



# Guide du big data

## *Les acteurs à suivre pour accompagner vos projets*

Acteurs historiques ou nouveaux entrants, ils maîtrisent la donnée depuis les infrastructures jusqu'aux services, en passant par l'organisation, l'exploration, le traitement, l'analyse et l'automatisation de grands volumes de données. Une sélection de la rédaction pour exploiter de manière optimale vos big data.

# Le big data : un défi pour les PDG

Toutes les grandes entreprises françaises n'ont pas encore pris conscience de l'enjeu stratégique du *big data*, qui n'est pas une question technologique. Tel est le point de vue de **François Bourdoncle**, président de la société FB&Cie et co-fondateur d'Exalead.

PROPOS RECUEILLIS PAR CHRISTOPHE GUILLEMIN



## Alliancy, le mag. Les entreprises françaises ont-elles bien pris conscience des enjeux autour du big data ?

**François Bourdoncle.** Il faut être clair, les grands groupes français n'ont pas encore tous pris la mesure de la révolution industrielle que représente le big data. Ils estiment qu'il s'agit en grande partie d'une question de technologie. Mais le big data n'est pas tant une révolution technologique qu'un bouleversement complet du fonctionnement des entreprises. Elles passent d'une organisation en silos à une structure transversale, où les informations circulent mieux et sont largement partagées. Cette révolution va mettre à mal nos industries si elles ne réagissent pas. Car tous les nouveaux usages sont basés sur l'analyse des données. La question n'est pas de dire : « *Je vais donner un coup de jeune à mon business historique grâce au big data* », mais : « *comment vais-je répondre à des concurrents qui maîtrisent l'analyse de donnée et seront bientôt en capacité de détruire mon business* ».

## Quels conseils donneriez-vous à une entreprise souhaitant évoluer vers le big data ?

Encore une fois, il ne faut pas commencer par appréhender le big data comme une question technologique. Il s'agit d'abord d'une nou-

velle organisation de l'économie et d'une nouvelle stratégie où l'usage client est au cœur du business. Il faut donc que ce soit la direction elle-même qui soit moteur dans cette transformation numérique. C'est au PDG et à ses collaborateurs les plus proches de se pencher sur la question et éventuellement de recruter de nouvelles compétences si nécessaire. Il faudrait plus souvent faire siéger des responsables expérimentés dans l'industrie du numérique dans les conseils d'administration. Une fois que le cap est donné, les projets technologiques peuvent alors être lancés. Pour résumer : devant le défi du big data, les dirigeants doivent se dire que c'est leur problème à eux, et pas celui de leurs équipes.

## En juillet 2014, vous avez présenté au gouvernement un plan big data, dont l'ambition était de faire de la France un leader dans ce domaine, face notamment aux géants américains. Où en le développement de ce plan ?

Nous sommes aujourd'hui en phase opérationnelle du plan. La feuille de route que nous avons établie a été rendue publique en janvier 2015. C'est une étape importante car elle donne de la visibilité aux mesures que nous préconisons et aux projets industriels qui en découlent. Je rappelle que ce plan compte quatre volets importants. Le pre-

mier est le volet industriel. L'objectif est de favoriser les échanges entre les grands groupes français autour du big data et de mettre en lumière des projets lancés par certains industriels. Tous les deux mois nous organisons des « points d'étape » de projets big data présentés par ces entreprises. Nous suivons pour l'instant trois ou quatre projets dans des domaines aussi divers que l'assurance, l'énergie ou les transports. Ces projets, labélisés « plan big data », sont étudiés et présentés à d'autres acteurs du secteur. Nous les faisons travailler par petits groupes afin de favoriser les échanges. Et les retours sont très positifs. Les expériences des uns, viennent nourrir les réflexions des autres.

## Et les autres volets ?

Le deuxième volet, dit « régalién », doit faire de l'État un exemple dans le domaine du big data, en lançant de grands projets dans les domaines de la santé et de l'emploi. Cela avance. Nous travaillons notamment avec le ministère de la Santé et la CNAM autour de l'exploitation des données de santé. De même, nous collaborons avec le ministère du Travail et Pôle Emploi autour des données de l'emploi. Nous étudions par exemple comment fluidifier le marché de l'emploi grâce aux outils numériques. Nous voulons également développer les compétences du big data via un



volet éducation et formation. Dans ce domaine, nous sommes en train de réunir différents acteurs, notamment des universités et des écoles d'informatique, pour donner une impulsion aux cursus autour du big data. Il y a un gros déficit de compétences en France et il faut réellement faire décoller ces formations.

**Le quatrième volet concerne l'adéquation du big data avec la Loi Informatique et Libertés. N'est-ce pas un domaine plus difficile à faire avancer ?**

C'est effectivement un domaine sensible. Mais nous travaillons efficacement avec la CNIL sur ce sujet. L'idée n'est pas de changer immédiatement la loi mais d'adopter une logique de « certification de processus » et de « label éthique » autour des solutions big data. Cela permettra aux industriels qui y auront recours, de promouvoir une approche éthique de l'utilisation des données personnelles et d'en faire un atout concurrentiel, notamment à l'exportation. En début d'année, nous avons organisé une première matinée de travail autour de l'écosystème de la mobilité en milieu urbain, et notamment des voitures connectées. Il y avait une cinquantaine de représentants, notamment des constructeurs automobiles, mais aussi des assureurs ou des logisticiens. Ils ont échangé sur l'utilisa-



*François Bourdoncle, co-chef de projet du « plan big data » de la nouvelle France industrielle, lancé par le gouvernement fin 2013 et président de la société FB&Cie.*

tion des données de manière transversale, c'est-à-dire partagée par différents acteurs pour assurer un même service. Cette rencontre a été un franc succès. Elle a permis à des acteurs qui ne se connaissaient pas forcément de se rencontrer. Et certains se sont rendus compte qu'ils s'autocensuraient plus que nécessaire dans l'usage des données des voyageurs. Nous allons poursuivre ces rencontres en 2015.

