

2017 / 2018

# GUIDE DU BIG DATA

L'ANNUAIRE DE RÉFÉRENCE  
À DESTINATION DES UTILISATEURS

BIG DATA

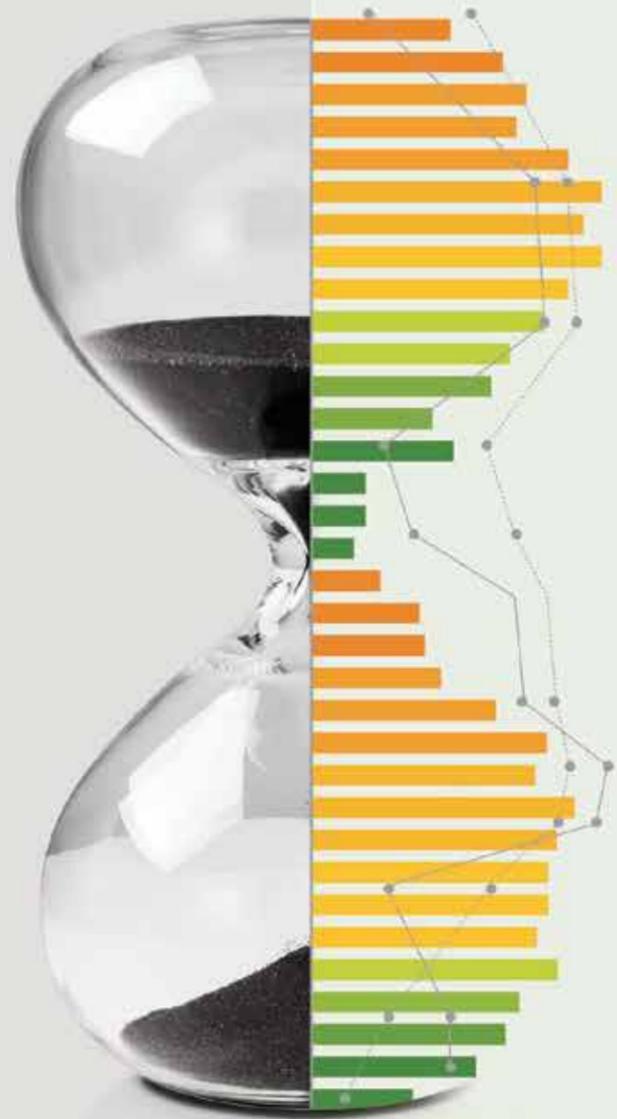


Nous pensons que la  
Donnée peut rendre ce qui  
est impossible aujourd'hui,  
possible demain.

**cloudera**<sup>®</sup>

---

Powering possibility



**Avec eux, le Big Data est une contrainte**

**Avec nous, le Big Data est un atout**

Les solutions traditionnelles vous obligent à utiliser des rapports pré-définis et à renouveler vos requêtes dès que vous changez l'angle de votre analyse. Qlik est la seule plateforme à vous permettre d'explorer vos données en toute liberté – qu'il s'agisse de 40 ou 40 milliards de lignes – et d'obtenir des réponses instantanément.

**C'est ce qui fait l'Associative Difference.™**

**Venez découvrir Qlik au Stand n°A3.**

Selon l'IDC, le marché du Big Data devrait atteindre **203 milliards de dollars de chiffre d'affaires en 2020, contre 130,1 milliards de dollars en 2016.**

Le marché américain représentera plus de **95 milliards de dollars de revenus et se placera devant le marché européen.** Ainsi, le Guide vous propose cette année un éclairage exclusif sur le développement du marché français comparé au marché américain sur plusieurs secteurs en croissance.

Les **entreprises de plus de 500 employés** seront les acteurs de premier plan de la croissance du secteur du Big Data et de l'Analytics, avec un revenu généré estimé à **154 milliards de dollars en 2020.** Cependant les **TPE et PME** ne seront pas en reste, on estime qu'elles produiront **plus d'un quart des revenus du Big Data dans le Monde.**

En termes de secteurs d'activité, **les banques & la finance, les télécoms, le retail, la santé et l'industrie** tireront toujours largement les investissements dans le Big Data et l'Analytics. Qu'il s'agisse d'informations géographiques, de données météorologiques, de données de recherche, de données sur le transport, de données sur la consommation d'énergie ou de données sur la santé, la nécessité de donner du sens aux «Big Data» mène à des innovations technologiques et au développement de nouveaux outils et nouvelles compétences dans des secteurs déjà avancés dans leurs investissements en Big Data. **Vous retrouverez donc des témoignages et cas pratiques dédiés à ces secteurs d'activité plus matures dans la partie Fournisseurs de ce Guide.**

Cependant, pour vous faire découvrir toujours plus **d'applications du Big Data et de l'Analytics,** nous nous sommes, **dans la partie éditoriale de ce Guide,** tournés vers **des secteurs d'activité** moins matures mais non moins innovants, qui présentent **un fort potentiel d'investissement** et de croissance. Ainsi, au travers d'interviews et notes de la rédaction dédiées au Big Data et à l'Analytics dans les secteurs de **du Droit, de l'Immobilier, de l'Agriculture, de l'Éducation et de l'eSport,** nous souhaitons vous apporter des clés sur leurs activités en cours de disruption et de l'inspiration quant à leurs nouveaux business model et manière de gérer les données.

Vous retrouverez également, **dans cette 5e édition, un état des lieux des avancées en matière d'éthique et d'ouverture des données en France,** dressé par des acteurs emblématiques qui contribuent chaque jour à structurer et dynamiser le marché du Big Data et de l'Open Data.

Bonne lecture !



**Marion PEAUDECERF**  
Directrice de publication

+33 (0)1 84 83 03 03 / mpeaudecerf@corp-agency.com

# Les tribunes du Big Data

## EDITO 03

★ DU BIG DATA  
À L'ANALYTICS AUX  
SCIENCES COGNITIVES

## 06

★ ETHIQUE DES ALGORITHMES  
Nozha BOUJEMAA  
INRIA 12

★ LEARNING ANALYTICS  
Ivan OSTROWICZ  
DOMOSCIO 60

★ ESPORT  
Stéphan EUTHINE  
FRANCE ESPORT  
LDLC EVENT  
48



★ IMMOBILIER  
Edouard LE GOFF  
KINAXIA 42



★ JUSTICE PREDICTIVE  
Louis LARRET-CHAHINE  
PREDICTICE 30

★ IMMOBILIER  
Thomas LEFEBVRE  
MEILLEURSAGENTS.COM 36



★ AGTECH  
Stéphane MARCEL  
SMAG 54

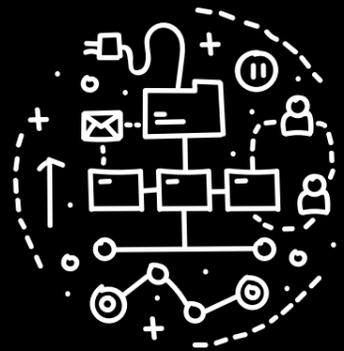
★ PÉPITES BIG DATA  
Xavier LORPHELIN  
SERENA CAPITAL 18

★ OPEN LAW  
Benjamin JEAN  
INNO<sup>3</sup>  
Association «Open Law,  
le Droit Ouvert» 24

42 CONSULTING	70
DIGITAL & DATA	70
AB INITIO SOFTWARE	72
ACCENTURE ANALYTICS	74
ADELEAN	76
ADVANCED SCHEMA	78
NODATA / GROUPE	79
ALTARES	80
AMAZON WEB SERVICES	82
ATTUNITY	84
BLUE DME	86
BLUESCALE	88
BMB SERVICES	90
BRENNUS ANALYTICS	92
CB EXECUTIVE	94
EXECUTIVE MASTÈRE DATA STRATEGY	95
CENTRALESUPÉLEC EXED	96
CISCO SYSTEMS	98
CLARANET	100
CLOUDERA	102
CLUB DECISION DSI	104
COHERIS	106
CONFLUENT INC	108
COUCHBASE	110
DATAIKU	112
DATASTAX	114
DENODO	116
DIGDASH	118
EBIZNEXT	120
EKIMETRICS	122
ELASTIC	124
SUNHOTELS	125
EULIDIA	126
EXPERIS	128
EXPERT SYSTEM	130
FINAXYS	132
GROUPE ESTIA	134
HORTONWORKS	136
HPE	138
HURENCE	140
HYPERCUBE	142
IDEATRANS	144
INDEXIMA	146
INVIOO	148
JEMS DATAFACTORY	150
KEYRUS	152
LANSROD BIG DATA	154
MAPR TECHNOLOGIES SAS	156
MATHWORKS	158
MICROSTRATEGY	160
MONGO DB	162
NE04J	164
ORACLE	166
ORANGE BUSINESS	168
SERVICES	168
OUTSCALE	170
OVH	172
PENTAHO	174
HITACHI VANTARA	175
PITNEY BOWES	176
SOFTWARE	176
PROVALIS RESEARCH	178
QLIK	180
SAAGIE	182
SAP	184
SAS INSTITUTE S.A.S	186
SCINETIK	188
SINEQUA	190
SKAPÁNĚ	192
SNOWFLAKE	194
SOFT COMPUTING	196
SOPRA STERIA	198
SPLUNK	200
SQLI	202
SYNALTIC GROUP	204
TABLEAU	206
TALEND	208
M2OCITY	209
TÉLÉCOM PARISTECH	210
TÉLÉCOM EVOLUTION	210
TERADATA	212
TIBCO SOFTWARE	214
TIMI	216
TOM SAWYER SOFTWARE	218
TRIFACTA	220
XEBIA	222
YSANCE	224
ZENIKA	226

# Sommaire

## Annuaire des Entreprises 68



# Du **Big Data** à l'**analytics** aux **sciences** **cognitives**

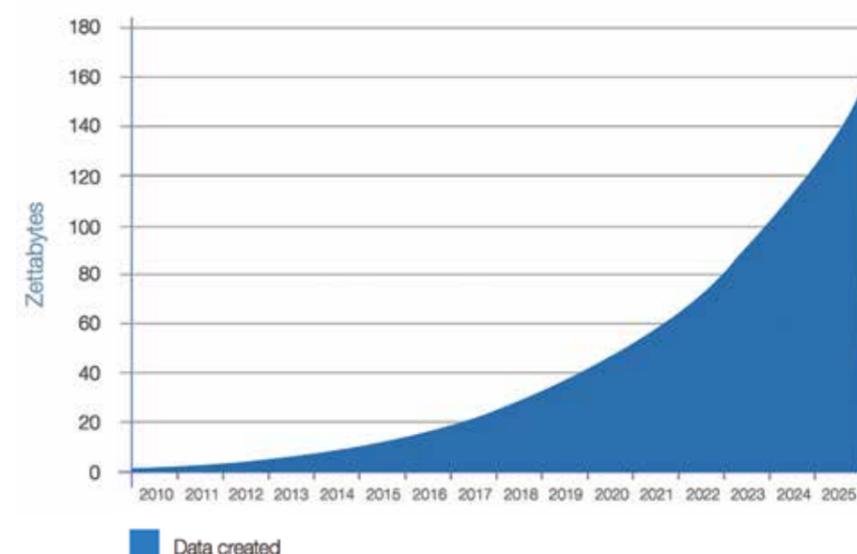


## Une **nouvelle** **ère** de la data

★ Nous vivons dans un Monde dominé par la data et les algorithmes. Selon une étude réalisée par la CNIL et l'IFOP, **80% des citoyens français sont conscients de la présence des algorithmes dans notre société**. Une société data-driven dans laquelle l'information est devenue une véritable devise monétaire. Les consommateurs utilisent quotidiennement des services gratuits des GAFAM en échange des données laissées sur leur passage et de l'autorisation donnée à ces géants du web de monétiser leur comportement sur Internet. Gartner prédit d'ailleurs que **d'ici 2021, 20% de toutes les activités individuelles impliqueront au moins un des sept géants de l'IT**, parmi lesquelles Google, Apple, Facebook, Amazon pour les États-Unis et l'Europe et Baidu, Alibaba et Tencent pour la Chine, générant toujours plus de données.

Plus globalement, l'IDC prédit que les **données produites en 2025 atteindront 163 Zettaoctets** (1 Zo équivalent 1 180 milliards de milliards d'octets), **soit 10 fois plus que les 16,1 Zo de données générées en 2016**. Face à cette augmentation considérable du volume de données, la question primordiale qui se pose n'est plus celle de la génération et collecte des données, mais bien celle de la qualité, de la **valorisation des données et des perspectives business qui peuvent en être retirées**.

Annual Size of the  
Global Datasphere



Source : IDC's Data Age 2025 study, sponsored by Seagate, April 2017

D'autant que l'essentiel des **datas générées, (plus de 60% selon IDC) le seront par les entreprises**, et proviendront de sources multiples telles que les systèmes embarqués et l'IoT, les systèmes cognitifs et d'Intelligence Artificielle, le mobile et les applications en temps réel etc...

Indépendamment de l'origine des données, les entreprises seront amenées à gérer plus de 97% de la « Datasphere » dans son ensemble.

Le mot Big Data n'est d'ailleurs plus totalement d'actualité, tant le volume est une problématique révolue. Gartner, lors de l'une de ses conférences aux États-Unis, a proclamé la mort de la Business Intelligence et l'avènement de la Data Analytics. Alors que la Business Intelligence consiste à examiner des sets de données collectées afin de tirer des conclusions sur les informations qu'ils contiennent et de prendre des décisions business plus éclairées, l'Analytics est une véritable data science. Les outils de Data Analytics sont déployés afin de prédire des comportements dans le futur et de construire des modèles prédictifs d'analyse.

**163 Zettaoctets,**

**c'est le volume estimé des**

**données produites en 2025,**

**soit 10 fois plus que les**

**16,1 Zo de données**

**générées en 2016**

## Trop de **data** !

★ Devant cette abondance de data, les opportunités sont nombreuses pour tirer profit du large volume de données et pourtant les entreprises n'ont pas encore les capacités et compétences nécessaires pour donner du sens aux données critiques, soit car la stratégie n'est pas adaptée, soit par manque de ressources ou de technologies. Selon un sondage réalisé par NewVantage Partners auprès des « Fortune 1000 entreprises », elles peinent encore à établir une culture data-driven : **69,4% d'entre elles ont lancé des initiatives en la matière mais seul 27,9% en voit les bénéfices. De même, plus de 45,2% d'entre elles n'ont pas encore commencé à monétiser leurs initiatives Big Data.**

Parmi les obstacles prioritaires rencontrés selon ce même sondage, **le manque d'alignement** entre les différents services au sein de l'organisation et la **difficile adoption et compréhension du Big Data par le middle management** sont cités en premier.

Le cabinet Gartner souligne ici **le rôle du Chief Data Officer**, une fonction encore peu développée en France comparé aux

États-Unis. L'intégration d'un Chief Data Officer (CDO) est pourtant **un indicateur clef qu'une organisation s'inscrit dans une culture data-driven.**

Le rôle du CDO consiste justement à rendre accessibles et compréhensibles les données par l'ensemble de l'entreprise, depuis la logistique jusqu'à la vente, en passant par les RH, le marketing, etc. Son rôle est donc lié à la gouvernance de ces informations, en relation avec les principaux décideurs de l'entreprise. La majorité des cadres interrogés par NewVantage estime d'ailleurs que les CDOs devraient reporter au CEO (53,4%) ou au COO (17,8%) plutôt qu'au CIO (15,6%). La relation entre le CIO et le CDO est souvent complexe, chacun revendiquant le périmètre de l'autre.

De même seulement 15% de l'ensemble des données sera taggué, et pourra donc être soumis à une analyse Machine Learning ou d'Intelligence Artificielle, et seulement 3% du total sera analysé par des systèmes cognitifs.

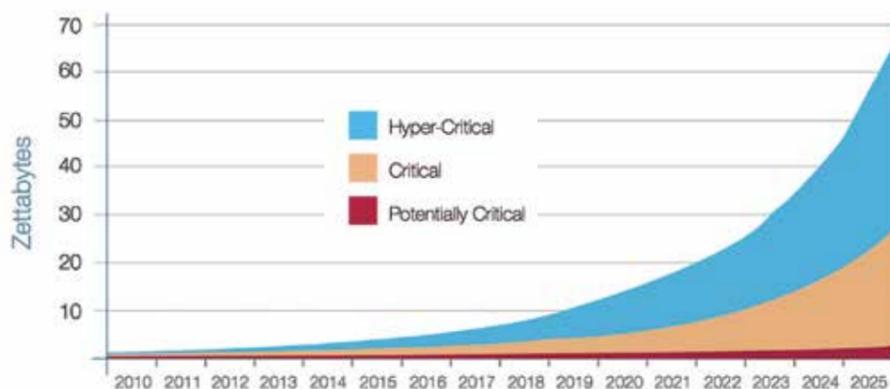
## Des données structurées aux données non structurées

★ Les données structurées sont celles dont l'ensemble des valeurs possibles est déterminé et connu à l'avance. Par exemple, dans une base de données rassemblant les résultats d'une enquête d'opinion, l'âge ou la catégorie socio-professionnelle des individus interrogés sont des données structurées, car les tranches d'âges ou la liste des catégories socio-professionnelles possibles sont déterminées a priori.

## Quelles datas ?

★ Naturellement, toutes les données ne sont pas adaptées ni disponibles à l'analyse. L'IDC estime ainsi que d'ici 2025, seulement **20% de la Datasphere sera essentielle** (« critical ») et seulement **10% sera absolument essentielle** (« hyper-Critical ») dans notre vie quotidienne.

### Data Critical Over Time



Data Type	CAGR 2015 to 2025
All Data. Includes all data in the global datasphere.	30%
Potentially critical. Data that may be necessary for the continued, convenient operation of users' daily lives.	37%
Critical. Data known to be necessary for the expected continuity of users' daily lives.	39%
Hypercritical. Data with direct and immediate impact on the health and well-being of users. (Examples include commercial air travel, medical applications, control systems, and telemetry. This category is heavy in metadata and data from embedded systems.)	54%

Source : IDC's Data Age 2025 study, sponsored by Seagate, April 2017

**90% des données sont soit semi-structurées soit non structurées**

Les données non structurées sont de façon générale des données textuelles qui ne rentrent pas dans des bases de données relationnelles. Elles existent sous de nombreuses formes (textes, images, vidéos, sons...), sont publiées sur de nombreux supports (articles de blog, fils Twitter, posts Facebook, réclamations téléphoniques auprès du service client...) et sont essentielles à la compréhension du business. **Alors que les données structurées décrivent une action ou un fait, les données non structurées donnent les raisons et le pourquoi !**

On estime aujourd'hui que **plus de 90% des données sont soit semi-structurées soit non structurées.** Jusqu'ici la valeur de ces données était largement sous exploitée et la priorité des entreprises était de mettre en place des initiatives sur les données structurées notamment parce que les outils nécessaires pour exploiter le volume des données non structurées n'existaient pas encore. **Les technologies de Machine Learning et de data visualisation rendent cela possible, de même que la baisse des coûts des technologies et la qualité de celles-ci** vont permettre aux entreprises de mettre l'analyse des données non structurées en haut de leurs priorités.

Par leur variété et leur volume, les données non structurées sont essentielles, par exemple pour améliorer la connaissance client. L'analyse linguistique et le text mining permettent désormais de reconnaître les mots et expressions soigneusement choisis pour décrypter les sentiments exprimés par l'internaute, le client ou le prospect.

## Des plateformes analytiques à l'analyse intégrée

★ L'analyse intégrée ou embarquée (embedded analytics) représente l'utilisation



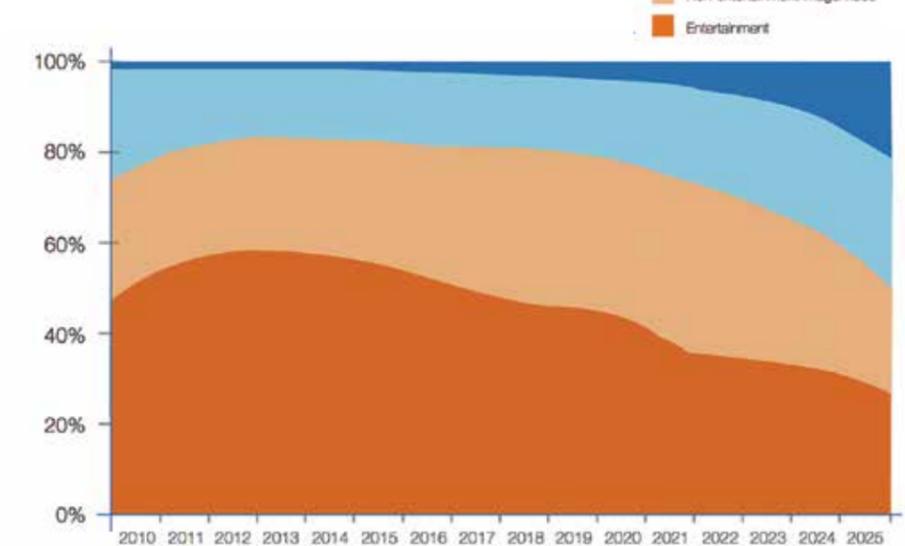
des capacités d'analyse au cœur même des applications logicielles. Les outils analytiques sont des composants natifs de l'application plutôt que des plateformes dédiées et distinctes de l'application. Ces capacités d'analyse peuvent résider en dehors de l'application, néanmoins elles doivent être accessibles de manière aisée et rapide par l'application, sans forcer l'utilisateur à passer par un autre système. Selon une étude Logi Analytics, **plus de 66% des équipes IT utilisent désormais l'analyse embarquée dans leurs organisations, et presque 30% envisagent de l'utiliser dans un futur proche.**

Les données embarquées sont produites par les objets connectés et les technologies Machine-to-Machine tels que les caméras de sécurité, compteurs intelligents, lecteurs RFID, wearables, les automates, les voitures & tous les moyens de transports connectés, les bâtiments connectés, implants médicaux etc. Selon Gartner, **26 milliards d'objets seront connectés dans le monde en 2020 et 6 milliards pourraient bénéficier d'Intelligence Artificielle dès 2018.**

Aujourd'hui, chacun d'entre nous détient en moyenne moins d'un objet connecté embarquant des capacités analytiques. Dans les 10 prochaines années, chaque personne en détiendra plus de 4 (Source : IDC) et les données des systèmes d'analyse embarqués représenteront presque 20% de l'ensemble de la « Datasphere » d'ici 2025.

Cependant, les données IoT ne provoqueront qu'une très légère augmentation de la demande de stockage (moins de 3%), d'ici 2018. La plupart des données IoT ne seront pas conservées.

### Data Creation Share by Type



Source : IDC's Data Age 2025 study, sponsored by Seagate, April 2017

Les données collectées de l'IoT et des systèmes embarqués sont de vrais gisements de valeur, pourtant, selon Forrester, seul un tiers des entreprises qui exploitent des équipements connectés savent utiliser les données générées. Il faut donc **instaurer un dialogue entre data-scientists et responsables métiers afin de faire parler les données collectées.**

L'essor des objets connectés et embarqués impliquera naturellement une explosion du volume de données mobiles et de données en real-time va augmenter une fois et demi plus vite que le volume des données globales.

## Du prédictif au prescriptif !

★ L'analytique pour les entreprises est passée par plusieurs phases, qui ont abouti aux modèles suivants :

- **L'Analytique descriptive** qui aide les utilisateurs à avoir une meilleure compréhension des données passées grâce à du reporting, du groupement et des tableaux de bord.
- **L'Analytique prédictive** qui utilise des techniques statistiques pour étudier les données passées et présentes afin de faire des prévisions.
- **L'Analytique prescriptive** qui permet aux utilisateurs de prendre les bonnes décisions par la simulation et l'optimisation des scénarios.

En 2017, de plus en plus d'entreprises se sont intéressées à l'analyse prescriptive mais seulement 10% des organisations dans le monde l'utilisent, un chiffre qui devrait atteindre 35% d'ici 2020. Gartner prédit par ailleurs que le marché devrait atteindre 1,1 milliards d'euros d'ici 2019.

## Des technologies Big Data aux technologies cognitives et d'Intelligence Artificielle

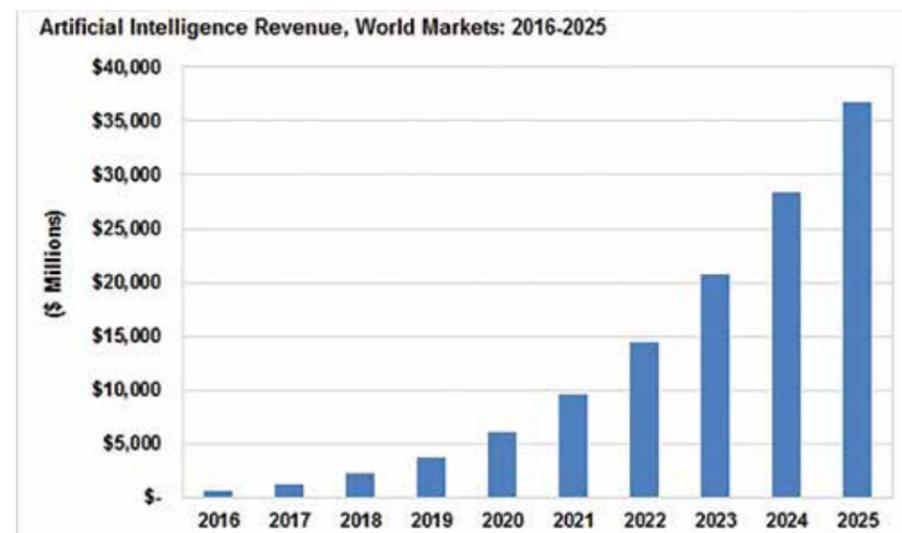
★ Alors qu'aujourd'hui, seulement 4% des organisations disposent d'un système cognitif opérationnel, les ¾ des organisations interrogées ont les data et les capacités analytiques requises pour implémenter des systèmes cognitifs. L'explosion du volume de données disponibles donne implicitement un avantage compétitif aux entreprises ayant intégré l'Intelligence Artificielle. 89% des early adopters des technologies cognitives sont plus rentables et plus innovants que leurs pairs et

concurrents. (Source : IBM)  
Outre les données nécessaires, les applications d'Intelligence Artificielle nécessitent des capacités de stockage en masse et une puissance de traitement que seules les toutes dernières technologies peuvent offrir. L'économie du Cloud jouera un rôle de premier plan dans le développement d'applications basées sur l'Intelligence Artificielle dans l'entreprise. Ainsi d'ici 2020, 67% des dépenses en infrastructures des entreprises IT seront basées sur des offres Cloud et les plateformes Cloud collaboratives utilisées conjointement par les entreprises vont tripler d'ici 2018 (Source : IDC).

Plus globalement, d'ici 2019, 40% des initiatives de transformation digitale et 100% des développements IoT seront supportés par les technologies cognitives et l'Intelligence Artificielle (Source IDC).

Le marché de l'Intelligence Artificielle pour les applications en entreprise est estimé à plus de 11 Milliards de dollars en 2024 et le revenu global est estimé à plus de 36 Milliards de dollars d'ici 2025 selon Tractica.

Une étude Accenture montre que l'Intelligence Artificielle a le potentiel d'augmenter le taux de profitabilité de 38% en moyenne et d'entraîner une relance de l'économie à hauteur de 14 milliards de dollars à travers 16 industries d'ici 2035. Les domaines de l'information et de la communication, de la finance et du manufacturing profiteront prioritairement de l'impact de l'Intelligence Artificielle.



Source : TRACTICA

## De la défense à la sécurité by design

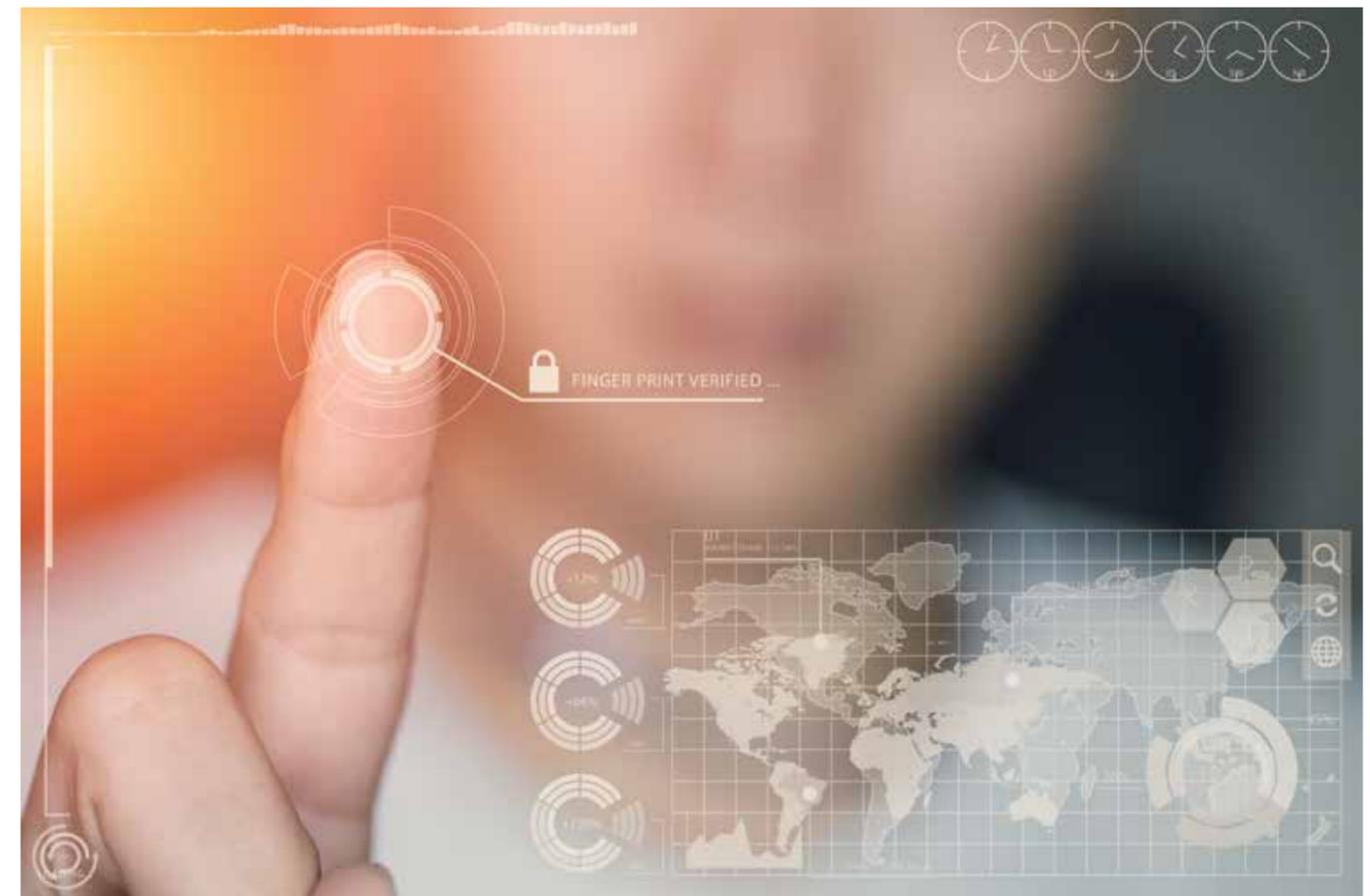
★ Avec de plus en plus de données personnelles à gérer, en real-time et mobiles, les entreprises se doivent d'investir massivement pour garantir la fiabilité et la confidentialité des données et manager les risques de sécurité. La plupart des données requièrent un haut niveau de sécurité, soit presque 90% d'entre elles d'ici 2025, à l'image des informations financières, des informations personnelles et médicales... D'autre part, une prise de conscience est nécessaire sur le fait que les challenges en termes de sécurité et de confidentialité ne sont pas uniquement dédiés aux équipes IT mais requièrent la participation de l'ensemble des salariés d'une entreprise.

Le Règlement européen sur la protection des données personnelles (RGPD) qui entrera en application le 25 mai 2018 vise notamment à renforcer les droits des personnes et responsabiliser les acteurs traitant des données (responsables de traitement et sous-traitants).

Nommer un référent Conformité Informatique & Libertés pour certifier les démarches, le former à ses missions de Data Protection Officer (DPO), former les collaborateurs à la collecte des données, faire des audits pour vérifier la conformité au regard des nouvelles directives, avoir le temps d'établir un plan d'actions, mettre en place des process et des plateformes SI "sécurisés" sont autant de recommandations de la CNIL pour être en conformité avec le règlement.

Néanmoins, la mise en conformité avec le règlement européen ne garantit pas pour autant la fiabilité, la transparence et l'éthique de l'ensemble des démarches Big Data et d'Intelligence Artificielle. Les algorithmes sont également remis en question. Selon une étude de la CNIL, plus de 72% des personnes interrogées en France estime que les algorithmes sont un véritable enjeu de société et plus de 64% considère les algorithmes comme une menace en raison de l'accumulation des données personnelles sur les choix, les goûts et les comportements de chacun. Il revient aux entreprises de s'associer aux institutions publiques pour instaurer la confiance des citoyens et consommateurs et définir des standards éthiques sur le sujet. TransAlgo interviewé ci-après intervient notamment dans ce but.

**La plupart des données requièrent un haut niveau de sécurité, soit presque 90% d'entre elles d'ici 2025**





# ★ Nozha BOUJEMAA

Directrice de recherche  
Conseillère du président de  
l'INRIA sur le Big Data

# INRIA



Nozha Boujemaa dirige l'institut DATAIA (Données, Intelligence Artificielle et Société) récemment accepté par l'état comme Institut Convergence. Elle est coordonnatrice du projet de plate-forme scientifique nationale TransAlgo, dédiée au développement d'outils et de méthodes pour la transparence et la responsabilité des systèmes algorithmiques.

Elle est également conseillère internationale du programme «Advanced Core Technologies for Big Data Integration» pour la JST (Japan Science and Technology Agency) et est par ailleurs membre du conseil d'administration de la BDVA (Big Data Value Association) et General-Chair du European Big Data Value Forum 2017.

# De l'éthique des algorithmes

## Référentiel du **Big Data** et de l'**Intelligence Artificielle**

★ Le Big Data est une chance, une véritable opportunité. Il s'agit d'une avancée technologique significative qui améliore la vie des citoyens, des sociétés, des entreprises et qui génère de la valeur. Les choses évoluent très vite. Il y a deux-trois ans, on a beaucoup parlé de Big Data, parce qu'on s'est rendu compte que les données étaient partout, dans la sphère personnelle et dans le milieu professionnel. On a donc mis le Big Data au premier plan mais **parce que les données sont partout, les algorithmes le sont également pour en assurer leur gestion, leur traitement et leur analyse.** Il y a en ce moment une prise de conscience de l'aspect diffus et pervasive des algorithmes.

Dans l'inconscient collectif, le Big Data représente uniquement un grand volume de données. Pourtant, quand on prend la définition du Big Data avec tous les V possibles, un seul V concerne le volume, or tous les autres, vitesse, vélocité etc... concernent des actions algorithmiques. **Considérer uniquement le volume est extrêmement limitatif.**

Savoir s'il s'agit de Big Data, de Small Data ou encore de Smart Data n'est pas le plus important. Il est en revanche essentiel de prendre en compte la dualité des données et algorithmes. **Les données brutes sans algorithmes sont une masse inerte et les algorithmes sans données sont des processus creux.** Données et algorithmes sont vraiment deux faces de la même pièce.

Le Big Data, c'est tout ce qui peut générer de la valeur à partir des données : c'est l'exploitation des données, la génération de connaissances avec le traitement de l'information, les algorithmes de protection de données mais c'est également comprendre et faire comprendre à travers la dataviz. À partir de ces socles technologiques, il est possible d'atteindre le stade de l'Intelligence Artificielle opérant

comme des services logiciels dans les domaines applicatifs métiers. L'engouement actuel pour l'Intelligence Artificielle, le séparant des technologies Big Data, est surprenant car celles-ci sont naturellement liées à l'apprentissage, la génération de connaissances et de décisions à partir des données. **L'IA, dans son essor actuel, repose principalement sur les technologies qui exploitent les données.** L'IA est orientée vers des services dans des domaines applicatifs particuliers. Les technologies Big Data en tant que telles sont agnostiques au domaine applicatif. Quand on parle d'apprentissage, de moteur de recherche d'information, cela peut avoir plusieurs domaines d'application. La reconnaissance d'objet, la reconnaissance du langage naturel exploitent essentiellement des technologies Big Data.

En revanche, les agents conversationnels utilisés pour des applications particulières dans le domaine de la santé, des services publics ou encore les algorithmes d'identification visuelle par ordinateur pour la détection d'humains, d'obstacles, d'autres véhicules, de signaux routiers mis à profit des véhicules autonomes relèvent de l'Intelligence Artificielle.

Il est important de voir **ce continuum, entre d'une part les technologies socles Big Data et d'autre part l'Intelligence Artificielle qui est orientée vers des domaines applicatifs métiers.** Le Big Data permet l'essor des services applicatifs de l'Intelligence Artificielle.

L'IA dans son cadre général peut dépasser l'aspect data. Mais ce qui fait aujourd'hui l'émergence de cette vague IA actuelle, c'est le Deep Learning, ce sont les technologies d'apprentissage, les capacités avancées qu'offrent les algorithmes d'analyse de contenus selon leurs secteurs respectifs.

## QU'EST-CE QU'UN ALGORITHME ?

Plusieurs communautés essaient d'en donner des définitions, néanmoins nous retiendrons la définition d'Interstices.

Un algorithme est défini comme une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre un problème ou d'obtenir un résultat.

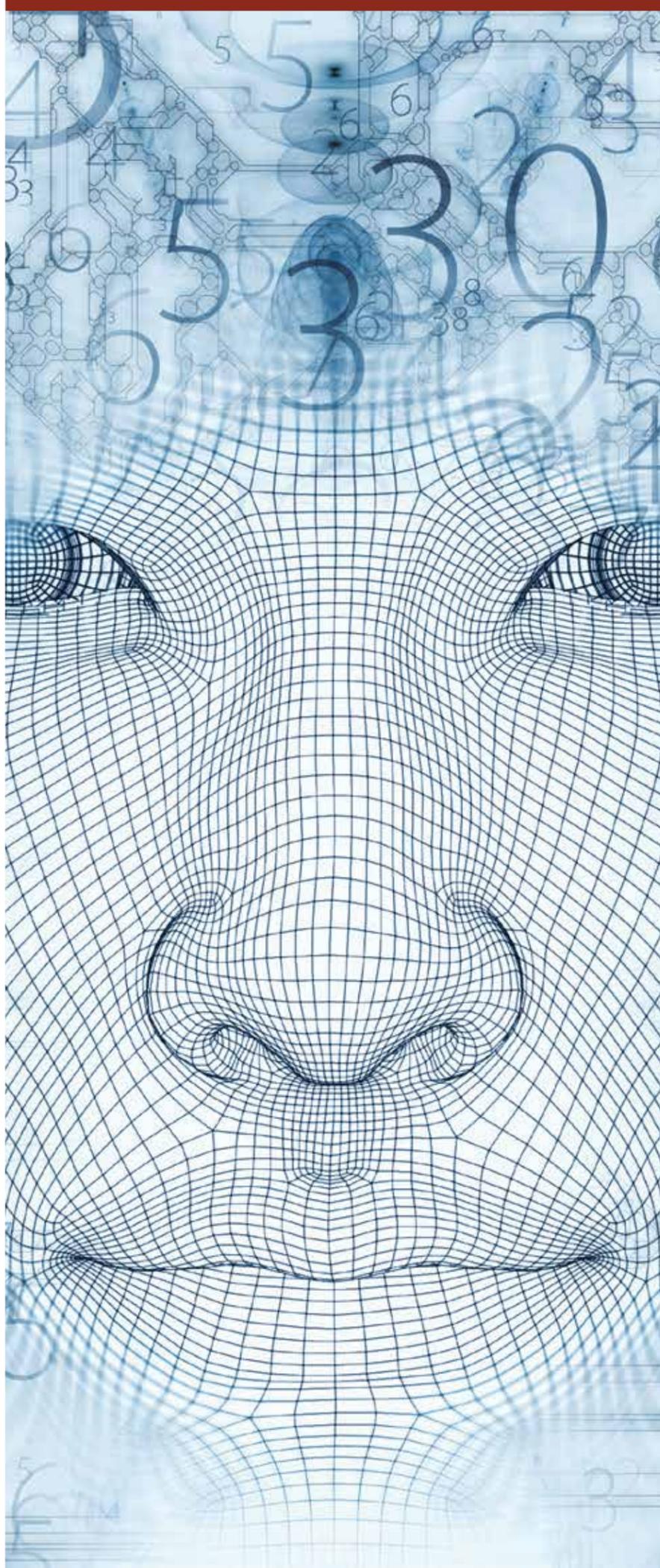
La vertu essentielle des algorithmes est de permettre l'exécution optimisée de procédés répétitifs, essentiellement grâce à la formalisation et à la description des enchaînements logiques à un niveau plus abstrait. L'algorithme est le squelette abstrait du programme informatique, sa substantifique moelle, indépendante du mode de codage particulier qui permettra sa mise en œuvre effective au sein d'un ordinateur ou d'une machine mécanique.

**Les données brutes sans algorithmes sont une masse inerte et les algorithmes sans données sont des processus creux**

## Pourquoi l'éthique des algorithmes est-elle actuellement en question ?

★ La compréhension de l'écosystème de la donnée évolue et on se rend compte que les données sont partout et que de facto les algorithmes sont partout. Plus on déploie les technologies Big Data, plus on est impacté par les algorithmes. La première vague était plutôt optimiste et utopiste vis-à-vis des technologies Big Data. On pensait que le Big Data allait tout permettre (une sorte de baguette magique qui permet de tout résoudre) mais c'est le déploiement à large échelle qui a généré cette prise de conscience qu'il pourrait y avoir des effets de bord pas toujours maîtrisables.

Ce ne sont pas les algorithmes dans l'absolu qui sont à remettre en question, il s'agit d'une véritable évolution technologique. L'analyse des données n'est pas une nouvelle discipline mais elle est en train d'évoluer de la description du passé vers l'analyse prédictive et de plus en plus vers une analyse prescriptive qui elle doit être questionnée en particulier dans les contextes à fort impact sur l'individu et la société.



## Comment l'Europe se positionne-t-elle par rapport aux autres continents en matière de réflexion éthique ?

★ Comprendre le positionnement vis à vis des algorithmes et de l'éthique dépend beaucoup des sphères culturelles diverses dans une géopolitique variable à travers le monde.

### La naissance d'un questionnement éthique aux États-Unis

En France et en Europe, nous sommes toujours dans une posture où l'humain doit rester maître des décisions. Ce n'est pas le cas aux États-Unis où la frontière du prescriptif a été franchie il y a peu. On peut désormais faire une consultation chez le «docteur Watson» sans médecin. Le prescriptif envahit peu à peu la société, dans les services privés comme publics, dans des domaines comme la Justice Prédictive, les décisions d'octroi de prêts ou encore le recrutement.

La première vague de questionnement, d'ailleurs, vient des États-Unis, notamment au travers des ouvrages *The Black Box Society - The Secret Algorithms That Control Money and Information* (Frank Pasquale) et *Weapons of Math Destruction* (Cathy O'Neil, Crown Publishers). La Commission fédérale du commerce a également produit un rapport sur « Les Big Data : une chance pour l'inclusion ou une source d'exclusion » sur plusieurs pans de la société. L'office scientifique et technologique de la Maison Blanche a enfin publié en 2016 un rapport sur les opportunités du Big Data en lien avec les droits civiques.

Le domaine privé s'empare également de ces questions éthiques. Amazon, Apple, DeepMind, Google, Facebook, IBM et Microsoft ont récemment initié un « Partnership on AI to benefit people and society » qui consiste en une plateforme ouverte visant à étudier et formuler les meilleures pratiques sur les technologies de l'IA, faire progresser la compréhension par le public de l'IA et servir de plate-

forme ouverte de discussion et d'engagement sur l'IA et ses influences sur les personnes et la société. Ces acteurs réunis souhaitent développer une politique concertée sur l'éthique, la transparence des algorithmes et des études sur l'impact de l'IA sur la société et les individus. Cette initiative récente, portée par le business, souhaite ainsi lever les verrous et les freins des citoyens sur ces thématiques éthiques. Aucune loi ne les y oblige, c'est plutôt le marché qui les y oblige et les motive à donner des garanties et à offrir de la confiance aux citoyens. Ils sont en effet bien conscients qu'à service égal, celui qui sera le plus transparent et responsable sera celui qui remportera la course.

Quand les gens n'ont plus confiance ou ont des doutes, ils peuvent ne pas adhérer aux offres technologiques même si elles sont innovantes. Fournir des garanties de confiance et de transparence peut être extrêmement favorable au business.

### Des réflexions initiales orientées vers la protection des données personnelles en Europe

En Europe, les réflexions ont surtout été orientées vers la protection des données personnelles. La CNIL a été très active et maintenant le RGDP vient encadrer l'utilisation des données au niveau européen. Le RGDP est clairement une bonne avancée concernant la garantie des droits des citoyens par contre il ne garantit pas nécessairement l'explicabilité et la transparence des services numériques.

La démarche européenne, plus généralement, s'oriente davantage vers la protection, d'une part car les algorithmes n'ont pas encore franchi ce caractère prescriptif et, d'autre part, car les grandes plateformes comme Google ou Facebook qui se nourrissent de données européennes ne sont, elles, pas européennes. L'Europe a donc été naturellement dans une réaction défensive, de contrôle et de régulation des données personnelles, carburant de ces plateformes.

Le respect de la protection des données peut cependant donner des chances aux acteurs européens et leur donner des arguments compétitifs vis-à-vis d'acteurs américains. Naturellement, il ne faut pas que la régulation freine les fournisseurs de technologies européens vis-à-vis d'acteurs qui n'ont pas cette réglementation qui pèse sur eux et qui de fait ont plus de marge de manœuvre.

Dans tous les cas, le déploiement de services numériques, incluant des algorithmes de traitement de contenus, faisant de la proposition de décision, en soutien aux décisions humaines ou remplaçant même l'intervention humaine, impose de définir des garanties éthiques et c'est ce qui a motivé l'émergence de TransAlgo.

La Commission Européenne a, par ailleurs, récemment sorti un appel d'offres sur 18 mois, une étude sur la transparence et la redevabilité, ce que l'on appelle «accountability» des algorithmes. Il y a donc une vraie prise de conscience de cette nécessité pour les entreprises de rendre compte et de pouvoir disposer d'algorithmes auditables. Il s'agit là, désormais, de la responsabilité sociale et juridique des entreprises. Ce volet a des implications sur la nécessité d'adaptation et d'évolution des formations des étudiants et ingénieurs en science des données. Le concept essentiel derrière ces initiatives est l'asymétrie informationnelle. Par exemple, quand on nous recommande un bien ou un service, est-ce que la recommandation est vraiment en fonction des traces laissées par le citoyen et au bénéfice du citoyen, au sens de la personnalisation du service, ou est-ce que c'est au bénéfice de l'entreprise qui cherche à vendre ce bien ou ce service plutôt qu'un autre ? Est-ce que le livre que l'on me recommande est celui qui est fait pour moi ou est-ce le produit que l'industriel cherche à écouler ? Est-ce que l'algorithme est loyal pour celui qui consomme le service ou est-ce qu'il est loyal pour son concepteur ?

À service égal,

ce sont les acteurs

les plus transparents

et responsables qui

remporteront la

course

Quand on parle  
de transparence des  
algorithmes, cela ne veut  
en aucun cas dire qu'il faut  
mettre à disposition le  
code source

## Quelle est l'origine de TransAlgo ?

★ La Loi pour une **République Numérique** consiste en une véritable avancée par rapport aux autres pays européens et c'est elle qui a été le déclencheur de l'initiative TransAlgo. Le précédent gouvernement, via Axelle Lemaire, avait missionné le Conseil Général de l'Économie pour établir un rapport nommé « Modalités de régulation des algorithmes de traitement des contenus ». La mission était clairement orientée vers la régulation des algorithmes. Il y a eu une très large consultation, réunissant académiques, industriels, associatifs... qui a abouti à 5 recommandations prioritaires. La première recommandation opérationnelle de ce rapport était de **lancer une plateforme scientifique pour l'éthique des algorithmes**. Une fois énoncé par la loi que le citoyen a le droit à l'explication des décisions algorithmiques dans le cadre des services numériques publiques, comment aller plus loin et appliquer cette loi ? Dans les faits, tous les algorithmes ne sont pas explicables scientifiquement et techniquement aux citoyens, et il est particulièrement difficile d'appliquer cette loi. INRIA a donc proposé **TransAlgo en tant que plateforme pour le développement de la transparence et de la responsabilité des systèmes algorithmiques**. Inria s'est vu confier le rôle d'opérateur de TransAlgo, avec le soutien du Conseil national du numérique (CNNum) et de l'Institut Mines Télécom (IMT), et est chargé de jouer le rôle de catalyseur de la dynamique scientifique avec d'autres partenaires académiques, notamment le CNRS. Outre l'expertise scientifique, Inria apportera l'aide au développement logiciel.

## Les missions de TransAlgo ?

★ TransAlgo aura pour mission d'améliorer la **prise de conscience collective des enjeux liés aux algorithmes**, ce que la Commission Européenne définit par la « **Transparency Awareness Rising** ». Ces questions sont souvent orientées citoyens et grand public, mais les mêmes questions se posent dans un contexte inter-entreprises. Certaines interrogations peuvent même se transformer en question de sou-

veraineté nationale. Dans le domaine de la sécurité intérieure, les entreprises qui travaillent sur toutes les données sensibles ont intérêt à avoir toutes les garanties de transparence et d'auditabilité.

TransAlgo a pour vocation de développer des recherches sur la neutralité du classement, l'explication des méthodes d'apprentissage, les biais des données et des algorithmes d'apprentissage. Développer des recherches autour de ces sujets renforce la maîtrise des technologies.

## Quels enjeux pour des algo- rithmes éthiques ?

★ **Un algorithme peut être très puissant mais s'il est entraîné avec des données biaisées ou des données non à jour, pas assez représentatives, bruitées, il ne sera pas digne de confiance.** De même un algorithme d'apprentissage est un système à réaction positive car il apprend les données du passé et a tendance à perpétuer le passé à travers la projection ou la généralisation construite par le modèle. On parle alors de bulle filtrante car **le passé peut renforcer l'avenir**, c'est une source de risque et d'exclusion. La Justice Prédictive est très critiquée pour ce risque. Il y a donc de **nombreuses sources de questionnement, non sur les algorithmes eux même, mais sur la manière dont ils sont déployés et utilisés.**

Une raison supplémentaire de vigilance réside dans le nombre croissant d'algorithmes de Big Data et d'IA en open source de plus en plus déployés comme TensorFlow. Celui qui déploie ces technologies n'en est pas le concepteur et les données d'apprentissage, de paramétrage ne peuvent donc pas être maîtrisées.

**Il est donc primordial de bien former les futurs Data Scientists sur la réutilisation de codes et sur la responsabilité de la réutilisation de codes.** Une étude (Mobilitics) a été menée, par la CNIL et une équipe de recherche de l'INRIA, autour d'une application sur un appareil mobile. Cette application demandait le consentement de l'utilisateur pour partager ou non sa position GPS mais l'étude a mesuré les informations entrantes et sortantes et a été ainsi en mesure de prouver que l'application mobile outrepassait le consentement de l'utilisateur en communiquant sa

position GPS quelque-soit la réponse de l'utilisateur. En l'occurrence, les responsables n'étaient pas au courant et ont dû diligenter une enquête interne pour comprendre d'où venait le problème. L'application avait été développée par réutilisation de code open source mais ses fuites informationnelles n'avaient pas été maîtrisées. Cela interpelle sur **la nécessité de former les Data Scientists à l'économie et au droit des données ainsi qu'à l'éthique des données.** Cela responsabilisera les ingénieurs et ne pourra qu'alléger la responsabilité des dirigeants. Il est également nécessaire de **construire des algorithmes auditables par construction** et non de faire appel à du rétro engineering par exemple.

Il est important de distinguer la fiabilité et l'éthique d'un algorithme. La fiabilité d'un algorithme peut être vérifiée sur le plan informatique pour vérifier qu'il n'y a pas de bugs. La vérification d'un code source est plutôt bien développée depuis des années mais en étudier la transparence et le comportement est très différent. Un code source peut être fiable et vérifiable et pour autant n'être ni responsable ni éthique. **Quand on parle de transparence des algorithmes, cela ne veut en aucun cas dire qu'il faut mettre à disposition le code source.** Les acteurs industriels étaient inquiets de la démarche TransAlgo, car ils pensaient qu'on allait les obliger à publier leur code source. La démarche de TransAlgo n'est pas de violer les secrets industriels mais de **comprendre le comportement des algorithmes.** C'est une approche véritablement win-win pour les industriels.

## Peut-on contraindre les acteurs business à respecter les aspects éthiques des algorithmes ?

★ Je ne suis pas en faveur de la régulation des algorithmes par TransAlgo. La mission de TransAlgo est de mettre à disposition des outils d'auditabilité des algorithmes afin d'élaguer ceux qui ne posent pas de questions et d'identifier rapidement ceux qui demandent plus d'investigations. Il reviendra ensuite au régulateur de faire son travail dans le cadre des lois existantes.

## Les différentes phases de développement de TransAlgo

1 Dans un premier temps, notre objectif est de **créer un centre de ressources** dont l'ouverture est prévue dans le courant de l'automne 2017 à destination du grand public. Nous avons donc constitué un comité scientifique et éditorial pluridisciplinaire, réunissant informaticiens, mathématiciens, sociologues, économistes, juristes... pour clarifier, conceptualiser et définir un référentiel des terminologies utilisées telles que la transparence, la loyauté, la responsabilité, la non-discrimination des algorithmes.

Nous souhaitons répondre à un certain nombre de questions telles que :

- Quand dit-on d'un algorithme qu'il est loyal et par rapport à qui ?
- L'équité d'un algorithme est-elle absolue ? Ou l'équité individuelle peut-elle s'opposer à l'équité collective ?
- La non-discrimination d'un algorithme peut-elle être remise en cause par la discrimination positive au bénéfice du citoyen ?
- Quelle distinction fait-on entre explicable et éthique, juridiquement responsable et éthique ?

Par exemple, les algorithmes dans le domaine du marketing politique ne sont pas interdits par la loi. Pourtant l'impact du marketing politique sur le Brexit ainsi que sur les élections américaines divise (Cambridge Analytica), on peut se poser la question de l'éthique de ces algorithmes qui ont pour but explicite d'influencer les opinions. Ce n'est pourtant pas illégal dans l'absolu même s'il peut être utile en permettant d'atteindre des cibles qui ne l'étaient pas auparavant.

L'effet indésirable, c'est quand un algorithme peut changer l'opinion d'une personne et a fortiori son vote qui aura un impact durable sur la vie politique d'un pays.

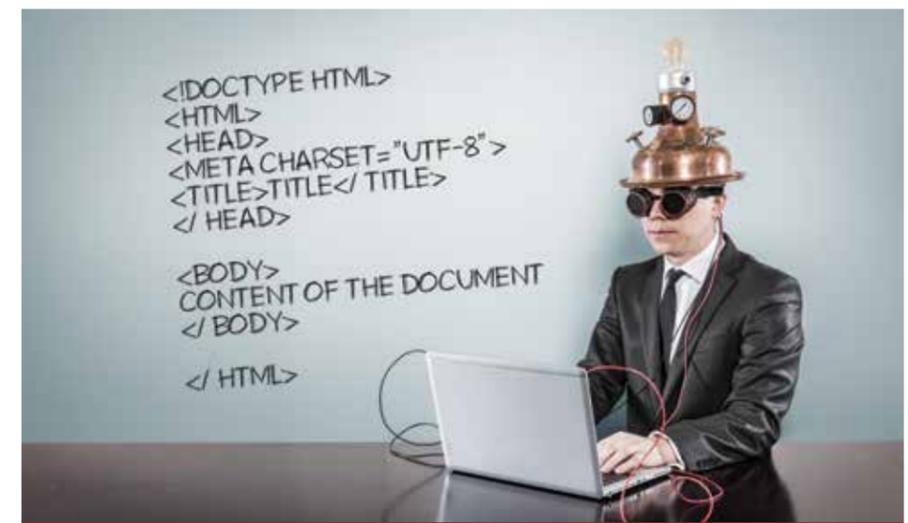
Afin d'étayer ces questionnements éthiques, nous collectons des ressources, des livres blancs, des rapports, des articles scientifiques et grand public, des jeux de données et enfin des algorithmes en open source, avec trois points d'entrée :

- Domaines applicatifs : police, justice, médecine, finance...
- Typologie des systèmes analytiques : algorithmes de classement, de recommandation etc...
- Questions éthiques et juridiques : responsabilité, redevabilité, opacité, discrimination, loyauté, asymétrie informationnelle etc.

**Nous travaillons également sur un accès facile et libre à ces ressources grâce à un moteur de recherche pour devenir un véritable centre de référence online sur les sujets, de manière totalement ouverte.** Afin que tous puissent échanger sur le comportement des algorithmes, ce centre sera ouvert à l'enrichissement par la communauté scientifique, les professionnels mais également le grand public au travers d'un débat modéré.

2 TransAlgo souhaite également **organiser des challenges pour encourager le développement de la recherche** et répondre à des problématiques clairement exprimées par la DGCCRF telles que la discrimination des prix ou la neutralité de classement et qui émanent directement de problématiques citoyennes.

3 Nous souhaitons enfin **créer des Moocs** pour différents publics : **grand public, scientifique** mais également pour les **autorités de régulation et pouvoirs publics** afin d'aider à la montée en compétences. Il est primordial de fournir aux autorités de régulation du numérique comme l'Autorité de la concurrence, la répression des fraudes des outils afin de leur permettre de s'approprier les enjeux de la transformation numérique.





★ Xavier  
**LORPHELIN**  
Managing Partner

SERENA  
CAPITAL



Xavier Lorphelin est associé fondateur du fonds d'investissement Serena Capital, créé en 2008, aux côtés de Marc Fournier et Philippe Hayat.

Serial entrepreneur comme ses associés, Xavier Lorphelin possède plus de 20 ans d'expériences dans le développement et l'investissement dans des entreprises innovantes et technologiques. Il a travaillé successivement dans la Silicon Valley, pour Flextronics, leader mondial de la sous-traitance en fabrication électronique, et JSourcé, startup spécialisée dans les solutions logicielles basées sur la technologie Java et dédiée aux cartes à puce et aux systèmes de paiement, qu'il a co-fondée en 1997.

Xavier Lorphelin a également co-fondé l'incubateur Kangaroo Village avant de gérer l'activité capital risque de la Société Générale pendant 6 ans.

Il est enfin diplômé de l'École Polytechnique, de l'École Nationale des Ponts et Chaussées et de l'Université de Stanford.

# À la recherche des pépites du Big Data et de l'IA

Quelle est la **génèse** du lancement du **fonds Serena Data Ventures** ?

★ Quand nous avons fondé Serena Capital, nous avons lancé en 2008 un premier fonds d'investissement qui ciblait des sociétés innovantes qui avaient déjà validé leur business model et recherchaient des fonds pour accélérer leur développement commercial. Nous ne couvrons donc pas le financement « early-stage » des startups.

Suite à des échanges avec Bertrand Diard, fondateur de Talend et Jamal Labeled, fondateur d'Easyvista, tous les deux dans le domaine du Big Data, nous avons compris qu'il était primordial de mettre en place un écosystème dédié à l'accompagnement et au financement des startups, car le Big Data était en train de bouleverser profondément toutes les industries. Nous avons donc lancé, fin 2016, Serena Data Ventures, **le premier fonds de venture capital européen dédié aux jeunes startups spécialisées dans le Big Data, le Machine Learning et l'Intelligence Artificielle**. Nous avons bouclé un premier closing de 80 millions d'euros début 2017.

Notre vocation est de financer des **startups « data-driven »**, qui développent des solutions et applications à partir du traitement et de l'intelligence qui peuvent être retirés de la data.

Notre champ d'intervention est très large. Nous accompagnons aussi bien **des startups qui développent des technologies intrinsèques au domaine de la data**, comme Dataiku qui a développé des outils utilisés par les data scientists, que **des startups qui développent des applications plus verticales** avec des algorithmes et des jeux de données pertinents par rapport à une industrie donnée. Les applications peuvent être dans des domaines très divers comme

la cybersécurité, la finance et l'assurance, l'énergie, le médical, l'automobile etc. Il s'agit alors d'un véritable gisement de valeur car **elles vont permettre**, la plupart du temps, **de développer de nouveaux services, propres à une industrie**.



Comment se positionne **la France** en matière de **Big Data** et **d'Intelligence Artificielle** ?

★ Nous avons en France une vraie carte à jouer car nous disposons de **fondements technologiques forts notamment en matière de data et d'algorithmie**. Pour pouvoir développer de nouveaux services et usages, il faut pouvoir faire tourner des

**Nous avons en France une vraie carte à jouer car nous disposons de fondements technologiques forts notamment en matière de data et d'algorithmie**

algorithmes performants. En France, nous disposons d'un très bon capital technique et avons toujours eu **un vivier d'ingénieurs et de mathématiciens assez unique**.

Je pense que nous n'avons pas à rougir d'une comparaison internationale et le fait que des laboratoires d'Intelligence Artificielle s'implantent en France n'est pas neutre quant à l'excellence française en la matière. Certaines startups françaises croissent d'ailleurs de manière exponentielle aux États-Unis comme Dataiku, que nous avons déjà cité.

L'Angleterre a également une vraie expertise autour de la data, mais très concentrée autour de la Fintech. Nous avons la chance en France d'avoir un **tissu industriel fort et multi-sectoriel**, ce qui permet de **booster l'innovation et de favoriser l'éclosion de startups** dans des domaines spécifiques mais variés comme **l'énergie, l'industrie, le BTP, le retail, les cosmétiques et la fashion tech, l'aéronautique, le médical, le pharmaceutique** etc. Les startups peuvent ainsi bénéficier du soutien des leaders du marché comme l'Oréal, Carrefour, Auchan, Airbus... Pour être performantes, elles doivent avoir accès aux datas pour faire tourner leurs algorithmes et celles-ci sont majoritairement aux mains des gros industriels et corporates. Les startups bénéficient donc d'un accès privilégié à la data grâce aux partenariats avec ces acteurs, ce qui leur permet même d'être à pied d'égalité avec Google ou encore Facebook qui n'ont pas facilement accès à ces datas sectorielles. **L'omniprésence de Google ne semble donc pas être une véritable menace dans la mesure où les startups françaises sont plutôt verticales**. Google essaye d'ailleurs de développer des accords, par exemple, dans le secteur médical, pour obtenir des données.

Par ailleurs la France devient de plus en plus compétitive en matière de financement publics et privés. **L'écosystème français des Venture Capitalists n'a qu'une vingtaine d'année en France, contre 40 ans aux États-Unis, mais il est cependant de plus en plus structuré et performant**.

## Quels sont les secteurs porteurs, les principaux secteurs financés ?

★ La **première vague** de financement a plutôt été tournée vers des startups dans le domaine du **marketing, orientées sur la connaissance, l'acquisition et la rétention client**.

La **seconde vague** d'investissements a touché le domaine de la **finance**, tirée par l'engouement pour la Fintech. Les technologies et algorithmes Big Data sont particulièrement adaptés à la collecte des données en temps réel, afin de prendre de meilleures décisions de trading dans le cas des Robo-Advisors. Le modèle dématérialisé de la finance et du **secteur bancaire** a ainsi amplement tiré l'innovation.

De même, dans le **secteur de l'assurance**, le modèle de l'actuariat est un modèle par essence algorithmique. Les assurances intègrent de plus en plus de datas personnelles mais également des datas d'usage, par exemple celles des véhicules connectés, ce qui va accélérer la mutation dans le domaine automobile.

Au-delà de la finance et de l'assurance, on distingue plusieurs secteurs porteurs pour les startups data-driven :

**1** Le digital et la data impactent tout d'abord le **secteur de la santé** et du médical. Malgré des contraintes réglementaires fortes, de nombreuses disruptions sont en cours dans ce secteur qui commence à intégrer du prédictif dans les diagnostics et le suivi des patients. L'entreprise **INATO** s'illustre par exemple en début de la chaîne, en mettant le prédictif au service du lancement de nouveaux médicaments et en améliorant les processus de tests grâce au Big Data.

**2** Le **secteur automobile** est également très intéressé par l'usage de l'Intelligence Artificielle et du machine learning dans le cadre des **véhicules autonomes et connectés**.

**3** Dans le secteur du retail, les **sujets de traçabilité alimentaire et d'optimisation des stocks pour les retailers** représentent un vrai vecteur de croissance. La société **VEKIA** propose par exemple des algorithmes permettant de prédire les stocks afin d'**améliorer les process de supply chain** et de limiter les invendus.

**4** La **Smart City et la domotique** n'en sont également qu'à leurs débuts. L'IOT dans ce domaine doit être vu plus sous l'angle data que hardware. Les capteurs ne sont finalement qu'une commodité pour récupérer des données mais de **nombreux modèles algorithmiques sont encore à développer** avec des champs d'application très larges.

**5** Les secteurs traditionnels sont également porteurs, notamment **l'agriculture, le BTP, l'énergie** où les enjeux de productivité et de rentabilité sont très forts et où la data peut permettre des gains importants. De même **l'immobilier** est en train d'être disrupté avec des modèles qui permettent de déterminer quel est le meilleur emplacement pour un commerce dans une rue donnée.

## Des exemples de startups déjà sélectionnées et financées par Data Ventures ?



### ★ **CYBERSÉCURITÉ À PROPOS DE CYBELANGEL**

**CYBELANGEL** propose un service de détection de fuites de données sur Internet. Ils surveillent le Dark Web et les objets connectés pour identifier des menaces qui pourraient porter atteinte à votre entreprise. Ils proposent également une solution de lutte contre la contrefaçon sur Internet. Ils identifient les réseaux de contrefacteurs qui portent atteinte à l'image de marque d'une entreprise. La proposition de valeur dans le domaine de la cybersécurité parle directement aux grands groupes et le nombre de startups à surveiller dans le secteur est croissant.

### ★ **BTP À PROPOS DE FINALCAD**

**FINALCAD** est une startup qui propose de digitaliser tous les process liés au suivi de chantier. Elle permet de numériser et normaliser tous les processus sur le terrain, afin de pouvoir capitaliser sur les bonnes pratiques, faire vivre la maquette numérique sur le chantier (tel-que-construit) et faciliter la collaboration entre les différents intervenants. Elle propose notamment de collecter les datas afin de prédire les défauts potentiels d'une construction et de définir les risques à envisager au cours de celle-ci.

### ★ **FASHION & BEAUTY À PROPOS D'HEURITECH**

Fondée en 2013 par deux chercheurs en Intelligence Artificielle, **HEURITECH** a développé, grâce à des technologies de Deep Learning, une solution de reconnaissance automatique et en temps réel des objets, formes, textures, personnages, à l'intérieur de toute image ou photographie postée sur les réseaux sociaux et Internet. Cette innovation ouvre de nombreux cas d'applications, notamment dans le domaine du luxe et de la mode en permettant la détection des tendances vestimentaires pour la conception des collections et la gestion des achats/réassorts, l'optimisation des recommandations et du merchandising sur les plateformes d'e-commerce, la segmentation qualitative et quantitative des clients...



