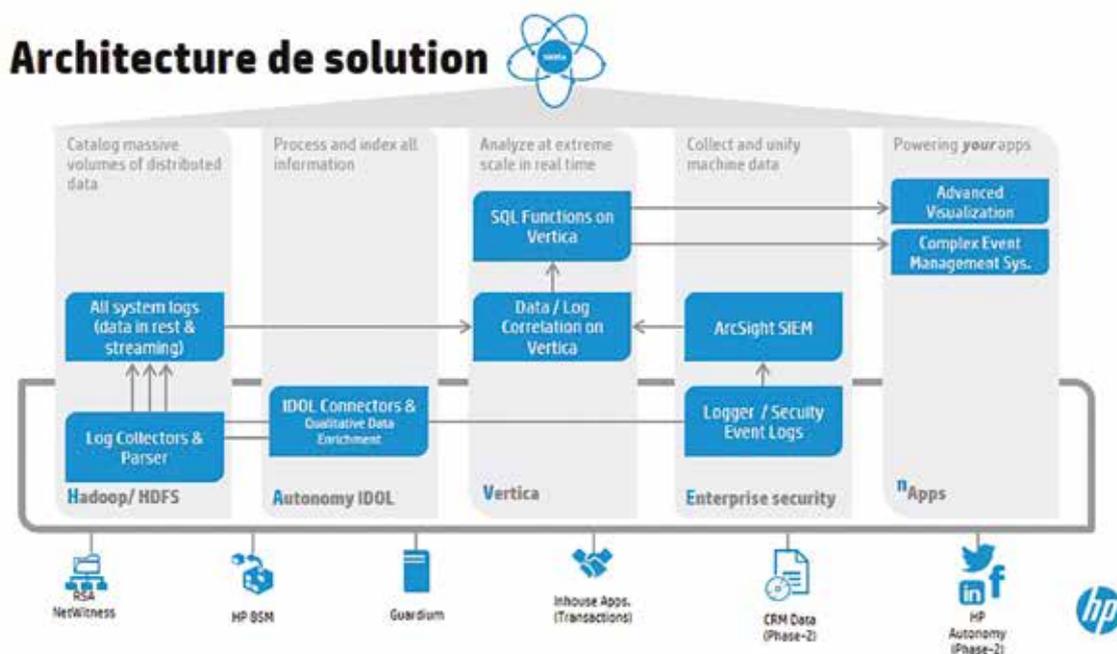
Sécurité, fraude et conformité

Cette banque européenne de 12 millions de clients, dispose de près d'un millier d'agences et gère plus de 100 milliards de capitaux en 2013. Elle détient une distinction pour la qualité de son centre d'appels et de ses plateformes internet, mobiles et réseaux sociaux.

La banque avait comme objectif d'accélérer l'analyse des risques de sécurité, la réponse aux requêtes et audits du régulateur, tout en se donnant les moyens d'améliorer l'efficacité de ses campagnes marketing.

Tous les types de données sont captés: applications métiers, DAB, agences, réseaux sociaux, logs qualité de service.

Le projet est implémenté sur la plateforme HAVen telle que décrite dans l'architecture ci-dessous :



Le projet se déroule en 3 phases qui répondent aux 3 enjeux auxquels doit faire face la société :

- **Conformité** - identifier les accès frauduleux aux comptes clients et réduire drastiquement le temps de réponse aux requêtes de conformité
- **Sécurité** - corréler les données entre les différents systèmes pour découvrir et mieux comprendre les potentielles failles de sécurité
- **Génération de revenus** - améliorer la connaissance clients pour optimiser l'efficacité des campagnes marketing et l'acquisition de nouveaux clients

Ce projet est innovant dans la diversité des sources et des formats de données utilisés.

Premiers résultats obtenus

- **2500x** plus rapide pour répondre aux requêtes complexes d'audit, de fraude et de sécurité
- **97%** de coût d'exploitation en moins pour opérer ces requêtes
- **28%** de TCO en moins par rapport aux solutions alternatives étudiées





COORDONNÉES

IBM FRANCE

17 avenue de l'Europe
92275 Bois Colombes

Tél +33 (0)1 58 75 00 00
Mail lucile.hyon@fr.ibm.com

CONTACT

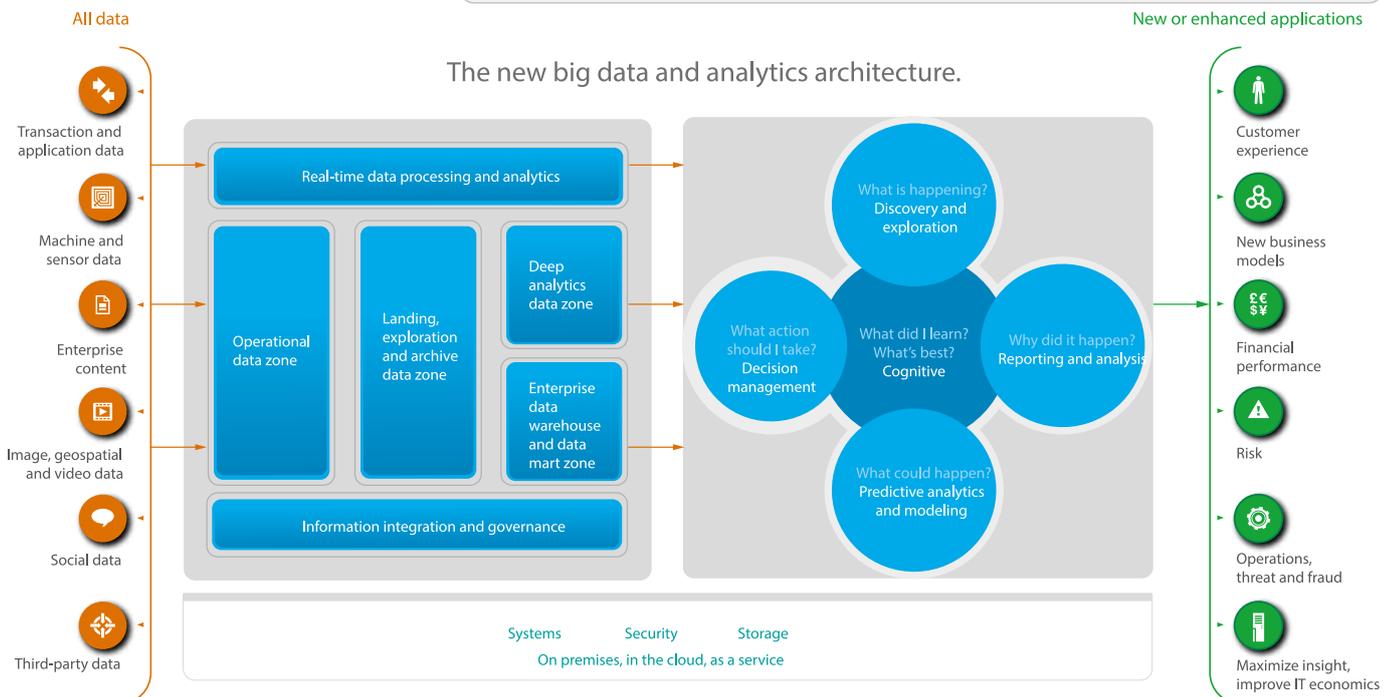
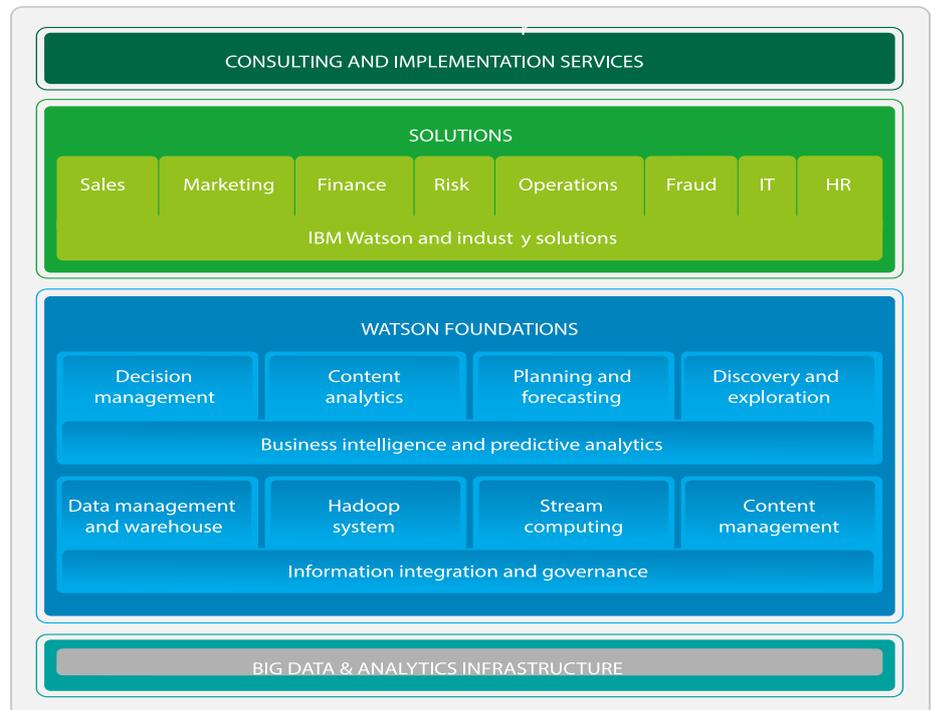
**Lucile
HYON-LE GOURRIEREC**

Senior Marketing Manager

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.

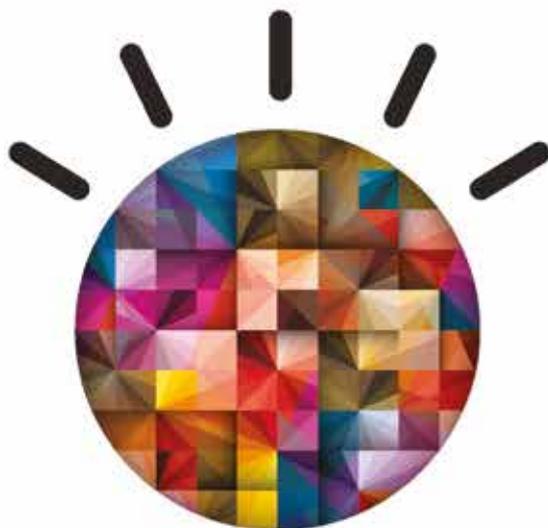


Big Data at the Speed of Business

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.

Gouvernance de l'information : intégration, gestion des données de référence, gestion du cycle de vie et sécurité.

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

<http://ibm.co/luxkBg1>



COORDONNÉES

KEYRUS

155 rue Anatole France
92300 Levallois-Perret - France

Tél +33 (0)1 41 34 10 00

Mail bigdata-keyrus@keyrus.com

CONTACT

**Xavier
DEHAN**

*Sales Director - Data Intelligence
Big Data Analytics*

Fort de près de vingt ans d'expérience au service de la Maîtrise et de la Valorisation de la Donnée, Keyrus aide les entreprises, en France et à l'international, 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 métier - 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 de plus en 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.

Afin d'atteindre ses objectifs d'excellence dans le domaine de la Data Science et en vue d'assurer l'efficacité maximale de ses équipes qui seront les garants de la bonne réalisation des projets qu'il conduira dans le domaine du Big Data, **Keyrus** a créé le 15 octobre 2014, avec l'École polytechnique, Orange et Thales, une Chaire pour former la prochaine génération de **Data Scientists**.

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.

www.keyrus.fr



COORDONNÉES

MICROPOLE

91-95 rue Carnot
92300 Levallois-Perret, France

Tél +33 (01) 74 18 74 18

Fax +33 (01) 74 18 74 00

Mail info@micropole.com

CONTACT

**Charles
PARAT**

Directeur recherche & Innovation

Micropole lance la 1^{ère} offre «Big Data as a Service» à destination des entreprises.*

L'offre **Data Science Platform** est le résultat d'une étroite collaboration entre de grands acteurs du marché : **Micropole**, entreprise internationale de conseil en technologies innovantes, spécialisée dans les domaines de la Transformation Digitale et du Pilotage de la Performance et **les entreprises de Cloud Souverain**, producteurs de puissance numérique pour les entreprises, au travers de leurs solutions d'infrastructure performantes et innovantes.

Inédite, l'offre de services **Data Science Platform** est destinée à démocratiser le Big Data en France en mettant à disposition de chaque entreprise sa plateforme Big Data sur-mesure. **Data Science Platform** intègre un studio de Data Science (qui combine des outils de gestion de données, de statistique, de visualisation et d'analyse prédictive), permettant d'intégrer, de manipuler, d'analyser et de présenter rapidement l'information issue de larges quantités de données.

Spécialistes métiers, les équipes Micropole accompagnent les équipes clientes tout au long du projet. Leur valeur ajoutée et leur vision fonctionnelle adaptée au Big Data, leur permettent de produire des cas d'usages métiers, gages de ROI et générateur d'avantages économiques ou concurrentiels pour l'entreprise.

Innovante et compétitive, l'offre **Data Science Platform** révolutionne les offres traditionnelles existantes sur le marché. Une réponse totalement nouvelle qui va permettre à toutes les entreprises de tirer parti de la puissance du Big Data.

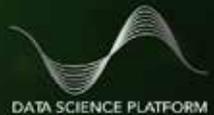
L'offre **Data Science Platform** se déploie en 24h au sein de nos infrastructures. Nos équipes vous accompagnent sur la définition de votre application Big Data, et tout au long du projet dans la prise en main des outils.

Pour des questions de sécurité et de transparence, notre offre est proposée sur les plateformes déployées au sein de Cloud Publics Souverains. Cela signifie que vos données sont protégées par la législation française, et localisées sur le territoire national.

*Pour des questions de réversibilité, l'offre **Data Science Platform** est basée sur des briques standards du marché et peut donc être installée sur le site de votre choix (on premise).

Vous démarrez sur le Big Data ?

Choisissez l'offre qui vous ressemble*.
Nous bâtissons votre solution, vous vous concentrez sur les usages.



Outil : Studio de Data Science

StORAGE
100 Go

Configuration
Cluster de 5 nœuds
(1 nœud principal, 3 nœuds de calcul, 1 nœud de services)

Volume de données
Entrantes : Illimité
Sortantes : 1 To

StORAGE
100-250 Go

Configuration
Cluster de 7 nœuds
(1 nœud principal, 5 nœuds de calcul, 1 nœud de services)

Volume de données
Entrantes : Illimité
Sortantes : 1 To

StORAGE
250-500 Go

Configuration
Cluster de 9 nœuds
(1 nœud principal, 7 nœuds de calcul, 1 nœud de services)

Volume de données
Entrantes : Illimité
Sortantes : 2 To

StORAGE
Jusqu'à 1 To

Configuration
Cluster de 11 nœuds
(1 nœud principal, 9 nœuds de calcul, 1 nœud de services)

Volume de données
Entrantes : Illimité
Sortantes : 2 To

Quelle que soit la taille de votre projet, Data Science Platform vous permet de bénéficier d'une offre de services intégrant le conseil technique et fonctionnel, la formation aux outils de Data Science, une plateforme sécurisée déployée sur un Cloud Souverain, ainsi que d'une tarification mensuelle unique.

Big Data et connaissance client : projet dans le Retail

Historiquement les données présentes au sein des entreprises sont organisées dans des silos d'information, le Big Data permet de décloisonner cette information et de l'exposer différemment.

Il est alors possible, sur un sujet métier tel que la connaissance client, de créer une vision 360° des données, en mettant en corrélation les données issues de différents canaux (point de vente, sites internet, centre d'appels, etc.), de se construire une vision inédite du client et d'arriver ainsi à visualiser le parcours client omnicanal complet, plus rapidement.

Pour notre client, un acteur reconnu de la distribution spécialisée, les objectifs fixés dans le cadre du projet étaient les suivants :

- Accélérer l'omnicanalité pour accroître le chiffre d'affaires de l'enseigne ;
- Mieux analyser les comportements web-to-store, dans les deux sens ;
- Prendre des décisions stratégiques, de manière instantanée, qui correspondront mieux aux attentes de leurs clients ;
- Découvrir des types de comportements pour anticiper l'effet « churn » ;
- Proposer de nouveaux services ou produits avant même que le client ait conscience de son besoin, générer de la valeur pour le client ;
- Être plus réactif dans le traitement de l'information, quelle que soit sa forme ;
- Industrialiser la démarche de réconciliation des données.

La démarche de mise en œuvre de Micropole autour du sujet de l'omnicanalité, consiste dans un premier temps à intégrer au sein d'un référentiel de données unique Big Data, l'ensemble des informations issues de tous les canaux de notre client, afin de créer une vision décloisonnée.

Dans un second temps, les données sont mises en cohérence en vue de leur traitement dans des solutions dites de « DataScience », dans le cadre d'une approche de type Datalab.