**Il était également question de mettre en place un ou plusieurs centres de ressources technologiques permettant aux start-up d'accéder aux données de grandes structures, publiques ou privées, ainsi qu'aux infra-**

**structures nécessaires pour les exploiter, comme les moyens de calcul. Où en est cette création ?**

L'appel à manifestation d'intérêt a été lancé par la DGE l'été dernier. Cet appel d'offres a été dépouillé en novembre. Il s'appelait « challenges big data » et propose donc de récompenser des start-up ayant un projet particulièrement innovant à accéder à des ressources de grandes entreprises ou d'institutions publiques pour les aider dans leur projet. Il y a cinq centres de ressources en cours d'élaboration dont deux à Paris. Un de ces centres de la capitale est déjà opérationnel depuis novembre 2014 et est dédié au secteur des assurances. Il est sponsorisé par le groupe Axa.

## Acteurs historiques

### Teradata, une architecture de données unifiée

Teradata est une société fondée en 1979 en Californie. Revendiquant la position de « leader mondial des solutions analytiques », elle est spécialisée dans l'analyse big data, les applications marketing intégrées et les services. Présent dans 43 pays, Teradata emploie plus de 10 000 collaborateurs. Ses clients sont de grandes entreprises des secteurs des biens de consommation, des services financiers, de la santé, de l'automobile, de l'énergie, du voyage, du transport, de l'industrie ou des télécommunications.

Son offre UDA, pour « Unified Data Architecture » (architecture de données unifiée), combine « toutes les technologies et l'expertise développées par Teradata pour accompagner ses clients dans les étapes de gestion et d'analyse de ses données. Cette offre intègre des solutions complètes comprenant des plateformes analytiques de données, des entrepôts de données, des outils d'analyse ainsi que des applications marketing ».

UDA intègre notamment les solutions d'infrastructure de Teradata, pour stocker les données, des technologies d'Aster (société acquise en 2011) pour des analyses avancées et la gestion de données multi-structurées. L'architecture intègre également des passerelles vers d'autres solutions, dont Hadoop. « UDA s'adapte à tous les outils de l'entreprise. Cette architecture va connecter tous les éléments permettant de mener à bien un projet big data », résume Teradata.

fr.teradata.com

**TERADATA.**

### EXALEAD OneCall : quand le big data transforme la relation client

Marque Dassault Systèmes depuis 2010, EXALEAD propose des solutions d'exploration et d'analyse de données basées sur un moteur d'indexation et de fonctionnalités sémantiques avancées. Depuis 2013, le groupe de haute technologie propose la solution EXALEAD OneCall, dont le principe est d'utiliser la puissance de sa technologie de traitement du big data pour améliorer l'interaction avec le client et in fine sa satisfaction, en équipant les collaborateurs de centres d'appels, agences... mais aussi les back offices et le management. « Quel que soit le type de données ou leurs sources, EXALEAD OneCall agrège, analyse et restitue une information riche et contextualisée aux conseillers, agents, responsables d'équipe... », explique-t-on chez Dassault Systèmes. « En fournissant une véritable vue 360° de l'information sur une seule interface, elle permet de mieux interagir avec le client, d'avoir une discussion plus efficace, de pouvoir anticiper et proposer et ainsi d'améliorer l'expérience client ».

OneCall a été conçu sur la technologie EXALEAD CloudView qui consolide, analyse et enrichit dans un même index des données de sources différentes (sites Web, serveurs de fichiers, CRM, bases de données, annuaires...), structurées ou non. OneCall bénéficie de la haute performance de l'index et des capacités de traitement sémantique avancé et de Web Mining développées depuis 15 ans. Cette solution est non intrusive et se déploie au-dessus des systèmes et applications en place dans l'entreprise.

**DASSAULT SYSTEMES**

### EY, transformer et créer de la valeur avec le big data

Le cabinet d'audit et de conseil EY (anciennement Ernst & Young) a développé une offre de conseil pluridisciplinaire visant à accompagner la transformation des entreprises pour faire de la data un levier de performance et de création de valeur pérenne. « Le big data ne relève pas tant d'enjeux techniques que d'un défi de transformation des organisations et de leur modèle économique. L'exploitation big data relève de la capacité à convertir la data en connaissances, en innovation et en valeur », indique-t-on chez EY.

Dans cette optique, le cabinet propose une offre big data pluridisciplinaire, autour de 4 grands domaines : Marketing/ Ventes - Business Operations - Finance et Risque - Stratégie, Gouvernance et Architecture Data. Elle est portée par un large éventail de compétences, allant des data scientists, aux consultants en stratégie, en passant par des experts en sécurité informatique et des avocats spécialisés en droit des nouvelles technologies et en protection des données à caractère personnel (privacy). Cette pluridisciplinarité permet de proposer aux entreprises une approche transversale de leurs enjeux et ainsi d'activer des leviers d'actions concrets pour une exploitation optimale de la data.