Quels sont les **critères de sélection** et d'**évaluation** des **startups Big Data** et **IA** financées et soutenues ?

**1** Un des premiers critères est celui de l'**expertise technologique de l'équipe**. Il est nécessaire pour une startup d'avoir un vrai savoir-faire technologique, une barrière technologique suffisante à l'entrée qui permette de présumer que cet acteur pourra s'imposer durablement et potentiellement devenir un leader mondial.

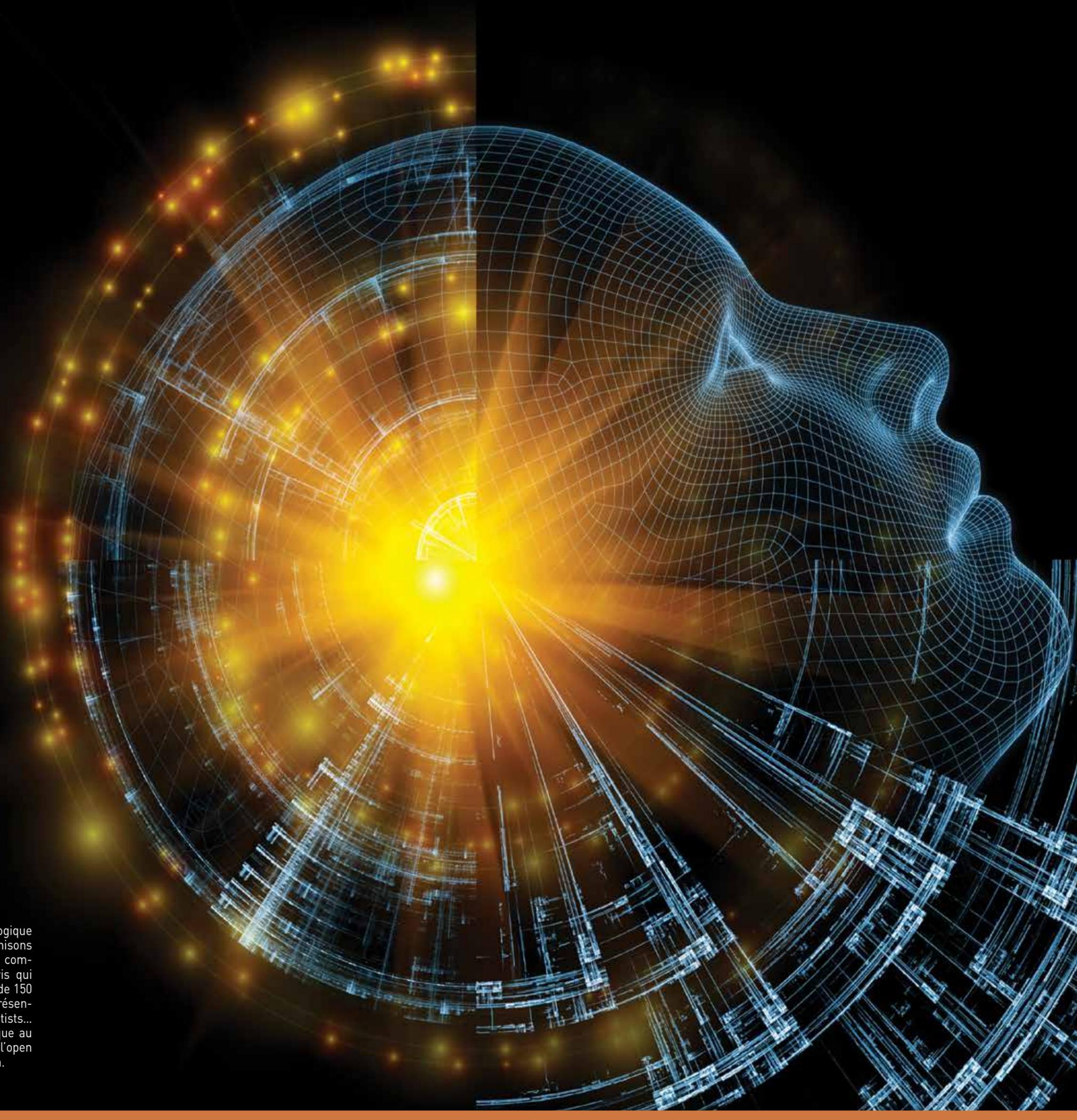
**2** Au-delà de l'expertise technologique, il est primordial que l'équipe ait **une bonne connaissance des enjeux business et des besoins clients** afin de pouvoir s'assurer que la solution développée par la startup réponde à un vrai problème.

**3** Il est essentiel pour une startup dans le domaine du Big Data de pouvoir valider très en amont la pertinence de sa solution sur le marché adressé. Nous privilégions des startups qui ont déjà réussi à mettre en place un ou plusieurs **Proof of Concept** avec des clients ou partenaires potentiels.

Quel **écosystème** mettez-vous en place **pour détecter de nouveaux champions** de la data ?



Nous sommes dans une logique d'écosystème ouvert et nous organisons plusieurs événements ouverts à la communauté, comme Data Driven Paris qui réunit sur un rythme mensuel plus de 150 participants : entrepreneurs, représentants de grands groupes, data scientists... Cela permet de créer une dynamique au sein de l'écosystème et de favoriser l'open innovation dans le secteur de la data.





☆ Benjamin  
**JEAN**

Open Data & Open  
Innovation Legal Expert

**Inno<sup>3</sup>**

CoFondateur et Président  
«Open Law,  
le Droit Ouvert»

**Benjamin Jean** est juriste de formation, spécialisé en propriété intellectuelle et expert en modèles ouverts. Fort d'un background juridique et informatique, il travaille sur les nouveaux usages collaboratifs introduits par le numérique au sein de chaque secteur et a rédigé plusieurs ouvrages sur le sujet dont « Option Libre, du bon usage des licences libres ».

Il a créé le cabinet Inno<sup>3</sup>, qui revêt un rôle de facilitateur et de catalyseur au profit de la création et de l'innovation dans des entreprises et organisations publiques souhaitant adopter ou prendre en compte les modèles partagés et collaboratifs (Open Source, Open Data, Open Innovation) au sein de leurs politiques d'innovation. Il est président de l'association Open Law, consultant en cabinet d'avocats, enseignant en grandes écoles et universités, et chercheur associé au Centre de Recherches Juridiques (CRJ).

En 2016, il a pris la coprésidence de l'Open Source Summit ; édition qu'il a conduit sur la thématique : « Empowering Open Innovation ».



# Open law, le droit ouvert

Pourquoi vous êtes-vous intéressé aux sujets tels que **l'Open Source, l'Open Data & l'Open Innovation** ?

★ Au cours de mes études de droit, je me suis assez rapidement forgé l'opinion que **l'Open Source, l'Open Data et plus largement l'Innovation Ouverte sont une manière de créer et d'innover de plus en plus pertinente à l'ère du numérique**. La transformation numérique se généralisant à tous les secteurs, elle amène progressivement l'ensemble des acteurs à utiliser le même langage et les mêmes outils. Ces méthodes collaboratives apparaissent naturellement comme **une voie nouvelle de communication et de coopération, voire de « coopération »**, entre des acteurs qui ne se parlaient pas auparavant. C'est particulièrement le cas pour **le modèle de l'Open Source** qui rend possible le développement de logiciels entre des centaines voire des milliers de parties prenantes grâce à un cadre juridique et une gouvernance clairs, transparents et équilibrés. Certains acteurs économiques éprouvent encore des difficultés à trouver leur place dans cette nouvelle façon de faire, mais la situation évolue néanmoins inexorablement et **les banques** sont par exemple de plus en plus investies sur ce sujet. La Société Générale est particulièrement active en ce moment, opérant sa propre transformation et soutenant des initiatives telles que l'Open Source Summit ou encore, plus récemment, la production d'un MOOC ouvert et dédié au sujet de l'Open Source.

D'autres secteurs, tel celui de la mobilité, représenté par la Fabrique des mobilités en France, vont encore plus loin et investissent massivement dans l'émergence et la pérennisation d'écosystèmes ouverts.

Sur le même modèle, **l'Open Source Hardware** applique ces logiques de partage et de collaboration au **développement de ressources matérielles**. Un phénomène qui touche les objets 3D triviaux mais aussi les voitures et satellites. Dans le secteur de la santé, il est intéressant de considérer des projets tels qu'EchOpen qui réunissent médecins et passionnés pour permettre le développement de matériels médicaux professionnels.

**L'Open Data** est un sujet plus récent qui s'est particulièrement développé sous l'impulsion des dernières réformes législatives et qui **vient alimenter les démarches de Big Data ou encore Data Lakes**. C'est néanmoins un mouvement plus ancien et plus large qui se présente souvent comme la seule alternative pérenne face aux géants d'Internet que sont Google, Apple, Microsoft et Amazon. Ainsi le projet international **OpenStreetMap**, sorte de wikipedia des données géographiques fondé en 2004, a permis de collecter des données dans le monde entier sur les routes, voies ferrées, les rivières, les forêts, les bâtiments et bien plus encore. Les données cartographiques collectées sont réutilisables sous licence libre OdbL qui est un cadre particulièrement intéressant qui permet l'usage par tous et impose comme obligation principale de partager ses propres améliorations sous la même licence dans une logique de réciprocité.





Mêlé à la transformation numérique, l'Open Innovation trouve un potentiel nouveau en fournissant un environnement neutre propice au dialogue, au partage, à l'innovation et au développement de nouveaux usages en capacité d'éliminer les impasses, de surmonter les résistances, de décompartmenter les métiers.

Un dernier exemple, encore dans le domaine de la santé : le **laboratoire Roche** s'est récemment illustré en matière d'**Open Data et d'Open Innovation** en lançant le programme **Epidemium**. Organisé en partenariat avec le laboratoire communautaire La Paillasse en vue de **tirer parti des méthodes d'analyses du Big Data et du crowdsourcing**, il a permis d'améliorer la recherche dans l'épidémiologie du cancer. Pour ce faire, il a conçu une base de données ouverte sur le cancer, baptisée Roche Open Database, afin de mettre à disposition ces informations à des équipes d'étudiants, de statisticiens, de biologistes et d'experts en data visualisation. Encore un exemple d'une ressource qui s'enrichit du partage qui en est fait.

Vous êtes cofondateur du **projet Open Law**, quelles sont vos motivations à encourager **l'ouverture du droit ?**

★ La combinaison de mon parcours professionnel et de mon domaine d'activité m'a rapidement amené à réfléchir à l'intérêt des modèles ouverts, et plus largement **l'application des principes de l'innovation ouverte au champ du droit**.

L'initiative a été lancée en s'appuyant sur l'adage « Nul n'est censé ignorer la loi », alors qu'à l'époque **l'accès aux bases de données de Légifrance était soumis à des conditions commerciales rédhitoires pour beaucoup d'acteurs**. Notre ambition était alors de soutenir le mouvement de l'open data afin que tout acteur puisse exploiter les données et créer une valeur ajoutée à celles-ci. Les missions d'Open Law se sont ensuite précisées au fil des programmes :

**1** Promouvoir des actions favorisant **l'ouverture et l'interopérabilité des données juridiques, et l'innovation collaborative autour de ces données**, ainsi que tous travaux de recherche scientifique portant sur des technologies favorisant l'utilisation ou l'accès au droit.

**2** Réunir et animer une communauté **d'innovateurs du droit**, qui participe à la construction d'un écosystème pérenne favorisant le décloisonnement des métiers et la co-création dans les milieux juridiques et judiciaire.

**3** Développer des ressources juridiques numériques ouvertes, pour soutenir l'Open Data et les mouvements Open Source.

**4** Fournir un environnement neutre propice au dialogue, au partage, à l'innovation et au développement de nouveaux usages.

Pour le reste, nous nous sommes appliqués à nous-mêmes les préconisations

classiques en termes de projet de co-création et avons fait en sorte de soutenir cette dynamique forte en accompagnant le nombre croissant de contributeurs rencontrés. L'idée est simple : **plus le cadre est clair et précis, plus les gens sont libres et développent cette collaboration**.

**Le rôle d'Open Law est donc de définir un socle de confiance**, notamment par l'usage d'outils contractuels tirés de l'Open Source : - l'usage systématique de licences **Open Source ou Open Data** sur nos créations afin de permettre juridiquement l'utilisation plurielle et le développement collectif de ces ressources. - un statut de **contributeurs rémunérés aux communs** par lequel les contributeurs d'une ressource jugée centrale sont rémunérés pour diffuser leurs contributions en Open Source, Open Data, etc. Cette démarche participe à une **mutualisation de ressources numériques**.

Au sein d'Open Law, nous comptons aujourd'hui de nombreuses personnes physiques et des organisations comme AXA, Dentons, la DILA (La Direction de l'information légale et administrative), les Editions Lefebvre Sarrut, Wolters Kluwer France, SVP, CMS. Certains sont concurrents, voire en situation client-prestataire, mais tous ont fait le pas de **soutenir cette démarche d'ouverture et de transparence au profit de l'écosystème dont ils font partie**.

## Quelles sont les **victoires** récentes concernant **l'ouverture des données ?**

★ La Loi pour une République Numérique est à cet égard La grande réforme, la dernière des quatre réformes du précédent quinquennat, faisant de **la France un pays précurseur avec une véritable réflexion sur le bénéfice que peut tirer le secteur privé de cette ouverture des données**. Cette Loi se combine par ailleurs à la présidence française en 2016 de l'OGP, le Partenariat pour un Gouvernement Ouvert, qui a donné lieu à une sollicitation forte sur ces sujets des différents ministères et de la société civile. De nombreux nouveaux acteurs vont certainement monter sur le sujet dans les prochains mois.

Ainsi, notre 5ème programme au sein d'Open Law était consacré à **l'ouverture de la Jurisprudence**, dans l'idée d'anticiper et de préparer l'ouverture programmée dans le domaine par la Loi pour une République Numérique. Etalab, le Ministère de la Justice, la Cour de Cassation, la Cour des Comptes, le Conseil d'Etat, l'Ordre des Avocats de Paris, la CNIL et l'INPI ont été réunis à l'initiative de l'association Open Law **pour échanger et expérimenter l'ouverture des données juridiques. Les avocats sont les premiers concernés par l'ouverture de la Jurisprudence**. La force de l'ouverture, c'est que tous les avocats auront accès au même niveau d'information. On constate actuellement une réelle asymétrie de l'information avec des indépendants ou petits cabinets qui n'ont pas les moyens de s'abonner à une grande revue juridique leur donnant accès à l'information. Si les avocats sont les premiers bénéficiaires de cette ouverture au regard de leur pratique, **l'objectif était de partager les bonnes pratiques dans le secteur afin de permettre une ouverture réussie au profit de tous et d'anticiper les risques potentiels**.

Dans la lignée, un certain nombre d'acteurs privés et publics, tels qu'**ELS, la Cour de Cassation, la DILA, Wolters Kluwers ou encore l'IMT, travaille à la conception d'un jeu de données ouvert d'entraînement d'algorithmes**. Accompagné d'une réflexion éthique afin d'évacuer les biais, comme en matière d'évaluation de pensions alimentaires qui diffèrent en fonction

des époques et selon le sexe du demandeur. Ce nouveau jeu de données permet ainsi d'agréger un nombre important de données, disponibles ou ouvertes à cette occasion, qui sont « alignées » et formatées de sorte à pouvoir facilement être exploitées par des algorithmes classiques tels les projets Open Source de Google ou Facebook.

C'est un travail qui a naturellement été réalisé en open innovation et open source entre acteurs publics et privés même concurrents pour **permettre à des porteurs de projet d'avoir accès à un maximum de données**, une matière première souvent difficilement accessible pour un nouvel entrant. En cela, il **réduit au maximum les obstacles et coûts d'entrée pour tout nouvel acteur économique et permet d'accélérer la mutation du secteur**.

## L'Open Law, déjà une réalité dans le monde du droit ?

★ On ne peut pas dire que tous les acteurs du droit y contribuent, mais nous y travaillons dans la démarche d'inclusion et de transparence qui est la nôtre. De fait, nous constatons un intérêt croissant des acteurs, traditionnels et nouveaux, pour ces idées et méthodologies. De plus en plus de jeunes praticiens, avocats, juristes, notaires, etc., sont intéressés par l'idée de **repenser leur métier au regard de ce qu'apportent les nouvelles technologies dans le domaine du droit**. C'est un mouvement qui prendra nécessairement encore du temps et **force est de constater que le métier d'avocat n'est pas encore directement entièrement transformé par l'open data**, ni même globalement conscient d'une nécessité de « transformation numérique » même si quelques cabinets internalisent des développeurs et incubent des startups.

Open Law, dans son 3ème programme, a d'ailleurs insufflé une réflexion sur la stratégie économique et business entre avocats et autres professions réglementées (éditeurs juridiques, éditeurs de logiciels, grands clients, et institutions) afin de les aider à repenser leur organisation mais également leurs rôles à l'ère du numérique. Cela a un intérêt aussi bien pour les acteurs traditionnels, originaires du droit, que pour les porteurs de technologies qui

s'intéressent au droit comme marché depuis quelques années.

On a pu observer initialement une lutte entre ces deux types d'acteurs, **les nouveaux étant souvent qualifiés de « braconniers du droit »**, mais l'on commence à recenser des projets communs. **Le groupe Editions Lefebvre Sarrut** a par exemple conclu un partenariat avec **Rocket Lawyer**, acteur de référence américain pour développer une offre commune sur le marché français. **Wolters Kluwers** a également conclu un partenariat avec **Predictice** dont vous retrouverez l'interview ci-après.

Ces rapprochements témoignent de la volonté du secteur du droit d'entrer dans une logique d'innovation ouverte pour développer de nouvelles offres. **Petits et grands acteurs bénéficient réciproquement les uns des autres**.

C'est peut-être une des forces de notre mouvement en France que d'avoir réussi à accompagner ces LegalTech au sein d'une **initiative conjointement soutenue par le secteur privé et public. Cela n'existe ni aux États-Unis, ni dans les autres pays d'Europe, où le secteur des nouvelles technologies essaye de disrupter le secteur juridique**. Les initiatives y restent donc majoritairement privées. Notre force repose ainsi sur la maturité des acteurs et la manière de collaborer, plus que sur les seules technologies encore majoritairement américaines.

Une des forces de notre mouvement en France est d'avoir réussi à accompagner ces LegalTech au sein d'une initiative conjointement soutenue par le secteur privé et public

Les  
**algorithmes**  
vont-ils  
**transformer**  
la **fonction**  
d'**avocat**,  
de **juriste**, de  
**magistrat** ?

★ Tous les sujets liés à l'**Intelligence Artificielle et aux chatbots** sont en train de faire bouger les lignes dans le domaine du droit. **La technologie permet notamment d'automatiser et d'accélérer des tâches ultras répétitives** comme l'affectation d'un dossier à une chambre pour un tribunal ou encore la gestion des formalités administrative en matière de données à caractère personnel. En revanche, **cela ne remplace en rien le rôle du professionnel du droit en matière de conseil et d'éthique.**

Un changement de culture et de rôle pour les fonctions juridiques est nécessaire. Les attentes évoluent et le rôle du juriste aussi, il sera beaucoup plus stratège et ses compétences pourraient même en faire la cheville ouvrière de projets de collaboration réunissant une grande diversité d'acteurs.

**Les technologies doivent permettre aux professionnels du droit de s'épanouir et de travailler plus en confiance et non de réduire la profession à un rôle de prestataire de services.** Lorsque l'on considère le marché « classique » du droit, il y a fort à parier que de plus en plus de juristes vont se lancer dans des projets entrepreneuriaux, des projets LegalTech, et vont faire du droit l'objet du service proposé. L'arbitrage en ligne, à l'image de la plateforme **eJust** à destination des entreprises ou dans une moindre mesure **Refundmyticket** spécialisée en automatisation de la résolution de litiges en matière de vols, va ainsi changer profondément notre rapport à la justice et soulager les tribunaux (1 à 2 jours de tribunaux du ressort de Roissy et d'Orly sont aujourd'hui consacrés à la résolution de ces petits litiges). **SupraLegem**, qui offre aux juristes un nouvel angle de vue sur la jurisprudence, **en appliquant l'Intelligence Artificielle à des centaines de milliers de jurisprudences administratives**, en est un autre. **Le moteur de recherche juridique doctrine.fr**, qui se présente comme un « **Google du Droit** », est un autre exemple d'une nouvelle façon d'appréhender le droit.

Quels sont selon  
vous les **futurs**  
enjeux du **Big**  
**Data**, de l'**Open**  
**Data** et de l'**IA**  
dans le droit ?

★ **La formation du juriste de demain est cruciale.**

C'est l'objet du dernier programme d'Open Law qui réunit tous les acteurs concernés par le sujet pour réfléchir ensemble aux nouveaux besoins auxquels devra répondre le juriste, en termes de compétences, de savoir être et de savoir-faire, afin de se lancer sereinement dans le monde numérique qui est le nôtre.

Réciproquement, nous traitons l'axe inverse qui consiste à identifier les réponses que le numérique apporte en termes de pédagogie et de support à la formation du juriste : Moocs, Serious Games, etc.

Tout cela sera expérimenté et documenté pour que chacun puisse ensuite s'approprier ces idées. Les sujets du Big Data, de l'Open Data, les algorithmes, etc. seront naturellement dans nos axes de développement.

**Les données à caractère personnel restent un sujet sensible** dans le monde du Big Data. **La réforme RGPD est imminente et va imposer un comportement plus responsable et éthique.** Les produits et applications devront être conçus en conformité à cette réglementation, dans un esprit proche des réflexions que nous menons en matière de **Legal Design**. L'objectif est de **conceptualiser by design les produits et services juridiques pour mieux répondre aux besoins et redonner confiance aux utilisateurs.**

Enfin, **la Blockchain trouve progressivement sa place dans le domaine du droit**, notamment comme registre infailible, utile pour prouver que tel document était bien possédé par telle personne à tel moment ou que telle somme a bien été transférée à telle date à une personne donnée. Les Smart Contracts intégrant la Blockchain permettront le suivi numérique et l'exécution automatique d'engagements définis au sein d'un contrat traditionnel, parfois particulièrement utile et innovant lorsque les règles peuvent être « codées » simplement qu'il s'agisse d'une Loi, par exemple en matière de compatibilité publique ou de RGPD, ou d'un contrat, tel qu'un contrat permettant la circulation d'œuvres d'art numériques.



# ★ Louis LARRET-CHAHINE

Co-founder

## Predictice



À 25 ans, Louis Larret-Chahine a déjà plusieurs casquettes : élève-avocat, créateur d'une galerie en ligne spécialisée dans l'art urbain, et co-fondateur de Predictice, start-up créée en janvier 2016 et consacrée à la Justice Prédictive, dont l'objectif est de donner aux avocats et aux juristes, l'accès aux probabilités de réussite des cas qu'ils plaident.

Predictice a développé un logiciel qui permet d'optimiser sa stratégie contentieuse en extrayant des analyses précises et fiables des recherches de jurisprudences.

# Le prédictif à l'assaut du droit

Quel **constat** faites-vous de l'**avancée** du **Big Data** et de l'**Intelligence Artificielle** dans le secteur **juridique** ?

★ Si l'on prend du recul, on distingue **deux vagues dans l'essor de la LegalTech**. La première s'est portée sur l'**automatisation de tâches à faible valeur ajoutée**, telles que la rédaction de contrats ou la mise en relation. Cette première vague a très bien fonctionné et il y a un marché qui s'est créé notamment aux États-Unis. Les **levées de fonds totalisent presque 1 milliard d'euros dans ce domaine**. Cela correspond à 90 % des sociétés nouvellement créées sur 2016 sur ce segment croisant l'informatique et le droit.

La seconde vague, restée pendant longtemps relativement marginale représente aujourd'hui environ **10% des sociétés nouvellement créées. Celles-ci placent le Big Data et l'Intelligence Artificielle au cœur de leur activité**, cette fois non dans un but d'automatisation ni de remplacement du professionnel du droit, mais avec l'objectif d'**améliorer la performance et d'aider à la prise de décision**. Il s'agit principalement de sociétés qui analysent des données pour en extraire les informations permettant de renforcer la pertinence, l'analyse, le savoir-faire du professionnel du droit, qu'il soit avocat, juriste ou magistrat. Le profil des sociétés et les usages évoluent en fonction de la technologie, qui elle-même influe sur l'usage. Cela crée un cercle d'influence positive.

Le domaine juridique est très longtemps resté concentré sur lui-même. Et, parce qu'il y a eu des résistances, le secteur cède comme de l'eau percerait un barrage.

Contrairement au secteur du e-commerce qui a mis 10-12 ans à se digitaliser, le **secteur du droit le fait en 5 ans**. Il subit une sorte de rattrapage à grande vitesse.

Les **LegalTech** sont principalement nées aux **États-Unis** mais **la France rattrape-t-elle son retard** en la matière ?

★ Les États unis ont été précurseurs et les premières sociétés à avoir fait de l'Analytics dans le domaine du droit sont américaines. **Les modèles technologiques ont été inventés et développés aux US, leur soft power est très fort dans le milieu informatique**. Certains des projets peuvent servir de source d'inspiration, comme Lex Machina ou Ravel Law, deux projets d'ailleurs rachetés par le groupe LexisNexis (cf. encadrés ci-contre).

Néanmoins, il ne faut pas oublier que les deux systèmes juridiques - Common Law et droit civil - sont si différents qu'une grande partie de leur système applicatif n'est tout simplement pas transposable. D'autre part, les solutions analytiques dans le domaine juridique sont peu scalables. **La France est donc relativement protégée par rapport à des sociétés américaines**.

En outre, nous sommes favorisés car il y a, d'une part, **un secteur juridique fort porté par de très bons**

juristes et, d'autre part, d'excellentes écoles d'ingénieurs, ainsi qu'un écosystème favorisant l'entrepreneuriat très dynamique. Le croisement entre l'analytique, les mathématiques et le droit est en grande partie aidé par l'excellence française dans ces deux domaines : sur le terrain on observe notamment une éclosion des **pratiques analytiques** témoignant d'une appropriation très rapide de ce système alors même qu'il est encore balbutiant dans tous les pays sauf aux États Unis.

S'agissant de Predictice, un nombre croissant de cabinets d'avocats s'abonnent. L'Etat français a par ailleurs conclu un partenariat avec nous pour faire tester le logiciel dans plusieurs juridictions. De plus, certains étudiants et chercheurs ont accès à notre outil grâce aux partenariats conclus avec les universités, comme Science Po Paris, l'Université de Nîmes ou des écoles de formation d'Avocats. Nous avons donc un cercle d'utilisateurs très dynamique, qui correspond à quasiment la totalité des acteurs du monde du droit et qui témoigne de l'appropriation rapide des technologies analytiques.

Contrairement au secteur

du e-commerce qui

a mis 10-12 ans à se

digitaliser, le secteur du

droit le fait en 5 ans

## ★ À PROPOS DE RAVEL LAW

Ravel Law, plate-forme juridique de recherche, d'analyse et de visualisation qui vise à transformer les avocats en analystes de données, s'est vendue cinq ans après sa fondation en 2012 à Lexis Nexis, le géant de la recherche juridique de New York. Ravel, fondé dans les dortoirs de Stanford Law School a recueilli plus de 14 millions de dollars de financement au cours des années, y compris de X Fund, New Enterprise Associates, North Bridge Venture Partners, Work-Bench et Ulu Ventures.

## ★ À PROPOS DE LEX MACHINA

La plate-forme Legal Analytics de Lex Machina, intégrée par LexisNexis dans Lexis Advance Online Legal Information Solution, est une nouvelle catégorie de technologies juridiques qui modifie fondamentalement la façon dont les entreprises et les cabinets d'avocats concourent dans les affaires et la pratique du droit. Livré en tant que Software-as-a-service, Lex Machina fournit des informations stratégiques aux juges, avocats, parties et plus encore, extraites de millions de pages d'informations juridiques. Cela permet aux cabinets d'avocats et aux entreprises de prédire les comportements et les résultats que différentes stratégies juridiques produiront. Lex Machina a été nommé «Best Legal Analytics» par les lecteurs de The Recorder en 2016, 2015 et 2014, et a reçu le prix «Meilleur nouveau produit de l'année» en 2015 par l'Association américaine des bibliothèques de droit.

## Quelle utilité de l'outil analytique au service du droit ? Quel impact sur l'activité humaine et quel potentiel pour l'utilisateur ?

★ Les solutions analytiques au service du droit permettent de comprendre très vite une grande masse de données et d'informations. Avec un outil comme Predictice, il est possible de choisir la meilleure argumentation, la meilleure stratégie et d'avoir une analyse de risque avec une granularité très fine.

Tout le travail des jeunes collaborateurs ou stagiaires qui consistait à cruncher les données à la main est potentiellement réduit. Cela pourrait impliquer un rétrécissement de la pyramide hiérarchique avec moins de personnes occupées à faire ces tâches et un recentrage sur la stratégie, la rédaction, la relation client...

Pour les assureurs de protection juridique, qui sont des clients historiques de Predictice, l'outil sert à établir un profil de risques et établir les chances de succès en 8 à 30 secondes d'une procédure, dans le but d'éviter un contentieux qui peut être traumatisant et onéreux. Dans ce cas-là, l'analytique est un outil de déjudiciarisation. Le juriste devient donc en grande partie médiateur et négociateur, le flux de contentieux diminue devant les juridictions, qui peuvent se concentrer sur les cas qui nécessitent vraiment leur attention. Cela pourrait donc contribuer à résoudre l'engorgement judiciaire.

S'agissant des magistrats, l'outil leur fournit une sorte de grille souple de préconisation de sentence beaucoup plus fine qu'un barème, qui pourrait permettre d'une part une forme d'homogénéisation des points de vue sur une même question, et d'autre part un renforcement de la nécessité de motiver les décisions d'un tribunal ou d'une juridiction qui déciderait de s'éloigner de ce qui est préconisé par la somme de l'ensemble des décisions de

ses pairs. Il ne s'agit pas là d'automatiser la justice. L'outil est présenté de telle sorte qu'on essaye d'éviter l'effet performatif, ce que l'on trouve dans de nombreux systèmes algorithmiques, où la préconisation agit comme prophétie auto-réalisatrice et a tendance à tordre le réel pour l'ajuster à sa propre prévision. On trouve cet effet performatif dans des systèmes qui à la fois suggèrent des décisions mais également la prennent, en flow automatique. Dans le cadre du droit, c'est loin d'être le cas car il y a bien souvent une présence des avocats et toujours d'un magistrat.

## Qu'est-ce qui vous a poussé à lancer Predictice ?

★ En cabinet d'avocat, j'ai été chargé de sortir 300 décisions de justice sur un thème donné et de faire des moyennes et des médianes sur un cas d'indemnité pour un client. Cela a fait tilt, il a semblé que ces tâches très chronophages devaient pouvoir être automatisées.

Avec deux amis, ayant la même croyance dans les possibilités de la technologie pour le monde du droit (Alexandre Chéronnet, qui pilotait le rapport sur les LegalTech pour la clinique du droit de SciencePo, et Thomas Baduel, ingénieur avec plusieurs expériences significatives), nous avons décidé de lancer le projet.

Le milieu professionnel est bien moins homogène qu'on peut le penser et il y a certaines spécialités qui sont particulièrement techniques. Nous avons donc décidé d'aller voir les meilleurs professionnels dans chacune des spécialités pour être sûrs de développer un outil utile, un outil créé par les professionnels du droit pour les professionnels du droit.

Nous avons sollicité toute une série de partenaires, de très gros cabinets d'avocats, de petites boutiques spécialisées, des assureurs, l'État, des facultés etc... pour nous assurer de l'utilité de la plate-forme. Ces partenaires ont également contribué à l'enrichissement des données.

## Qu'est-ce que la Justice Prédicative ?

★ La Justice Prédicative est la projection dans le futur d'une relation observée entre les éléments de fait ou de droit des décisions de justice passées, dans le but d'évaluer et d'optimiser les aspects quantitatifs et qualitatifs des futurs judiciaires possibles.

**Nous nous appuyons sur ce qui a été fait pour déduire ce qui sera fait plus tard.**

En réalité, le fait de savoir ce qui va se passer est bien moins utile que de savoir comment on va l'optimiser. Nous ne nous contentons pas de donner un chiffre, mais nous guidons également dans le choix de la meilleure stratégie. L'outil est notamment capable d'identifier les arguments qui ont le plus d'influence sur la prise de décision juridictionnelle et comment cela va impacter statistiquement la décision. Par exemple si vous allez plaider devant la Cour d'appel de Lyon et que vous ne connaissez pas bien cette juridiction, l'outil va vous permettre de dire que cette Cour d'appel est très peu sensible au harcèlement moral mais très sensible à d'autres arguments.

La justice prédictive

est la projection dans

le futur d'une relation

observée entre les

éléments de fait ou de droit

des décisions de justice

passées, dans le but

d'évaluer et d'optimiser les

aspects quantitatifs et

qualitatifs des futurs

judiciaires possibles

FACIAL RECOGNITION  
STATUS: FAILED  
NO FACES DETECTED  
TRY AGAIN



DATABASE COUNT: 6.391.554.897  
MATCHING RESULTS: 0



2,5 à 3,5 millions, c'est le  
nombre de décisions de  
justice qui seront  
consultables chaque année

## Existe-t-il une probabilité d'erreur dans les prédictions d'issues des litiges rendues ?

★ **Le Machine Learning est par définition faillible.** Un outil comme celui développé par Predictice a donc une marge d'erreur théorique infranchissable. Sur la partie mathématique, l'algorithme n'a pas de marge d'erreur : lorsqu'on lui demande de faire une médiane sur les indemnités accordées dans 12 000 décisions, le calcul est exact. L'outil peut en revanche faire des erreurs dans la compréhension du langage naturel. Il peut mal comprendre une triple négation ou une forme d'ironie dans un argument exposé. **L'outil est à 91% de compréhension du langage naturel juridique**, ce qui n'est pas très loin du maximum théorique.

C'est marginal mais cela doit pousser le professionnel à maintenir une forme de vigilance. Les analyses chiffrées peuvent être questionnées, ce qui devrait permettre de ne pas trop subir l'effet uniformisateur.

## Quelles sont les données exploitées ?

★ Globalement toutes données confondues, nous exploitons **une base de 15 millions de documents avec un flux attendu d'environ 3 millions de décisions de justice supplémentaire par an.**

Les données sont issues de trois sources :

**1 L'Open Data** - Un Big Bang judiciaire est en train de se produire, notamment avec les articles 20 à 21 de la loi pour une République numérique, votée en octobre 2016, qui obligent à mettre en ligne gratuitement les décisions de justice. Nous allons passer d'un

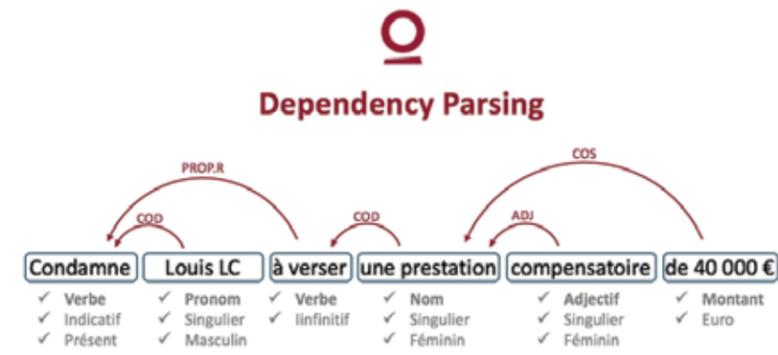
système avec actuellement 2,5 millions décisions de justice consultables, pour l'ensemble de l'histoire judiciaire, à un système avec 2,5 à 3,5 millions de décisions chaque année. Une personne qui commence sa carrière juridique aura très rapidement **20 à 30 millions de décisions potentiellement consultables**. Il y a un véritable changement de paradigme avec l'explosion des données et le développement des technologies pour les traiter.

**2 Les fonds privés de certains éditeurs** et notamment les **éditeurs juridiques** avec qui nous avons noué des partenariats forts à la fois technologiques et commerciaux. Par exemple, Wolters Kluwer nous aide à améliorer l'outil en intégrant des thésaurus, des jeux de données colossaux.

**3** La troisième source de données, ce sont **nos clients** et notamment **les plus grands cabinets ou entreprises qui ont fait l'effort de collecter des données et de structurer leurs bases**. Certains possèdent plusieurs centaines de milliers de documents. Ces partenariats demandent une certaine vigilance sur les conditions de sécurité. Nous avons, pour cette raison, décidé de stocker toutes nos données en France, avec un hébergeur français, pour garantir une réversibilité parfaite.

## Quels sont les fondements techniques de Predictice ?

★ Une des briques applicatives fondamentales de l'outil est ce qu'on appelle un **Dependency Parsing**. Ce qui est complexe dans le langage juridique c'est qu'il y a parfois plusieurs pages avec une ponctuation très particulière, voire inexistante. Il faut donc entraîner l'algorithme afin qu'il soit capable de comprendre le langage juridique tel qu'il est employé par les juridictions dans les décisions de justice. **Un dependency parser relie les mots en fonction de leurs relations.** Il vient extraire la force grammaticale entre les mots et phrases, en les jumelant à un thésaurus juridique pour pouvoir jouer avec les sens, les homonymes, les antonymes etc.



Dans cet exemple, l'algorithme est capable de comprendre qu'il y a eu un montant de 40 000 euros versé au titre de prestation compensatoire par M. X, jugement qui a été prononcé par la Cour d'Appel.

## Les algorithmes vont-ils uberiser la fonction d'avocat ou de juriste ?

★ Je n'aime pas le flou de ce néologisme. En outre la vision française « d'uberiser » semble impliquer un remplacement dans la douleur. Je serais plus enclin à accepter une autre racine, celle de l'Uber-

man dont le sommeil polyphasique permet de réduire incroyablement son temps de repos. Cette piste conduit au concept de l'homme augmenté, et dans ce cas-là, j'espère que tous les juristes auront accès à des informations qui étaient cachées dans la masse de données.

Si c'est uberiser dans le sens de combiner l'homme et la machine afin de mieux conseiller les clients et d'être sûr de ne pas rater une information cruciale, alors oui, le monde du droit doit être uberisé.

**Si uberiser signifie créer un système bis qui va remplacer l'homme par la machine dans la violence, alors ce n'est pas la vision que nous avons de l'évolution du secteur juridique.**

un enjeu qui nous tient à cœur et qui pose de nombreuses questions notamment sur les questions de ré-identification des personnes citées dans les données publiques.

Il y a également un vrai enjeu autour de l'effet performatif, de la pérennisation d'un état de fait qui est indésirable. Aux Etats Unis, il existe un algorithme qui permet d'avoir des recommandations de peine pour éviter la récidive. L'algorithme s'appuie sur des données antérieures et ce qui se passe dans les faits c'est qu'une personne de couleur va être plus souvent condamnée. Quand on touche à des sujets régaliens, il faut absolument traiter ces enjeux éthiques.

**Le sujet de la fracture numérique est également préoccupant** : le déploiement des outils numériques performant, des compétences qui vont avec et la récolte efficace des données doit se faire sur l'ensemble du territoire. Le numérique doit être une chance pour tous.

## ★ À PROPOS DU DEPENDENCY PARSING

L'analyse par dépendance est la tâche de reconnaître une phrase et de lui attribuer une structure syntaxique. La structure syntaxique la plus utilisée est l'arbre d'analyse qui peut être généré à l'aide de certains algorithmes d'analyse. Ces analyses sont utiles dans diverses applications comme la vérification de grammaire ou, plus important encore, elle joue un rôle essentiel dans l'analyse sémantique.

## Quels futurs enjeux et applications du Big Data et de l'Intelligence Artificielle dans le droit ?

★ Un des enjeux cruciaux est celui éthique. Nous avons d'ailleurs contribué à la **création du Comité éthique et scientifique de la justice prédictive**, composé d'un panel paritaire de professionnels, à la fois universitaires et avocats. **Ce comité travaille notamment sur des recommandations éthiques et sur le déploiement de l'outil dans le domaine du droit pénal.** C'est





★ Thomas  
**LEFEBVRE**

Directeur scientifique

MeilleursAgents.com



**Thomas Lefebvre** est titulaire d'un Master en mathématiques appliquées obtenu à l'Université de Besançon. Il a ensuite obtenu un Doctorat à l'Université Paris Dauphine en conciliant ses études et un emploi chez MeilleursAgents.com. Recruté comme Analyste immobilier doctorant, il est devenu en 2015 responsable de la R&D, puis Directeur scientifique de MeilleursAgents.com après la soutenance de sa thèse sur l'économie de l'immobilier.

Très attentif aux liens qui unissent le monde de l'entreprise et celui de la recherche, Thomas est aujourd'hui Chercheur associé à Paris-Dauphine. Il enseigne parallèlement la finance et l'économie de l'immobilier aux étudiants en Master de cette même Université.

# Big data allié ou ennemi de l'agent immobilier ?

En quoi le **Big Data** a-t-il **impacté** le secteur de **l'immobilier** et **Meilleurs-Agents.com**

★ La culture de la donnée aux États-Unis est beaucoup plus développée qu'en France. L'ensemble des transactions immobilières y sont accessibles et disponibles depuis plus de 30 ans. En France, l'ouverture des données immobilières est très récente.

MeilleursAgents a été créé, en 2008, par Sébastien de Lafond, Pascal Boulenger, Jordan Sanial et Julien Cheyssial, qui ont fait le constat que **le marché de l'immobilier était un marché opaque où les transactions immobilières reposaient sur un ensemble de connaissances et d'échanges essentiellement informels non structurés et peu quantifiés**. La masse des données sur l'immobilier augmente alors de manière exponentielle du fait de la migration progressive des données de l'ensemble de la profession vers le net. Cela a entraîné un véritable **foisonnement de la donnée immobilière**. Le lancement de Google Maps en février 2005 a également été un déclencheur pour le lancement de l'activité de MeilleursAgents, basé sur la cartographie des prix de l'immobilier.

MeilleursAgents a donc été créé avec pour objectif de fournir des informations pour aider les particuliers à prendre des décisions immobilières basées sur les éléments les plus rationnels et les plus objectifs possibles, par exemple être capable d'objectiver le prix au m<sup>2</sup> dans tel quartier de Paris.



**Meilleurs-Agents.com** n'est pas un acteur traditionnel de l'immobilier : quel en est le modèle ?

★ MeilleursAgents est une plateforme d'intermédiation dont le business est de mettre en relation des particuliers, qui viennent pour rechercher des informations sur le marché, le prix, et des agences. C'est ce flux de porteurs de projets immobiliers, acheteurs et vendeurs, que nous mettons en relation avec des agences immobilières qui payent un abonnement pour être affichées sur notre site web. Aujourd'hui, plus de 7000 agences sont visibles sur le site de MeilleursAgents, ce qui représente plus d'1/4 des agences du territoire français.

Nous proposons aux agences d'être visibles sur l'ensemble de notre site et de valoriser leur savoir-faire au travers

de e-vitrines auprès de particuliers ayant un projet immobilier. Nos offres de visibilité sont de véritables outils marketing permettant aux agents immobiliers d'obtenir des mandats de vente. **La question première que va se poser un particulier quand il va vouloir vendre son bien c'est d'en connaître la valeur. Nous répondons à cette question grâce à notre data et à nos outils d'estimations.** Nous sourçons donc des prospects vendeurs avant tout le monde et les redirigeons vers les agences partenaires. En France nous voyons **50% des vendeurs** en avant-première. En Ile-de-France et sur Paris, cela représente plus de **80% des vendeurs**.

**La data est donc véritablement le socle de l'entreprise**, c'est grâce à la data qu'il y a un flux de visiteurs transformés et envoyés vers les agences.

# Quelles sont les données récoltées ? Parle-t-on véritablement de Big Data dans l'immobilier ?



Pour répondre à la question de la valeur d'un bien, il faut mettre à disposition des outils pour permettre d'obtenir un prix hyper fiable et précis. Ce n'est pas un exercice simple car le volume de données immobilières est limité, il s'agit d'un univers fini. On recense environ 800 000 transactions par an dans l'immobilier ancien, ce qui ne peut pas forcément être qualifié de Big Data. Certains acteurs, du e-commerce par exemple, peuvent récolter les données de millions de transactions et des millions de flux de visiteurs par seconde. Dans l'immobilier, on parle plus de Smart Data, parce que l'actif immobilier et les données qui en découlent sont très hétérogènes. Il n'y a pas deux biens immobiliers qui se ressemblent.

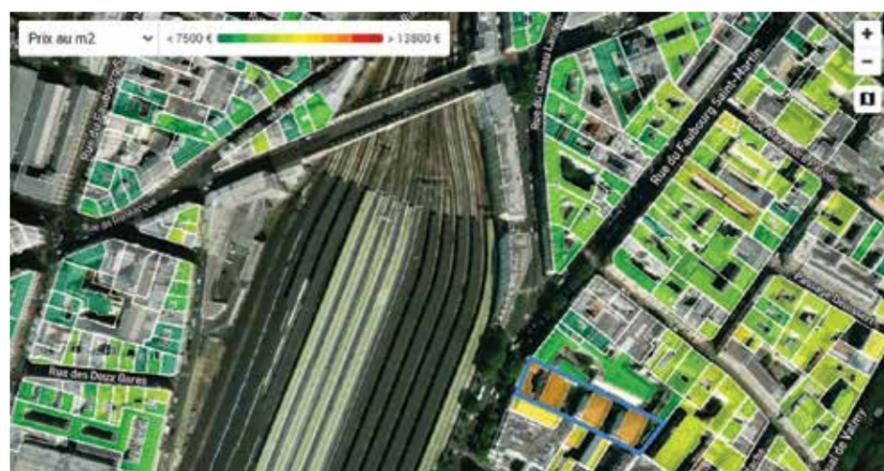
**1** La base BIEN en île de France est une de nos premières sources de données. Elle recense l'ensemble des transactions immobilières enregistrées par les notaires. Le site Immo Notaires permet à un usager d'accéder, après rétribution, à ces informations détaillées de vente de biens immobiliers. A la différence de MeilleursAgents, le site des notaires fournit uniquement des informations sur les transactions réalisées. Un particulier qui a un appartement de 100m2, situé au 4ème étage, va pouvoir obtenir de cette application des informations sur la dernière transaction réalisée dans l'immeuble, par exemple un 30m2 situé au rez-de-chaussée et vendu il y a deux ans. Mais comment utiliser cette information pour estimer le prix de mon bien, deux ans plus tard, alors que mon bien n'a rien à voir avec celui déjà vendu ?

Notre valeur ajoutée est d'exploiter ce flux de données et de casser cette hétérogénéité structurelle, due au marché, et la rareté des informations sur les transactions (800 000 transactions sur 30 Millions d'unités d'habitations). On estime le stock de biens à vendre par le flux. Mettre un prix sur un appartement qui n'a pas été vendu, parfois depuis plus de 30 ans, est très complexe. Il faut donc faire parler les données mises à disposition. Au sein d'un même immeuble par exemple, les prix peuvent varier du simple au double entre les étages.

**Communiqué réalisé par MeilleursAgents sur l'impact des ascenseurs et des étages sur les prix**



À Paris par exemple, MeilleursAgents cherche donc à mettre un prix sur chaque immeuble parisien. La complexité vient de l'hétérogénéité des biens d'un immeuble à l'autre, vous pouvez passer d'un immeuble haussmannien à un immeuble années 70, donc les prestations et les prix sont différents. Il faut réussir à modéliser cette réalité de marché. Par exemple, le 138 rue du Faubourg Saint Martin est situé proche de la Gare de l'Est et subit par conséquent les mêmes nuisances que les autres immeubles avoisinants et pourtant on distingue trois immeubles sur la carte derrière la façade du 138, qui bénéficient d'une meilleure vue et de moins de nuisances que les immeubles avoisinants et de meilleures prestations.



L'objet de la data, c'est donc de pouvoir offrir cette granularité de prix et une estimation la plus fiable possible. Un prix est estimé sur chaque immeuble sur l'ensemble de l'île de France et dans les 20 plus grandes villes de France. Pour les autres villes, les estimations sont faites à l'échelle d'un quartier ou d'une commune. Le modèle de MeilleursAgents dépend véritablement du contexte informationnel et des données dont ils disposent dans chaque ville.



**2** Nous utilisons les données des transactions, remontées par les agences partenaires du site, soit 7000 agences sur la plateforme qui remontent les biens vendus. C'est un système particulièrement vertueux car les agences se créent des e-vitrines sur notre site. Elles sont visibles dans l'annuaire, on peut consulter leurs annonces, les avis clients et la liste des biens vendus. Leur but est de montrer qu'elles sont actives sur leur secteur, de rassurer le consommateur et de montrer qu'elles sont pertinentes pour vendre un bien. Elles ont donc tout intérêt à remonter leurs biens vendus.

**3** Les annonces remontées par les agences sont également utilisées, il ne s'agit pas dans ce cas d'une donnée de transaction mais d'un premier signal, analysé pour en déduire un prix de transaction. Les annonces permettent d'estimer le prix des biens. Par exemple, sur un marché hyper dynamique, le délai de commercialisation sera court et la décote faible. Sur un marché baissier en revanche, comme de 2012 à 2014, la décote est plus importante et le marché moins dynamique. Le modèle permet d'anticiper l'augmentation d'un bien à partir de l'observation du marché et donc d'estimer le prix mais également le délai de commercialisation. Les annonces publiées sur notre site le sont sans adresse naturellement en raison de la concurrence entre les agences. Nous avons donc développé des modèles de géolocalisation comme pourrait le faire un humain. Le modèle va lire une annonce, et grâce à un référentiel géographique va être capable de lire les différents mots clefs, les différents points géographiques et va trianguler la localisation de l'annonce sur la carte.

**4** Enfin nous utilisons les données publiques de l'INSEE, par exemple sur le revenu, le chômage, l'âge des habitants. La relation entre le revenu et le prix de l'immobilier est très pertinente. Ces données seront utilisées aussi pour combler le manque d'informations. Les données de l'INSEE ont également été utilisées pour modéliser le marché de la location grâce à des algorithmes de diffusion. Nous avons constaté, en utilisant 15 critères sur la base de données de l'INSEE, qu'il y avait des villes similaires au sens des critères socio-démographiques des habitants : mêmes âges, mêmes revenus, même taux d'employabilité et de chômage. Nous avons donc utilisé des algorithmes de Machine Learning pour clusteriser les villes et appliquer les données de loyers à des villes sur lesquelles nous n'avions pas d'informations. Des algorithmes de diffusion et de clustering de villes combinés ont permis de couvrir environ 90% du territoire français. Il s'agit là de techniques assimilées Big Data.

Cela représente un certain volume de données hétérogènes sur lequel nous utilisons des techniques Big Data et de Machine Learning. Dès le départ MeilleursAgents a fait le pari de la recherche et a énormément investi en R&D. Nos modèles sont issus de nos propres travaux de R&D plutôt que d'utiliser des briques existantes.

## Une culture de la recherche à l'origine de l'innovation



Je suis arrivé dans l'entreprise il y a 7 ans en tant que doctorant et analyste immobilier. Depuis la soutenance de ma thèse, j'occupe le poste de Directeur Scientifique et pilote l'ensemble des activités de R&D de MeilleursAgents. J'embauche de jeunes doctorants ou travaille avec des chercheurs pour nous permettre de continuer à être moteur en R&D dans le secteur de l'immobilier en France. Il y a une vraie complémentarité avec l'équipe data qui va travailler sur les flux de données, leur qualité mais surtout sur les outils permettant de tester nos modèles ainsi que leur industrialisation. Notre équipe scientifique travaille, elle, sur les modèles mathématiques et économiques permettant de comprendre et appréhender le marché immobilier. Les équipes travaillent naturellement

ensemble pour créer des cartes, des indices et l'ensemble des outils innovants de MeilleursAgents. En face des données, la partie scientifique et recherche est primordiale. **Nous avons noué plusieurs partenariats de recherche** avec l'université Paris Dauphine, le laboratoire d'économie de Cergy Pontoise, la Toulouse School of Economics, ou encore le CNRS. Et nous faisons appel aux thèses CIFRE, ce sont des dispositifs mis en place par l'Etat pour promouvoir l'embauche de doctorants dans les entreprises.

**On constate de plus en plus, avec l'avènement du Big Data notamment, de conventions industrielles de formation financées par la recherche d'entreprise.** Il s'agit de partenariats entre l'entreprise, le laboratoire de recherche et le doctorant, pour travailler sur un sujet de recherche qui pourra être appliqué de manière opérationnelle et business. C'est également économique pour les entreprises qui embauchent les doctorants car elles touchent 15 000 euros par an, plus des réductions d'impôts pour la recherche. Cela permet d'avoir des salariés spécialisés (Bac + 8) en lien avec les laboratoires, véritable vivier d'innovation pour les entreprises. Pour les laboratoires cela permet également d'avoir des liens avec les entreprises pour récupérer des données et créer des sujets de recherche.

## Quels sont les fondements de vos modèles algorithmiques ?

★ Il s'agit de modèles mathématiques et économétriques issus de la littérature sur la finance et l'économie du logement :

- **Le modèle ou l'approche hédonique** qui consiste à décomposer le prix d'un bien en valorisant ses caractéristiques (étage, nombre de salles de bains, qualité de l'environnement urbain, terrasse...)

- **Le modèle des ventes répétées** qui prend pour base, sur un intervalle plus ou moins long, l'ensemble des biens ayant fait l'objet d'au moins deux ventes. On mesure alors pour chacune des paires ainsi répertoriées, le taux de croissance du prix entre le moment de la première vente et celui de la deuxième. Une méthode statistique adaptée permet d'agrèger l'ensemble des taux observés et d'en déduire un indice d'évolution des prix.

- **Le modèle de statistiques spatiales** : la géostatistique est appliquée aux données immobilières pour déterminer une estimation non biaisée du prix, donner une meilleure prédiction et déterminer les segmentations du marché immobilier. L'économétrie spatiale est utilisée pour déterminer le degré de corrélation spatiale, estimer les indices ou pour évaluer la variation de prix liée à une externalité.

En agriculture par exemple des chercheurs travaillent avec des prélèvements qui permettent de déterminer la fertilité du sol en certains points, les statistiques spatiales permettent de prédire la fertilité de l'ensemble du terrain. C'est la même chose avec les données immobilières, **à partir de transactions en certains points, nous déterminons l'influence sur un bien pour estimer un autre bien de manière intelligente.** Le modèle doit être capable d'appréhender les vecteurs de diffusion s'appliquant à un bien. Par exemple, un bien situé à Nice sur la promenade des anglais avec vue mer ou avec vue sur cour, et donc sur la gare, n'a pas la même valeur. Le modèle doit être capable d'appréhender ces différentes caractéristiques.

Aujourd'hui, une 20aine de caractéristiques suffit, parmi lesquelles la présence d'un ascenseur, l'étage, la surface, le nombre de pièces, la présence d'un balcon, les DPE etc... En revanche, nous pouvons encore avoir une marge de progression sur l'évaluation de l'état du bien qui est subjectif et déclaré par le particulier. Les 20 critères objectifs sont observés dans les bases de données. Ces caractéristiques sont estimées pour faire varier les critères de scoring. On sait par exemple que l'impact du rez-de-chaussée est estimé à 30% de décote. Nos modèles de scoring sont également recalibrés tous les ans.

Les données de la Smart City pourraient à termes également être incluses dans nos modèles, cela fait partie des vecteurs de diffusion. Par exemple les données de pollution ne sont pas encore prises en compte, il y a pourtant une décote naturelle si vous habitez en étage peu élevé sur la rue la Fayette à Paris.

## Quel est l'impact, selon vous, de la data sur les métiers de l'immobilier ? Les agents immobiliers sont-ils voués à disparaître ?

★ Depuis le départ, les agents immobiliers sont au cœur de notre modèle. MeilleursAgents n'envisage pas la suppression des agents. Aux vues des montants engagés et le fait qu'un particulier n'achète qu'un ou deux biens dans sa vie, il y a un vrai besoin de conseil et une recherche de réassurance dans l'estimation et la négociation du bien.

**La data ne va pas, selon nous, uberiser le marché. Elle doit au contraire permettre aux agents de se concentrer sur leur cœur de métier, sur la réassurance, la fixation et la négociation du prix et l'accompagnement pendant la transaction.** La data donne une indication du prix mais l'agent immobilier est indispensable pour plus de précision dans l'estimation.

Il doit s'aider de la data pour augmenter ses performances et la recherche de mandats. Si aujourd'hui plus de 90% des français débutent un projet immobilier sur internet et si 1 vendeur sur 2 réalise une estimation sur MeilleursAgents, ils sont toujours une majorité à conclure avec un agent immobilier.

Le vendeur est donc mieux informé sur les prix mais aussi sur les agences, particulièrement celles visibles sur des logiques de recherches géolocalisées et qui jouent le jeu des avis clients. C'est important pour les agents immobiliers de bien appréhender cette complémentarité fondamentale.

## Quels sont les prochains défis du Big Data pour les acteurs de l'immobilier ? Va-t-on vers du prédictif dans les estimations immobilières ?

★ Il n'y a pas encore de prédictif dans l'immobilier. En revanche, nous avons construit l'ITI, un indicateur de tension immobilière qui va permettre de comprendre et d'anticiper les évolutions du marché. Nous retraçons les visites, sur notre site web, des personnes qui font des demandes d'estimations, à peu près 150 000 par mois. Quand ils font une demande d'estimation, ils se déclarent comme acheteur, vendeur ou simple particulier qui souhaite une estimation de la valeur

de son bien. En retraçant ce flux, on peut par exemple dire que sur cette zone on recense 4 acheteurs pour un bien en vente. En suivant ces informations dans le temps ainsi que l'évolution des taux par exemple en France, nous sommes capables d'anticiper les évolutions du marché. Le marché parisien entre 2009 et 2011 a par exemple subi une hausse de 40% et s'est ensuite retourné en juin 2011 : nous aurions été capables d'anticiper ce retournement avec cet indicateur.

Le marché de l'immobilier n'est pas forcément le plus avancé en matière de data et d'analyse, cela vient de la culture française car historiquement l'immobilier est un actif sans risque qui n'a pas forcément besoin d'être optimisé. La mentalité a changé mais c'est également la technologie qui a donné un coup de pouce (Google Technologies, Google Maps etc...) et le consommateur qui est de plus en plus en demande.

**Les futurs enjeux résident dans l'ouverture des données, la compréhension du marché mais également la transparence. L'accessibilité de l'information pour les particuliers est une vraie nécessité, le marketing est un vrai challenge pour les professionnels.**

La data ne va pas, selon

nous, uberiser le marché.

Elle doit au contraire

permettre aux agents de

se concentrer sur

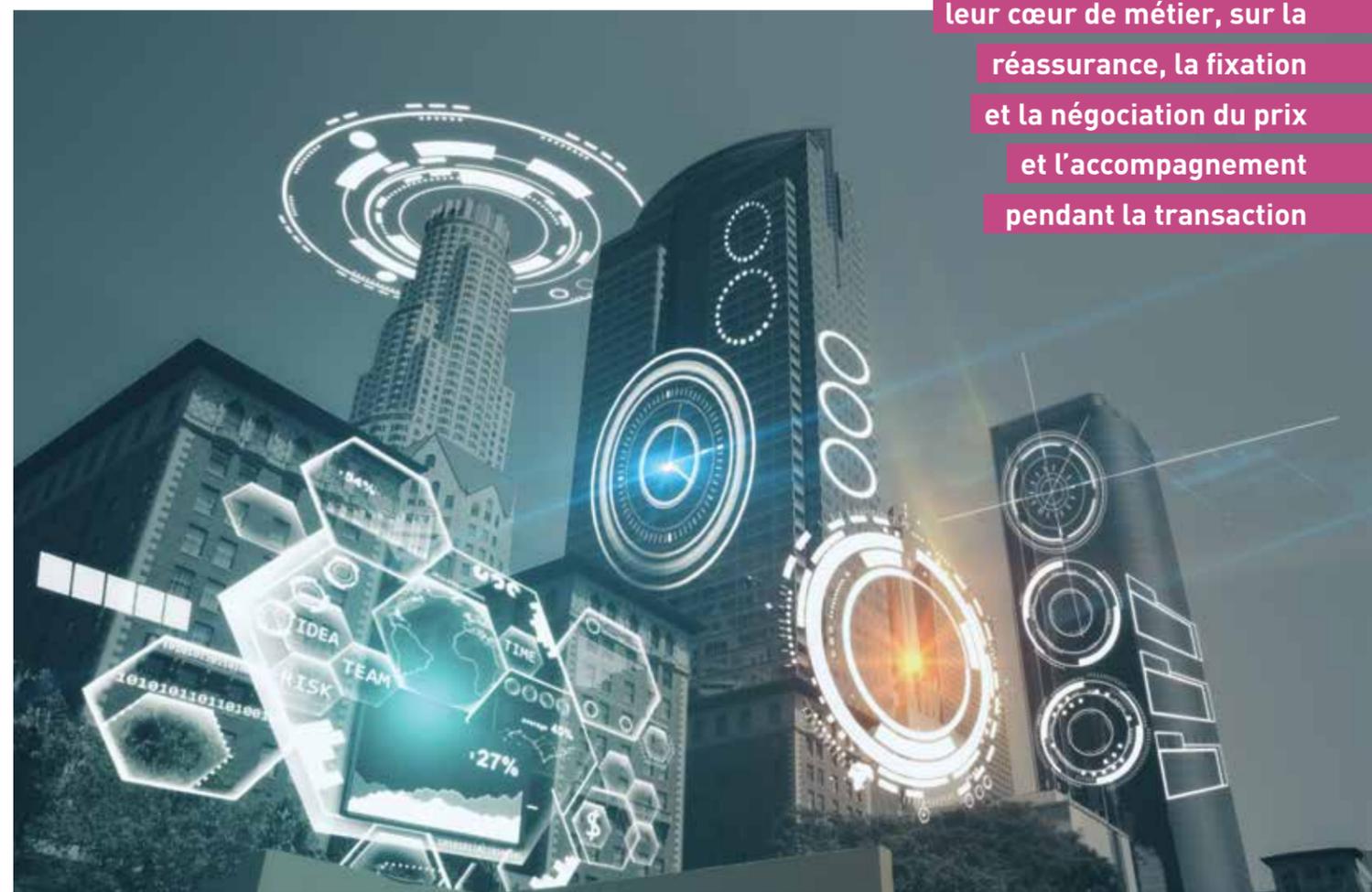
leur cœur de métier, sur la

réassurance, la fixation

et la négociation du prix

et l'accompagnement

pendant la transaction





☆ Edouard  
**LE GOFF**

Fondateur & Président

Kinaxia



Edouard Le Goff a fondé Kinaxia, "Kindunos" signifiant en grec "danger" et "Axia" signifiant "valeur", en 2010, sur l'intuition que l'aide à la décision allait se développer rapidement avec le Big Data, l'Open Data et l'Intelligence Artificielle. Kinaxia a donc été lancée dans le but de créer de la valeur à partir de la connaissance du risque. Edouard Le Goff a d'abord fondé preventimmo.fr, spécialisé dans la fourniture dématérialisée de documents d'urbanisme réglementaire. Fort de cette expertise, il a lancé le 15 mai 2017 CityScan.fr, une nouvelle solution web de géo-décision immobilière qui permet d'évaluer n'importe quelle adresse en France métropolitaine, en identifiant ses points forts et ses points faibles, grâce à plus de 70 indicateurs facilement compréhensibles.

# Big Data et géo-décision immobilière

Quel **constat** faites-vous de la **digitalisation** du **marché de l'immobilier** ? La **data** a-t-elle un **impact** aussi **prégnant** que dans d'autres secteurs d'activité ?

☆ Le secteur a pris en compte le digital mais le numérique et les technologies n'ont pas encore véritablement transformé le secteur de l'immobilier qui a longtemps été très protectionniste. Les métiers et les usages ont freiné ces évolutions. La data a bouleversé presque toutes les industries mais l'immobilier reste en marge, car les technologies ne sont pas encore adaptées au secteur d'activité.

Jusqu'ici, l'immobilier sur le web avait repris les méthodes d'antan avec essentiellement des moteurs de recherche et des portails qui dominaient par le nombre plus conséquent d'annonces publiées.

La véritable transition n'a pas encore eu lieu mais le développement de l'Open Data notamment dans le contexte de la Smart City, du Big Data et de l'IA va changer radicalement la manière de faire de l'immobilier. La donnée va permettre de contextualiser les biens, d'enrichir l'expérience, d'étayer les informations sur le bien et autour du bien. La valeur du Big Data dans l'immobilier résidera plus dans l'information créée que

dans l'information fournie. Il s'agira de récolter de l'information afin d'affiner la connaissance de l'utilisateur et de matcher ces informations avec la connaissance d'un produit. L'Intelligence Artificielle au sens de l'auto-apprentissage et du Machine Learning sera indispensable pour améliorer la connaissance et la pertinence du matching.

Le marché de l'immobilier, jusqu'ici plutôt statique et tiré par l'offre, va muter d'ici 5 à 10 ans vers un marché dynamique tiré par une demande en attente d'accompagnement. CityScan s'inscrit dans ces évolutions et souhaite être un des acteurs de ces mutations. Nous sommes en contact avec de nombreux réseaux historiques de l'immobilier qui partagent cette vision et anticipent également cette dynamique qui leur offrent de nouvelles opportunités.

Quant aux métiers, je pense que l'agent immobilier aura toujours un rôle très important mais pas le même qu'aujourd'hui. La relation humaine est importante car l'acte d'achat est parfois celui d'une vie, il s'agit d'un choix souvent anxiogène. L'accompagnement d'un agent immobilier est primordial, en revanche le secteur doit se professionnaliser et donner des gages de confiance. Le secteur se doit d'adopter des outils pour se concentrer sur l'accompagnement du client et l'intermédiation. Il doit se doter des méthodes et technologies digitales et data pour trouver grâce aux yeux du grand public.

**Parle-t-on** véritablement de **Big Data** dans l'**immobilier** ?

☆ L'immobilier reste un secteur et métier sensible dans lequel les informations ne sont pas partagées de manière naturelle, ceux qui les détiennent les protègent. Or c'est en libérant la donnée stratégique qu'un cercle vertueux va se créer autour des données immobilières. Il y a par ailleurs un véritable problème de définition du Big Data, je penche pour une approche tournée sur l'utilité de la donnée pour les utilisateurs. On peut parler de Big Data dans le cas de services et outils qui permettent de réutiliser de la donnée utilisateurs pour permettre de créer de nouvelles données pour améliorer les services aux utilisateurs. Dans l'immobilier, on parle plutôt de Smart Data : de deux données médiocres peut naître une information pertinente si elles sont croisées. Le Big Data il faut le créer !