La construction du Datalab s'appuie sur une méthode Agile. Les étapes menant à la validation du périmètre Datalab sont :

- Le cadrage des besoins ;
- Le cadrage des données ;
- La présentation des résultats.



Comment une approche agile a aidé notre client à exploiter rapidement la valeur de ses données.

Le cycle de « cadrage du besoin » permet d'avancer rapidement sur les besoins à remplir et le travail des sources associées. L'avantage est que les découvertes faites pendant l'étude détaillée des données sont répercutées sur les besoins initiaux. Ce processus est répété autant de fois que nécessaire.

À l'issue de cette phase, les cycles de « cadrage des données » débutent. Une phase de réalisation technique a lieu entre chaque point de cadrage et chaque présentation.

Pendant la présentation des résultats, les équipes fonctionnelles peuvent suivre et prévalider les réalisations. Les équipes techniques peuvent remonter des problèmes rencontrés pendant les développements. Dans les 2 cas, les points discutés sont inclus dans le cadrage des données. Il est possible de remonter au cycle précédent si la situation l'exige.

La phase de validation du périmètre clôture le processus. Elle prend un temps minime, compte tenu du travail effectué en amont.

Les résultats

Sur une phase d'expérimentation de 3 mois, il ne nous a fallu que 3 semaines sur un gisement de données de plusieurs centaines de gigaoctets, pour mettre en évidence les premiers comportements clients et définir les parcours clients associées.

L'usage d'un logiciel de Data Science (solution combinant des outils de gestion de données, de statistique, de visualisation et d'analyse prédictive), a permis de modéliser facilement les parcours clients, d'industrialiser le processus et de présenter les résultats en utilisant une solution de Data Visualisation (l'utilisation de techniques visuelles et interactives pour valoriser ses données).

À l'issue de la phase d'expérimentation, la solution a été déployée en production. Les outils de Data Science utilisés dans le cadre du Datalab ont quant à eux naturellement intégré le panorama des outils standard du service Marketing pour la production des analyses basées sur la plateforme Big Data.





COORDONNÉES NEO TECHNOLOGY

Tél +33 (0)8 05 08 03 44

Mail Cedric.Fauvet@neotechnology.com

CONTACT

**Cedric
FAUVET**

Business Development France

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 connexions 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 vous rencontrez 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 Walmart, SFR, Cisco, eBay et Glowbl.

(graphes) – [:SONT] –> (partout)

- **Événements:** neo4j.com/events
- **Training:** neo4j.com/graphacademy
- **Neo4j dans votre société:** neo4j.com/brownbag
- **Télécharger Livre:** neo4j.com/books
- **Nos clients:** neo4j.com/customers
- **Télécharger le produit:** neo4j.com/tryneo4j
- **Suivez-nous sur Twitter:** @Neo4jFR
- **Rencontrer la communauté:** meetup.com/find/?keywords=Neo4j

**Venez rencontrer
à Paris l'équipe
française de Neo4j
à l'occasion du salon
Big Data 2015,
stand 404!**

LA SOCIETE

Spécialisée depuis 2009 dans la livraison rapide d'achats réalisés sur internet, c'est tout naturellement que la société londonienne Shutl a été rachetée par eBay, qui en a fait sa plateforme nouvelle génération pour redéfinir le e-commerce en accélérant et en simplifiant la livraison des commandes passées sur Internet ou appareil mobile. eBay élimine ainsi le plus gros obstacle entre vendeurs et acheteurs en ligne en proposant une livraison le jour même.

LE DÉFI

Le service de livraison le jour même de Shutl s'est développé de manière exponentielle et couvre aujourd'hui jusqu'à 85 % du Royaume-Uni. Sa plateforme de services a dû être entièrement repensée afin de prendre en charge l'explosion du volume de données et les nouvelles fonctionnalités. Le problème était que les jointures MySQL utilisées auparavant ont créé une base de codes trop lente et complexe, proposant des temps de réponses à requêtes de 15 minutes pour certaines.

Développeur senior pour eBay, Volker Pacher et son équipe pensaient qu'en ajoutant une base de données de graphe à l'AOS et à la structure de services, il serait possible de résoudre les problèmes de performances et d'évolutivité.

POURQUOI NEO4j?

Le choix s'est porté sur Neo4j, considérée comme la meilleure solution pour sa flexibilité, sa vitesse et sa simplicité d'utilisation. Fort de son modèle de graphe à propriétés parfaitement compatible avec le domaine modélisé, et avec sa base de données ne nécessitant pas de schéma, Neo4j a simplifié son extensibilité et a permis d'accélérer la phase de développement.

Cypher a de son côté permis d'exprimer les requêtes sous une forme très compacte et intuitive, ce qui a encore accéléré le développement. L'équipe a ainsi pu exploiter le code existant à l'aide d'une bibliothèque Ruby pour Neo4j prenant également en charge Cypher. Comme aime à le dire Volker Pacher : « Notre système fonctionne sur 7 lignes de Cypher ».

LES AVANTAGES

Basée sur jRuby, Sinatra, MongoDB et Neo4j, la nouvelle plateforme garantit des transactions rapides avec des performances relativement constantes. Par ailleurs, son modèle de données permet aux requêtes de rester localisées à l'intérieur de leurs portions respectives du graphe.

« Notre solution Neo4j est littéralement des milliers de fois plus rapide que la solution MySQL précédente, avec des requêtes qui nécessitent de 10 à 100 fois moins de code. Dans le même temps, le côté flexible de Neo4j nous a permis d'implémenter l'ensemble en à peine une année et il nous permet encore d'ajouter des fonctionnalités jusqu'ici impossibles, ce qui permettra l'accompagnement du service eBay Now dans ses futurs développements. »

– Volker Pacher, eBay

Aujourd'hui et grâce à Neo4j, la plateforme Shutl orchestre les livraisons entre les boutiques, les coursiers et les acheteurs 24 h/24 et 7 j/7. Les envois s'effectuent directement depuis les points de vente. Le service organise la collecte et la livraison des articles selon les préférences des clients, généralement dans un délai de 2 heures, ou dans une fenêtre de livraison d'une heure choisie par les clients. Le résultat : une prestation innovante qui améliore la qualité de service pour les clients ainsi que la productivité des partenaires revendeurs et transporteurs. Tous les acteurs sont gagnants : les clients disposent de plus de choix pour la livraison, les coursiers ne perdent plus de temps à attendre et les boutiques peuvent proposer des services supplémentaires à leurs clients sur Internet.





COORDONNÉES PENTAHO

Paris - France

Tél **+33(0)6 38 38 06 33**
 Fax 1 407 517-4575
 Mail ebrochard@pentaho.com

CONTACT

**Emmanuel
BROCHARD**

Responsable France

Pentaho construit le futur de l'analytique. Sa plate-forme analytique intégrée, moderne et complète est conçue pour les diverses exigences de l'analyse d'entreprise, et notamment celles du Big Data.

Pentaho Business Analytics pour exploiter plus rapidement et à moindre coût les Big Data grâce à :

- Pentaho Big Data Analytics Center
- Une continuité totale depuis l'accès aux données jusqu'à la prise de décision
- Un développement visuel et une exécution distribuée
- Une analyse interactive et instantanée : aucun codage ni compétences ETL requis

Instaview : Cette première application interactive et instantanée 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
- Une meilleure productivité de l'équipe informatique grâce à la réutilisation des compétences existantes ; sans formation particulière pour les informaticiens, analystes d'entreprise et scientifiques spécialistes des données
- Une protection contre les risques : fonctionne avec toutes les principales distributions Hadoop, les bases de données NoSQL et les entrepôts de données traditionnels
- 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

Analyse et visualisation NoSQL : Pentaho supporte nativement les bases de données NoSQL les plus utilisées, notamment Apache Cassandra/DataStax, HBase, MongoDB/10gen et HPCC Systems, et offre une productivité de développement et une performance d'exécution multipliées par 10 ou par 15 par rapport au codage manuel.

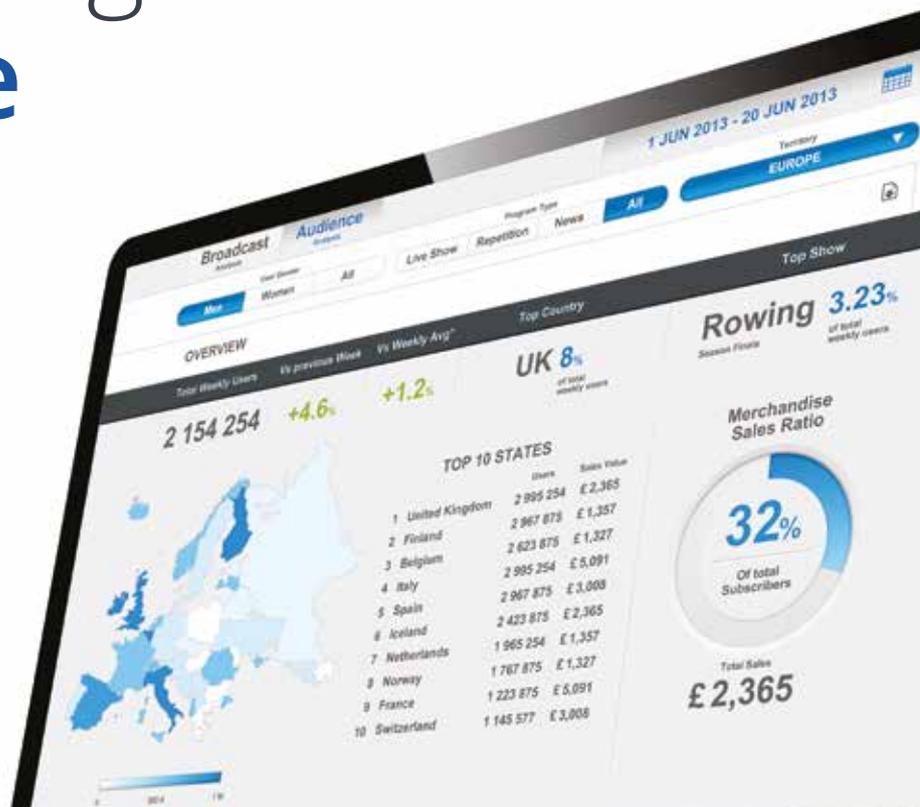
Big Data Analytics Center : www.pentahobigdata.com pour en savoir plus sur Pentaho et Big Data. Pour évaluer gratuitement "Pentaho Business Analytics" : www.pentaho.fr/download.



Bring Your Big Data to Life

Big Data Integration and Analytics

Optimisation d'Hadoop et plus encore.
 Découvrez comment sur pentaho.fr



Monétiser les Big Data des banques de détail commence par une meilleure expérience client

Par Emmanuel Brochard, responsable de Pentaho en France

Les banques de détail commencent à considérer les Big Data comme des actifs prometteurs à même de générer de nouveaux flux de revenus. Elles ont besoin de se projeter au-delà de la vente de produits complémentaires et d'examiner comment l'analyse des Big Data peut rétablir la confiance et améliorer l'ensemble de l'expérience client.

Plus que jamais, les banques de détail ont besoin d'utiliser la valeur de leurs données pour créer des offres plus concurrentielles. Simultanément, le pouvoir d'achat des ménages français à la baisse ou au mieux stable depuis 2007 conduit les consommateurs à utiliser des ressources telles que des sites comparatifs de services bancaires ou des logiciels de gestion pour être plus actifs dans le suivi de leurs finances. A divers degrés, les banques commencent à intégrer dans une vue unique les détails de leurs produits phares, tels que les comptes d'épargne et de crédits immobiliers. Certaines déjà inquiètes pour leur réputation, ont eu peur de pousser plus loin leur intégration par crainte que leurs données ne s'échappent des traditionnels silos informatiques et menacent la confidentialité et sécurité des données. D'autres banques craignent qu'une plus grande transparence et un apport de plus d'informations aux clients les incitent à aller voir ailleurs.

L'intégration et l'analyse des Big Data pour concevoir une proposition de valeur commerciale

Soyons clairs, monétiser les Big Data ne se limite pas à vendre plus facilement des assurances additionnelles sur des crédits immobiliers. Il s'agit d'offrir des services et expériences personnalisés exceptionnels qui conduisent à terme à une proposition de valeur commerciale, ou de maximiser la valeur totale d'un client au travers de toutes ses interactions et transactions. En déterminant les différents besoins des clients et en intégrant cette connaissance dans toutes ses transactions, une banque devrait être capable d'améliorer le service client et la fidélisation ainsi que d'augmenter sa propre rentabilité en optimisant ses prix en fonction de la valeur client. Aussi simple que cela puisse paraître, ceci a échappé aux banques de détail pendant des années parce qu'elles ont établi leurs prix sur des suppositions trop génériques et se sont davantage concentré sur la croissance des revenus plutôt que sur l'enrichissement de la valeur client. Contrairement à d'autres marchés tels que les fournisseurs d'accès internet qui offrent des services à leurs clients, les banques de détail sont les seules à détenir des données concrètes sur ce que leurs clients ont exactement dépensé, quand et à quelle fréquence. Elles disposent donc des meilleures données pour établir des profils de clients détaillés et concevoir en conséquence des produits et services sur mesure.

Fournir une vue des clients à 360° à double usage

Le service fondamental qui profite à la banque de détail et à ses clients est la capacité à fournir cette vue intégrée à 360 degrés du portefeuille complet de chaque client décrit précédemment. Cette vue nécessite de fonctionner de deux façons ! En effet, les banques devraient avoir cette vue holistique de leurs clients, mais simultanément les clients demandent la même visibilité de leurs produits et services. Cela implique de pouvoir utiliser un seul mot de passe pour se connecter et de tout visualiser au travers d'un tableau de bord clair et simple. Les clients préférant traiter en face à face ou par téléphone devraient obtenir une expérience intégrée similaire sans être transférés vers d'autres services utilisant des systèmes informatiques non intégrés. Cette vue à 360° est le prérequis qui permettra de monétiser les données avec plus de profit. Comme évoqué plus haut, de nombreuses banques craignent que leurs données s'échappent de leurs silos sécurisés si elles tentent de les intégrer à de nouvelles applications qui améliorent l'expérience des clients ou des responsables de la banque. Cependant, de nouveaux outils d'intégration de données sophistiqués permettent aux banques de mélanger les données à la source sans devoir les transférer au préalable dans une « zone de transit ». Ces mêmes outils permettent également d'établir des règles simples qui garantissent que les données soient traitées selon les règles de gouvernance des données locales ou européennes, ce qui éliminent virtuellement les risques de failles de conformité et de sécurité.

Mesurer la perception des clients

La plupart des banques mesurent la satisfaction des clients un peu au hasard, d'une façon dépassée. Elles raffolent d'enquêtes où leurs clients sont invités à compléter un formulaire après chaque transaction. Une façon plus révélatrice et moins intrusive de connaître la perception des clients est d'associer moins fréquemment des enquêtes plus détaillées avec des contenus des réseaux sociaux et internet que les clients publient de leur plein gré. Des outils analytiques pour les Big Data permettent de mixer ces différentes sources de données pour aider les banques à concevoir en permanence de meilleurs services.

Les banques de détail ont certes un travail considérable à accomplir pour bâtir les infrastructures et transformer leurs cultures de sorte qu'elles puissent fournir des vues intégrées des clients à 360° qui serviront de fondation à leur proposition de valeur commerciale. Néanmoins, elles disposent aussi d'atouts considérables. Par exemple, lorsque des clients ont suffisamment confiance en leur banque pour leur permettre de partager des données avec leurs commerçants préférés, les données pourraient avoir une valeur monétaire très élevée. Ou encore, la tendance omni-canal des services bancaires qui se développent via les « DAB » intelligents, murs d'écran tactiles, applications et kiosques mobiles, sont autant d'opportunités de créer des expériences de services sophistiqués pilotés par des données intégrées de haute qualité sont pratiquement sans limite. Cependant, les banques doivent déterminer tout d'abord comment monétiser leurs propres données avant même de penser à les commercialiser à l'extérieur !





COORDONNÉES

PIVOTAL / EMC

80 Quai Voltaire - CS 21002
95870 Bezons - France

Tél +33 (0) 1 39 96 96 37

Mail info@pivotal.io

Farid
AADIM

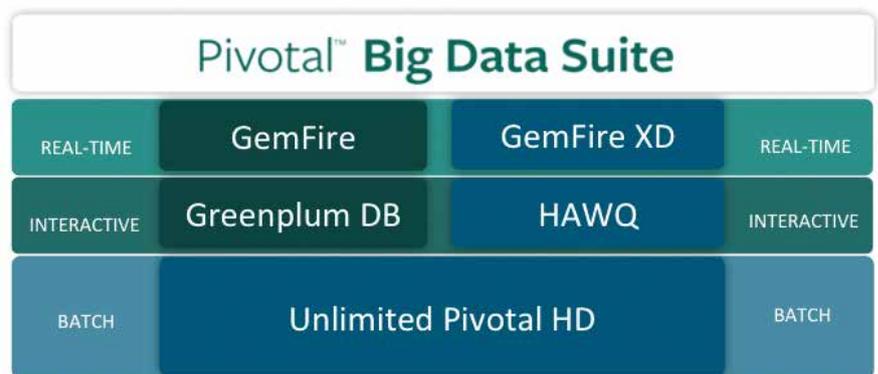
EMEA Inside Sales Manager

Pivotal : Extraire la vraie Valeur Business de toutes vos Données !