Pour en savoir plus sur ces leviers et les missions Big data menées par EY : <http://www.ey.com/Fr/BIGDATA>



**EY** Building a better working world

## Acteurs historiques

### Hitachi Content Platform, le stockage sécurisé des données dans le cloud

Fondé en 1991, Hitachi Data Systems (HDS) est une division appartenant à 100 % au groupe japonais Hitachi. Elle fournit des technologies, services et solutions informatiques dans plus d'une centaine de pays. Dans le domaine du big data, elle propose principalement des solutions de stockage et de virtualisation dédiées à l'analyse de gros volumes de données.

Au cœur de son offre : la solution Hitachi Content Platform (HCP), une plateforme de stockage objet dans le cloud. « Elle permet le stockage massif de données et leur analyse par leurs métadonnées, en support du métier et des applicatifs », explique HDS. Fin janvier 2015, une nouvelle version d'HCP a été lancée. HCP S10 offre « de plus grandes capacités de stockage et une protection accrue des données grâce à la technologie Erasure Coding (EC) », indique Hitachi. EC est une méthode de protection des données dans laquelle des données sont divisées en fragments stockés à différents emplacements, tels que des disques, des nœuds de stockage ou des lieux géographiques.

La société nipponne intègre également à son offre big data des solutions d'autres éditeurs grâce à divers partenariats. HDS propose ainsi d'intégrer la solution d'analyse de données : SAP-HANA, du géant allemand SAP, leader européen des ERP. HDS propose également des solutions basées sur les technologies Hadoop, OpenStack, Microsoft, Oracle, Symantec ou VMware.

### Intersystems change la base de données et l'analyse temps réel

L'analyse de données complexes est la spécialité de la société originaire du Massachusetts, depuis sa création en 1978. Avec Caché, Intersystems veut permettre aux entreprises de réaliser des « Breakthroughs in big data ». Avec ce système avancé de gestion de base de données associé à un environnement de développement rapide des applications, Intersystems fournit les conditions de l'innovation aux entreprises, à travers un socle technologique qui se veut à la fois robuste et agile. « Les données ne sont décrites qu'une fois dans un seul dictionnaire de données intégré et sont disponibles immédiatement en utilisant un accès par objet, un SQL haute performance et un puissant accès multidimensionnel - qui peuvent tous accéder simultanément aux mêmes données » explique l'entreprise. Le système peut être renforcé grâce à DeepSee, une technologie d'analyse en temps réel combinée des données structurées et non structurées. Intersystems précise que ses outils sont aisément intégrables dans les applications, DeepSee ne demandant par exemple pas d'entrepôts de données. La société compte parmi ses principaux clients des acteurs du monde de la santé, de la finance, de l'énergie et de l'aéronautique, tous grands consommateurs d'analyses temps réel de grandes quantités de données.

### Keyrus : du traitement des données à la création de valeur pérenne

Selon Keyrus, « appréhender le phénomène big data c'est savoir collecter, décrypter, traiter les Big Data pour activer les leviers de croissance et de performance des entreprises quels que soit leur taille ou secteur économique. C'est aussi construire de nouveaux Business Models pour assurer leur pérennité et leur développement. C'est enfin créer des avantages concurrentiels durables, en bénéficiant de l'analyse fine, et de la capacité d'anticipation, voire de prédiction issues de la corrélation d'informations hétérogènes ».

Keyrus a développé sa propre plateforme big data as a service, et mobilise les compétences nécessaires à la réussite de tout projet big data : la connaissance métiers et Data visualisation, l'expertise analytique et la maîtrise des technologies avancées (Hadoop, Spark...), autour d'une approche pragmatique Proof of Value (POV) ciblée sur des projets concrets et réalistes.

Keyrus se positionne comme le partenaire stratégique pour les projets big data des entreprises, s'appuyant sur des références multi-sectorielles de projets big data mis en œuvre en France et à l'international et des alliances avec les éditeurs majeurs du marché.

Acteur historique du conseil et de l'intégration de solutions de BI, Performance Management et valorisation de données, Keyrus est par ailleurs co-fondateur de la chaire Data Scientist de l'Ecole polytechnique.

[www.keyrus.fr](http://www.keyrus.fr)

**HITACHI**  
Inspire the Next

**INTERSYSTEMS**

**Keyrus**

# Haven, la plateforme big data HP



HP décline sa stratégie big data au travers d'une plateforme qui permet de traiter 100 % des informations à disposition des clients, qu'elles soient issues de l'entreprise (Bases de données, courriels, documents internes) ou externes (réseaux sociaux, pages web, documents en ligne, vidéo, audio ...). Baptisée Haven, il s'agit d'une plateforme intégrée mais modulaire basée sur quatre composants.

Le premier est sans surprise la technologie open source Hadoop, désormais incontournable dans la majorité des projets big data. Hadoop permet le stockage et l'analyse d'importantes quantités de données structurées ou non structurées, à moindre coût. HP est l'un des contributeurs majeurs de cette solution à code ouvert dont le développement est encadré par l'Apache Software Foundation. Pour enrichir Hadoop, HP propose des moteurs et des connecteurs qui permettront aux clients d'accélérer la valorisation des données.