La donnée va permettre de contextualiser les biens, d'enrichir l'expérience, d'étayer les informations sur le bien et autour du bien

## Comment est né CityScan ?

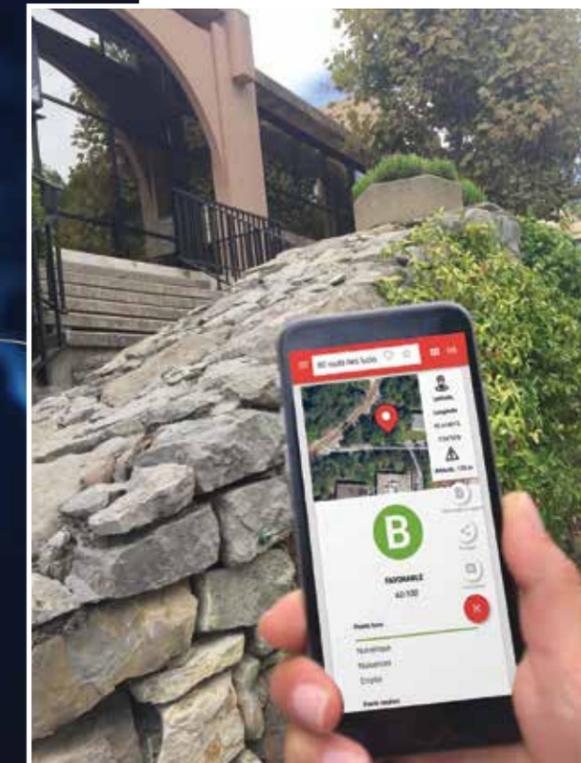
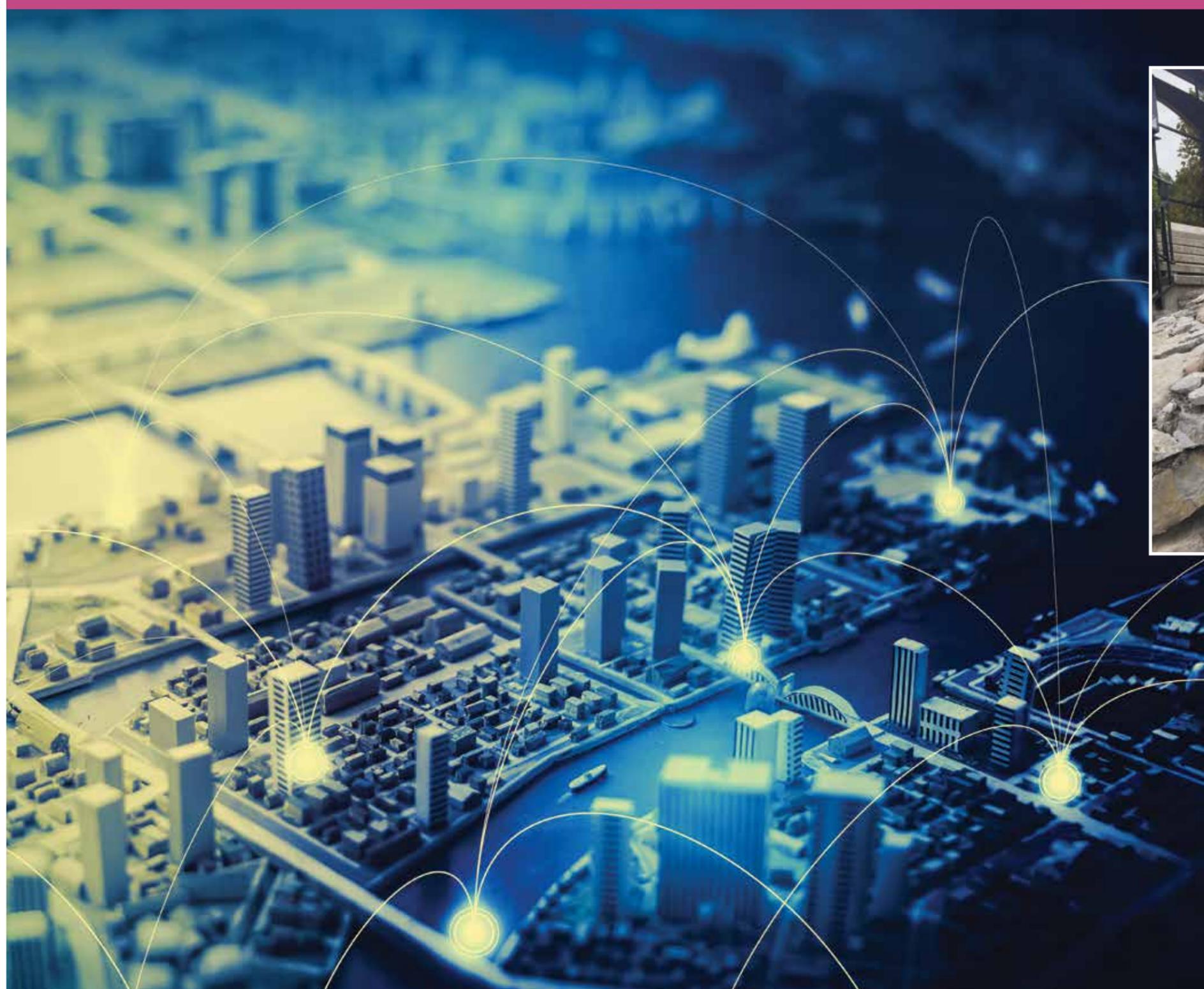
★ Au lancement de Kinaxia, nous avons tout d'abord souhaité répondre à une réalité d'entreprise et de marché. En effet, la data, il y a 7 ans, n'était pas encore disponible et le marché immobilier n'était pas prêt à accueillir une solution comme CityScan. Nous avons donc décidé de nous appuyer sur l'obligation réglementaire de fournir des documents faisant état des risques Naturels, Miniers et Technologiques (ERNMT), pour créer preventimmo.fr. Ce type de documents était à l'époque particulièrement erroné et présentait des problèmes de qualité, de mise à jour de la donnée, de localisation etc.

Les documents étaient établis de manière artisanale. L'immobilier pouvait alors être considéré comme le parent pauvre du diagnostic. **Nos premiers développements se sont donc tournés vers l'automatisation de la production de documents d'urbanisme, la création de bases de données propres et de couches de géolocalisation.** Notre solution, s'adressant aux diagnostiqueurs immobiliers, permet d'établir les états des risques en moins de 2 minutes en ligne. Aujourd'hui, presque 25% des ERNMT (sur presque 3M /an) sont faits par le biais de preventimmo.fr.

Il s'agissait d'un premier pas vers des prestations plus larges dans le domaine de l'urbanisme réglementaire. Nous avons ainsi développé un service permettant d'industrialiser la production d'informations s'adressant aux notaires. Ils ont, en effet, une obligation d'informer les futurs acquéreurs des différents droits à construire et limitation au droit de mutation tels que des soumissions à des servitudes ou encore des droits de préemption. Ces informations étaient également produites de manière artisanale avec une qualité, un délai et des prix extrêmement variables. Enfin, nous avons également automatisé les documents et la production d'information concernant l'État des risques de pollution des sols (ERPS).

**Nous fournissons donc à ce jour plus de 60 000 documents d'urbanismes par mois et comptons entre 3000 et 4000 clients uniques chaque jour.**

Le succès de preventimmo.fr nous a permis de réinvestir dans une politique de Recherche & Développement plus poussée, avec toujours pour ambition globale de fournir de l'information environnementale à destination des acteurs de la chaîne de la transaction immobilière. Nous avons travaillé en équipe restreinte sur une nou-



velle vision de l'immobilier sur internet qui a mené au lancement de CityScan, testé en beta notamment sur les clients de preventimmo.fr entre Janvier et Mai 2017, et accessible à tous depuis le 15 mai.

## Qu'est-ce que CityScan ?

★ CityScan a été développé dans le but de permettre à un utilisateur d'évaluer gratuitement l'adresse d'un bien via un moteur de recherche. L'ensemble du faisceau décisionnel immobilier est mis à disposition sur une adresse à savoir :

- L'accessibilité aux transports, écoles, commodités, au numérique...

- La dynamique économique du quartier
- L'exposition aux nuisances
- L'exposition aux risques d'inondation, de mouvement de terrains, de tout autre risque naturel ou technologique...

Nous devons donc traiter des problématiques complexes en termes scientifiques pour rendre des informations complexes accessibles à tous et permettre à tout un chacun de pouvoir décider de leurs investissements. Nous utilisons ainsi des indicateurs avec des codes couleur et des notations de A à E, identiques aux barèmes alimentaires ou de DPE. C'est la représentation la plus commune et la plus admise.

## Quel est votre business model ?

★ Nous nous adressons au grand public en premier lieu. Il s'agit de notre client final à qui le service est dédié. **Nous dénombrons déjà plus de 30 000 inscrits et plus de 70 000 adresses testées sur le site deux mois après l'ouverture du service.** Cependant nous développons de nouveaux services destinés au BtoB sur un modèle

d'abonnement Premium pour les professionnels. Nous avons packagé plusieurs fonctionnalités pour permettre au professionnel d'avoir un outil d'accompagnement objectif dans l'ensemble de son processus commercial, afin de lui permettre d'identifier le profil de son acquéreur, ses besoins, ses contraintes. La personnalisation du service va permettre d'évaluer la pertinence d'une adresse en fonction des besoins de son prospect. L'évaluation à terme ne prendra en compte que ce qui est important pour le client final.

Le professionnel pourra appliquer ce profil à l'ensemble de son portefeuille immobilier et faire émerger de manière objective les biens en adéquation avec les besoins et contraintes de son client. Il s'agit d'un élément objectif d'accompagnement, de justification, de confiance pour le professionnel de l'immobilier, souvent remis en cause.

CityScan va se positionner en **tiers de confiance pour accompagner l'agent et le client dans la recherche de l'emplacement immobilier qui lui correspond, de son adresse idéale.**

## Quelles sont les **données récoltées** ?

★ **Les données sont le nerf de la guerre mais également ce qu'il y a de moins important. La valeur réside dans le traitement et la sémantisation qui est faite des différentes données. Les données brutes n'ont de valeur que si elles sont croisées et enrichies avec d'autres données.** Nous travaillons sur des algorithmes capables de modéliser chacun des indicateurs à partir de datas brutes.

Nous nous sommes d'abord appuyés sur notre savoir-faire, notre base de données et notre expertise sur l'urbanisme numérique, croisées avec des données d'Open Data public et privé au niveau national et mondial. Mais cela n'était naturellement pas suffisant, les indicateurs sont variés et nécessitent des expertises et savoirs scientifiques spécifiques.

Certains partenariats sont commerciaux et d'autres relèvent de projets de R&D collaboratifs où chacun investit notamment par le biais de thèses ou de post doctorants.

Nous nous appuyons donc sur des partenariats de R&D secteur par secteur et indicateur par indicateur, afin de bénéficier de l'expertise de partenaires référents sur leur expertise. Concernant par exemple les nuisances olfactives, nous avons mis en place un plan de R&D sur 5 ans avec l'université de Nice, permettant d'améliorer sa cartographie. Nous prenons ainsi en compte les typologies et la propagation de ces nuisances, les obstacles à celles-ci etc... Concernant la pollution de l'air, nous nous appuyons sur un programme européen afin d'utiliser les données satellites pour enrichir nos indicateurs...

Enfin, nous discutons actuellement sur la mise en place de partenariats privés avec des majors dans le domaine des assurances, de la construction, des télécoms, de l'énergie et ou du numérique.

Quelles sont les **technologies Big Data utilisées** ?

## ★ Il y a actuellement un vrai mouvement dans le Big Data autour du NoSQL. Nous utilisons Cassandra pour ce qui relève du back-end et les technologies Google pour ce qui relève du front-end.

Nos véritables défis résident dans la nécessité d'effectuer les calculs d'indicateurs en temps réel, la réactivité attendue, le volume de nos bases de données et le nombre croissant d'utilisateurs. NoSQL et Cassandra nous permettent d'atteindre la performance souhaitée vis-à-vis de ces contraintes.

Comment se positionne la **France** par rapport aux **États-Unis** ? Est-ce que vous avez pu vous inspirer d'**acteurs Outre-Atlantique** ?

## ★ Trulia ou encore Zillow représentent le graal en la matière. Les informations y sont accessibles, l'expérience utilisateur y est extrêmement forte. CityScan souhaite tendre vers ce type de services mais nous ne pouvons pas dupliquer ce type d'applications à l'identique. **Les États-Unis ont depuis toujours adopté une philosophie d'ouverture des données.** La libéralisation de la donnée est tellement forte que cela permet au secteur d'aboutir à des usages très innovants. **En France on doit recréer, contourner l'information pour créer des applications dans un modèle pas prévu pour cela.** De plus, la taille nationale du marché français n'est pas comparable à celle des États-Unis. L'écart en termes de services est donc particulièrement flagrant.

CityScan pourrait-il être **dupliqué** à **l'étranger** ou à **d'autres secteurs** d'activité ?

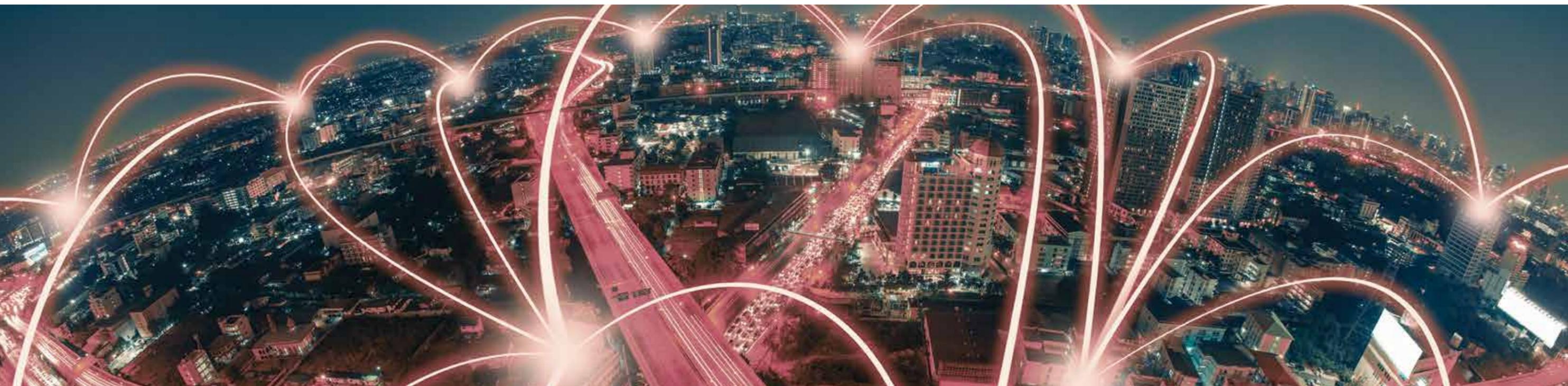
★ Nous avons choisi l'immobilier car le time to market se situe dans un futur très proche. De plus, nous connaissons bien le marché mais ces technologies et ce savoir-faire pourraient s'appliquer également au secteur du tourisme.

L'immobilier est un secteur particulièrement intéressant cependant il garde ses spécificités nationales. CityScan s'appliquera donc difficilement à d'autres territoires. Il n'existe d'ailleurs pas à ce jour de plateforme européenne dans le secteur de l'immobilier.

La transparence, la confiance et l'intelligence de ces services seront primordiales dans les futures applications françaises, afin de fournir de l'information pertinente dans les outils délivrés aux utilisateurs.

La transparence, la confiance et l'intelligence de ces services seront primordiales dans les futures applications françaises, afin de fournir de l'information pertinente dans les outils délivrés aux utilisateurs.

Les données brutes n'ont de valeur que si elles sont croisées et enrichies avec d'autres données





☆ Stéphan  
**EUTHINE**

Président

France eSport

COO

LDLC Event



Stéphan Euthine est responsable du sujet eSport pour le groupe LDLC. Il a travaillé sur le sujet pour le groupe pendant 9 ans avant la création en 2016 de l'entité dédiée LDLC Event. Il est également Président de l'Association loi 1901, France eSport, qui fédère les principaux acteurs du secteur.

# La data dans l'arène de l'esport

Pourquoi avoir créé une **filiale 100% esport** et quid de l'origine de **l'Association France eSport** ?

★ La réflexion et l'engouement pour le sujet ont débuté il y a une dizaine d'années. Nous avons réfléchi au sein du groupe LDLC à l'amélioration de la visibilité de notre marque à travers différents outils dans le domaine de l'esport. Nous avons débuté nos investissements par du sponsoring d'événements et de la visibilité média dans le but de tester la cohérence de ces dispositifs avec notre clientèle. L'activité s'est naturellement développée et les retombées ont dépassé nos attentes. Nous avons eu une **véritable prise de conscience** lorsqu'en 2014 la team LDLC de pro gamers a remporté la DreamHack Winter, l'équivalent de championnats du monde dans le domaine du sport. Notre équipe avait été éliminée en quart de finale mais a été repêchée suite à la disqualification de l'équipe adverse. Nous étions donc dans un scénario avec un **storytelling parfait, qui a généré plus de 500 millions de tweets**. Cela a généré plus de tweets que la sortie du trailer de Star Wars.

Nous avons réalisé le potentiel de l'activité sur les réseaux sociaux mais également pour les médias grands publics. Nous avons donc créé LDLC Event, spécialisé dans la communication, dans le but de pouvoir proposer à d'autres marques, endémiques ou non de l'esport, notre savoir-faire en matière de formation, conseil, merchandising, licensing et sponsoring. L'optique est naturellement d'offrir aux marques une plus grande visibilité à travers l'esport.

**La loi pour une République Numérique du 7 octobre 2016 a marqué la reconnaissance, par le gouvernement français, de l'esport et du métier de joueur dans le domaine du sport électronique, c'est à dire du jeux vidéo.** Cela signifie que l'État octroie une existence juridique à des jeunes qui pratiquent,

pour gagner leur vie, des jeux comme Call of Duty, Fifa, Counter-Strike ou encore League of Legends. Une discussion s'est alors engagée dans le secteur et un groupe de travail a été identifié sur le sujet du esport à l'initiative du gouvernement. **L'association France eSport a donc été créée dans le but de représenter le gouvernement sur ce sujet.** Au début 10 acteurs, membres fondateurs, ont été réunis autour de la table. Il aura fallu plus d'un an pour fédérer également les éditeurs de jeux, les promoteurs et joueurs et réunir ainsi l'ensemble de l'activité au sein de l'Association dont j'ai pris la Présidence.

Que représente **le marché** de l'esport en France ?

★ Concrètement « **l'esport désigne la pratique compétitive (affrontement en temps réel) sur internet ou en LAN d'un jeu vidéo, seul ou en équipe, par le biais d'un support électronique (ordinateur, console, smartphone, VR, tout support d'aujourd'hui ou de demain)** ». Il faut cependant définir des standards communs en matière d'esport afin de pouvoir le mesurer. Une question essentielle est par exemple de savoir si l'esport désigne uniquement les gamers professionnels ou également les influenceurs et joueurs amateurs ?

De plus l'esport n'est pas homogène. Il existe plusieurs jeux, plusieurs communautés et il est nécessaire de pouvoir descendre dans les chiffres pour segmenter les communautés et pouvoir mieux les expliquer.

**Depuis 2011, la pratique des jeux en ligne ou en réseaux a explosé, notamment avec le mode spectateur qui permet d'enregistrer une partie et de la streamer** (la diffuser en ligne). C'est ce format qui a donné une grande ampleur à l'activité. Twitch propose par exemple une plateforme de diffusion de parties. Les outils adaptés à ce format live

sont arrivés à maturité et offrent un tel niveau de visibilité que les médias grand public s'y sont intéressés. Ils ont notamment fait le constat que le jeune public ne pouvait plus être touché à travers les médias télévisés.

**Les jeux vidéo et l'esport se placent en tête de liste des médias les plus plébiscités par les Millenials.** Ce média est donc devenu naturel pour la plupart des marques. **Les 12-25 représentent plus de 50% de la cible eSport.** Il s'agit d'une véritable source de visibilité pour les marques qui cherchent à rajeunir leur image ou qui veulent toucher un public jeune sur Twitch, Youtube ou Dailymotion.

La loi pour une République

Numérique du 7 octobre

2016 a marqué la

reconnaissance, par

le gouvernement français,

de l'esport

## A PROPOS DE TWITCH

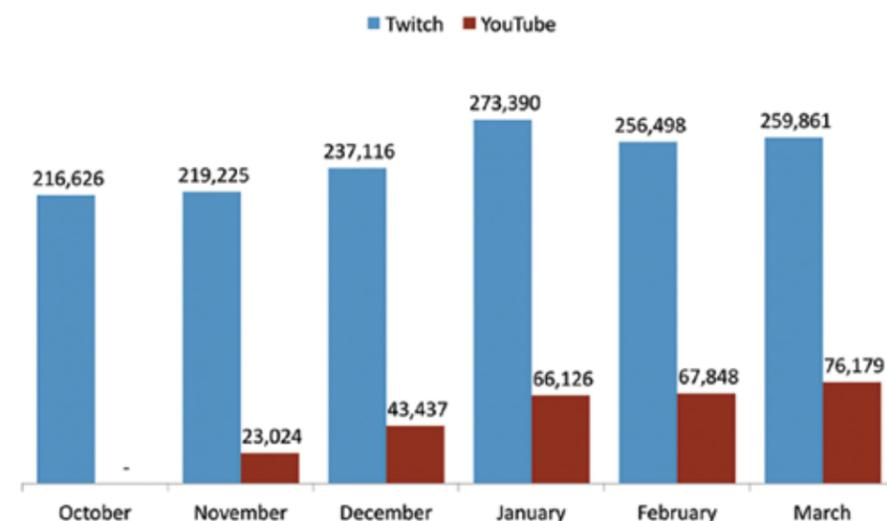
Twitch est une plate-forme vidéo de live streaming appartenant à Twitch Interactive, une filiale d'Amazon.com.

Le site se concentre principalement sur les jeux vidéo : la diffusion live de partie ou la retransmission de compétitions eSports par exemple. Le contenu sur le site peut être visualisé en direct ou en VOD.

La popularité de Twitch éclipserait celle de Youtube Live. Twitch a même été considéré comme la quatrième source de trafic Internet haut débit aux États-Unis.

### Streamlabs' Streamers On Twitch And Youtube

Monthly Active Streamer\* Growth, Global, 2017



Note: Only includes broadcasters that use Streamlabs.  
Source: Streamlabs, 2017

BI INTELLIGENCE

L'eSport existe d'ailleurs depuis 93 et l'activité arrive à maturité avec deux générations de gamers : les parents, situé entre 25 et 35 ans qui représentent 30% de la cible, et leurs enfants qui jouent à leur tour. Il s'agit d'une cible qui vieillit mais qui s'étoffe par le bas avec une très faible déperdition de public.

Il est difficile de donner des chiffres exacts sur le secteur et des études doivent être menées en la matière mais il faut savoir que pour quelques finales mondiales, ce sont des stades de 50.000 personnes qui se remplissent pour suivre en direct les «combats» des gamers professionnels.

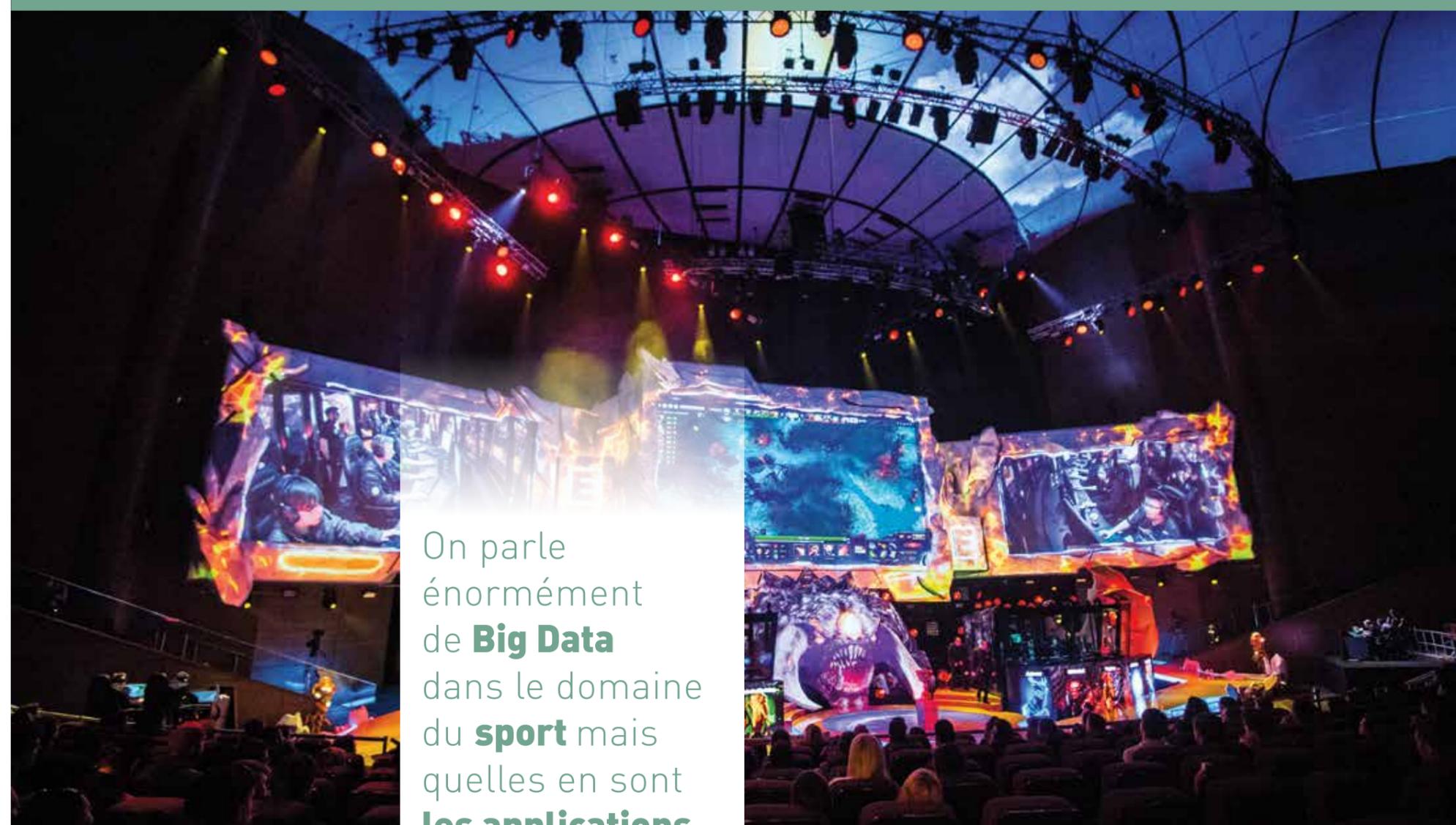
Comment fonctionne l'écosystème esport ?

Quels en sont les acteurs majeurs ?



L'écosystème de l'esport est véritablement tranché avec des acteurs distincts :

- **Les Éditeurs** qui investissent massivement dans le domaine de l'esport et détiennent les licences sur l'ensemble des jeux ;
- **Les Joueurs** qui achètent les licences de jeux, stream des parties etc...
- **Les Promoteurs d'événements sportifs** qui essaient de créer des écosystèmes financiers mais qui cherchent encore l'équilibre. L'équilibre des forces est plutôt du côté des éditeurs. En effet les promoteurs pro-



On parle énormément de **Big Data** dans le domaine du **sport** mais quelles en sont les **applications** dans l'**esport** ?  
Quelles sont les **données** qui peuvent être **récoltées** et **analysées** ?



De nombreux organismes travaillent sur le sujet pour délivrer des études mais les chiffres sont encore contestés. De mauvaises comparaisons sont faites entre les données du sport et de l'esport. Un match de finale de tennis a pu être comparé maladroitement à trois jours de compétition esport. L'un des principaux rôles de France eSports est donc de définir une étude nationale, un baromètre annuel pour mesurer les pratiques dans le secteur.

La Data est une grande problématique de l'esport pour connaître son activité, ses pratiquants, la géographie des pratiquants...

Les données démographiques et économiques sont primordiales pour définir :

- Quels sont les acteurs ? Sur quels territoires sont-ils présents ? Quels sont les chaînes de valeurs dans l'esport ?
- Qui pratique (les éditeurs peuvent fournir leurs chiffres sur ce point) ? Comment et pourquoi l'esport est-il consommé (stream, plateforme youtube, VOD) ?

La concaténation des données des consommateurs permet d'étudier leurs modes de consommation et de comprendre comment les toucher en tant qu'annonceur ou promoteur. La pratique des joueurs est très généraliste, il faut donc étudier également le comportement des joueurs amateurs et spectateurs.

L'enjeu Big Data de demain est de récolter également des données sur leur mode d'utilisation. Twitch détient une des bases de données les plus riches en la matière. C'est d'ailleurs la raison du rachat de Twitch par Amazon.

Par ailleurs, le **Big Data** fait partie intégrante de la stratégie des Leagues qui investissent dans des outils leur permettant d'améliorer la performance de leurs joueurs. En tant que Club, les données sont essentielles afin d'aller chercher de la statistique sur les jeux et parties et de faire du scouting afin de détecter de nouveaux talents. Les outils analytiques existants sont essentiellement développés en interne mais des développements plus poussés sont possible dans ce domaine.

★ **WORLD GAMING FEDERATION**, startup française, a développé une plateforme sociale pour les 2 milliards de joueurs de jeu vidéo et des services de Big Data pour les marques.

Ils souhaitent aider l'industrie du jeu vidéo à monétiser et faciliter les relations avec les utilisateurs. Ils collectent des données sur l'ensemble des joueurs afin d'uberiser leur utilisation des jeux vidéo.

Ils vont notamment au-devant de leurs besoins en leur proposant par exemple de faire des compétitions.

Pour une League, l'enjeu est également de bien qualifier son public car elle vend aux marques la possibilité d'atteindre le public souhaité.

Enfin, **des modèles capables d'analyser des parties de jeux et de récupérer des informations sur d'autres permettent de prédire les résultats d'un match d'esport**. Le PMU travaille par exemple sur ce format avec la société GamoLoco et étudie la partie prédictive du jeu.

## Quelles opportunités pour les entreprises de l'IT ?

★ Dans le domaine, **quatre acteurs majeurs vont s'affronter : Amazon avec Twitch, Youtube avec Youtube Gaming pour du format live, Samsung et enfin Microsoft** qui investit notamment dans une plateforme de diffusion concurrente à Twitch.

Ils se placent sur ces formats digitaux de diffusion notamment pour pouvoir produire et récupérer de grands volumes de données. Ces données et bases de données de joueurs recoupées avec d'autres données pourront être vendues et louées.

## Quels sont les challenges du secteur en matière de Big Data ?

★ L'esport est une activité par essence digitale, online, **le challenge réside en revanche dans l'ouverture des données du secteur**.

Certains éditeurs sont très ouverts, comme **Valve avec CounterStrike**. Par le biais de leurs APIs, ils permettent de récupérer la donnée et de travailler sur des modèles prédictifs. Certains jeux en revanche ne fournissent pas d'accès libre à leurs données et il faut davantage gratter pour récupérer de la donnée. **L'éditeur de jeu Blizzard a relativement fermé l'accès aux données** car il souhaite notamment se laisser une marge de développement sur la data et des applications potentielles en BtoB.

Un des modèles économiques possible pour un éditeur de jeux est par ailleurs la commercialisation de ses données.

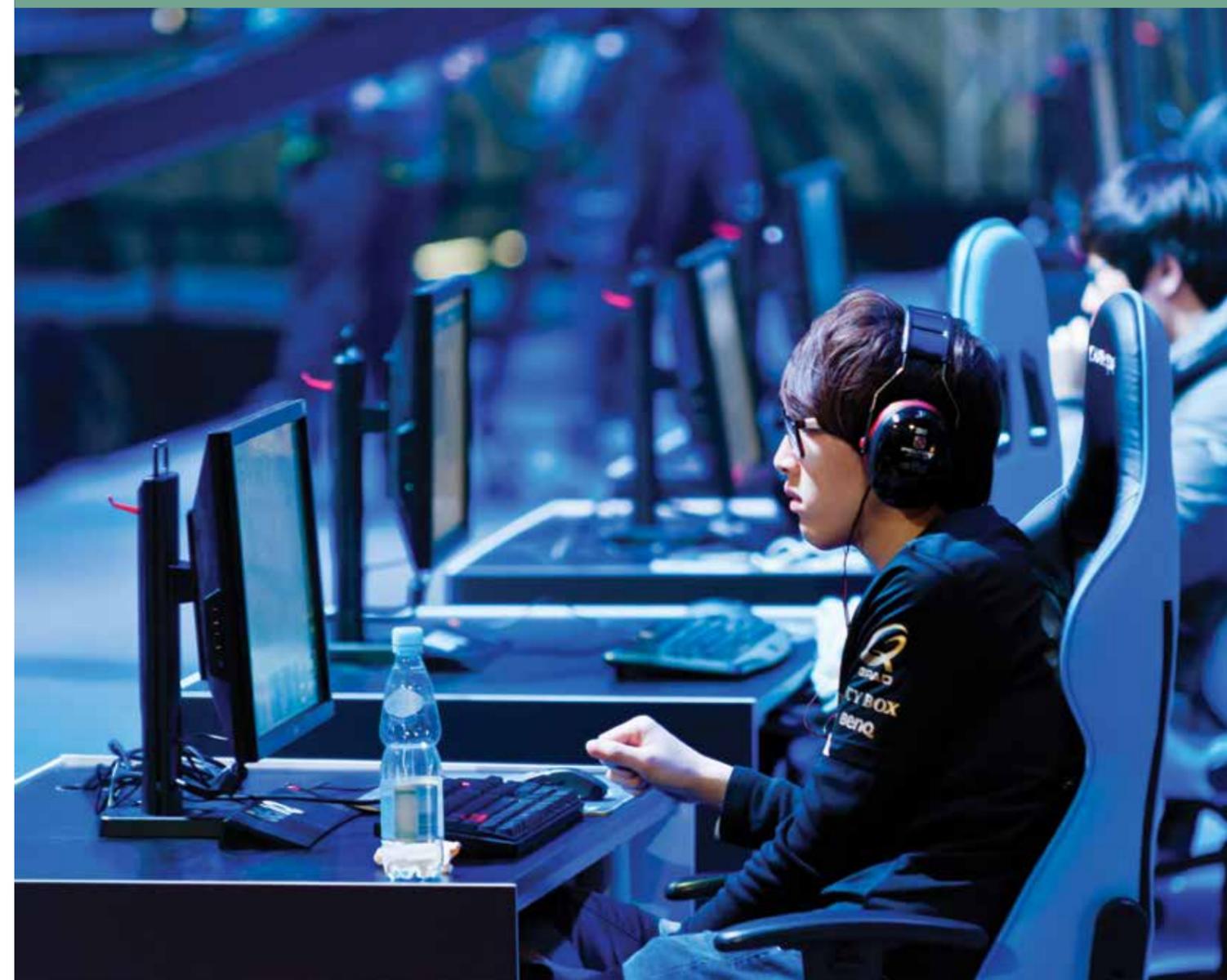
Le secteur français subit donc la stratégie des éditeurs qui maîtrisent les produits. C'est une des principales différences avec le sport, qui est beaucoup plus libre.

**Le modèle est encore fragile. Plusieurs startups réfléchissent à des modèles basés sur le Big Data. La data appartient soit à l'éditeur soit aux promoteurs**. Il est donc difficile pour une startup de développer un modèle économique alors même qu'elle n'est pas propriétaire de la data.

Certains startups tirent pourtant leur épingle du jeu et récoltent même des données que les éditeurs de jeux n'ont pas mais **il est primordial que les éditeurs et ayant droit sur les jeux lâchent du lest sur l'ouverture de leurs données afin d'encourager l'innovation en matière de Big Data dans le secteur**.



L'esport est une activité par essence digitale, online, le challenge réside en revanche dans l'ouverture des données du secteur





☆ Stéphane  
**MARCEL**

Directeur Général

SMAG  
Smart Agriculture



Directeur général de SMAG, Stéphane Marcel est également Directeur Général du fonds d'investissement privé InVivo Invest. Le fil conducteur de ses activités est le numérique, avec de nombreux enjeux et business cases autour de la gestion de la data (Big Data agricole).

# AGTECH : fleuron français de la donnée ?

Quels enjeux et défis pour l'agriculture connectée ?  
Un secteur en avance ou en retard dans l'utilisation du Big Data ?

★ Nous faisons actuellement face à une véritable révolution industrielle et digitale qui touche tous les secteurs d'activité. Cette révolution est complexe à absorber pour les acteurs économiques car c'est un modèle de rupture dans la façon de valoriser les données, de structurer les entreprises, de conduire la R&D, de manager. Nous passons d'un paradigme descriptif à un paradigme prédictif et inférentiel. Les métiers sont bousculés et l'agriculture n'est pas en marge de ces changements. Nous sommes passés d'une agriculture raisonnée, fondée sur des méthodes de bilan et des constats, à une agriculture mesurée et pilotée en permanence par la donnée.

L'agriculture vit sa propre révolution numérique et une véritable mutation structurelle. Cela oblige l'ensemble des acteurs à se remettre en cause.

On constate que des agriculteurs migrent petit à petit de leur métier de paysan vers une fonction de chef d'entreprise. Ils se doivent d'être polyvalents, poly compétents, afin de manager des préoccupations telles que l'augmentation des marges, la réglementation, les fluctuations des marchés de matière première, le respect de l'environnement, les aléas climatiques réguliers... Dans ce contexte, l'agriculture revendique l'utilisation des technologies nouvelles pour lui permettre de mieux piloter son activité.

Le secteur agricole n'est pas en retard sur la digitalisation. 95% des exploitants utilisent internet tous les jours. Le taux de pénétration des smartphones est très élevé sur cette population ultra consommatrice de nouvelles technologies. À l'image du tracteur qui est venu bouleverser le secteur d'activité, c'est désormais le digital et le Big Data qui viennent métamorphoser les pratiques culturelles, les processus de production et la prise de décisions. Les machines agricoles sont aujourd'hui bardées de capteurs qui produisent et récoltent des données en temps réel. Les technologies d'observation de la terre, basées sur les satellites et les drones, fournissent une foultitude de data avec une forte hétérogénéité (structurées, non structurées etc.). L'agriculture est donc un excellent candidat à l'utilisation du Big Data.

Nourrir 9 milliards de bouches sur la planète passera forcément par l'optimisation des rendements des parcelles agricoles et non uniquement par l'augmentation des surfaces exploitées. Le premier facteur de succès réside dans la R&D génétique mais le digital reste essentiel. Il faudra des outils d'aide à la décision faisant appel au Big Data pour atteindre cet objectif.

## ★ A PROPOS DE SMAG

Depuis 16 ans, SMAG accompagne les professionnels des filières agricoles et agro-industrielles à l'échelle nationale et internationale dans le déploiement de systèmes d'informations simples et performants, pour optimiser la gestion et le pilotage technico-économique des exploitations et des productions agricoles. L'objectif est de produire plus et mieux grâce à l'adoption des nouvelles technologies et en particulier grâce à l'exploitation de la donnée.

Le digital et le Big Data

viennent métamorphoser

les pratiques culturelles, les

processus de production et

la prise de décisions



## Comment la France se positionne-t-elle dans le secteur ?

★ Les AgTech représentent un segment particulièrement dynamique. Les plus gros players sont naturellement américains, à l'image de Google Ventures qui a largement investi dans le secteur. Israël est également particulièrement actif sur le sujet mais **l'AgTech française n'a pas à rougir**.

L'agriculture française a par ailleurs des spécificités et ne répond pas à la maxime « one product fits all ». Il existe des référentiels locaux, des cultures différentes dans le secteur et **il n'est ainsi pas possible d'adresser le marché agricole avec une seule application**. Il faut redescendre sur les spécificités agricoles même si l'enjeu est international.

## Quels sont les différents champs d'application du Big Data dans l'agriculture ?

★ Ils sont multiples. Par exemple, il est omniprésent dans le monde du **Farm Management software**, avec une forte implication de solutions logicielles dans l'exploitation des cultures ou l'élevage bovin... Ces outils offrent la possibilité de suivre les coûts de production sur différentes cultures ou parcelles et de surveiller les bénéfices dans l'ensemble de l'exploitation agricole pour optimiser les processus et améliorer la prise de décision.

Le Big Data est également exploité dans le domaine de **l'Agriculture de précision**. L'utilisation de modèles agronomiques et mathématiques permettent de moduler les interventions à l'intérieur même d'une parcelle agricole. En effet une parcelle agricole n'est pas homogène et certaines zones sont plus ou moins productives.

Jusqu'alors, on pratiquait l'agriculture à la parcelle. Aujourd'hui, les technologies analytiques permettent d'identifier les zones hétérogènes au sein même des parcelles. On parle donc d'intra-parcellaire. Grâce aux mesures réalisées par les drones et satellites, on est désormais capables de déterminer si un sol est de meilleure qualité au sein de la parcelle, quelle est la teneur en chlorophylle de la plante etc.

Cela nécessite des infrastructures hautement scalables pour faire tourner les modèles agronomiques afin de déterminer par exemple un degré d'azote à mettre sur chaque parcelle.

Ces informations peuvent même être communiquées directement à la console du tracteur. L'optimisation des intrants permet de produire plus (maximisation des cultures) et mieux (moins, d'azotes, moins de produits chimiques). Ce type de business case n'est possible que grâce à la data. BE API, l'une des structures intégrées au Pôle Agro-Digital d'InVivo est d'ailleurs intégralement dédiée au développement de l'agriculture de précision. Elle propose aux coopératives adhérentes InVivo différentes étapes qui permettent la modulation des semis, la fertilisation et l'usage des pro-

duits phyto en fonction de l'hétérogénéité intra-parcellaire et du besoin réel de la culture.

Le Big Data permet aussi de faire de **la prédiction de rendement et d'identifier la qualité des cultures avant la moisson**. Les modèles de prédiction permettent par exemple de connaître, à partir du mois d'avril, les rendements de culture de blé ainsi que sa teneur en protéines.

**Les stations météo connectées qui font partie des IoT** (Internet of Things ou Internet des objets) permettent la captation de données qui alimentent les algorithmes de météorologie et déterminent les plages de traitement idéales sur une parcelle en fonction des conditions météorologiques. L'exploitant gagne ainsi en temps (plus besoin de se déplacer sur sa parcelle si la météo n'est pas suffisamment favorable) et en argent (plus de traitement réduit à néant s'il se met à pleuvoir juste après l'intervention)

En somme, grâce au Big Data, on peut **augmenter la productivité et le rendement des cultures tout en diminuant l'impact environnemental**.

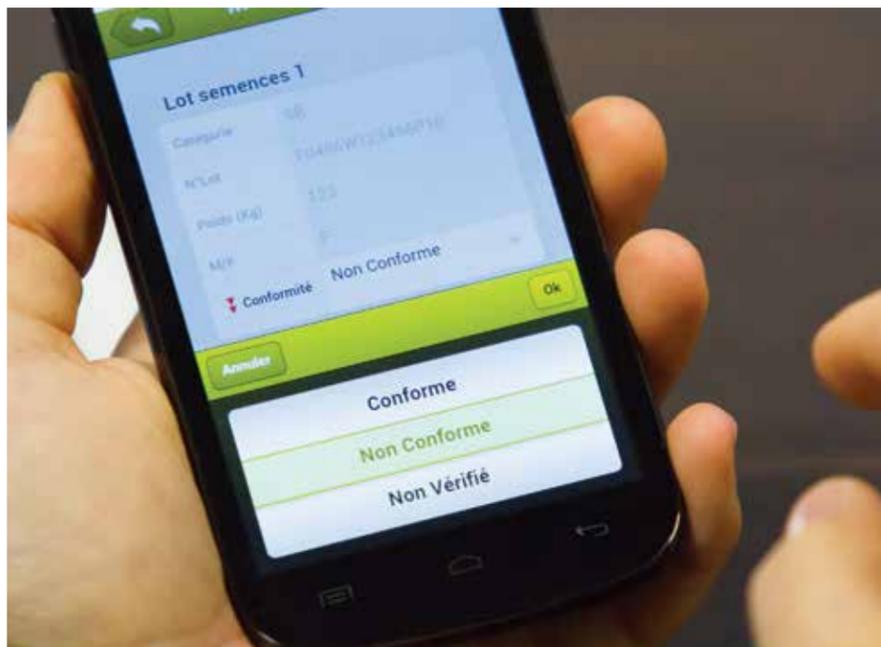


## Quels sont les différents types de données collectées ?

★ Elles portent sur le sol (température, humidité, qualité, éléments organiques), l'air (température, hydrométrie, vent, ensoleillement), les plantes (stress hydrique, vigueur, taux de chlorophylle) ainsi que sur les parcelles (traitements, fertilisation aux champs etc...) et l'historique des cultures d'une campagne à l'autre.



Toutes les données récoltées alimentent notre Data Lake qui nécessite une forte capacité de traitement. Sur une surface agricole utile de 30 millions d'hectares en France (10 millions en prairie et 20 millions en culture), nos plateformes gèrent plus de 10 millions d'hectares. Au total, ce sont plus de 25 000 agriculteurs qui utilisent nos solutions. Plus nous disposons de données et meilleurs sont les conseils donnés à nos utilisateurs.



## Quel est le business model du Big Data dans l'agriculture ?

★ C'est un business model propre au métier d'éditeur. Les applications sont installées sur des serveurs distants et disponibles sous forme d'abonnements en mode SaaS (Software as a Service) en fonction de ce qu'ils consomment. En somme, un logiciel en ligne, hébergé dans le Cloud. Rien n'est installé sur l'ordinateur de l'exploitant, une simple connexion internet lui suffit pour accéder à ses outils.

SMAG œuvre ainsi dans le domaine des objets connectés, de l'agriculture de précision, des applications mobiles, de la saisie d'information agronomique, de la prise de photo (image mining & data mining) et de l'aide à la décision (OAD). Les couches logicielles sont produites en interne tandis que le développement des objets connectés revient à des partenaires. L'intelligence logicielle est donc au cœur du business.

Vers une mise à disposition des données agricoles en open-data ?

★ Les plateformes construites dans l'agriculture sont globalement assez ouvertes grâce à des APIs gratuites ou payantes. Un certain nombre de données sont déjà en Open Data mais cela ne va pas assez loin. Il y a encore des avancées à faire en matière de réglementation notamment concernant les opérateurs publics de données agricoles qui ne sont pas encore en Open Data.

Ils détiennent des données de déclaration comme les données de la PAC (Politique Agricole Commune). En France, nous sommes encore loin du développement actuel de l'Open Data qui s'opère aux États-Unis et qui donne accès aux données sur les types de sols ou la géographie par exemple. Le développement de l'Open data entraînera plus d'Open Innovation et naturellement plus de création de valeur.

Le monde agricole saura-t-il également s'emparer de l'Intelligence Artificielle ?

★ Nous ne faisons pas encore d'Intelligence Artificielle mais cela ne saurait tarder. Nos développements sont essentiellement liés aux business cases des clients de la filière. Le secteur saura s'approprier l'IA si le besoin se fait ressentir. L'agriculture s'est, par exemple, emparée depuis bien longtemps de la robotique et il n'y a aucun frein de principe à ce qu'elle puisse mettre à profit également l'IA au sens pur.

C'est aux acteurs du secteur qu'il revient d'écrire le futur de l'AgTech



Quelles futures tendances pour l'AgTech ?

★ De nouvelles technologies vont trouver leur place dans l'agriculture, comme la Blockchain ou encore l'impression 3D.

La reconnaissance et le traitement d'images vont progresser avec, par exemple, des applications permettant d'identifier certaines maladies grâce aux smartphones. Cela pourra notamment faire appel à l'IA en termes de traitement.

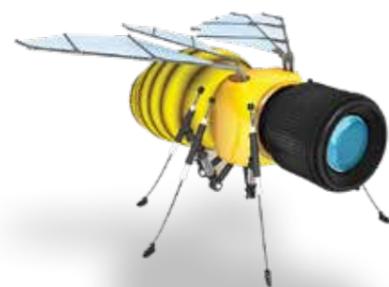
L'agriculture urbaine ou Urban Farming sera particulièrement consommatrice de technologies qui permettront de maîtriser l'ensemble des paramètres d'exploitation sur des surfaces réduites. L'utilisation du Big Data y sera d'autant plus importante.

Globalement le digital dans l'agriculture est créateur de valeur économique, environnementale, sociétale et agronomique. Il permet de renouveler les modèles éco-

nomiques un peu ancestraux du secteur. Pour faire émerger de nouveaux produits ou services digitaux liés à l'agriculture de précision, les objets connectés et le Big data, nous avons créé l'année dernière un «Studio» agro-digital de 600 m2 à Montpellier.

Il faut cependant que les acteurs du secteur acceptent de changer de modèle. C'est à eux qu'il revient d'écrire le futur de l'AgTech. Ceux qui expliquent que l'on pourra se passer de l'agriculteur ont tort, car ce sera toujours à lui de prendre les décisions. Cependant, grâce au digital, ce sera plus simple, plus rapide et davantage anticipé. Le digital permettra demain de mieux maîtriser les risques de l'agriculture. C'est pour cela qu'il est tant au cœur des débats actuels.

★ A PROPOS DE L'URBAN FARMING C'est la pratique de la culture, de la transformation et de la distribution de nourriture dans ou autour d'un village, d'une ville ou d'une métropole. L'agriculture urbaine peut également impliquer l'élevage, l'aquaculture, l'agroforesterie, l'apiculture urbaine et l'horticulture.





★ Ivan  
**OSTROWICZ**

Co-Founder & CEO

**DOMOSCIO**



Ivan Ostrowicz est ingénieur diplômé de l'Université Polytechnique de Catalogne et de l'ENSIMAG et diplômé du MBA de Neoma BS. Il a une expérience à l'international dans le management des organisations et en des systèmes d'informations en entreprise (La Caixa, SFR, Air France) et en cabinet de conseil (EY).

# Learning analytics : vers une révolution de l'éducation ?

Nous sommes une startup Edtech spécialisée dans le Big Data et l'Intelligence Artificielle combinés avec les sciences cognitives afin de proposer des solutions d'apprentissage innovantes d'Adaptive Learning. Fondée en 2013, Domoscio signifie en espéranto «Maison de la Connaissance». Ce nom révèle beaucoup sur la volonté de la startup d'inventer une nouvelle approche du savoir en transformant le processus d'apprentissage.

Quel **constat** faites-vous tout d'abord de la **digitalisation** de l'**éducation** ?

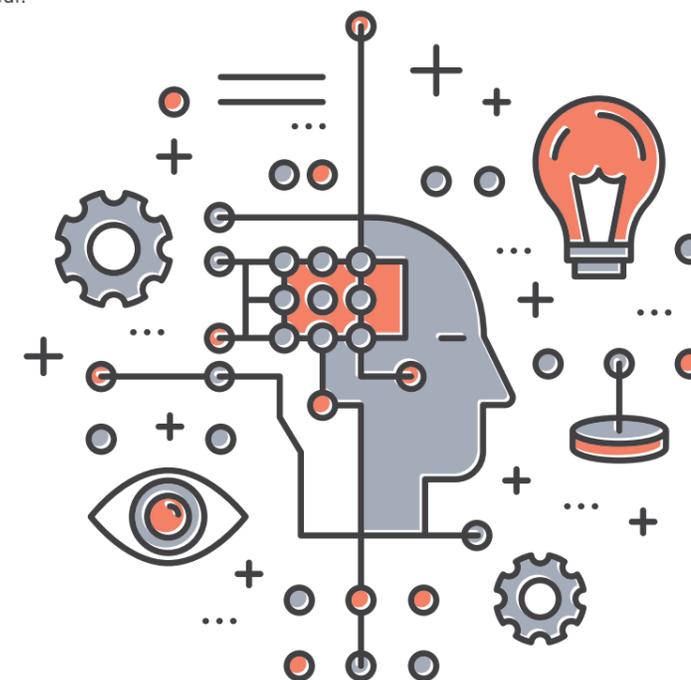
Quels sont **les enjeux français** en termes d'**éducation en ligne** ?

★ On constate une différence d'évolution entre, d'une part, le marché de l'éducation primaire, secondaire et universitaire, et d'autre part, le marché de la formation professionnelle. Ce dernier, qui s'adresse aux entreprises et organismes de formation, a été poussé à évoluer de par la réforme de la formation.

Le marché, il y a encore 4 ans, était très peu actif, mais il doit désormais s'adapter au besoin de personnalisation des méthodes d'apprentissage. Cette personnalisation nécessite de nombreuses ressources, soit 1 formateur par apprenant, soit l'utilisation de la technologie. Aujourd'hui, certains organismes de formations privilégient une approche **blending learning**, c'est-à-dire une **balance équilibrée entre présentiel et distanciel**. Nous sommes convaincus que le présentiel a toujours sa place dans le modèle éducatif et de la formation. D'autant plus que nombre d'outils permettent de revigorer le présentiel et de l'animer. De même, coté distanciel, différentes formes de formations sont nées : les MOOC, et surtout l'**Adaptive Learning**.

Les algorithmes permettent un meilleur pilotage de la formation et une réduction des temps de formation. En moyenne, les solutions d'Adaptive Learning nécessitent **37% de ressources en moins** pour arriver à un même objectif et trois mois après, **les apprenants retiennent en moyenne 79% de ce qu'ils ont appris**. C'est la recherche de performance, de résultats et le besoin de KPIs sur les formations qui fait bouger les lignes du secteur.

**Les algorithmes  
permettent un meilleur  
pilotage de la formation  
et une réduction des temps  
de formation**



## Qui sont les précurseurs de cette révolution de l'apprentissage ?

★ Les États-Unis, avec un marché très actif et un investissement de plus de 150 millions de dollars dans des startups, sont naturellement précurseurs sur le sujet. Il faut cependant se confronter à la réalité de marché. La taille du marché implique des solutions très hétérogènes, des plateformes de e-learning très différentes, des visions de la formation et des contenus différents.

Le Royaume-Uni, la Norvège et les Pays Bas sont également assez avancés en termes de solutions de Learning Analytics et d'Adaptive Learning, mais le marché français n'est pas forcément en retard et les éditeurs français ont fait un saut énorme ces trois dernières années. Néanmoins le marché est assez restreint et récent. D'après l'Observatoire de l'EdTech, on recense plus de 254 sociétés françaises sur le marché. Et 75 d'entre elles déclarent faire de l'Adaptive Learning, mais certains utilisent ce terme à des fins purement marketing.

Nous constatons également que de nombreuses solutions d'Adaptive Learning fonctionnent avec leur propre plateforme et parfois leur propre contenu. Cela limite largement les interconnexions avec les plateformes de e-learning internes aux entreprises, qui sont en demande de solutions d'Adaptive Learning.

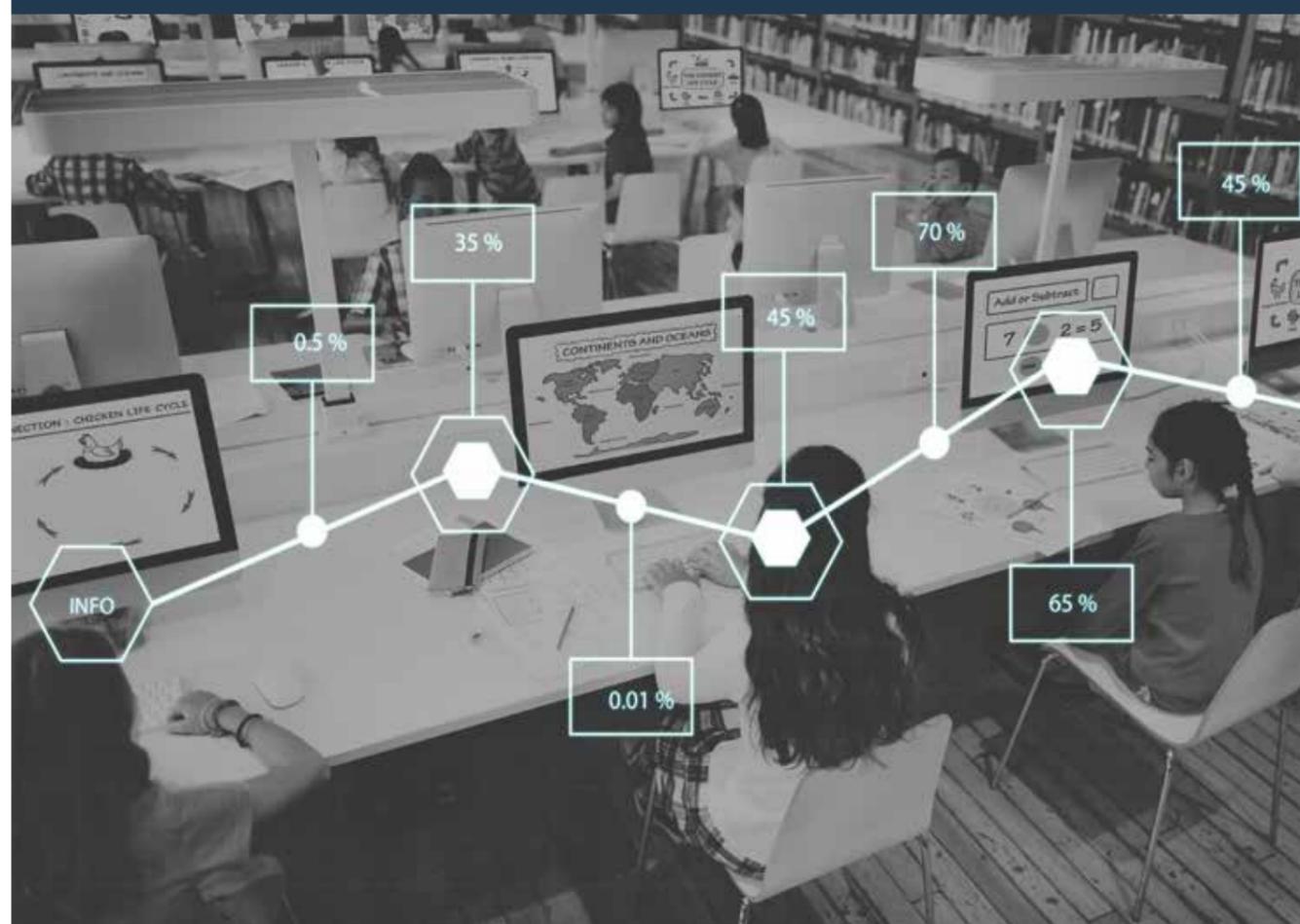
## Quel est le fondement de Domoscio ?

★ L'origine vient d'un besoin particulier de personnalisation de la formation ressenti par Benoit Praly, ingénieur spécialisé en mathématiques à l'ENSIMAG de Grenoble en tant qu'étudiant. La première version de la solution développée, intégrant des algorithmes, était donc à des fins d'utilisation personnelle. Nous sommes donc partis du premier constat que nous apprenons tous différemment, que chacun apprend à son rythme et à sa manière mais également

du second constat que les apprenants vont oublier ce qu'ils ont appris. Le système scolaire primaire et secondaire, avec un professeur pour 30 élèves, et l'enseignement supérieur, avec un professeur pour 200 élèves ne sont pas adaptés à l'individualité de l'apprentissage et à l'engagement nécessaires de l'apprenant. Le système actuel de l'éducation et de la formation commence à avancer pour proposer des solutions de personnalisation du processus d'apprentissage et consolidation des acquis.

Le taux d'engagement du e-learning, avec par exemple, de simples mises à disposition de vidéos de 2h en ligne, est faible. Les Moocs représentent une avancée certaine dans l'élargissement de la formation à une plus large population, mais l'assiduité est également critiquée par le manque de personnalisation au besoin de chacun.

En effet, l'objectif d'un apprenant sur un Mooc n'est absolument pas pris en compte. Une personne peut n'avoir pour objectif que l'étude d'un chapitre, on ne devrait alors pas forcément parler d'abandon. La personnalisation permet d'aller plus loin sur la mesure de l'engagement des apprenants.



## ★ À PROPOS DES MOOCS

Une étude de l'université de Pennsylvanie donnait déjà l'alarme en 2013 : le taux de rétention, c'est-à-dire le pourcentage d'inscrits qui suivent un MOOC jusqu'au bout, est particulièrement bas, de l'ordre de 5 % à 10 % en moyenne. La part des inscrits aux MOOC du MIT et de Harvard en 2012-2013 ayant suivi plus de la moitié des cours s'élève à seulement 9% et à peine 5 % ont validé leur formation.

L'éducation présente donc deux challenges prioritaires : comment personnaliser l'expérience d'apprentissage et comment permettre à l'apprenant de retenir ce qu'il a appris.

Nous nous sommes fondés sur les concepts et techniques issus des sciences cognitives, neurosciences et de la psychologie cognitive tels que les courbes de l'oubli ou encore la répétition espacée mais également d'études prouvant l'impact de la personnalisation sur l'apprentissage. Pour appliquer ces méthodes à l'informatique, nous nous basons naturellement sur la donnée. Le Big Data et les algorithmes d'Intelligence Artificielle, spécialement de Machine Learning, sont naturellement au cœur de Domoscio. Nous utilisons des algorithmes de Machine Learning, afin de traiter les données, couplés aux sciences cognitives, pour faire ressortir la meilleure méthode d'apprentissage pour chaque individu.

Nous combinons donc trois solutions d'Adaptive Learning pour recommander la bonne méthode à la bonne personne en fonction de son objectif d'apprentissage. Nous envoyons les données vers notre API, qui analyse, fait des calculs, avant de les renvoyer à la plateforme pour faire des recommandations personnalisées.

Nous avons donc développé trois solutions technologiques pour trois étapes d'apprentissage :

**1 Premier apprentissage :** une solution permettant de collecter les données d'apprentissage de l'apprenant et, en retour, de lui proposer à l'instant T, et selon ses objectifs pédagogiques, les contenus, exercices et formations les plus adaptés.

**2 Consolidation :** une solution d'ancrage adaptatif® qui répond au besoin de rétention d'information avec un planning de révisions et des piqûres de rappel adaptés à chaque apprenant. Le système d'Intelligence Artificielle est notamment capable de mesurer à quelle vitesse l'apprenant oublie.

**3 Application :** une solution d'analyse des données présentant toutes les Learning Analytics sous forme de tableaux de bord pour permettre d'analyser l'avancement de l'apprenant, la manière dont le contenu est utilisé et consommé, d'alerter sur les apprenants qui ont bientôt terminé ou au contraire ceux qui sont en risque de décrochage. Ce troisième volet permet des feedbacks indispensables et recherchés dans le domaine de l'éducation et bien sûr de la formation.

## Qui sont les acteurs concernés par l'Adaptive Learning ?



L'adaptive Learning adresse 4 types de publics :

- Les éditeurs de contenus, éditeurs de manuels scolaires ou de formation professionnelle (comme les éditions Hatier, Didier, Hachette, Edebé...)
- Les professionnels de la formation professionnelle : les directions formation, université d'entreprise, RH
- Les Institutions académiques et organismes de formation tels que l'université Paris Descartes, IFIS (santé), Abilways...
- Les plateformes de e-learning, d'apprentissage en ligne et LMS : Xperteam par exemple dans le domaine des plateformes propriétaires ou Moodle pour une plateforme Open Source.

En se positionnant comme acteur complémentaire aux acteurs de l'éducation, Domoscio propose ainsi, soit d'utiliser leur plateforme de e-learning, soit de s'interfacer avec les plateformes existantes des clients, propriétaires ou open source.

temps a-t-il passé sur le contenu, le résultat des exercices, s'il a consulté des contenus lorsqu'il faisait un exercice etc...

Nous utilisons ensuite des algorithmes de Machine Learning, qui permettent de mettre à jour en continu nos modèles pour fournir des recommandations personnalisées en fonction des nouvelles données recueillies. Pour faire tourner ces algorithmes, nous fonctionnons avec une infrastructure en SaaS. La puissance de calcul nous est offerte par des technologies Microsoft et nous avons fait le choix de serveurs à l'international afin d'être capable de répondre aux besoins à un niveau mondial.

L'Adaptive Learning fait aussi du « clustering ». Notre algorithme propose notamment un match-making automatisé pour une mise en relation des personnes de niveaux homogènes ou ayant les mêmes méthodes d'apprentissage afin de générer de l'entraide entre les participants. Mais l'algorithme peut également proposer des profils experts dans certaines catégories ou l'apprenant bloque, dans le but de favoriser la collaboration entre les apprenants. L'Adaptive Learning peut donc être particulièrement utile :

- À l'apprenant qui grâce aux Learning Analytics peut voir sa progression, ses points

forts, ses points à travailler etc.

- Aux éditeurs pour analyser les contenus les mieux compris, les moins utilisés, les adapter si besoin

- Aux ressources humaines et fonctions formation des entreprises pour analyser l'engagement sur la plateforme, les risques de décrochage sur une thématique, les blocages etc... mais également le ROI des investissements dans la formation.

- Aux managers des entreprises. Grâce au Machine Learning, on connaît le profil de l'apprenant, sa typologie, on sait quelles formations il suit, qu'est-ce qui a bien fonctionné pour lui dans le passé, et on est en mesure de l'appliquer au présent pour permettre à ses salariés d'acquérir les bonnes compétences, dans un contexte professionnel où la tendance est au turn-over horizontal.

## Quelles seront, selon vous, les futures applications du Big Data et de l'IA dans l'éducation



Il y aura de plus en plus de numérique dans la formation, dans l'éducation. Cependant, nous sommes convaincus que les modèles présentiels et online ne s'opposent pas. Les applications interactives, la gamification et la personnalisation de parcours vont être indispensables aux nouvelles méthodes d'apprentissage mais l'humain ne sera pas laissé de côté. Nous envisageons l'humain en symbiose avec le numérique dans le parcours d'apprentissage. Le rôle du professeur et du formateur va naturellement évoluer vers un rôle d'accompagnateur.

L'Intelligence Artificielle sera complémentaire, car aujourd'hui, elle n'est pas capable d'interagir avec l'apprenant dans tous les types d'interactions humaines possibles. Il y a par exemple des moments de rencontre incontournables pour compléter l'apprentissage informatique notamment dans les métiers manuels comme l'ébénisterie. Les algorithmes vont par ailleurs créer des moments de rencontre entre des individus complémentaires ou similaires pour faciliter l'apprentissage.

Le Big Data a également vocation à entrer dans la personnalisation de la gamification de l'apprentissage. La gamification peut être liée au contenu et doit être différente en fonction de la cible. Les enfants seront friands de médailles. On va chercher à ce qu'ils s'investissent plus même s'ils font plus d'exercices en privilégiant le côté pédagogique. Dans la formation professionnelle, la gamification doit être orientée vers l'acquisition de compétences, les challenges et la réussite professionnelle entre collègue. L'Adaptive Learning est plus que jamais d'actualité. Plus on aura de numérique et de technologies, plus on pourra développer de formations numériques et plus les intelligences artificielles seront puissantes et efficaces dans la personnalisation de l'apprentissage.

Plus on aura  
de numérique et de  
technologies, plus on  
pourra développer  
de formations numériques  
et plus les intelligences  
artificielles seront  
puissantes et efficaces  
dans la personnalisation de  
l'apprentissage.