La prise de décision en temps réel est essentielle au succès de l'entreprise. Pourtant, les données de votre entreprise continue de croître de façon exponentielle d'année en année, ce qui rend l'analyse plus difficile. Pour transformer les données structurées et non structurées en intelligence décisionnelle, votre entreprise a besoin d'une intelligence efficace pour exploiter le Big Data. Pivotal propose, au travers de Pivotal Big Data Suite, une solution logicielle de Business Data Lake permettant le stockage et le polymorphisme de traitement de tout type de données. General Electric, American Express, Axel Springer, NTT, Monsanto et SouthWest Airlines comptent parmi nos références.

Pour plus d'information sur Pivotal Big Data Suite :

<http://www.pivotal.io/big-data/pivotal-big-data-suite>



Pivotal : Accélérateur d'innovation

Pivotal se positionne comme moteur de votre innovation. En complément de cette solution Data, la stratégie est complétée par 2 composants majeurs : l'Agilité et la Rapidité.

Avec Pivotal Labs, venez développer avec nous, votre application mobile dans notre Lab Agile comme Twitter, Facebook, LinkedIn ou Groupon l'on fait avant vous.

Le PaaS est l'élément essentiel de votre chaîne DevOps qui vous permet de réduire à la fois le temps de développement de votre application (de 6 mois à 6 semaines d'après Warner Music) et le temps de vos 'Opérations' (plus de 90% de gain d'après Rakuten). Pivotal Cloud Foundry est la solution PaaS adaptée au monde de l'Entreprise grâce son écosystème unique (Fondation Open Source).

Pour plus d'information sur le PaaS :

<http://www.pivotal.io/platform-as-a-service/pivotal-cf>

Pour plus d'information sur l'Agilité offerte :

<http://www.pivotal.io/agile>

A propos de Pivotal :

Né d'une spin-off de EMC et de VMware, Pivotal est un éditeur de logiciel offrant une approche technologique moderne afin de permettre aux Entreprises d'innover dans de nouveaux business. A l'intersection du Cloud, Big Data et du développement Agile, Pivotal permet de mieux utiliser ses données, d'accélérer la création d'application et de réduire les coûts, tout en proposant à ces Entreprises, la vitesse et la démultiplication d'exécution dont elles ont besoin pour être plus compétitifs.

Un opérateur telecom fédère qualité réseau et qualité client avec Pivotal Big Data Suite

La solution Pivotal comprend : Pivotal RTI Spring RabbitMQ et Pivotal Big Data Suite - GemFireXD, Pivotal HD (Hadoop) et Pivotal HAWQ

Notre client est l'un des 10 plus grands opérateurs télécoms mondiaux. D'après les études indépendantes sur la qualité des réseaux et l'expérience client, l'optimisation réseau et les améliorations sur les processus métiers sont structurants et sont devenus réalité grâce à notre solution Pivotal Big Data Suite.

Dans ce contexte, nous avons 2 challenges :

- Réduire les temps de latence, qui impactent la qualité de service :
 - qualité des appels (interruption d'appel, grésillements sur la ligne),
 - qualité du réseau (impossibilité de se connecter au réseau),
 - l'expérience client.
- Améliorer les actions/campagnes marketing par le biais d'analyse de données en temps-réel, sur les statistiques du trafic réseau.

Notre Solution a permis :

- D'augmenter l'ARPU (revenue moyen par utilisateur) grâce à la géolocalisation des clients finaux en temps-réel.

Comment créer de nouveaux services innovants et produits qui tirent parti de données essentielles venant du réseau ? Le premier projet en production de notre client, au Royaume-Uni est un projet de Géo-Marketing. A l'aide du temps réel, dès qu'un client rentre dans une zone prédéfinie « GeoFencing », un message marketing est automatiquement envoyé au client, ce qui permet d'éviter le spam.

- L'amélioration de la qualité de service du réseau : Procéder à des ajustements basés sur des événements en fonction de la puissance. Augmenter le débit en quelques minutes au lieu d'attendre le lendemain. Etre à même de gérer en temps-réel le réseau en fonction de la fréquentation de celui-ci, afin d'améliorer l'expérience client.

Notre client souhaitait connaître en temps-réel, les appels interrompus, les temps de latences pour télécharger des pages webs afin d'être capable d'augmenter rapidement la bande passante si nécessaire d'un endroit à l'autre en fonction des flux de personnes sur les zones géographiques.

- De mieux comprendre où les investissements doivent être fait en fonction de la base de données du réseau.

Dans cette industrie, le plus gros investissement (Capex) reste l'infrastructure réseau. Il est donc important de bien comprendre son réseau en fonction de ses utilisateurs. Notre solution leur permet de mieux cibler leur investissement 4G pour qu'il soit le plus efficace possible avec le meilleur retour sur investissement possible.



Composants	Description
Spring	Spring helps development teams build simple, portable, fast, and flexible JVM-based systems and applications. Spring is the most popular application development framework for enterprise Java.
RabbitMQ	Increase application speed, scalability and reliability by delivering asynchronous messaging to applications. A message broker for applications.
GemFire	Pivotal GemFire is a distributed data management platform designed for many diverse data management situations, but is especially useful for high-volume, latency-sensitive, mission-critical, transactional systems. <ul style="list-style-type: none">• Scale-Out Performance• Consistent database operations across globally distributed applications• High availability, resilience, and global scale• Powerful developer features• Easy administration of distributed nodes
GemFireXD	<ul style="list-style-type: none">• Data Consistency with Cloud Scalability• Extreme Performance and Continuous Uptime with Predictable Performance• Data Aware Parallel Function Execution• Data Stream with Enterprise Data Store Correlation• Relational Technology Based on Apache Derby• Referential Integrity• Standards-Compliant Powerful Querying Engine - ANSI• Data-Dependent and Data-Aware Java Stored Procedures
Pivotal HD (Hadoop)	<ul style="list-style-type: none">• Simple and Complete Cluster Management: Command Center• Big Data + Big Computing: GraphLab on OpenMPI• Hadoop In the Cloud: Pivotal HD Virtualized by VMware• Spring Data: Build Distributed Processing Solutions with Apache Hadoop
HAWQ	<ul style="list-style-type: none">• A Fast, Proven SQL Database Engine for Hadoop• Big Data Analytics Capability and Productivity• Parallel Data Flow Framework - Dynamic Pipelining™• Extension Frameworks with Hbase, Hive, etc - PXF• Big Data Analytics Capability and Productivity - MADlib





**COORDONNÉES
QLIK**

93, avenue Charles de Gaulle
92200 Neuilly-Sur-Seine

Tél **+33 (0) 1 55 62 06 90**
Mail infoFR@qlik.com

CONTACT

**Stéphane
BRIFFOD**

Director Presales

Qlik (NASDAQ : QLIK), leader en matière de Data Discovery, propose des solutions intuitives de visualisation en libre-service et d'analyse guidée. Environ 33 000 sociétés s'appuient sur les solutions de Qlik pour extraire des informations essentielles de sources diverses et explorer des relations invisibles qui donnent naissance à des idées novatrices. Le groupe, dont le siège se trouve aux Etats-Unis, possède des bureaux dans le monde entier et compte plus de 1 700 partenaires dans plus de 100 pays. En France, la filiale a été créée en 2007 et compte parmi ses clients des sociétés telles que 3M, la Société Générale, le CNRS, Bayer Healthcare, Meetic, Essilor, RueDuCommerce, le Groupe SEB, etc.

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

Le modèle orienté applications des solutions Qlik est compatible avec les solutions de BI existantes et offre une expérience basée sur l'immersion et la collaboration et intégrant une dimension collaborative et mobile.



Qlik rend le Big Data accessible aux utilisateurs métiers.

Libérez votre intuition.

www.qlik.com/fr



UNE NOUVELLE APPROCHE DU DECISIONNEL

A l'inverse des solutions de BI traditionnelles, basées sur une approche descendante et des chemins d'analyses prédéfinis, les solutions Qlik de Data Discovery donnent toute liberté d'analyse à l'utilisateur métier, dans le respect d'une gouvernance établie par l'IT. La puissance et la richesse de la technologie associative de Qlik apportent :

- Des perspectives à la portée de tous : Avec les solutions Qlik, tout le monde peut créer des perspectives à partir de ses données métier.
- Mobilité : Retrouvez vos applications sur un appareil mobile. Vos données métiers sont disponibles partout et tout le temps.
- Des analyses instantanées : Les solutions Qlik réduisent considérablement les délais nécessaires pour générer des perspectives et obtenir des résultats instantanés.
- Modèle inspiré des applications grand public : Les solutions Qlik permettent de développer et de déployer rapidement des applications simples, dédiées à un sujet ou généralistes, et intuitives qui sont faciles à réutiliser.
- Remixage et réassemblage : Les utilisateurs métier peuvent facilement réassembler des données dans de nouvelles vues et créer de nouvelles visualisations pour une meilleure compréhension.
- Réseaux sociaux et collaboration : A tout moment, les utilisateurs peuvent partager des perspectives et résultats au sein de l'application ou par l'intermédiaire de plateformes de collaboration permettant de créer une communauté d'utilisateurs.

Les solutions Qlik peuvent être déployées sur site, en mode cloud, sur un ordinateur portable ou un périphérique mobile, pour un seul utilisateur ou bien à l'échelle de grandes multinationales.





COORDONNÉES SAP FRANCE

Tour SAP - 35 rue d'Alsace
92300 Levallois-Perret

Tél 0805 800 023
Mail jean-michel.jurbert@sap.com

CONTACT

**Jean-Michel
JURBERT**

Directeur de Marché Solutions SAP
HANA, BI - Big Data



Didier Mamma

Leader des applications d'entreprise, SAP (NYSE : SAP) fournit aux entreprises de toutes tailles et de tous secteurs les moyens logiciels d'une meilleure gestion. Du back office à la direction, de l'entrepôt à la vente, quel que soit le terminal d'utilisation (ordinateur, terminal mobile), les applications SAP permettent aux collaborateurs et aux organisations de travailler plus efficacement ensemble et de s'appuyer sur des tableaux de bord personnalisés afin de conserver leur avance dans un contexte concurrentiel. Les applications et services SAP offrent à plus de 281 000 clients la possibilité d'être rentable, de s'adapter continuellement et de croître durablement

Didier Mamma

Nouvelle approche et vision stratégique : comment le Big Data bouleverse notre économie

Didier Mamma est Vice-President of Big Data Sales EMEA chez SAP. Fasciné par le « pouvoir » des données, il tient à en étudier la portée stratégique et les grands blocs fonctionnels. Le Big Data n'est certainement pas une nouvelle technologie mais bien un moyen de repenser son avantage concurrentiel. Comme l'a remarquablement décrit Jeremy Rifkin nous sommes entrés de plain-pied dans la 3^e révolution industrielle, digitale et hyperconnectée. On ne peut plus appréhender le présent avec les paradigmes du passé.

La digitalisation remet également en cause tous les modèles économiques. Amazon, Netflix, Nest, eBay, Apple, Critéo, Houzz, Google... sont quelques exemples d'entreprises qui sont en train d'établir un nouvel ordre économique trans-industries.

La question demeure dans la capacité des entreprises à capter et traiter cette multitude d'informations désormais à leur disposition. Comment imaginent-elles l'avenir de leur business ? Dans ce contexte, la répartition des dépenses des principaux groupes du CAC40 semble alors déséquilibrée. En moyenne, ils dédient 72% de leurs investissements à la « maintenance » de leur système d'information. Dramatiquement, une faible partie des dépenses est allouée à l'innovation.

Pour libérer une part de cet important budget, l'idée de SAP est de simplifier l'existant. C'est l'objectif de SAP HANA. La Plateforme HANA est capable d'accueillir et de traiter la variété des données de l'entreprise, analytiques ou transactionnelles, structurées ou non. Finie la frénétique duplication des données et l'accroissement du plat de « spaghettis » informatique. Parmi les 3V qui définissent le Big Data, on se focalise souvent sur le V de volume, or c'est la notion la plus simple à gérer. La complexité des traitements ou calculs, la vitesse des opérations suffisent à justifier une approche Big Data. Comme le dit Didier, « on peut faire du Big Data avec des gigabites ! ». En définitive, le Big Data « en mouvement » nécessite quatre dimensions. La première est liée à la puissance de calcul pur et rapide. La seconde, la capacité d'extraire l'information « cachée » des données. La troisième rend le résultat des analyses grâce à la visualisation. Et la dernière dimension est l'expert qui connaît la nature de la donnée et sa valeur potentielle.

Frédéric PUCHE

Le big data, richesse d'applications et efficacité d'utilisation

Frédéric Puche est en charge des innovations chez SAP France. Il a organisé et animé les présentations dans le cadre du SAP Big Data Tour en Juin dernier. Rendre concret le Big Data et en démontrer la richesse des domaines d'applications, c'est son objectif. Il nous présente ici quelques beaux projets menés chez SAP. On associe souvent Big Data à volume de données et stockage adéquat. Mais le Big Data c'est aussi et surtout la capacité de capter les données pertinentes issues des divers canaux, des média sociaux aux objets connectés.

HANA, la solution Big Data de SAP, comporte trois volets qui répondent à cet objectif. Le premier est une base de données en mémoire. Le second, un moteur d'analyse sémantique et de sentiment puissant, capable d'effectuer des prédictions dynamiques. Le troisième et dernier, un moteur de règles systématisant des alertes et apportant agilité à l'entreprise. L'avantage ? Centraliser l'intégralité des données sur une même plateforme, de-dupliquer les intermédiaires et faciliter les process. Avant, on faisait du temps réel, mais chaque Business Unit disposait de ses propres données et outils. Avec Hana, on fait du Big Data, et les projets innovants naissent. L'Executive Briefing Center de SAP en France est justement là pour rendre compte concrètement de ces projets. Dans le sport, un domaine qui a récemment animé l'actualité du Big Data, l'introduction d'objets connectés et l'analyse des flux vidéo apportent une nouvelle approche. La Mannschaft par exemple, stocke, décortique, analyse et construit ses statistiques grâce à SAP HANA. Semelles intelligentes, capteurs d'accélération, mesure des pulsations... Grâce au Wifi, les objets connectés apportent au sport finesse d'analyse et réactivité. Le sport, c'est aussi un public, dont on peut observer le comportement via des puces, applications mobiles ou tickets intelligents, et à qui on peut recommander le service ou produit adéquat.

L'automobile est un autre secteur fortement impacté par le Big Data. L'introduction de capteurs et des objets connectés dans les voitures se démocratise et engendre une nouvelle approche de cette industrie. Pour les assureurs plus particulièrement, connaître le kilométrage effectué et obtenir des informations sur le comportement de l'automobiliste révolutionne le modèle économique, en instaurant le Pay As You Drive. Le machine to machine permet également de faire communiquer entre eux deux véhicules. BMW s'intéresse actuellement aux analyses prédictives pour la maintenance mais aussi l'anticipation des bouchons, avec une possibilité d'interaction avec l'affichage public pour faire remonter l'information et alerter les automobilistes.

Dans l'aéronautique, la maintenance est un point sensible. L'utilisation de capteurs, couplés à une analyse algorithmique, offre la capacité d'établir la probabilité de pannes en amont. Avec Hadoop et un historique de 3 à 6 mois voire davantage, on peut alors affiner les modèles mathématiques. On connaît davantage le Big Data dans le secteur BtoC. Récemment apparues, les Smart Vending Machine sont la version upgradée du distributeur automatique, et disposent d'un écran digital intelligent. La machine, grâce à l'utilisation de cartes de fidélité, apprend à connaître chaque consommateur, et lui pousse le bon produit ou service au bon moment. Historique d'achat, horaire, météo ou données externes non-structurées permettent de disposer d'une information précise et in fine apportent un chiffre d'affaire incrémental. Revenons aux capteurs et à leur utilisation. Pour la grande consommation par exemple, la remontée d'informations issue de chariots de supermarché peut apporter à un gérant une vision en temps-

réel de l'affluence dans son supermarché, lui permettre de relier cette information aux données externes comme la météo et prévoir les réactions adaptées. Un modèle mathématique lui dira alors si réorganiser ses rayons d'une certaine façon lorsque survient la pluie en période de forte affluence pourra améliorer son chiffre d'affaire.