Autonomy permet d'interpréter et indexer les informations humaines, non structurées. Ce logiciel recherche, analyse et enrichit les données issues d'une grande variété de sources : textes, mails, réseaux sociaux, audio, vidéo. Il utilise pour cela la technologie Machine Learning IDOL (Intelligent Data Operating Layer). IDOL est un algorithme statistique qui interprète le contenu indépendamment des langues et fournit plus de 400 connecteurs pour un millier de formats.

Pour les analyses structurées et semi-structurées, HP propose la solution Vertica, « solution analytique temps réel (architecture en colonnes) capable d'ingérer de très grands volumes de données et de traiter des requêtes complexes 50 à 1000 fois plus rapidement qu'une base de données traditionnelle », assure HP. Vertica permet notamment d'appliquer des modèles de statistiques prédictives en s'appuyant sur les bi-

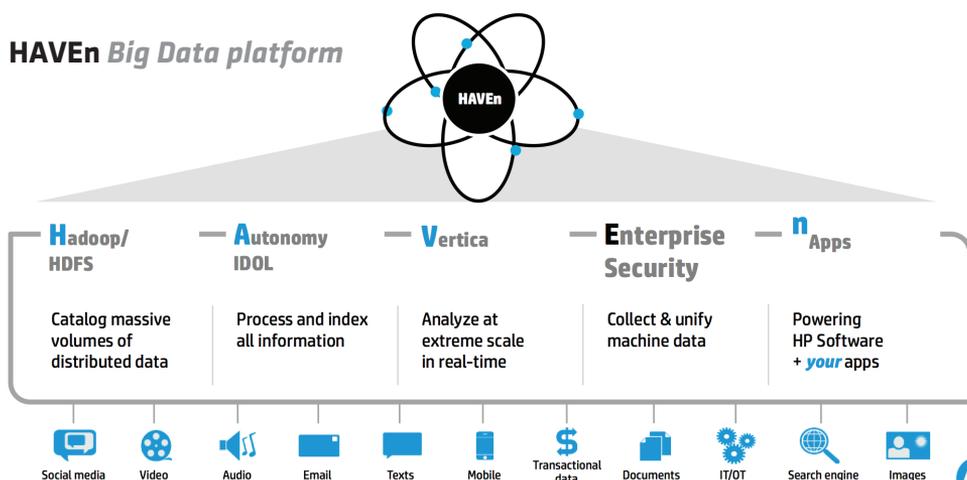
bliothèques R, Python, SQL. Une innovation récente permet de lancer des requêtes SQL directement sur Hadoop. Selon Laurent Ridoux, Chief Technologist big data chez HP France, Vertica présente un optimum en terme de performance, de scalabilité et de coût qui l'ont fait choisir par les acteurs majeurs du web comme Facebook, mais également par des industriels dans les télécommunications, la banque, la santé, la grande distribution, l'énergie, le transport.

Pour traiter les logs machines, HP a également intégré à sa plateforme une brique dédiée à la collecte et à l'analyse temps réel des données machines et des logs d'applications (capteurs, internet des objets, ...), baptisée Enterprise Security (Arcsight Logger). Elle peut traiter 1 million d'évènements par seconde et fournit 350 connecteurs en standard.

HP et son écosystème d'éditeurs et d'intégrateurs développent les n applications métiers qui permettent aux clients de tirer immédiatement parti des solutions construites sur Haven.

La plateforme Haven est disponible soit sur site client, soit en mode hébergé, mais également « On demand ». L'ensemble de cette plateforme est supporté par des infrastructures standards X86, spécifiquement adaptées au big data, qui contribuent à sa compétitivité économique. « Fort de ses 15 ans d'expérience dans le calcul scientifique, HP a innové pour outiller ces architectures massivement parallèles et réduire leur coût d'exploitation » assure Laurent Ridoux.

## HAVEn Big Data platform



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. Intel, le logo Intel, Xeon et Xeon Inside sont des marques de commerce d'Intel Corporation aux Etats-Unis et dans d'autres pays.

« Nous proposons une méthodologie en trois étapes », Laurent Ridoux, Chief Technologist big data chez HP France.

### Quelle est votre principale clientèle en France autour de la solution Haven ?

Laurent Ridoux : Notre solution s'adresse à tous les secteurs d'activité et à tout type d'entreprise. Il y a encore aujourd'hui une maturité hétérogène des entreprises en matière de big data. Historiquement, les précurseurs ont été les télécoms, la banque, les assurances et bien entendu les acteurs du web. Pour ces derniers, le big data est leur principal outil industriel. C'est le cas en France avec 1000Mercis ou BlaBlaCar, mais aussi dans le monde avec Facebook ou Zynga. Zynga est un éditeur de jeux en ligne qui utilise 5 Peta de Vertica en temps réel. Il n'est pas nécessaire d'être un géant pour faire du big data.

En France, les précurseurs hors web sont des entreprises qui ont fait face à une rupture de leur modèle de vente ou qui l'anticipent. L'arrivée de Free a accéléré la volatilité des clients. Pour faire face, les opérateurs ont dû adapter leurs offres, mais aussi les services associés. L'amélioration des services, la compréhension du client, la lutte contre leur attrition sont passées par de l'analyse big data. C'est la même chose du côté des banques et des assurances qui font face à de nouveaux comportements clients et voient arriver la concurrence des GAFA (Google Amazon Facebook Apple) qui se matérialise en Amérique du Nord. Il est donc crucial pour ces acteurs de garder le leadership sur la valeur ajoutée offerte au client.