## Quelles sont les données qui peuvent être récoltées dans l'enseignement ? Pour quelles applications ?



S'il y a des données existantes, nous récoltons en amont toutes les données d'apprentissage, de recherches, les données administratives ou de CRM disponibles dans les universités par exemple qui pourraient être intéressantes pour les solutions d'adaptive learning. Nous récoltons par la suite toutes les données de parcours de l'utilisateur. Sur quel contenu l'utilisateur est passé, combien de

# GAMIFICATION



# L'Annuaire

	 Analytique	 API	 Applications	 Consultants/SSII Intégrateurs	 Dataviz	 Infrastructure	 Institutionnels
42 CONSULTING DIGITAL & DATA	.	.	.	.	.	.	.
AB INITIO SOFTWARE	.	.	.	.	.	.	.
ACCENTURE ANALYTICS	.	.	.	.	.	.	.
ADELEAN	.	.	.	.	.	.	.
ADVANCED SCHEMA	.	.	.	.	.	.	.
ALTARES	.	.	.	.	.	.	.
AMAZON WEB SERVICES	.	.	.	.	.	.	.
ATTUNITY	.	.	.	.	.	.	.
BLUE DME	.	.	.	.	.	.	.
BLUESCALE	.	.	.	.	.	.	.
BMB SERVICES	.	.	.	.	.	.	.
BRENNUS ANALYTICS	.	.	.	.	.	.	.
CB EXECUTIVE	.	.	.	.	.	.	.
CENTRALESUPÉLEC EXED	.	.	.	.	.	.	.
CISCO SYSTEMS	.	.	.	.	.	.	.
CLARANET	.	.	.	.	.	.	.
CLOUDERA	.	.	.	.	.	.	.
CLUB DECISION DSI	.	.	.	.	.	.	.
COHERIS	.	.	.	.	.	.	.
CONFLUENT INC	.	.	.	.	.	.	.
COUCHBASE	.	.	.	.	.	.	.
DATAIKU	.	.	.	.	.	.	.
DATASTAX	.	.	.	.	.	.	.
DENODO	.	.	.	.	.	.	.
DIGDASH	.	.	.	.	.	.	.
EBIZNEXT	.	.	.	.	.	.	.
EKIMETRICS	.	.	.	.	.	.	.
ELASTIC	.	.	.	.	.	.	.
EULIDIA	.	.	.	.	.	.	.
EXPERIS	.	.	.	.	.	.	.
EXPERT SYSTEM	.	.	.	.	.	.	.
FINAXYS	.	.	.	.	.	.	.
GROUPE ESTIA	.	.	.	.	.	.	.
HORTONWORKS	.	.	.	.	.	.	.
HPE	.	.	.	.	.	.	.
HURENCE	.	.	.	.	.	.	.
HYPERCUBE	.	.	.	.	.	.	.
IDEATRANS	.	.	.	.	.	.	.
INDEXIMA	.	.	.	.	.	.	.
INVIOO	.	.	.	.	.	.	.

# Entreprises & Labels

	 Analytique	 API	 Applications	 Consultants/SSII Intégrateurs	 Dataviz	 Infrastructure	 Institutionnels
JEMS DATAFACTORY	.	.	.	.	.	.	.
KEYRUS	.	.	.	.	.	.	.
LANSROD BIG DATA	.	.	.	.	.	.	.
MAPR TECHNOLOGIES SAS	.	.	.	.	.	.	.
MATHWORKS	.	.	.	.	.	.	.
MICROSTRATEGY	.	.	.	.	.	.	.
MONGO DB	.	.	.	.	.	.	.
NEO4J	.	.	.	.	.	.	.
ORACLE	.	.	.	.	.	.	.
ORANGE BUSINESS SERVICES	.	.	.	.	.	.	.
OUTSCALE	.	.	.	.	.	.	.
OVH	.	.	.	.	.	.	.
PENTAHO	.	.	.	.	.	.	.
PITNEY BOWES SOFTWARE	.	.	.	.	.	.	.
PROVALIS RESEARCH	.	.	.	.	.	.	.
OLIK	.	.	.	.	.	.	.
SAAGIE	.	.	.	.	.	.	.
SAP	.	.	.	.	.	.	.
SAS INSTITUTE S.A.S	.	.	.	.	.	.	.
SCINETIK	.	.	.	.	.	.	.
SINEQUA	.	.	.	.	.	.	.
SKAPÂNÊ	.	.	.	.	.	.	.
SNOWFLAKE	.	.	.	.	.	.	.
SOFT COMPUTING	.	.	.	.	.	.	.
SOPRA STERIA	.	.	.	.	.	.	.
SPLUNK	.	.	.	.	.	.	.
SQLI	.	.	.	.	.	.	.
SYNALTIC GROUP	.	.	.	.	.	.	.
TABLEAU	.	.	.	.	.	.	.
TALEND	.	.	.	.	.	.	.
TÉLÉCOM PARISTECH TÉLÉCOM EVOLUTION	.	.	.	.	.	.	.
TERADATA	.	.	.	.	.	.	.
TIBCO SOFTWARE	.	.	.	.	.	.	.
TIMI	.	.	.	.	.	.	.
TOM SAWYER SOFTWARE	.	.	.	.	.	.	.
TRIFACTA	.	.	.	.	.	.	.
XEBIA	.	.	.	.	.	.	.
YSANCE	.	.	.	.	.	.	.
ZENIKA	.	.	.	.	.	.	.

# 42 CONSULTING DIGITAL & DATA

LUDOVIC MAUX

Manager du Pôle Big Data

01 45 18 33 00

data@42consulting.fr

2 bis avenue Foch  
94160 Saint-Mandé France

www.42consulting.fr



42  
CONSULTING  
DIGITAL & DATA



## HISTORIQUE

42 Consulting est une société de conseil en technologies et systèmes d'informations. Depuis 2006, au fil de nombreux projets réalisés pour ses clients grands comptes, l'entreprise a développé une croissance constante et une expertise pointue. Cette dernière, à la fois métier, technologique et méthodologique, permet à 42 Consulting d'assurer sa mission : livrer, pour chaque projet, des solutions opérationnelles, efficaces et innovantes qui répondent aux problématiques et aux besoins de ses clients. Le « delivery » est donc au cœur de l'approche de 42 Consulting, qui propose des dispositifs sur-mesure permettant à ses clients de réussir leur transition numérique. 42 Consulting déploie son savoir-faire sur trois grands pôles d'activité : Digital & Data - Finance & Assurance - Media & Telecom. Forte de ses 200 collaborateurs en France et à l'international, l'entreprise répond aux besoins de ses clients à travers le monde. Ses bureaux sont situés à Paris, Luxembourg, Amsterdam, Bruxelles et Casablanca. En outre, 42 Consulting déploie des partenariats avec les écoles, notamment l'Estiam (Ecole supérieure des technologies de l'information appliquées aux métiers). Ce lien privilégié apporte aux acteurs du secteur une solution à leur problématique de recrutement de spécialistes du numérique.

## FOCUS : L'EXPERTISE DIGITAL & DATA

42 Consulting Digital & Data conçoit et développe des applications numériques multi-devices et active les données pour améliorer les services et ainsi accélérer la transformation numérique de ses clients. 42 Consulting a, entre autres, développé une offre pour monitorer et optimiser les infrastructures techniques, notamment :  
- La conception et le développement Web & Mobile full-stack  
- L'architecture Big Data, la Data Ingénierie, la Data Science et la Data Gouvernance  
42 Consulting Digital & Data intervient sur l'identification des use cases, le design d'architecture, le management de projet, le développement agile, l'hébergement, l'intégration, le testing & la validation, ainsi que la maintenance.

## EXEMPLE DE REALISATION

L'un des principaux opérateurs de télécommunications français a récemment fait appel à 42 Consulting pour améliorer la qualité de service de sa plateforme VOD (vidéo à la demande). L'objectif principal du client dans cette réalisation était, par le biais de données techniques délivrées par l'infrastructure - données non explorées jusque-là - de détecter, prédire et diagnostiquer les problématiques de fonctionnement des plateformes plus rapidement et plus efficacement. Ceci a permis, notamment, de réduire la durée des incidents et des diagnostics, et de détecter des incidents jusque-là inexplorés. D'autre part, la collecte des données et leur historisation a permis d'établir des études de tendances sur les incidents, de valider l'effet des améliorations de la plateforme et de construire des modèles statistiques pour prédire les incidents.

## Les phases de développement du projet

**Exploration des données** : La première phase du projet consistait à initier une exploration des données : collecte des logs de la plateforme de services, définition des indicateurs métiers pertinents pour monitorer la qualité de service, processing des logs et visualisation de dashboards « quantitatifs ». Ce travail exploratoire a

permis de dégager les premières tendances et de mettre en lumière les indicateurs les plus pertinents, dans le contexte spécifique de l'opérateur. S'est alors avérée nécessaire la création de nouveaux indicateurs - non existants dans les données sources - mais pertinents, pour quantifier l'amélioration de la qualité de service.

**Prototypage** : La deuxième phase du projet consistait à prototyper, définir et dimensionner une architecture Big Data capable de traiter les volumétries de logs issues de la plateforme (soit par jour : 10 milliards de lignes de logs en entrée, 50 millions d'événements indexés). Une fois cette plateforme définie, une phase de Data ingénierie allait permettre, sur un échantillon de données, de développer les modules d'imports des données, de mettre en place les scripts pour processor et enrichir les données, de les indexer et de développer les dashboards définis lors de la phase d'exploration. Permettant alors aux équipes métiers de l'opérateur de mesurer la qualité de service.

**Industrialisation** : Après cette phase de prototypage a débuté la phase d'industrialisation de la solution. Cette étape clé devait permettre de préparer la mise en production de la solution et de répondre aux problématiques liées à la mise à l'échelle, la sécurisation et l'automatisation de bout en bout. Pour ce projet, 42 Consulting Digital & Data a utilisé l'ensemble de la Stack Elastic (Logstash, Elasticsearch, Kibana) ainsi que Spark pour le processing des données.



**Ce qu'a permis l'expertise 42 Consulting Digital & Data** : Les experts Data de 42 Consulting ont accompagné l'opérateur, en appui avec les experts métiers, dans toutes les phases du projet depuis l'exploration des logs. Grâce à un travail en synergie avec les experts métiers de l'opérateur, ils ont su appréhender la complexité technique de la plateforme, y associer leur expertise en architecture et ingénierie des données, pour créer les indicateurs de monitoring pertinents.

**Les résultats pour le client** : Grâce à cette solution développée par 42 Consulting, l'opérateur dispose d'une plateforme de monitoring de la qualité de services de son service de VOD. Les indicateurs mesurés étaient jusqu'alors présents dans des logs non exploités et de nouveaux indicateurs ont pu être créés, permettant un suivi détaillé des différents composants techniques et la remontée d'anomalies.

## LES ATOUTS DE 42 CONSULTING

**Expertise** : 42 Consulting est une société de conseil spécialiste des technologies core business de ses clients. Le savoir-faire métier constitue l'ADN de la marque depuis sa création.

**Agilité** : 42 Consulting ne se contente pas de mettre en œuvre des process agiles pour ses clients. La société elle-même est agile depuis ses débuts, capable de se réinventer en permanence.

**Responsabilité** : Experte des technologies numériques, 42 Consulting s'estime responsable de ses actions sur le plan humain, social et écologique.

**Passion** : C'est le moteur qui fait tourner 42 Consulting : tous ses consultants sont des passionnés de technologies, qui prennent un réel plaisir à se dépasser sur chaque projet.

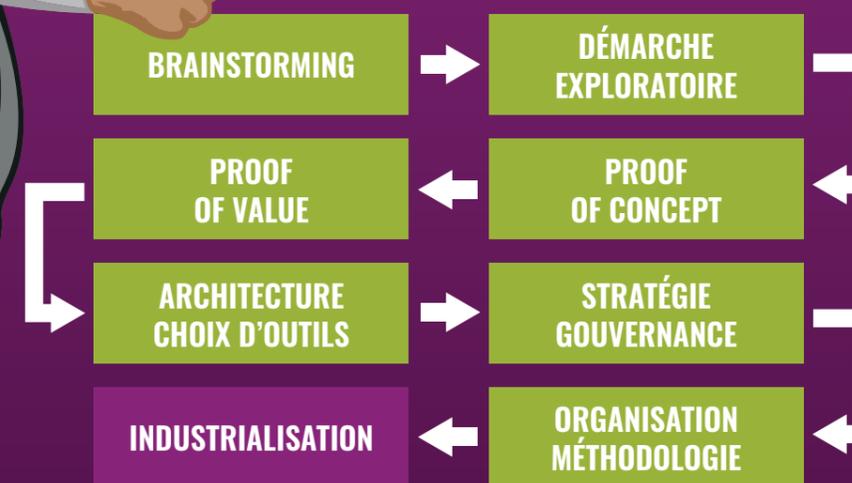
**Humour** : Les collaborateurs 42 Consulting sont sérieux sans se prendre au sérieux.

# 42

## CONSULTING DIGITAL & DATA

### Où en êtes-vous dans vos projets

# BIG DATA?



42 Consulting accompagne vos projets **BIG DATA**  
et vous apporte son expertise en :



Data  
Architecture



Data  
Ingénierie



Data  
Science



Data  
Gouvernance

data@42consulting.fr  
www.42consulting.fr

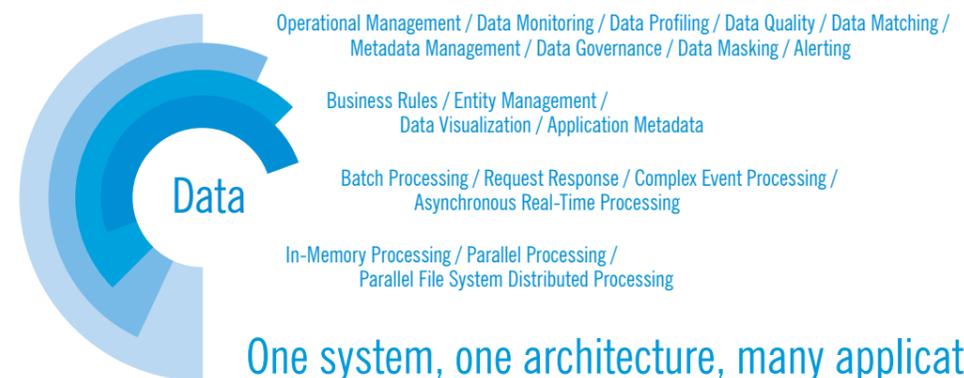


## Why all these companies turned to Ab Initio

Not only is Ab Initio software intuitive and easy to use, it also stands up to the most complex application logic and to huge amounts of data. And it does this with high performance and remarkable robustness. This combination is unique.

For more than 20 years, Ab Initio has focused on the processing and management of corporate data. Many of the largest companies in the world have chosen Ab Initio to solve the challenges of making the most out of their data. That's why Ab Initio offers a complete solution for reading, writing, processing, and querying the "big data" stored in Hadoop. This solution uses all the strengths for which Ab Initio is known—for example:

- A completely integrated product suite
- A highly productive and maintainable graphical development platform
- The ability to run applications in parallel, with complete scalability, on heterogeneous clusters of all major platforms (Unix, Linux, Windows, mainframe)
- High performance for both batch and real-time processing
- Support for and integration with enterprise metadata via Ab Initio's metadata solution



## Ab Initio and Hadoop

Ab Initio can handle Hadoop data with the same power and speed it offers when processing other kinds of data—capabilities far beyond those available from other vendors.

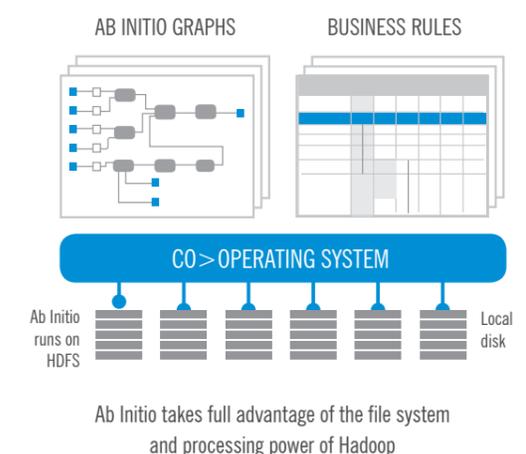
An Ab Initio application can run as many ways parallel as the data is distributed on the Hadoop cluster, maintaining data locality for optimal processing performance.

Ab Initio integrates with Hadoop sequence and text files and supports Parquet, RCFile, Avro, and ORC formats out of the box. In addition, Ab Initio jobs can run under the control of YARN and can be fully monitored by YARN.

Ab Initio software is certified to run on all standard Hadoop distributions (MapR, Cloudera, Hortonworks).

## Federated Queries

Ab Initio's Query>It lets you issue federated queries that can include Hadoop and all databases and files that Ab Initio can access.



Ab Initio software is a general-purpose data processing and metadata management platform. It has a single architecture for processing files, database tables, message queues, web services, and metadata. This architecture enables virtually any technical or business rule to be graphically defined, shared, and executed. It processes data in parallel across multiple processors, even processors on different servers. It can run the same rules in batch and real-time, and within a service-oriented architecture. It supports distributed checkpoint /restart with application monitoring and alerting. And it enables end-to-end metadata to be collected, versioned, and analyzed by nontechnical users.

The world's largest companies have chosen Ab Initio for their key applications. Here are some examples:

### Telecom

One of the world's largest phone companies uses Ab Initio software to process call detail information for call rating, data usage tracking, and network monitoring. Many billions of call-detail records per day are processed, as are millions of usage queries.

### Banking

At a major European bank, switching to Ab Initio reduced the time required to add new data sources to a data lake from 20 person-days of effort per source to under 5 person-days of effort — a 75% reduction.

### Retail

A major retailer uses Ab Initio software to coordinate all its inventory and sales-tracking applications. Ab Initio enables all the different software packages to seamlessly work together, allowing for better stock management and significantly improved theft prevention.

### Credit Cards

A major credit card network uses Ab Initio as its data backbone, processing and passing all transactions to back-end systems in batch and real-time. They accumulate over a petabyte of transaction data per year in an Ab Initio data retention system that supports customer service call centers with sub-second response times to queries.

### Stock Exchange

One of the world's largest stock exchanges converted millions of lines of Cobol code for mission-critical operations into Ab Initio applications. The solution connects to the real-time trading bus and processes transactions at over 500,000 messages per second. Overall performance is dramatically improved, and the stock exchange has opened new lines of business providing trading behavior data for stocks.



## CRÉDIT AGRICOLE CONSUMER FINANCE – MISE EN PLACE ET INDUSTRIALISATION D'UN DATALAKE EN « DATA BACKBONE »

### CONTEXTE ET ENJEUX

Crédit Agricole Consumer Finance est un acteur majeur du crédit à la consommation en Europe. Ancré au sein du groupe Crédit Agricole S.A., il occupe sur le marché français une position de premier plan sur l'ensemble des métiers du crédit à la consommation.

« CA-CF 3.0 » est son plan de **transformation digitale**, dont l'une des ambitions est de devenir une entreprise « **Data Centric** », notamment par l'implémentation d'un **Datalake**. Cette plateforme vise à répondre aux nouveaux enjeux métiers comme la personnalisation des relations client, en valorisant les données dans l'entreprise, et en facilitant l'implémentation d'usages opérationnels, décisionnels et DataScience.

### PRÉSENTATION DU PROJET

Le projet s'est déroulé en 2 temps.

**1) Implémentation de la plateforme BigData** (infrastructure et logiciels), en collaboration avec SILCA (entité groupe spécialisée). La plateforme est construite autour d'un cluster **MapR** (Hadoop), **d'HPE Vertica** pour BDD Analytics, et **d'Elasticsearch** pour BDD indexées. **Talend** complète le dispositif, notamment parce qu'il est déjà utilisé dans les projets BI par les Centres de Services Accenture. Avec ses administrateurs qualifiés et les pratiques standardisées, l'utilisation de Talend dans ce contexte a permis de **faciliter la transition des équipes** internes issues de la BI traditionnelle vers l'univers BigData, en simplifiant et en **accéléralant les développements** de traitements distribués, déployés industriellement.

**2) Industrialisation en un « Data Backbone » pour accélérer les projets Métiers.** Cette industrialisation a été opérée suivant la méthodologie « SCRUM Agile ». Elle s'est déclinée concrètement en : 1. définition des capacités nécessaires pour remplir les objectifs du « Data Backbone » ; 2. priorisation/planification de celles-ci sur une roadmap ; 3. réalisation des chantiers associés.



Figure 2 : Schéma d'Architecture Applicative CIBLE simplifiée

### LE RÔLE D'ACCENTURE

#### A. Priorisation et Réalisation des chantiers d'Industrialisation.

Une **équipe SCRUM** mixte (architectes internes, exploitants SILCA, experts BigData Accenture), définit l'architecture et accompagne les projets Métiers, mais surtout construit un **SOCLE** d'assets : normes (noms, arborescence applicative...) • procédures (livraison, déploiement, sauvegarde & restauration...) • référentiels (flux données, catalogue données, volumes...) • composants **applicatifs génériques** (ingestion, gestion logs, gestion tables, purge fichiers...)

#### B. Implémentation des projets Métiers.

Pour permettre des délais courts et des coûts maîtrisés, Accenture a proposé un dispositif ambitieux : une équipe d'ex-

perte à Paris, appuyée par nos centres de services de Nantes et de l'île Maurice. En huit mois, les résultats étaient : **8 projets en Production** : Monitoring Web, Ingestion données DMP, Segmentation client, Calcul Règles Métiers... • **une vingtaine prévus** : Décommissionnement de DWH/Datamarts, souscription par réseaux sociaux, Data Lab...

### RETOUR D'EXPÉRIENCE

Plusieurs facteurs sont clefs pour la réussite de projets BigData.

**Investir sur l'Expertise Technique** : C'est le prérequis des projets BigData : **il faut intégrer très tôt de l'expertise pointue**, capable de discuter les options d'architecture, de prototyper rapidement, et de développer les traitements les plus complexes.

**L'optimisation des traitements SPARK** est ainsi un domaine particulièrement complexe, pour lequel Accenture a développé **une expertise spécifique d'identification et d'optimisation** des lenteurs. Sur certains traitements, les **performances ont été améliorées de 30% à 60%**.

**Un SOCLE industrialisé** : Ce type de projet, dans un contexte de relative nouveauté (beaucoup d'Open Source, briques logicielles faiblement intégrées...), souffre généralement d'un manque d'industrialisation. Il est donc **important de construire très tôt un SOCLE** d'assets qui : **assure plusieurs niveaux de cohérence** (technique, méthodes, données) entre les différents projets • **réduit leurs délais et coûts** pour rapidement créer de la valeur Métier (au-delà des capacités technologiques). • **simplifie** les activités de **maintenance** et de **supervision**.

Crédit Agricole Consumer Finance s'est ainsi appuyé sur des outils comme Talend et sur des équipes de Delivery expertes comme celles d'Accenture.

**Une équipe Mixte** : Le passage aux technologies BigData s'accompagne aussi d'une transformation des compétences internes (nouvelles technologies, méthodes Agile...). L'organisation en équipe mixte, ici CA-CF / Accenture, est un facteur de succès qui : • **permet la co-construction dans un contexte technologique délicat et évoluant très vite** • **garantit la pérennité en favorisant la montée en compétences des ressources internes** • **fluidifie l'intégration et administratif par un accès direct aux ressources internes** (interlocuteurs clefs, documents legacy, procédures...), car les nouvelles architectures BigData sont souvent sources de blocages (données, flux, réglementation...)

#### Priorisation des Projets Métiers

C'est d'abord par les projets que le couple « Product Owner » / « Scrum Master » doit prioriser l'ingestion de nouvelles données, suivant une approche « MVP » incrémentale. On évite ainsi l'écueil du Datalake « à tout prix » pour fournir une valeur métier rapide : c'est au fur et à mesure des projets qu'on va intégrer de nouvelles données pour constituer **un SOCLE élargi avec des données cohérentes, gouvernées et stockées de manière optimisée**.

Enfin, les **bons résultats des premiers projets** en Production permettent de créer une **dynamique positive**, et incitent de **nouvelles directions métiers** à se lancer.

## LES ALGORITHMES DE RECONNAISSANCE VISUELLE S'AMÉLIORENT TRÈS VITE ET L'ANALYSE AUTOMATIQUE DE CE BIG DATA VIDÉO EST UNE SPÉCIALITÉ EN PLEIN BOOM. ACCENTURE ACCÉLÈRE AINSI L'INDUSTRIALISATION ET LA COMMERCIALISATION DE SON OFFRE VIDÉO ANALYTICS SERVICE PLATFORM (VASP)

### POURQUOI CETTE ACCÉLÉRATION SUR LA VIDÉO ANALYTICS ?

Nos centres de R&D, notamment ceux de Sophia Antipolis et Singapour, travaillent depuis plusieurs années sur le développement de cette plateforme, qui rencontre une demande en forte croissance. Une part de cette demande vient d'abord des villes et des services publics qui investissent massivement dans la supervision automatisée des caméras placées dans les lieux publics, afin de lutter contre la criminalité et le terrorisme. Par ailleurs, il est certain que les travaux récents sur le deep learning ont considérablement amélioré les capacités de l'analyse vidéo, aussi bien sur la qualité et la précision des algorithmes que sur l'extension du champ des possibles. Enfin, la démocratisation de l'intelligence artificielle libère la créativité des entreprises qui envisagent désormais le traitement des vidéos par l'IA comme un outil mature, ce qui ouvre de nouveaux cas d'usage, y compris dans des secteurs assez éloignés de la sécurité.

### POUVEZ-VOUS NOUS DÉCRIRE EN QUELQUES MOTS L'OFFRE ACCENTURE « VIDÉO ANALYTICS SERVICE PLATFORM » ?

Cette plateforme pré-intègre toutes les briques technologiques nécessaires à la récupération des flux vidéo, leur traitement de façon scalable par des modules spécialisés et évolutifs (par exemple : lecture de plaque, reconnaissance d'objets ou de sons...), le traitement des alertes et métadonnées générées, et enfin la présentation de ces événements, par exemple sur une carte. Entièrement virtualisées en containers Docker, toutes ces briques peuvent s'installer sur un cloud public, privé, ou directement dans l'environnement de nos clients : les flux vidéo requièrent souvent de rester sur un réseau isolé pour des contraintes de sécurité. Accenture Digital a récemment intégré cette plateforme au sein de son offre « Analytics as a Service » : Accenture Insight Platform (AIP). Elle bénéficie ainsi des standards de scalabilité, supervision, d'un provisionnement automatisé, mais aussi du haut niveau de sécurité que propose AIP.

### EN QUOI SE DIFFÉRENCIE-T-ELLE DES PRODUITS CONCURRENTS QUI EXISTENT SUR LE MARCHÉ ?

Justement ce n'est pas un produit ! Notre conviction, c'est que les modules de traitement de la vidéo s'améliorent tellement vite que seule une plateforme évolutive permet de proposer une solution qui reste à jour. Par exemple, dans deux ans, le meilleur module de reconnaissance de plaque d'immatriculation ne sera sans doute plus celui d'aujourd'hui. Construire un produit à partir d'algorithmes « maison », n'est plus compatible avec la vitesse d'évolution de ce marché. VASP pré-intègre donc des modules open source, des modules commerciaux, voire des modules « maison » lorsqu'il n'existe pas de solution prête à l'emploi. Dans de nombreux cas, nous construisons également des modules spécifiques pour nos clients. L'architecture interne de la plateforme et les différents algorithmes ont également été particulièrement étudiés pour optimiser les performances ce qui en fait une solution assez peu gourmande en CPU/GPU ou mémoire. Un serveur peut par exemple gérer jusqu'à 16 flux vidéo.

### ET QUELLE EST LA PART DE CES DÉVELOPPEMENTS SPÉCIFIQUES SUR VOS PROJETS ?

La plateforme offre un cœur robuste et stable, avec de nombreux algorithmes performants qui ne nécessitent pas de développement spécifique. Néanmoins il faut parfois ré-entraîner les réseaux de neurones pour s'adapter aux configurations spécifiques (lumière, angle, qualité de l'image), ou développer un algorithme particulier pour répondre à une demande qui ne serait pas standard. Certains clients peuvent aussi vouloir développer leur propre réseau de neurones pour un module qui leur serait spécifique : la plateforme peut intégrer ce modèle, simplifiant la mise en œuvre du deep learning qui est encore peu standardisée.

### AVEZ-VOUS DÉJÀ QUELQUES RÉFÉRENCES DONT VOUS POURRIEZ PARLER ?

Le secteur qui a le premier investi dans notre solution est celui de la sécurité des villes et des manifestations, et nous avons plusieurs retours d'expérience publiques dans ce domaine, pour la ville de Singapour ou la police de la ville de Mons en Belgique par exemple. En France, en collaboration avec la Police Nationale, le Service des Technologies de l'Information de la Sécurité Intérieure, ST(SI)<sup>2</sup> a été parmi les premiers à tester la plateforme lors de la Braderie de Lille. En proposant une nouvelle solution technologique en appui aux policiers, en améliorant son système actuel et en permettant une prise de décision en temps réel, le ST(SI)<sup>2</sup> a non seulement contribué à améliorer la sécurité publique, mais également avancé dans la construction des dispositifs de sécurité publique de demain.

### QUELS SONT LES DIFFÉRENTS CAS D'USAGE POSSIBLES DE CETTE OFFRE ?

Ils vont bien au-delà de la surveillance des villes. Nous avons plusieurs projets en cours dans des secteurs très différents. Par exemple en grande distribution : la reconnaissance démographique des clients (sexe, âge, etc.) et l'analyse de leurs comportements permettent de revoir l'agencement d'un magasin pour l'optimiser par rapport aux comportements types. Quatre grands domaines restent à court terme les plus prometteurs pour VASP : la ville intelligente, la distribution (agence, centre commerciaux, boutiques...), les transports (gares, aéroports...), et la sécurité (Police, gestion des frontières, ou des infrastructures critiques).

**Vous souhaitez en savoir plus ? Vous souhaitez rejoindre notre équipe de Data Scientists, mettre en pratique vos compétences Deep Learning jusqu'à la mise en œuvre de vos algorithmes ? Venez nous voir au Salon Big Data, ou contactez-nous !**

Jean-François GUILMARD  
j.guilmarde@accenture.com  
+33 6 32 27 70 22



## CINQ CAS D'UTILISATION ELASTIC ET LES SOLUTIONS ADELEAN

Depuis quelques années, Elasticsearch s'est imposé comme la technologie de choix pour les moteurs de recherche et l'analyse de données. Adelean intègre des solutions Lucene, Elasticsearch et leur écosystème depuis 2010 à l'époque où le marché de ces technologies était encore une niche. Avec, parmi vous, une vingtaine de clients par an depuis 2013 nous rencontrons autant de cas d'utilisations différents. Certains se distinguent par leur originalité, leurs challenges ou leurs solutions innovantes. Nous allons vous présenter ici cinq d'entre eux, cas d'usage et solutions qui vont vous donner un aperçu des capacités des suites logicielles d'Adelean et Elastic.

### L'APPROCHE ADELEAN

Pour qu'un moteur de recherche ou d'analyse de données soit pertinent et performant, Adelean prend en compte trois aspects :

- L'indexation et la modélisation des données
- La construction des requêtes de recherche et les services
- La configuration de la plateforme Elasticsearch elle-même

Le tout piloté par une API transverse et une Interface d'Administration puissante : la Business Console.

Perdre de vue un seul de ces aspects mène à des solutions partielles. Sans modélisation ni préparation des données avant l'indexation dans Elasticsearch, les possibilités de recherche et d'analyse sont limitées. Le langage technique d'Elasticsearch (le « Query DSL ») est très puissant, mais peut être aussi très complexe. Or la qualité de la requête a un impact important sur les performances et la pertinence de la recherche. Pour finir, une bonne configuration de la plateforme (analyseurs de texte, mapping et types de données, cache, etc.) améliore de beaucoup le résultat.

## QUELQUES CAS D'UTILISATION ET LES SOLUTIONS ADELEAN

### L'APPLICATION DE GESTION ET LA RECHERCHE DANS LA BASE CLIENT

Nous rencontrons souvent ce cas d'utilisation chez nos clients dans le monde de l'assurance, qui avec la banque correspond à 30 % de l'activité d'Adelean. Il s'agit de clients qui migrent leur architecture monolithique construite autour d'une base de données relationnelle vers une architecture plus ouverte, basée sur des microservices et sur un ensemble de systèmes de stockage choisi en fonction des besoins métier. On parle alors de persistance polyglotte. Introduire une technologie NoSQL comme Elasticsearch dans l'architecture d'entreprise présente de multiples avantages :

- Les performances sont maximisées impliquant une fluidité accrue des applications,
- L'implémentation est plus facile et les applications plus riches en fonctionnalités
- Le système est scalable et peut accueillir de plus en plus de données.

Adelean a construit des systèmes rendant possible, dès les premières phases d'un projet, le passage des Clients aux Contrats et aux autres Objets Métier puis à la GED ou même aux réseaux sociaux et internet. Une vision 360° est alors offerte aux utilisateurs de l'application.

### LA CATÉGORISATION ET L'ANALYSE DES TENDANCES

Nous utilisons Elasticsearch et son interface graphique Kibana pour des problématiques parfois réservées à l'intelligence artificielle : la catégorisation automatisée des contenus et l'analyse des tendances et sentiments (aussi appelée « Sentiment Analysis »). Dans ce cas d'utilisation, Elasticsearch a deux atouts majeurs :

- il est scalable et peut traiter des très gros volumes de données rapidement ;
- il bénéficie de toute la puissance de la technologie Lucene pour l'analyse du texte. Pour implémenter ce cas d'utilisation, nous utilisons trois fonctionnalités d'Elasticsearch que nous citons pour les plus techniques d'entre vous :
- Les agrégations « Terms » et « Significant Terms », ou leur alternative propriétaire « Graph API » de X-Pack, pour détecter les « top terms », mais aussi les fameux « signaux faibles » ou « maximum relatifs ».
- La « Percolate Query » utilisée pour tester des milliers de requêtes stockées dans un index spécifique et permettant de catégoriser chaque document.
- L'interface Kibana pour donner la main à l'utilisateur métier pour le pilotage, l'ajustement et le suivi des algorithmes et de leur pertinence.

Le savoir-faire d'Adelean permet l'utilisation pertinente et innovante de fonctionnalités avancées d'Elasticsearch.

### LE CAS D'UTILISATION « ENTERPRISE SEARCH APPLIANCE »

Ce cas d'utilisation est large. Il représente par ailleurs environ 30 % de nos projets. Pour un meilleur service, nous avons construit notre solution Adelean Enterprise Search sur trois piliers :

- **L'intégration dans votre SI.** Beaucoup de données d'entreprise sont difficiles d'accès. Les sites intranet sont développés sans contraintes SEO, si bien que leur indexation relève du « reverse engineering ». Les données sont éparpillées dans des sources hétérogènes : bases de données, SharePoint, partages réseau, e-mails, sites intranet, etc. Enfin, l'accès aux données est souvent protégé par des systèmes d'authentification et de gestion des droits. Adelean propose surtout des solutions qui s'adaptent à l'entreprise.

- **La sécurité.** Adelean prend ce point au sérieux. Les données dans une entreprise sont précieuses et doivent être protégées du vol ou de la divulgation. Des droits d'accès sont définis à tous les niveaux et dans tous les systèmes. Adelean Enterprise Search intègre les droits d'accès dans ses APIs de recherche. Un utilisateur n'aura accès qu'aux documents et informations auquel il a accès dans le système source.

- **L'enrichissement.** En collectant les données à la source, Adelean conserve la sémantique des données. Là où un site internet ou intranet ne présente souvent que du texte, Adelean peut se connecter à la base de données qui a généré le site et mieux identifier la nature du texte. Il est ainsi capable de distinguer un menu, un paragraphe, un article de blog ou une carte de visite.

### LE CAS D'UTILISATION E-COMMERCE

C'est peut-être le cas d'utilisation le plus classique et le plus ancien des moteurs de recherche d'entreprise. Et pourtant il laisse encore la place à beaucoup d'innovation. Il représente aussi 30 % des cas d'utilisation d'Adelean. Les particularités des moteurs de recherche e-commerce sont multiples. De nombreux facteurs entrent dans le calcul de la pertinence des résultats, en plus du texte de la fiche descriptive. La popularité, la nouveauté, le stock, la marge, le chiffre d'affaires ou les accords avec les fournisseurs sont tout aussi importants.

**Un moteur de recherche e-commerce est un moteur vivant.** La saisonnalité, les lancements des campagnes ou de nouveaux produits, la fluctuation du stock en temps réel sont autant des facteurs qui agissent sur la base des documents indexés ainsi que sur la pertinence et les performances du moteur. C'est pourquoi la **Business Console** d'Adelean permet de surveiller et d'ajuster en permanence les réglages du moteur pour de meilleurs résultats. Grâce à la couche de **Services**, Adelean permet de croiser les données du moteur de recherche avec des profils utilisateur transformant ainsi le moteur de recherche en moteur de recommandation.

Pour de meilleurs résultats, Adelean a mis l'accent sur la couche d'**Indexation**. Elle alimente le moteur avec toutes les données nécessaires que ce soit les fiches produit, les stocks, les prix, mais aussi l'historique des recherches. En croisant les données avant l'indexation, et non pendant la recherche, elle garantit des temps de réponse très faibles. De plus, dans un contexte de fournisseurs de type « Marketplace » la couche d'Indexation permet d'homogénéiser les fiches descriptives de produits provenant de plusieurs fournisseurs. Il est même possible d'indexer d'autres types de contenus, comme du contenu éditorial, à côté des produits.

### L'ANALYSE DE DONNÉES ET LA BI

Elasticsearch et son écosystème (comme Logstash, Beats ou Kibana) sont aussi utilisés dans l'analyse de données et l'analyse « temps réel ». Ils se sont d'ailleurs créés une forte réputation dans l'analyse des logs. Avec des fondateurs venant du monde de la BI (et ayant développé les produits de Business Objects et SAP), Adelean maîtrise le traitement des données et leur visualisation de manière optimale. Plutôt que de se contenter d'envoyer les données à Elasticsearch de la manière la plus directe possible, nous menons un travail de préparation et modélisation des données en amont. La quasi-totalité de nos projets compte des mini-ETLs s'occupant de l'agrégation, de la consolidation, du nettoyage, de l'enrichissement et de la modélisation de données avant l'indexation par le moteur de recherche. Notre « tagline » est « Extract, Transform and Search ». Et c'est pour cela que nos clients et partenaires reconnaissent nos moteurs comme encore plus performants et pertinents.

## DE L'INTÉGRATEUR ELASTICSEARCH AU FUTUR (ET MEILLEUR) « GOOGLE SEARCH APPLIANCE »

### COMMENT SE DÉFINIT ADELEAN ? UNE SOCIÉTÉ DE CONSEIL ? UN INTÉGRATEUR ? UN ÉDITEUR DE LOGICIELS ?

Nous sommes tout ceci à la fois. Nous avons démarré en 2010 en tant que société de conseil autour des moteurs de recherche. Très vite nous avons misé sur la technologie Elasticsearch et son écosystème dont nous sommes devenus l'intégrateur de référence en France. Nous développons actuellement une solution verticale pour l'indexation, la recherche et l'analyse de données d'entreprise.

### QUI SONT LES CLIENTS D'ADELEAN ?

Depuis 2013, nous avons une vingtaine de clients par an. Nous sommes basés en Île-de-France et desservons toute la France de Bordeaux à Lille et de Rennes à Metz. Nous avons également quelques clients en Belgique et Canada. Nous travaillons pour de grands groupes, mais également pour des startups, des associations. Parfois, des freelances ou de très petites entreprises nous sollicitent aussi. Nous sommes très impliqués dans la relation avec nos clients. Nous aidons nos sponsors à évoluer dans leurs projets et fonctions, les startups à lever des fonds, les petites entreprises à avancer.

### QUELLE EST LA RELATION D'ADELEAN AVEC ELASTIC ?

Nous suivons Elasticsearch et Elastic depuis très longtemps. Nous nous sommes spécialisés sur l'intégration et le conseil autour de la Suite Elastic depuis 2011 et, depuis la même époque, les projets autour de ces technologies et leur écosystème (Open Source ou commercial) représentent 90 % de notre activité. Nous avons vu grandir la technologie, et la société Elastic nous a permis de grandir à son tour. Nous participons régulièrement aux événements et formations qu'elle organise. C'est nécessaire pour rester à jour compte tenu du rythme soutenu des développements et la fréquence de sortie des versions. Adelean a été sponsor des deux éditions de la conférence Elastic{ON} à Paris et a fait trois fois le déplacement à l'édition de San Francisco. Cela dit, Adelean n'est pas affilié à Elastic et reste un acteur indépendant. Notre point de vue externe est très apprécié par nos clients. Il nous arrive de préconiser ou de mettre en œuvre des technologies concurrentes lorsqu'elles sont plus pertinentes.

### ELASTIC EST-IL EN TRAIN DE DEVENIR UN CONCURRENT D'ADELEAN ET VICE-VERSA ?

L'équipe d'Elastic s'est étoffée ces dernières années, surtout au niveau commercial. De l'éditeur fournissant des logiciels, du support et de la formation, Elastic propose maintenant des prestations de consulting et a ses propres équipes d'architectes solutions. A l'inverse Adelean, historiquement positionné sur le conseil et l'intégration, propose maintenant une solution logicielle, du support et de la formation. Mais nos offres restent complémentaires : Adelean Enterprise Search étend Elasticsearch et Elastic fourni le cœur du moteur d'Adelean. Elastic propose des formations publiques et exhaustives, Adelean des ateliers intra-entreprise et adaptés à des cas d'utilisation spécifiques. Elastic commercialise un support de production 24/7, Adelean fourni un support de proximité et réservé aux clients ayant auparavant fait appel aux prestations de conseil et intégration.

### EST-IL VRAI QU'IL N'Y A PAS DE FONCTION COMMERCIALE CHEZ ADELEAN ?

Chez Adelean nous mettons l'accent sur l'excellence technique. C'est la qualité du travail de nos ingénieurs qui fait vendre nos prestations, souvent par le bouche-à-oreille. Cela dit, nous dépendons beaucoup des forces commerciales de nos partenaires. Ce sont nos clients, nos collègues ainsi que nos partenaires qui développent les aspects commerciaux, permettant à Adelean de rester concentré sur la technique.

### EN QUOI LA SOLUTION D'ADELEAN EST MEILLEURE QUE L'ANCIENNE « GOOGLE SEARCH APPLIANCE » ?

Google a appris aux entreprises à rendre leur site performant et « responsive », à utiliser HTTPS, à ne pas dupliquer les contenus... Il a mis à leur disposition la « Search Console », un outil référençant les erreurs et donnant les clés pour leur correction. Google a surtout appris aux entreprises comment s'adapter à son moteur. Dans le monde des données d'entreprise, ceci fonctionne moins bien. Les données d'une entreprise se trouvent dans des sources hétérogènes (bases de données, SharePoint, partages réseau, e-mails, sites intranet, etc.) souvent impossibles à modifier ou à adapter à un quelconque crawler, indexeur ou moteur de recherche. C'est donc au moteur de s'adapter à cette diversité, à faire le maximum pour extraire les données de là où elles sont sans exiger des modifications à la source. Et c'est ce que notre solution Adelean Enterprise Search fait. Arrivent ensuite les contraintes de sécurité qui sont beaucoup plus fortes en entreprise.

### EN QUOI LES CONTRAINTES DE SÉCURITÉ COMPLEXIFIENT-ELLES LA VIE D'UN MOTEUR DE RECHERCHE ?

Lorsque nous indexons les données internes d'une entreprise, des problématiques de sécurité se posent sérieusement. Il y a d'abord l'accès aux données : souvent les mécanismes d'authentification sont complexes et hétérogènes (il faut naviguer entre SSO, authentification HTTP, certificats, SAML, JWT, Kerberos, etc.). Il faut ensuite reproduire la hiérarchie de rôles et droits d'accès dans le moteur d'indexation pour que, lors de la recherche, les résultats soient filtrés selon les droits de l'utilisateur qui l'a lancée. Certaines problématiques se posent aussi dans le cas extrême où le moteur lui-même n'a pas le droit de stocker une donnée en clair. Il y a, dans ce cas, un compromis à faire entre les fonctionnalités du moteur et la sécurité des données.

### QUELS SONT LES PLUS D'ADELEAN POUR L'INDEXATION DES DONNÉES D'ENTREPRISE ?

Nous avons déjà la capacité de nous connecter à beaucoup de sources de données. Si un connecteur nous manque ou si une source nécessite une attention particulière et spécifique, nous le développons avec notre client. La vision « bout en bout » (indexation, configuration, recherche) avec une console d'administration (Business Console) est aussi un atout pour nous. Nous pouvons ensuite mentionner l'indexation « temps réel » de certaines sources de données, la génération des aperçus et vignettes, l'extraction des entités nommées, l'OCR, la catégorisation et le rajout de la sémantique. Ainsi que toutes les fonctionnalités de type ETL et notamment le nettoyage et l'homogénéisation des données hétérogènes. Pour finir, nous avons fait le chemin inverse de Google : nous sommes partis des moteurs de recherche pour le e-commerce, les bases de données client, les GED – des données souvent très structurées – pour arriver aux données non structurées ou semi-structurées. Notre premier réflexe est donc de structurer et de rajouter de la sémantique dans les données que nous indexons : les étapes en amont et en aval du moteur de recherche lui-même font partie intégrante de notre solution.

Lucian Precup est CEO d'Adelean et développe des solutions basées sur la Stack Elastic depuis 2011. Adelean est partenaire intégrateur d'Elastic depuis 2013 et développe des solutions verticales pour l'indexation, la recherche et l'analyse de données d'entreprise. Auparavant et avec une équipe de chercheurs de l'INRIA, Lucian a développé des logiciels pour l'intégration de données en temps réel chez Business Objects et SAP.

# ADVANCED SCHEMA

**HERVÉ SEBAN**

Président

+ 33 (0)1 85 09 82 90  
contact@advanced-schema.com  
19 rue Galilée  
75116 Paris France  
www.advanced-schema.com



**advanced**  
schema



# NODATA / GROUPE ADVANCED SCHEMA NICOLAS BRIGITTE-ALPHONSINE

Directeur Général / Directeur R&D

INTERVIEW



INTERVIEW

**ADVANCED SCHEMA** est une Entreprise internationale de Services du Numérique (ESN), présente en Europe et en Amérique du Nord, qui intervient sur 5 domaines d'activités stratégiques : la Business Intelligence, le Data Warehousing, le Big Data, le CRM (Gestion de la Relation Client) et le Digital.

Ainsi, et depuis plus de 15 ans, ADVANCED Schema accompagne et conseille ses clients dans leurs projets de refonte de leurs systèmes d'information et dans la mise en œuvre de systèmes de pilotage et de reportings performants.

**ADVANCED SCHEMA** se positionne ainsi comme une « end-to-end company » intervenant de l'hébergement au conseil en passant par le développement d'applications, l'intégration et l'édition logicielle.

Dans un univers toujours plus digitalisé où la donnée est au cœur des problématiques métiers, le Groupe ADVANCED Schema s'est structuré afin de pouvoir proposer à ses clients des offres adaptées à ces nouveaux enjeux.

Une organisation nouvelle, illustrée notamment par la création d'un pôle R&D en 2011 et par l'acquisition de la société **Kernel 42** spécialisée dans le Digital en 2014.

Afin de proposer une offre qui soit la plus pertinente, notre équipe R&D décline son savoir-faire à travers deux offres distinctes :

- le « Lab as a Service », véritable plateforme répondant aux besoins d'expérimentation et de flexibilité de ses clients.
- le mode « Tailored » pour un développement à façon d'applications décisionnelles et Big Data.

Dernière-née du groupe, la société NODATA, spécialisée dans l'édition de logiciels, présente **NODATA une solution intégrée d'outils autour de la donnée**. Alliant modularité, performance et maîtrise des coûts, la suite NODATA est le fruit d'un mariage entre le savoir-faire historique d'ADVANCED Schema et des compétences nouvelles développées au sein de son pôle R&D.

## LORS DE L'ÉDITION 2017 DU SALON BIG DATA, ADVANCED SCHEMA PRÉSENTAIT LA PLATEFORME NODATA, ALORS EN PHASE DE PRÉ-COMMERCIALISATION. POUVEZ-VOUS NOUS (RE)PRÉSENTER LA PLATEFORME NODATA ?

ADVANCED Schema a créé la Plateforme NODATA dans le but de couvrir l'intégralité des besoins analytiques des entreprises en matière de data. NODATA est une suite complète dédiée à la manipulation de bout en bout de la donnée. De l'intégration à la restitution en passant par la cartographie des données.

Entièrement créée par la R&D d'ADVANCED Schema, en plaçant les dernières innovations technologiques au cœur de l'architecture de la solution, la suite NODATA est une plateforme permettant l'industrialisation des modèles data driven.

En plus de ses performances inégalées et de son approche ergonomique, NODATA apporte un ensemble d'outils qui facilite la collaboration autour des indicateurs clés des entreprises. Le modèle de pricing et de distribution est lui aussi innovant.

Comme nous participons aujourd'hui à la transformation digitale des outils de pilotage de nos clients, les départements IT peuvent s'appuyer sur NODATA pour mettre en place les nouvelles architectures orientées DATA.

## QUELLE EST LA MATURITÉ DE L'OFFRE NODATA ? OÙ EN ÊTES-VOUS DANS LE PROCESSUS DE COMMERCIALISATION ?

L'engouement est tel que 6 grands groupes ont déjà adoptés NODATA dans sa phase de pré-commercialisation dont 3 sont ambassadeurs de la solution, en adhérant au programme d'« early adoption » : RCI Bank & Services, Renault et Swiss Life.

NODATA a relevé les challenges d'un usage intensif et d'environnements de production complexes. Les contraintes imposées par les normes de sécurité et les standards de disponibilités ont facilement été remplis par la solution.

NODATA évolue perpétuellement dans le sens des usages de nos clients. L'interaction avec nos ambassadeurs et nos clients nous permet de challenger la réponse qu'apporte NODATA à leurs problématiques quotidiennes.

## COMMENT S'INSCRIT NODATA DANS LA STRATÉGIE D'ADVANCED SCHEMA ?

NODATA est essentielle dans notre stratégie. Depuis 16 ans, pour délivrer des projets répondant au mieux aux besoins des clients, nous avons toujours privilégié le choix des outils les plus adaptés. C'est la raison pour laquelle nous sommes partenaires de la plupart des grands éditeurs de logiciels. Or, ces dernières années, nous avons vécu plusieurs ruptures technologiques (temps réel, adressage de très fortes volumétries, traitements distribués...). Elles ont rendu quasiment tout possible en matière d'analytique. Il fallait donc une nouvelle génération d'outils qui soient tout à la fois intégrés, modulaires et permettant en natif, le développement en mode agile et le travail collaboratif. Il fallait aussi associer à cette évolution technologique, une distribution et un type de pricing innovant.

Parce qu'il ne s'agit pas pour nous de devenir mono-technologie, nous continuons à intégrer les solutions de nos partenaires éditeurs. Néanmoins, parallèlement, nous avons désormais une offre alternative, NODATA. Cette solution, portée par une société d'édition éponyme, a elle aussi vocation à avoir plusieurs partenaires intégrateurs, ADVANCED Schema en étant logiquement le premier.

## FOCUS SUR LES FILIALES DU GROUPE ADVANCED SCHEMA

**nodata**

Réinventez l'exploitation de la donnée avec **NODATA**, nouvelle filiale du groupe dédiée à l'édition de logiciels. La suite **NODATA (Not Only Data)** offre la possibilité d'allier innovation technologique, performance et expérience utilisateur pour des coûts maîtrisés.

Cette solution complète intégrée et modulaire couvre aussi bien la transformation de la donnée, la modélisation et la data visualisation...



ADVANCED Schema lance Connexin, sa filiale spécialisée dans la santé connectée.

Première réalisation : une plateforme éponyme qui permet la communication entre professionnels de santé.

Elle a été développée afin de faciliter la communication dans le lien ville-hôpital entre professionnels de santé en utilisant les canaux digitaux.



**Kernel 42** étend l'expertise d'ADVANCED Schema aux projets digitaux : Web, e-commerce, mobilité et Big Data.

La combinaison de ses expertises digitales et Data (Big Data, Business Intelligence) permet de concevoir des dispositifs digitaux intelligents et parfaitement intégrés aux Systèmes d'Information.



CONTACT

**DE LA NÉCESSITÉ D'ABORDER LES ENJEUX LIÉS À LA DATA DANS L'ENTREPRISE AVEC UNE APPROCHE « OPEN MIND » SOUS PEINE DE PASSER À CÔTÉ DE SON SUJET.**

Les problématiques en termes de Data ont beaucoup évolué depuis quelques années, c'est une évidence. On le constate au quotidien : le simple fait d'évoquer la question avec nos clients puis dans un second temps d'aller dans le détail de ses besoins provoque l'ouverture d'une succession de portes que ce dernier ne pensait même pas franchir.

Certains ignoraient même jusqu'à l'existence de ces portes. Par ailleurs, dans le domaine de la data d'entreprise, les enjeux sont rarement simples, à moins de rester en surface et d'ignorer le contexte opérationnel du (des) métier(s) de son client.

Acteur incontournable de la finance, une filiale spécialisée dans le crédit-bail de ce grand compte s'est tournée vers nous à l'origine pour pouvoir répondre à une problématique de conquête commerciale, sous-tendue par le constat que sa base de prospects était loin d'être performante.

*Pour des raisons stratégiques, notre client, J. D., a préféré témoigner de façon anonyme.*

J.D. : « Le -à l'origine- est important car passées les phases d'échanges itératifs puis d'audit, les réflexions (menées par une task force mixte client/Altares) nous ont dévoilé des possibilités d'amélioration insoupçonnées. J'ai beaucoup milité en interne pour aborder ce projet stratégique en étant « open-minded » ; ce n'était pas évident car certains pragmatiques comme le DSI ou le directeur financier ne voyaient pas ce saut dans l'inconnu (technique et financier pour ce qui les concerne) d'un très bon œil ! L'autre difficulté pour moi résidait dans le fait que la data « n'appartient » à personne en particulier dans l'entreprise et personne n'en est responsable ou comptable, comme pourrait l'être un Chief Data Officer dans d'autres structures.

Néanmoins, en tant que directeur commercial data-compatible, j'ai pu convaincre, assez rapidement et je l'en remercie, la Direction Générale d'aller dans cette voie. »

Comme on peut le voir, quand on traite avec une entreprise peu aguerrie dans le domaine, un projet qui implique de la data, on sait quand ça commence...mais c'est à peu près tout ! Altares s'est souvent retrouvé en face de prestataires concurrents extrêmement compétents mais spécialisés dans une seule technologie ou une approche unique et donc limitée. Notre force est d'abord d'avoir des collaborateurs spécialistes des métiers des clients qui viennent nous solliciter et ensuite de disposer, pour des problématiques de prospection commerciale ou même d'animation marketing d'une agence « in-house » avec Manageo qui pratique le data marketing depuis plus de 15 ans.

J.D. : « Au-delà de sa connaissance de la gestion de la data, c'est en effet cette combinaison de connaissance métiers et de savoir-faire éprouvé dans l'optimisation de la prospection commerciale qui nous

a poussé à choisir Altares. La façon de travailler en mode ateliers de cadrage (parfois assez éprouvants pour nous), puis l'accompagnement sur la durée et à chaque étape, nous a également rassuré. »

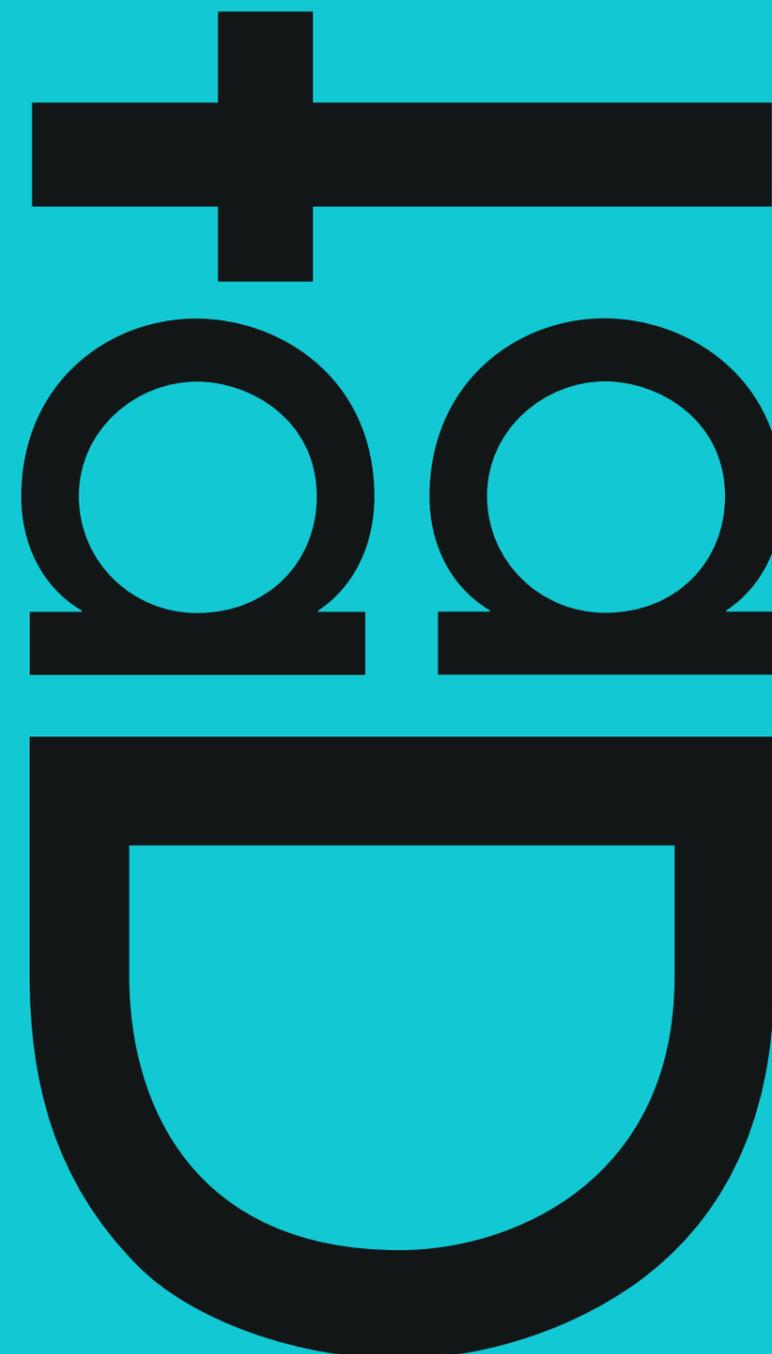
L'autre valeur ajoutée qui a eu son importance et qui a été décisive dans la satisfaction de notre client est l'apport et l'agrégation de data externes et à jour dont dispose Altares et qui constitue, par ailleurs, son activité historique. Ces données, sur les secteurs, les marchés, les entreprises et leurs contacts respectifs ont permis de construire des modèles et cartes de score d'appétence.

La deuxième phase a consisté à paramétrer et à activer les outils de data-marketing pour la prospection proprement dite, via la nouvelle solution mList de Manageo, filiale du Groupe Altares.

Le troisième volet de la collaboration a consisté à développer un outil de data-visualisation afin que les managers et leurs équipes commerciales puissent avoir en temps réel un reflet de leurs actions et la possibilité de piloter ces dernières en fonction des priorités, des segmentations et des options tactiques et stratégiques.

J.D. : « La seule dataviz n'aurait pas suffi à prendre les bonnes décisions ; au point où nous en étions en termes de connaissance data, il était facile de faire fausse route. Les équipes d'Altares ont donc appuyé le déploiement de leur outil par des conseils réguliers sur la façon de générer du lead et de prioriser nos actions. Pour résumer, les équipes d'Altares ont construit le camion, nous ont appris à le conduire et maintenant, ils nous laissent les clés, ça nous va bien ! »

On voit donc émerger très régulièrement le triptyque data science/data marketing/data visualisation qui semble répondre à une grande partie des problématiques liées à la data dans les entreprises, le ciment de tout cela étant assuré par du conseil et de l'accompagnement.



**La data n'a aucune valeur...  
si nous n'en faisons rien**



## GÉRARD GUINAMAND, DIRECTEUR D'ENGIE DIGITAL, RACONTE...

« Nous faisons face à une véritable révolution du monde de l'énergie qui vient bouleverser notre business model. ENGIE a anticipé cette évolution et a engagé une véritable transformation pour devenir le leader mondial de la transition énergétique. Le digital est l'une des clés pour atteindre notre objectif, via les outils s'appuyant sur l'IoT, le big data... L'utilisation du Cloud est un levier essentiel pour proposer de nouvelles applications et les déployer à l'international, c'est pourquoi nous avons retenu AWS, Amazon Web Services, leader incontesté dans son métier. »

### PROBLÉMATIQUES

- Accélérer le time-to-market des nouveaux produits et applications
- Disposer de services à la demande, rapides et flexibles
- Pouvoir déployer des services facilement, à grande échelle, à l'international

### QUAND TRANSITION ÉNERGÉTIQUE ET TRANSITION DIGITALE VONT DE PAIR

Pour réaliser sa transformation digitale, ENGIE se base sur 4 piliers : proposer des produits et des applications à haute valeur digitale à ses clients, s'associer à des partenaires stratégiques pour faciliter leur mise en œuvre, être Cloud et API « first » pour être flexible, et enfin veiller à la cybersécurité des produits développés.

« Nous avons logiquement choisi AWS afin de nous accompagner dans cette transition, explique Gérard Guinamand. Nous sommes un grand groupe international, il faut que nous travaillions avec les meilleurs experts de chaque secteur pour bénéficier de leur expérience. Pas de place à l'expérimentation, nous devons nous assurer d'avoir le partenaire adéquat pour nos projets cloud, afin de les déployer rapidement et efficacement, dans le monde entier. »

### RÉDUIRE LE « TIME-TO-MARKET »

Le Cloud AWS permet désormais au Groupe de développer des applications et des produits accessibles au plus grand nombre, dans tous les pays où il est présent, de façon universelle. « Nous avons également besoin de flexibilité, d'une capacité à voir grand et à donner un accès sécurisé à nos développeurs, partout dans le monde », précise Gérard Guinamand.

À ces besoins s'ajoute la rude concurrence à laquelle se livrent les acteurs du marché de l'énergie. « Il faut pouvoir mettre des produits en service très rapidement, avant les autres. Grâce à notre plateforme de développement Cloud qui s'appuie sur l'infrastructure AWS, nous sommes désormais capables de sortir de nouvelles solutions très vite et à grande échelle. À titre d'exemple, le développement de nos cinq dernières applications a débuté en début d'année. 6 mois plus tard, elles étaient toutes mises sur le marché, un vrai changement par rapport au passé ! », juge Gérard Guinamand.

Parmi ces services, ENGIE propose par exemple Clara Domus en Italie, dédiée au smart building, mais aussi une application destinée aux clients disposant d'un compteur Linky en France, afin de maîtriser leur consommation d'électricité, ou encore Delta, qui analyse le comportement des centrales thermiques de façon à augmenter le rendement et à faire de la maintenance prédictive. « Les outils d'analyse mis en place nous permettent par exemple de trouver les causes de différences de performances entre deux centrales électriques », précise Gérard Guinamand.

### DÉPLOYER DES SERVICES IOT AVEC AWS : L'EXEMPLE DE CLARA DOMUS

Clara Domus permet aux gestionnaires d'immeubles de suivre en temps réel le fonctionnement d'un bâtiment équipé de capteurs et de savoir comment il se comporte au quotidien, pour ainsi relever les anomalies et en tirer des plans d'action. 500 bâtiments sont actuellement connectés, avec plus de 16 000 capteurs. À terme, ils seront plus de 3 500 bâtiments connectés avec 160 000 capteurs. Clara Domus s'appuie sur l'infrastructure AWS ainsi que sur C3 IoT, partenaire d'AWS en matière de big data et d'analytics. « Nous utilisons C3 IoT comme PaaS pour construire nos propres applications complétées par AWS IoT, qui collecte les data, prend en charge une partie de l'intégration des données, avant de se connecter à C3 IoT, détaille Gérard Guinamand. Le choix du Cloud permet de maîtriser l'effort financier lié à une telle transformation digitale : « les services s'adaptent à nos besoins, sans surconsommation ».

Enfin, avoir une vraie logique de partenariat est un point capital pour ENGIE Digital. Des experts AWS sont présents en permanence pour former les équipes et transférer la compétence. « L'objectif est de faire en sorte qu'au bout des 3 ans de contrat, nous puissions faire croître et embellir ce que nous avons construit avec une certaine autonomie. »

### DES SERVICES ADAPTÉS

« Nos derniers services n'auraient clairement pas pu sortir avant que l'on n'utilise le service Cloud d'AWS. Nous avons nos propres serveurs, donc il aurait fallu tout mettre en place, à la fois en termes de plateforme et d'infrastructure... On voit tout de suite le temps que ça peut prendre », conclut Gérard Guinamand.

### BILAN

**5 nouveaux produits sortis en seulement 6 mois**

**À terme, 160 000 capteurs installés sur des smart buildings italiens pour en recueillir les données**

**Un investissement global d'1,5 milliard d'euros sur le digital**

# 80% des entreprises du CAC40 ont un point commun

## Leur choix du Cloud





## ATTUNITY - GESTION DU BIG DATA ET SOLUTIONS D'INTÉGRATION DE DONNÉES

Provides Universal Big Data Availability to a Fortune 100 Automotive Maker



### 4,500 OLTP, ERP, CRM Systems

#### RÉSULTATS

- Consolidation d'un Data Lake global
- Surveillance centralisée de toutes les tâches de réplication
- Minimisation du travail et réduction des coûts
- Réalisation d'analyses rapides et avantage concurrentiel

#### AVANTAGES SUPPLÉMENTAIRES

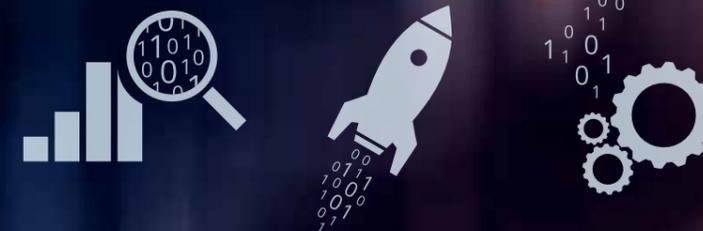
- Historical Insight
- Maintenance planifiée
- Innovation Design
- Primes Assurance
- Modèles de pricing
- Garanties

- Real-Time Analytics
- Maintenance prédictive
  - Ville connectée
  - Infotainment
- Conduite autonome
- Optimisation des trajets

Right Data.  
Right Place.  
Right Time.

## Data Integration & Big Data Management

- Hadoop & Big Data
- Real-time streaming data into Kafka
- Databases & Data Warehouses
- On premise & in the cloud





Blue DME



### BLUE DME AMBITIONNE DE DEVENIR LE PREMIER ACTEUR DE L'ANALYSE PRÉDICTIONNELLE DANS LES VENTES COMPLEXES B2C.

Nous éditons une plateforme SaaS d'intelligence artificielle pour améliorer la performance des forces marketing et commerciales pour les entreprises qui ont un cycle de vente complexe.

### MAIS QU'EST-CE QU'UN CYCLE DE VENTE COMPLEXE ?

Un produit complexe ? Notre solution est efficace pour toutes les entreprises qui ont : des équipes commerciales qui qualifient / négocient / contractualisent, un produit qui n'a pas réellement de prix catalogue et multi-caractéristiques (voir une offre personnalisée), qui demande un investissement important de l'acheteur et donc une phase de recherche et de réflexion, une approche multi-canal avec des parcours clients complexes (site web, applications, boutiques physiques, call-center), etc.

### QU'EST-CE QUE VOS ÉQUIPES ET VOTRE ENTREPRISE Y GAGNENT ?

Notre solution permet d'agir très opérationnellement sur la marge commerciale (pricing dynamique), d'augmenter les taux de conversion pour les commerciaux (score d'appétence sur les produits et services associés, recommandations temps réel personnalisées pour les acheteurs finaux), de réduire les coûts d'acquisition (score de joignabilité) et plus globalement d'améliorer l'efficacité opérationnelle par une connaissance client à 360°.