Il s'agit là presque de considérations organisationnelles. Le Port de Hambourg, mondialement connu comme l'un des principaux Hub mondiaux, est actuellement confronté à une problématique liée à sa taille. L'optimisation du trafic y est donc primordiale. Il faut anticiper l'arrivée des bateaux, les géolocaliser en temps-réel et intégrer les facteurs externes pour prévoir l'arrivée des camions pour le déchargement du fret au moment idéal.

Dernier exemple, la sécurité. Le Big Data peut par exemple permettre de détecter un mouvement suspect. Dans un aéroport, lieu hautement fréquenté et stratégique, on peut alors repérer un individu au comportement que le modèle considérera comme anormal. Seul, sans téléphone et immobile pendant plus de 40 minutes, il pourra entraîner une alerte au sein du service de sécurité. L'intérêt est là encore de pouvoir stocker les informations et faire appel à l'historique des données ou au replay.



Frédéric PUCHE





**COORDONNÉES
SAS INSTITUTE**

Domaine de Grégy
Grégy-sur-Yerres
77257 BRIE-COMTE ROBERT CEDEX

Tél **01 60 62 11 11**
Fax 01 60 62 11 20
Mail comsas@fra.sas.com

CONTACT

**Serge
BOULET**

Directeur Marketing

SAS, votre atout performance

SAS, leader mondial des solutions de business analytics, propose des outils analytiques puissants, 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.

Vous avez besoin d'une solution spécifique adaptée à votre secteur ? D'une solution fonctionnelle transverse pour répondre à vos challenges ? Quelle que soit votre problématique, SAS a une réponse.

Nous vous aidons à valoriser votre relation client, à piloter vos risques, à combattre la fraude et à optimiser vos réseaux informatiques. Nos solutions s'appuient sur un socle technologique avec trois composantes clé : la gestion des données, les outils d'analyse et la restitution de l'information.

Avec SAS, la business analytics est accessible et simple : SAS® Visual Analytics vous permet d'explorer visuellement et rapidement vos données, de créer vos rapports et partager les résultats sur le web ou sur tablette.

Big Data, Hadoop et analytique

Les enjeux analytiques des projets Big Data sont à la fois d'ordre technologique et métier avec des problématiques liées à la nature, au volume et à la localisation des données, à des architectures extensibles, des compétences variées, une agilité et une précision d'analyse, et à de nouveaux besoins relatifs à ces données.

La nouvelle plateforme analytique de SAS permet d'analyser interactivement toutes les nouvelles sources de données pour créer des modèles prédictifs intégrés à un cycle analytique complet. Elle permet aux analystes de travailler en faisant appel à une riche bibliothèque d'algorithmes et de méthodes analytiques pour créer et modifier dynamiquement les modèles prédictifs.

Cette plateforme obtient des performances remarquables car les traitements s'exécutent en-mémoire, au plus près des données et en mode distribué lorsque possible.

Hadoop représente la plateforme majoritairement utilisée pour les projets Big Data, par le faible coût de stockage, l'architecture évolutive, l'agilité de laboratoire qui la caractérise. Votre succès est notre succès.

SAS et Hadoop relèvent
le défi des Big data.
Et ils le remportent.



Big Data Analytics

Que faire avec des volumes massifs de données si vous ne pouvez pas vous en servir ? Ou si leur analyse doit prendre des jours ou des semaines ?

Combiner le pouvoir analytique de SAS aux capacités d'Hadoop, dans un environnement unique et interactif, vous livre les décisions clé en main - pour des résultats rapides et une valeur maximale.



Lire le rapport TDWI
sas.com/tdwi

Macys.com sublime l'expérience client avec les solutions SAS sur Hadoop

Macy's s'appuie sur une plateforme Hadoop Cloudera et les solutions d'analyse SAS pour améliorer la performance de ses ventes en ligne.

Les acteurs traditionnels de la grande distribution doivent désormais être à même de proposer des services concurrentiels à ceux offerts par les purs acteurs du commerce électronique. C'est ainsi que Macys.com, filiale e-commerce de Macy's Inc., icône de la grande distribution aux États-Unis, avec plus de 80 ans d'existence et un chiffre d'affaires de 20 milliards de dollars, a mené un projet destiné à améliorer l'expérience des clients internautes et accroître sa rentabilité.

Porteurs du Projet

Ce projet a été dirigé par Kerem Tomak, Vice-Président Analytics de Macys.com, avec son équipe d'analystes.

Objectifs du projet

L'objectif était de mesurer l'efficacité des campagnes marketing et les paramètres de navigation sur le site, connaître la Valeur Client, identifier les clients les plus fidèles, et adresser des promotions ciblées aux meilleurs d'entre eux.

Présentation du projet

a. Nature des données traitées

Macys.com exploite des données variées : données clients, suivi de campagnes d'e-mailings de promotion, suivi de navigation, préférences clients (ce qu'ils aiment et n'aiment pas).

b. Ressources et méthodologie

Toutes les données à analyser sont stockées sur une plateforme de données Hadoop Cloudera. Elles sont analysées avec les solutions analytiques de SAS pour segmenter finement les clients et calculer les scores propres à chaque segment.

L'équipe analytique est aussi responsable du reporting à destination des équipes marketing et finance. L'automatisation du cycle analytique a permis de réduire le temps de production de ces rapports et dégager du temps pour générer plus d'analyses et en accroître la finesse et la pertinence.

c. Calendrier de déploiement

Ce projet qui a démarré en 2012 est en production depuis plusieurs mois et en perpétuelle amélioration : ainsi, Macys.com a pu réduire le taux de désabonnement à ses campagnes marketing de 20% et accroître la productivité des analystes. Le gain est estimé à 500 000 USD par an.

Aspect Novateur

La meilleure segmentation des clients a permis à Macys.com d'optimiser les campagnes marketing (en réduire la fréquence et le nombre des clients ciblés) et accroître leur efficacité (taux de souscription et taux de réponse).

Valorisation observée de la donnée

La capacité de comprendre le comportement multicanal des clients est critique pour améliorer l'expérience client et SAS est utilisé pour analyser les données et créer les modèles analytiques qui valident et adaptent les algorithmes de « up et cross sell ».

« Cette capacité à analyser en profondeur et à tirer parti des données est importante pour supporter la stratégie et le pilotage des évolutions du sites Macys.com, » commente Kerem Tomak.

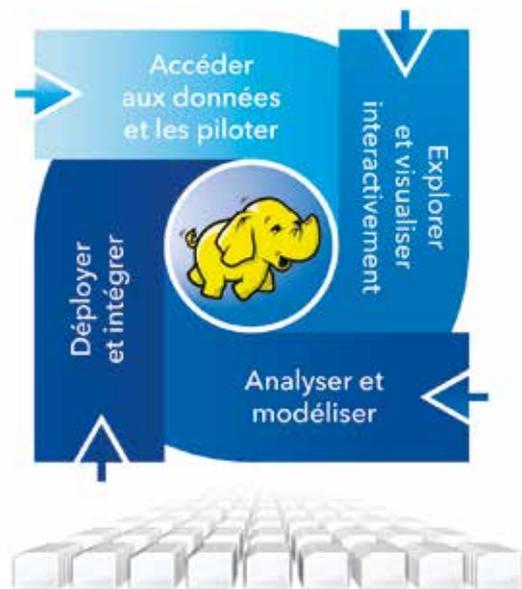
Etapes à venir et démarche Big Data dans l'entreprise

L'équipe analytique de Macys.com étudie l'adoption de solutions d'optimisation des processus métier pour maximiser l'efficacité des campagnes promotionnelles en améliorant encore le ciblage et en accroissant la réactivité.

La volonté de Macys.com est de générer des informations plus précises et en temps réel sur les préférences des clients afin de personnaliser dynamiquement les promotions.

Plusieurs clients SAS utilisent SAS et Hadoop conjointement pour valoriser leurs sources de données (web, log machines, réseaux sociaux, verbatim Service Clients etc.), par exemple dans des projets d'amélioration d'utilisation de machines industrielles (Predictive Asset Maintenance), pour fournir des recommandations ciblées (produits, contenu) ou pour accroître l'efficacité des messages publicitaires et des campagnes promotionnelles par un ciblage précis.

SAS Institute, société privée créée en 1976. 70 000 sites clients, dont 2600 en France, plus de 3 milliards de dollars de CA et 14 000 employés.





COORDONNÉES SCINETIK

78 rue Taitbout
75009 Paris

Tél. **+33 673 804 708**
Mail lively@scinetik.com

CONTACT

**Xavière
TALLENT**

CEO

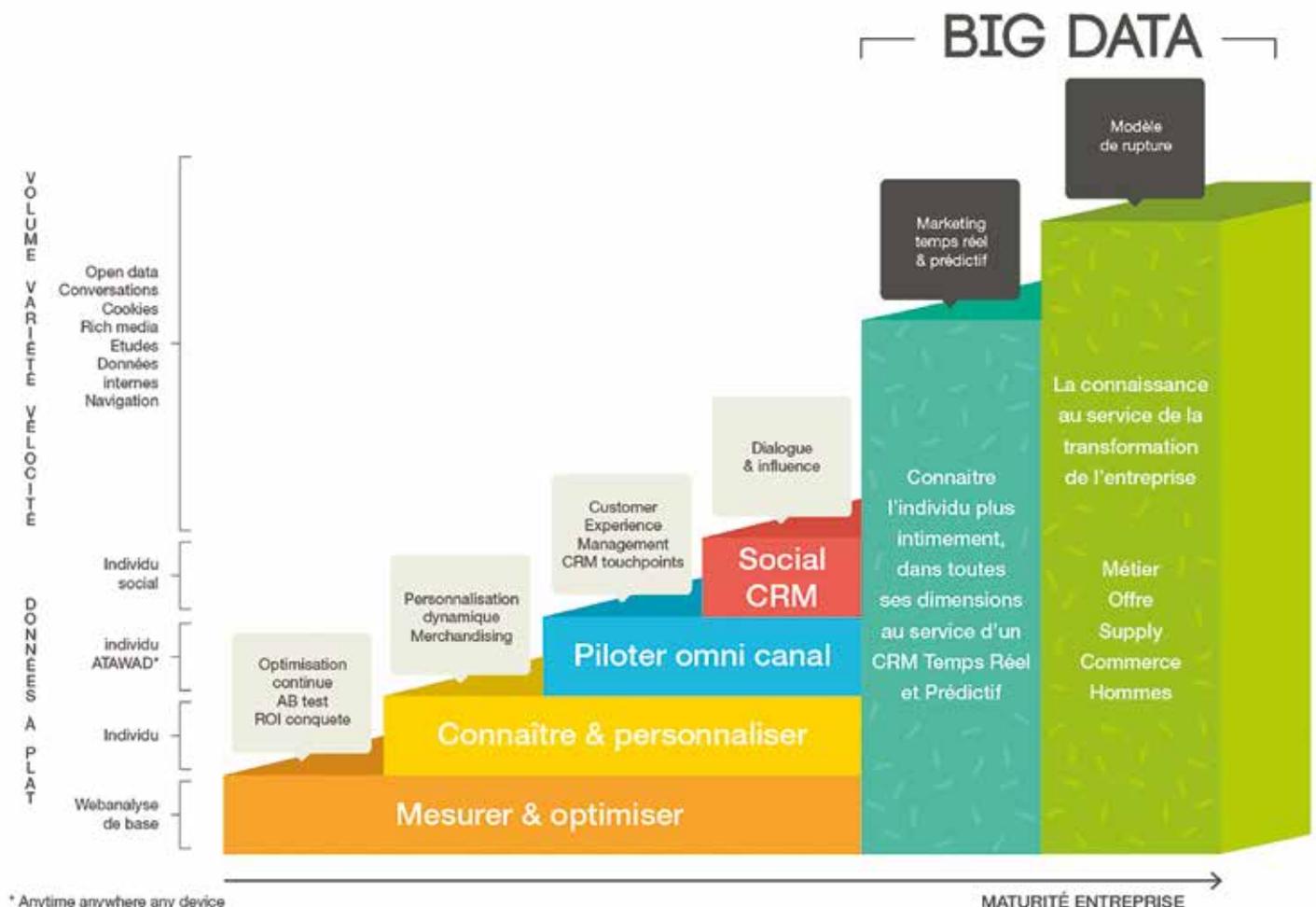
Scinetik est un cabinet de conseil dont la mission est d'assurer pour ses clients, le développement pérenne et rentable de leur activité de commerce connecté.

Notre mission est d'accompagner nos clients dans la définition et la mise en œuvre de stratégies de développement sur l'ensemble des drivers de business de la conquête à la fidélisation en passant par la conversion. Pour y parvenir et afin de s'assurer de l'efficacité des actions, Scinetik fonde l'ensemble de ses recommandations sur l'analyse de la donnée. Cette donnée vient de multiples sources : le webanalytic, les comportements d'achat, les données CRM, des données externes...

Scinetik a été créé en réaction au manque de pilotage des activités de commerce connecté au sein des retailers. Beaucoup d'ambitions, peu d'expertise data ou un manque de perception de l'enjeu vital de l'exploitation de la donnée avant même tout enjeu de branding et d'image de marque.

Redonner du sens au business, telle est notre mission.

Naturellement cet engagement dans la donnée nous a amené à développer une expertise forte en matière de « Big Data ». Mais pour nous le concept « big data » n'est pas utilisé comme un effet de mode mais comme la conviction profonde que nous sommes arrivés à un nouvel âge du marketing et du business où nous devons réinventer la façon même d'exploiter la donnée. Nous nous devons également d'accepter avec enthousiasme le défi que nous oppose les variétés des sources et types de données, l'opportunité des open data et tout ceci en temps réel pour un business toujours plus rapide, une obsolescence des produits et des promesses relationnelles toujours plus grande. Scinetik travaille pour de grands groupes internationaux dans le domaine du retail, de la grande consommation, de la mode, du sport.



Le Big Data au service de la transformation de l'entreprise

Le contexte du Client

Après plusieurs années de veille sur la montée du Big data, beaucoup de lecture, la participation à des conférences, la rencontre avec des entreprises ayant démarré leur réflexion big data en France mais surtout à l'étranger, notre client a décidé de mettre son entreprise sur le chemin du big data.

La question posée à Scinetik

«*En quoi le big data peut-il impacter véritablement l'entreprise dans ses différentes dimensions ? Comment ne pas limiter le Big Data à un super CRM comme c'est le cas dans un grand nombre d'entreprises, notamment en France ?*»

Le défi

Concilier ambition stratégique majeure, pragmatisme économique et opérationnel. Le Big Data n'est pas prioritairement un sujet technologique mais un sujet stratégique et organisationnel.

Un défi Stratégique car il va modifier potentiellement le business model de l'entreprise, son offre, son merchandising, son marketing et bien entendu son CRM.

Un défi organisationnel car beaucoup d'entreprises se réfugient derrière le « data scientist », sorte de profil hybride entre data/métier/technique. Or si l'ensemble des équipes de l'entreprise n'est pas partie prenante dans le projet alors les outputs du big data, aussi brillants soient-ils ne se diffuseront pas dans les différentes strates de l'entreprise, n'amèneront pas les hommes et les femmes à travailler différemment en faisant de la donnée un atout majeur au quotidien et dans la durée.

La démarche

- Création d'un groupe de réflexion highlevel, ayant à la fois la vision de l'entreprise dans son ensemble et ses particularités et également le pouvoir de faire bouger les lignes.
- Définition d'une vision stratégique centrale
- Déclinaison de cette vision sur les différents métiers de l'entreprise
- L'idée force : Création de prototypes. Chaque prototype adresse une dimension de la big data dans une dynamique vertueuse et apprenante
- Définition de la méthodologie de récupération et d'analyse des données internes et externes
- Définition de différents scénarios pour la mise en place de l'infra, l'enjeu n'étant pas de lancer de lourds investissements dès le démarrage mais de tester plusieurs options pour permettre les meilleurs choix à terme
- Lancement des prototypes

Les résultats à date

Démarrage opérationnel des prototypes :

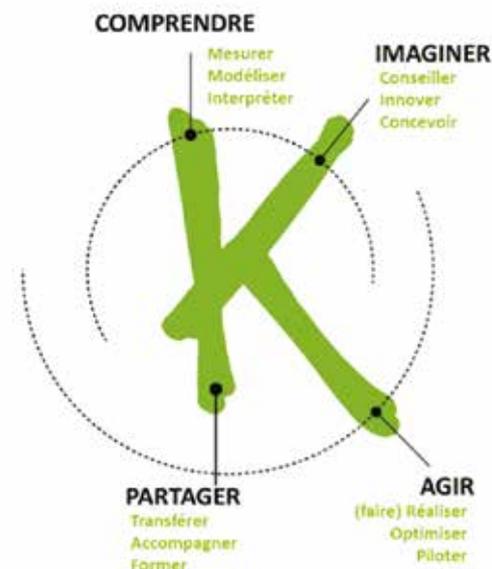
- cadrage des prototypes et mise en place des équipes multicompetences internes-externes
- identification des sources de données notamment externes, et pour chaque source la donnée utile
- identification des outils/solutions adaptées et potentiels partenaires techniques (IT).
- .../...