« Il n'est pas nécessaire d'être un géant pour faire du big data. »

**Laurent Ridoux**  
Chief Technologist Big Data  
chez HP France



### Parmi vos clients figurent le leader européen du covoiturage BlaBlaCar et le groupe 1000mercis, pionnier de la publicité et du marketing interactifs. Quels usages font-ils du big data ?

Sans rentrer dans les détails, 1000mercis fait du ciblage marketing sur les canaux digitaux de manière individualisée pour proposer le bon message, à la bonne personne, au bon moment, sur le bon canal. Cette prestation se déroule avec des campagnes marketing, dites programmatiques, basées sur du RTB (Real Time Bidding). La plateforme analytique repose sur Hadoop. BlaBlaCar, quant à lui, utilise notre base de données Vertica pour faire ses calculs d'optimisation de covoiturage.

### Quelle méthodologie vous permet de couvrir des projets big data allant des plus simples aux plus complexes ?

Nous avons élaboré une démarche en trois étapes. L'objet de la première étape est de faire émerger et identifier les cas d'usages qui pré-

sentent la plus forte valeur pour le client. Nous regardons les données dont il dispose, celles qu'il peut collecter en complément et les projets déjà identifiés par les métiers. Nous complétons cette vision avec les cas d'usages pertinents qui ont déjà été déployés. C'est une phase de cadrage et d'identification du coût et des bénéfices du projet.

La deuxième étape consiste à valider, dans le cadre d'un « Proof of Concept », la faisabilité technique et les bénéfices des cas retenus. Nous utilisons notre Centre d'innovation big data de Grenoble, en partenariat avec Intel, pour modéliser les analyses attendues.

Si ce Proof of Concept est concluant, le client peut passer alors en mode projet pour mettre en œuvre sa solution.

Cette méthodologie s'adapte très facilement à la maturité du client. Elle permet de maîtriser l'avancement et l'investissement engagé par le client. Chaque étape est calibrée en fonction d'un scope métier et fait l'objet d'un Go/No Go pour la phase suivante.

## Acteurs historiques

### Micropole, des projets lancés en 24 heures

Fondée en 1987, Micropole est un groupe international de conseil en technologies innovantes.

Sa solution Big Data 24 permet, en seulement 24 heures, de se « lancer dans l'expérimentation concrète de projets big data, en s'affranchissant des contraintes technologiques », explique le groupe. Concrètement, Big Data 24 est une plateforme big data complète proposée en mode SaaS. Le client n'a donc pas à déployer d'outils big data sur son propre système d'information. Et pour un plus grand « respect des données sensibles », la solution est hébergée sur le cloud public souverain Cloudwatt, « avec lequel nous avons passé un accord de partenariat », souligne Micropole. Côté fonctionnalités, Big Data 24 intègre les outils nécessaires pour lancer un projet big data, allant de la gestion et collecte des données, à leur analyse et visualisation. « Ces outils peuvent être adaptés sur-mesure selon les besoins et les solutions utilisés par les clients », précise le groupe.

Enfin, Big Data 24 est une offre en mode locatif, payée au mois et sans engagement sur la durée. Le client peut donc l'utiliser uniquement le temps nécessaire et ainsi maîtriser son budget. « Cette solution s'adresse principalement aux directions marketing et IT qui souhaitent bénéficier facilement, rapidement et sans lourd investissement des opportunités offertes par le big data », conclut-on chez Micropole.

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

**MICROPOL**<sup>®</sup>  
Driving Distinction

### MicroStrategy, le big data en mobilité

Éditeur américain de logiciels de Business Intelligence, MicroStrategy propose depuis 1989 des solutions d'analyse de données et d'aide à la décision. Sa plateforme MicroStrategy Analytics est utilisée par plus de 4000 grandes entreprises et institutions, dans une trentaine de pays. Elle permet d'analyser de vastes volumes de données et de distribuer des informations opérationnelles via des rapports, des tableaux de bord et des analyses avancées ad hoc. « Notre plateforme se connecte à toutes les sources de données de l'entreprise et dispose de connecteurs natifs pour les principales bases de données, dont celles spécifiques aux big data », explique l'éditeur.

Son offre se distingue notamment par la solution MicroStrategy Mobile qui complète sa plateforme de base. Cette option permet aux entreprises d'accéder aux analyses et rapports, par exemple sur les performances de la société, depuis un smartphone ou une tablette. MicroStrategy Mobile se présente comme un éditeur d'applications mobile que va utiliser l'entreprise pour créer sa propre solution pour les plateformes Android ou iOS. « Il n'est pas nécessaire de disposer de compétences en langage de programmation », souligne l'éditeur. Concrètement, MicroStrategy Mobile fonctionne selon le principe du WY-SIWYG (What you see is what you get). Il suffit donc de glisser-déposer des éléments de l'interface pour créer de visu l'application mobile.

**MicroStrategy**<sup>®</sup>

### Pentaho, une plateforme open source

Éditeur français de logiciels, Pentaho a fêté ses dix ans en 2014. Il propose une plateforme big data intégrant deux modules : Pentaho Business Analytics, dédiée à l'analyse des données, et Pentaho Data Integration, qui sert à collecter et croiser les données. Disponible en version 5.3 depuis février 2015, la plateforme Pentaho se distingue tout d'abord par sa version communautaire gratuite. L'éditeur exploite en effet des briques en open source pour sa solution big data, et reverse également ses contributions à la communauté. Selon Pentaho, ce mode de développement à code ouvert est un gage de qualité et d'efficacité. Et surtout, la version communautaire peut être téléchargée et essayée gratuitement par les clients. Si elle apporte satisfaction, il est alors possible d'opter pour la version commerciale, qui intègre des services de conseil, d'intégration et de support.