### LE PRICING : VOTRE LEVIER LE PLUS PUISSANT SUR LA MARGE. MAINTENANT REINVENTÉ PAR LE BIG DATA ET INDUSTRIALISÉ.

La plateforme Blue DME vous permet de travailler sur le levier le plus intéressant pour votre business : un pricing dynamique, entre les mains des équipes opérationnelles pour une échelle industrielle et basée sur l'ensemble des données que vous possédez.

### METTRE RAPIDEMENT LES RESULTATS DE DATA SCIENCE ENTRE LES MAINS DES EQUIPES MARKETING ET COMMERCIALES.

Le prochain défi des équipes marketing et commerciales n'est pas d'expérimenter sur des projets de data science et des modèles prédictifs mais bien d'industrialiser la mise en production de ces modèles. Passer d'un cycle mensuel ou semestriel à une dynamique de déploiement automatisée pour laisser le métier mettre les bons « insights » entre les mains des forces commerciales.

### L'EXEMPLE D'UN CONSTRUCTEUR AUTOMOBILE : ESHOP, RÉINVENTER L'EXPERIENCE CLIENT OFFLINE POUR MIEUX VENDRE.

Lors du projet d'un constructeur automobile français en Inde, Blue DME a accompagné la mise en place de boutiques eShops en shopping-malls pour ajouter aux points de vente physiques une plateforme « digitale » d'accélération des ventes. Ainsi, la solution développée par Blue DME est proposée comme élément central de la plateforme de collecte et d'analyse des événements concernant tous les visiteurs et prospects de Renault susceptibles de visiter ces eShops. La solution permet ainsi aux équipes commerciales et marketing du constructeur d'augmenter leur performance de conversion et de réduire leur délai de vente par une meilleure personnalisation de l'expérience client à partir des données prospects et client online et offline récupérées en temps-réel.

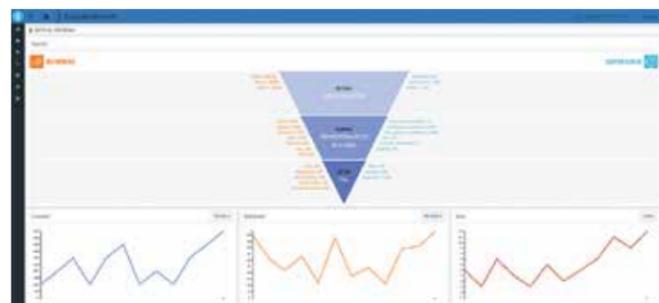
### L'EXEMPLE DE PGA MOTORS, DISTRIBUTEUR AUTOMOBILE : ADRESSER TOUTES LES ETAPES DU CYCLE DE VENTE POUR CRÉER UNE EQUIPE COMMERCIALE ORIENTÉE RESULTAT.

La solution déployée collecte en temps réel les données omnicanal (web, call center, concessions automobiles), et met en production de puissants algorithmes de matching de données. Le data lake de PGA Motors est alors « augmenté » automatiquement de modèles prédictifs pour fournir des insights impactant l'intégralité du processus de vente : quels types de véhicules intéressent le prospect (algorithmes de recommandation) ? Le prospect est-il prêt à conclure une vente dans les prochains jours (scores de chaleur) ? Quand faut-il l'appeler, sur quel canal (scores de joignabilité) ? Ces « insights » s'intègrent directement aux outils opérationnels existants de PGA Motors (Dealer et Lead Management system, site web autosphere.fr, etc.). Blue DME a créé des modèles prédictifs qui complètent la solution : scores de réachat, segmentation client, pricing, module data privacy, etc.

### VERS LE COMMERCIAL AUGMENTÉ ET LES COLLABORATIVES BOTS.

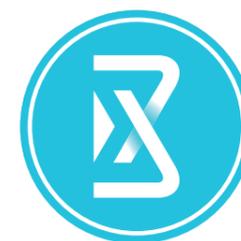
Notre prochain objectif ? Accompagner vos équipes de vente et marketing avec des « collaborative bots » qui pourront prendre la main lorsque c'est nécessaire et absorber la charge (pic d'activité, problématique de temps de traitement d'un prospect).

### EN ROUTE VERS LE SMART SELLING !



# Et si l'intelligence artificielle améliorerait les performances de votre équipe commerciale ?

## #DealClosing #DynamicPricing #NextBestActions



# Blue DME

WE BUILD YOUR AI-DRIVEN BUSINESS

#Assurance #Automobile #Industrie #ProduitsComplexes



@Blue\_DME



bluedme.com



### BLUESCALE AIDE SES CLIENTS A METTRE EN ŒUVRE LES NOUVEAUX LEVIERS D'INNOVATION TOUT AU LONG DE LA CHAÎNE DE VALEUR DU BIG DATA

Bluescale est composé d'une équipe unique de près de 60 consultants, développeurs et ingénieurs de la donnée, maîtrisant les dernières technologies et les nouveaux paradigmes de la data, avec pour seul objectif de mettre en place de façon industrielle et agile les briques technologiques de la chaîne d'information de demain.

### Bluescale intervient tout au long du cycle de vie de vos projets :

- Accompagnement fonctionnel et gouvernance de la donnée
- Conception et mise en œuvre d'architectures big data assurant performance, qualité, sécurité, traçabilité, agilité et innovation
- Mise en place de processus d'industrialisation de prototypes ou de modèles
- Intégration de solutions de dataviz, de mise en valeur et partage de la connaissance
- Développement d'applications agiles Data Driven opérationnelles
- Conduite du changement, formations

**Bluescale, c'est plus de 10 ans d'expérience dédiés à la data et à l'innovation.**

### BLUESCALE A DEVELOPPE DEPUIS PLUSIEURS ANNEES UN HAUT NIVEAU D'EXPERTISE SUR SPOTFIRE, AU POINT DE DEVENIR UN ACTEUR INCONTOURNABLE SUR CETTE SOLUTION.

Les consultants de Bluescale, au profil de data engineer, disposent des compétences clés pour appréhender un outil aussi riche que Spotfire :

- une **expertise dans les technologies Big Data** (Hadoop, Impala, Spark...) ou dans les systèmes de stockage plus traditionnels, comme les moteurs relationnels
- une expérience forte dans la **collaboration avec les data scientists**, mise à profit pour implémenter les analyses statistiques, développées en R et exécutées au travers du moteur intégré nativement à l'outil
- une **polyvalence technique**, permettant de tirer entièrement partie de l'outil en y intégrant des scripts très élaborés (Python, R, Javascript)

Les consultants de Bluescale interviennent sur toutes les étapes de la **transformation digitale des entreprises** : l'expérience en data management est un véritable atout lors de la mise en place d'un projet de data visualization.

Au-delà de leurs compétences techniques, les collaborateurs de Bluescale s'approprient rapidement les enjeux métiers de leurs clients afin d'être force de proposition sur de nouveaux axes de valorisation de leurs données.

### TIBCO SPOTFIRE ET LE BIG DATA

**Tibco Spotfire, outil de visualisation, d'analyse et d'exploration de la donnée, apparaît sur un nombre croissant de projets Big Data comme une brique incontournable pour l'analyse des données du data lake.**

En plus des visualisations classiques que l'on retrouve dans la plupart des outils, **certaines aspects de Spotfire le distinguent de ses concurrents de façon décisive** :

- Des fonctions statistiques embarquées (segmentation, prévision, corrélation...)
- La grande modularité des composants et des configurations, permettant des analyses comparatives
- La capacité d'enrichir l'interface utilisateur et de proposer des analyses offrant une forte interactivité (HTML, JavaScript, Python)
- Les plugins riches et nombreux (visualisations poussées en D3.js par exemple)
- Les représentations cartographiques avancées
- La richesse des connecteurs

L'outil dispose de nombreuses solutions permettant d'**optimiser les performances** lorsque la volumétrie des données à analyser est trop importante pour être montée en mémoire :

- en utilisant la fonctionnalité de data-on-demand quand c'est possible, permettant de réduire le volume de données à analyser
- en déléguant les calculs au moteur de stockage, ce que permettent les connecteurs Spotfire
- en s'appuyant sur le serveur statistique qui permet de distribuer les calculs sur des données stockées sur Hadoop

Enfin Spotfire peut se connecter à la plupart des systèmes via les connecteurs que proposent toutes les technologies de stockage et d'analyse dans l'écosystème Big Data.

### RETOURS D'EXPERIENCE BLUESCALE

#### Analyse de marché dans le secteur automobile

Pour la direction marketing d'un grand groupe automobile, mise en place d'une application s'appuyant sur les données du data-lake, permettant d'analyser le marché du véhicule neuf sur un historique de 10 ans et pour 15 pays : profil des acheteurs, performance des marques, tendances du marché, etc.

#### Analyse de rendement / Étude de corrélations et effets de tests

Pour un grand groupe industriel, mise en place d'un tableau de bord permettant d'avoir une vue d'ensemble du rendement des jours précédents et de zoomer sur les anomalies. Les ingénieurs peuvent ensuite accéder au détail des données et en analyser les causes. Une étude de la corrélation entre les données de tests et les facteurs de production permet de tirer parti des fonctions d'analyses et du moteur TERR de Spotfire.

#### Prévision de consommation électrique

Implémentation de modèles de prévision de consommation électrique de data scientists permettant une analyse interactive par les utilisateurs avec possibilité de paramétrer des scénarios de croissance sur différents critères (emplois, démographie, industrie).

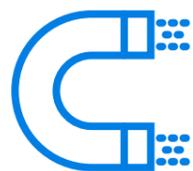
#### Détection de fraude

Pour un acteur majeur de la construction, la fraude aux achats représente un risque majeur. Aussi, à partir de règles métier, Bluescale a implémenté des rapports permettant d'effectuer une analyse des transactions inhabituelles.

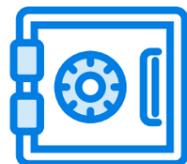
#### Guide de bonnes pratiques et optimisations

Pour une société d'assurance-crédit, Bluescale met en place les bonnes pratiques de développement et d'industrialisation des rapports Spotfire. Bluescale participe également à l'optimisation et à l'amélioration du rendu de certains rapports clé, notamment dans l'analyse de la performance des recouvrements.

## WE ARE DATA ENGINEERS



Collecte et Processing



Organisation des données



Architecture Data



Valorisation de la donnée

**YANN MANEIX**  
Directeur Technique

+33 6 31 70 74 65  
partner@bmb-services.com  
11 rue Bague  
75015 Paris France  
www.bmb-services.com



**BMB SERVICES EST UN CABINET DE CONSEIL QUI ACCOMPAGNE SES CLIENTS DANS LA TRANSFORMATION DU SYSTÈME D'INFORMATION AUTOUR DES INNOVATIONS SAP. DES ENTREPRISES UTILISATRICES NOUS SOLLICITENT AUJOURD'HUI DANS L'ÉLABORATION DE CAS D'USAGE, OU DANS LA MISE EN ŒUVRE DE PROTOTYPES SUR DES SCÉNARIOS À CONCEPTUALISER.**

A ce jour, la mise en œuvre de plateformes Big Data en est encore à ses débuts et se matérialise souvent par des besoins de « data tiering » (gestion de données « chaudes » versus données « froides ») ou de croisement des données structurées et non-structurées.

Dans ce contexte de forte volumétrie, et afin de limiter l'empreinte du stockage sur la base de données SAP HANA adossée à son ERP, l'un de nos clients nous a mandaté pour définir une architecture qui permette la réplication de données les plus anciennes du système SAP ECC vers un datalake Hadoop.

Cette architecture doit permettre :

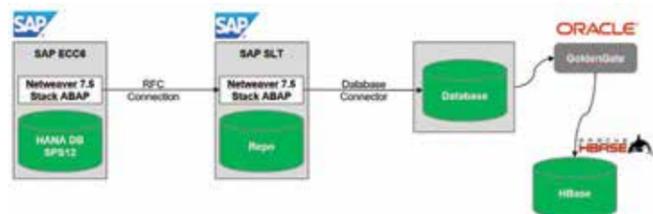
- de répliquer les données historiques
- un accès simple et instantané à ces données
- une maîtrise des coûts contraint par le modèle de licensing SAP HANA.

Pour cela 3 scénarios ont été proposés et des maquettes réalisées pour évaluer la solution la plus adaptée.

Il est à noter que les choix d'architectures retenus étant liés aux contraintes et à l'historique du client, ces options pourraient être différentes dans un autre contexte.

### Scénario 1 : Flux en deux étapes

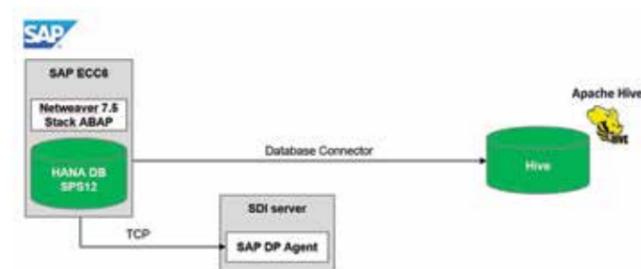
Le premier scénario consiste en la connexion d'un système SAP ECC vers une base de données relationnelle intermédiaire en mode Change Data Capture (CDC) puis réplication vers Apache HBase.



Le prototype a permis de mettre en avant les avantages et les inconvénients de cette architecture. Elle a l'avantage de fournir un mécanisme natif de réplication des données en mode temps réel. Elle permet également la réplication automatique des structures. Néanmoins, l'absence de connecteur SAP natif vers le monde Hadoop nécessite une étape intermédiaire.

### Scénario 2 : SAP HANA vers Hadoop

Le deuxième scénario consiste à utiliser les mécanismes de connexion natifs de SAP HANA (Smart Data Integrator/Smart Data Access) vers Hadoop en mode temps réel.

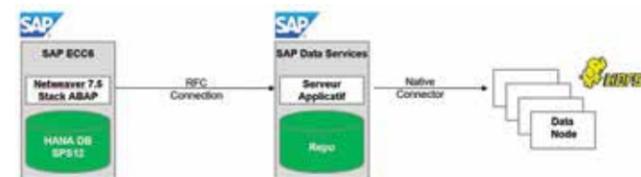


Cette solution permet une simplification de la mise en place en raison de la présence de connecteurs Hadoop au niveau de la base HANA.

En revanche, la complexité de la matrice de sécurité ainsi que l'obligation d'existence des tables à répliquer dans l'environnement cible constituent des points négatifs.

### Scénario 3 : SAP ECC vers Hadoop

Le troisième scénario consiste en la connexion d'un système SAP ECC vers une cible HDFS (ou Hive), avec une réplication des tables métiers SAP vers Hadoop au travers de SAP Data Services.



La solution SAP Data services permet de se connecter en standard à des sources et des cibles SAP et non SAP, ce qui constitue un réel avantage.

Nonobstant cette technologie s'avère complexe à mettre en œuvre pour des flux en temps réels.

La flexibilité offerte par Hadoop permet aujourd'hui de répondre à différents besoins métiers. La conceptualisation et la mise en place des architectures big data reste fortement contrainte par les contextes du client.



NOTRE APPROCHE

# Aligner technologies et métiers pour créer plus de valeur

# PRÉVOYEZ AVEC LE BIG DATA

**Offrez-vous de la haute performance et de l'agilité sur vos données avec les technologies Hadoop® du moment.**

Faites-vous accompagner dans votre orientation stratégique pour mieux analyser, prédire et comprendre l'information qualitative de demain.

## CONTACTEZ-NOUS

Stéphane Parisis  
Directeur associé  
Mobile : 06 22 61 87 64  
stephane.paris@bmb-services.com

William Grosjean  
Directeur associé  
Mobile : 06 84 77 02 78  
william.grosjean@bmb-services.com



Bâtiment B  
11, rue Bague  
75015 Paris- France

[www.bmb-services.com](http://www.bmb-services.com)





# BOOST YOUR PROFITS

## BOOSTER SA PERFORMANCE COMMERCIALE A L'AIDE DU PRICING INTELLIGENT

### POURQUOI S'INTÉRESSER AU PRICING ?

Le pricing est le **levier n°1 de rentabilité** des entreprises : un bon pricing, c'est 5 à 15% de marge en plus ! Et pourtant déterminer le bon prix de vente, celui qui maximise la marge ou les ventes dans un contexte donné, reste une tâche extrêmement complexe et chronophage.

### QU'APPORTE LA SOLUTION PROPOSÉE PAR BRENNUS ANALYTICS ?

Chez Brennus Analytics, nous sommes convaincus qu'il est grand temps pour les entreprises de s'affranchir de cette complexité ! C'est pourquoi nous avons développé Brennus™, une solution d'optimisation des prix basée sur l'**Intelligence Artificielle**. Outil d'aide à la décision qui s'adapte aux besoins de chaque client, Brennus™ assiste au quotidien directions et vendeurs en quête de performance commerciale. Notre objectif : vous aider à **améliorer vos marges et garder un coup d'avance sur vos concurrents**.

### A QUI S'ADRESSE T-ELLE ?

La solution Brennus est particulièrement adaptée aux **environnements complexes et concurrentiels** (beaucoup de produits, beaucoup de clients), comme l'industrie ou la distribution. Elle est toutefois applicable à une vaste étendue de secteurs, puisque nous travaillons aussi pour des sociétés de service.

### COMMENT FONCTIONNE LE LOGICIEL DE PRICING BRENNUS™ ?

Brennus™ **analyse vos données** pour analyser, comparer et prédire les comportements d'achat de vos clients. La solution détermine ensuite le meilleur prix pour **atteindre vos objectifs stratégiques** (par exemple augmentation des parts de marché ou maximisation de la marge), tout en tenant compte des **contraintes propres à votre entreprise** (production, stocks, contrats, etc.). Grâce à Brennus™, vous pouvez visualiser et analyser votre politique pricing, mais aussi simuler des changements ou encore construire vos prix à partir des recommandations de l'outil.

### QUEL EST VOTRE SOCLE TECHNOLOGIQUE ?

Notre solution (SaaS) s'appuie sur des technologies avancées d'Intelligence Artificielle, permettant de traiter une quantité massive de données et d'intégrer **un nombre quasi infini de variables et critères d'optimisation**. Brennus™ utilise notamment des systèmes multi-agents adaptatifs en apprentissage continu, particulièrement adaptés au **pricing dynamique et temps réel**, et qui présentent l'avantage d'être une technologie « **white box** ».

## EXEMPLE DE CAS CLIENT

### CONTEXTE

- Distributeur spécialisé
- Marché ultra concurrentiel
- Ventes physiques/web
- 30 000 produits
- 10 m de transactions/an

### PROBLEME

- Chute de l'activité
- Mauvaise image prix
- Gestion tarifaire et promotionnelle « intuitive »

### SOLUTION

- Segmentation produits dynamique
- Introduction d'un pricing différencié (KVIs/autres)
- Recommandations auto. de prix tarifs et promos

### RESULTAT

€ +10% de marge sur le périmètre optimisé

## BRENNUS ANALYTICS

Fondée en 2015 par plusieurs docteurs en Intelligence Artificielle rejoints par des spécialistes du pricing, la startup Brennus Analytics propose aux entreprises d'augmenter leur marge en optimisant leurs prix de vente. Brennus Analytics a été incubée par Telecom Paris Tech, puis a rejoint les programmes d'accélération BPI le Hub et WAI Boost en septembre 2017. Elle est soutenue par Scientipôle et la BPI depuis ses débuts, ainsi que par le fonds d'investissement ZTP et des investisseurs privés.

Brennus™ is a superior pricing solution based on an exclusive AI technology

Managing complexity for you, it delivers optimized prices in line with your business strategy

Analyze & Visualize

Automate & Simulate

Predict & Optimize

Price Lists

Rebates, discounts

Transactional Pricing



## BRENNUS ANALYTICS

### AI-POWERED PRICING SOLUTION

brennus-analytics.com | contact@brennus-analytics.com | +33 (0) 9 70 09 10 09

Brennus™ is a SaaS Price solution provided by Brennus Analytics



### ENTREPRISE

MediaSchool Group, créé en 2002 par Franck Papazian, est un groupe d'enseignement supérieur privé indépendant, spécialisé dans les domaines de la communication, du journalisme et du digital. Avec ce champ de compétences unique, 30 écoles - ECS, IEJ, #SUPDEWEB, SUPDEPROD et Paris School of Luxury - situées dans onze villes, et deux centres de formation continue - dont CB Executive dont il est question ici - MediaSchool Group forme tout au long de la vie à des métiers de passion.

### SAVOIR-FAIRE

CB Executive est le service de formation continue et Executive de MediaSchool Group. Il décline les expertises thématiques des écoles - communication, journalisme et digital - en formations pour les professionnels en activité. Le digital requiert de nouvelles connaissances et compétences, pour pouvoir rester indispensable dans son activité. Dans notre offre de formations, nous allons de la sensibilisation à l'approfondissement opérationnel.

### PRESTATIONS

Nous proposons deux Executive Mastères (diplômants) à temps partiel, mais aussi des séminaires courts, formations sur-mesure, cycles certifiants et des voyages d'études consacrés aux tendances digitales de demain et d'ailleurs (Etats-Unis, Chine, Canada, Israël). Les formats sont variés pour permettre à chacun de trouver son espace de respiration intellectuelle ou de travail sur des compétences opérationnelles, tout en conservant une activité intense.

### SECTEURS CIBLES

Notre thème central est le digital, des outils aux stratégies de communication/marketing, en passant par la conduite du changement. Il touche tous les secteurs. Historiquement implanté dans l'univers des médias, MediaSchool Group a ouvert ses secteurs cibles en Executive Education, avec des participants provenant de l'industrie, du luxe, de la santé, de la protection sociale, de l'IT, du consulting, etc.

### COUVERTURE GEOGRAPHIQUE

MediaSchool Group est implanté en France à : Paris, Strasbourg, Toulouse, Marseille, Nice, Reims et Rennes. Et à l'international à : Londres, Bruxelles, Barcelone, Shanghai. Les formations pour les professionnels peuvent avoir lieu soit dans ces villes, soit dans vos locaux - en France ou à l'étranger - ou encore dans nos locaux du 16e arrondissement de Paris.

### ACTUALITE

Nos deux Executive Mastères innovants sont toujours : «Communication, Marketing et Transformation digitale» et «Data Strategy». Sur ces Executive Mastères nous recrutons des professionnels de très bon niveau, en recherche d'expertise renforcée, d'évolution, d'ouverture vers des compétences nouvelles (pour le Data Strategy) ou parfois simplement souhaitant valoriser par un diplôme reconnu de niveau I une expérience riche mais non «certifiée». Ce type de formation représente également pour certains l'occasion de réfléchir à leur positionnement professionnel, et, tout au long du cursus, outre les échanges fertiles entre les participants et avec les intervenants, nous proposons un accompagnement personnalisé à la gestion de carrière et au Personal Branding.

Nous proposons également cette année un parcours de trois jours sur-mesure sur la programmation et la Data.

### AVANTAGES CONCURRENTIELS

Dans le domaine du digital, CB Executive tire sa légitimité de l'ancienneté de son réseau puisqu'en formation initiale le M1 'Communication digitale' a été créé en 2009. Notre vivier d'intervenants est de ce fait extrêmement riche, composé d'experts praticiens qui interviennent dans nos écoles ou dans l'Institut Multi-Médias (pour cadres dirigeants) existant depuis 1983. Les formations sont évaluées et enrichies chaque année, les formateurs sélectionnés pour leur expertise, leur capacité à être en prise sur les enjeux actuels et leur pédagogie.

Un Conseil scientifique contribue à la définition des programmes nouveaux et à leur parfaite adéquation aux besoins des professionnels en activité. Les membres sont des professionnels de haut niveau des secteurs de la communication, des médias, du digital et des RH. Enfin CB Executive s'est bâti sur « l'esprit start-up », pour faire preuve de souplesse, de réactivité et de créativité.



### CHARLES DADI, FONDATEUR PARTNER, FACTONICS - ET ENSEIGNANT DANS L'EXECUTIVE MASTÈRE DATA STRATEGY JULIEN MURESIANU, FONDATEUR DE JALGOS - START-UP ET CENTRE DE R&D EN INTELLIGENCE ARTIFICIELLE - ET ENSEIGNANT DANS L'EXECUTIVE MASTÈRE DATA STRATEGY

#### POURQUOI SE FORMER A LA DATA STRATEGY ? (CHARLES DADI)

Aujourd'hui les données (data) permettent un fort développement de l'activité car elles peuvent être mieux et plus rapidement traitées. Pour 60% des dirigeants (selon le Boston Consulting Group, 2014), l'utilisation des données est une priorité (70% des dirigeants dans la grande distribution). Il ne s'agit donc pas de théoriser sur le 'Big Data' mais d'étudier les technologies qui permettent de les collecter, les trier, les exploiter et les protéger, afin d'accroître la performance de son organisation. Il faut également connaître les différentes catégories de data, structurées, non structurées, sociales, open data... pour pouvoir fonder une stratégie pertinente. Avec la data strategy, on parle finalement de stratégie d'entreprise. Les données en sont simplement un élément qui devient crucial car générateur de valeur et différenciant. Toutefois la plupart des entreprises ont du mal à appréhender la stratégie data car celle-ci implique un changement d'organisation, une étude fine de ROI, et surtout d'avoir en interne des personnes compétentes pour traiter ces sujets. Or, les compétences manquent cruellement dans le domaine. D'où notre travail sur la mise en place de cette formation Executive. Il faut rapidement former des managers capables de piloter la stratégie data avec une feuille de route claire, précise et chiffrée. Cela repose sur une bonne connaissance de la data science, appliquée à des cas business.

#### QUELS SONT LES ATOUTS DE L'EXECUTIVE MASTÈRE DATA STRATEGY ? (CHARLES DADI)

Au-delà des aspects statistiques, technologiques, stratégiques et des outils à connaître, dans un positionnement très « business/marketing », cette formation insiste aussi sur les savoir-être à mobiliser dans le cadre d'une activité liée aux data : le leadership et la collaboration - entre les différents métiers et fonctions (IT, finance, marketing...). Nous marions la science des data avec la culture de l'agilité pour former les meilleurs managers de demain. Ceux qui auront une « double casquette » avec une expertise technologique et une excellente connaissance de leur activité Business. Ils seront aussi des communicants pour expliquer leur démarche et les résultats attendus, capables de faire du Data Design pour fédérer les différentes ressources internes. La data strategy mobilise des compétences larges. C'est le seul programme à ce jour qui forme des professionnels à l'ensemble de ces compétences. Son format est également un atout : l'Executive Mastère DATA STRATEGY totalise 23 journées de formation (184 heures). Les enseignements ont lieu pendant la semaine, à raison de deux à trois jours par mois, dans nos locaux du 16e arrondissement. Un MOOC est proposé en amont pour commencer à s'immerger dans les notions de data science indispensables. Les participants pratiquent les outils, traitent des études de cas, sont mis en situation. La formation est très pratique et se conclut par la production d'une note opérationnelle.

#### A QUI S'ADRESSE CETTE FORMATION ? (JULIEN MURESIANU)

A des professionnels expérimentés ayant au moins cinq ans d'expérience professionnelle. Plus précisément, nous souhaitons une promotion diverse dans les métiers d'origine car les « data strategies » doivent avant tout avoir une grande curiosité et une appétence pour les nouvelles technologies. Qu'ils viennent du marketing, de l'IT, de la stratégie ou de la production, nous recherchons des personnes très motivées, souhaitant devenir des chefs de projet data efficaces. Aujourd'hui selon le niveau de maturité des entreprises, la responsabilité de la stratégie data et son application incombent soit à la direction marketing soit à la DSI. Mais le chef de projet Data doit avoir une vision très large. Nous formons les interfaces entre les purs Data Scientists et les métiers.

#### QU'ENSEIGNEZ-VOUS DANS L'EXECUTIVE MASTÈRE « DATA STRATEGY » ? (JULIEN MURESIANU)

J'enseigne les grands principes de l'analyse statistique (statistiques traditionnelles, modèles de base et utilisation en entreprise), la data science, l'intelligence artificielle et le Model-thinking. J'explique le plus simplement possible les familles d'algorithmes et leur utilisation : optimisation, prédiction, clustering, matching et design d'algorithmes. Mes enseignements sont conçus pour être accessibles à des non-scientifiques. J'utilise beaucoup d'images et d'exemples pour faire comprendre la data science à ces cadres qui viennent de différents secteurs d'activité. J'anime également une séquence sur l'éthique car le Big Data pose un certain nombre de dilemmes éthiques. L'idée n'est pas de donner une réponse toute faite mais de sensibiliser les participants au réflexe de se poser des questions en la matière. Nous ne voulons pas qu'ils fassent tout et n'importe quoi ensuite avec les données.

#### QUE PENSEZ-VOUS DES PARTICIPANTS A CE PROGRAMME ET COMMENT LES VOYEZ-VOUS EVOLUER ? (JULIEN MURESIANU)

Je suis très agréablement surpris par le niveau de la promotion sélectionnée cette année. Ils sont hyper investis, très dynamiques, passionnés par le sujet. Ils font des choses, posent beaucoup de questions, veulent comprendre, c'est vraiment très plaisant. L'ambiance est excellente et j'ai vraiment senti entre mes deux interventions le progrès qui a été fait. Ils ont acquis des réflexes et sont bien partis pour être à la sortie d'excellents Chief Data Officers ou équivalents.



CentraleSupélec Exed est l'entité dédiée à la formation continue de CentraleSupélec. En accompagnant les professionnels aux différentes étapes de leur carrière dans le champ des sciences de l'ingénieur et du management, nous prolongeons notre mission première : celle de former les leaders de culture scientifique et technologique, et des innovateurs, capables de relever les grands défis de leur temps.

Les formations qualifiantes et diplômantes proposées par CentraleSupélec Exed s'articulent autour de quatre grands types de formations : les Formations Mastère Spécialisé®, les Executive Certificates, les formations courtes et les formations sur-mesure.

### LA FORMATION BIG DATA – ENJEUX ET OPPORTUNITÉS

A l'ère du numérique, le déluge de données crée de nouvelles opportunités économiques pour les entreprises. La possibilité d'analyser ces masses de données représente un avantage non négligeable en termes de compétitivité, et de nouveaux métiers apparaissent comme celui de Data Scientist. 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. L'objectif de cette formation est de dresser un état des lieux et un panorama des technologies disponibles pour permettre aux entreprises de garder la maîtrise de ces données et de la valeur qu'elles peuvent générer, et ainsi, de gérer la transformation vers l'entreprise numérique. L'organisation centrée données permet ainsi de placer les données au centre de la gestion opérationnelle et décisionnelle, et de réduire le temps de prise de décision.

### 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



## BIENVENUE DANS L'ÉCOLE DU FUTUR

Au cœur du pôle Innovation & Recherche de Paris-Saclay, CentraleSupélec Exed forme les cadres aux métiers de demain et certifie la progression réelle de leurs compétences.

### Avancer dans un monde digital

Plongez vos hauts potentiels dans un écosystème entrepreneurial innovant et permettez-leur d'acquérir les compétences indispensables pour réussir à l'ère du numérique et de l'usine 4.0.

Le nouveau catalogue des formations CentraleSupélec Exed est paru. Recevez votre exemplaire sur simple demande.

### Catalogue des formations CentraleSupélec Exed

Leadership  
& Transformation

Systèmes d'Information  
& Technologies du Digital

Excellence  
Opérationnelle

Technologies  
& Systèmes



Pour recevoir l'édition 2018,  
flashez ce QR Code ou appelez-le :  
**+33 (0)1 75 31 75 00**

[www.exed.centralesupelec.fr](http://www.exed.centralesupelec.fr)

## ÉLEVER SES COMPÉTENCES POUR RELEVER LES NOUVEAUX DÉFIS D'AUJOURD'HUI !

### 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 des données et Big Data,  
confess your Data / Cadrage et management de projets Big Data

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



[www.exed.centralesupelec.fr](http://www.exed.centralesupelec.fr)  
[info.exed@centralesupelec.fr](mailto:info.exed@centralesupelec.fr)  
+33 (0) 1 75 31 60 95

Le monde change  
Nous vous aidons à changer



SYSTÈMES  
D'INFORMATION  
ET TECHNOLOGIES  
DIGITAL



## QUANTIUM S'EMPARÉ D'UNE NOUVELLE NICHE DANS LE MARCHÉ DE L'ANALYTIQUE

En alliant une distribution MapR Hadoop à l'infrastructure Cisco UCS, le spécialiste de l'analytique diminue ses temps de requête de 92% et améliore la précision de ses résultats.

### LES BESOINS

Les consommateurs australiens sont parmi les plus avancés technologiquement. Valorisant la personnalisation de leur expérience, ils s'attendent à recevoir des messages, des recommandations et des promotions adaptées à chacun. De telles observations apportent une pierre à l'édifice de Quantum, société spécialisée dans la data analytique basée à Sydney. Depuis 15 ans, Quantum offre à ses clients un aperçu des besoins consommateurs, de leurs comportements, de leurs habitudes d'achats, des media qu'ils consultent en appliquant des techniques rigoureuses de data science aux données transactionnelles. Forte de son expérience dans les secteurs du retail, de la finance et des médias, l'entreprise travaille avec des sociétés telles que Woolworths, National Australia Bank ou Coca Cola. A l'origine, Quantum analysait uniquement les données fournies par ses clients. Un business très lucratif pour ses dirigeants, qui ont souhaité innover et acquérir leurs propres sets de données. L'objectif : offrir de nouveaux services à plus forte valeur ajoutée par le biais d'une corrélation entre les données internes de ses clients et des informations externes anonymisées sur le comportement d'achat des consommateurs.

Mais pour passer du concept au déploiement, Quantum a dû faire face à un obstacle de taille : la plate-forme analytique existante. Au fur et à mesure que la complexité et le périmètre de leurs analyses progressaient, la plateforme de base de données existante et l'architecture serveurs ne pouvaient plus faire face à la demande. Quantum a donc eu besoin d'une nouvelle plate-forme analytique reposant sur des serveurs fiables, évolutifs, et hautement performants, ainsi que sur un framework logiciel de classe entreprise.

### LA SOLUTION : UNE DISTRIBUTION HADOOP MAPR SUR UNE INFRASTRUCTURE CISCO

Pour relever le défi, Quantum a constitué un groupe de data-scientists afin de créer leur propre cahier des charges et d'évaluer les solutions logicielles et matérielles du marché. «Les décisions concernant la nouvelle plate-forme affecteront nos activités pour les années à venir. Nous avons donc beaucoup investi dans le processus de sélection», commente Alex Shaw, Responsable des opérations technologiques. Quantum s'est rendu compte qu'une solution Big Data était nécessaire non seulement en raison du volume de données, mais surtout pour répondre aux lourdes exigences d'analytique. Après s'être décidé sur Hadoop pour la partie logicielle, s'est posée la question de la distribution. Après un POC de trois mois durant lequel différents use-cases ont été testés sur des sets de données et de requêtes réelles, notre partenaire MapR, qui bénéficie du label « Cisco Preferred Solution », a remporté la mise. Niveau matériel, Quantum a souhaité short-lister son fournisseur historique, les équipements Cisco, et un fournisseur OEM de serveurs x86 dit « de commodité ». Si les résultats en matière de performance ont été relativement similaires, les serveurs Cisco UCS ont fait la différence tant sur leur faible consommation énergétique que sur leurs capacités de gestion centralisée.

Le portefeuille de Quantum se développant rapidement, l'entreprise a dû augmenter sa capacité de calcul en conséquence. L'architecture en cluster de la nouvelle plate-forme lui permet d'évoluer facilement, tout en conservant un TCO inférieur aux architectures précédentes : Quantum pouvant affiner son niveau de performance par un simple ajout de serveurs au cluster, sans avoir besoin de les remplacer par des modèles plus puissants. Comparativement, les outils de gestion de la plateforme Cisco/ MapR réduisent jusqu'à 30% les délais d'administration d'un système multi-tenant, où les données provenant de plusieurs clients résident sur un même cluster. Si l'approche « multi-tenant » réduit les coûts et optimise l'utilisation du matériel, elle introduit un risque d'accès non autorisé à l'information par d'autres utilisateurs du même cluster. « Pour garantir la sécurité des données de nos clients, nous nous reposons sur les fonctionnalités telles que le partitionnement de données qui permet de séparer logiquement les sets de données tout en optimisant le stockage », précise Alex Shaw.

### DES PERFORMANCES QUI DÉPASSENT LES OBJECTIFS DÉFINIS

Au cours de la rédaction du cahier des charges, Quantum s'est fixé comme objectif une augmentation de ses performances par un facteur dix. Cet objectif a été largement dépassé puisque la plateforme Ciso/MapR affiche une diminution du temps de traitement des requêtes de 92%, soit une progression en termes de performances par un facteur 12,5. La performance est cruciale pour Quantum qui souhaite fournir des résultats complets et rapides à ses clients. «Le fait d'accéder à des sets de données externes, que nous combinons aux données apportées par nos clients, nous confère une position unique sur le marché ». Alex Shaw détaille : « Une enseigne retail peut disposer d'une image assez bonne du comportement de ses clients dans ses propres magasins, mais connaît-elle la fréquence à laquelle ses clients achètent chez un concurrent ? En faisant se rejoindre des données concernant la consommation en magasin ou en ligne, avec des habitudes de consommation en matière de media, nous sommes capables de rendre les campagnes publicitaires plus efficaces. »

A mesure que les sets de données ont augmenté et que l'analytique s'est faite plus complexe, les méthodes d'échantillonnage sont devenues plus longues et compliquées. Mais grâce à la nouvelle solution Cisco/MapR, les data scientists peuvent dorénavant concevoir des requêtes complexes, les exécuter sur des volumes de plusieurs Terabytes, et obtenir des résultats plus précis en quelques minutes à peine. En permettant de tester des scénarios alternatifs rapidement et avec précision, et en raccourcissant les délais de développement, la plateforme contribue à favoriser l'innovation. « Nous avons beaucoup de gens intelligents mais qui ont été handicapés par la technologie et son incapacité à accompagner et à mettre en œuvre leurs idées. Maintenant qu'ils ont amélioré la façon de procéder à nos analyses, nous ouvrons la voie à la création de services nouveaux et innovants » conclue Alex Shaw.

Pour en savoir plus

- Les solutions Cisco pour Big Data, visitez : [cisco.fr/big-data](http://cisco.fr/big-data)



# Vos données sont partout. Augmentez la valeur métier de vos données.



Source IDC 2016



Retrouver les résultats de l'étude IDC :





### A PROPOS DE PUBLICIS ETO, DATA DRIVEN AGENCY

Publicis ETO est une agence du Groupe Publicis qui met la Data au cœur des dispositifs marketing et communication. Créée en 1986, composée de 300 personnes, elle rassemble tous les métiers du data marketing, du CRM, de la communication et du marketing : consulting, build et run de plateforme IT & Digital, design de customer journey, création Omnicanal, Data Management et Data Science. Véritable architecte de solutions de Connaissance Client, fournissant les meilleurs outils d'intégration, de pilotage et de restitution de la donnée., Publicis ETO accompagne les marques dans la conception et l'animation de leur parcours clients omnicanaux et gère près de 50 plateformes (RCU, Datalake, DMP, programmes de fidélisation, e-crm...), dont une bonne part à l'internationale.

Le métier de la Data chez Publicis ETO est organisé autour de deux pôles de compétences, le pôle Data Management et le pôle Data Science. Les équipes Datascience ont vocation à explorer l'ensemble des données disponibles pour analyser et fournir des modèles permettant de transformer et valoriser la donnée client en données activables. Les équipes Data Management quant à elle délivrent des plateformes industrialisées permettant l'activation effective des données et des modèles définis (Segment, Cluster, Score, Recommandation). Ces deux activités nécessitent des expertises quant à la collecte, le processing et l'exposition de ces données activables.

Compte tenu des évolutions technologiques, les plateformes Big Data ont révolutionné l'approche de la Data chez Publicis ETO, de par notamment la multiplication des capacités de stockage, de calcul et d'analyse. Publicis ETO s'est équipé en interne de 2 Datalakes : un Datalab dédié aux travaux des Datascientists et un datalake dédié aux équipes Datamanagement pour le développement de projets, de pilotes, et assurer la veille et R&D technologique.

Plus particulièrement, afin de permettre à chaque Datascientist de disposer d'un environnement de travail adapté à ses besoins en ressources, le datalab Publicis ETO a été construit sur un modèle qui se veut flexible, évolutif, sur la base de ressources mutualisées. Et ce en respectant les règles de cloisonnement et de sécurité inhérentes à la nature même de nos projets.

Cette approche permet à Publicis ETO de proposer à ses clients des espaces de travail collaboratifs et sécurisés, ayant pour objectif d'éprouver rapidement ces nouvelles technologies tout en maîtrisant ses coûts. Ce modèle, basé sur un usage des standards innovants et open-source Hadoop, permet pour le client une reprise des travaux produits. Alors, Publicis ETO, de par sa maîtrise des enjeux métiers et technologiques, sera à-même de l'accompagner dans l'activation opérationnelle des modèles et, au-delà, dans la mise en œuvre de sa plateforme Big Data.

Claranet accompagne Publicis ETO depuis près de 15 ans dans l'hébergement et l'infogérance de ses plateformes. C'est donc tout naturellement que Publicis ETO a choisi d'être accompagné par les équipes de Claranet Critical Data pour l'infogérance et la configuration de ses plateformes Big Data. La combinaison de l'expertise Métier de Publicis ETO et de l'expertise Technique de Claranet offrent à nos clients tous les atouts de réussite dans le cadre d'une mise en œuvre d'un projet Big Data.

### A PROPOS DE CLARANET CRITICAL DATA

Les experts Critical Data de Claranet sont en charge de l'infogérance des plateformes data, et des plateformes nécessitant un haut niveau de conformité (hébergement de données de santé, de données bancaires et de données personnelles).

Leur mission est de mettre à profit leur expertise sur les aspects sécurité et cloud public, afin de garantir à nos clients des niveaux élevés de qualité de service, d'industrialisation, et de respect des engagements contractuels de leur plateforme.



### LA DATA AU CŒUR DES ENJEUX MÉTIERS DES ENTREPRISES

La Practice Critical Data de Claranet est en charge de l'infogérance des plateformes data, et des plateformes nécessitant un haut niveau de conformité (hébergement de données de santé, de données bancaires et de données personnelles).

### COMMENT ÉVOLUE LE MARCHÉ DU BIG DATA SELON VOUS ?

Les enjeux métiers autour de la donnée sont au cœur des stratégies digitales des entreprises. On assiste à une véritable démocratisation du « Big Data », mais les problématiques techniques (disponibilité, performance, et évolutivité des plateformes), combinées aux contraintes réglementaires et aux exigences de sécurité, restent cruciales pour la bonne réussite de ces projets. En conséquence, beaucoup de projets Big Data ont d'abord été montés « on-premise » par nos clients. Aujourd'hui ces derniers font face à des difficultés budgétaires, techniques et d'industrialisation. C'est pourquoi la plupart d'entre eux souhaitent les externaliser dans le cloud. De ce fait, on assiste à une véritable cloudification des plateformes Big Data.

### AVEZ-VOUS DES SOLUTIONS SPÉCIFIQUES POUR RÉPONDRE À CETTE CLOUDIFICATION ?

Chaque projet Big Data possède des contraintes qu'il faut prendre en compte. Le choix du cloud provider est dépendant de ces contraintes. Migrer ses données dans le cloud, même si elles peuvent être anonymisées, représente notamment des enjeux de sécurité et de conformité. Notre Practice Critical Data est experte dans les sujets de sécurité des données, que ce soient des données bancaires, personnelles ou données de santé par exemple. Notre mission est donc de mettre à profit notre expertise sur les aspects sécurité et cloud public, afin de garantir à nos clients des niveaux élevés de qualité de service, d'industrialisation, et de respect des engagements contractuels de leur plateforme. Dans ce but, Claranet s'est associé avec Cloudera, éditeur de Plateforme Big Data Open Source, pour lancer un service évolutif, sécurisé et managé dans le cloud, basé sur la solution Cloudera Entreprise. Nous assurons le déploiement automatisé des plateformes dans le cloud, et surtout leur maintien en conditions opérationnelles.

### COMMENT RÉPONDEZ-VOUS AUX EXIGENCES RÉGLEMENTAIRES ET SÉCURITAIRES ?

En fonction de la typologie des données hébergées sur les plateformes, nous mettons en place les niveaux de sécurité adéquats : données de santé, données personnelles, données bancaires. Les agréments et certifications (PCIDSS, HADS, ISO27001) dont Claranet bénéficie, facilitent la mise en place de projets aux contraintes spécifiques. Je pense notamment aux projets Big Data du secteur des banques et assurances. Le fait que Claranet soit certifié PCI DSS, et que Cloudera soit la première solution ayant permis la certification PCIDSS d'une plateforme Big Data, nous permet d'amener des garanties sur notre capacité à héberger et infogérer des données bancaires. C'est une vraie valeur ajoutée pour ces acteurs.

### LES PROJETS BIG DATA DEMANDENT-ILS DES COMPÉTENCES PARTICULIÈRES ?

Il faut bien comprendre que les plateformes Big Data sont des plateformes de production, avec des données sensibles. Il est indispensable de s'assurer de leur maintien en conditions opérationnelles avec une supervision fine - technique et métier-, de la gouvernance, un suivi des SLAs et du reporting. Mais surtout, il faut des experts qui vont accompagner les datascientists et datastewards de nos clients, sur la performance de leur plateforme, de leurs requêtes, et de leurs traitements récurrents. Les plateformes Big Data apportent des vrais leviers pour les entreprises sur leurs enjeux métiers. Elles vont leur permettre de faire des analyses prédictives, mieux connaître leurs clients ou de faire de la détection de fraude par exemple. En cas de problèmes de performance, il est indispensable de pouvoir faire le lien entre les gens qui développent et le monde des opérations. C'est pourquoi nous avons mis en place une équipe data dédiée, composée de DataOps, et de Data Architects sur la partie NoSQL / Big Data. Ayant des connaissances sur les langages de programmation comme Spark, Scala, Python, ils échangent avec les développeurs pour s'assurer de la bonne performance des traitements qu'ils réalisent, et accompagnent les Data Scientists sur l'exploitation de leurs données. Ce sont des métiers relativement récents. La plupart de ces profils ont beaucoup travaillé dans le monde de la base de données relationnelle par le passé, puis se sont orientés vers l'écosystème Big Data. Ils sont à la fois très proches du système et de la donnée, et ils font le lien entre l'aspect fonctionnel, la donnée, et les opérations. Ils apportent leur vision sur l'importance de la donnée, sur la gestion des modèles de données, et sur l'optimisation des traitements.

### LES PROJETS BIG DATA DEMANDENT-ILS UNE GESTION DE PROJET SPÉCIFIQUE ?

Un projet Big Data se décompose de façon assez classique en 3 phases : le Think, le Build et le Run. En revanche, dans la phase de Think, le DataOps, l'équipe conformité de Claranet, et un chef de projet vont travailler ensemble au dimensionnement de la plateforme et à son architecture, y compris son architecture sécurité. Ils vont également travailler conjointement avec l'équipe client sur les limitations d'accès à la plateforme et pour s'assurer que les contrôles d'accès soient bien respectés, mais aussi sur le modèle de données et sur l'identification des sources de données qui vont être intégrées dans cette plateforme. Il y a donc toute une phase d'accompagnement, de dimensionnement, de préparation durant cette phase de Think, avant de commencer le Build de la plateforme.

Pour en savoir plus sur notre offre Big Data, rendez-vous Stand n° A37



### CLUDERA LANCE SDX, SHARED DATA EXPÉRIENCE, UNE EXPÉRIENCE DE DONNÉES PARTAGÉES POUR LES ENTREPRISES ET AIDE À UNIFIER LES CAPACITÉS DE GOUVERNANCE ET DE SÉCURITÉ DANS LE CLOUD.

Cloudera, Inc. (code NYSE : CLDR), la plate-forme moderne de Machine Learning et d'Analytics, optimisée pour le Cloud, lance Cloudera SDX, un framework logiciel modulaire permettant aux entreprises de bénéficier d'une « expérience de données partagées ». En appliquant un framework centralisé et cohérent aux schémas, à la sécurité, la gouvernance, l'ingestion de données et plus encore, SDX permet à des dizaines d'applications client différentes d'utiliser simultanément des ensembles de données partagées. **SDX facilite le développement d'applications de données multifonctions avancées de Machine Learning et d'Analytics, réduit le coût de leur déploiement et rend leur sécurisation plus cohérente.**

Différentes applications complexes fonctionnant ensemble sont aujourd'hui nécessaires pour mener à bien les projets stratégiques autour de la cybersécurité, de la connaissance client et l'intégration des produits et des services, et plus encore. Les meilleures plateformes permettent l'utilisation simultanée du Machine Learning et du traitement en temps réel. Dans de nombreux cas, ces clusters d'applications multifonctions sont hébergés dans des infrastructures Cloud ou s'appuient sur des données générées dans le Cloud.

Le développement et la livraison de ces applications complexes sont un défi dans les clusters multi-tenant et sont particulièrement difficiles dans les environnements Cloud, car les politiques et les contextes de données ne persistent pas dans les environnements transitoires. La plupart des services dans le Cloud sont en fait des applications cloisonnées fonctionnant dans des clusters isolés.

**SDX permet aux clients de créer, gérer et sécuriser ces applications d'Analytics multifonctions à forte valeur.** SDX facilite également les solutions partenaires certifiées Cloudera qui intègrent et améliorent encore davantage les capacités de la plate-forme.

Un composant clé de Cloudera SDX est le **catalogue de données partagées sur lequel des fonctions cohérentes de sécurité, de gouvernance et de gestion** peuvent être établies et exploitées par des applications d'Analytics éphémères et persistantes. Le catalogue partagé fournit également un accès en libre-service à des métadonnées métiers. Le service informatique n'a plus besoin de recréer les contextes et les contrôles pour chaque nouveau cas d'utilisation.

« Les entreprises citent souvent la sécurité, la gouvernance et la complexité parmi leurs principales raisons de ne pas migrer leur charges de traitement opérationnel vers le Cloud, » déclare Tony Baer, principal analyst chez Ovum. « Cloudera s'est imposé en intégrant des fonctionnalités de sécurité et de gouvernance des données permettant aux entreprises de mettre leurs charges de traitement des données en production, en toute confiance. Cloudera SDX s'appuie sur la propriété intellectuelle de la société et son expérience des données partagée dans les environnements sur site et dans le Cloud. »

Cloudera SDX répond aux défis critiques suivants pour les entreprises :

- **Données cloisonnées** - Les clusters en libre-service dans le Cloud ne partageant pas naturellement les données et les métadonnées, les clusters individuels deviennent de fait des silos. En partageant des métadonnées et des données persistantes entre les applications à la demande et les clusters transitoires, les clients de Cloudera conservent leur agilité et s'assurent que chaque cluster isolé ne nécessite ni administration ni contrôle individuel, sans subir de coûts supplémentaires liés à la réplication et au stockage des données.
- **Faibles de sécurité** - Sans contrôles de sécurité centralisés, les administrateurs sont forcés de réappliquer continuellement la sécurité et les politiques d'accès à plusieurs exemplaires de données cloisonnées, entraînant un travail supplémentaire et un risque accru de fuites d'informations confidentielles. Avec Cloudera SDX, la sécurité est appliquée uniformément au niveau des données. Les politiques sont omniprésentes et n'ont pas besoin d'être modifiées ni d'être réappliquées lorsque les données sont déplacées ou utilisées dans une nouvelle application d'analyse.
- **Problématiques de gouvernance/non-conformité** - Un catalogue de métadonnées et de données partagées est aujourd'hui impératif pour la conformité HIPAA et PCI-DSS, et de futures exigences telles que le RGPD. Un catalogue de données partagées permet de facilement et rapidement trouver et comprendre le contexte des données, utiliser des applications en libre-service, et fournir des fonctionnalités d'audit et de lignage inaltérables.

« SDX est « l'ingrédient miracle » de Cloudera Enterprise qui accélère la science des données, le Machine Learning et l'Analytics, » précise Mike Olson, founder and chief strategy officer chez Cloudera. « Les données sont la ressource la plus précieuse au monde. C'est le carburant qui donne vie aux idées et aux machines, et qui résout des problèmes impossibles. L'objectif de Cloudera a toujours été d'aider les entreprises à maximiser la valeur de leurs données. Cloudera SDX simplifie cette mission pour tous les utilisateurs. »

Cloudera SDX prend en charge plusieurs configurations de Cloud public, privé et physique, et est disponible dès aujourd'hui pour les clients possédant des licences de Cloudera Data Science and Engineering, Analytic DB, Operational DB ou Enterprise Data Hub. Pour plus d'informations, consultez [cloudera.com/sdx](http://cloudera.com/sdx).

#### À propos de Cloudera

Nous pensons que les données peuvent rendre ce qui est impossible aujourd'hui, possible demain. Nous facilitons la transformation de données complexes en connaissances claires et exploitables, en offrant la plate-forme moderne de Machine Learning et d'Analytics optimisée pour le Cloud. Les plus grandes entreprises au monde font confiance à Cloudera pour résoudre leurs plus grands défis métiers. Pour en savoir plus : [cloudera.com](http://cloudera.com).

### EN EXPLOITANT LA PLATE-FORME DE MACHINE LEARNING DE CLUDERA POUR OPTIMISER LA COLLECTE DES DONNÉES ET LES STRATÉGIES D'ACTIVATION, 100MERCIS GROUP MAXIMISE L'EFFICACITÉ, LA PERSONNALISATION ET LA PERTINENCE DE LA COMMUNICATION MULTICANAL PROGRAMMATIQUE DE SES CLIENTS.

- Services aux entreprises, Adtech & Martech
- Siège social : Paris, France

**Défi** : Alors que le monde est de plus en plus orienté vers les données et hyper connecté, les entreprises spécialisées en technologies de marketing telles que Numberly se basent sur leur capacité à fournir les meilleures informations à partir d'ensembles de données volumineux et divers, pour acquérir un avantage concurrentiel. Générer le plus gros retour sur investissement marketing et publicitaire dépend de leur capacité à fournir des messages ciblés, pertinents et au bon moment. Pour maintenir sa position en tant que leader mondial, Numberly doit capturer, traiter et concilier des ensembles de données larges et divers – comme les données statiques issus des CRM et les flux de données des capteurs – et rendre ces données accessibles à la data science et aux analyses, le tout en temps réel.

**Solution** : L'écosystème de collecte des données, de machine learning et d'activation de Cloudera alimente les analyses marketing en temps réel, les enchères publicitaires et les moteurs de recommandation à grande échelle et représente un atout essentiel de Numberly. « Cloudera dynamise tous nos algorithmes de machine learning et nos compétences internes. » Alexys Jacob, CTO à Numberly. Numberly apprécie particulièrement l'outil de Cloudera qui supporte les besoins cruciaux de sécurité et gouvernance. « Nous travaillons avec de nombreux secteurs d'activité comme le retail, les banques et assurances, qui sont tous très préoccupés par les questions de sécurité. C'est la raison pour laquelle notre travail s'est toujours concentré sur la sécurité des données que nous exploitons. »

**Mise en place** : Numberly avait besoin d'une plate-forme de données moderne pour dynamiser le machine learning. Cloudera lui permet de se focaliser sur l'innovation sans se soucier de l'intégration et du support de chaque détail d'une infrastructure technologique complexe.

Numberly évalue Cloudera Enterprise sur :

- L'accès direct et intuitif à une plate-forme d'experts internationaux
- La visibilité sur l'écosystème environnant
- La gouvernance, la confidentialité et la traçabilité des données (via Cloudera Navigator)
- L'accès à la boîte à outils du Cloudera Data Science Workbench

« Nous voulions améliorer la maturité et la qualité de notre plate-forme et c'est pour cela que l'on s'est associé à Cloudera », déclare Alexys. « Nous avons une vision claire de tout l'écosystème et des outils que nous utilisons. Les analyses de configuration, les rapports et la supervision nous aident à comprendre l'activité exacte du cluster, comment et quand la réduire. Ce sont des questions importantes pour une entreprise à forte croissance. »

**Résultats** : Les clients de Numberly profitent directement de la qualité supplémentaire que Cloudera apporte à sa plate-forme de machine learning. Par exemple :

- Une grande entreprise internationale de divertissement a observé une augmentation de 35% du taux de clics grâce à un

algorithme propulsé par Cloudera qui permettait à Numberly d'optimiser ses stratégies programmatiques en se basant sur les signaux des utilisateurs, notamment les analyses de navigation,

- Le Stade de France comprend mieux et optimise le flux des participants pendant les événements en analysant les traces de connexions Wi-Fi, la radio-identification (RFID) et les données des capteurs, en plus des données hors-ligne.
- Une entreprise internationale de banque et assurance fait confiance à Numberly pour tirer profit de ses données comportementales afin de réduire les coûts d'acquisition de ses nouveaux clients. Grâce à un algorithme de machine learning innovant, les coûts ont été réduits de 20%.

#### Applications commerciales supportées :

- Achats programmatiques de bannières optimisés grâce à des algorithmes en temps réel
- Personnalisation et moteurs de recommandations
- Optimisation de la collecte des données et des stratégies d'activation
- Personnalisation et mise en évidence de l'optimisation des achats programmatiques

#### Impact :

- Accroissement des ventes des clients
- Augmentation des taux de clics publicitaires
- Meilleur trafic dans les stades et magasins de détail
- Placement des produits en rayon plus efficace

#### Sources de données :

- Données CRM PII
- Sources de données localisées comme les réseaux sociaux par exemple
- Les données non personnelles (Non-PII) comme le suivi d'une campagne média
- Les capteurs comme les RFID et les balises iBeacons

#### Solution :

- Une plate-forme de données moderne : Cloudera Enterprise
- Charges de travail : Ingénierie des données
- Composants clés : Apache Spark, Cloudera Navigator

#### Échelle Big Data :

- Plus de 2 pétaoctets (2PB+)



Alexys Jacob, CTO, Numberly  
« Cloudera dynamise tous nos algorithmes d'apprentissage automatique et nos compétences internes. »

# CLUB DECISION DSI

**VÉRONIQUE DAVAL**

Présidente du Club

01 53 45 28 65

club@decisiondsi.com

33 rue Galillée  
75116 Paris Feance

www.clubdecisiondsi.com

Analytique API Applications Consultants/SSI Intégrateurs Dataviz Infrastructure Institutionnels

**CLUB DECISION DSI**  
1<sup>er</sup> Club Français de décideurs informatiques & télécoms



# Devenez Membre du Club Décision DSI

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

## LE CLUB DÉCISION DSI, PREMIER CLUB FRANÇAIS DÉDIÉ AUX DIRECTEURS INFORMATIQUES

Aujourd'hui le Club Décision DSI réunit 1 250 membres «directeurs Informatique & Télécoms» de sociétés du secteur privé et du secteur public dont le parc informatique est supérieur à 200 pc. Il rassemble des hommes et des femmes d'expérience, animés par la volonté de mener des actions qui soutiennent d'autres directeurs des systèmes d'information (DSI) dans la réalisation de leurs projets professionnels. Le club est administré par Véronique Daval, la présidente, le vice-président, ainsi que par un conseil central composé de neuf directeurs informatiques.

### Chaque année le club organise :

- 20 manifestations thématiques par an.
- 7 événements accessibles aux partenaires (éditeurs de logiciels, constructeurs, opérateurs télécoms, SSI)
- 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 1250 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.

## LES MEMBRES DU BUREAU ET AMBASSADEURS DU CLUB



Armand ASSOULINE  
CIO  
MSC FRANCE



Gilles BERTHELOT  
RSSI  
GROUPE SNCF



Christophe BOUTONNET  
Sous-directeur  
Schéma  
Directeur et  
Politique SI



Christian DOGUET  
DSI - CHAÎNE THERMALE  
DU SOLEIL



Damien GRIESSINGER  
CTO  
EPP0



Christophe GUILLARME  
DSI  
GROUPE AB TÉLÉVISION



Trieu HUYNH-THIEN  
DSI ADJOINT - CENTRE  
GEORGES POMPIDOU



Bertrand LENAIN  
DSI  
CCI DE SEINE ET MARNE



Stéphane MALGRAND  
DSI - LABORATOIRE  
NATIONAL DE MÉTROLOGIE  
ET ESSAIS



Lionel ROBIN  
DSI  
GROUPE LA RESERVE



Dominique TROUVE  
DSI  
HÔPITAUX AVICENNE



Claude YAMEOGO  
ARCHITECT SI  
ALSTOM



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



**Club Decision DSI**



### COHERIS ACCOMPAGNE LES ENTREPRISES DANS LEURS ENJEUX DE MISE EN CONFORMITÉ DE LEURS DONNÉES PERSONNELLES

L'échéance de l'entrée en vigueur du Règlement Général sur la Protection des Données (RGPD) arrive à grand pas ! Plus qu'un nouveau cadre de protection des données, c'est une véritable remise en question de la manière d'appréhender la donnée qu'accompagne la mise en application de la RGPD, et ce, dans toutes les directions métiers et SI. Coheris, expert en matière de gestion de la relation clients et de traitement de la donnée, propose 4 offres complémentaires alliant consulting et solutions logicielles, pour accompagner les entreprises dans la gestion et le traitement de leurs données personnelles et sensibles.

### CARTOGRAPHIE DES DONNÉES SENSIBLES DANS LES CHAMPS TEXTES

C'est une offre de service qui permet d'identifier dans les zones de commentaires libres de toute application, les données sensibles au sens de la réglementation, et d'établir des tableaux de bord associés.

Grâce à leur expertise en analyse de données et en s'appuyant sur la solution de datamining Coheris Spad, nos Datascientists vous accompagnent dans la mise en conformité de vos données issues des champs textuels de toutes vos applications.

S'appuyant sur un dictionnaire de mots et expressions classés par thématiques et correspondant à la définition des données sensibles du RGPD, nos Datascientists réaliseront l'analyse complète de vos champs textes qui identifieront les verbatims problématiques et les classeront par niveau de criticité.

Grâce à la puissance de la solution de Datamining Coheris SPAD, nos Datascientists peuvent traiter une volumétrie de verbatims importante (plusieurs millions), vous permettant d'éviter la suppression de ces verbatims ou la charge de relecture exhaustive par vos équipes.

Notre méthodologie permet de traiter les exceptions liées à votre secteur d'activité ou votre contexte métier.

### MODULE COHERIS RGPD TEXT-CONTROL

La solution logicielle Coheris RGPD Text-Control est utilisable sur toutes vos applications (CRM, ERP, RH, ...) pour :

- contrôler les espaces de textes libres de ces applications et éviter la saisie de données dites à caractère sensible (santé, vie sexuelle, opinions politiques, etc. ou encore des commentaires subjectifs ou des jugements de valeur).
- permettre aux responsables de traitement de contrôler ces zones de risques pouvant nuire à la conformité de votre entreprise à la nouvelle réglementation RGPD.

### Deux modes de contrôle sont disponibles : temps réel ou différé

Le contrôle en temps réel de la présence d'expressions sensibles dans un champ texte vous permet de détecter en temps réel au moment de sa création, si un verbatim contient des expressions sensibles lors la saisie par un utilisateur au sein d'une application métier.

Le contrôle régulier en **mode différé** de la présence d'expressions sensibles dans des champs textes permet d'appliquer le calcul d'un score de criticité à de nouveaux verbatims.

### Formation « Données & RGPD »

Coheris propose une formation pratique d'une journée pour permettre aux entreprises d'appréhender simplement le RGPD et de bien démarrer leur mise en conformité dans la gestion de leurs données et des traitements associés. Au programme de cette formation, différents sujets seront abordés, notamment :

- Le contexte et les enjeux de la réglementation
- L'enjeu des données personnelles
- Démarche structurée dans l'entreprise : exemple d'un CRM



# DONNEES et RGPD

## Appréhender simplement le RGPD

## Bien démarrer la mise en conformité de vos données et les traitements associés



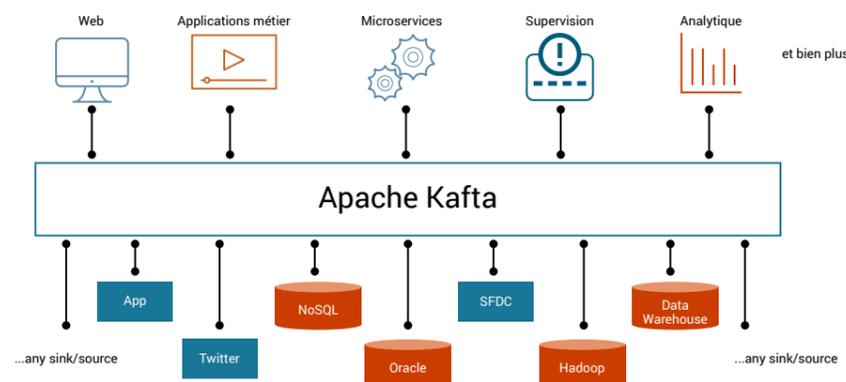


## Confluent: Apache Kafka pour l'entreprise

La réactivité en temps réel a été possible dans le passé, mais elle a toujours été limitée à un seul système afin de garantir une limitation de la latence. Lorsque le temps réel peut être distribué, prendre en charge de très gros volumes de données et être fiable et tolérant aux erreurs, des changements vraiment importants commencent à apparaître.

Les entreprises découvrent qu'elles peuvent créer de nouveaux débouchés commerciaux, ainsi que rendre leurs activités actuelles plus efficaces à l'aide de données disponibles en temps réel.

Apache Kafka™ est une plateforme de Streaming conçue pour résoudre ces problèmes dans une architecture moderne et décentralisée. À l'origine envisagée comme une file de messagerie décentralisée rapide et évolutive, elle s'est rapidement développée pour devenir une plateforme de Streaming à grande échelle, capable non seulement de publication et d'abonnement, mais également de stockage et de traitement des données dans le flux.



### Un système nerveux central

La première plateforme de Streaming est **Apache Kafka™**, une technologie open source utilisée par les géants technologiques, notamment **LinkedIn, Uber, eBay, Netflix**, et **Airbnb** en tant que système nerveux central qui relie leur entreprise. Les multinationales ont de plus en plus recours à Kafka : des entreprises comme **Goldman Sachs, Capital One** et **Target** se tournent toutes vers Kafka pour encourager la transformation des activités.

### La platform Confluent

Les créateurs de Kafka ont depuis fondé Confluent et offrent maintenant une plateforme de Streaming adaptée aux entreprises qui permet de gérer les charges de travail critiques. La plateforme Confluent offre une expérience plus complète pour les développeurs, ainsi que des fonctionnalités d'exploitation et d'administration de l'entreprise, ce qui permet de déployer et d'exploiter Kafka à grande échelle. La plateforme Confluent propose une série de points de départ en fonction des besoins spécifiques d'une organisation.

— PLUS DE —  
**1/3** DES ENTREPRISES DU  
**FORTUNE 500**  
UTILISENT KAFTA DANS LA PRODUCTION

# Le pouvoir de l'instant

## LES BONNES CHOSES ARRIVENT À CEUX QUI N'ATTENDENT PAS.

Faites fonctionner votre business en temps réel avec la plateforme pour entreprise de Confluent.

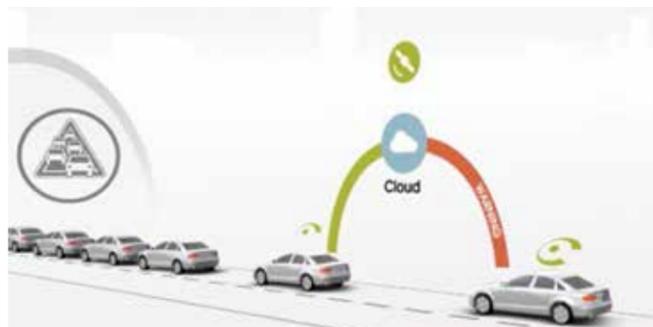
La plateforme entreprise de Confluent basée sur Apache Kafka est une manière plus complète de se lancer dans le stream de données en temps réel.