L'enjeu de ces prototypes est également de faire monter en compétence les équipes internes et de dessiner la future architecture SI de metadonnées.

En fonction des résultats attendus nous serons à même d'utiliser différentes techniques d'analyse et de modélisation des données en s'appuyant sur une équipe de statisticiens spécialisés Big data.

Le défi que nous voulons relever est de répondre à de réels besoins métier en utilisant les mines d'informations existantes mais peu ou pas exploitées et en s'appuyant sur de nouvelles techniques et méthodes d'analyse.

Nous sommes convaincus que nos clients peuvent tirer de la valeur des metadonnées. C'est un vaste domaine d'exploration et d'expérimentation qui s'offre à nous.





COORDONNÉES SEMDEE

15, Rue Jean Baptiste BERLIER
75013 PARIS

Tél **+33 (0)1 55 43 79 60**
Mail contact@semdee.com

CONTACT

**Brice
HOARAU**

CEO

Qu'elles proviennent directement de l'entreprise (CRM, ECM, CMS, RSE, back office, E-mails, dossiers textes, voir même des DATA WAREHOUSE) ou de l'extérieur (Sites Internet, Blogs, Forums, Réseaux Sociaux) les données non structurées sont partout. Confronté à ce déluge les Entreprises entendent parler de « formidable opportunité », de « gisement de valeur » ou de « création de valeur ».

Comment passer de cette situation d'avalanche de données à une création de valeur ? C'est à cette question que répond SEMDEE éditeur de technologies sémantiques au travers de son offre « SEMDEE SMART DATA FACTORY ».

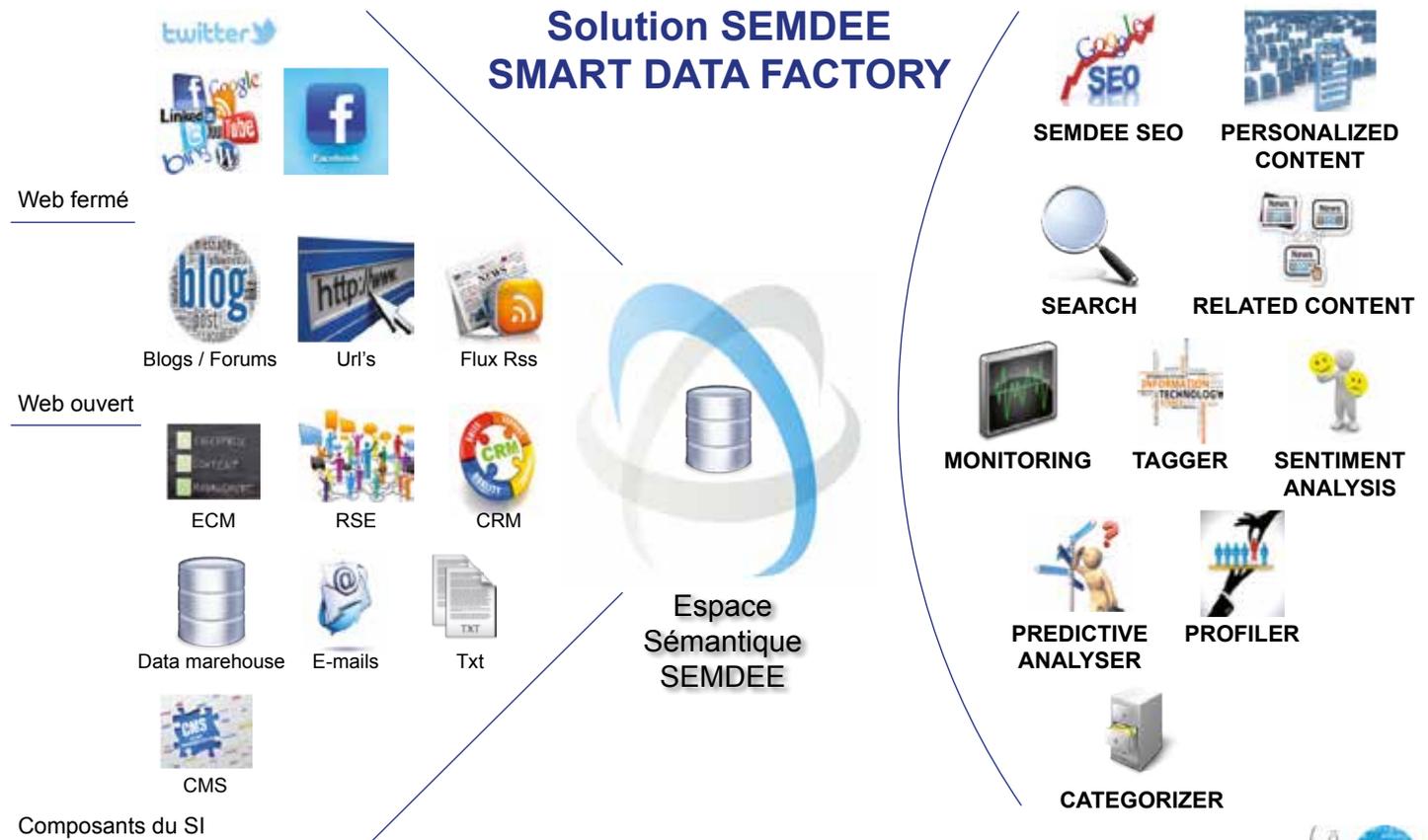
La technologie SEMDEE est issue des Sciences Cognitives et de travaux sur la mémoire épisodique. Utilisant les récents développements en Intelligence Artificielle, SEMDEE met à disposition de ses utilisateurs les ultimes avancées dans le domaine de la sémantique. Basée sur une approche statistique l'utilisation du « Machine Learning » permet une amélioration constante de la précision et de la pertinence.

Inspiré du fonctionnement du cerveau humain l'Espace Sémantique, au cœur de la gamme, indexe les données qui vont être utilisées. Les principales étapes cognitives modélisées, chacune sous forme de module fonctionnel, constituent l'ensemble de la gamme.

Ces modules sont utilisables de manière autonome pour répondre à un simple besoin technique (recherche, classification etc...) ou d'une manière plus complète dans le cadre de la solution « SEMDEE SMART DATA FACTORY ».

D'une manière générale, « SEMDEE SMART DATA FACTORY » se propose d'industrialiser l'acquisition et l'utilisation de la connaissance. SMART DATA FACTORY s'adresse directement aux clients et aux intégrateurs qui souhaitent constituer leur propre plate-forme d'analyse de données.

Solution SEMDEE SMART DATA FACTORY



Comment détecter et anticiper le «churn» avec SEMDEE

Porteur du projet : Murat AHAT

Objectifs

Aider un grand groupe à anticiper le départ de ses clients.

Mettre en place un système d'alerte au « churn » dans le cadre d'un plan de rétention des clients.

Présentation

Voici les principales étapes qui seront menées dans le cadre de ce projet :

- A) Récupération des historiques des données des clients qui ont quitté le service.
Ces données peuvent provenir :
 - a. Du web
 - b. Des réseaux sociaux
 - c. Des plates-formes de service
 - d. Des emails
 - e. Des posts dans les forums
- B) Agrégation des données par client
- C) Indexation sémantique de ce contenu
- D) Pondération des contenus en fonction des connaissances métiers
- E) Génération de l'ensemble des profils des clients
- F) Détermination d'un seuil d'alerte (proximité par rapport à ces profils)
- G) Récupération de contenus émis par les clients actuels du service dans
 - a. Les formulaires
 - b. Les emails
 - c. Les réseaux sociaux
 - d. Les forums
 - e. Les plates-formes de services
- H) Agrégation des données par client
- I) Indexation sémantique des contenus
- J) Lorsqu'un client actuel du service se rapproche du seuil d'alerte
 - a. déclenchement d'une alerte auprès du service client
 - b. Le service client peut alors traiter en direct
- K) Outils supplémentaires mis à disposition du service client
 - a. Lorsqu'un opérateur consulte un profil susceptible de quitter le service, le système propose automatiquement la liste des profils qui lui sont le plus proche
 - b. L'opérateur peut effectuer des recherches en mode sémantique pour approfondir sa connaissance

Compétences humaines :

Dans le cadre de ce projet, il sera nécessaire de disposer des ressources suivantes :

- Une compétence métier
- Un consultant SEMDEE pour installation et paramétrage
- Une compétence d'intégration pour personnaliser la restitution auprès des utilisateurs

Calendrier de déploiement :

Les durées sont données à titre indicatif car elles dépendent évidemment du volume de données, de la diversité des supports et de la qualité de la connaissance métier mis à disposition.

A partir du moment où les sources de données sont précisément identifiées :

- Extraction des données (1 à 2 semaines selon complexité et le nombre de support)
- Indexation sémantique et paramétrage (1 semaines).
- Paramétrages des profils et alertes (1 semaine)
- Mise en place des modules connexes permettant un meilleur confort (recherche, enrichissement de l'information, suggestion de profils etc (1 semaine/module)
- Intégration, personnalisation dans l'environnement utilisateur (intégrateur)

Aspect Novateur

Fonctionnement instantané

(pas de besoin de dictionnaire, de lexique etc)

Fonctionne sur l'ensemble du contenu (pas limité à un groupe de mots ou de phrases mais sur des textes entiers) - Fonctionne dans toutes les langues. Machine learning : plus le système va fonctionner plus il va devenir précis.

Toutes les étapes cognitives sont intégrées en une seule solution = Cohérence cognitive accrue.

Valorisation observée de la donnée

ROI directement calculable.

Combien vaut un client ? Combien coûte la perte d'un client ?

Combien de clients en moyenne perdent les entreprises sur le même secteur d'activité ?

Combien l'entreprise avait elle perdue de clients l'année précédente ?

Etapas à venir et démarche Big Data dans l'entreprise

Fonctionne sur toutes les autres problématiques de la même manière :

Détection d'opportunités de ventes - Amélioration de la connaissance des usages client, du client, des tarifs et des concurrents

Détection de nouveaux besoins...





COORDONNÉES

SENTELIS

31 Rue du pont
92 200 Neuilly-sur-Seine

Tél +33 (0)6 76 09 82 46

Mail i.regnier@sentelis.com

CONTACT

**Isabelle
REGNIER**

Associé Fondateur

SENTELIS, smart solutions for smarter enterprises

Sentelis est un cabinet de conseil en stratégie, gouvernance et architecture de systèmes d'information.

Nous accompagnons les décideurs SI dans leur transition numérique vers l'entreprise 3.0 (CIO, CDO, CTO, CMO...), de la définition de leur stratégie à la concrétisation opérationnelle.

Nous les aidons à identifier, à optimiser et à déployer les capacités stratégiques du système d'information 3.0 en réponse à l'équation digitale qu'ils doivent résoudre pour conjuguer innovation et industrialisation, pour intégrer et tirer bénéfices des nouvelles réalités numériques comme le « Cloud Computing », le « Mobile Computing », les « Social Medias », l'« Internet-of-Things », le « Digital Working », le « Crowd-Sourcing » et bien sûr le « Big Data ».

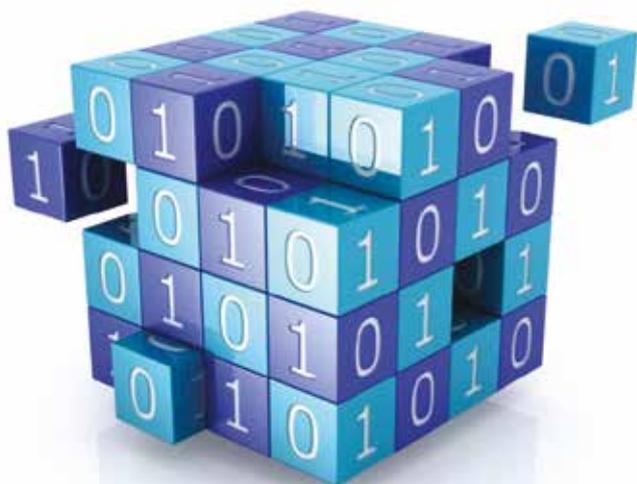
Nos solutions concrètes, pragmatiques et différenciées répondent aux besoins d'efficience (Time-to-market) et d'efficacité (Cost-to-income) du SI. Elles sont imaginées et déployées au travers d'un cadre méthodologique, architectural et de gouvernance innovant et unique sur le marché, dont Sentelis est l'éditeur : smartfoundations™.

Sentelis smartfoundations™ couvre l'ensemble du cycle de vie des fondations du SI Digital, qu'ils s'agissent de composants SI ou de pratiques SI transverses. Il garantit des fondations industrielles, pérennes et à fort retour sur investissement.

Nous avons un savoir-faire reconnu et éprouvé sur la plupart des fondations de l'entreprise digitale, aussi bien des composants du SI 3.0 (Big Data & Business Analytics, SOA/API Management & Cloud Integration, MDM, BPM/ACM, eCRM, Enterprise App Stores & Social Networks), que des pratiques de la DSI 3.0 (Fast IT/DevOps, IT Factory, SSC) et des dispositifs transverses métiers-DSI de l'entreprise 3.0 (Data Labs, Data Governance, Social Enterprise Architecture).

Nous intervenons en toute indépendance vis-à-vis des éditeurs sur ces fondations de l'entreprise 3.0 (étude opportunité, preuve-de-concept métier & technologique, industrialisation) y compris sur les aspects gouvernance et conduite du changement.

Vos données en savent plus que vous. Faites-les parler.



Entreprise 3.0 (n.m) :

Entreprise qui conjugue au présent innovation et industrialisation

Big Data (n.m) :

Fondation stratégique de l'entreprise 3.0

Smartfoundations (n.f) :

Cadre de référence Sentelis pour résoudre l'équation digitale du SI de l'entreprise 3.0

Sentelis (n.f) :

Créateur et intégrateur de fondations stratégiques SI et DSI de l'entreprise 3.0

Architecture 3.0 : FONDATION BIG DATA pour une Mutuelle d'Assurance

Porteur du Projet : Sébastien LAYER

Objectifs

Mettre en place un socle Big Data cross-métiers pour servir les nouveaux usages 3.0 :

- Analyse réactive (contextualisation temps réel de la relation client sur tous les canaux)
- Analyse historique du portefeuille client sur l'ensemble des données à disposition (micro-segmentation, qualité et évolution de la relation, segmentation comportementale)
- Analyse prédictive du portefeuille client (détection d'opportunités, prévention des risques)

Présentation

Intégration dans un lac de données d'une très grande variété de données représentatives du système d'information (données transactionnelles, événementielles, conversationnelles, sociales référentielles, décisionnelles...) pour un volume avoisinant plusieurs milliards de données.

Réalisation en 2 étapes majeures pour sécuriser la mise en œuvre de la fondation Big Data :

- Identification de cas d'usage candidats.
- Conduite d'une preuve-de-concept à valeur métier et architecturale sur une infrastructure Commodity Hardware et des composants Open Source de l'écosystème Hadoop

Réalisation d'une preuve de concept sur un usage emblématique (Vision 360° Client) au sein d'une équipe mixte métier-SI en mode agile (Utilisateur métier, Ergonome, Analyste Donnée, Statisticien, Architecte Big Data, Développeur Big Data) incluant :

- L'ingestion en masse des données dans une logique Schema-less/Store-first-Ask-Later
- L'ingestion de donnée unitaire à haute vélocité et sans latence
- Le traitement en masse des données pour corrélérer l'ensemble des données clients
- L'indexation en masse des données pour permettre leur exploration
- L'exposition de services pour banaliser l'accès aux données et aux traitements

Calendrier de déploiement

- 3 mois pour réaliser la preuve-de-concept
- 2 mois pour choisir les composants de la stack Big Data
- 6 mois pour industrialiser et déployer les premiers usages en production

Aspect Novateur

- Modèle d'architecture digitale 3.0
- Désilotage des données entre les métiers et des mondes opérationnel et analytique
- Amélioration d'usages existants (ex : vision 360 Client, Contrat) - Perspective de nouveaux usages (ex : analyse de la couleur et la teneur de la conversation client, corrélation des flux comptables) - Capacité d'innovation renforcée (Test & Learn, Data Lab)

Valorisation observée de la donnée

Le projet permet d'envisager une exploitation de toutes les données disponibles structurées et non-structurées comme jamais auparavant, sans frontière de temps ni d'espace (ex : analyse des commentaires clients dans les questionnaires de satisfaction et corrélation avec la notation ; détection de nouvelle segmentation via des algorithmes d'analyses non propriétaires).

Etapes à venir et démarche Big Data dans l'entreprise

Les étapes à venir sont : Industrialisation du socle Big Data (pratiques et solutions) - Mise en production des premiers usages.

Ces étapes s'accompagnent en parallèle de la définition de l'offre de service associée à la fondation Big Data et de l'organisation pour en assurer la pérennité et le déploiement dans l'entreprise.