Côté fonctionnalités, un des principaux atouts de la plateforme Pentaho est sa capacité à croiser les données à la source. La solution peut ainsi travailler plus rapidement sur les informations et sans rapatrier de trop gros volumes de données. « Pentaho permet aux analystes de créer des croisements nettoyés et structurés directement à partir des sources de big data avec la facilité d'utilisation et l'accès en temps réel exigés dans les environnements analytiques agiles d'aujourd'hui », explique l'éditeur.

 **pentaho**<sup>®</sup>

## Acteurs historiques

### SAS, l'analytique s'adapte au big data

SAS est un des principaux éditeurs mondiaux de solutions de business analytics. Créé en 1976 aux États-Unis, SAS a déployé des logiciels décisionnels sur plus de 75 000 sites dans le monde, dont 2600 en France. Parmi son offre big data : la solution SAS Visual Statistics. « Cette solution permet de créer et modifier des modèles prédictifs basés sur d'importants volumes de données, en utilisant des méthodes analytiques de régression et d'estimation, de classification et de regroupement », poursuit l'éditeur. Les résultats de ces analyses sont présentés sous forme graphique aux utilisateurs, avec une interface interactive à base de glisser-déposer. Autre solution : SAS In-Memory Statistics for Hadoop. Il s'agit d'un environnement de programmation analytique pour le célèbre framework open source. La solution de SAS permet à un grand nombre d'utilisateurs simultanés de gérer, explorer, analyser et modéliser des données sur Hadoop. « Dès que les données Hadoop sont chargées en mémoire, SAS In-Memory Statistics for Hadoop permet à plusieurs utilisateurs d'effectuer simultanément et en continu de multiples analyses au sein d'une même session d'analyse », explique Oliver Schabenberger, Senior Director, Analytic Server R & D chez SAS. « Cette approche en mémoire limite le brassage de données induit par les entrées-sorties sur disque devenues très contraignantes lorsqu'on analyse des big data. »

### Simpana CommVault, une plateforme unifiée de sauvegarde de données

CommVault est une société américaine de logiciels de sauvegarde de données. Fondée en 1996, elle s'est fait connaître avec son produit phare: Simpana. Son principal atout est d'être une plateforme totalement unifiée. Tous les éléments de la solution partagent en effet la même base de code et sont accessibles via une interface unique. « La plateforme Simpana comporte des modules sous licence individuelle pour analyser, répliquer, protéger, archiver et rechercher les données. Puisque ces modules partagent un jeu commun de services principaux, ils communiquent sans effort entre eux par l'intermédiaire de la plateforme pour résoudre une multitude de problèmes liés au stockage ainsi qu'à l'accès aux données et informations », explique-t-on chez CommVault.

Sa version 10, dernière en date, se veut résolument orientée big data en intégrant une indexation « intelligente » des données sauvegardées. Toutes les informations sont ainsi envoyées vers un magasin de contenu (ContentStore) où leur sont associés plusieurs niveaux de métadonnées facilitant leur analyse. Et ce ContentStore possède des API permettant à des logiciels tiers d'accéder à son index, notamment des outils big data. Enfin, toutes les données stockées dans ce magasin sont chiffrées, dupliquées et leur accès filtré selon la politique de sécurité de l'entreprise, précise-t-on chez CommVault.

### WISEO, pour une approche « Test'n Learn »

Fondée en 1999, WISEO se veut une alternative aux intégrateurs, agences et cabinets de conseil. Avec une expertise dans les domaines du digital, des technologies, des process et de la data, WISEO a renforcé sa position sur le big data en novembre 2014 par le rachat d'Isthma.

« Ce rapprochement illustre notre volonté de nous renforcer sur le marché de l'analyse de données. WISEO dispose désormais de toutes les compétences pour accompagner ses clients dans leur stratégie big data », déclarait à l'occasion de cette acquisition Olivier Dhonte, PDG de WISEO. La société parisienne possède aujourd'hui des compétences dans les trois principaux domaines du big data : le conseil, l'analytique (équipe de data scientists) et la business intelligence.

Dès la phase amont d'un projet big data, elle propose une approche « Test'n Learn ». Le principe est de mettre en place des projets sous-jacents au projet principal dont on va mesurer régulièrement l'efficacité. A chaque retour d'expérience, ces chantiers sont affinés et les éventuels points faibles corrigés. S'il se révèle rapidement concluant, le chantier se poursuit en phase supérieure. Dans le cas contraire, il est abandonné. Chez le client, la méthodologie de WISEO s'organise autour d'ateliers animés par des data scientists permettant de valider au plus tôt, sur des échantillons représentatifs, les modèles de traitement et d'analyse des données.

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



## Acteurs historiques

### OpenText, le big data pour améliorer la chaîne logistique

Depuis 1991, cet éditeur américain propose des solutions de gestion de l'information d'entreprise (EIM). L'objectif est de mieux organiser les flux d'informations transitant sur le SI de l'entreprise, du partage de fichiers entre collaborateurs à l'échange de documents avec les clients et fournisseurs.