VENEZ NOUS VOIR AU STAND 29  
POUR EN APPRENDRE PLUS



**COYOTE INFORME SES 5 MILLIONS D'UTILISATEURS EN TEMPS RÉEL GRÂCE À LA TECHNOLOGIE COUCHBASE MOBILE**

Coyote, assistant d'aide à la conduite communautaire, a fait appel à Couchbase, éditeur de logiciel spécialisé dans la technologie de base de données, pour informer du trafic ses 5 millions d'utilisateurs en temps réel.



Le système Coyote avertit plus de 5 millions d'utilisateurs des limitations de vitesse, des situations de dangers, des alertes climatiques ou encore des perturbations routières à travers de multiples plateformes (dispositifs connectés, applications, systèmes embarqués).

Ce sont près de 13 millions d'alertes qui sont remontées quotidiennement par le système avant d'être partagées aux millions d'utilisateurs. Ces données doivent être synchronisées, traitées et diffusées en temps réel afin d'offrir un service premium à cette clientèle exigeante.

Couchbase Mobile intervient dans la récolte, le traitement et la synchronisation de ces données que ce soit en mode On ou Offline. Sans perte de temps à rechercher du réseau, les alertes sont instantanément envoyées sur les milliers de devices permettant ainsi d'optimiser l'expérience des utilisateurs avec une solution complète combinant agilité, gain de temps et efficacité.



« Là où nous passions des heures à synchroniser les données depuis le serveur sur les plateformes, avec Couchbase, d'une simple manipulation, les données sont déployées instantanément sur les millions de devices !

Grâce à Couchbase, nous pouvons aujourd'hui consacrer plus de temps au développement de nouveaux services et au déploiement de notre outil à l'échelle mondiale.

Couchbase n'est pas seulement un fournisseur mais c'est surtout un partenaire » se félicite **Mathieu Brun Bellut, CIO de Coyote**

Avec l'utilisation de Couchbase Mobile, Coyote se différencie en proposant des services personnalisés. En effet depuis l'intégration de cette solution, Coyote a développé l'outil de data visualization 'Mes Stats' qui permet aux utilisateurs de prendre connaissance de leurs données (nombre de kilomètres parcourus, vitesse moyenne, temps passé dans les bouchons, etc.).

**À PROPOS DE COUCHBASE**

Couchbase propose des technologies de données NoSQL à des entreprises axées sur l'innovation numérique. Dans le monde entier, des organisations choisissent Couchbase pour la flexibilité de son modèle de données, la facilité de son redimensionnement, ses performances et sa haute disponibilité pour construire des applications web, mobiles, et IoT d'entreprise. La plate-forme Couchbase inclut Couchbase Server, Couchbase Lite, première base de données NoSQL mobile, et Couchbase Sync Gateway. Couchbase est conçu pour réaliser des déploiements mondiaux, avec réplique configurable entre datacenters afin de renforcer la proximité géographique et la disponibilité des données. Tous les produits Couchbase sont des projets open source.

Couchbase compte parmi ses clients des leaders de l'industrie tels qu'AOL, Amadeus, AT&T, Cisco, Comcast, Concur, Disney, eBay, General Electric, Marriott, Nordstorm, Neiman Marcus, PayPal, Ryanair, Rakuten/Viber, Tesco, Verizon, Wells Fargo, ainsi que des centaines d'autres grandes entreprises.

Accel Partners, Adams Street Partners, Ignition Partners, Mayfield Fund, North Bridge Venture Partners et WestSummit Capital font partie des investisseurs Couchbase.

**Contact presse**

Agence Point Virgule - Khéoudja SADAoui  
+33 (0)6 52 49 85 80 / +33 (0)1 73 79 50 72  
ksadaoui@pointvirgule.com

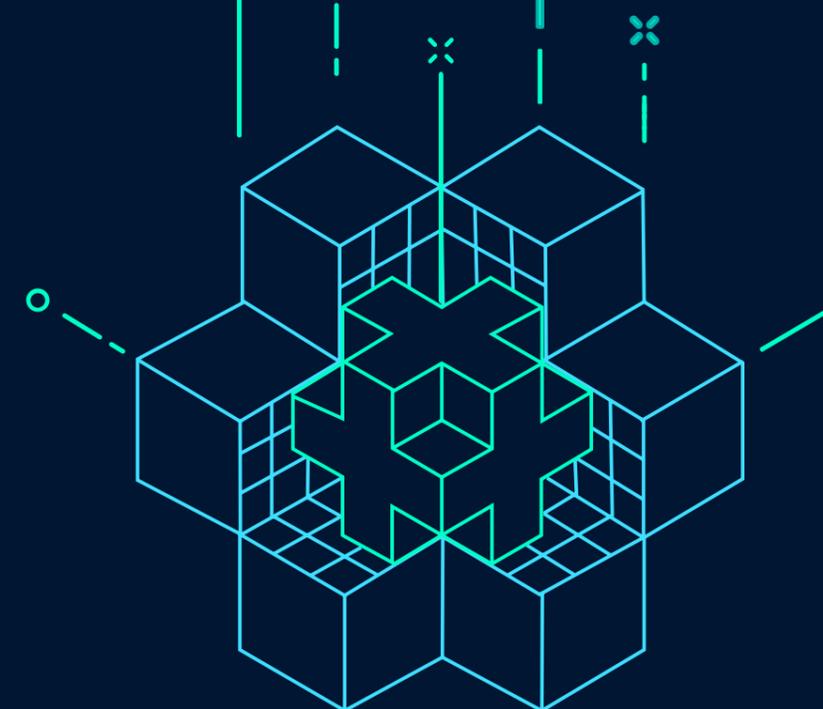
# DE L'EXPÉRIENCE CLIENT À L'EUPHORIE CLIENT

Bienvenue sur la première *Engagement Database* au monde



Consultez **couchbase.com** dès aujourd'hui pour en savoir plus.

START A REVOLUTION



# EXPERIENCE DATA SCIENCE AT SCALE

## STAFF

data scientists  
and analysts  
that collaborate

## STRUCTURE

reproducible,  
clear workflows

## ACCELERATE

data cleaning  
and wrangling

## AUTOMATE

data projects  
for maximum  
efficiency

## DEPLOY

to production  
quickly for  
business impact

With Dataiku, the unique advanced analytics solution that was named a visionary in the Gartner 2017 Magic Quadrant for Data Science Platforms. Dataiku enables companies to build and deliver their own data products more efficiently.

GRDF est un des clients historiques de Dataiku. Retrouvez dans cet échange entre Marc Batty, co-fondateur de Dataiku et Paul Fournier, Responsable Innovation numérique et dataLab chez GRDF, comment une entreprise a pu utiliser le big data pour s'attaquer à de véritables problématiques business.

### PAUL, COMMENT EST ORGANISÉE VOTRE DÉMARCHE BIG DATA CHEZ GRDF, ET POUR RÉPONDRE À QUELS DÉFIS ?

Nous avons un dataLab au sein de la DSI composé de Data Scientists qui utilisent Dataiku en proximité avec les métiers et des équipes SI. Cette équipe travaille sur des expérimentations data en mode agile.

Une de ces expérimentations porte sur la prévision de la charge d'appels des plateaux téléphoniques du service Urgence Sécurité Gaz. En cas de suspicion de fuite ou de coupure de gaz, GRDF met à disposition un numéro d'appel gratuit pour les particuliers. Derrière ce numéro, il y a un plateau téléphonique où 140 professionnels de l'Urgence Sécurité Gaz se relaient 24h/24 7j/7 pour répondre aux appels. Cela représente un volume d'environ 700 00 appels/an.

### LA QUESTION QUI A ÉTÉ POSÉE AU DATALAB EST LA SUIVANTE : EST-IL POSSIBLE D'AMÉLIORER LA PRÉVISION DU NOMBRE D'APPELS POUR OPTIMISER LA PLANIFICATION DES RESSOURCES ET DES ACTIVITÉS DE CES ÉQUIPES ?

Les données historiques des appels des dernières années étant accessibles, nous avons pu rapidement tester nos algorithmes de Data Science. **Après quelques essais, nous avons obtenu une amélioration de la prévision à moyen terme de 15% par rapport à la méthode existante.** La prochaine étape consiste à intégrer au modèle les incidents réseaux et ainsi prévoir l'impact d'un incident sur le nombre d'appels à court terme (d'une heure à quelques jours) en fonction de sa nature, sa localisation ou encore de l'heure à laquelle il survient. Je vous cite cet exemple pour montrer qu'il est possible d'obtenir des quicks wins à partir du moment où les données sont accessibles et que l'on a les outils pour les traiter.

**Marc :** Intéressant! Un autre exemple de cas d'usage où vous avez utilisé des algorithmes de Machine Learning avec succès ?

**Paul :** Oui tout à fait. Comme vous le savez, **la donnée est en effet un véritable « actif » pour l'entreprise. C'est à la fois une source de valeur et une source de risque notamment du fait de l'application de la GDPR dès 2018.**

C'est dans ce contexte que notre Chief Data Security Officer nous a sollicité pour analyser les données saisies dans les zones de commentaires libres. Ces zones présentent en effet le risque de contenir des informations sensibles ou des informations sur la vie privée. Des règles strictes encadrent donc leur utilisation. La CNIL préconise la réalisation d'audits et le recours à des outils automatisés vérifiant les mots contenus dans ces zones. Elle fournit ainsi une liste de 337 mots "interdits".

On a récupéré 1 million de commentaires libres issus d'une de nos applications CRM. La première approche a été de chercher les mots "interdits".

Les limites de la méthode sont vite apparues :

- Certains mots interdits peuvent apparaître dans un commentaire conforme – par exemple : « le client est incapable de déter-

miner sur place quel est le bon compteur en cave. Le compteur se situe près du vide-ordure”

- Certains commentaires non conformes peuvent ne contenir aucun mot interdit.

Nous avons donc entrepris de construire un modèle capable de classer les commentaires : conforme ou non conforme. Notre Chief Data Security Officer a classé "à la main" un échantillon de commentaires qui a servi de base d'apprentissage à notre modèle. Le modèle a appris ce qu'était un commentaire conforme ou non conforme sur ces exemples et il sait maintenant classer automatiquement des commentaires qu'il n'a jamais "vus".

Après plusieurs étapes d'optimisation, les résultats sont au rendez-vous. Par rapport à la simple recherche de mots "interdits" 7 fois plus de commentaires non conformes sont détectés et le modèle se trompe 2 fois moins.

En d'autres termes : **en utilisant le Machine Learning via la plateforme Dataiku, nous avons une méthode qui améliore significativement la qualité et la précision de la détection.**

**Marc :** Selon toi, quelles sont les clés pour réussir une démarche data au sein de l'entreprise ?

**Paul :** Avec le recul que nous avons aujourd'hui, j'ai tendance à dire que tout ceci a été possible grâce à :

- Une organisation agile : équipe pluridisciplinaire / tests terrains / indicateurs de performance
- Une acceptation à tous les niveaux de l'entreprise du droit à l'erreur (on s'autorise à travailler sur un problème sans savoir si nous allons trouver une solution)
- Une démarche orientée "cas d'usages" plutôt qu'une course effrénée à la mise en place d'infrastructures Big Data

Cela nous a permis de positionner l'activité du dataLab au cœur du projet d'entreprise de GRDF, sur des sujets tels que la performance opérationnelle, l'expérience client réussie et la sécurité des réseaux.

Enfin, pour conclure sur notre partenariat fructueux avec Dataiku, j'utiliserai la formule suivante :

« GRDF est une grande entreprise qui a pensé son dataLab comme une start-up. Dataiku est une start-up qui a tous les atouts pour devenir une grande entreprise »





The power behind the moment.



The power behind the moment.

## Join DataStax at Big Data Paris

**MARCH 12 TO 13**

**HALL PASSY | ZONE A | STAND 25**

### LE SERVICE 6PLAY DE M6 WEB UTILISE DATASTAX ENTERPRISE POUR OFFRIR LA MEILLEURE EXPÉRIENCE PERSONNALISÉE À SES UTILISATEURS

En charge des offres digitales et interactives du Groupe M6, M6 Web accompagne les marques et chaînes du Groupe dans leur développement sur les nouveaux écrans (PC, IPTV, tablettes, smartphones) avec des offres de contenus innovantes.

M6 Web développe et exploite 6play, la plateforme de divertissement TV du Groupe. Elle permet de retrouver gratuitement un bouquet de chaînes de télévision en ligne : M6, W9, 6ter ainsi que 6 chaînes exclusivement on-line. 6play est l'offre de télévision digitale la plus accessible avec une présence sur tous les écrans (ordinateur, smartphone, tablette, TV, Xbox) et près de 3 millions de vidéos vues par jour : elle compte 7 millions d'utilisateurs à ce jour.

### LE CHALLENGE

Pour son service de rattrapage 6play, M6 Web devait se doter d'une base de données fiable et flexible permettant de stocker un grand nombre de préférences utilisateurs et de suivre au plus près leur consommation vidéo. « M6 Web collecte et stocke les préférences utilisateurs et suit leur consommation vidéo afin de pouvoir personnaliser au mieux l'affichage de leurs espaces et améliorer l'expérience utilisateur. Pour cela, nous avons besoin d'une base de données NoSQL dotée d'une scalabilité horizontale très simple », explique François Jomier, Responsable systèmes et réseaux chez 6Play.

Afin de choisir son système, M6 Web a effectué une évaluation approfondie des différentes bases de données à travers une veille technologique. Le choix s'est porté sur DataStax Enterprise, la version certifiée pour la production de la base données Apache Cassandra qui supporte les traitements transactionnels et analytiques tout en garantissant un service continu et une montée en charge linéaire. « DataStax Enterprise nous a convaincu pour son environnement de production, l'accompagnement et la sécurité, son prix abordable, sa simplicité d'administration et sa comptabilité avec notre environnement technique (PHP). » précise François Jomier. « Nous avons déjà mis en place un projet MySQL au périmètre similaire à DataStax il y a 3 ans, mais ce système ne nous avait pas permis de monter en charge. » ajoute-t-il.

### LA SOLUTION

Désormais utilisateur de DataStax Enterprise, M6 Web a pu constater la tenue en charge de la base de données. Les mécanismes de haute-disponibilité intégrés au produit, qui ne nécessitent pas la mise en place d'architecture d'équilibrage de charge, correspondent parfaitement à leurs attentes.

« Nous avons besoin d'un déploiement simple et rapide, la solution DataStax Enterprise s'est montrée très efficace. En moins de 3 semaines et avec l'aide d'une seule personne, nous avons pu mettre en place les environnements de développement et de pré-production », se réjouit Olivier Mansour, responsable R&D Backend et API chez 6play.

« Nous sommes ravis d'avoir été choisis par la marque 6play pour la mise en place du projet et la validation du modèle », se réjouit Xavier Guérin, Regional Vice-Président Europe de l'Ouest chez DataStax. « Nous leur avons également conseillé un environnement de production OpsCenter pour le backup et le monitoring. » poursuit-il. « DataStax apporte à nos clients un véritable confort, une sécurité mais aussi un accompagnement technique en plus d'une performance sans égal », conclut-il.

### LES RESULTATS

Grâce à DataStax Enterprise, la plateforme digitale 6play est capable de stocker et de traiter de vastes quantités de données sur les préférences utilisateurs mais également de suivre au plus près les consommations vidéo de plus de 7 millions d'utilisateurs à date. Le dispositif mis en place est vital aux projets de personnalisation de 6play.

« Le support de PHP par DataStax a été un élément déterminant dans la réussite de notre projet » ajoute Olivier Mansour, Responsable R&D Backend et API chez 6play. « Nous ne pourrions réellement pas nous passer de Cassandra pour notre projet. DataStax Enterprise fait partie des produits qui ont le bon goût de se faire oublier », conclut-il.



# Nagez sans risque dans votre Data Lake et accélérez votre projet Big Data

## PROFIL DE LA SOCIETE

Autodesk est un leader dans la conception 3D, d'ingénierie et de logiciels de divertissement. Depuis son introduction du logiciel AutoCAD en 1982, Autodesk continue à développer le plus vaste portefeuille de logiciels 3D pour les marchés mondiaux. Les clients à travers la fabrication, l'architecture, la construction et les médias et les industries du divertissement, utilisent les logiciels Autodesk pour concevoir, visualiser et simuler leurs idées avant qu'elles ne soient construites ou créées. Compte tenu de l'évolution des modes de consommation du logiciel, la société a décidé de transformer son modèle d'affaires de licences perpétuelles classiques à un modèle plus moderne de licence par abonnement pour augmenter les profits et propulser la croissance.

## BESOIN D'AFFAIRES

Le système existant Business Intelligence (BI) d'Autodesk ne pouvait pas soutenir ce changement crucial pour le modèle de revenus. La transition a affecté d'une part la capacité de la section des finances à suivre les abonnements, les renouvellements et les paiements, et d'autre part le système de BI, qui, avec son entrepôt de données opérationnelles, ne pouvait pas répondre aux demandes des parties prenantes des entreprises nécessitant de plus en plus des données de haute qualité en temps réel. Autodesk a rapidement décidé que l'évolution vers une architecture agile BI 2.0 était nécessaire avec un entrepôt de données logique à sa base. Plus précisément, cela impliquerait le déplacement de leur modèle physique existant pour une approche plus logique d'intégration des données. Suite à la décision, un entrepôt de données logique en utilisant la plate-forme Denodo pour la virtualisation de données a été mis en place pour répondre à ce changement.

## LA SOLUTION: ENTREPOT DONNEES LOGIQUE AVEC BIG DATA

Chez Autodesk, les données proviennent de multiples sources disparates de données telles que OLTP, des fichiers plats, des données géo-spatiales, des flux de données de médias sociaux, et les journaux Web. Les systèmes d'intégration traditionnels incluent l'ETL pour le traitement par lots de données provenant de bases de données relationnelles et Kafka (CSE) pour l'intégration des flux de données. En outre, Autodesk avait un lac de données d'entreprise, qui a logé des grandes données de plusieurs sources de données telles que Spark et Scala, tandis que OLTP et les données géo-spatiales étaient ETL-d et stockées dans un entrepôt de données opérationnelles. L'infrastructure a été lente et inefficace, et elle a été incapable de répondre aux demandes de renseignements d'affaires et de fournir l'agilité nécessaire pour répondre à leur transition vers un nouveau modèle de licence. La plate-forme Denodo pour la virtualisation de données a été introduite au service des finances d'Autodesk, qui permet de suivre les abonnements, les renouvellements et les paiements. La virtualisation des données a permis d'abstraire des données financières de leurs utilisateurs professionnels sans devoir déplacer ou transformer physiquement des données. En outre, Autodesk a réalisé la nécessité de protéger l'ensemble des données sensibles appartenant à l'entreprise. Il y a actuellement plusieurs gouvernances, le risque et la conformité des initiatives (GRC) conçues pour aider Autodesk à mieux gérer le traitement des données sensibles, ainsi que pour atténuer les risques. Sarbanes-Oxley (SOX) contrôle également la

plupart des initiatives internes d'Autodesk. La plate-forme Denodo permet la gestion de celles-ci et de plusieurs autres initiatives de conformité qui correspondent à d'autres lois sur la confidentialité. En utilisant la plate-forme Denodo comme un entrepôt de données logique, Autodesk a créé un seul point d'accès pour toutes les données utilisées dans l'entreprise (Fig. 1)



## AVANTAGES

Autodesk a transformé avec succès son modèle d'affaires en utilisant la plate-forme Denodo pour la virtualisation de données. L'avantage commercial majeur de la mise en œuvre Denodo a facilité la transition d'un modèle à base de licence perpétuelle à un modèle plus moderne sur abonnement. Ce changement a amélioré la performance de l'entreprise d'Autodesk à travers l'organisation, a renforcé la collaboration en partageant l'information en temps réel dans les entreprises et l'informatique, et modernisé leur entreprise à tous les niveaux afin d'assurer une plus grande agilité, performance et rentabilité. En outre, la solution a permis à des partenaires externes d'Autodesk de se connecter directement à l'application SAP ECC d'Autodesk à travers des services Web. Cette capacité leur a permis de traiter les paiements à proximité en temps réel. Autodesk n'a plus besoin de déplacer des données - ils peuvent utiliser moins d'ETL. Cela implique la réduction des copies de données. En outre, pour la première fois, Autodesk est en mesure d'effectuer le point d'application unique pour la sécurité et un environnement uniforme pour l'accès aux données en place. La solution Denodo a fourni à l'équipe de développement la flexibilité nécessaire pour comprendre ce dont ils ont besoin pour construire avant de pouvoir effectivement construire. Cela a permis à l'équipe de développement de planifier à l'avance, et voir ce que leurs équipes commerciales nécessitent à partir du produit avant de commencer. Cette approche a fourni l'agilité et a rendu la virtualisation de données attrayante pour la société. La solution de virtualisation de données a également aidé les équipes d'affaires et de développement d'Autodesk à devenir plus collaboratives.



Pour en savoir plus, visitez [denodo.com](http://denodo.com) ou contactez-nous au numéro +33 (0) 1 42 68 51 27



**DIGDASH ENTERPRISE EST UN LOGICIEL DE BUSINESS INTELLIGENCE AGILE (DATAVIZ, DATA DISCOVERY ET ANALYTIQUE) CENTRÉ SUR L'UTILISATEUR PERMETTANT AUX ENTREPRISES DE FÉDÉRER LEURS ACTIVITÉS DÉCISIONNELLES EN MODE COLLABORATIF.**

Aujourd'hui, plus de données signifie plus de possibilités, à condition de pouvoir les traiter et les comprendre. Les rapports d'activités mensuels et les solutions orientées "experts" ne suffisent plus à faire face aux très gros volumes de données (Big Data). DigDash Enterprise permet aux utilisateurs non informatiques de faire parler leurs données instantanément et de prendre les bonnes décisions au bon moment.

DigDash a été fondé par une équipe d'experts en informatique décisionnelle. Réunis autour d'une vision plus agile de la Business Intelligence, DigDash édite DigDash Enterprise, l'un des logiciels de DataViz les plus polyvalents du marché permettant de créer des tableaux de bord performants et simples à utiliser.

DigDash Enterprise se connecte facilement à tous types de données structurées ou non. Grâce à ses multiples connecteurs, le logiciel fonctionne aussi bien avec des sources de données structurées, semi structurées et non structurées. Il complète également les solutions décisionnelles déjà en place.

Les utilisateurs visualisent leurs données sans l'aide du service informatique et communiquent entre eux directement via les tableaux de bord. Les fonctionnalités avancées sont disponibles en standard.

En vous donnant accès à l'état de votre activité et à vos objectifs en temps réel, DigDash Enterprise utilise intelligemment vos données pour améliorer vos performances.

### PROJET : LE CEA CHOISIT UN DÉCISIONNEL AGILE

**LE CEA A MIS EN ŒUVRE DIGDASH ENTERPRISE POUR DISPOSER D'UN DÉCISIONNEL CENTRALISÉ TECHNIQUEMENT MAIS LAISSANT DE L'AUTONOMIE AUX UTILISATEURS FINAUX.**

Le Commissariat à l'énergie atomique et aux énergies alternatives (CEA) est un établissement public à caractère industriel et commercial à vocation de recherche et développement non seulement dans l'énergie (au départ, la seule énergie atomique) mais aussi dans les domaines connexes (y compris l'IT). Ses dix sites abritent 20 000 collaborateurs. En 2006, l'outil de dashboarding DashBoard de Business Object y a été déployé pour couvrir les besoins en décisionnel agile des utilisateurs finaux. Mais le produit a été arrêté en 2008 et n'est plus commercialisé depuis 2011.

Pour le remplacer, l'architecte applicatif du CEA, Damien Poulain, a longtemps cherché une solution garantissant une architecture centralisée pour gérer des données variées (finances, RH, achats...) et parfois sensibles (données relatives à l'énergie nucléaire, applications militaires, recherche...) mais permettant aux utilisateurs de concevoir leurs propres rapports. Il souhaitait en effet qu'il y ait une seule solution pour tout l'établissement sur tous ses sites. Deux solutions ont été testées auprès d'un panel de 23 utilisateurs représentatifs. Finalement, c'est DigDash qui a été choisi par ceux-ci.

Les très gros volumes de données et le nombre important d'utilisateurs ne posent pas de difficulté, la propagation des éléments partagés graphiques ou fonctionnels est rapide et le rafraîchissement des données du Datawarehouse peut être automatisé tandis que les utilisateurs finaux peuvent télécharger eux-mêmes leurs propres données.



## De la visibilité sur des milliards de données.



Pour en savoir plus : [www.digdash.com](http://www.digdash.com)





### EBIZNEXT, ACTEUR DE RÉFÉRENCE DANS LA FAST DATA

ebiznext s'est forgée une solide expérience dans la transformation digitale de ses clients, grâce à la complémentarité de ses consultants spécialisés dans la « fast data », de ses experts technologiques et de ses Data Scientists.

Au travers d'une démarche itérative et pragmatique, l'équipe ebiznext accompagne depuis plusieurs années ses clients dans la mise en œuvre de solutions destinées à approfondir la connaissance de leurs marchés (BtoB, BtoC, BtoBtoC...) et maximiser ainsi leur performance digitale.

ebiznext conseille et met en place une offre innovante de Datalake à la carte (de type DMP pour les marketeurs) qui s'appuie sur des technologies à la fois modernes et éprouvées : Spark / Mesos / Akka / Couchbase / Kafka / Elasticsearch

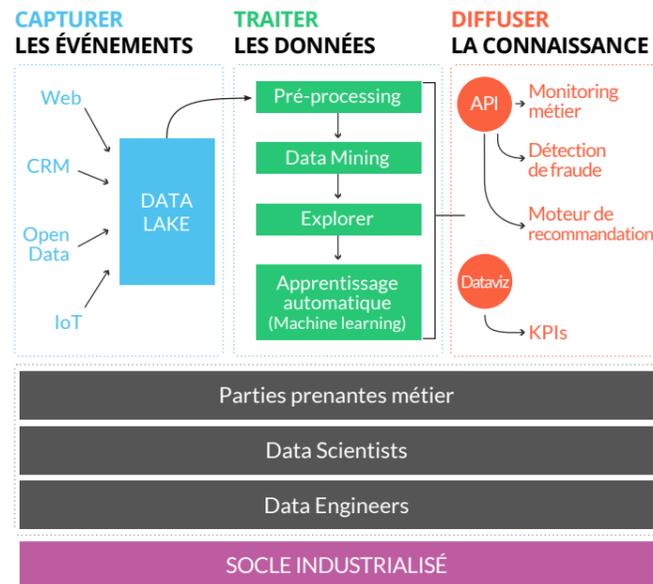
À titre d'exemple : ebiznext a conçu et mis en place un Datalake industrialisé apte à répondre à des cas d'usages nécessitant des données capturées, traitées et diffusées en temps réel.

### PROJET FAST DATA

L'objet de ce projet est de réaliser des calculs de mesures d'audiences à partir de données issues de différentes plateformes digitales (portail entreprise, réseaux sociaux, etc.)

L'architecture fonctionnelle se décompose comme suit :

- les services de mesure chargés de collecter la donnée,
- les services métiers chargés de calculer les agrégats à la volée et de persister les données en base,
- les services de présentation chargés de restituer les indicateurs aux utilisateurs finaux.



### PRÉCONISATION D'ARCHITECTURE

Nous proposons une architecture sur la base de briques OpenSource. Celle-ci permet, dans une approche Fast Data, de :

- répondre à des problématiques de forte volumétrie et de garantie de résilience aux pannes,
- restituer « au plus tôt » des indicateurs.

L'architecture applicative générique ci-dessous appliquée par ebiznext dans le cadre de ses projets se décompose en trois phases qui correspondent comme suit aux services fonctionnels attendus :

- les services de mesure sont implémentés par la couche de capture des événements,
- les services métiers sont implémentés par la couche de traitement des données,
- les services de présentation sont implémentés par la couche de présentation.

### Capture des événements (services de mesures)

- Les points de vigilance sont les suivants :
- Sécurité : s'assurer que les composants clients sont autorisés à émettre des messages vers le Datalake.
  - Validité des données : s'assurer que les schémas des événements transmis par les composants respectent le schéma de données attendu par le Datalake.
  - Performance : s'assurer qu'un pic d'émission d'événements ne va pas perturber les traitements

La suite Confluent incluant le middleware Kafka, un schéma registry et un proxy HTTP/REST, répond aux problématiques listées ci-dessus.

### Traitement des données (services métiers)

Nous proposons un pipeline de traitement basé sur le framework Kafka Streams de la plateforme Confluent. La framework Kafka Connect est mis en place pour pousser ces agrégats vers une ou plusieurs bases tel que Elasticsearch ou encore vers le système de fichiers distribués tel que HDFS.

### Diffusion de la connaissance (Portail utilisateurs)

Les informations métiers sont ensuite diffusées via une API de présentation ou un outil de DataViz.

### CONTRIBUTIONS EBIZNEXT

Parallèlement à la réalisation de projets fast data, les consultants ebiznext animent différents types d'événements :

- talks et workshops (ScalalO, Devovx, PSUG, JUG, Data Marketing Paris...),
- meetups intra-entreprises et hackathons (thématiques : IoT, Machine Learning...),
- formations qualifiantes (Data ScienceTech Institute, ENSAI, Centrale Paris...),
- formations certifiantes (Lightbend, Couchbase...).

ebiznext apporte par ailleurs des contributions majeures sur [blog.ebiznext.com](http://blog.ebiznext.com) et [github.com/ebiznext](https://github.com/ebiznext).





### HISTORIQUE

Ekimetrics est une société de conseil en Data Science. La société a été fondée en 2006 à Paris et compte à présent plus de 180 consultants basés à Paris, Londres, New York, Hong Kong et Dubaï. L'approche pionnière et rigoureuse dans le traitement de la data a permis à Ekimetrics de devenir le leader européen indépendant du ROI et du conseil en Data Science.

### SAVOIR-FAIRE

Ekimetrics couvre l'ensemble de la chaîne de valeur de la data et déploie une expertise avancée dans le conseil en technologie, la gouvernance des données, le cadrage stratégique et l'utilisation des algorithmes de machine learning à des fins business et marketing. Les consultants d'Ekimetrics disposent d'une double expertise business et mathématiques appliquées, pour identifier et mettre en place les méthodologies les plus adaptées aux questions posées.

### PRESTATIONS / PRODUITS / SERVICES PROPOSES

Chaque modèle créé adresse une question spécifique soulevée par les clients et tient compte des particularités business en décollant : marché, chaîne de valeur, accessibilité des données, etc. La mesure de performance est pensée en amont de chaque projet. Ekimetrics privilégie une approche agile, avec des boucles de modélisation et de test & learn.

### SECTEURS CIBLES

Une approche agile et des méthodologies propriétaires permettent à Ekimetrics de s'adapter à une grande variété de problématiques et d'industries. Le cabinet conseille depuis plusieurs années des leaders européens et internationaux issus de secteurs aussi variés que le luxe, l'automobile, la banque-assurance, la cosmétique, la grande consommation ou l'énergie.

### COUVERTURE GEOGRAPHIQUE

Historiquement Ekimetrics est très présent en Europe, le cabinet a su évoluer et couvrir différentes zones pour accompagner

ses clients dans leur stratégie internationale. A présent doté de cinq bureaux : Paris, Londres, New York, Hong Kong et Dubaï ; Ekimetrics bénéficie d'une réelle couverture à l'international et compte parmi ses salariés plus de 25 nationalités.

### NOUVEAUTES PRODUITS ET SERVICES

L'équipe Data Operation Team qui pilote les projets BI & Big Data au sein d'Ekimetrics se renforce continuellement pour assurer la réalisation et le déploiement de projets Analytics on premise et sur le cloud. Ekimetrics est actuellement à la pointe dans l'utilisation des distributions Hadoop et Spark. Cette année Ekimetrics a aussi étendu ses équipes en Machine Learning et en product engineering pour répondre aux besoins croissants de ses clients dans les domaines algorithmiques et applicatifs - applications web, dataviz, outils d'aide à la décision. La capacité à concevoir et déployer des architectures Big Data complexes conjuguée à la maîtrise du machine learning et des langages applicatifs permet à Ekimetrics d'offrir une expertise unique en Data Science dans l'hexagone et en Europe.

### AVANTAGES CONCURRENTIELS

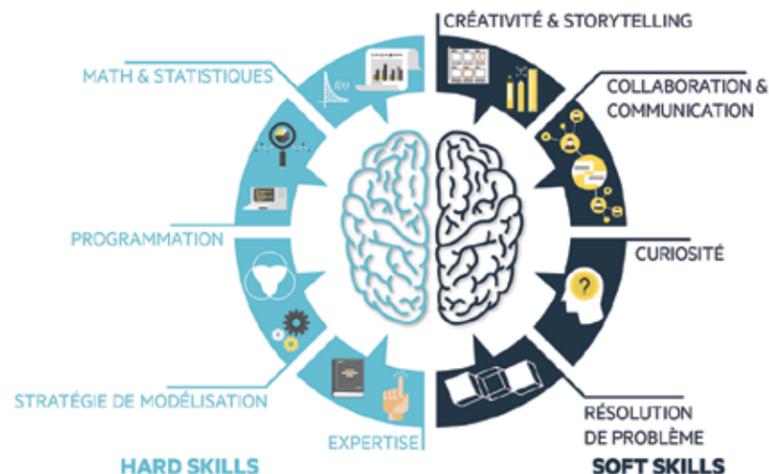
Ekimetrics est indépendant de tout organisme de médias et d'éditeurs de logiciels, ce qui permet au cabinet une grande agilité dans le choix des outils et des méthodes, ainsi qu'une réelle impartialité dans les recommandations. Les 180 consultants ont tous un double profil consultant / statisticien-Data Scientist (X, ENSAE, Telecom, Mines, Columbia, Imperial College). Ekimetrics centre son approche sur les besoins business des clients pour construire la méthodologie et l'architecture technique adaptée. Grâce à une triple expertise stratégique, statistique et technologique, les consultants sont à même de livrer des recommandations actionnables par les CIOs, CDOs et les CMOs. Ekimetrics accompagne ses clients pour optimiser l'ensemble de la chaîne de valeur de la data et joue un rôle d'interface entre les métiers, la DSI et les analytics sur des projets data transversaux.

“ Un projet Big Data bien mené nécessite la mise en place d'une organisation facilitant la bonne connexion entre l'équipe Big Data et les métiers. ”

François Poitrine  
DG et Co-fondateur Ekimetrics



### Dans le cerveau d'un Data Scientist



### Voir clair. Voir loin. Voir grand.

Pionnier et référent en datascience depuis 2006, notre métier est d'aider les grands groupes à identifier leurs leviers de croissance les plus importants, en nous appuyant sur l'analyse fine de leurs données. En 10 ans, nous avons généré plus de 1 milliard d'euros de gains cumulés pour nos 350 clients, dans 40 pays.



Connaissance clients



Optimisation des processus



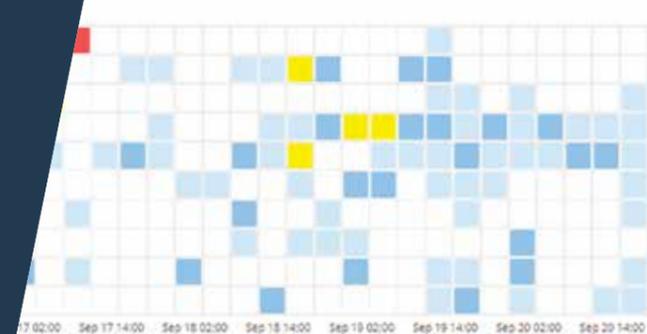
Développement nouveaux services



Détection des fraudes



elastic



Fondé en 2012, Elastic est le premier fournisseur mondial de logiciels permettant d'exploiter des données structurées et non structurées en temps réel, et ce de façon flexible. Ingérez, convertissez et enrichissez des données issues de n'importe quelle source ; puis recherchez, analysez et visualisez-les en temps réel.

### UN DÉPLOIEMENT ADAPTÉ À TOUS VOS BESOINS

#### La Suite Elastic :

- Le coeur open source composé d'Elasticsearch, Kibana, Beats et Logstash
- Les fonctionnalités commerciales X-Pack : security, monitoring, alerting, reporting, Graph et machine learning

#### Elastic Cloud :

- La Suite Elastic en quelques clics
- Les dernières versions logiciels disponibles immédiatement
- Kibana et backups inclus et gratuits
- Infrastructure surveillée 24h/24 et 7j/7

#### Elastic Cloud Enterprise (ECE) :

- Un cloud Elastic dans votre data center
- Gestion de plusieurs cas d'usage
- Fournir de l'Elastic as a service à votre organisation
- Backup et restauration des clusters avec l'API Elasticsearch Snapshot

### NOS CLIENTS

Les produits Elastic activent plusieurs cas d'usage stratégiques tels que les applications de recherche, le logging, les analyses sécuritaires, les analyses d'indicateurs et le développement d'applications de données évolutives en temps réel. Nos produits sont adoptés par des milliers d'entreprises telles que la SNCF, Renault, Airbus, Société Générale, BNP Paribas, Orange, Carrefour, Enedis, BlaBlaCar, BUT, ContentSquare et plus encore.

### CRÉATION DE VALEUR AJOUTÉE POUR TOUS LES CAS D'USAGE

Pour toute entreprise, la Suite Elastic s'avère être l'outil le plus performant pour collecter et analyser des ensembles de données volumineux, souvent hétérogènes, et en dégager de la valeur en temps réel. De nos jours, l'abondance et le volume de données disponibles signifient qu'il est d'autant plus vital pour les développeurs de créer des applications capables d'offrir une valeur immédiate à leur entreprise. Grâce à la Suite Elastic combinée avec X-Pack (fonctionnalités pour la sécurité, l'alerting, le monitoring, le reporting, l'analyse de graphes et le machine learning), les développeurs et utilisateurs finaux sont en mesure d'injecter de la valeur quel que soit le cas d'usage.

#### Search :

- Tirez profit de la recherche plein texte en temps réel
- Recherchez à travers tous types de données, qu'elles soient structurées ou non structurées
- Optimisez l'expérience de recherche de vos utilisateurs avec des informations sur les termes les plus recherchés, les délais de réponse et les types de résultats
- Choisissez comment ajuster la pertinence, mettre en œuvre des outils d'analyse et booster les résultats

#### Log Analytics :

- Optimisez la performance du système, identifiez les problèmes en temps réel et gérez l'évolutivité
- Respectez les SLA de service client, garantisiez la disponibilité MTTI/MTTR, les opérations de développement à distance
- Exploitez les données et créez de nouvelles opportunités de revenus à l'aide d'indicateurs.

#### Security Analytics :

- Détectez automatiquement les anomalies et les tendances cachées au sein des données qui sont susceptibles de poser un risque de sécurité
- Autorisez la détection des menaces en temps réel pour les initiatives de sécurité informatique et de lutte contre la fraude
- Respectez les normes de conformité ; limitez les risques financiers et les risques encourus par les utilisateurs

#### Business Analytics :

- Boostez la visibilité au sein de votre entreprise
- Prenez des décisions éclairées basées sur des analyses en temps réel
- Détectez facilement les KPI anormaux, tels qu'une chute soudaine des commandes ou du taux de clics

#### Metrics Analytics :

- Scalez sans être soumis à des restrictions de coût
- Associez vos indicateurs à d'autres données, telles les logs d'exploitation
- Réduisez vos coûts de fonctionnement et centralisez vos analyses
- Allez au-delà de l'index inversé pour des recherches plus rapides et une utilisation réduite de la mémoire et de l'espace disque

### SUITE ELASTIC + MACHINE LEARNING

Avec des ensembles de données complexes et dynamiques, il est pratiquement impossible de repérer des problèmes d'infrastructure, des intrusions ou des problèmes métier en utilisant des règles spécifiques ou un personnel scrutant des tableaux de bord. Les fonctionnalités de machine learning de X-Pack modélisent automatiquement le comportement de vos données Elasticsearch (tendances, périodicité, etc.), et ce, en temps réel afin d'identifier les problèmes plus rapidement, de rationaliser l'analyse des causes fondamentales et de réduire les faux positifs.

Pour découvrir comment Elastic peut vous aider et en savoir plus à propos de nos formules d'abonnement, n'hésitez pas à nous contacter. Ou testez Elastic Cloud gratuitement pendant 14 jours pour découvrir la Suite Elastic en pleine action : [elastic.co/fr/cloud](http://elastic.co/fr/cloud)

### COMMENT LA CROISSANCE EXPONENTIELLE DES DONNÉES ENTRANTES A-T-ELLE AFFECTÉ VOTRE ENTREPRISE ?

L'activité de Sunhotels, société européenne de services de voyages en ligne a toujours été fondée sur l'acquisition de données. Par conséquent, l'adoption des technologies open source, en particulier Elasticsearch, nous aide à mieux comprendre nos données client et à améliorer nos services en conséquence.

Récemment, le volume des données que nous traitons a connu une croissance exponentielle. Nous traitons les demandes de recherche et de réservation de plusieurs milliers d'agences de voyages, fournisseurs B2B et agrégateurs tiers à travers toute l'Europe. En deux ans, le nombre moyen de requêtes par seconde est passé de 600 à plus de 4 000, soit 250 millions de demandes et des milliers de réservations par jour.

### COMMENT LES TECHNOLOGIES ELASTIC VOUS ONT-ELLES AIDÉ À GÉRER LA RÉVOLUTION DES DONNÉES ?

Depuis plusieurs années, Elasticsearch nous permet d'étudier les temps de réponse des recherches, de trouver un moyen de différencier les types de recherche et leurs résultats, et de capturer différents flux de métadonnées afin de mieux comprendre chaque interaction et de réagir plus rapidement en cas de problème. Nous avons commencé à consigner 15 champs différents sur chaque recherche et réservation, notamment le temps de réponse, la destination, le type de client, le contrat du produit (tiers ou direct).

Lorsque nous ajoutons toutes ces données brutes dans Elasticsearch, nous voyons les temps de réponse globaux à un niveau granulaire, et pouvons analyser et comprendre tous les facteurs susceptibles d'affecter la fréquentation d'un hôtel : faible disponibilité, saisonnalité et sensibilité aux prix. Grâce à ces informations, nous suggérons des solutions, telles que modification des tarifs, amélioration de la sous-traitance/cartographie/disponibilité ou connexion/déconnexion aux hôtels, destinations, fournisseurs ou autres clients.

Avec SQL, il était difficile de suivre les recherches en même temps que les types de réservations effectuées, et de trouver des angles morts de ventes. Maintenant, toute l'entreprise utilise les mêmes tableaux de bord que ceux du service d'exploitation, si bien que nous disposons d'informations pertinentes pour créer des stratégies à long terme.

### EN QUOI CELA EST-IL UTILE POUR LA STRATÉGIE D'INNOVATION ET DE CROISSANCE DE SUNHOTELS ?

Nos partenaires et nous-mêmes tirons profit de cette infrastructure de gestion des données de deux façons. Tout d'abord, nous détectons les changements de comportement plus rapidement et réagissons de façon proactive. En fonction des tendances qui se dégagent au fil du temps, nous sommes en mesure de prédire les comportements et d'adapter le service en conséquence.

Pouvoir prédire les comportements à partir de l'historique des données signifie que nous pouvons être proactifs et offrir à notre équipe commerciale l'information dont elle a besoin pour agir rapidement. Si nous prévoyons une demande élevée, cela peut se

concrétiser par la multiplication des propriétés disponibles dans une certaine zone ou à un certain prix. Cela peut également entraîner une augmentation de l'activité dans des régions plus calmes, notamment en améliorant les réservations et en avertissant les partenaires des périodes calmes ou chargées.

Deuxièmement, notre capacité à traiter plusieurs téraoctets de données dans Elasticsearch allège considérablement notre organisation. L'industrie du voyage - particulièrement en B2B - est très traditionnelle avec de grandes équipes chargées de réaliser des analyses approfondies. Elasticsearch facilitant le traitement du volume de données, notre équipe peut se concentrer sur les résultats des analyses pour améliorer le service, ce qui réduit considérablement les délais de traitement et le nombre d'heures de travail.

### QUEL RÔLE LE MACHINE LEARNING JOUERA-T-IL DANS LA STRATÉGIE DE SUNHOTELS ?

Puisque l'entreprise compte sur nous pour lui fournir des informations plus poussées, les fonctionnalités machine learning dans X-Pack, nous permettent d'automatiser une importante couche d'analyse actuellement gérée par 12 robots conçus sur mesure, eux-mêmes supervisés par une toute petite équipe d'ingénieurs. Nous pouvons aujourd'hui consacrer beaucoup de temps à la surveillance et à l'optimisation des paramètres, alors que nous avons dix fois plus de données, de connexions externes et de requêtes entrantes que nous n'en avons l'année dernière. Nous avons donc trouvé un moyen d'automatiser à grande échelle des analyses complexes.

Si un de nos clients fait 150 réservations par jour, nos robots peuvent examiner certains des comportements et des anomalies associées à ces transactions, mais, pour émettre des hypothèses sur la relation entre les points, un processus de codage complexe doit être exécuté. Sans le machine learning, il serait très difficile d'effectuer des analyses croisées et de sélectionner des indicateurs de mesure provenant de plusieurs index.

Nous voulons également que l'infrastructure prenne automatiquement en compte la saisonnalité. En temps normal, nous devons ajouter manuellement ce paramètre à tous les robots, ce qui prend beaucoup de temps et est très technique. Le machine learning nous aide à lancer des analyses conformément à des règles cohérentes, implémentées automatiquement.

Une multitude de clusters et de nœuds étant en jeu, nous savions que d'importantes sommes d'argent devraient être engagées. C'est pourquoi, après avoir enquêté sur un certain nombre de solutions de machine learning, nous avons été ravis de voir que cette fonctionnalité soit ajoutée à X-Pack. Celle-ci peut être facilement déployée dans l'écosystème Elastic existant sans nécessiter une intégration particulière ou des informations de connexion distinctes.



**MARKETING MIX MODELING - AGROALIMENTAIRE**  
 « Optimiser l'allocation des budgets marketing en associant Advanced Analytics & Management de la Performance »

**Quand la Data Science booste le "MMM" pour doper les ventes**  
 L'émergence des médias en ligne a largement complexifié le sujet de l'allocation des budgets et rendu la tâche difficile aux équipes marketing pour établir leur stratégie d'investissement média. L'enjeu de notre mission a consisté à faire évoluer les pratiques d'allocation budgétaire pour maximiser l'impact combiné des canaux sur les ventes.

**Choisir le bon dispositif pour réduire le "time to value"**  
 Nous avons opté pour la mise en œuvre d'un "Proof of Value" industrialisable pour des résultats tangibles et rapides, en fédérant le Marketing & la DSI et pour réduire le "time to value". Cette démarche s'appuie sur plusieurs de nos méthodes et meilleures pratiques, sur les compétences de notre squad analytics aidée des solutions technologiques de nos partenaires : Dataiku DSS & Python. Pour supporter ce use case, notre équipe a délivré un modèle de prescription, appuyé sur un algorithme d'optimisation spécifiquement créé pour l'occasion.

**Data to Action !**  
 Plus qu'un simple algorithme, nos insights sont activables et éclairent les prises de décision du service marketing. Le processus d'attribution s'est enrichi en permettant aux utilisateurs d'accéder à un moteur de recommandation et de simulation.

**Création de valeur**  
 • Une meilleure compréhension de l'impact des campagnes et optimisation des dépenses marketing  
 • Une autonomie du marketing dans l'exercice de répartition budgétaire en s'appuyant sur un moteur de recommandation et de simulation (what-if Analysis)

**DÉTECTION DE FRAUDES - ASSURANCES**  
 « Utiliser le levier de la Data Science pour accroître la précision et la productivité dans la détection des dossiers frauduleux »

**La fraude en assurances pèse sur les marges et pénalise les assurés honnêtes**

La maîtrise du phénomène de fraude constitue un enjeu majeur de préservation de la rentabilité des assurances et de la confiance des clients dans le réseau des professionnels de santé. Pour faire face, le contrôle systématique par l'humain basé sur l'intuition est non viable et improductif. Comment éviter des faux positifs ? Comment augmenter la productivité de ses équipes ? Comment faire face à l'évolution et à la sophistication du processus de fraude ?

**Passer de simples règles métiers à de la détection intelligente ...**  
 Certes meilleur qu'un choix aléatoire, le processus en place utilisait des règles métiers sophistiquées mais statiques pour identifier des dossiers suspects, limitant ainsi la capacité à détecter de nouveaux schémas de fraude. Nous avons innové avec notre client en lui proposant d'utiliser des algorithmes d'apprentissage avancés pour fiabiliser et augmenter la pertinence de la sélection des dossiers.  
 ... via une démarche d'innovation rapide

La success story débute par un "Proof of Value" conduit par Eulidia et ses partenaires : Dataiku & Teralab, providers d'un environnement de stockage des données sensibles souverain et d'un studio

facteur de productivité pour nos équipes dans la mise en œuvre du modèle. Après quelques semaines, l'infrastructure et le modèle seront déployés en production et l'équipe interne formée à la mise à jour et à la supervision du processus prédictif pour permettre aux auditeurs de prioriser leurs contrôles.

**Création de valeur**  
 • Une meilleure connaissance des variables explicatives de la fraude pour mieux manager le réseau de partenaires  
 • Une performance de détection 3 fois supérieure au dispositif précédent, facteur de productivité pour l'équipe d'audit  
 • Un transfert de connaissances aux utilisateurs et l'appui d'un studio Data Science pour la mise à jour et la supervision du modèle de manière autonome

**CHOIX D'INFRASTRUCTURE - RÉASSURANCE**  
 « Architecture cloud : une piste pour réduire les coûts et optimiser les performances des calculs scientifiques ? »

**Le calcul scientifique comme pierre angulaire du métier de réassureur**  
 Le cœur de métier du réassureur repose sur la capacité à modéliser les risques et à en déterminer les niveaux de couverture. Pour agir en tant que spécialiste des catastrophes naturelles et des risques non-assurables notre client s'appuie sur des algorithmes scientifiques lourds, exécutés sur une architecture de calculs distribués. Face à une croissance exponentielle des traitements et à une contrainte de scalabilité, un diagnostic a été lancé pour répondre à 3 enjeux : pérennité, scalabilité & optimisation des investissements.

**Fédérer Métiers et DSI autour des enjeux du calcul scientifique**  
 L'étude s'est articulée autour d'ateliers métiers et techniques pour tenir compte du fait qu'un changement de support de calcul dépasse le simple sujet technologique. Permettre la parallélisation des traitements de plusieurs équipes, garantir une scalabilité instantanée de la plate-forme, simplifier l'effort d'exploitation et contribuer à l'effort collectif de réduction des coûts informatiques ont constitué l'essentiel des drivers de choix.

**Tirer parti de l'agilité et du modèle du cloud**  
 Parmi les scénarios de remplacement de la grille, une option hybride on premise / cloud a été retenue, sur la base d'une séquence d'investissement progressive et réaliste. Offrant une élasticité importante et un business model "pay as you use" favorable au contexte de notre réassureur, ce dispositif présente également l'intérêt de libérer le client des contraintes d'administration de la plate-forme et de conforter l'IT dans sa capacité à apporter des services performants et agiles à des métiers stratégiques.

**Création de valeur**  
 • Réduction des coûts fixes de la grille de calcul  
 • Amélioration du ratio coût complet / puissance du dispositif



# eulidia

BUSINESS PERFORMANCE PARTNER

TURN DATA INTO  
**BUSINESS  
 VALUE**

Mixing Business Intelligence & Data Science



## EXPERIS CRÉE L'ASSISTANT NUMÉRIQUE DU RECRUTEUR, LA DIGITAL ASSESSMENT ROOM

Experis est la marque mondiale ManpowerGroup dédiée au conseil et au recrutement de cadres. Experis accompagne ses clients dans leur transformation digitale en répondant à leurs besoins de compétences et de conseil, selon les modalités les plus appropriées à leur contexte : recrutement, mise à disposition de ressources, sous-traitance, externalisation, etc.

### LE DATA LAB

Le Data Lab Experis, véritable accélérateur d'incubations de projets, s'appuie sur les principes de modélisation et de prédiction pour proposer une approche innovante aux métiers. Il met en perspective les relations entre les données et la façon dont elles vont répondre aux enjeux Business et créer de la valeur. Une idée, un besoin spécifique ? Le Data Lab Experis participe à la conception et à la réalisation de votre projet. En intégrant vos équipes, nos consultants s'inscrivent au cœur de votre stratégie afin de cerner au mieux vos attentes. Data Scientistes, Ingénieurs Etudes & Développement, Data Analystes, vous apportent les meilleurs outils et technologies pour valoriser vos données.

Nos clients, DRH et DSI en support de la transformation numérique, nous sollicitent pour apporter des solutions innovantes aux modes de gestion des Ressources Humaines afin d'optimiser les processus et générer de la valeur. Le Data Lab Experis a exploré les dernières avancées technologiques, au service des différentes étapes d'un recrutement, pour optimiser la précision du matching profil de poste / profil de candidat. C'est ainsi que notre prototype d'assistant numérique du recruteur, solution d'entretien virtuel, a vu le jour.

La Digital Assessment Room apporte de nouveaux éclairages pour optimiser les entretiens de recrutement. Fruit de notre expertise en transformation numérique et en évaluation, elle capitalise également sur les champs d'études de l'intelligence artificielle : data science, apprentissage automatique, traitement du langage naturel, reconnaissance de la parole et des visages. La Digital Assessment Room révolutionne l'expérience de l'entretien pour les candidats et les recruteurs, en proposant une analyse plus complète des compétences et des comportements.

### LA DIGITAL ASSESSMENT ROOM

#### • Le concept

La Digital Assessment Room propose une application permettant de faire passer des entretiens à des candidats avec un recruteur virtuel. Un avatar leur pose des questions, choisies en fonction du poste à pourvoir, comme il est d'usage en entretien de recrutement. Une caméra et un micro captent en temps réel leurs gestes, leurs sourires, leurs paroles et les tonalités de leurs voix. Ces données sont stockées et analysées. Des algorithmes basés sur l'analyse du langage corporel naturel évaluent différentes caractéristiques : la présence physique, la confiance... La solution retranscrit les réponses orales à l'écrit et permet d'identifier des mots clés et des champs lexicaux en rapport avec le profil de poste. Une fois l'entretien terminé, les candidats répondent à un questionnaire de personnalité, construit spécifiquement pour le métier concerné, qui permet de compléter l'évaluation. Le candi-

dat reçoit dans la foulée un rapport d'entretien détaillé avec des conseils pratiques, issus des analyses conjointes des enregistrements vidéo & audio, et des réponses au quiz de personnalité.

Notre approche est orientée « co-working », au plus proche des spécificités de nos clients, de manière à proposer des solutions personnalisées et évolutives, répondant efficacement aux besoins et attentes. Pour que l'intégration de l'analytique RH soit performante pour l'entreprise et les collaborateurs / candidats, il est bien évidemment nécessaire de veiller à respecter les règles du droit du travail et la réglementation française & européenne sur les données à caractère personnel. Direction Juridique et « Chief Data Officer / Data Privacy Officer », sont toujours associés à nos projets pour en assurer la régularité et la bonne gouvernance.

#### • Des retours d'expérience prometteurs

Un premier prototype de cette application a été présenté en avant-première lors du salon VivaTechnology 2017 en juin dernier. Les retours candidats sont prometteurs : 9 sur 10 déclarent avoir apprécié l'expérience et reconnaissent son côté novateur. Tous ont apprécié le côté ludique et les recommandations personnalisées, très opérationnelles. Tant et si bien que la Digital Assessment Room a été invitée à l'AI Hackademy Microsoft experiences17.

#### • Une solution pour accompagner candidats et recruteurs

Du côté des candidats, la Digital Assessment Room propose une expérience originale pour mieux se connaître en termes de présence, d'empathie et de personnalité, grâce à l'analyse de leur registre de communication. La conversion de l'audio en texte apporte des enseignements sur les cheminements de pensée et la richesse lexicale. En clair, avec cette solution, les candidats bénéficient d'un coaching très opérationnel pour mieux préparer leurs entretiens de recrutement, qu'ils soient jeunes diplômés, expérimentés en recherche d'emploi, en reclassement, ou encore en attente de mission (cas des intérimaires et des consultants en prestation de services).

Côté recruteurs, la Digital Assessment Room trouve des applications pour optimiser la gestion de campagnes de recrutement massives, quand les candidatures affluent, quand les compétences sont difficiles à évaluer sur la seule base du CV. L'entretien virtuel permet de fiabiliser et d'accélérer la phase de présélection des candidatures, sur la base des compétences propres à un métier, « hard & soft skills », évaluées en toute objectivité. En écartant les biais subconscients, par exemple sur l'âge, le genre ou l'origine, la solution se focalise sur la performance.

La Digital Assessment Room représente une parfaite illustration de la capacité d'innovation technologique de notre Data Lab, en lien étroit avec les métiers, sur toutes les étapes d'intégration, de stockage, d'analyse et de restitution des données.

#### Nos domaines d'expertise :

- Architectures BI et Big Data
- Développements applicatifs
- Process Mining
- Data visualisation

## BI • BIG DATA DATA SCIENCE

# Révélez la richesse de vos données

**Nos experts interviennent pour valoriser l'ensemble des données de votre entreprise et optimiser sa performance**

Diagnostic	Data Visualisation
Gestion de Projet	Reporting
Prototypage POC / POV	Data Management

Contactez-nous  
[www.it.experis-france.fr](http://www.it.experis-france.fr)



Experis IT  
Conseil et innovation en transformation digitale

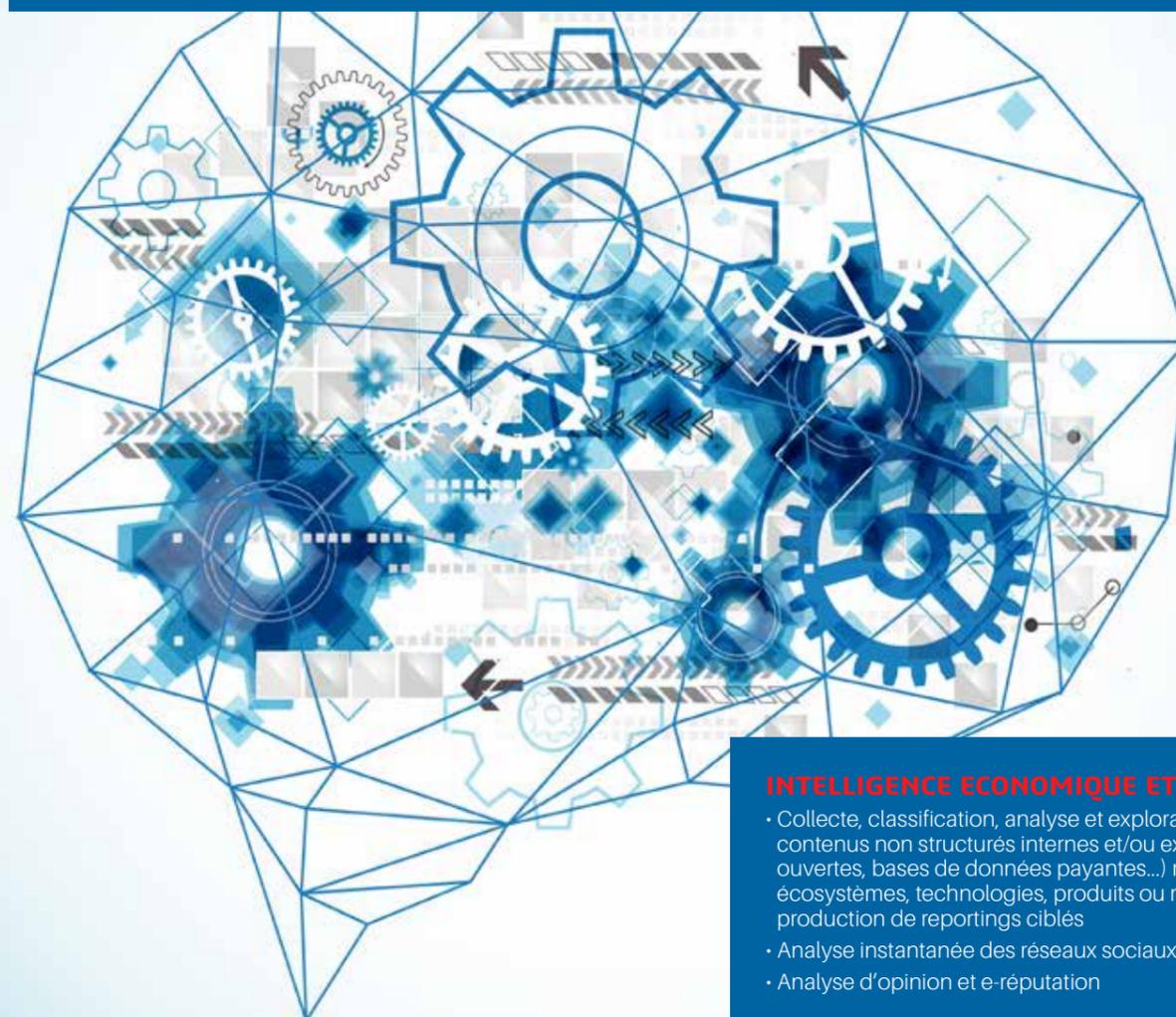
- MOA, gouvernance, gestion de projet
- Développement agile et architecture
- BI, big data, data science



Experis®  
ManpowerGroup



# REVELEZ L'INTELLIGENCE DES BIG DATA TEXTUELLES



## L'INTELLIGENCE ARTIFICIELLE INVESTIT LE SECTEUR BANCAIRE

Par Alain Biancardi, Vice Président Sales Expert System France

Aujourd'hui, l'intelligence artificielle permet de traiter un formulaire de demande en 5 secondes, plutôt qu'en 58 minutes. Appliquée à un secteur bancaire en pleine mutation, cette technologie a beaucoup à apporter à l'un des plus vieux métiers du monde qui doit se moderniser d'urgence afin de mieux répondre aux attentes de ses clients et de les conserver, mais aussi pour rester compétitif face aux fintechs qui se multiplient. La mutation est en marche.

### TRANSFORMER L'EXPÉRIENCE CLIENT

Aujourd'hui, pour la banque comme pour de très nombreux secteurs, exploiter l'information est devenu primordial : c'est une ressource inestimable pour un organisme financier. L'intelligence artificielle intervient dès ce stade : une partie des processus de compréhension sémantique et d'analyse des contenus textuels peuvent – et doivent – être traités par des technologies telles que l'IA et le machine learning.

La recherche en langage naturel transforme ainsi l'expérience client via les chatbots et les forums de questions-réponses intelligentes sur les sites web. Plus les questions traitées en langage naturel – par opposition à la recherche par mots-clés – sont adressées précisément, plus le client est facilement et rapidement satisfait, à la hauteur d'un service humain mais affranchi de tout risque de stress ou de conflit et en mode 24h/24 et 7j/7.

Parce que certaines catégories de clients ne passeront jamais au 100% digital, il est également possible de réduire la durée d'appel SAV en apportant une assistance intelligente aux agents. Les conséquences sont immédiates : d'une part, les coûts fondent et les performances des centres d'appels sont optimisées, et d'autre part, la satisfaction client connaît une forte croissance.

L'analyse multicanal de la voix du client permet quant à elle d'identifier les sujets d'intérêt des clients ou un potentiel risque d'attrition de leur part. Elle permet également d'adopter des actions correctives efficaces en cas de détection de problème et de capter les signaux faibles susceptibles de nourrir la stratégie marketing du groupe lors de la phase remodelage de ses offres de produits et de services. Cette reconnaissance par l'IA permet d'opérer une véritable révolution dans l'expérience utilisateur telle qu'elle existe aujourd'hui. Cette dynamique nous conduit vers un nouveau type de relation de proximité «client-banque» où le lien permanent n'est plus uniquement réservé aux banques en ligne, mais démocratisé à l'ensemble des établissements bancaires. Ou comment concurrencer ces fintechs tout en modernisant son image...

### UNE CAPACITÉ D'ANALYSE DÉMULTIPLIÉE

Déployée au service de l'intelligence économique et marketing, l'IA peut être un véritable atout pour les banques, notamment en leur permettant de construire une vision 360° de leurs clients. Pour cela, l'intelligence artificielle peut réaliser une veille sur le web et les réseaux sociaux afin d'évaluer l'e-réputation d'un établissement ou analyser les verbatim et les emails remontés par les agences afin de qualifier la relation client et de mesurer le taux de satisfaction associé. Elle peut également être un outil de choix pour atteindre un temps de réaction plus compétitif au niveau du traitement des retours des clients.

Cette veille doit aussi s'inscrire dans le cadre des benchmarks : c'est aujourd'hui un outil-clé pour étudier les produits et services, le type de communication ou encore le mode de relation-client adoptés par des établissements challengers, et adapter sa stratégie en conséquence.

Par ailleurs, l'assistance d'une intelligence artificielle peut être un atout concret dans un objectif de réduction des risques opérationnels et contractuels, notamment dans le cadre de la lutte contre le blanchiment d'argent. En effet, en analysant, selon des règles définies, des contrats signés afin d'en déduire un taux de risque associé dans le cadre d'un investissement ou d'un crédit par exemple, l'IA peut compresser le temps d'investigation requis en le réduisant de plusieurs semaines à quelques minutes ! L'intelligence artificielle sémantique peut ainsi apporter une réponse quasi instantanée grâce à l'analyse des données à partir de différentes sources privées et publiques. L'exposition au risque peut ainsi être réduite.

### AUTOMATISER LES PROCESSUS MÉTIER

À l'heure actuelle, la banque pourrait gagner à réduire le temps de traitement des demandes. En effet, des déclarations de type vol/ perte de cartes bancaires ou sinistres, doivent désormais devenir des processus automatisés, pour améliorer la productivité et optimiser les coûts. Une assistance IA peut également simplifier des processus internes comme le recrutement : en analysant automatiquement et en croisant les offres d'emploi publiées et les candidatures reçues, il est désormais possible d'accélérer la détection de candidatures pertinentes et donc d'accélérer les recrutements.

Parmi tous ces processus, l'intérêt que présente l'intelligence artificielle est la réduction drastique du traitement des dossiers, qui peut approcher les 50 %. En effet, la documentation reçue par les établissements bancaires, et notamment par ceux qui disposent d'une offre d'assurance, exige une analyse chronophage et coûteuse susceptible de nuire à la satisfaction de leurs clients. L'IA permet aujourd'hui d'extraire les informations importantes contenues dans un dossier de déclaration de sinistres, d'analyser le type de dommage en question, de comprendre s'il est couvert ou non par la police d'assurance. Les données pertinentes peuvent alors être importées dans une base de données de sinistres, et aider ainsi les agents à déclencher le processus d'indemnisation approprié. Néanmoins, l'IA n'a pas vocation à traiter ces dossiers sensibles : elle doit assister la prise de décision en diagnostiquant le pourcentage de risque qui pourra ensuite amener à une décision humaine.