Autres informations : La mise en place du lac de données a été un révélateur et un catalyseur de nouveaux usages pour l'ensemble des métiers, encouragés par des coûts et une performance de traitement sans commune mesure avec l'existant. Elle ouvre la place, au-delà de l'amélioration de la Business Intelligence et du développement de la Business Analytics, au développement de la Data Science pour tirer le maximum de valeur du capital informationnel de l'entreprise. Elle a mis en évidence la nécessité d'une gouvernance de la donnée resserée pour maîtriser l'explosion des données et des usages. Le lac de données est la première étape vers la construction d'une infrastructure fédérée des données entre monde opérationnel et monde analytique (Shared Data Backbone), dont la frontière disparaît.





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 pornain@sinequa.com

CONTACT

**Xavier
PORNAIN**

VP Sales & Alliances

Sinequa : Découvrez la valeur cachée de vos données

Le Search et l'Analyse de contenus structurés et textuels en temps-réel, au service de vos processus métiers stratégiques

Le défi du Big Data : Plus de 80% de ces données sont non structurées - des textes en beaucoup de langues différentes. Les professionnels de l'information perdent un temps considérable à regrouper l'information dont ils ont besoin pour accomplir leur travail. La qualité du service client faiblit, la R&D est freinée et les entreprises laissent échapper des opportunités lorsque leur personnel est submergé d'informations et incapable de trouver la bonne information au bon moment.

Notre mission :

- Offrir un accès en temps réel, personnalisable et sécurisé aux informations pertinentes à chaque utilisateur dans son contexte métier.
- Simplifier la création d'InfoApps (Search-Based Applications) par secteur et par métier

Notre solution :

- Une analyse de contenu unique, basée sur une analyse sémantique brevetée (traitement du langage naturel) et sur une expérience en traitement de données structurées provenant d'applications métier
- Une architecture informatique très performante et scalable, permettant de gérer des dizaines de milliers d'utilisateurs et des milliards de documents en temps-réel
- Une intégration sécurisée et homogène avec plus de 140 sources de données

Notre expérience :

- Deux décennies de recherche en matière de traitement du langage naturel

L'information pertinente pour chaque personne au bon moment

Analyse
approfondie
du Big Data
en temps réel

Tirez profit de votre Big Data pour

-  mieux décider
-  optimiser votre R&D
-  mieux servir vos clients
-  bénéficier de tout votre capital intellectuel

Vos données non-structurées sont essentielles : elles représentent jusqu'à 90% de vos données.

Search et Analyse du Big Data au service de la Relation Client

AMÉLIORATION DU SERVICE CLIENT ET RÉDUCTION DES COÛTS

L'un des principaux opérateurs de télécommunication en Europe utilise Sinequa pour délivrer instantanément aux employés de ses centres d'appels une vue 360° des clients. Il réduit ainsi de manière significative la durée moyenne des appels, diminue le coût d'accès à l'information et rend possible le développement d'Info Apps en mode agile pour répondre à l'évolution du business.

Porteurs du projet

Côté client, le projet a été porté par la DSI de la Relation Client, avec une implication forte du Métier de la Relation Client notamment pour son impact sur l'ergonomie des Chargés de Clientèle (CC).

Objectifs

Côté business, il s'agit avant tout d'améliorer la satisfaction client; côté informatique, de diminuer les coûts liés au fonctionnement du centre de contacts et de ses quelques 250 applications; enfin, l'accès unifié à l'information depuis une seule application permet de réduire drastiquement la durée de formation et de montée en compétence, et d'améliorer l'ergonomie des CC.

a. Améliorer la satisfaction client

La nouvelle interface de travail des CC leur permet de répondre directement à 80% des requêtes client. Libéré d'une navigation chronophage entre plusieurs applications (plus de 30 dans certains cas), le CC peut traiter la demande du client plus efficacement.

b. Réduire le coût de fonctionnement et gagner en performance

Le découplage entre la lecture des données et l'accès aux applications back-office améliore les performances et réduit les temps de réponse, même durant les pics d'activité.

Le système d'information gagne en agilité puisque des applications (Info Apps) peuvent être développées rapidement afin de s'adapter aux évolutions du business.

c. Diminuer le besoin en formation et améliorer l'ergonomie du CC

La complexité du Back-Office existant (ERP, CRM, etc.) s'efface devant la nouvelle interface du CC réduisant le besoin de connaissance des outils Back Office.

Présentation

a. Nature des données traitées

Il s'agit de données structurées venant d'ERP, du CRM, d'applications métiers spécifiques mais aussi de contenus non-structurés comme des contrats, des factures ou des emails.

b. Ressources et Méthodologie

Le projet a été conçu en méthode agile avec les utilisateurs finaux.

c. Calendrier de déploiement

Le pilote a été réalisé en 2 mois et la mise en production en 9 mois.

Aspect Novateur

Fédérer de la donnée structurée et non-structurée afin de créer une vue 360° client en temps-réel pour des milliers de CC est l'un des cas d'usage du Search et de l'Analyse de donnée Big Data les plus innovants réalisés aujourd'hui. Le Search joue désormais un rôle déterminant au cœur de processus métiers stratégiques. Offrant un véritable levier de compétitivité, il ouvre la voie à des innovations qui n'auraient pu voir le jour dans des architectures informatiques « classiques ».

En à peine neuf mois, le projet a été déployé auprès de 10 000 CC. Le ROI en satisfaction client et en temps gagné par appel s'est fait sentir très rapidement après le déploiement.

Valorisation Observée de la donnée

La mise en place du nouvel accueil client permet de réduire de près de 20% la durée moyenne des appels donnant ainsi aux 10 000 CC la capacité de traiter environ 40.0 000 appels supplémentaires par jour, et d'augmenter l'ARPU (Average Revenue Per User) grâce à une meilleure connaissance du Client (permettant de proposer les offres les plus adaptées).

Etapes à venir et démarche Big Data dans l'entreprise

A l'avenir, encore plus de sources de données pourront être indexées, y compris des données non structurées provenant par exemple des réseaux sociaux (Data-profiling, Data-scoring).

*Pour une vision à 360°
de vos clients*



*Chargé de
clientèle*



*Vue 360°
Client*



SINEQUA
REAL-TIME BIG DATA SEARCH & ANALYTICS





COORDONNÉES SPLUNK

Tour Ariane – 5 place de la Pyramide
92088 La Défense Cedex

Tél **01 55 68 12 61**

Mail emea_france@splunk.com

CONTACT

**Pierre
GOYENEIX**

Directeur Commercial EMEA South

Splunk Enterprise est la plateforme leader de l'intelligence opérationnelle en temps réel. Elle offre un moyen simple, rapide et sécurisé d'explorer, analyser et visualiser les flux massifs de données générés par vos systèmes informatiques et votre infrastructure technologique, qu'elle soit physique, virtuelle ou dans le Cloud.

Splunk Enterprise offre :

- De puissantes capacités d'analyse accessibles à tous à des vitesses extraordinaires
- Une expérience utilisateur entièrement repensée
- Un environnement de développement enrichi permettant de prolonger facilement la plateforme

Splunk Enterprise 6.1 est notre dernière version et offre des fonctionnalités conçues pour améliorer la manière dont vous analysez vos données et interagissez avec elles, tout en proposant une disponibilité continue de votre déploiement distribué de Splunk Enterprise.

Recueillez et indexez les données machines de toutes les sources ou presque en temps réel. Explorez, surveillez, analysez et visualisez vos données pour obtenir de nouveaux renseignements. Indexez toutes vos informations pour bénéficier d'une meilleure visibilité et renforcer vos capacités d'investigation et de dépannage. Travaillez plus intelligemment en partageant des recherches entre équipes et en accumulant des connaissances propres à votre organisation. Créez des rapports ciblés pour identifier des tendances ou répondre à des contrôles de conformité. Composez des tableaux de bord interactifs pour surveiller les incidents de sécurité, les niveaux de services et autres indicateurs clés de performance. Analysez les transactions des utilisateurs, le comportement des clients, celui des machines, les menaces de sécurité et les activités frauduleuses, le tout en temps réel. Indexez toutes les données

**QUI NE
SPLUNK
RIEN
N'A RIEN.**

Commencez avec des données machine et le logiciel Splunk.[®] Terminez avec un avantage considérable.

Splunk transforme les données machine en renseignements précieux pour votre entreprise. Découvrez la meilleure des plateformes d'information opérationnelle en temps réel.

Pour en savoir plus, rendez-vous sur fr.splunk.com

splunk > listen to your data™

Splunk® chez Karavel

IDENTIFICATION DES ERREURS EN TEMPS RÉEL ET DIMINUTION DU TEMPS DE RÉPONSE, POUR UNE MEILLEURE EXPÉRIENCE UTILISATEUR.

SECTEUR

Voyages (agence de voyages en ligne)

CAS D'UTILISATION SPLUNK

- Gestion des applications
- Intelligence numérique
- Intelligence opérationnelle

RÉSULTATS

- Meilleure expérience utilisateur avec une réduction de 82 % des temps de réponse des pages.
- Amélioration du référencement naturel par l'identification et la réparation ou la suppression des liens non trouvés.
- Identification en temps réel des divergences de configuration, permettant d'éviter des pertes de recettes.
- Relation proactive avec les opérateurs pour rectifier les erreurs humaines du processus.

SOURCES DE DONNÉES

- Logs du proxy inverse Varnish
- Logs applicatifs Tomca

Défis

Karavel attache une grande importance à la haute qualité de service offerte à ses clients, mais souffrait d'un manque de visibilité détaillée sur son architecture informatique. Des outils de surveillance étaient bien utilisés, mais ils ne fournissaient pas une vision suffisamment précise. Karavel pouvait par exemple connaître le temps de réponse moyen des pages principales d'un site donné, mais pas la distribution détaillée des temps de réponse ; il n'était donc pas possible de savoir combien d'utilisateurs recevaient un service inférieur aux accords internes sur les niveaux de service, ni d'identifier préventivement les situations où ceux-ci étaient sur le point d'être enfreints. Ce manque de visibilité avait en outre des conséquences sur le processus de réservation en ligne, car Karavel n'était pas en mesure de détecter certains problèmes : par exemple, des divergences entre les configurations d'un produit créé par l'opérateur, importé sur la plateforme de Karavel et offert aux clients via le site web. Certaines erreurs rendaient impossible la finalisation de l'achat par le client.

Intervention de Splunk

Karavel a choisi Splunk Enterprise pour garantir la qualité de ses services de réservation en ligne, mieux comprendre et améliorer les temps de réponses de ses pages web, et assurer le respect de ses accords sur les niveaux de service. Les logiciels Splunk sont maintenant utilisés tant par l'équipe technique que par l'équipe commerciale de Karavel, et servent à améliorer le service offert aux clients par toutes les marques de l'entreprise. L'équipe technique surveille en temps réel le processus de réservation en ligne, et utilise des alertes automatiques qui se déclenchent lorsque qu'un nombre anormalement élevé d'erreurs se produit. Elle examine d'autre part chaque semaine les rapports de Splunk exposant les temps de réponse des pages web. L'équipe commerciale observe elle aussi les rapports de Splunk, pour identifier et rectifier préventivement tout problème. En outre, les tableaux de bord de Splunk fournissent à l'équipe de direction des indices hebdomadaires de performances.

Karavel envoie à Splunk Enterprise les logs de son proxy inverse Varnish et de Tomcat. Un relayeur envoie ces données à deux indexeurs ayant une tête de recherche unique. Karavel utilise les logiciels Splunk conjointement avec AppDynamics, effectuant les recherches avec Splunk et utilisant ensuite AppDynamics pour des investigations en temps réel. Karavel a aussi construit des applications Splunk dédiées : une pour le processus de réservation en ligne et une pour l'accord de service de chaque site web.

Progrès

- Une meilleure vision relative aux accords sur les niveaux de service des sites web a diminué les temps de réponse de 82 %
- Identification en temps réel des divergences de configuration des réservations en ligne
- Optimisation du référencement naturel
- Équipe commerciale

En résumé, l'utilisation de Splunk Enterprise pour obtenir une intelligence opérationnelle détaillée sur son architecture web a apporté à Karavel une nouvelle compréhension approfondie de la performance de ses sites. Les équipes de toute l'entreprise, à la fois techniques et fonctionnelles, utilisent Splunk Enterprise pour améliorer la qualité de service fournie aux clients des sites du catalogue de Karavel.

« Nous proposons aujourd'hui une meilleure expérience de navigation sur notre catalogue de sites de voyages, avec des temps de réponse réduits et une identification immédiate de divergences de configuration potentiellement coûteuses. »

Architecte technique
Karavel





COORDONNÉES SYNCSORT

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

Tél +33 (0)1 46 67 46 10

Mail syncsortsamkt@syncsort.com

CONTACT

**Florence
RABUEL**

Regional Director

Fort de 40 ans d'innovation, d'expertise et de réussite, Syncsort aide les entreprises à travers le monde à intégrer les Big Data ... plus intelligemment.

Syncsort propose des solutions logicielles fiables, performantes, fonctionnant aussi bien sur Hadoop que sur le mainframe. Nous aidons nos clients, partout dans le monde à collecter, traiter et distribuer plus de données, en moins de temps, avec moins de ressources et en optimisant les coûts. Une majorité des entreprises du Top 100 sont clients de Syncsort et nos solutions sont utilisées dans plus de 85 pays pour off loader des applications legacy, onéreuses et inefficaces, accélérer les traitements dans votre data warehouse et votre mainframe et optimiser l'intégration de vos données dans le cloud.

En 1968, sept ingénieurs et mathématiciens se sont réunis pour réduire les coûts et la complexité liés aux données et à leur traitement. En perfectionnant une série d'algorithmes qui rationalisèrent le tri mainframe, utilisèrent moins de ressources, s'adaptèrent selon des variables d'environnement spécifiques et assurèrent une scalabilité pour faire face à des volumes de données toujours croissants, ils changèrent à jamais la manière dont les entreprises pouvaient envisager l'économie des données ... et ils fondèrent Syncsort.

Aujourd'hui, avec des milliers de déploiements à travers le monde sur toutes les plateformes majeures dont le mainframe, Syncsort fournit des solutions d'intégration Big Data aux entreprises gérant d'énormes quantités de données tout au long du Big Data Continuum, une manière plus ingénieuse de collecter, traiter une avalanche de données toujours croissantes..

Le siège social de Syncsort est situé au Nord Est du New Jersey. Syncsort est présent à l'international avec des filiales au Royaume-Uni, en France et en Allemagne, avec un centre de support aux Pays-Bas et des distributeurs. Syncsort dispose d'un réseau international de partenaires, revendeurs et distributeurs qui incluent ses solutions comme composants de leur offre informatique.

Syncsort est cette entreprise unique qui a brillamment fait évoluer son business model de l'ère du mainframe à celui d'Hadoop. En capitalisant sur cette expérience, Syncsort a développé des solutions sur mesure pour permettre aux entreprises d'optimiser leurs charges de travail à travers l'ensemble de leur architecture de données. Elles réalisent ainsi des économies et allègent la charge sur le datawarehouse et le mainframe, ce qui accélère les traitements par ceux-ci.

Notre portefeuille de solutions innovantes comprend notamment :

- ✓ DMX-h. Le premier moteur ETL qui fonctionne en mode natif au sein du cluster Hadoop, ce qui permet aux entreprises de mettre en place des flux de données complexes dans Hadoop et accélérer sans heurt les processus MapReduce — tels que HiveQL et Pig — sans devoir rédiger, générer ou optimiser du code manuellement.
- ✓ SILQ. Le premier et unique outil conçu pour aider les clients à comprendre des tâches complexes d'intégration de données SQL et à les offloader du datawarehouse vers Hadoop.
- ✓ Le moteur de tri et ETL le plus rapide pour mainframe, Hadoop et Linux/Unix/Windows.

syncsort



*Loune Jaffe,
PDG*

JE VOUS INVITE À EN APPRENDRE DAVANTAGE SUR NOTRE ENTREPRISE À

SYNCSORT.COM/FR



Efficacité accrue grâce au projet Hadoop

En tant qu'opérateur télécom majeur comptant plus de 21 millions de clients, SFR collecte quotidiennement une quantité énorme de données. L'objectif du projet Hadoop chez SFR était en premier lieu de réduire les coûts d'exploitation de ces données. Au fur et à mesure que le volume de données augmente, il est de plus en plus difficile et coûteux de stocker et traiter les données. En utilisant de façon optimale le cluster Hadoop et les outils puissants de tri et ETL de Syncsort, il est possible de limiter l'investissement dans du nouveau hardware.