Au catalogue de cet éditeur américain, deux solutions peuvent appuyer un projet big data. La première est Active Orders. Elle évalue les performances de la chaîne logistique. Sa fonction « *Gestion de la performance des fournisseurs* » permet par exemple d'évaluer les fournisseurs sur des paramètres tels que les taux d'exécution des commandes, la ponctualité de livraison ou l'exactitude des factures.

Seconde solution : Active Documents. Elle va analyser et indexer l'ensemble des documents commerciaux échangés entre l'entreprise et ses clients/fournisseurs. « *Active Documents fournit une visibilité pour de nombreux types de document, notamment les formats EDI, XML et propriétaires* », souligne OpenText. L'analyse de ces documents met en exergue les éventuelles erreurs de communication ou simplement les accusés de réception manquants ou en retard. Ces deux technologies ont été initialement développées par la société GXS, spécialiste des solutions d'intégration du commerce électronique interentreprises, rachetée en 2013 par OpenText.

OPENTEXT

## Nouveaux entrants

### IKO System, la prospection commerciale boostée par le big data

Depuis 2010, IKO System édite une solution de « Sales Intelligence » en mode SaaS. Le principe est d'utiliser les outils du big data pour collecter de gros volumes de données issues de nombreuses sources, notamment les réseaux sociaux, pour doper la prospection commerciale des entreprises. « *Notre technologie identifie en temps réel des signaux d'achats au travers de millions d'informations déstructurées provenant de sources publiques et privées (web, médias sociaux, actualités, interviews, études) puis score une liste de prospects à contacter en priorité permettant ainsi de multiplier par 8 le résultat des campagnes de prospection* », assure-t-on chez IKO System. Au total, les algorithmes d'IKO System surveillent plus de 60 000 sources d'informations internationales. Les informations internes de l'entreprise sont également prises en compte, dont les données structurées CRM. Pour analyser ces données structurées ou non-structurées, IKO System exploite des technologies sémantiques et des algorithmes d'analyse prédictive. Pour présenter les résultats, la solution attribue des scores de probabilité d'achat sur des prospects, pour prioriser les actions de prospection commerciale. Cette solution est aujourd'hui utilisée par les équipes marketing et commerciales de plus de 150 sociétés dans 25 pays européens dont Oracle, Adobe, HP, La Poste ou encore Talend.

IKO

### OpenDataSoft traite tous les formats et toutes les sources de données

Elue start-up « big data » la plus prometteuse par le CeBit fin 2013, OpenDataSoft s'est fait une spécialité de la mise en musique de données issues de nombreuses sources différentes. Depuis sa création en 2012, l'entreprise annonce réinventer le data-management à l'ère des API et des apps. La plateforme « *clef en main* » qu'elle propose permet aux entreprises d'utiliser des données de n'importe quels formats et sources, le tout quelle que soit leur volumétrie. Des données métiers jusqu'aux informations externes issues d'initiatives open-data ou de *crowdsourcing*, l'outil permettra de capter, de réutiliser et surtout d'industrialiser les services innovants que l'entreprise cliente voudra inventer. Grâce au Cloud, OpenDataSoft veut rendre beaucoup plus accessible la data visualisation et l'analyse prédictive. Jean-Marc Lazard, le PDG, est un ancien directeur de l'innovation d'Exalead, et David Thoumas, autre cofondateur et directeur technique, a lui passé 10 ans à la direction du développement des logiciels de la société du Groupe Dassault Systèmes. La SNCF, la RATP, l'Express, Ouest-France ou encore Veolia utilisent déjà la plateforme pour leurs projets.

OpenDataSoft

## Nouveaux entrants

### Dataiku démocratise le big data

Après à peine 2 ans d'existence, la start-up Dataiku est déjà une référence française dans l'univers du big data. En proposant son Data Science Studio (DSS), elle veut faciliter et accélérer les projets big data des entreprises. « *L'outil est un enabler, un couteau suisse technologique. Il permet de se connecter rapidement à tout l'écosystème technologique du big data qui s'est formé autour de Hadoop, NoSQL, etc.* » résume le CEO de la jeune pousse, Florian Douetteau, un ancien d'Exalead. Sur son site, avec le slogan « Machine Learning on Dirty Data », la start-up affiche déjà fièrement des clients comme vente-privee.com, Pages Jaunes, BlaBlaCar, Parkeon ou Capterio. Dataiku entend démocratiser l'analyse de données : le groupe type d'utilisateurs de DSS mélange ainsi des profils business, des analystes métiers, et des data scientists, plaçant les projets big data au cœur de l'entreprise. La dernière version du logiciel, sortie en février, s'inscrit dans cette logique en simplifiant le processus de déploiement et en rendant DSS compatible avec MacOS X, Docker et Amazon Web Services (AWS). Au programme également de cette récente évolution : la génération automatique de cartes géographiques, ainsi que des possibilités étendues de data visualisation et d'export des données.