En investissant le secteur de la banque, l'intelligence artificielle permet de déporter le travail fastidieux de compréhension du contexte vers la technologie pour que l'agent puisse se concentrer uniquement sur des tâches à valeur ajoutée. Réduire les erreurs humaines, accélérer la rapidité des réponses et la satisfaction client : l'IA est là pour assister l'agent dans sa prise de décision en identifiant d'éventuelles imprécisions, les risques de fraude, et en optimisant le ratio entre bénéfice, profitabilité et satisfaction client.

## COGITO®

Vos contenus non structurés sont précieux mais restent trop souvent inexploités. Expert System développe des solutions d'analyse cognitive fondées sur sa technologie sémantique Cogito qui permettent d'automatiser la compréhension de vos données textuelles et de vos processus métier.

### INTELLIGENCE ECONOMIQUE ET VEILLE

- Collecte, classification, analyse et exploration de contenus non structurés internes et/ou externes (sources ouvertes, bases de données payantes...) relatifs aux écosystèmes, technologies, produits ou marchés pour la production de reportings ciblés
- Analyse instantanée des réseaux sociaux
- Analyse d'opinion et e-réputation

### GESTION DES RISQUES OPERATIONNELS

- Analyse des risques et incidents industriels
- Analyse de rapports techniques
- Analyse de conformité (ex. RGPD 2018)

### AUTOMATISATION DES PROCESSUS METIER

Extraction d'informations issues de contenus non structurés ou de formulaires pour alimenter et accélérer les processus métier :

- demandes / déclarations (ex. gestion de sinistres)
- notices / contrats
- candidatures / CV

### OPTIMISATION DE LA RELATION CLIENT

- Analyse de verbatim
- Analyse d'emails
- Traitement automatisé des contrats (comparaison, classement...)

### EXPERT SYSTEM FRANCE

Tour Mattei  
207 rue de Bercy  
75 012 Paris  
T. 01 80 98 11 00

www.expertsystem.com/fr



NOTRE SINGULARITÉ ?  
NOUS SOMMES PLURIELS !

# FINAXYS

CONSULTANTS TRÈS AGILES  
IT FINANCE - BANQUE - ASSURANCE

**FINAXYS EN BREF :** Finaxys est un cabinet de consulting & intégrateur IT pour les acteurs de la Finance. Nos experts, passionnés par les métiers de la Finance, accompagnent nos clients dans la digitalisation de leur SI en apportant des conseils stratégiques et en implémentant des technologies innovantes.

**Notre intervention s'articule autour de 4 piliers :**

- Conseil SI et Fonctionnel • Étude et développement • Opérations et production • Gestion de projet à engagement de résultat

**QUELQUES CHIFFRES :** • 300 consultants • 10 ans d'existence • 20% de croissance annuelle • 35M euros de CA en 2016 • 3 implantations : Paris / Bruxelles / Londres

**NOTRE VISION :** Nous mettons au profit notre connaissance du secteur pour vous accompagner dans la transformation de votre métier : les enjeux réglementaires, les nouveaux acteurs, les technologies disruptives. Cette transformation s'opère par la mise en place d'une organisation Data Centric, la réappropriation de l'agilité entre les métiers et IT, la réduction des coûts et l'automatisation. Les technologies et méthodes que nous implémentons chez nos clients sont le reflet de nos convictions sur l'évolution du secteur Finance : le Big Data, le DevOps, le Cloud et le Machine Learning.

**LE BIG DATA CHEZ FINAXYS :** Nous sommes convaincus que les solutions Big Data apportent une réponse pertinente aux problématiques métiers dans le domaine bancaire : réglementaire, gestion des risques, détection de fraude, PnL, gestion de référentiel. Mais aussi IT & Digital : monitoring applicatif, analyse du comportement des investisseurs, vision client 360. C'est pourquoi nous sommes précurseurs, depuis 2013, dans l'utilisation des technologies de l'écosystème Hadoop pour la finance de marché, en les testant au sein du Lab Finaxys, puis en implémentant chez nos clients les architectures Big Data les mieux adaptées à leurs besoins. **Notre apport de valeur :** portée par une tribu de 30 experts, notre offre Big Data se différencie par sa complétude : nous conseillons les responsables IT dans la définition de leur stratégie en fonction de leurs cas d'usage métiers, nous formons leurs équipes à l'utilisation des briques de l'écosystème Hadoop (HBase, Spark, Flink, Kafka), nous coréalisons le socle cible sur la plateforme d'Hortonworks, Cloudera ou Confluent, enfin nous intervenons ponctuellement en phase de post production pour apporter une expertise sur une composante logicielle. Nos experts Big Data sont de profils variés : chefs de projets, architectes, business analysts, développeurs, administrateurs socle.

**NOTRE PROJET AVEC L'AUTORITÉ DES MARCHÉS FINANCIERS**

**Communiqué de presse AMF (publié le 25 janvier 2017) :** Décidée à rester à la pointe en matière de détection des abus de marché et autres manquements, l'Autorité des marchés financiers (AMF) initie un programme de modernisation de sa plateforme de surveillance pour une meilleure capacité de détection, d'analyse et d'investigation. Au lendemain de la crise financière de 2008, les législateurs européens ont réformé le cadre des marchés financiers. La directive révisée MiFID 2 et le règlement MiFIR, après le règlement EMIR sur les dérivés négociés de gré à gré, renforceront début 2018 la capacité des régulateurs nationaux à exercer leurs missions grâce à un périmètre de données à leur disposition plus large. Parallèlement, l'essor du trading algorithmique et à haute fréquence dans des marchés fragmentés a contribué à augmenter la volumétrie des ordres et transactions. La masse des données qui sera collectée demain par l'AMF est évaluée à 50 millions de messages par jour. Pour s'assurer du bon fonctionnement des marchés, l'AMF dispose d'ores et déjà de moyens humains et techniques performants, dans lesquels elle a investi au

fil des années. La plateforme de surveillance analyse quelque 1,6 milliard de comptes rendus de transactions et génère, en moyenne, 36 000 alertes par an. Après recoupement avec l'information disponible en lien avec les instruments financiers échangés, plus de 500 événements de marché sont analysés en profondeur chaque année par les équipes de l'AMF, donnant lieu à une trentaine d'ouvertures d'enquêtes.

**UNE PLATEFORME REPOSANT SUR DES TECHNOLOGIES BIG DATA**

ICY (prononcez « I see why... »), le nouvel outil de surveillance que l'AMF commence à développer en interne avec l'appui de la société Neurones, reposera sur des technologies de type « Big Data », qui permettent d'exploiter rapidement des données représentant des volumes importants et de natures diverses. A compter du second semestre 2017, ICY sera déployé progressivement afin d'être opérationnel pour recevoir les données MiFID 2 à compter du 3 janvier 2018. A terme, l'AMF bénéficiera ainsi d'une capacité démultipliée de stockage et d'analyse croisée des données tout en profitant de fonctionnalités innovantes. L'AMF souhaite ainsi renforcer sa faculté d'examen en temps réel pour une plus grande réactivité, mais aussi améliorer la détection des abus de marché grâce à des outils d'intelligence artificielle comme l'apprentissage automatique (Machine Learning). Certains modèles de détection, pour les manipulations de cours par exemple, pourront être ajustés automatiquement avec ce type de technologie. « Dans un contexte de marchés financiers de plus en plus automatisés et fragmentés, la détection des manquements d'initié ou des manipulations de marché nécessite des outils sophistiqués et innovants qui sont en mesure de décrypter une information complexe et non structurée, précise Alexandra Givry, directrice de la Division surveillance des marchés. ICY permettra à l'AMF d'exploiter les masses de données disponibles et ainsi de mieux exercer ses activités de surveillance des marchés et d'investigation, ainsi que de renforcer ses capacités d'analyse des tendances à l'œuvre sur les marchés et de l'impact de la réglementation ».

**Finaxys, filiale du groupe Neurones, participe à la construction et l'implémentation de la nouvelle plateforme Big Data de l'AMF**

**INTERVIEW D'INTERVENANTS FINAXYS TRAVAILLANT SUR LE PROJET**

**Lyes : Directeur Programme Finaxys :** « Je suis en charge de la direction transverse des équipes du groupe Neurones impliquées sur le programme ICY. Au quotidien, je m'assure de la tenue des objectifs de chaque équipe du groupe Neurones, de l'efficacité des échanges entre les intervenants avec l'AMF et du reporting des activités auprès du sponsor. Mon rôle consiste également à valider les réponses aux appels d'offre sur les différents marchés de l'AMF et à obtenir le staffing adéquat aux différentes missions. Je manage une équipe d'experts Finaxys : Architecte Technique, AMOA, MOE, Technical Leader, Chef de projet fonctionnel. »  
**Vivien : Chef de projet fonctionnel :** « Je suis chef de projet fonctionnel Finaxys sur le programme ICY et suis en charge du pilotage des équipes AMOA impliquées sur les différents projets en lien avec l'outil de surveillance. Au quotidien, l'équipe fonctionnelle réceptionne les spécifications de l'AMF et les qualifie pour l'équipe de développement. Nos analystes accompagnent également l'AMF à la rédaction de cas de tests. Nous travaillons notamment sur le projet réglementaire MiFID 2. »  
**Rimoun : Chef de projet technique :** « En tant que chef de projet technique Finaxys sur le programme ICY, mon rôle est de manager la réalisation des développements de l'équipe Finaxys demandés par l'AMF. Je suis également le garant du respect des exigences techniques sous la supervision de l'AMF. Enfin, je participe à l'industrialisation des développements, à l'accompagnement de l'équipe de développement et aide les consultants Finaxys à monter en compétences sur les différentes technologies Big Data de l'écosystème HADOOP Hortonworks. »



**RÉGLEMENTATIONS**



**GESTION DES RISQUES**



**DONNÉES DE RÉFÉRENTIEL**



**DÉTECTION DE FRAUDE**



**PNL**

**350 COLLABORATEURS**

Notre histoire commence en 2008 avec 5 collaborateurs, nous sommes fiers d'être à présent 350 à porter les couleurs de Finaxys.

**TOP 5 FOURNISSEUR**

Notre savoir faire unique allié à notre connaissance des métiers de la finance sont des atouts reconnus de nos clients banques d'investissement.

**25 CLIENTS**

Nos clients sont tous des acteurs du secteur financier, sur lequel nous avons choisi de nous spécialiser, depuis notre création.

**NOS PARTENAIRES BIG DATA**

cloudera

Hortonworks

confluent

dataArtisans

Couchbase



Turn data into business value

by Consort<sup>NT</sup>



# Nos clients utilisent la data science pour améliorer leurs performances, ET VOUS ?

## HISTOIRE

Le Groupe Estia est un cabinet de conseil spécialisé dans la transformation des données en actifs, travaillant sur l'intégralité du cycle de transformation et de valorisation de la data.

Le cabinet fondé en 2010 à Paris compte 100 consultants experts de la data (data scientists, statisticiens, consultants métier...).

En 2017 la société accélère son développement en s'appuyant sur un nouvel actionnaire de référence, le groupe Consort<sup>NT</sup> intégrateur et opérateur de solutions et d'infrastructures (2 000 collaborateurs, 140m euros de CA, 6 pays).

Le Groupe Consort<sup>NT</sup> et Groupe Estia proposent dorénavant une offre unique, continuum de services couplant technologies, infrastructures et capacités algorithmiques au service de la performance de leurs clients.

## NOTRE CŒUR D'ACTIVITE

Le Groupe Estia couvre l'ensemble de la chaîne de valeur data : de la transformation des données en actif au travers de leur exploitation opérationnelle par les métiers (pilotage, communication omnicanal, prévision...) en passant par la conception d'outils, d'applications analytiques, d'algorithmes & modèles mais aussi la conception du stockage des données...

Le couplage d'expertises métiers et de connaissances sectorielles permet au Groupe Estia de proposer une gamme de prestations de conseil complète dans le domaine de la data science (conception, AMOA, réalisation).

## SERVICES PROPOSÉS

Nous assistons les directions métiers dans leurs projets de transformation, d'organisation et d'optimisation ainsi que dans leurs opérations de fonctionnement courant.

Notre gamme de services s'axe autour d'offres métier : Pilotage des activités, Connaissance clients, Communication omnicanal, Modélisation à visée prédictive, Transformation vers une organisation data centric, Gestion de projets data... et de pratiques sectorielles.

## MULTI-SECTORIALITE

Le Groupe Estia dispose d'une expérience multisectorielle large. Nos équipes accompagnent depuis plusieurs années plus d'une trentaine de clients nationaux et internationaux issus de secteurs variés : Banque, Assurance, Crédit, Retail, Média & WEB, Energie,

Cette approche trans-sectorielle, avec comme pivot une expertise exclusive de la data, nous permet d'apporter un point de vue privilégié sur les best practices Data à travers près de 300 missions : data science, big data, analyse quantitative, dataviz...

## COUVERTURE GEOGRAPHIQUE

Le Groupe Estia dispose d'une implantation en France (Paris) et en Amérique du Nord (Montréal). Il bénéficie du réseau du Groupe Consort<sup>NT</sup> en France : Lyon, Sophia Antipolis, Toulouse, Nantes, Rennes, Lille comme à l'international : Munich, Stuttgart, Francfort, Cologne, Bruxelles, Luxembourg, Rabat, ...

## RÉALISATIONS

De l'analyse massive de données sur une plate-forme Big data dédiée, à l'analyse textuelle de données captées en web crawling, en passant par la création et l'exploitation au quotidien d'outils analytiques (Dataviz, Algorithmes, Machine Learning, IA, Segmentations) nous accompagnons au quotidien nos clients sur des thématiques multiples d'exploitation et de valorisation de leur actif data.

## DIFFÉRENTIATEURS

Le Groupe Estia est un pur player de la data et veille à conserver un positionnement exclusif sur ce savoir-faire. Nous portons la conviction que la donnée est un levier essentiel de création de valeur pour toute organisation. Nous nous attachons à faire émerger cette création de valeur de manière pragmatique en s'exonérant de tous effets de mode. Nos équipes sont focalisées sur les intérêts business de nos clients et privilégient dans ce sens les solutions présentant un ROI mesurable et rapide.

Nos équipes sont par ailleurs indépendantes des éditeurs de solutions logicielles, ce qui leur permet de proposer les solutions les plus efficaces et de faire preuve d'un haut niveau d'impartialité dans leurs recommandations.

Le Groupe Estia consacre depuis sa création des moyens conséquents à l'innovation (label Jeune Entreprise innovante obtenu en 2013) et propose à ses clients de s'associer à des programmes de recherche pour identifier ensemble de nouveaux leviers de croissance !



PILOTAGE DE VOS ACTIVITÉS



GESTION DE VOS PROJETS



CONNAISSANCE CLIENTS



CO-INNOVATION



EXPÉRIENCE CLIENT PERSONNALISÉE



TRANSFORMATION DE VOS ORGANISATIONS



ANALYSE PRÉDICTIVE

Une expertise unique pour faire de la data un levier de performance.



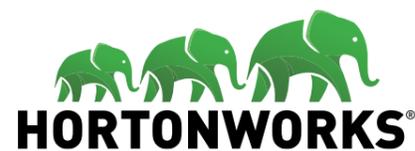
Turn data into business value

by Consort<sup>NT</sup>

www.groupe-estia.fr Groupe Estia, by Consort<sup>NT</sup>

Paris - Lyon - Sophia Antipolis - Toulouse - Nantes - Rennes - Lille - Montréal

Munich - Stuttgart - Cologne - Nuremberg - Wiesbaden - Bruxelles - Luxembourg - Rabat



### HORTONWORKS BIG DATA FOR CYBERSECURITY

The connected world creates a rate and volume of streaming cybersecurity data that is unprecedented, and attacks are increasingly sophisticated and multifaceted.

Existing security tools typically provide one facet of cybersecurity info, but not a single one was created for the big data generated by the digital world and it is unreasonably time-consuming to piece together data from multiple systems to assess the true nature of a single threat across an enterprise.

Hortonworks enables information security teams to embark a cybersecurity journey to leverage big data to improved economics and improved threat detection. • Hortonworks Connected Data Platforms help augment existing infrastructure with log analytics optimization • Hortonworks Connected Data Platforms, combined with open source Apache Metron expand the breadth of data sources accessible for security analytics with robust data ingest with data enrichment • Hortonworks cybersecurity solution, leverages Apache Metron to create a single holistic view of enterprise risk that encompasses an expansive and contextual view of enterprise data to enable machine learning, real-time streaming analytics for accelerated threat detection and improved SOC efficiency.

### BENEFITS OF HORTONWORKS BIG DATA APPROACH TO CYBERSECURITY - LOG ANALYTICS OPTIMIZATION

• Cost effective long term data capture • Increased visibility with expanded data types • Optimization of existing log analytics solutions • Efficient collection of machine data • Enables real-time data collection and analytics • Experts in leveraging big data to optimize collection and analysis of machine data • Experience services teams help optimize existing log analytics solutions

### DATA INGEST WITH ENRICHMENT

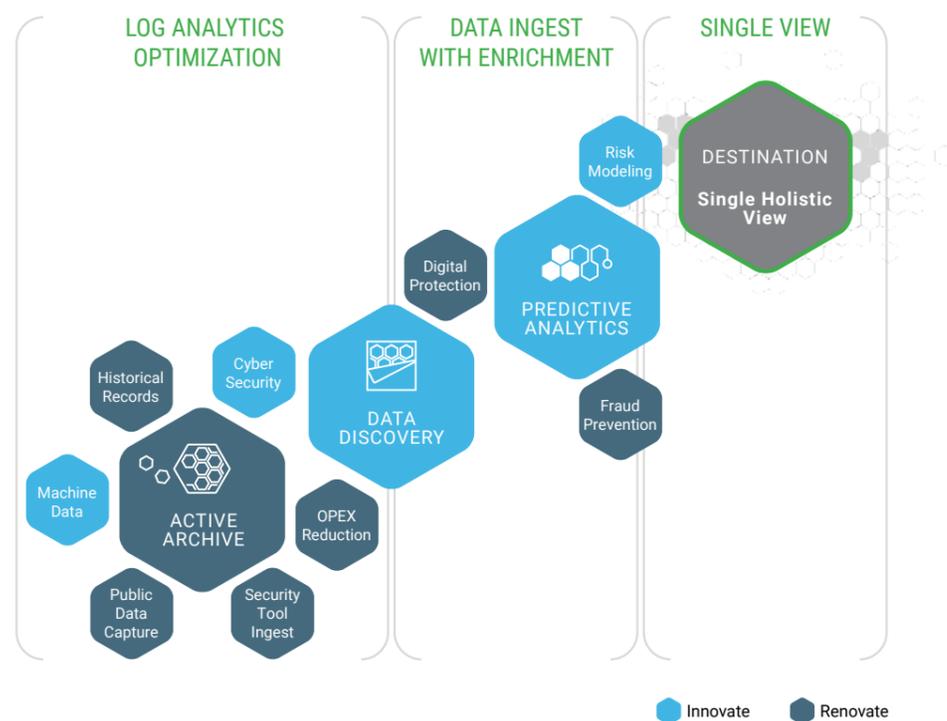
• Contextual view of data from a wide variety of real-time and batch data sources • Big data ingest and enrichment technology to capture full content and context of network traffic and machine data • Standard, consistent data formats • Enriched data ready for analysis • Reduces manual efforts and custom data preparation before analysis can be done • Experience that brings together data science and security for modern cybersecurity

### SINGLE VIEW OF ENTERPRISE RISK

• Accelerated threat detection measured in minutes, instead of months • More comprehensive datasets increase visibility of enterprise risk • Improve efficiency of SOC personnel • Faster threat analysis with stream analytics and machine learning • Reduce manual efforts necessary to triage the deluge of alerts generated today • Integrated view of data with contextual information necessary to detect dynamic threat vectors • Leading edge data scientists and professional service teams with deep expertise in streaming analytics, machine learning and data science as applied to cybersecurity

### COMPONENTS OF HORTONWORKS BIG DATA SOLUTION FOR CYBERSECURITY

• Hortonworks Data Platform (HDP®) • Hortonworks DataFlow (HDF™) with Apache NiFi • Apache Metron: Open source big data real time security analytics platform • Advanced data science tools for building and productionizing machine learning • Expert services to build models on custom use cases



# YOUR FASTEST WAY TO APACHE HADOOP® AND ITS ECOSYSTEMS



IT'S FREE. IT'S FAST. IT'S PORTABLE.



Use Sandbox in the cloud or in a virtual machine.



The Hortonworks Sandbox includes step-by-step tutorials to learn and explore.

EXPERIENCE APACHE HADOOP, APACHE NIFI, APACHE SPARK, SQL AND MORE

[hortonworks.com/sandbox](http://hortonworks.com/sandbox)



**HPE & BIG DATA : DES TÉMOIGNAGES UTILISATEURS CONVAINCANTS**

Les responsables de projets Big Data sont assez unanimes sur le fait qu'il faut trouver le bon tempo. Il ne sert à rien de partir trop vite. Comme l'a résumé Florence Laget, directeur des activités Big Data chez HPE France, le 'quick & dirty' n'est pas la meilleure option pour bien démarrer.

Ainsi chez BUT International, on reconnaît avoir au départ tâtonné. « Il faut éviter les prototypes techniques élaborés sans cas d'usage. Il faut une mobilisation des dirigeants (comme sponsors) et des managers, sans nécessairement leur parler Big Data », a résumé un responsable de la chaîne de magasins. « Il n'est pas nécessaire de créer un data-lake à tout prix. Il faut commencer par adapter son environnement et sa philosophie en matière de bases de données. Il faut identifier tel ou tel besoin, à partir de telles données et non pas créer une plateforme fourre-tout. Il faut acquérir une bonne maîtrise des technologies sous-jacentes, savoir différencier l'exploitation des données fraîches ou froides, par rapport aux besoins métiers. »



Une aide à la décision, des performances à la demi-seconde, avec un maximum de sécurité.

Le choix d'une architecture système robuste contribue largement au succès. BUT International, qui avait une exigence de performances à la seconde, pour une centaine de millions de calculs à la minute, a sélectionné SAP HANA sur HPE et a déjà déployé la solution auprès de 4.000 utilisateurs.

Au PMU, la collecte des données sur le fonctionnement des 22.000 bornes installées dans les bars-tabacs va conduire à une optimisation de l'ensemble du parc. Le dispositif permet à la fois d'anticiper d'éventuelles pannes sérieuses, et de mieux répartir les bornes en fonction de l'activité réelle



**SÉCURITÉ ET NOUVELLE RÉGLEMENTATION**

La question de la sécurité des données reste une pierre angulaire. Les intervenants mentionnent les obligations de la nouvelle réglementation européenne GDPR. « Cela conduit à s'interroger sur la finalité des croisements de données », souligne l'un d'eux. « Il faut bien segmenter et sécuriser en amont, et ne pas croiser n'importe quoi », ajoute un autre.

Et Florence Laget, directeur de l'activité Big Data de HPE France, précise: « Le principal enjeu reste celui de l'organisation et de la gouvernance de la donnée. Il faut obtenir l'adhésion des métiers et de la direction ».



Le projet SeqOIA de plateforme pilote de séquençage, porté par l'Assistance Publique - Hôpitaux de Paris, l'Institut Curie et Gustave Roussy, a été retenu dans le cadre du plan France Médecine Génomique 2025. La constitution de cette plateforme nationale de séquençage génomique très haut débit permettra l'intégration du séquençage au parcours de soin du patient. L'APHP a fait le choix de la solution HPE pour supporter ce projet stratégique, pour sa performance, sa capacité de croissance très aisée et illimitée en volume, et son ouverture vers le mode hybride.

Hewlett Packard Enterprise est une entreprise leader dans l'industrie des technologies qui accompagne ses clients dans leur transformation numérique. Avec les solutions Data Analytics et IoT, nous vous aidons à implémenter des plateformes de données qui vous permettront de collecter et transformer les données métier en actions concrètes pour améliorer vos process et la qualité de vos opérations, offrir une expérience utilisateur unique ou générer de nouveaux services. Avec l'offre la plus complète du marché, allant du Edge computing au Cloud Hybride, en passant par les solutions de mobilité et de travail collaboratif, nos technologies et nos services aident nos clients à travers le monde à améliorer l'efficacité, la performance et la sécurité de leurs systèmes d'information.

APOLLO / MACHINE LEARNING



Retail



Energie



Santé



Manufacturing



Transport



Villes

N'hésitez pas à venir rencontrer nos équipes sur le stand HPE pour en savoir plus sur nos cas d'usages ou nous contacter si vous ne pouvez pas participer au salon Big data. florence.laget@hpe.com



Vous avez maintenant l'habitude qu'à chaque nouvelle édition du Guide du Big Data, Hurence vous présente ses travaux phares de l'année. Cette année nous avons travaillé de manière récurrente sur une problématique très courante pour la plupart de nos clients qui évoluent dans le e-commerce.

### COMMENT RÉCUPÉRER LES DONNÉES DÉTAILLÉES DE NAVIGATION WEB TOUT EN CONSERVANT LES INDICATEURS STANDARDS CHEZ GOOGLE ANALYTICS?

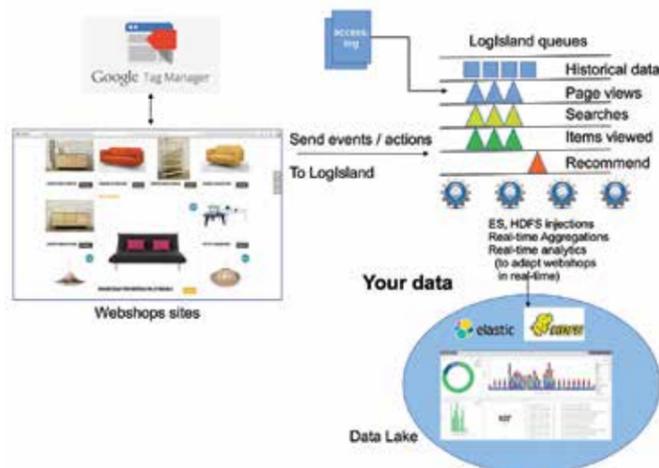
L'objectif est de récupérer ces données pour les mettre dans le Data Lake et les enrichir, ceci pour de multiples raisons. Une première raison est tout simplement liée à la réappropriation des données. Nos clients souhaitent avoir un accès total à leurs données (Google ne restituant qu'une vision partielle et/ou agrégée des données de navigation, et non pas l'ensemble des données brutes). Ils souhaitent également pouvoir mieux contrôler et protéger ces données business qui leur sont propres. La seconde raison consiste à mettre en place un traitement analytique en temps réel. Des recommandations / informations contextuelles sont proposées instantanément à l'utilisateur suivant son comportement. Ces contenus sont produits en temps réel à partir des données de l'utilisateur et de l'entreprise (en fonction des besoins, des marges, des stocks, des partenariats en cours, ...). Le troisième objectif vise à analyser les raisons de l'échec d'une vente, et à connaître les données qui ont manqué à l'utilisateur pour la concrétiser. D'où l'intérêt de rapatrier ces données de navigation pour les croiser et les enrichir avec d'autres sources de données externes/métiers (ERP, PIM, base utilisateur, infos concurrentielles ...).

Nous avons mis en place des chaînes de traitement basées sur notre produit open source **LogIsland** pour permettre la récupération temps réel des données de navigation tout en conservant une partie chez Google Analytics. Le site Web de l'entreprise n'est pas modifié car nous utilisons le même tag manager que pour Google Analytics (Google Tag Manager): les instructions d'envoi d'événements alimentent à la fois Google et le Data Lake de façon transparente. Ceci a permis à nos clients de rendre leurs sites Web beaucoup plus réactifs (très proche de ce que fait Amazon), pour exploiter au mieux les opportunités de vente. Il est important de détecter qu'un utilisateur a une intention d'achat et n'arrive pas à la concrétiser.

Nous avons implémenté ces projets sur la base de notre outil LogIsland, capable d'analyser des logs aussi bien pour le web analytics que pour la cyber-sécurité, l'IoT, et l'Usine du futur (aussi appelée Industrie 4.0). Nous pouvons décliner cette solution au-dessus de n'importe quelle infrastructure Big Data des clients classiques de Hadoop. Il existe d'autres possibilités d'implémentation mais la nôtre a l'avantage d'offrir une grande facilité d'utilisation (avec des composants clés en main). Elle est de plus extensible (plugins java), libre et en **open source**.

### LOGISLAND, LA SCALABILITÉ D'UNE SOLUTION BIG DATA CLÉ EN MAIN.

Nous utilisons Kafka pour la gestion des événements et des alertes et Spark Streaming pour l'analyse en temps réel des logs ou autres données. Grâce à ses plugins dédiés, **LogIsland** s'intègre parfaitement à votre cluster Big Data Hadoop et à votre moteur de recherche Elasticsearch, sans dépendance. **LogIsland** peut collecter vos logs à travers Nifi ou Logstash ou tout autre ETL, déverser ces logs dans votre Data Lake, les faire indexer à des fins de recherche, en plus de les analyser à la volée et les transformer en événements ou alertes. **LogIsland** peut même directement récupérer des événements par protocole IoT standard ou, comme nous l'avons vu plus haut, via un Tag Manager pour le web analytics. Il est scalable grâce à Spark; ses performances sont incomparables, son architecture offre un énorme potentiel.



### L'APPROCHE HURENCE EST TRÈS PRAGMATIQUE.

Elle part de la problématique de nos clients qui consiste souvent à optimiser les ventes et faire en sorte que l'expérience utilisateur du site web soit la plus personnalisée possible. Les utilisateurs sont uniques et il est nécessaire d'anticiper leurs besoins (que l'on a pu prédire) pour que l'expérience de navigation soit simplement extraordinaire.

La force de Hurence c'est de projeter la vision Big Data de l'entreprise sans se bloquer sur les aspects stratégiques avec une première implémentation novatrice. LogIsland pose ainsi les bases d'un futur beaucoup plus temps réel et beaucoup plus réactif, une première étape incontournable.

Plus d'informations sur [hurence.com](http://hurence.com)



**hurence**  
CONSULTING - FORMATIONS - PRODUITS

# le big data sur mesure

Depuis 2011, nous vous accompagnons quelle que soit votre taille et la complexité de vos systèmes d'information dans vos développements de solutions Big Data





## DEPUIS LA PRÉCÉDENTE ÉDITION DU GUIDE, QUEL CONSTAT FAITES-VOUS DE L'ÉVOLUTION DU BIG DATA ?

La terminologie même de « Big Data » commence à prendre quelques rides tant elle a été galvaudée, interprétée selon les enjeux et le référentiel de chacun. Et pourtant, nous observons une énergie toujours plus forte autour de la data, des initiatives et des projets permettant d'en tirer la valeur tant attendue. L'émergence des architectures Server Less et l'acceptation du cloud permettent notamment à de plus en plus de clients de gérer de grandes quantités de données à des coûts raisonnables. Il n'existe peu ou plus de secteurs d'activité n'ayant pas engagé des ressources sur ces sujets et l'envie de ne pas passer à côté des opportunités promises surpasse les premières déceptions ou difficultés rencontrées. Nous évoquons l'an dernier « l'âge de raison » du Big Data pour célébrer les 7 ans d'apparition du terme. C'est aujourd'hui encore plus vrai, les décideurs restent engagés mais plus exigeants et en même temps raisonnables sur le retour de leurs investissements.

## QUELS SONT, SELON VOUS, LES NOUVEAUX CHALLENGES TECHNIQUES, TECHNOLOGIQUES ET/OU STRATÉGIQUES DU BIG DATA ?

C'est une trivialité de le rappeler, mais les sujets 'Big Data' ne peuvent rester des initiatives de R&D cantonnées dans des laboratoires affranchis des enjeux et contraintes opérationnelles. Au-delà de mobiliser efficacement les ressources rares en data science sur les sujets à plus forte valeur ajoutée, la capacité d'industrialiser les projets est certainement le principal challenge aujourd'hui entre les mains des DSI. La réalité des systèmes d'informations et de leur dette technique est trop souvent négligée en amont des travaux. Combien d'applications métiers sont APIées ou capables de s'affranchir des dépendances python/scala pour opérer des moteurs prédictifs ? Par ailleurs, la question de la volumétrie devient progressivement un enjeu pour les non pure players du web grâce notamment à l'Internet des objets et la capacité nouvelle de collecter à l'infini les mesures relevées par un ensemble de capteurs. Enfin, le nouvel eldorado de la donnée non structurée reste à conquérir. La plupart des expérimentations restent décevantes, principalement en raison des difficultés d'ordre ontologique, linguistique ou de mise en œuvre opérationnelles des techniques d'apprentissage profond.

## LES END-USERS SONT-ILS PLUS MATURES EN LA MATIÈRE ? QUELS SECTEURS SONT VRAIMENT MOTEURS POUR VOUS ?

Les décideurs sont définitivement de plus en plus matures. Les expérimentations ont permis de faire la part des choses entre promesses et réalité. Les efforts d'évangélisation des vendeurs du secteur ont également aidé les décideurs à y voir plus clair dans la jungle technologique actuelle. Il n'est pas évident de discerner un secteur plus moteur qu'un autre. Nous observons que les secteurs de l'industrie ou des services publics sont désormais prêts à rattrapper leur retard même si les Telecoms ou la Finance gardent une longueur d'avance.

## L'INTELLIGENCE ARTIFICIELLE : EN VOIE DE DÉMOCRATISATION CHEZ VOS CLIENTS ?

L'IA se démocratise jour après jour et devient une réalité opérationnelle chez nos clients. De la mise en œuvre de Chatbots pour

un premier niveau de support informatique, à l'optimisation de l'allocation de ressources de contrôle, en passant par la traduction automatique de contenus, les champs d'application quotidienne sont infinis et d'ailleurs pas toujours explicitement mis en avant comme issus de l'IA. Le plus difficile reste néanmoins de faire garder les pieds sur terre aux non-initiés, car malgré les progrès considérables en la matière, et quitte à bousculer le discours des vendeurs de solutions, il n'y a rien de magique ou miraculeux, même avec l'apprentissage profond !

## QUELLES SONT LES ÉVOLUTIONS DE VOS OFFRES DATA ?

Nous poursuivons la mise en œuvre de notre stratégie data via notamment notre plateforme HyperCube. Celle-ci s'enrichit continuellement pour offrir tous les outils utiles aux data scientists et autres analystes pour conduire leurs travaux quotidiens d'analyse ou de modélisation de la donnée. Nous avons notamment récemment intégré une offre plateforme massive et élastique capable d'optimiser dynamiquement les ressources serveurs. Nous déclinons également cette plateforme en « verticaux » à même d'adresser les enjeux spécifiques de certains secteurs/fonctions (outils de demand planing, de pricing, industrie 4.0...). Nous développons enfin notre savoir-faire et nos expériences dans l'accompagnement de nos clients dans leurs réflexions autour des enjeux d'organisation, de gouvernance ou d'architecture data.

## ENFIN, QUELLE EST LA PLUS-VALUE APPORTÉE AUX CLIENTS VOUS CHOISSANT ?

Nos clients trouvent en nous un interlocuteur rare capable de mobiliser un large spectre d'expertise depuis le développement, le design, l'architecture ou encore la data science que nous mêlons étroitement avec le savoir-faire reconnu des consultants en management BearingPoint ! Notre engagement est de mettre en œuvre opérationnellement la data science. Grâce à nos solutions et nos savoir-faire, la mise en production des modèles HyperCube se fait en quelques secondes !



Marc Hispa  
HYPERCUBE

# Data.

For you.

#welovedata

HyperCube



**IDEATRANS, AGENCE DE TRANSFORMATION DATA DIGITALE & INNOVATION ! PARIS - TOULOUSE**

**Promesse : Pertinence & Performance !** Efficacité renforcée & Innovation naturelle, Différenciation personnalisée & Adéquation anticipée, iDeaTrans vous aide à développer et transformer vos métiers et votre business par la Data, le Digital et l'Innovation, en adressant les 5 piliers de Valeur suivants :

- **Révéler, clarifier et transmettre votre maturité Data-Digital :** stratégie et enjeux Data dans l'organisation ? Conscience, connaissance et compréhension, capacité à fédérer et transformer ...
- **Gouverner, libérer et valoriser la Data-Digital en toute sécurité :** piloter, enrichir, ouvrir les Data en protégeant sa valeur, les organisations et les personnes ? (incl RGPD, DPO/CIL, l'anonymisation) ...
- **Sourcer vos besoins Data-Digital :** quelles Data utiles ? interne, externe ? Open - Shared Data...
- **Se confronter aux métiers 360° et faire émerger vos opportunités Data-Digital :** réunir un panel 360°, créer les conditions d'émergence ? une organisation appropriée ? l'adéquation marché et besoins ↔ Data ? ...
- **Transformer la Data-Digital en éléments de différenciation métiers et de création de valeur :** efficacité ? KPIs ? analytics ? nouveaux paradigmes ? ...

**Marchés :** iDeaTrans s'engage avec et envers les Organisations qui cherchent à répondre à une volonté et à des problématiques de Transformation, d'Innovation (améliorer ou réinventer), qu'elles soient techniques, métiers, ou organisationnelles, en s'appuyant sur les ressources nécessaires, du fort potentiel Data-Digital (360° interne - externe), sur une forte expertise, ainsi que sur une méthode reconnue (Label d'OR) d'Innovation ouverte.

**Clients :** Evolution xy (plateforme Data et nouveaux services pour une coopérative de bovins), SSI - Aibus (initialiser un projet Data pour améliorer la maintenance sur les moteurs d'avion et la rendre prédictive), Laboratoire Pharma (direction technique d'un projet de transformation Digitale Monde), HackInAdour (accompagner l'ouverture, la mise en commun et l'usage des Data territoriales en Adour dans le cadre d'une démarche HIA pour redynamiser les activités en Adour), Monoprix (créer le schéma directeur Data, Référentiels User Centric, Omni Canal), QLIK (accompagner le challenge Open Data France), BEL (analyser le marché de la Data Marketing Fromage en Algérie), IPSEN (révéler la Data Digital dans la e-Santé, Start-Ups ; exploiter les Logs Big Data), AGORA (créer de la Valeur Data au sein de communautés), C3 COP21 (organiser et utiliser l'Open Data EDF-ERDF-Météo France-IGN-CNES-IRSTEA-ETALAB-MEDDE pour instruire la problématique Climat et faire émerger 29 projets dans toutes la France), GfK IFR (créer et gouverner la Data & Digital SI pour développer le Business Clients et la Transformation du Groupe, diriger les Opérations monde IFR), UBISOFT - LECLERC (Conseil autour des référentiels MDM), POLE EMPLOI (révéler et clarifier la maturité Data afin d'initier un projet Big Data), MNMS (contribuer à faire émerger des solutions innovantes dans les organisations, créer les conditions du Shared-Data (EDF, AG2R La Mondiale, SYNERTHON ATM Toulouse,)), de l'idée au marché), etc...

**Savoir-faire :** +25 années d'expertise dans la valorisation de la Data (Marketing, Retail, Open Shared Big Smart Data), identification, collecte, transformation, enrichissement, création de valeur et d'expériences..., de la source au client, des contenus aux usages

métiers protégés et sécurisés. +20 ans de transformation des organisations en s'appuyant sur le Digital-Data, l'Innovation Métiers (Plateformes SIRH, ERP, ETL, MDM, CRM, DWH, SIEM, BI BigData, DataViz Analytics, DMP, LMS e-Learning, etc). Accompagnement à l'ouverture et partage de Data pour de grandes démarches de développement territorial et sociétales. Conseil & investissement dans des Start-Ups depuis 1997 (de l'idée au marché). Organisation d'événements Data - Digital - Innovation (ex : www.lanuitdubigdata.com )

**Contributions :** AFNOR, AFDEL TECHIN, IDC, Gartner, CIO, CIOnet, IT for Business, 01, IT Expert, Mag IT, CIGREF, IBM, Oracle, Informatica, QLIK, Business Objects SAP, AFAL, AFCDP, Finance Innovation, EBG, Alliance Big Data, Big Data Paris, Cap Digital, ESSEC, Novancia, AFM, CMIIT, Distributique, Printemps des Etudes, G9+, Total, Micropole, Neurones, B&D, Linkbynet, etc...

**Organisation innovante :** iDeaTrans réunit des acteurs pertinents et des experts métiers (indépendants) depuis +15 ans, organisation originale ouverte, et collaborative. iDeaTrans a choisi la méthode et démarche d'innovation MyNeedMySolution (MNMS s'appuyant sur l'intelligence collective depuis + 15 ans) pour révéler, faire émerger les idées et besoins, et les transformer en solutions concrètes et pérennes.

**Fondateur CEO :** Fabrice Benaut (Paris 2), un activiste du numérique éthique, révélateur et accélérateur de la transformation et de l'innovation, fondateur de la société iDeaTrans, Fabrice est aussi associé dans MNMS et dans plusieurs startups. Intra-entrepreneur innovant, Fabrice a été Directeur des Systèmes d'Information et Opérations, en charge des Projets Métiers, membre du COMEX de GfK IFR Global (60 pays), et CIL. Ses 25 années d'expérience internationale dans le Business de la Data (360° métiers, Marketing, Marché, Distribution, Usages, Comportements, Traces, etc) font de lui un expert en matière de Big Data et de Transformation (numérique) de l'entreprise (startup et grands groupes). Fabrice est également Commissaire aux Données pour de grands projets territoriaux, HIA www.hackinadour.com (2017), www.c3challenge.com COP21 C3 (2015).

Proche du monde des start-ups depuis 1997 (co-working) où il intervient en qualité de conseil et d'investisseur (Privowny, Stample, PCB xField Paintball, Citymeo, Integra, Keakr, etc...), Fabrice anime des ateliers « de l'idée au marché » et de co-design (Focus Group) avec MNMS.

Conférencier reconnu, Fabrice Benaut est président de l'Alliance Big-Active Data, et membre -très - actif dans plusieurs associations regroupant le monde du Digital Data Marketing, DSI-CIO et Business, professionnels du monde numérique. Il aime partager et enrichir son savoir Data Digital en organisant des événements, en intervenant dans des écoles, universités et des groupements professionnels. Fabrice est aussi Expert Judiciaire Eco-Fi Etudes de Marché Data (Versailles).

*Transformer les organisations, les métiers et le business par la Data, le Digital et l'Innovation*

*Gouverner, Libérer, et Valoriser la data et le digital en toute sécurité*

*Confronter vos métiers 360° et faire émerger vos opportunités data-digital*

*Transformer la data-digital en éléments de différenciation métiers et de création de valeur*

*Révéler, clarifier et transmettre votre maturité data-digital*

*Sourcer vos besoins data-digital (open, shared, big data, interne, externe)*

**iDEATRANS**  
Innovation Data Digital

*transmettre des savoirs et des savoir-faire.*

**iDeaTrans**  
Agence de transformation data, digitale & innovation

Partenaires



MyNeedMySolution  
Agence d'innovation



Alliance Active Data  
Première communauté du Big Data

Paris • Toulouse • fabrice.benaut@ideatrans.net  
www.ideatrans.net



### MAPPY NE FAIT PLUS DE COMPROMIS ENTRE COÛT, PERFORMANCES ET VOLUMÉTRIE GRÂCE À INDEXIMA

Spécialiste du calcul d'itinéraire et des services de cartographie, Mappy est reconnu comme le leader français de la recherche locale par la carte sur Internet, mobiles et tablettes.

Comme dans toutes les organisations, mieux comprendre ses services est un enjeu majeur les différentes directions fonctionnelles de Mappy.

Par exemple comprendre finement le comportement de ses utilisateurs pour la direction marketing afin de proposer des offres adaptées ou optimiser le parcours client. Mettre en place des KPIs pour la direction technique pour optimiser les performances ou défaillances potentielles en production.

Depuis plus de deux ans et grâce à Indexima les équipes marketing de Mappy peuvent désormais suivre et analyser, l'ensemble des visites sur les différents produits avec une finesse jamais obtenue auparavant.

### UNE CHAÎNE BI TRADITIONNELLE TROP RIGIDE ET COÛTEUSE

Avant 2014, l'infrastructure BI offrait peu de flexibilité et d'agilité dans l'analyse : cela entraînait de nombreux aller-retour entre les métiers et l'équipe informatique pour obtenir les statistiques d'usage en amont des axes d'analyses à analyser. Il fallait attendre un temps important avant de pouvoir analyser les usages. D'autre part, l'information ainsi agrégée perdait en précision et cela limitait les analyses.

Cette architecture BI reposait sur des traitements développés en Python pour gérer les 200 Go de log quotidiens enregistrés par les 700 serveurs en production. Stockés dans Microsoft SQL Server, les quelques 10 millions de lignes de statistiques étaient présentées via des cubes et consommées avec Excel.

### INDEXIMA APPORTE SIMPLICITÉ ET ÉCONOMIE EN CAPEX ET OPEX POUR L'IT ET AGILITÉ POUR LES ANALYSTES

Le premier chantier fut de monter une infrastructure de stockage de masse : Hadoop fut privilégié. Puis le choix d'une solution datavisualisation pour entrer dans l'ère de la BI self-service. Guidé par un choix de conserver toute la richesse des logs. Les analyses n'ont alors plus de limites et les lignes de statistiques se démultiplient. Courant 2015, la donnée agrégée dépasse le milliard de lignes.

### UNE SOLUTION BLUFFANTE PAR SA TECHNOLOGIE

C'est alors que se pose deux sérieux problèmes.

Les requêtes générées par la BI self-service sont coûteuses, surtout lorsque la stratégie d'exécution se base sur des principes de « full scan ». Demander aux analystes de patienter plusieurs minutes ou plusieurs heures avant d'avoir la réponse à leurs interrogations devient intolérable. D'autre part, stocker de grands volumes de données est une chose mais dépenser des fortunes en mémoire et en serveur pour assurer la performance n'est pas une option.

Les solutions existantes pour améliorer cette performance sur les volumétries de Mappy ne se sont pas révélées satisfaisantes. Celles s'appuyant sur des concepts similaires aux cubes étaient trop peu flexibles pour les utilisateurs métiers, celles reposant sur des traitements massivement parallèles étaient trop chères pour la DSI.

L'équipe BigData de Mappy était au pied du mur. La solution BI répondait aux besoins d'agilité et de finesse de ses données mais avec des problèmes de performance sur Hadoop qui semblaient insurmontables.

### « IL Y A UN AVANT ET UN APRÈS INDEXIMA »

C'est en testant Indexima que Mappy a enfin pu lever ce conflit FLEXIBILITE / COÛT / PERFORMANCE. Les temps de réponse d'affichage sont retombés en dessous de la centaine de millisecondes. En combinant trois innovations : des hyper Index, un moteur Orienté Colonne distribué et un mécanisme de génération automatique des Hyper Index, Indexima adresse l'analyse de très gros volumes sans rogner sur les performances. Elle apporte à Mappy toute la souplesse et la finesse dont elle avait besoin, sur une infrastructure très raisonnable. En 2017, les collaborateurs Mappy utilisent leur outil d'exploration de données sur plus de 25 milliards de lignes, sur un cluster de 6 machines.

L'accès à la donnée est quasi instantané : les requêtes SQL sont exécutées en moins de 100 ms.

L'infrastructure est totalement scalable et peu coûteuse : Indexima est installé sur 6 serveurs du cluster Hadoop et supporte l'analyse du parcours des 12 millions de visiteurs uniques et 5 millions de points d'intérêt. La volumétrie des données générées a été multipliée par 25 en moins de deux ans, sans le moindre ajout de serveur de calcul. Le ROI est immédiat, surtout comparé à des solutions de type Spark qui nécessiteraient 1000 fois plus de machines.

Et pour une qualité de service à l'utilisateur incomparable : chaque analyste peut croiser les données qu'il souhaite et obtenir des réponses immédiates sur l'ensemble des données de masse. Enfin la préparation de donnée est également facilitée puisque qu'il n'y a plus besoin de couches multiples de datamart.

### INDEXIMA, UN ÉLÉMENT CLÉ POUR L'EXPLOSION DES DONNÉES ENCORE À VENIR

Indexima a permis à Mappy d'accélérer sa propre roadmap et le développement de son équipe Big Data. Les collaborateurs de Mappy n'hésitent pas à parler d'un changement de métier et d'une nouvelle façon de voir le big data et la BI.

Indexima est devenu un élément clef de l'architecture BI de Mappy, et permet d'appréhender sereinement l'arrivée des centaines de To de données à venir, sans surprise.

« Nous n'avons toujours pas atteint les limites d'Indexima, nous n'avons même pas complètement cerné l'ensemble du potentiel et de la puissance de la solution. Aujourd'hui il n'y a plus aucune limite aux questions que l'on peut se poser autour des données et de la BI (des questions et des challenges que l'on n'osait pas se poser par le passé). Indexima a révolutionné notre état d'esprit ! »



# Indexima



## The Smart Data Hub



### Data Space

Intelligence des données



### Hyper Index

Performances extrêmes



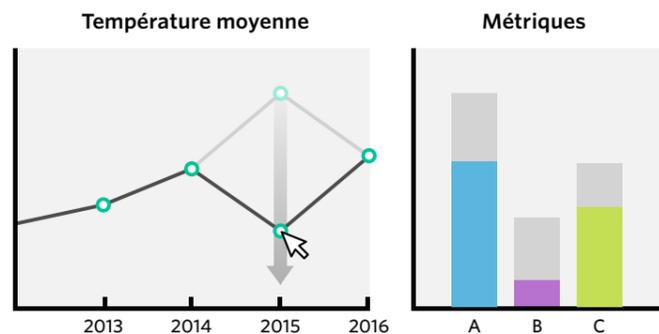
**ET SI LA VALORISATION DES DONNÉES ÉTAIT ACCESSIBLE À TOUS ?**

Si les data scientists sont indispensables pour mener les projets data les plus avancés, ce sont en réalité les métiers qui sont les mieux placés pour valoriser vos données au rythme de votre activité. En effet, l'expertise métier et la connaissance fine du contexte sont indispensables pour donner du sens à ces dernières. C'est pourquoi nous avons développé INVIOO, un outil radicalement plus intuitif pour rendre le travail sur les données accessible aux professionnels sans compétences informatiques ou mathématiques. INVIOO permet de préparer, nettoyer, explorer, enrichir, calculer, et prédire des données en remplaçant le code par la manipulation directe des graphiques. Vous utilisez probablement déjà des graphiques et tableaux de bord pour présenter de l'information, INVIOO capitalise sur cette familiarité pour vous permettre de travailler directement la donnée sur ces graphiques. Il en résulte une expérience radicalement plus intuitive car vous comprenez mieux l'impact de ce que vous faites.

Pour ce faire nous avons développé une technologie unique d'interaction graphique : l'utilisateur peut manipuler directement les éléments graphiques représentant les données, et les modifier en contrôlant le résultat de ses actions en temps réel.

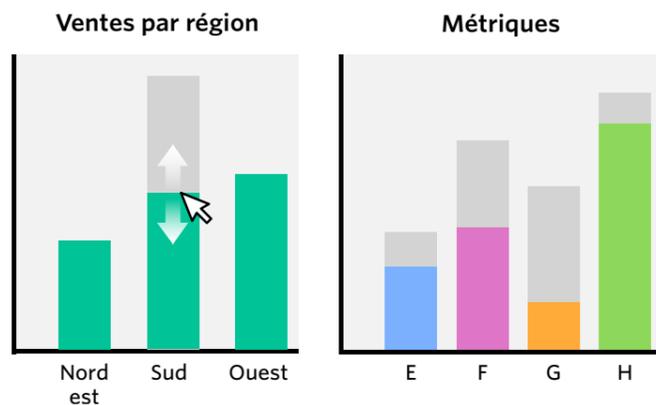
**EXEMPLE 1 :**

Vous avez modélisé le lien entre l'évolution de la température moyenne annuelle et des métriques sur votre activité : il vous suffit de cliquer et déplacer directement un point de température pour voir les métriques se recalculer sous vos yeux.



**EXEMPLE 2 :**

Autre exemple : vous pouvez directement faire varier la somme de vos ventes dans une région donnée (qui est l'agrégation de millions de transactions) pour voir l'impact éventuel sur une série de métriques. Il suffit de cliquer sur l'arête de la barre concernée et de la déplacer pour modifier la valeur agrégée. Ces métriques sont recalculées en temps réel à partir des transactions non agrégées, qui ont elles-mêmes été impactées par votre manipulation de la somme.



Pour plus d'informations visitez [www.invioo.com](http://www.invioo.com) ou contactez-nous sur [contact@invioo.com](mailto:contact@invioo.com)

# Radicalement plus intuitif.



# JEMS DATAFACTORY

**HAMID LAGHRISSE**  
Business Unit Manager

01 47 89 63 10  
hlaghrissi@jems-group.com  
8 rue de l'Hôtel de Ville  
92200 Neuilly-sur-Seine France  
www.jemsdatafactory.com



## PRÉDIRE LE RENDEMENT ET LA QUALITÉ DES BLÉS AVEC LE BIG DATA

Un retour d'expérience d'Olivier Descroizette, Directeur Services Agro-Environnement, SMAG (filiale INVIVO)

### QUELS ÉTAIENT LES BESOINS INITIAUX ET LES OBJECTIFS DU PROJET ?

SMAG, leader des systèmes d'information et du web agricoles et filiale du premier groupe coopératif français INVIVO, opère un pivot de son modèle économique en mettant en place une architecture Big Data. L'objectif est ambitieux : mettre en place une stratégie data-driven pour ses clients et développer une gamme de nouveaux services à très haute valeur ajoutée pour le monde agricole, basés sur la valorisation qu'offrent les technologies Big Data.

JEMS datafactory, leader des solutions Big Data, nous a naturellement accompagné dans le développement et le déploiement de DATA CROP, premier projet d'une telle envergure et répondant à plusieurs objectifs technologiques et fonctionnels :

- Développer le premier service français de prédiction de rendement et de qualité du blé
- Industrialiser un algorithme de haute complexité de prédiction de rendement et de qualité des cultures à partir de données internes et externes
- Proposer une solution avec une répercussion économique forte pour les acteurs de la filière

### POURQUOI LE CHOIX DE JEMS DATAFACTORY ?

Le choix s'est rapidement porté sur JEMS datafactory du fait de l'adéquation entre leur offre et nos besoins.

Nous recherchions un acteur capable de nous accompagner de bout-en-bout dans la mise en œuvre d'une plateforme Data Lake. Celle-ci constitue désormais le socle de la fouille de données et des traitements algorithmiques pour nos Data scientists. L'expertise de JEMS datafactory nous a permis de mettre en place une plateforme inédite dans le monde agricole.

Par ailleurs, notre ambition est de devenir le pionnier de la Data Agricole en France et à l'international. Il nous faut anticiper les usages et services de demain, basés sur des données de plus en plus nombreuses et complexes à analyser. Fort d'une solide expertise sur l'hébergement et l'exploitation de plateformes Cloud, nous avons identifié JEMS datafactory comme le partenaire de projet capable de nous hisser vers cet objectif.

En prenant en compte tous ces aspects nous avons sélectionné JEMS datafactory sur la base de leur proposition de valeur. Leur démarche globale structurait notre projet sur une réponse évolutive pour construire un Data Lake répondant immédiatement à nos besoins de services, et prenant en compte les objectifs business majeurs. JEMS datafactory nous a aidé à adresser les sujets primordiaux de supervision, sécurisation et gouvernance de la donnée

### QUELS DÉFIS STRATÉGIQUES ET TECHNOLOGIQUES ?

Le secteur agricole opère une mutation d'ampleur depuis plusieurs années : le taux d'équipements technologiques des agriculteurs est en forte hausse, les données générées sont d'une grande variabilité et la quantité de données produites explose.

Les principaux défis technologiques du projet sont d'industrialiser et de mettre en production dans un contexte innovant :

- Récupérer des données volumineuses de sources internes et externes à l'entreprise
- Consolider ces données dans une plateforme unique et les mettre à disposition des utilisateurs finaux et aux datas scientifiques
- Calculer un algorithme de prédiction de rendement complexe sur des données volumineuses
- Rendre accessible aux Data scientists la fouille et la manipulation de données, ainsi que l'élaboration et l'optimisation des algorithmes
- Mettre en place une chaîne d'intégration de données performante, de l'acquisition des données jusqu'à leur exposition

### UNE SOLUTION SUR-MESURE ?

Pour répondre à ces défis technologiques, nous avons mis en place une plateforme de Data Manufacturing organisée en plusieurs couches :

- Une couche « data\_lake » dans laquelle sont stockées toutes les données brutes en fonction du propriétaire de la source de données
- Une couche « data\_products » dans laquelle sont stockés les résultats des algorithmes organisés par produit
- La couche « data\_labs » : environnement type « bac à sable » pour permettre des analyses et des tests sur les données et les regroupements de données
- Une couche « data\_view » permettant le requêtage et la visualisation des données par les utilisateurs finaux

### QUELS SONT LES RÉSULTATS ET PERSPECTIVES À CE JOUR ?

Le socle du premier Data Lake agricole est mis en place pour l'entreprise. Le modèle DATA CROP est en production sur des clients pilotes et réalise une montée en charge et couvre déjà :

- Près de 30 000 exploitations agricoles
- Plus de 160 000 parcelles de blé
- Une surface d'environ 1 million d'hectares (20% de la production française)

Les données satellites sont intégrées et traitées quotidiennement avec des résolutions d'images de 10 mètres. La prévision des rendements est affinée chaque jour sur l'ensemble des parcelles et les résultats sont exportés pour chaque client.

Pour ce modèle pilote, nous traitons à ce jour plus de 320 Go de données / jour avec l'ambition de passer à 1 To / jour d'ici la fin de l'année 2017 par l'intégration de nouveaux clients, et de nouvelles sources de données.

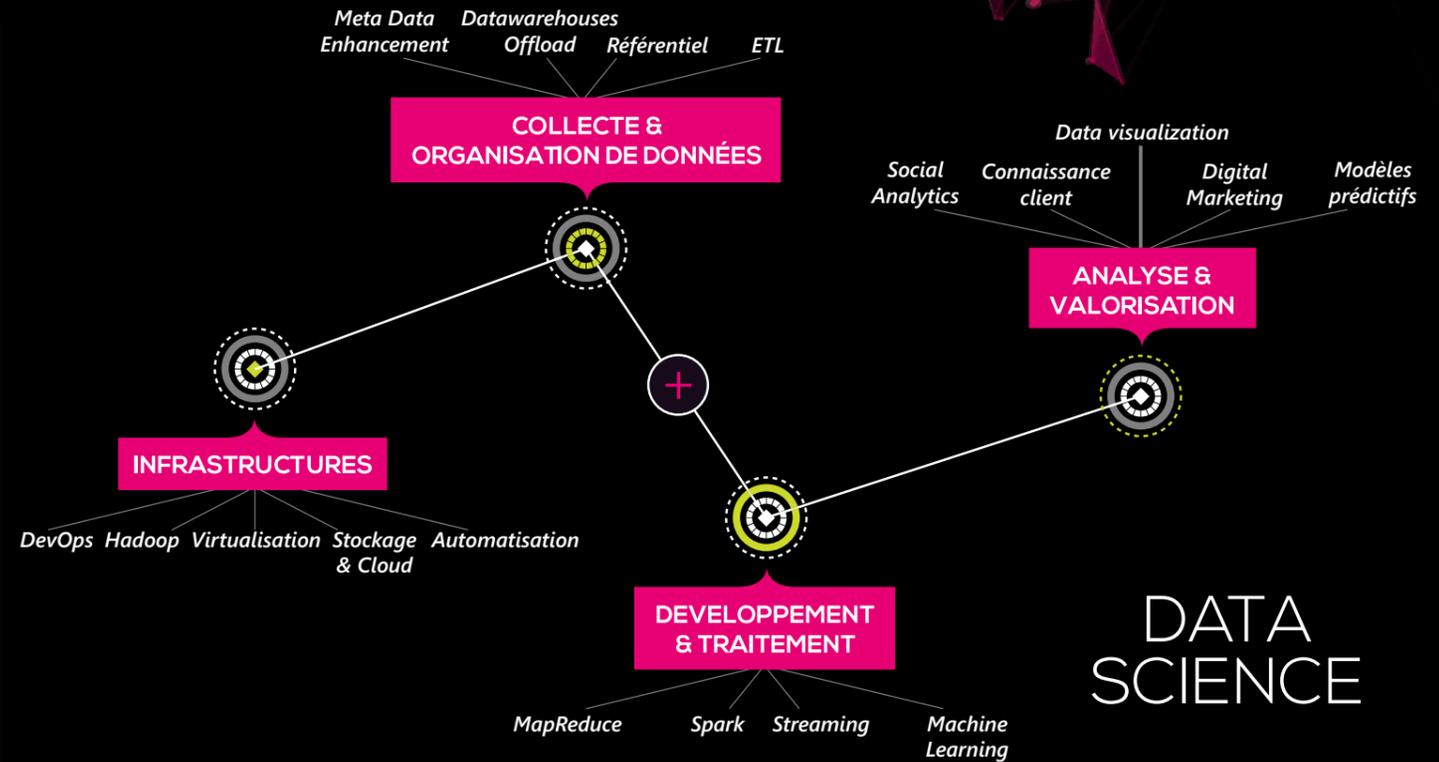
Les perspectives du modèle sont fortes avec de multiples retombées attendues pour la filière :

- Pour les organismes de collecte de céréales, DATA CROP permet d'optimiser les coûts de stockage et de gestion logistique, et d'organiser les plans de moisson
- Pour le trading agricole, il permet d'anticiper les productions et donc la volatilité des prix de vente des céréales françaises sur les marchés mondiaux
- Pour les exploitants agricoles, il optimise les itinéraires culturaux (fertilisation, gestion des traitements phytosanitaires, gestion de la récolte et de sa commercialisation)

L'expansion internationale de SMAG et les opportunités offertes par le réseau du groupe INVIVO permettront d'accéder à de nouveaux marchés potentiels : semences, agrofourniture, biocontrôle, logistique des grains, assurances agricoles, trading.



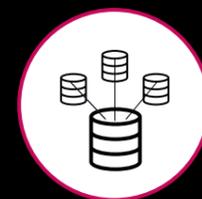
# DATA ENGINEERING



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



GDPR



MDM



IoT



Datalake



Vision 360



Risks & Frauds





### KEYRUS, CRÉATEUR DE VALEUR À L'ÈRE DE LA DATA ET DU DIGITAL

Acteur international du conseil et des technologies, spécialiste de la Data et du Digital, Keyrus a pour mission d'aider les entreprises à tirer profit du paradigme de la Donnée et du Numérique pour accroître leur performance, faciliter et accélérer leur transformation et générer de nouveaux leviers de croissance, de compétitivité et de pérennité.

Keyrus développe une proposition de valeur unique sur le marché autour d'une offre novatrice qui s'appuie sur la combinaison de trois expertises majeures et convergentes :

**MANAGEMENT & TRANSFORMATION**  
Stratégie & Innovation  
Pilotage de la Performance  
Transformation Digitale  
Accompagnement des Projets

**DATA INTELLIGENCE**  
Data Science  
Big Data & Cloud Analytics  
Business Intelligence  
CPM/EPM

**DIGITAL EXPERIENCE**  
Innovation & Stratégie Digitale  
Marketing Digital & CRM  
Commerce Digital  
Performance Digitale

### DÉCOUVREZ C2X DE KEYRUS SOLUTIONS : UNE ANALYSE AVANCÉE ET PRÉDICTIVE DE VOS DONNÉES POUR BOOSTER VOS REVENUS !

C2X, Customer Experience Excellence, est une solution orientée données, créée en collaboration avec les leaders dans le domaine du Big Data. Destinée aux métiers de la relation client, C2X permet de collecter et d'exploiter un large éventail de données sur vos clients afin de vous offrir une vision enrichie de leurs interactions et besoins.

Grâce à C2X, les acteurs de la relation client peuvent découvrir de nouvelles opportunités tout en favorisant les montées en gamme et les ventes croisées jusqu'alors non identifiées. D'une part, ces nouvelles données agrégées et exploitables autour de vos clients permettent de mieux identifier et anticiper les risques de défections. D'autre part, les scores d'appétences offerts par C2X vous aident à prioriser vos actions commerciales.

En partenariat avec



### CHIFFRES CLÉS



**227,9 M€**  
CA 2016



**+ 2 800**  
Collaborateurs



**+ 15**  
Pays



**+ 100**  
Collaborateurs Big Data & Analytics en France

### NOS RÉFÉRENCES DANS LE BIG DATA

#### MARKETING TEMPS RÉEL & OMNICANAL

Nos experts Big Data ont répondu aux attentes de la filiale du **leader mondial de l'eau** qui désirait concevoir et réaliser une plateforme d'engagement client :

- > Basée sur l'analyse du comportement client
- > Capable de détecter des changements de comportements et d'adopter des stratégies d'engagement adaptées
- > Omnicanal : web, mobile, email et courrier via une solution éditique automatisée

#### CONNAISSANCE CLIENT 360°

Le **leader mondial de l'assurance** a fait appel à notre expertise pour :

- > La conception et réalisation d'une plateforme de consolidation de la vision 360° client en temps réel
- > La mise en œuvre de solutions innovantes autour de partenaires technologiques de pointe
- > Une approche entièrement en cloud dans le respect des normes de sécurité et de chiffrement de référence

#### MARKETING PERSONNALISÉ

Nous avons accompagné un **célèbre quotidien gratuit** destiné aux usagers de transports en commun dans :

- > La création d'un moteur de recommandations de contenus et l'optimisation de sa connaissance client à travers un Boot Camp Big Data
- > Le développement d'un moteur de recommandations hybride multi paradigmes : content based & collaborative filtering et multi stratégies
- > La conception d'une architecture temps réel, réactive et scalable permettant de délivrer les recommandations

#### SMART GRID/IOT

Nous avons répondu aux besoins de la filiale distribution du **leader mondial de l'électricité** :

- > La conception et pilotage d'un système de gestion de flotte de plusieurs milliers de véhicules électriques
- > L'optimisation de l'usage de la flotte afin de prioriser les véhicules électriques sur leurs trajets
- > Enjeu de sécurisation : maîtriser les appels de puissance électrique de la recharge
- > La conception et pilotage d'un système de gestion de flotte de plusieurs milliers de véhicules électriques

## LE BIG DATA

## UNE RÉALITÉ À VOTRE PORTÉE !

**LEADER DU BIG DATA, KEYRUS POSSÈDE À LA FOIS LA CONNAISSANCE MÉTIER, L'EXPERTISE ANALYTIQUE ET LA MAÎTRISE TECHNOLOGIQUE INDISPENSABLES AU SUCCÈS DE VOS PROJETS BIG DATA.**

- > Valoriser les données massives en les transformant en connaissances exploitables
- > Augmenter les capacités d'analyse pour une meilleure anticipation et prédiction
- > Améliorer la performance des dispositifs digitaux de nouvelle génération
- > Optimiser et faire évoluer les architectures décisionnelles

### KEYRUS VOUS ACCOMPAGNE SUR L'ENSEMBLE DE VOS PROJETS BIG DATA ET ANALYTICS

#### VOS BESOINS

#### NOS OFFRES

Mieux comprendre les opportunités et enjeux du Big Data



FORMATIONS ET SÉMINAIRES

Définir et planifier votre feuille de route Big Data



STRATÉGIE, PROCESS ET MÉTHODOLOGIE

Analyser le potentiel de vos données  
Réaliser un POC



CONDUITE DE POC

Faire monter vos équipes en compétence  
Renforcer vos projets Big Data par de l'expertise



ACCOMPAGNEMENT MÉTIERS & DATA

Disposer d'une architecture Big Data  
Déployer vos POC



INDUSTRIALISATION

· MARKETING PERSONNALISÉ  
· MARKETING TEMPS RÉEL ET UBIQUITAIRE

· CONNAISSANCE CLIENT 360°  
· PRICING DYNAMIQUE

· LUTTE CONTRE LA FRAUDE  
· GEOFENCING

· MAINTENANCE PRÉDICTIVE  
· USINE DU FUTUR

· SMART CITIES  
· SMART GRID  
· IOT / IDO  
· MTOM



### LA PRIORITÉ DE CONSTRUCTION D'UN DATA LAKE

Le retour d'expérience de nos nombreuses interventions clients sur des projets Big Data a fait émerger un constat clair : dans toutes les approches clients, il y a la volonté de traiter leurs données dans un Data Lake Hadoop.