Un projet qui ne se termine jamais

SFR a décidé de lancer le projet Hadoop à la fin de 2013. Dans une première phase, le choix s'est porté sur la distribution Hadoop Cloudera et début 2014, il a été décidé de travailler avec DMX-h de Syncsort pour loader et offloader des données à partir du cluster Hadoop. SFR utilisait déjà l'outil ETL DMX de Syncsort pour quatre autres projets, pour lesquels il avait fait ses preuves. Les très bonnes performances de DMX-h de Syncsort ont ensuite pu être démontrées dans le cadre du Proof of Concept pour le projet Hadoop. "Nous travaillons encore en mode batch," explique François Nguyen, en charge de la division 'Connaissance Client' chez SFR. " Cela implique que nous disposons d'une fenêtre de traitement réduite. Nos engagements précédents avec Syncsort se sont révélés très efficaces. Leurs capacités de tri n'ont pas d'équivalent sur le marché, et nous sommes très heureux que Syncsort soit l'un des véritables innovateurs sur le marché Hadoop. Leur mécanisme de tri est incorporé à la distribution Hadoop que nous utilisons.

Au cours de l'été 2014, les premières sources de données ont été loadées dans le cluster Hadoop, et davantage de sources seront ajoutées toutes les deux ou trois semaines. "Il ne s'agit pas d'un projet avec un horizon déterminé, nous allons continuer à ajouter des sources de données en permanence," explique François Nguyen.

A l'heure actuelle, la division Connaissance Client gère une base de données très volumineuse, qui n'arrête pas de croître. Les données traitées dans le cadre du projet Hadoop sont principalement des données structurées mais on y retrouve également des données semi-structurées provenant de logs.

En primeur

Le projet chez SFR est l'une des premières véritables expériences Big Data et Hadoop sur le marché français. Le projet permettra à SFR de réaliser des économies, grâce à la performance des solutions implementées.

De premiers résultats rapides

L'une des premières réussites du projet a été la capacité à concrétiser les premières étapes dans le délai imparti et selon le budget, ce qui est une prouesse compte tenu qu'il s'agit de l'introduction de technologies de pointe. La mise en route aussi rapide du projet est principalement due au fait que SFR n'a pas eu à effectuer une phase de tuning avant de commencer à utiliser DMX-h pour manipuler les données ni à réaliser de codage supplémentaire. Les résultats seront évalués en permanence au fur et à mesure que de nouvelles sources de données s'ajoutent au cluster Hadoop.

L'utilisation des puissantes capacités de tri de Syncsort DMX h n'ont pas rendues nécessaires l'acquisition de hardware supplémentaire.



TABLEAU SOFTWARE



COORDONNÉES TABLEAU SOFTWARE

Blue Fin Building
110 Southwark Street
SE1 0SU London - ROYAUME-UNI

Tél +33 9 70 44 93 95

Mail jsigonney@tableausoftware.com

CONTACT

**Julien
SIGONNEY**

Sales Area Manager

Connectez-vous à n'importe quelle source de données, quelle que soit son type ou sa taille. Analysez et créez des tableaux de bords et partagez-les en quelques clics. Tout cela en respectant votre architecture de données et vos protocoles de sécurité.

Rencontrez le nouveau visage des solutions décisionnelles : Tableau Software. Tableau aide les utilisateurs à visualiser et à comprendre leurs données. Quel que soit la source, le format ou le sujet de vos données, le logiciel primé de Tableau vous permet d'obtenir rapidement des analyses poussées et visuelles. Le résultat ? Chacun peut obtenir facilement des réponses, sans notion de programmation. Des tableaux de bords pour exécutifs aux rapports ponctuels, Tableau vous permet de partager vos analyses interactives en quelques clics. Plus de 21 000 entreprises et organisations, incluant certaines des plus importantes au monde, font confiance à Tableau pour obtenir rapidement des résultats, au bureau ou en déplacement.

Laissez de côté les solutions de décision traditionnelles et optez pour la plus facile d'utilisation, la plus rapide et la plus agile des solutions décisionnelles du marché. Avec tableau, les employés obtiennent ce qu'ils désirent et votre service informatique économise du temps et de l'argent.

Tableau Software concentre son offre autour de trois produits principaux pour les entreprises : Tableau Desktop, Tableau Server et Tableau Online. Nos équipes ventes et marketing France sont à votre disposition pour répondre à vos questions et vous aider à choisir la solution la mieux adaptée à vos besoins comme elles l'ont fait pour d'autres clients prestigieux tels que : La Croix Rouge, Unilever, BNP Paribas, Nokia ou encore eBay.

Répondez aux questions aussi vite qu'elles vous viennent 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, visualiser et créer facilement des tableaux de bord interactifs que vous pouvez ensuite partager en quelques clics. Et tout cela sans rien programmer !

Tableau se connecte en direct à la majorité des bases de données et feuilles de calculs et offre également une architecture en mémoire révolutionnaire qui augmente la rapidité des analyses. Quiconque sachant se servir d'Excel peut apprendre rapidement à utiliser Tableau.

Faciliter l'analyse de vos données.

Pour les analystes, les cadres, les informaticiens...
Pour tous...

- 10 à 100 fois plus rapide que les solutions disponibles sur le marché
- Créez facilement des tableaux de bord intelligents, quel que soit le type de données utilisées

La conception de l'avenir de l'une des principales sociétés de services financiers d'Europe

Partner : BNP Paribas

Contact : Pierre Thebault (BNP Paribas) et Edouard Beaucourt (Tableau Software)

Principaux objectifs du projet

Innover pour développer des parts de marché, remporter de nouveaux clients et augmenter la part du portefeuille de clients existants.

Présentation du projet

Le géant bancaire français a misé sur Tableau pour recueillir et analyser ses données en seulement quelques minutes. La division BNP Paribas du Sud-Ouest de la France dispose maintenant de la capacité de visualiser les tendances de géocodage utilisées dans le cadre des campagnes commerciales et marketing de la banque. Et l'aperçu partagé et détaillé de la performance de 2 000 agences à travers la F contribue à renforcer la compétitivité.

Mise en avant des innovations incluses dans le projet

Pierre Thebault, géoanalyste, est chargé d'analyser et d'identifier les tendances de géocodage au sein d'une région. Grâce à Tableau, il traque les opportunités de prospector de nouveaux clients et de développer de nouvelles campagnes marketing extrêmement ciblées : « Tableau est le moyen le plus rapide de passer des données aux décisions ». Il est à même de localiser et de visualiser le nombre de prospects dans la région de Toulouse, et de les segmenter par revenus, par risques et selon les services financiers de BNP Paribas dont ils disposent déjà. S'ils ont souscrit un prêt immobilier, ils constituent un candidat potentiel pour la souscription d'une assurance habitation ou d'une carte de crédit. Tableau permet de visualiser l'ensemble des données en quelques secondes, ce qui permet son équipe de transmettre ce public cible segmenté aux équipes commerciales internes de la banque pour un suivi immédiat.

BNP Paribas recherche constamment des manières de différencier ses services en offrant une expérience plus localisée. Tableau permet à la banque d'étudier la localisation des distributeurs automatiques de billets des concurrents et d'identifier les sites potentiels pour de nouveaux distributeurs : « Comparer la localisation de nos clients avec la disponibilité du réseau existant de distributeurs automatiques de billets nous permet de cibler de nouveaux emplacements de distributeurs », déclare Pierre Thebault.

Malgré l'expansion des services bancaires en ligne, les français restent attachés aux agences physiques. Tableau représente la source primaire pour les rapports sur la performance des 2 000 agences BNP Paribas française. Directeurs et responsables partagent un aperçu détaillé de la performance de chacune des agences : chiffre d'affaires, rentabilité, personnel, adoption des services bancaires en ligne de la part des clients, etc. « L'accès rapide à des données fiables, grâce à Tableau, se traduit par une prise de décision vive et fiable concernant les agences », affirme Pierre Thebault. Il poursuit, « Avant Tableau, nous utilisions des feuilles de calcul et une base de données Access. Cela nous prenait des heures, des semaines, voire des mois, pour trouver les données correctes et les réponses dont nous avions besoin. Désormais, grâce à Tableau, l'intégration homogène et la simplicité d'utilisation signifient que nous obtenons les réponses en quelques minutes. Cela rend-il la banque plus perspicace ? Oui, sans l'ombre d'un doute. »

Prochaines étapes

La majeure partie de l'inspiration pour ce programme couronné de succès appartient à Mydral, le partenaire d'exécution de BNP Paribas. La vision, la perspicacité et l'expérience de l'équipe de Mydral ont contribué au lancement de Tableau en quelques jours. Mydral a également fourni à BNP Paribas trois jours de formation et de transfert des connaissances afin de garantir que l'équipe soit opérationnelle presque immédiatement.

Ce projet attire désormais l'attention d'autres services de la banque. BNP Paribas déploie actuellement Tableau sur les sites régionaux de Paris et Nantes dans le cadre d'un projet de segmentation de la clientèle. Ces données sont utilisées pour comprendre les tendances, concevoir de nouvelles campagnes, et maximiser la valeur de BNP Paribas tirée des données marketing.





COORDONNÉES

TIBCO JASPERSOFT

25, rue de Balzac 75008 Paris France

Tél + 33 (0)1 44 51 70 90

Mail gcarbonnel@jaspersoft.com

CONTACT

Leïla

BOUTALEB BROUSSE

EMEA Marketing Manager - Analytics

TIBCO Software Inc. (NASDAQ : TIBX) est un leader mondial dans le domaine des logiciels d'infrastructure et décisionnels. Qu'il s'agisse d'optimisation des stocks, de vente croisée ou de prévention des crises, TIBCO offre un concept unique nommé Two-Second Advantage®, autrement dit la possibilité de capturer la bonne information au bon moment et de réagir en fonction de cette information plus vite que la concurrence. Grâce à son vaste éventail de produits et services novateurs, TIBCO est un partenaire technologique stratégique plébiscité par les entreprises dans le monde entier. Pour en savoir plus sur TIBCO, rendez-vous sur www.tibco.com.

TIBCO, Jaspersoft, Two-Second Advantage et TIBCO Software sont des marques commerciales ou des marques déposées de TIBCO Software Inc. et/ou de ses filiales aux Etats-Unis et/ou dans d'autres pays. Amazon Web Services, AWS et Redshift sont des marques commerciales d'Amazon.com, Inc. ou de ses filiales aux Etats-Unis et/ou dans d'autres pays. MongoDB est une marque commerciale de MongoDB, Inc. Tous les autres noms de produits, d'entreprises et de marques mentionnés dans ce document appartiennent à leur propriétaire respectif et ne sont mentionnés qu'à des fins d'identification.



YOUR ACCESS TO
FAST DATA
BEGINS NOW.



Integrate
people, processes,
and systems



Analyse
for actionable
insights



Act
on key events
in real time

Ericsson utilise Tibco Jaspersoft pour l'analyse Big Data sur MongoDB

La société ajoute le décisionnel en libre-service à sa solution Ericsson Multiscreen TV.

Le défi : les requêtes Big Data dans MongoDB

Les périphériques mobiles sont devenus partie intégrante de nos habitudes de travail et de loisirs. Pour garantir leur compétitivité, les fournisseurs de services de télécommunications/par câble doivent assurer à leurs clients un accès aux contenus vidéo sur tous leurs périphériques.

Leader du secteur des télécommunications, Ericsson a répondu aux besoins de TV multi-écran des opérateurs grâce à sa solution Multiscreen TV : elle permet en effet aux fournisseurs de services télévisuels de contrôler et de gérer la fourniture des contenus, de manière centralisée, vers quasiment tous les types de périphériques grand public.

Ericsson a spécifiquement conçu sa solution Multiscreen TV pour aider les fournisseurs de services télévisuels à réduire les coûts et générer de nouvelles sources de revenus. Cette solution inclut des rapports intégrés sur deux volets :

- Opérationnel – données sur l'utilisation et l'intégrité du système pour l'optimiser.
- Marketing – données sur le contenu consulté et utilisé par les clients afin de soutenir les prises de décision concernant les offres et le marketing, ainsi que pour calculer les redevances.

Ericsson a remplacé sa précédente base de données relationnelles par MongoDB (open source NoSQL). Cela lui a permis de réduire les délais de réponse, les temps de latence, la géoredondance et les indisponibilités lors des mises à jour.

Lorsqu'Ericsson a développé la toute dernière version de sa solution Multiscreen TV, la société avait besoin d'un moteur décisionnel puissant, capable d'interroger de grands volumes de données dans MongoDB. Auparavant, Ericsson utilisait JReport, qui ne permettait pas l'intégration avec MongoDB ni ne produisait les types de formats attendus par les fournisseurs de services télévisuels : les fichiers CSV.

« Nous recherchions un produit qui permettrait une intégration directe avec MongoDB », explique Jon Anderson, responsable de la stratégie produits pour Ericsson Multiscreen TV. « Nous souhaitons pouvoir produire de beaux rapports tout en disposant de capacités d'exportation conséquentes. Nombre de nos clients exportent vers un entrepôt de données plus vaste. »

L'intégration Jaspersoft/MongoDB sans ETL

Si Ericsson a choisi Jaspersoft, c'est pour pouvoir proposer des fonctionnalités de reporting plus riches dans sa solution Multiscreen TV. Avec JasperReports Server, la société remplit un grand nombre de ses objectifs :

- Compatibilité avec MongoDB : le connecteur Jaspersoft intègre MongoDB à la suite décisionnelle Jaspersoft, pour fournir des fonctionnalités flexibles et abordables de reporting, d'analyses ad hoc et de tableaux de bord, à partir de données MongoDB.
- Intégration sans ETL : Jaspersoft permet l'intégration directe avec MongoDB, sans environnement ETL (Extract, Transform, Load), ce qui accélère et facilite l'obtention des données depuis Jaspersoft.

FAST FACTS

CUSTOMER
Ericsson AB

SECTEUR
Technologie

REPÈRES TECHNOLOGIQUES

- JasperReports Server et iReport Designer
- Redhat Enterprise Linux sur VMware
- Interface utilisateur sur Ruby on Rails
- Serveur d'applications JBoss
- Base de données MongoDB
- Javascript

LE DÉFI

Ericsson recherchait un puissant moteur décisionnel de reporting pour permettre aux utilisateurs de sa solution Multiscreen TV d'interroger de grands volumes de données dans MongoDB.

LA SOLUTION

JasperReports Server répond aux exigences de l'environnement technologique de la société ; il offre des fonctionnalités de reporting Big Data et d'intégration avec MongoDB, sans ETL.

RÉSULTATS

- Economies de coûts : la société économise des milliers de dollars et d'heures de travail en s'affranchissant de la maintenance de deux bases de données distinctes, MongoDB et une base de données relationnelles.
- Création rapide de rapports : en quelques heures seulement, les développeurs Ericsson peuvent créer un rapport à l'aide d'iReport Designer.
- Vision anticipée : dans certains cas, les clients peuvent obtenir une vision anticipée de plusieurs semaines grâce au reporting ad hoc en libre-service.





COORDONNÉES

TALEND

9 rue Pagès 92150 Suresnes

Tél +33 1 46 25 06 00
 Fax +33 1 42 04 36 67
 Mail ccornavin@talend.com

CONTACT

**Charlotte
 CORNAVIN**

Marketing Manager

Les solutions d'intégration de Talend aident les entreprises à tirer le meilleur parti de leurs données. A travers le support natif des plates-formes modernes de Big Data, Talend réduit la complexité de l'intégration, tout en permettant aux départements informatiques de répondre plus rapidement aux besoins métiers, le tout pour un coût prévisible. Reposant sur des technologies open source, les solutions hautement évolutives de Talend répondent à tous les besoins d'intégration, actuels et émergents.

Support natif de Big Data

A la différence des solutions d'intégration héritées, Talend réside nativement dans un environnement Hadoop, sans empreinte de déploiement. Nativement intégré dans les distributions Hadoop majeures comme Cloudera, Hortonworks ou MapR, Talend utilise la sécurité native de Hadoop et fait de la qualité de données dans Hadoop.

Des solutions ouvertes

Comme Hadoop, Talend est engagée sur de l'Open Source et des standards ouverts et les avantages que ceux-ci apportent : un écosystème innovant, pas de dépendance vendeur, développement plus rapide et plus agile, ainsi que le support d'une vaste communauté. L'utilisation des standards et du large écosystème de Talend signifie que les développeurs peuvent très facilement adopter les solutions Talend et que les ressources Talend peuvent facilement être trouvées sur le marché.

Prévisibilité

Le modèle de prix des souscriptions sans conteneur d'exécution permet de faire évoluer de manière prévisible les données et les projets, sans avoir à faire évoluer le coût de l'intégration. La conformité aux standards Java, Eclipse et Big Data réduit le temps de développement et de maintenance des projets. Ainsi, les coûts opérationnels sont également prévisibles.

A l'épreuve du futur

Comme l'aventure Big Data passe des tests à l'analytique et au temps réel/aux cas opérationnels, Talend livre une plateforme d'intégration unique, complète et unifiée répondant à tous les besoins d'aujourd'hui et de demain.