### Outscale, une plateforme Cloud au service du big data

Fondé en 2010 sous l'impulsion de Dassault Systèmes, Outscale est un fournisseur de service Cloud de type Infrastructure-as-a-service (IaaS). Sa solution est destinée à accueillir les outils de collecte et d'analyse du big data, mais aussi des solutions de calcul 3D ou HPC. Les clients peuvent choisir d'utiliser l'infrastructure de Cloud public d'Outscale ou de déployer une plateforme VPC (Virtual Private Cloud) sur leur propre système grâce à la solution TINA OS. Cette dernière est un cloud manager permettant aux clients de : « *piloter et automatiser leurs ressources Cloud grâce aux API et à l'interopérabilité avec les standards du marché (AWS EC2 et S3)* », précise Outscale. Son cloud public est quant à lui construit sur la base des solutions Cisco UCS avec des processeurs Intel et sur des volumes de stockage NetApp. Il dispose de plus de 18 000 vcores, 100 To de RAM, 3 Po de stockage et 10 Gbps de bande passante réseau. Les data centers sont localisés en Europe, aux États-Unis et en Asie. Le client peut choisir la localisation des données, par exemple en France, où l'infrastructure d'Outscale a reçu la certification de sécurité ISO 27001-2013 garantissant la sécurité et la traçabilité des données. « *Cette infrastructure offre aux entreprises les capacités instantanées de calcul, stockage et réseau en France et dans le monde, basées sur une infrastructure Cloud Computing haute performance de Classe Entreprise* », conclut Outscale.

### Quinten, un algorithme unique sur le marché

Quinten est une société de conseil stratégique et opérationnel spécialisée dans la valorisation des données. Créée en 2008, elle a effectué une centaine de missions pour plus de 25 clients récurrents. Historiquement, Quinten s'est lancé dans le domaine complexe qu'est la recherche médicale. Elle a ensuite étendu ses compétences à d'autres secteurs tels que les cosmétiques, la finance, les médias ou l'agroalimentaire. Parmi ces clients figurent : Sanofi, Danone, L'Oréal ou M6 publicité. Quinten emploie 25 data scientists, dédiés à la collecte, au traitement et à l'analyse de donnée dans le cadre de projets big data. Outre ces compétences, la société se démarque par l'utilisation d'une technologie conçue et développée au sein de son département R&D : l'algorithme Q-Finder. « *Cet algorithme est capable d'exploiter tout type de données de façon exhaustive et sans aucun a priori, afin d'identifier les configurations à forte valeur ou à fort risque en fonction de la problématique* », explique Quinten. Exemple : Q-Finder permet d'identifier au sein d'une base de données de points de vente, les combinaisons optimales d'investissements générant la meilleure performance pour chacun de ces points de vente. « *L'entreprise peut ainsi affiner l'allocation de ses investissements en les adaptant à ses contraintes budgétaires et à sa stratégie de développement* », indique-t-on chez Quinten.

[www.quinten-france.com](http://www.quinten-france.com)



# Notre sélection des 50 qui font le *big data*

	Infra-structure	Data management	Analytics	Décision	Conseil	Autres
<b>Action</b>			●			
<b>Business &amp; Decision</b>	●	●	●	●	●	
<b>Capgemini</b>					●	
<b>Cloudera Inc</b>		●	●	●		
<b>CommVault</b>		●				
<b>Conexance</b>		●	●			
<b>Criteo</b>			●			
<b>Dassault Systèmes</b>		●	●	●		
<b>Data Publica</b>					●	
<b>Data&amp;Data</b>			●			
<b>Dataiku</b>	●	●	●	●	●	
<b>datascience.net</b>						●
<b>Deloitte</b>		●	●	●	●	
<b>DreamQuark</b>	●	●	●	●	●	●
<b>EY (ancien Ernst &amp; Young)</b>		●	●	●	●	
<b>HP France</b>	●	●	●	●	●	●
<b>Hitachi Data Systems (HDS)</b>	●	●	●			
<b>Hortonworks</b>	●	●	●		●	
<b>IBM</b>	●	●	●	●	●	
<b>IKO System</b>				●	●	
<b>Informatica</b>		●				
<b>Infotel Conseil</b>		●	●	●	●	●
<b>InterSystems</b>	●	●	●	●	●	
<b>ITS Group</b>	●		●		●	

	Infra-structure	Data management	Analytics	Décision	Conseil	Autres
<b>Keyrus</b>	●	●	●	●	●	
<b>Linkurious</b>			●			
<b>Makazi</b>		●	●	●		
<b>Micropole</b>			●	●	●	
<b>Microsoft</b>	●	●	●	●	●	
<b>MicroStrategy</b>			●			
<b>OpenDataSoft</b>		●	●			
<b>OpenText</b>		●	●	●		
<b>Oracle</b>	●	●	●	●	●	
<b>Outscale</b>	●					
<b>Pentaho</b>		●	●	●	●	
<b>PredicSis</b>			●			
<b>Proxem</b>			●	●		
<b>Quantmetry</b>			●		●	
<b>Quinten</b>		●	●	●	●	
<b>Rocket Fuel</b>			●			
<b>salesforce</b>			●			
<b>SAS</b>		●	●	●		●
<b>Sollan</b>		●	●	●	●	
<b>Squid Solutions</b>	●	●	●			
<b>Talend</b>		●	●	●		
<b>Tapvalue</b>			●	●		
<b>TellMePlus</b>			●	●		
<b>Thales</b>	●	●	●	●	●	
<b>WISEO</b>		●	●	●	●	
<b>Ysance</b>		●	●	●	●	

Et pour aller plus loin :



Téléchargez la version complète du Guide du big data sur [ALLIANCY.FR/guide-du-big-data](http://ALLIANCY.FR/guide-du-big-data)