Talend facilite l'intégration des Big Data:

- Mettez-les au service de vos enjeux, processus et prises de décision
- Simplifiez leur prise en main, leur partage et minimisez leurs coûts de gestion
- Maîtrisez leur qualité, leur accessibilité, leur gouvernance

Les solutions d'intégration et de qualité de Talend permettent aux entreprises de tirer profit de toutes leurs données quels que soient leur source, type ou taille.

Buffalo Studios optimise la valeur des données pour accroître sa compétitivité dans le secteur des jeux sociaux



Porteurs du Projet

Barry Sohl, Chief Technology Officer chez Buffalo Studios et son équipe informatique

Objectifs

- Obtention d'une plus grande flexibilité au niveau de l'intégration des données
- Prise de décision plus rapide et efficace

Présentation

Dans le monde du jeu virtuel sur les réseaux sociaux, chaque activité effectuée par un joueur est automatiquement enregistrée et transmise à l'éditeur ce qui lui permet ensuite de mieux orienter les offres envoyées aux joueurs (crédit supplémentaire, bonus, outils, etc). En termes de volume, cela représente plus de 10 000 événements traités toutes les secondes et la prise en charge de plus de 100GB d'informations par jour.

Buffalo Studios utilise un Data Warehouse de haute performance pour lancer ses rapports SQL sur les données de joueurs. Dans leur cas, ils utilisent Apache Flume pour ingérer le million de lignes de données qu'ils récupèrent chaque jour (environ 100GB), le charger dans leur Data Warehouse grâce à Talend pour ensuite le réutiliser dans un système de reporting et de visualisation.

Aspect Novateur

La nouvelle implémentation de Talend dans le système d'information de Buffalo Studios leur permet désormais de connaître en quasi temps-réel le comportement d'une nouvelle fonctionnalité dans le jeu. Cela leur a aussi permis de réduire les fraudes : « Nous pouvons maintenant détecter une activité frauduleuse en moins de 10 minutes » déclare Barry Sohl. Avant cela, si nous avions un bug et le laissions en place une journée entière, cela avait des conséquences catastrophiques pour notre activité mais c'est maintenant résolu ».

Valorisation observée de la donnée

Aujourd'hui Buffalo Studios utilise la solution d'intégration Big Data de Talend, et a obtenu la flexibilité qui lui manquait auparavant. « Quand nous avons des besoins d'analyse spécifiques, nous pouvons maintenant écrire nos propres composants Java et les incorporer à la solution Talend » explique Sohl. « Et quand il s'agit d'ajouter de nouveaux points de mesure des données, nous ne bloquons plus le processus à notre niveau, car nos développeurs travaillent simplement avec les outils sans avoir à effectuer manuellement de laborieuses modifications. »

Ce type de gestion allégée, à son tour, permet à Buffalo Studios d'ajouter de nouveaux points de données chaque semaine. Le problème des blocages au niveau des fichiers logs bruts est désormais totalement résolu.

Cet avantage revêt une importance capitale pour Buffalo Studios. En effet, si leur service technique n'arrive pas à livrer les informations cruciales à la direction, les décisionnaires restent dans l'incertitude. « Nous avons besoin de mesurer la performance de nos fonctionnalités, qu'elle soit bonne ou mauvaise, presque en temps-réel. Plus l'exploitation des données sera simple pour nous, plus l'équipe BI pourra en tirer parti. Talend nous aide à optimiser les processus, à améliorer la rapidité et à accroître la valeur que nous pouvons tirer de nos données, » précise Sohl.

De plus, les ingénieurs de l'équipe technique gagnent en productivité, puisqu'ils n'ont plus de problèmes d'intégration à résoudre et peuvent de nouveau se concentrer sur la construction de technologie innovante. C'est ce qui assure la croissance et la fidélisation des clients.

Etapes à venir et démarche Big Data dans l'entreprise

« Nous portons beaucoup d'intérêt à l'exploration des nombreuses possibilités offertes par Talend en termes de qualité des données. Je pense qu'il y aura un vrai potentiel à exploiter en élargissant le savoir-faire de Talend à d'autres applications, et que nous pourrions alors bénéficier de nombreux avantages au fur et à mesure que nous progresserons, » déclare Sohl. Buffalo Studio annonce que l'implémentation de Talend fonctionne si bien que d'autres sociétés du groupe Cesar Interactive Entertainment réfléchissent à le mettre en place.

Autres informations

Plus d'informations :

<http://fr.talend.com/resources/customer-reference/buffalo-studios-maximizes-data-value-to-build-competitive-edge-in>





COORDONNÉES

TERADATA FRANCE

Antony Parc I : 2-6 place du Général de Gaulle - 92160 Antony

Tél +33 (0)1 81 89 15 00

Mail severine.vigneron@teradata.com

CONTACT

**Séverine
VIGNERON**

Marketing Manager Teradata France

Teradata helps companies get more value from data than any other company. Our big data analytic solutions, integrated marketing applications, and team of experts can help your company gain a sustainable competitive advantage with data. Teradata helps organizations leverage all of their data so they can know more about their customers and business and do more of what's really important. With more than 10,000 professionals in 43 countries, Teradata empowers organizations to become data-driven businesses that exploit data for insight and value. More than 2,500 customers worldwide span industries including automotive and industrial; communications; consumer goods; e-commerce; financial services; government; healthcare and life sciences; hospitality and gaming; insurance; manufacturing; media and entertainment; oil and gas; retail; transportation, distribution, and logistics; travel; and utilities.

A future-focused company, Teradata has driven innovation for thirty-five years. Teradata's workload-specific platforms integrate an organization's data into a unified view of the business. Our database software includes the unparalleled Teradata® Database, and, for big data analytics, the Teradata Aster® Discovery Platform and Hadoop Portfolio. The unique Teradata Unified Data Architecture™ helps customers discover and operationalize insights by integrating multiple technologies, including Hadoop, into a robust, hybrid architecture. Teradata's marketing and analytic applications—available on premise or in the cloud – leverage data to improve marketing effectiveness, determine profitability, and forecast demand. These powerful solutions, along with years of hands-on experience working with the world's leading enterprises, enable customers to maximize the value of their data and serve their customers better. Teradata is recognized for technological excellence, sustainability, corporate social responsibility, ethics, and business value.

When the world gets smaller, the data gets bigger

When you're able to collect, unify, and analyse all of the data that surrounds your business, you can uncover the insights that matter most. It could be an insight that changes the way one customer sees your company or how your company sees the world. Learn how our analytic platforms, applications, and services can help you know more so you can do more at Teradata.com/DataDriven.

What would you do if you knew?™

TERADATA®

Big Data: Practicalities, Realities and Getting Started

Big Data is real and it's growing at mind-dazzling speed

Big Data refers to the vast spectrum of digital information arising from an increasing variety of data sources which come in wide ranging and ever-expanding data formats.

This includes structured data derived from common business systems, and unstructured data, or more correctly 'multi-structured' data, in its diverse variations – from weblogs and sensor data, to data in social networks, text documents, or all forms of digital images and video content. And the list goes on... Teradata's view on Big Data however goes even a step further.

We believe, fundamentally, that ALL data has relevancy for companies to make better, more informed business decisions. Identifying the valuable information amongst the irrelevant data is not always obvious without initial analysis, and often what has no value to one purpose is critical to another. As a result it becomes more expensive to decide what is valuable than to retain all data and consider later how it can be used to create value – this is the "data lake" concept.

With increasing scope for analytics, and recognition of the competitive advantage of being data-driven, it becomes ever more important for companies to find efficient processes and supporting technologies to derive business value from the "data lake" in the timeliest possible manner. This is exactly where Teradata comes in.

Teradata combines the most advanced, highly scalable Big Data technologies under one umbrella in an integrated manner. We call this the "Unified Data Architecture", blending in with what leading analyst firm Gartner refers to as the "Logical Data Warehouse".

With the Teradata Database, Hadoop, and the Aster Discovery Platform the Teradata Unified Data Architecture provides organisations with a rich set of capabilities to support them in their pursuit of deriving value from analytics based on ALL the data.

How Companies are Innovating with Big Data

Cardinal Health : Cardinal Health has used data to optimize the supply chain ensuring that the complicated array of raw materials that are essential for high quality healthcare arrive in good condition, when the patients need them.

McCain Foods : McCain Foods is a global food giant, with great pizzas, vegetables, appetizers and desserts. And they've transformed their culture with data. McCain has taken more than 22,000 reports and 3,000 personal reporting systems and put the data in one place. "We took 30,000 versions of the truth, and brought it down to one version of the truth."

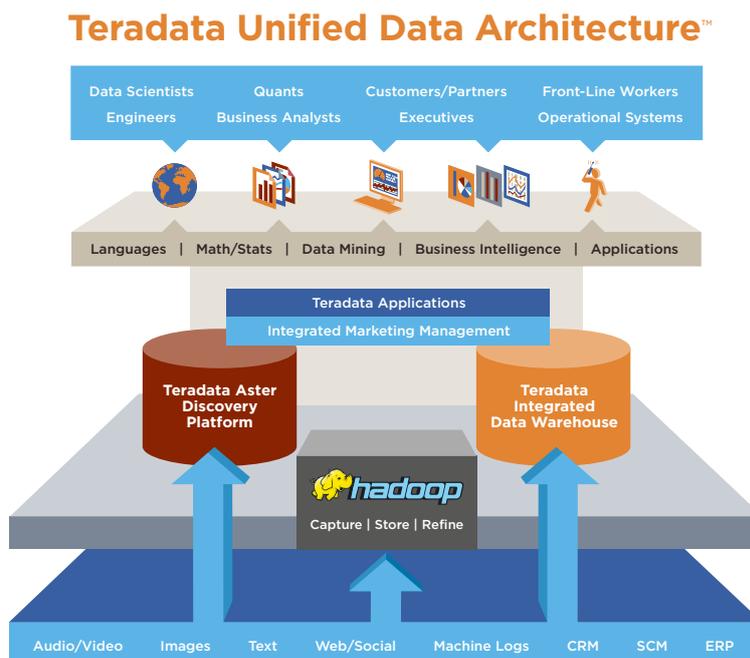
Verizon Wireless : How does the biggest wireless carrier in the United States keep their customers happy? Provide the best network with the best customer experience and listen. How does Verizon Wireless listen? With an Unified Data Architecture.

Barnes and Noble : Taking data from more than 1,300 stores and one of the largest loyalty programs in the US, Barnes and Noble combines it with data from their e-commerce site. Using data analytics, Barnes and Noble knows exactly who their customers are, and what they want to buy – or read – next.

Discover More Insights and Outcomes through our customer "data driven business Success Stories":

<http://blogs.teradata.com/customers/> or visit [Teradata.com](http://www.teradata.com)

Contact us to learn more on Teradata, solutions and our unique approach to Big Data.





COORDONNÉES

WISEO

38 rue de Ponthieu – 75008 Paris

Tél **01 56 56 71 00**

Mail contact@viseo.com

CONTACT

**Yves
COINTRELLE**

*Directeur de la stratégie et
du développement Business Intelligence*

En s'appuyant sur ses expertises dans les domaines du digital, des technologies, des process et de la data, VISEO vous accompagne dans vos chantiers de transformation et vous aide à faire émerger de nouvelles idées, de nouvelles offres, de nouveaux modes de collaboration, de nouvelles manières de produire et de vendre.

Avec 1 100 collaborateurs répartis sur quatre continents et un chiffre d'affaires supérieur à 110 millions d'euros, VISEO vous propose une alternative aux intégrateurs, agences et cabinets de conseil, en imposant une approche unique, souple, innovante et structurante. VISEO vous apporte ainsi une réponse globale et vous aide à mettre en œuvre des projets complexes pour faire face aux enjeux numériques qui impactent les systèmes d'information, les métiers ou les organisations

Adresser aujourd'hui les enjeux structurants de demain : VISEO mobilise l'ensemble de ses expertises pour vous permettre de comprendre et anticiper les sujets clés dans l'évolution de vos activités.

VISEO et le BIG DATA : et si nous parlions conseil plutôt que Technologie ?

- Le Big Data offre un nouveau regard sur les données en ouvrant de nouveaux moyens pour les saisir, les stocker, les gérer, les traiter et les analyser.
- Grâce à la combinaison unique de l'ensemble de ses compétences Big Data (Conseil, Data Scientist, Business Intelligence), VISEO accompagne ses clients autour de 3 grands axes :
 - Découvrir le potentiel et identifier les cas d'usage pertinents créateurs de valeur
 - Tester par la mise en place de POC
 - Déployer & Exploiter les données
 - Accompagner les évolutions organisationnelles



La solution dédiée aux *data scientists*

Des fonctions poussées d'exploration de donnée
Réconcilie statistiques et data visualisation en un seul outil

Plus d'informations : contact@deltametric.info

À tester sur

www.deltametric.net



Réussir son projet Big Data : la méthode

Les motivations qui amènent une entreprise à initier un projet Big Data sont souvent multiples. Une direction générale soucieuse de s'assurer que le capital informationnel de l'entreprise recèle de la valeur inexploitée ; une démarche exploratoire consécutive à une prise de conscience de possession d'importants gisements de données en sommeil et souvent en silo ou encore des directions métiers soucieuses d'améliorer leurs connaissances et leurs capacités d'analyse et de prise de décision ; souhaitant investiguer de nouvelles pistes d'optimisations, d'amélioration de productivité ou de rentabilité.

Quelle qu'en soit l'origine, la démarche associée à la mise en œuvre d'un projet Big Data est semée d'embûches et d'échecs. Les données ne sont pas toujours disponibles, exhaustives ou n'ont pas toujours la qualité espérée pour étayer les analyses souhaitées. Les budgets à mettre en regard des objectifs à atteindre ne sont pas toujours couverts par les gains procurés par la mise en œuvre des systèmes.

Pour éviter ces écueils, valider les modèles ou les ROI, sans périr ou engloutir des budgets pharaoniques, VISEO a développé une méthodologie mêlant ateliers de détermination des objectifs et data scientist outillés, permettant de valider au plus tôt (sur des échantillons représentatifs) les modèles.

Dès la phase amont nous privilégions une approche « Test'n Learn » permettant de se conforter dans la capacité des objectifs à être tenus. Réduire les risques, tenir les promesses, lancer des projets crédibles.

Conscients que les technologies sous-jacentes sont certes importantes mais restent une conséquence de l'objectif, nous proposons un panorama des technologies et des architectures. Nous tenons compte des choix technologiques déjà opérés par l'entreprise, des contraintes techniques, budgétaires mais également humaines. Pas de choix préétabli : chaque projet est différent et doit être abordé en tenant compte de ces multiples paramètres.

EXPLOITEZ LA RICHESSE DE VOS DONNÉES

- Analyse de vos données par nos data scientists
- Etudes d'opportunités métier
- Préconisations d'architectures et outils
- Mise en œuvre



Congrès & Expo

Rendez-vous les
10 & 11 mars 2015
CNIT Paris La Defense



**Retrouvez les leaders du secteur sur
l'événement majeur du Big Data en France**

- *2 jours de conférence et d'exposition*
- *150 intervenants experts français et internationaux*
- *120 exposants*
- *4000 professionnels du Big Data*

4^e édition

www.bigdataparis.com by **corp.**
in Corporations we Trust



GÉREZ

des millions de clients
individuellement.

Imaginez un instant que vous puissiez entretenir une relation privilégiée avec chacun de vos clients, aussi nombreux soient-ils. Avec les solutions analytiques SAP, d'immenses volumes de données structurées ou non sont transformés en un puissant outil d'intelligence prédictive qui vous permet d'étonner vos clients, d'aller au-devant de leurs désirs. Où qu'ils soient, en magasin, en ligne ou sur une application, ils ressentiront ainsi l'attention particulière que vous leur portez.

Avancez plus que jamais sur saprulikeneverbefore.com/france

PLUS DE RÉSULTATS.



CONSTRUISEZ VOTRE HUB DE DONNÉES D'ENTREPRISE AVEC

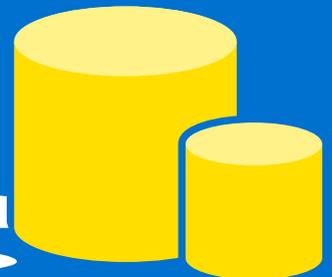
syncsort

ALLÉGEZ VOTRE
MAINFRAME

ALLÉGEZ VOTRE DATA
WAREHOUSE

NOUS SOMMES
LES SPÉCIALISTES
DE L'OFFLOAD !

✓ RAPIDE
✓ FIABLE
✓ SECURISE



DÉCOUVREZ COMMENT L'ACCÈS À TOUTES VOS
DONNÉES SUR HADOOP AU MOYEN D'UN SEUL ET
MÊME OUTIL PEUT FAIRE UNE DIFFÉRENCE ÉNORME !

TÉLÉCHARGEZ NOTRE GUIDE GRATUIT :

syncsort.com/guide



— Cinq étapes pour —
OFFLOADER
— VOTRE ENTREPÔT DE DONNÉES —
— AVEC HADOOP —

UN GUIDE DE DEMARRAGE RAPIDE