Le développement de Data Lake s'est accéléré avec la convergence du besoin de plateformes fédératrices dans les entreprises et de nouveaux moyens techniques économiques apportés par les technologies Big Data.

Un Data Lake est un référentiel de données permettant de stocker une très large quantité de données brutes dans le format natif pour une durée indéterminée. Cette méthode de stockage permet de faciliter la cohabitation entre les différents schémas et formes structurelles de données, généralement des blobs d'objets ou des fichiers.

Au sein d'un seul Data Lake, toutes les données de l'entreprise sont stockées. Les données brutes, y compris les copies des données système source, côtoient les données transformées.

Le Data Lake regroupe les données structurées en provenance de bases de données relationnelles en couloir ou en colonne, les données semi-structurées telles que les CSV, les logs, les XML, les JSON, et les données non structurées telles que les emails, les documents et les PDF. On y trouve même des données binaires telles que des images, des fichiers audios ou des vidéos.

Le point commun à toute construction d'un Data Lake est d'implémenter d'une part des chaînes d'acquisition de traitement et d'autre part des données de tous types dans ce même Data Lake. Ainsi, avant même de pouvoir exploiter et exposer les données du Data Lake, nous avons constaté des délais parfois considérables pour construire ce Data Lake. Ces délais sont souvent couplés à une indisponibilité de ressources expertes maîtrisant les technologies permettant le développement des chaînes d'acquisition et d'ingestion des données. Il est fréquent de constater des délais de plusieurs mois voire années avant d'intégrer l'ensemble des données.

Il faut rappeler qu'au-delà du stockage, l'un des enjeux du Data Lake est de pouvoir très facilement traiter et transformer l'information afin d'accélérer les cycles d'innovation, et ainsi être un support aux différentes initiatives data.

Malheureusement, cet enjeu important se trouve complètement obstrué par les difficultés rencontrées lors de sa mise en place. De ce constat est né l'idée de proposer DM360, un outil permettant d'ingérer rapidement tous types de données, depuis des sources externes (tous types de fichiers plats (xls, xml, json, ...), BDD SQL et Nosql, web services, open data, CRM, ...) vers le Data Lake, de façon automatisée, entraînant un double avantage pour l'entreprise : réduire considérablement les délais et les coûts d'ingestion des données dans le Data Lake.

Pour illustrer concrètement cet apport, il suffit de prendre en exemple un cas rencontré au sein d'un grand groupe : 2500 fi-

chiers étaient à intégrer dans le Data Lake, nécessitant pour chaque fichier le développement d'un job Spark. 10 data engineer ont été mobilisés durant une période de plus d'un an.

Avec l'utilisation de DM360, les délais et les coûts sont divisés par 10, ce qui représente un enjeu considérable. Chaque job Spark est généré automatiquement en prenant en compte les paramètres du fichier, permettant son ingestion de façon rapide et automatisée dans le Data Lake.

Mais ce n'est pas tout, car le même travail, une fois les données ingérées, est à effectuer en sortie du Data Lake. Il est effectivement nécessaire de créer un job Spark pour les exposer (établir des rapports, visualiser les données, analyser les données ou le Machine Learning).

De même, la modification de toutes ces chaînes (par exemple suite à l'intégration d'une nouvelle donnée ou d'un nouvel indicateur) entrainera à nouveau l'intervention d'un data engineer en charge de modifier la chaîne de bout en bout.

En prenant en compte le cycle de vie des données et les nombreuses interventions nécessaires à l'enrichissement et l'optimisation du Data Lake, le coût devient une contrainte importante, malgré les technologies open source.

En résumé, notre volonté est d'apporter une agilité, flexibilité et rapidité dans le management des données END-TO-END du Data Lake.

Les 5 grandes fonctionnalités présentes au sein de DM360 sont les suivantes :

- Une ingestion automatisée des données depuis tout type de sources externes vers le Data Lake.
- Un moteur d'agrégation, de jointures, d'enrichissement et de calcul.
- Un choix d'exposition (que ce soit dans hdfs, hive, elastic) ou d'exportation vers une BDD Sql externe.
- Un ordonnanceur pour la gestion des cycles de vie des données.
- Une interface utilisateur via Angular 4 pour une simplicité d'utilisation et un rendu visuel optimal.

### COMMENT VOYEZ-VOUS L'ÉVOLUTION ET LA MATURITÉ DU BIG DATA ?

Hadoop et les technologies NoSQL ont une dizaine d'années aujourd'hui. Après une période d'attente au sein de plusieurs entreprises, marquée par la volonté de résultats prouvés dans leur domaine avant de se lancer, nous constatons dans presque tous les secteurs une évolution importante : ces grandes entreprises ont rejoint les groupes précurseurs déjà convertis depuis de nombreuses années au Big Data. Et si ce n'est pas encore le cas, cela fait partie de leur stratégie à moyen ou à long terme. De façon générale, notre ressenti est que le Big Data devient l'enjeu stratégique majeur dans le domaine IT. Une fois la décision prise de se lancer dans le Big Data, il existe diverses manières d'adopter ces technologies. La tendance étant aujourd'hui d'utiliser un Data Lake, ce que nous partageons. En effet, dans une démarche de data management, le Data Lake permet de collecter tout type de données, de raffiner celles-ci pour offrir une vision 360. Le Data Lake est basé sur des technologies qui permettent le traitement in-situ des données. Le fait de disposer de puissance de calcul directement associée au stockage permet de raffiner un flux de données, et ainsi de facilement en créer les déclinaisons métier attendues. La richesse des outillages intégrés permet ensuite de tirer parti des données.

### QUELS SONT LES CHALLENGES ET DIFFICULTÉS QUE RENCONTRENT LES ENTREPRISES DANS LA GESTION DE LEURS DONNÉES ?

Une des difficultés rencontrées par les entreprises est l'offre pléthorique dans le domaine Big Data, rendant le choix très complexe. Il en ressort que les entreprises se sentent perdues et démunies face à une offre de plus en plus grande. Lors de la mise en place d'un Data Lake, nous rencontrons souvent deux problèmes récurrents : d'une part, le nombre de projets de mise en place de Data Lake couplés à une pénurie de data engineer entraîne une mise en route difficile et compliquée dans le parcours de mise en place d'un Data Lake. D'autre part, il est fréquent qu'une entreprise ait des milliers de sources à intégrer dans son Data Lake. Pour chaque source, une chaîne de traitement (ingestion, mais aussi calcul et exposition) doit être développée. C'est un travail titanesque à l'échelle de plusieurs milliers de sources. Enfin, les solutions mises en place au sein de l'entreprise font face au manque de connaissances des équipes métiers, matérialisé par une maîtrise imparfaite des outils ainsi mis à leur disposition et destinés à leur donner accès au Big Data. Pour aider nos clients dans cette logique, nous avons mis au point une solution de data management, DM360, visant à apporter plus d'autonomie aux entreprises.

### COMMENT RÉPONDEZ VOUS AUX DIFFÉRENTS CHALLENGES ET PROBLÈMES CITÉS ?

Concernant la pénurie d'ingénieurs ayant une expertise Big Data, auquel nous avons nous-même été confrontés dès notre entrée dans le Big Data, nous avons créé Lansrod Data Factory, un centre de compétence unique permettant de former des ingénieurs issus de grandes écoles au métier du Big Data. Au-delà de leur cursus de haut niveau, ils ont en point commun une grande motivation. Ils ont à disposition des clusters « tests » pour évaluer les technologies en toute sécurité. Ils sont mis à contribution sur des pro-

jets de grosse volumétrie de type industriel et non « PoC ». Ceci implique qu'ils rencontrent les difficultés inhérentes et présentes sur les projets clients. Cela leur confère un avantage important par rapport aux autres ingénieurs Big Data. Une certification Big Data Hortonworks HDP® vient valider ses acquis. Cette stratégie nous a permis de répondre qualitativement et quantitativement aux besoins de nos clients en ressources Big Data, et a été validée par nos clients : en effet, 5 ingénieurs constituaient le Lansrod Data Factory lors de la création du centre en 2014. Ils sont aujourd'hui 50, sans compter les 30 ingénieurs ayant intégré des projets clients. Notre application DM360 répond à l'autre problématique rencontrée lors de l'intégration et lors de la gestion des milliers de sources dans le Data Lake. Celle-ci se résume simplement en plusieurs fonctionnalités :

- La capacité à permettre l'ingestion automatique de tous types de flux.
- Un moteur d'agrégation, de jointures, d'enrichissement et de calcul intégré.
- Un choix d'exposition ou d'exportation vers une BDD Sql externe.
- Un ordonnanceur pour la gestion des cycles de vie des données.
- Pour terminer, une interface utilisateur pour gérer l'ensemble de ces fonctionnalités. En conséquence, il n'est plus nécessaire de faire appel à un data engineer pour développer ces chaînes de traitement spark. Le gain, rapporté à l'ensemble des flux, est énorme. L'objectif est d'apporter une meilleure autonomie aux entreprises, et de leur permettre de se concentrer sur des problématiques métiers.

### QU'EN EST-IL DE VOTRE STRATEGIE DE LA CULTURE DATA ?

Les experts choisissent déjà LANSROD pour son approche unique du Big Data et sa capacité à leur proposer des projets complexes mais aussi une structure incluant un support important. L'organisation de l'entreprise basée sur la prise d'initiative et l'innovation sont valorisées et encouragées. Il y a une verticalité importante, où les experts échangent fréquemment avec un groupe de data engineer, afin de capitaliser sur les acquis et connaissances de chacun. J'y ai moi-même connu une belle progression en étant aujourd'hui architecte et techlead Big Data au sein d'un leader de la finance.

**Fahd ESSID**  
fahd.essid@lansrod.com  
+33 6 58 78 10 95



## HISTORIQUE

Fondée en 2009 à San José en Californie, MapR Technologies est un éditeur logiciel visionnaire qui a créé une plate-forme unique permettant de gérer n'importe quel type de données à travers l'ensemble des Clouds (privés, publics ou hybrides).

Nous permettons à nos clients d'exploiter la pleine puissance de leurs données grâce au premier système de données nouvelle génération, la « MapR Converged Data Platform ».

## COEUR DE METIER

Avec la Converged Data Platform de MapR, les entreprises peuvent effectuer de l'analytique et des applications simultanément au fur et à mesure que les données sont générées afin de créer de nouvelles applications, plus modernes et plus intelligentes, pour surpasser leurs concurrents.

MapR aide ses clients à traiter des problématiques telles que les vues client 360°, la détection de fraude, la gestion de risque, les moteurs de personnalisation ou de recommandation, l'analyse prédictive ou les objets connectés, pour n'en citer que quelques-unes.

## SERVICES / PRODUITS PROPOSES

La Converged Data Platform de MapR permet le traitement immédiat et direct des fichiers, tables et streams sur la même plate-forme.

Contrairement aux environnements « inter-connectés » qui demandent l'intégration et l'orchestration complexe d'un grand nombre de technologies, la convergence offre une architecture simplifiée permettant une vision des données en temps réel, un environnement de sécurité unifié et cohérent à travers l'ensemble des composants, une meilleure utilisation des ressources et une charge d'administration réduite.

De plus, au-delà des services de support assurés en 24x7, MapR propose également des services de consulting ainsi que des formations en ligne ou sur site.

## SECTEURS CIBLES

Les marchés adressés par MapR sont essentiellement les services financiers (banque de détail, banque d'investissement, assurances), la grande distribution, l'industrie, les médias, télécoms, services publics, et bien évidemment le secteur des nouvelles technologies.

## COUVERTURE GEOGRAPHIQUE

MapR est présent dans plus de 20 pays à travers les 3 principaux continents (Etats-Unis, Europe et Asie-Pacifique). Cette présence permet à MapR d'être au plus près de ses clients, et ainsi de leur fournir les meilleurs services en terme de consulting, ainsi bien sûr qu'un support 24x7x365 en « follow the sun ».

## AVANTAGES CONCURRENTIELS

Historiquement, MapR est reconnu pour proposer la seule distribution Hadoop hautement disponible et dédiée aux applications critiques. En plus de cette proposition de valeur historique, MapR a intensifié ses efforts de recherche et développement pour réduire la complexité d'intégration des nombreux composants nécessaires à la mise en œuvre d'une plate-forme de gestion de données, tels que Hadoop, Spark, le NoSQL, et Streaming et les moteurs SQL. En faisant converger toutes ces technologies dans une unique plate-forme, MapR réalise un tour de force en simplifiant l'intégration de ces composants essentiels tout en garantissant la haute disponibilité et la sécurité de l'ensemble. En se positionnant bien au-delà de Hadoop, MapR offre à ses clients la seule plate-forme de gestion de données conçue pour supporter leurs projets les plus ambitieux et les plus critiques.

Les plus grandes entreprises, tous secteurs confondus, font confiance à MapR pour résoudre leurs plus importants défis et construire leur Data Platform de demain.

# ALL DATA ONE PLATFORM EVERY CLOUD

## LIMITLESS POSSIBILITIES

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





### FREIGHTOS RÉALISE DU BIG DATA POUR LA LOGISTIQUE DE FRET EN LIGNE AVEC MATLAB ET GOOGLE BIGQUERY

Freightos a développé un marché de fret en ligne, soutenu par un puissant système d'itinéraire et de tarification en ligne qui remplace les processus et les outils manuels lourds, éliminant bon nombre des inefficiences et des erreurs qui contribuent à des pertes estimées à 650 millions de dollars chaque année pour l'industrie. Le système Freightos software-as-a-service (SaaS) utilise Google® BigQuery pour gérer et stocker plusieurs bases de données pour des milliers de contrats de fret, des millions de devis de fret et un large éventail d'autres données d'expédition de certains des plus grands fournisseurs de logistique au monde.

**DÉFI :** Avec des millions de lignes de données stockées dans plusieurs bases de données sur le cloud, les analystes de Freightos ont constaté que réaliser leurs analyses était laborieux, rendant les idées clés pour les tâches opérationnelles et de marketing difficiles à identifier. Bien qu'ils puissent exécuter des requêtes sur les données dans le cloud pour les rapports statiques, ils ont besoin d'explorer et d'analyser de façon interactive les données en temps réel pour en extraire des idées. Comme la plupart des grands jeux de données, celui de Freightos était trop important pour être téléchargé pour un traitement en local.

**SOLUTION :** Freightos a travaillé avec le service Consulting de MathWorks pour développer une approche automatisée pour extraire des données de BigQuery et les importer dans MATLAB. Parce que Freightos s'appuie sur des sources externes pour une grande partie de ses données, les analystes doivent nettoyer les résultats de la requête avant de les analyser. L'équipe a développé des scripts MATLAB qui identifient et corrigent les valeurs aberrantes dans les données. Ils ont utilisé la Statistics and Machine Learning Toolbox™ pour effectuer une analyse statistique complexe des données nettoyées. En travaillant avec MATLAB, Freightos et les consultants MathWorks ont généré des diagrammes de dispersion, des arborescences et d'autres visualisations pour mieux comprendre les résultats de l'analyse. Ils ont également utilisé les résultats analytiques pour créer des rapports qui aident les cadres exécutifs Freightos à prendre des décisions et à définir des stratégies marketing.

**RÉSULTATS :** Les analyses sont effectuées en quelques minutes au lieu de quelques heures. « Avec notre intégration de MATLAB et de BigQuery, les plus gros rapports, tels que les rapports trimestriels sur un jeu de données complet, peuvent être générés en 10 minutes », explique Leonid Hatskin, Data Analyst chez Freightos. « Dans le passé, cela aurait pris au moins une journée complète de travail. MATLAB a libéré au moins une journée par mois du temps de travail de notre équipe données. L'intégration de BigQuery a été rapidement implémentée : « Lorsque nous avons demandé au service Consulting de MathWorks d'avoir une assistance pour l'intégration de BigQuery, nous avons obtenu un support rapide d'une personne réelle » note Eytan Buchman, Vice President of Marketing chez Freightos. « Ce soutien a transformé MATLAB d'outil puissant à outil irremplaçable ». Des informations acquises 20 fois plus rapidement. « Par rapport à notre approche précédente, notre temps pour obtenir des informations depuis les données est 20 fois meilleur avec MATLAB et BigQuery », a déclaré Eytan Buchman. « L'amélioration du délai de mise sur le marché réalisée n'a pas de prix ».

### BAKER HUGHES, UNE SOCIÉTÉ DE GE (BHGE), DÉVELOPPE UN LOGICIEL DE MAINTENANCE PRÉDICTIONNELLE POUR DES ÉQUIPEMENTS D'EXTRACTION DE GAZ ET DE PÉTROLE À L'AIDE DU DATA ANALYTICS ET DU MACHINE LEARNING

Les camions BHGE sont équipés de pompes à déplacement positif qui injectent un mélange d'eau et de sable à des pressions élevées profondément dans les puits forés. Avec des pompes représentant 100 000 dollars du coût total des 1,5 millions de dollars du camion, Baker Hughes devait déterminer quand une pompe devait tomber en panne.

**DÉFI :** Si un camion sur un site actif a une panne de pompe, BHGE doit immédiatement remplacer le camion pour assurer un fonctionnement continu. L'envoi de camions de rechange sur chaque site coûte à l'entreprise des dizaines de millions de dollars de son revenu que ces camions pourraient générer s'ils étaient utilisés activement sur un autre site. L'incapacité à prédire avec précision quand les soupapes et les pompes nécessiteront une maintenance étaye les autres coûts. Une maintenance trop fréquente gaspille les efforts et entraîne le remplacement de pièces encore utilisables, et une maintenance trop rare risque de nuire aux pompes au-delà de la réparation.

**SOLUTION :** En travaillant avec MATLAB, les ingénieurs de BHGE ont développé un logiciel de surveillance du bon fonctionnement des pompes qui applique des techniques de machine learning en temps réel pour prédire l'instant idéal pour effectuer la maintenance. Ils ont traité et analysé jusqu'à un téraoctet de données collectées à plus de 50 000 échantillons par seconde à partir de capteurs installés sur 10 camions opérant sur le terrain, identifié les paramètres utiles pour prédire les pannes et créé et entraîné un réseau de neurones pour utiliser les données des capteurs afin de prédire les pannes de pompe.

**RÉSULTATS :** Des économies de plus de 10 millions de dollars ont été projetées : « En une seule année, nous pouvons consacrer une quantité importante de nos revenus uniquement à la maintenance et au remplacement des composants internes de la pompe, tels que les soupapes, les sièges de soupape, les pistons et les joints », explique Thomas Jaeger, responsable produit senior chez BHGE. « Nous estimons que le logiciel que nous avons développé avec MATLAB réduira les coûts globaux de 30 à 40%, ce qui s'ajoute aux économies que nous verrons en éliminant le besoin de camions supplémentaires. » Des temps de développement réduits par dix : « MATLAB nous a permis d'effectuer nos analyses et traitements souhaités, y compris du machine learning », selon Gulshan Singh, Reliability principal and team lead for drilling services chez BHGE. « Avec un langage de plus bas niveau, vous ne pouvez pas toujours trouver les bibliothèques dont vous avez besoin et réaliser le projet dans un délai alloué de quelques semaines. Si nous devons écrire notre propre code en utilisant des bibliothèques avec un langage de bas niveau pour toutes les fonctionnalités déjà intégrées dans MATLAB dont nous avons besoin, il aurait probablement fallu bien plus de temps pour compléter ce projet. » Plusieurs types de données accessibles facilement : « MATLAB a facilité la combinaison de plusieurs types de données en une seule application d'analyse », explique Gulshan Singh. « Nous avons même pu utiliser des données de capteurs utilisant un format de fichier propriétaire ».

### QUELLES SONT LES FONCTIONNALITÉS OFFERTES PAR LES PLATEFORMES DE DATA SCIENCES ET QUELS SONT LES AVANTAGES DE L'UTILISATION DE CE TYPE DE PLATEFORME ?

Les plateformes de data sciences rendent les outils de data sciences accessibles à un plus grand nombre d'utilisateurs. Elles comportent des outils pour se connecter à des sources de données, de prétraiter ces données, de combiner des données provenant de diverses sources, d'appliquer des techniques de machine learning et de déployer des analyses dans des systèmes de production. Traditionnellement, ces outils n'ont été accessibles qu'à ceux qui possèdent de solides compétences en programmation informatique et n'ont pas été conçus pour être facilement intégrés dans des environnements plus larges. En fournissant ces outils dans un package facile à utiliser, ces plateformes offrent un accès aux data sciences à un public plus large. Elles aident les organisations à faire face aux difficultés liées au recrutement de data scientists. Les plateformes de data sciences servent également de norme interne pour les processus de développement de data sciences, facilitant la collaboration entre les équipes sur ces projets.

### QUELS TYPES D'ENTREPRISES POURRAIENT UTILISER UNE PLATEFORME DE DATA SCIENCES ?

Nous observons un intérêt pour les data sciences dans de nombreuses industries, depuis les entreprises spécialisées en équipements industriels qui cherchent à utiliser les data sciences pour faire de la maintenance prédictive, jusqu'aux entreprises du secteur des services financiers qui utilisent les data sciences pour développer de nouvelles stratégies commerciales. Dans chacune de ces applications, les données sont disponibles ou peuvent être collectées pour résoudre des problèmes de longue date. Par exemple, dans le cas de la maintenance prédictive, les opérateurs d'équipements industriels coûteux ont constaté qu'ils peuvent prédire de manière fiable la durée de vie utile restante d'un équipement et optimiser les horaires de service en appliquant des techniques de machine learning sur les données générées par les capteurs. Les plateformes de data sciences permettent aux équipes d'ingénieurs qui développent et maintiennent ces équipements de tirer parti de la richesse de leurs connaissances sur la façon dont l'équipement devrait fonctionner. Cette idée de monter en compétence les ingénieurs, ou « experts domaine », est souvent plus attrayante que le recrutement de data scientists qui ont peu de connaissances sur le fonctionnement de l'équipement.

### DANS QUELS CAS LES PLATEFORMES DE DATA SCIENCES SONT APPROPRIÉES, ET DANS QUELS CAS NE LE SONT-ELLES PAS ?

Les plateformes de data sciences sont adaptées aux analyses de données exploratoires telles que la compréhension des tendances, la suppression des valeurs aberrantes, l'analyse statistique, ainsi qu'à la construction de modèles de machine learning. Pour offrir un processus complet, certaines plateformes de data sciences incluent des fonctionnalités permettant de développer un algorithme complet contenant un modèle de machine learning comme composant important. Par ailleurs, les plateformes

de data sciences peuvent fournir des outils spécifiques au type de données utilisées - par exemple, des techniques de traitement d'image pour les données image, des techniques de traitement du signal pour les données issues de capteurs et des analyses de traitement de texte pour les données textuelles.

Les plateformes de data sciences ne conviennent pas à la création de code de production pour des applications telles que l'infrastructure de réseau ou le développement web. Même si certaines plateformes de data sciences peuvent s'en rapprocher, elles n'ont pas pour vocation de remplacer pleinement les environnements de développement intégrés au niveau du développeur.

### SELON QUELS CRITÈRES LES ENTREPRISES DEVRAIENT-ELLES CHOISIR UNE PLATEFORME DE DATA SCIENCES ?

Les entreprises devraient considérer qui, dans l'organisation, utilisera la plateforme de data sciences. Les business units auront une expertise dans leurs secteurs d'activité respectifs, mais auront probablement besoin d'un outil plus facile à utiliser par rapport à une équipe de data sciences centralisée. Les organisations devraient également tenir compte des types de données qu'elles utiliseront. Beaucoup de plateformes de data sciences ont été conçues pour travailler avec des données de marketing et de vente, et ne s'adaptent pas aux sources de données plus récentes telles que les images, les vidéos, l'audio et les données capteurs. Un autre facteur à considérer est l'importance, pour l'organisation, de se différencier par ses programmes de data sciences. Dans les marchés concurrentiels, les organisations devraient chercher à adopter des outils qui offrent plus de flexibilité pour personnaliser l'analyse afin qu'elles ne finissent pas à être en concurrence avec les données seules.

### QU'EST-CE QUI DISTINGUE VOTRE PLATEFORME DE DATA SCIENCES ET QU'EST-CE QUI LA REND DIFFÉRENTE DE LA CONCURRENCE ?

MATLAB fournit une grande variété de façons d'accéder aux données provenant de plusieurs sources, y compris les données commerciales issues de bases de données, d'entrepôts de données et d'Hadoop, mais aussi les données d'ingénierie provenant de capteurs, de logiciels historiens et suivant des protocoles spécifiques à l'industrie. L'innovation se produit souvent lorsque diverses sources de données sont combinées. La connexion aux données, quel que soit le format, devient donc très importante. Nous mettons également l'accent sur l'accessibilité des outils utilisés par les data scientists aux experts domaine. Nous rendons ceci possible en mettant à disposition des fonctions et applications faciles à utiliser pour le machine learning, le deep learning, la vision par ordinateur, le traitement du signal, l'optimisation numérique et autres technologies d'analyses avancées. Enfin, MATLAB propose plusieurs solutions de déploiement pour l'exécution des analyses déployées sur des dispositifs embarqués, ainsi que sur une infrastructure informatique ou un cloud. La possibilité de choisir entre plusieurs solutions de déploiement continue de gagner en importance, car de nouvelles tendances comme l'internet des objets entraînent les équipes à réévaluer les processus afin de déterminer si le traitement doit être embarqué au niveau de l'objet, ou réalisé dans le cloud ou de façon hybride.



Jean-Pascal Ancelin,  
Vice-Président Europe du Nord  
et Directeur Général  
MicroStrategy France

### ANALYTICS : MICROSTRATEGY, LA PLATEFORME D'ENTREPRISE QUE LES MÉTIERS CHOISSENT

« L'analytique aujourd'hui est un écosystème, dont MicroStrategy est la clé de voûte. »

### EN QUOI L'ANALYTIQUE A-T-IL CHANGÉ DEPUIS 30 ANS ?

Les entreprises ont débuté avec des besoins en reporting, tableaux de bord dynamiques, requêtes ad hoc, analyse multidimensionnelle. Elles attendent aujourd'hui une plateforme analytique « On-premise » ou Cloud, puissante et ouverte vers les métiers. Ce qui a changé est sans conteste le volume de données dont disposent les sociétés aujourd'hui, et leur caractère non structuré, qu'il faut pouvoir intégrer. Monitoring, suivi des process, multiplication des capteurs ont généré plusieurs transformations derrière le potentiel de croissance : celle du stockage tout d'abord, qui ne peut plus se satisfaire de bases de données traditionnelles et très onéreuses. Les modes d'accès à l'information, ensuite, ont contribué à modifier notre consommation de data et l'usage que l'on en fait. Enfin, l'hyper concurrence des marchés, dont la réponse adaptée exige de la réactivité, et par conséquent, la capacité de l'entreprise à savoir transformer cette matière première protéiforme en information de valeur, immédiatement. L'analytique aujourd'hui est un écosystème, dont MicroStrategy est la clé de voûte.

### COMMENT MICROSTRATEGY ORCHESTRE-T-ELLE L'ANALYTIQUE EN ENTREPRISE ?

MicroStrategy relève de la partie visible des projets décisionnels, l'interface utilisateur, celle qui lui permettra d'étudier des problématiques, simples ou très complexes et d'en tirer les bonnes conclusions. Pour accéder à la donnée utile, MicroStrategy intègre une gestion multi-sources (ERP, bases de données, systèmes métiers, logistique, RH...). Nous proposons plus d'une centaine de connecteurs natifs et certifiés. En outre, MicroStrategy adresse tous les profils d'utilisateurs dans une entreprise, c'est aussi ce qui la distingue de certains de ses concurrents. Vraie plateforme d'entreprise, elle est tout autant faite pour les consommateurs de rapports, qui souhaitent recevoir l'information dynamique en temps réel, via mobile par exemple, au bon moment, que pour les utilisateurs avancés, en charge de bâtir les modèles mathématiques et les indicateurs de performances de l'entreprise. Les fonctions de dataviz et de self-services ont apporté aux directions fonctionnelles une grande autonomie.

### DATADRIVEN ENTREPRISE OU ANALYTIQUE DRIVEN ENTREPRISE ?

Oui, nous avons tous entendu parler de l'entreprise pilotée par sa data. Au demeurant, la donnée brute ne sert à rien. Seule sa transformation importe, comme l'on transforme le colza aujourd'hui en énergie. Ou encore sa reconversion, comme le font les entreprises de recyclage avec nos bouteilles en plastique. La donnée brute est encombrante. L'analytique est un patrimoine.

« La donnée brute ne sert à rien. Seule sa transformation importe. »

MicroStrategy offre une richesse fonctionnelle telle qu'elle peut accompagner l'entreprise dans sa progression et sa compréhension du big data et de l'analytique d'entreprise au fil de la montée en maturité de ses équipes. Transactionnel, data mining, prédictif et data visualisation, marche après marche, et ce jusqu'au machine learning, l'entreprise transforme ses usages, ses techniques et surtout ses expertises. Pour aboutir, un jour, à l'automatisation complète de sa chaîne de process.

### EN QUOI MICROSTRATEGY SÉDUIT AUTANT LES MÉTIERS, CECI AVEC L'AVANTAGE DE LEUR DSI ?

Du point de vue de la DSI, MicroStrategy a l'avantage d'être une solution totalement intégrée « On-premise » ou Cloud et déployable à l'ensemble des directions fonctionnelles. Nous développons nos propres solutions, qui offrent toute la cohérence nécessaire dans un univers data amplement hétérogène. Comme nous n'avons racheté aucune technologie tierce, les migrations sont simples et notre référentiel de métadonnées MicroStrategy garantit la bonne gouvernance de la donnée. Par ailleurs, nos clients peuvent s'appuyer sur un maillage de partenaires technologiques et intégrateurs, dont l'expérience et le savoir-faire sur MicroStrategy sont reconnus et certifiés. Sous l'angle métier, MicroStrategy accorde beaucoup d'importance à l'expérience utilisateur. L'esthétique de l'affichage des graphiques et des résultats, la beauté des applications créées, n'est pas un luxe, c'est un incontournable à l'adoption d'une solution. Nos clients relèvent la rapidité avec laquelle les métiers s'emparent de l'outil et le plaisir qu'ils y prennent.

« L'esthétique n'est pas un luxe, c'est un incontournable à l'adoption. »

Cette appropriation, c'est aussi le pont vers le self-service, cette manière de se servir des données disponibles et les corréler, les comparer, en toute liberté, avec des recherches faciles, une visualisation par filtres et sélecteurs. La mobilité et l'analyse temps réel s'ajoutent au succès que la plateforme remporte auprès des utilisateurs. Enfin, avec MicroStrategy 10.9, l'interface utilisateur marque une étonnante rupture avec les portails web traditionnels. Très proche de la philosophie d'affichage des applications mobiles, la nouvelle interface se veut aussi accessible qu'un livre, faite de chapitres, de pages à tourner et de sommaires. Avec Dossier pour les analystes et Librairie, pour les métiers, nous pouvons affirmer que l'information coule de source ! Cette évolution est le fruit des retours de nos clients, qui souhaitaient en outre pouvoir commenter en direct les rapports de leurs data-scientists et formuler de nouvelles demandes. C'est aujourd'hui chose faite. MicroStrategy offre un environnement de travail collaboratif très optimisé pour rapprocher toujours plus les services entre eux.

## Valorisez vos investissements Big Data avec MicroStrategy.

### CONNEXION NATIVE AUX BIG DATA

Visualisez et analysez toutes vos données avec des connecteurs prêts à l'emploi.

### PERFORMANCE INÉGALÉE

Traitement parallèle, cubes en mémoire et partitionnement offrent des temps de réponse ultra-rapides sur des milliards de lignes.

### ANALYTICS POUR TOUS

Déployez vers tous les profils d'utilisateurs, analystes, architectes, développeurs ou consommateurs d'information.

Découvrez pourquoi les entreprises choisissent  
MicroStrategy pour analyser leurs Big Data.

[microstrategy.fr/bigdata](http://microstrategy.fr/bigdata)

Contactez-nous pour plus d'information :  
info-fr@microstrategy.com +33 1 47 55 56 00



CONTACT

**RETHINKING THE CUSTOMER EXPERIENCE AT METLIFE FROM STALLED TO SUCCESS IN THREE MONTHS**

MetLife wanted to harness Big Data to create a **360-degree view** of its customers so it could know and talk to each of its more than 100 million clients as individuals. But the Fortune 50 company had already spent many years trying unsuccessfully to develop this kind of centralized system using relational databases.

**Which is why the 145-year old insurer in 2013 turned to MongoDB. Using MongoDB's technology over just 2 weeks, MetLife created a working prototype of a new system that pulled together every single relevant piece of customer information about each client.**

Three short months later, the finished version of this new system, called the 'MetLife Wall,' was in production across MetLife's call centers. **The Wall** collects vast amounts of structured and unstructured information from MetLife's more than 70 different administrative systems. After many years of trying, MetLife solved one of the biggest data challenges dogging companies today. All by using MongoDB's innovative approach for organizing massive amounts of data.

**INSIDE THE METLIFE WALL**

Today, when a customer calls MetLife to ask about a claim, add a new baby to a policy, or dig into coverage details, customer representatives use the Wall to pull up every bit of information they need in seconds – name, address, policies, and life events.

Using a touchscreen and a design based on how Facebook dishes up information, The Wall is instantly familiar to MetLife's call center operators. Which means customer reps can quickly and efficiently answer questions, handle claims, suggest new services, or offer promotions, while slashing wait and call times. MetLife now understands and serves each customer individually.

**POWER OF THE FLEXIBLE DATA MODEL**

Like most companies, MetLife has scores of data systems created or acquired over the years. MetLife's systems contain a huge array of structured and unstructured data, including policy and customer information and transactional history about everything from claims to payments. Few are connected and many are on mainframes with cumbersome interfaces. Ripping out its administrative systems and replacing them with one unified system wasn't an option for MetLife. So the company had tried over the years to use relational databases, which require a common schema and strict mapping of data sources. Adding each new system was an expensive and time consuming process of changing schemas, and extracting, cleaning, and matching data – one that MetLife never won.

**METLIFE'S CHALLENGE: DATA VARIETY**

MetLife's 70+ administrative systems contain a massive variety of structured and unstructured data, of two main types. One includes 50 million policies and 118 million customers. Another represents transactional history about payments and claims, and includes about 190 million documents. MetLife needed a way to pull this all together in a single view.

**RELATIONAL DATABASE DIDN'T WORK**

A relational database resembles many Excel spreadsheets. It has a highly structured table organization (its schema). When it co-

mes to customers, for example, you know different things about each one. You may have just the name and email of some, while for others you also know their phone number and different shipping addresses. To make this fit into a spreadsheet, you need to create lots of columns, many of which will be empty. This database becomes unwieldy and difficult to manage.

**MONGODB'S DOCUMENT APPROACH WORKED**

MongoDB stores information like a series of Word documents. A set of data is stored in a document that has its own schema. When you add a field to a given dataset, you can do so without having to add that field to all of the other documents. For instance, when it comes to managing customer data, you'd use a document for each individual. Everything you know about that person is stored in the document. Some documents have just a few fields, while others contain a lot of information. Adding new information about one customer doesn't require updating all the other documents. What makes MongoDB different is its flexible data model. MongoDB looks at data more naturally, making it easy to evolve schemas in real time. If relational databases are like Excel spreadsheets – where data is organized into sheets but where you add a column every time you add a field, creating a structured but unwieldy project – MongoDB is a series of Word documents. Each entry is a document that can have its own schema.

**FLEXIBLE, SCALABLE, USER FRIENDLY**

**MongoDB's horizontal scaling via automatic sharding provides reliable partitioning of massive amounts of data across multiple servers. And it's flexible, allowing organizations to leverage multiple data centers and multi-temperature storage techniques. Just as crucial for productivity and agile application development is the ease of use MongoDB provides developers. Developers can interact with the database in the same programming language they use to write the application, whether Java, Ruby, Python, or something else. Which means they can focus on building apps instead of wrestling with ORMs.**

**And MongoDB provides a variety of rich features, such as integrated search, geospatial, and native analytics, that don't exist in a traditional database. Giving companies the right resources they need to get projects done quickly.**

**A CATALYST FOR INNOVATION**

Working with MongoDB led not just to The Wall's success, but a wave of innovation inside MetLife. Harnessing Big Data is letting the insurance giant do things it never could do before.

**NOW, METLIFE IS:**

- Rolling out The Wall internationally
- Developing a version of The Wall for its sales force
- Creating a real-time analytical system that predicts customer attrition rates, prompting customer reps to offer alternative products or promotions
- Considering adding social data and data from mobile apps into The Wall to gain an even better understanding of customers, something that's easy to do with MongoDB. With MongoDB, MetLife joined the vanguard of the Big Data revolution. By quickly and successfully cracking one of the biggest problems it faced, MetLife changed the expectations of what is possible.

# Optez pour une base de données nouvelle génération.

Soyez prêt à relever les plus grands défis de votre entreprise.



**LA NASA UTILISE NEO4J POUR EXPLOITER SA BASE DE DONNÉES DE RETOURS D'EXPÉRIENCES**  
« Lessons Learned », la base de connaissances et d'expériences des missions de la NASA depuis 1950, peut dorénavant être consultée, enrichie et traitée grâce à l'utilisation de la base de données de graphes Neo4j.

Depuis 1950, la NASA compile dans une gigantesque base de données appelée « Lessons Learned » l'ensemble des connaissances qu'elle a tirées de ses missions passées. Constamment mise à jour, l'agence spatiale s'appuie sur cette base pour planifier les futurs projets et expéditions dans l'espace. Cette base contient des informations détaillées de chaque mission remontant jusqu'aux années 1950 et chaque archive est révisée et validée avant d'être intégrée. Outre le personnel de la NASA, des milliers de scientifiques, d'ingénieurs, d'enseignants et d'analystes du secteur privé et du secteur public accèdent chaque mois à cette base de données. Sa taille s'étant amplifiée, l'interface utilisée en interne pour interroger le jeu de données - une recherche par mots clés basée sur un algorithme de type PageRank - devenait peu maniable.

Par exemple, lorsqu'un ingénieur essayait d'effectuer une recherche à partir de 23 mots clés dans cette base de données de près de 10 millions de documents, les archives positionnées en premières lignes des résultats se trouvaient là parce qu'elles étaient le plus souvent consultées, non parce qu'elles contenaient la bonne information. En outre, après une telle recherche, l'ingénieur se retrouvait avec plus de 1 000 documents qu'il aurait fallu consulter individuellement pour savoir s'ils contenaient les informations nécessaires.

**Passer des données à la Connaissance, le bénéfice de la base de données de graphes**

La NASA est partie d'un principe simple : l'information n'est pas le pouvoir. C'est la connaissance qui l'est. Ainsi, peut importe le volume d'informations qu'une entreprise ou une organisation peut détenir, si elle ne sait pas la convertir en connaissance exploitable, alors, dans les faits, elle n'a rien.

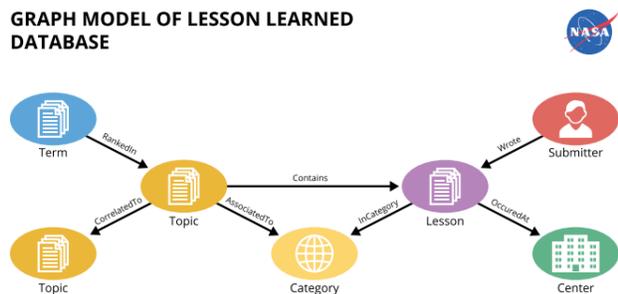
C'est sur ce principe que l'équipe de David Meza, Chief Knowledge Architect de la NASA, a entrepris le développement de cette nouvelle base de données, en combinant gestion des connaissances, maîtrise de l'informatique et de la science de la donnée. La gestion des connaissances est utilisée pour la stratégie, l'informatique pour automatiser le traitement des données et la science de la donnée fournit, quant à elle, les algorithmes et les méthodologies permettant de convertir ces données en connaissances exploitables pour les utilisateurs finaux.

**Transformer les données en graphes pour accéder à la connaissance**  
Dans un premier temps, la NASA a donc importé l'ensemble de ses données dans Neo4j pour bâtir cette nouvelle base.

La NASA a pris les métadonnées directement dans le fichier source et les a converties en une base de données documentaire. Les données sont présentées par une visualisation naturelle qui permet au cerveau de mieux saisir les schémas.

Au delà des données elles-mêmes, la NASA a pris tous les textes issus des apprentissages des projets et les a soumis à un algorithme de modélisation des sujets, l'allocation de Dirichlet latente qui a renvoyé près de 27 sujets. Ce qui a alors permis de disposer d'un nouvel élément de métadonnées qui a pu être ajouté aux apprentissages afin d'améliorer la recommandation : le sujet. Dans la mesure où de nombreuses métadonnées étaient associées à ce type d'apprentissages, l'Agence spatiale a pu interconnecter les sujets par leurs catégories respectives. Elle a pu ainsi non seulement disposer des apprentissages avec leur sujet, mais aussi des corrélations entre les différents sujets. La création de cette base de données a ainsi permis d'offrir à la NASA une visualisation interactive des données qui montre tous les différents sujets générés, ainsi que les termes au sein de ces sujets. Ce qui est particulièrement utile pour un analyste, mais peu utile pour les utilisateurs finaux. Il restait donc à l'agence spatiale à développer un modèle lui permettant de présenter les données à ses utilisateurs finaux. Il s'agit d'un modèle de données de graphes simple basé sur des référentiels spécifiques sur lesquels la NASA s'est concentré parce qu'elle a de très nombreux groupes et que chacun détient son propre jeu de données d'intérêt. Les collaborateurs peuvent ainsi voir un apprentissage (nœud violet) qui a été écrit par celui qui l'a soumis (nœud rouge) dans un centre donné (nœud vert). Il se trouve dans une certaine catégorie (nœud jaune) et correspond à un sujet (nœud orange) lui-même corrélé à un autre sujet qui contient plusieurs termes (nœud bleu).

**GRAPH MODEL OF LESSON LEARNED DATABASE**



NASA's knowledge management graph model.

David Meza, Chief Knowledge Architect à la NASA, se félicite : « **Neo4j nous a permis de gagner 2 ans de travail et de faire économiser 1 million de dollars d'argent public à l'Etat** »  
Prochaine étape pour la NASA : réformer sa base de données des 90000 commentaires des cosmonautes sur les missions dans la station spatiale internationale des 15 dernières années.

Neo4j  
Stand A49



# LES GRAPHES CONNECTENT LES BIG DATA

## LE BIG DATA MÉRITE QU'ON PENSE DIFFÉREMMENT

Il est temps d'exploiter le pouvoir de vos données connectées avec l'aide des bases de données de graphes.

De nombreuses entreprises utilisent déjà Neo4j parmi lesquelles **Airbus, Meetic, eBay, adidas** ou encore **Société Générale**.

RDV sur le **stand A49** pour découvrir ce que Neo4j peut faire pour votre entreprise :

- Big Data
- Détection de fraude
- Recherche basée sur les graphes
- Gestion des identités et des accès
- Gestion des données de référence
- Opérations IT & Réseaux
- Recommandation

Ne manquez pas notre atelier produit Neo4j le Lundi 12 Mars à 14h30 en Salle C.





## L'ANALYTICS DE DEMAIN

### SOLUTIONS D'ENTREPRISE

Utilisez les fonctions d'analyse et les applications du Big Data pour transcender l'intelligence réactive avec des analyses proactives destinées à la gestion des expériences client, des fraudes, de la logistique et des talents.



#### « MARKETING MODERNE »

Offrez des expériences client cohérentes, personnalisées et omnicanal en termes de marketing, de ventes et de services. Améliorez votre visualisation et votre analyse prédictive des comportements avec le Big Data, à la fois au sein et en-dehors de votre entreprise, afin d'optimiser les offres et les actions qui s'imposent en temps réel.



#### ATTÉNUEZ LES RISQUES

Prévenez les comportements frauduleux vis-à-vis des produits et services grâce à l'utilisation du Big Data à une échelle considérable, souvent en temps réel. Combinez Hadoop, NoSQL et des technologies relationnelles pour développer des analyses prédictives plus précises et les déployer plus rapidement de manière à réaliser des économies et à protéger la réputation de votre entreprise.



#### CHAÎNE D'APPROVISIONNEMENT CONNECTÉE

Aujourd'hui, le monde de l'entreprise évolue à un rythme sans précédent : le nombre de données, d'appareils et de connexions ne cesse d'évoluer. Utilisez des données géospatiales pour prendre des mesures intelligentes et personnalisées tout au long de la supply chain, au sein et en dehors de votre entreprise, généralement en temps réel.



#### UN SYSTÈME RH INNOVANT

Recrutez, faites évoluer et conservez les meilleurs talents en capturant et en utilisant le Big Data, quels que soient les points de contact des employés au sein et en-dehors de votre organisation. Gagnez en productivité et en efficacité, avec des informations plus précises, souvent en temps réel, qui vous permettent de mieux analyser et prévoir les performances des employés à chaque niveau de compétence.

### SOLUTIONS SECTORIELLES

Veillez à ce que votre entreprise profite des opportunités du Big Data grâce à des solutions, des applications et une architecture sectorielles.



#### LE BIG DATA DANS...

##### LES TÉLÉCOMMUNICATIONS

Déployez et monétisez rapidement vos services Digital Lifestyle, proposez une expérience client incomparable avec l'analyse géospatiale et rationalisez vos opérations.

##### LA RECHERCHE CLINIQUE

Utilisez les technologies Big Data innovantes pour optimiser la R&D clinique, limiter les risques, faire avancer les soins et améliorer les résultats des patients.

##### LA TECHNOLOGIE

Bénéficiez d'une flexibilité opérationnelle et d'une innovation continue tout en gérant les risques et la conformité.

##### LE SECTEUR PUBLIC

Répondez aux besoins des citoyens de manière plus efficace, suivez et analysez la consommation du service et assurez une meilleure planification.

##### LA VENTE AU DÉTAIL

Proposez une expérience d'achat unique qui dépasse les attentes et les besoins de vos clients. Alignez les individus, les processus et les technologies afin d'accroître la puissance de vos données.

##### LES SERVICES PUBLICS

Proposez des services fiables, sécurisés et évolutifs qui répondent aux demandes croissantes de vos clients, tout en surveillant l'offre et le service afin de garantir l'efficacité des opérations.



#### UN SYSTÈME RH INNOVANT

Fournissez de nouveaux services innovants, faites face à un nombre croissant d'exigences réglementaires et gérez les données disséminées dans des organisations complexes et globales.

# Intelligence artificielle, robotique et Cloud Analytics intégré

Une approche intégrée du Cloud permettra aux entreprises de capitaliser sur l'intelligence artificielle et la robotique pour réinventer leurs modes de fonctionnement.

**AVANTAGE :** Avec notre plateforme Analytics Cloud et des services informatiques intégrés, les entreprises disposent d'un environnement de développement et de tests d'applications qui leur donnera l'avantage dans la course à l'innovation.

**INDISPENSABLES :** Pour exploiter tous les avantages du Cloud et expérimenter de nouveaux services innovants comme l'intelligence artificielle, l'analyse prédictive, les entreprises doivent impérativement intégrer les plateformes Cloud et les services associés.

**INNOVATION :** Quant à celles qui sous-estimeront la puissance du Cloud intégré, elles courent le risque de se voir freinées par des systèmes obsolètes et distancées par leurs concurrents. D'après une enquête, 60 % des sondés pensent que la rigidité de leur infrastructure informatique nuit à leur capacité d'innovation. D'où l'importance de faire le point sur leur approche de l'innovation au-delà de la simple adoption des technologies Cloud.

**RÉSULTATS :** Démultipliez la valeur de l'ensemble de votre portefeuille d'applications grâce au composant d'intelligence artificielle d'Analytics Cloud orienté données. Notre approche unique de conception s'appuie sur des connecteurs prédéfinis, des sources de données tierces enrichies, une science décisionnelle dynamique et des résultats connectés pour améliorer votre activité pour des coûts et des efforts minimaux.

**INTELLIGENCE :** La solution Adaptive intelligence d'Oracle permet le regroupement de l'ensemble de données tierces enrichies et données propriétaires, et exploite la science décisionnelle dynamique pour fournir des résultats intelligents et adaptables dans le contexte des activités professionnelles.

**ÉVOLUTION :** Avec l'Adaptive Intelligence, les applications pourront réagir, apprendre et s'adapter en temps réel.

## Big Data Cloud Service

Oracle Big Data Cloud Service est un service automatisé qui fournit un environnement de haute puissance sur mesure pour faire avancer les capacités analytiques des entreprises. Grâce à la gestion automatisée du cycle de vie et à la sécurité d'un clic, Big Data Cloud Service est conçu pour gérer de manière optimale et sécurisée une grande variété de grandes charges de données et de technologies tout en simplifiant les opérations.

### DÉVOUÉ

Les instances dédiées, les réseaux et les disques connectés directement offrent des performances élevées.

### OPTIMISÉ

La configuration optimisée basée sur les formes de calcul spécifiques permet un temps de valeur rapide.

### GARANTI

Clusters Hadoop entièrement sécurisés et cryptés, avec une capacité supplémentaire à étendre la sécurité de la base de données Oracle à Hadoop. Comprehensive comprend l'écosystème Cloudera Hadoop (y compris Apache Spark) entièrement pris en charge. Logiciel Oracle inclus: intégration de données, distribution R, spatial et graphique.





### QUELS ENJEUX POUR LES ENTREPRISES DANS L'INTERNET DES OBJETS ET L'ANALYSE DES DONNÉES ?

Dans le domaine du Big Data et de l'internet des objets, la dynamique est incontestablement là : Orange gère 14 millions d'objets, contre 10 millions il y a un an. Nous traitons également 160 millions de données par minute versus 60 millions l'année dernière. C'est pour cela que, chez Orange Business Services, nous sommes présents à tous les stades de la vie des données, depuis le choix de l'objet connecté qui va collecter la donnée jusqu'à son traitement. Nous avons ainsi lancé en 2015 Datavenue, une offre IoT et data analytics à la carte pour accompagner les professionnels de bout-en-bout dans leurs projets. Nous les aidons à sélectionner les bons objets et les bons capteurs, qui vont ensuite acheminer la donnée sur un réseau sécurisé, adapté à leurs besoins et à leurs contraintes. Une fois récupérée, la donnée doit également être triée et traitée pour devenir une information utile : c'est pourquoi nous mettons à disposition de nos clients des plateformes de gestion et de traitement des données, simples d'utilisation et disponibles directement depuis les applications métiers.

C'est tout l'enjeu pour les entreprises : produire, à partir de la donnée brute, des informations utiles pour leur business, pour leur permettre d'améliorer l'expérience client, de faciliter le travail des équipes en interne, et ce bien évidemment en toute sécurité. 700 experts data et IoT travaillent chez Orange pour garantir cette expertise à nos clients, sans compter Orange Cyberdéfense, notre division spécialisée dans la protection des données et la sécurité numérique des entreprises.

### QUELS SONT LES SECTEURS CONCERNÉS ?

Tous les secteurs sont concernés, mais nous avons développé une expertise verticale dans 4 secteurs particuliers :

- Dans l'automobile, le déploiement des réseaux mobiles haut débit et la généralisation de la connectivité embarquée dans les véhicules neufs permettent de proposer de nombreux services auprès des constructeurs automobiles, des gestionnaires de flottes, des compagnies d'assurance et des clients finaux : télématique embarquée, mise à jour logicielles Over the Air, aide à la navigation, info-tainment... l'explosion des données produites par la connectivité embarquée prépare également le terrain pour l'arrivée prochaine des véhicules autonomes sur le marché.
- Dans l'industrie, les réseaux très basse consommation tels que Lora permettent, en remontant les données de façon fréquente et à moindre coût, de gagner en réactivité et d'optimiser les opérations de maintenance, par exemple grâce aux relevés automatiques de compteurs ou de réservoirs. Nous avons également la capacité tracer de façon plus efficace les flux logistique, d'approvisionnement, grâce à des capteurs positionnés en indoor ou outdoor en milieu urbain de façon beaucoup plus performante qu'avant, et donc d'aller chercher de l'information qui permette d'optimiser les processus opérationnels et de délivrer un meilleur service client.

Les progrès permis par l'IoT et la data analytics ne s'arrêtent pas aux portes de l'usine, puisque les produits, une fois sortis des entrepôts, se retrouvent ensuite en rayon dans les magasins et ali-

mentent notre quotidien. Ces mêmes magasins qui bénéficient par ailleurs également des progrès du big data et de l'IoT, que ce soit dans la gestion de la relation client, ou dans l'analyse des comportements des consommateurs.

Nous avons par exemple mis en place un système de gestion des files d'attente chez Leroy Merlin, afin de réduire l'attente en caisse et d'améliorer la satisfaction des clients du magasin. Mais les cas d'usages dans le commerce et la grande distribution sont variés.

- L'IoT et l'analyse de données ont également des conséquences directes dans la vie des citoyens: de plus en plus de projets de villes intelligentes se déploient dans le monde. Orange a d'ailleurs déjà opéré en France et au Moyen-Orient dans la ville de MSheireb. Grâce à l'analyse de données, nous sommes en mesure d'apporter au gestionnaire ou à la collectivité locale des outils et des informations qui lui permettent de gérer la ville plus efficacement, et d'améliorer le confort des habitants. Les données sont également disponibles directement auprès du citoyen, qui peut ainsi obtenir des informations en temps réel sur le menu de la cantine, la disponibilité d'un vélo en libre accès, ou acheter sa carte de transport à distance. Des outils précieux pour mieux vivre sa ville.

- Les objets de la sphère personnelle sont là aussi de plus en plus connectés : à la maison, on assiste à l'émergence de concepts ou d'objets qui peuvent avoir du sens, comme l'assistant personnel à qui on demande tout ce que l'on veut- ou presque, depuis le film que l'on veut voir ce soir à la liste de courses qu'il faut compléter. Djingo, qui est une véritable intelligence artificielle commercialisée par Orange, permet par exemple de passer un appel, de commander ses objets connectés, de demander une information et même d'être conseillé sur son compte Orange Bank. N'oublions pas non plus le secteur de la santé, où l'allongement de la durée de vie et l'augmentation des maladies chroniques sont un challenge pour les professionnels du secteur médical. Là encore, l'IoT et le Big Data permettent de faciliter la prise en charge des patients et d'améliorer son confort.

### EN CONCLUSION, QUEL IMPACT POUR LA STRATÉGIE DES ENTREPRISES ?

La relation client est transformée : elle se prolonge désormais dans chaque moment de vie du consommateur. L'enjeu pour les entreprises, c'est de se projeter vers « comment je prolonge ma relation client vers le quotidien du client ». Nous sommes bien évidemment présents pour accompagner les entreprises dans cette démarche.



Olivier Ondet / Directeur de programme Datavenue & IOT B2B  
Orange Business Services



# Transformer mes données en information utile.



Avec Datavenue, vous bénéficiez d'une offre complète et modulaire pour exploiter tout le potentiel des objets connectés et de l'analyse de données. Réinventez votre business.





CONTACT

## HISTORIQUE

Lancée en 2010 à l'initiative de Dassault Systèmes, Outscale fournit des services de Cloud Computing de Classe Entreprise, répondant aux exigences réglementaires et locales, aux entreprises souhaitant augmenter leur agilité business et déployer rapidement des business modèles à valeur ajoutée, en interne et pour leurs clients.

## CŒUR D'ACTIVITÉ

Présente en Europe, Amérique du Nord et Asie, Outscale met à disposition de ses clients un riche catalogue d'offres et services répondant aux problématiques Big Data des entreprises les plus agiles.

Outscale, propose le meilleur des technologies Big Data aux intégrateurs et délivre des plateformes prêtes à l'usage, pour lancer leurs projets Big Data en moins de 5 minutes. Intégrant la plateforme de données convergente de MapR, le fournisseur de Cloud apporte une solution clé en main, flexible et économique pour accompagner les entreprises vers l'ère du Big Data. Du déploiement de simples PoCs au prototype et à la mise en production des applications, les plateformes dédiées au Big Data d'Outscale allient performance, résilience et haute disponibilité.

## DATA PLATFORM AS A SERVICE – SOLUTION BIG DATA À LA DEMANDE

Réussissez vos projets clients Big Data avec Data Platform as a Service disponible en moins de 5 minutes.

Pour traiter et valoriser de très gros volumes de données numériques de leurs clients, Outscale met à disposition des intégrateurs sa plateforme spécialisée dans le déploiement d'infrastructures de projets Big Data : **Data Platform as a Service**. Avec Data Platform as a Service, basée sur les technologies Hadoop et MapR, Outscale permet aux intégrateurs de déployer leurs architectures Big Data les plus complexes sur des infrastructures souveraines. Flexible, élastique et performant, le Cloud Outscale est la solution clé en main pour les intégrateurs à la recherche d'un Cloud de classe entreprise et un accompagnement sur-mesure pour supporter le déploiement de leurs infrastructures de traitement de données.

Les données collectées et transformées sont hébergées en France dans notre solution de stockage optimisée pour garantir une haute performance et la sécurité de données.



L'offre Data Platform as a Service, basée sur MapR, permet aux intégrateurs d'accompagner leurs clients dans la création d'applications hautement distribuées et fortement scalables, grâce à la technologie MapR associée à la puissance du Cloud Outscale.

### Disponible en moins de 5 minutes !

Avec l'offre Data Platform as a Service, les intégrateurs démarrent leurs plateformes Big Data en choisissant la taille de cluster la mieux adaptée aux besoins de leurs clients. En moins de 5 minutes, la plateforme MapR dédiée est opérationnelle et les ressources sont disponibles.

### MapR, un Hadoop production-ready !

Afin de permettre au marché d'avoir confiance dans le Big Data, MapR s'est appuyé sur ses équipes de R&D pour transformer Hadoop en une plateforme prête pour la production. Avec MapR-FS, MapR-DB et MapR-Streams, les projets Big Data ne sont plus cantonnés à des simples PoCs, ils deviennent le cœur des stratégies de développement et de digitalisation des plus grands acteurs.

### DES SOLUTIONS FLEXIBLES ADAPTÉES À TOUT TYPE DE BESOIN

Deux types de configuration de Data Platform as a Service sont disponibles pour choisir le cluster le mieux adapté à chaque projet Big Data :

#### DataLab

- 1 noeud avec 16 cœurs et 64 Go de RAM
- Stockage utile : 100 Go

Plateforme dédiée aux Data Scientists pour affiner leurs algorithmes et parfaire leurs compétences.

#### Enterprise

- 5 noeuds avec 16 cœurs et 64 Go de RAM par noeud
- Stockage utile : 2 To

Plateforme prête pour la production associant performance, haute-disponibilité et forte résilience.

### BÉNÉFICES DE DATA PLATFORM AS A SERVICE



**Facilité et rapidité de mise en place**  
Un cluster Hadoop correspondant à vos besoins en moins de 5 minutes



**Facturation de licence à l'heure**  
Un modèle de facturation agile vous donnant accès au support MapR



**Sécurité des données**  
Un Cloud Provider respectant la protection et la sécurité des données en France

En savoir plus sur l'offre Data Platform as a Service : [fr.outscale.com/dataplatform-as-a-service](http://fr.outscale.com/dataplatform-as-a-service)



Réussissez vos projets clients Big Data avec Data Platform as a Service, disponible en moins de 5 minutes !





# Scaled for more than just #BigData

## 8 REASONS TO CHOOSE OVH



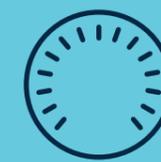
NO. 1 PUBLIC CLOUD PROVIDER  
IN EUROPE, 2<sup>ND</sup> IN NORTH AMERICA\*



OVER 1.2 MILLION  
CUSTOMERS WORLDWIDE



22 DATACENTRES ACROSS  
7 COUNTRIES & 4 CONTINENTS



11+ TBPS GLOBAL  
NETWORK CAPACITY



PRESENCE  
IN 19 COUNTRIES



270,000 PHYSICAL  
SERVERS



€1.5 BILLION INVESTMENT  
OVER 5 YEARS



32 POINTS OF PRESENCE  
WORLDWIDE

### PRÉDIRE L'ISSUE D'UNE ACTION EN JUSTICE GRÂCE AU BIG DATA

Fondée par deux élèves avocats technophiles et deux ingénieurs en informatique, la startup française Predictice exploite les techniques du machine learning pour prédire les chances de succès d'une procédure judiciaire et optimiser la stratégie contentieuse des avocats, en s'appuyant notamment sur les décisions de justice antérieures.

Zoom sur ce projet soutenu par OVH, qui convainc de plus en plus de cabinets d'avocats et services juridiques d'entreprise, et vaut à Predictice d'être considéré comme la startup phare de la legaltech française.

### DE BONNES DÉCISIONS DE JUSTICE GRÂCE AU TRAITEMENT MASSIF DES DONNÉES

Dans le cadre d'une politique globale en faveur de l'open data, depuis quelques années environ les juridictions sont incitées à mettre à disposition du public les décisions judiciaires devenues définitives, dans le respect de la vie privée des personnes concernées.

La base de données de Predictice comprend aujourd'hui un million de lignes, sans compter les différents codes (les articles de loi) et les textes de doctrine (les commentaires de codes et de décisions judiciaires par les universitaires) sur lesquels s'appuie l'entreprise et qu'il lui faut traiter.

Le droit utilise un langage très normé, ce qui représente ici un atout ; il y a peu de synonymes et les tournures de phrase utilisées pour rendre un jugement à Lille ou à Marseille sont similaires. En conséquence, il est possible d'indexer et d'enrichir les données de façon automatisée, en y ajoutant des métadonnées.

Ces métadonnées rendent compte des caractéristiques du litige : quelle a été l'indemnité réclamée et celle effectivement obtenue ? S'agissait-il d'un recours ou du jugement de première instance ? Et bien d'autres éléments encore. Cela permet à Predictice de proposer à ses utilisateurs un moteur de recherche capable de fournir toutes les informations utiles à propos d'un litige (textes de loi, jurisprudences, doctrine...).

### COMMENT ANALYSER CETTE MASSE DE DONNÉES ?

À partir de cette base documentaire, des algorithmes entrent en jeu. Deux litiges n'étant jamais identiques, l'objectif est d'identifier la corrélation entre un facteur ou une combinaison de facteurs (analyse multivariée) et la résolution d'une affaire. Pour cela, Predictice exploite l'algorithme SyntaxNet, développé par Google et rendu open source le 12 mai 2016. Cet outil d'analyse aide les machines à comprendre le langage humain et permet, dans le cas de l'entreprise, de repérer les dépendances entre les mots pour en extraire du sens. Le texte ainsi enrichi est soumis à des algorithmes de classification/régression (Vapnik's SVM) et des règles d'association pour créer des modèles prédictifs complexes. En appliquant ces modèles aux caractéristiques de son litige, l'avocat est ainsi capable d'évaluer les probabilités de succès.

Viennent ensuite l'évaluation et la comparaison de plusieurs stratégies contentieuses pour que l'avocat puisse construire, en fonction des caractéristiques variables de l'affaire, l'argumentation qui a statistiquement le plus de chances de réussir.

**PREDICTICE S'APPUIE SUR LES SOLUTIONS OVH**

Les fondateurs de Predictice ont choisi dès le départ d'appuyer leur activité sur un serveur dédié OVH virtualisé avec un hyperviseur ESXi de VMware pour isoler leurs différents services (frontaux web, bases de données MySQL et Elasticsearch, back office, monitoring) au sein de plusieurs machines virtuelles.

Predictice a rejoint le Digital Launch Pad, le programme d'accompagnement de startups d'OVH, pour bénéficier de conseils approfondis sur l'infrastructure et accéder aux serveurs GPU (processeurs graphiques permettant de réaliser des traitements massivement parallèles) parfaitement adaptées aux traitements de type deep learning. Cela permet d'améliorer la vitesse d'apprentissage des algorithmes de Predictice en déployant une architecture CUDA (Computer Unified Device Architecture) utilisant un processeur graphique pour effectuer les calculs à la place du CPU.

### LES PRINCIPAUX DÉFIS À RELEVER POUR L'ENTREPRISE

Le premier défi à relever tient à la « scalabilité » de l'infrastructure, qui doit supporter une croissance rapide du nombre d'utilisateurs, ainsi qu'une charge importante car l'usage de l'application est intensif. Lorsqu'un avocat travaille sur un dossier, il va effectuer des recherches pendant parfois plus d'une heure en lançant des requêtes gourmandes en ressources.

Grâce à la virtualisation du vRack (réseau privé d'OVH), il sera par la suite assez facile de répartir les machines virtuelles sur plusieurs serveurs dédiés interconnectés pour augmenter la capacité de traitement et constituer une architecture n-tiers plus résiliente.

En savoir plus : <https://predictice.com> et [OVH.com](https://www.ovh.com)





**HITACHI**  
Inspire the Next



### AVEC PENTAHO, LA SACEM A POUR OBJECTIF D'OFFRIR À SES UTILISATEURS MÉTIERS DES SERVICES D'ANALYSE DE BIG DATA À LA DEMANDE

#### A PROPOS DE LA SACEM

La Société des auteurs, compositeurs et éditeurs de musique (Sacem) a pour vocation de représenter et défendre les intérêts de ses membres en vue de promouvoir la création musicale sous toutes ses formes (de la musique contemporaine au jazz, rap, hip-hop, chanson française, musique de films, musique à l'image, œuvres dédiées au jeune public...) mais également d'autres répertoires (humour, poésie, doublage-sous-titrage...)

Sa mission essentielle est de collecter les droits d'auteur et de les répartir aux auteurs, compositeurs et éditeurs dont les œuvres ont été diffusées ou reproduites. Organisme privé, la Sacem est une société à but non lucratif gérée par les créateurs et les éditeurs de musique qui composent son Conseil d'administration.

Elle compte 161 170 membres dont 19 650 créateurs étrangers issus de 166 nationalités (4 206 nouveaux membres en 2016) et représente plus de 118 millions d'œuvres du répertoire mondial. La Sacem compte 500 000 clients qui diffusent de la musique en public : commerces, bars, discothèques, organisateurs de concerts ou d'événements en musique.

#### CHALLENGES

L'écoute musicale massive sur internet et la globalisation ont complexifié le processus de la gestion collective des œuvres musicales par la Sacem et ses partenaires à l'étranger. Il devient alors indispensable d'optimiser ce processus, afin d'éviter la répétition des tâches et de faciliter la collecte de droits. Pour garder une compréhension d'un marché en constante évolution, obtenir une corrélation juste entre la collecte et la répartition des droits, la Sacem a besoin d'analyses de plus en plus précises sur des volumes de données de plus en plus grands. À l'issue de la collecte de tous les droits qu'elle effectue, la Sacem cherche à réduire les délais de traitement, et par conséquent à accélérer le paiement aux artistes, en simplifiant les processus de consolidation. L'ensemble de ces problématiques nécessite un outil d'intégration et d'analyse de données dans un environnement Big data existant.

#### LA SOLUTION

La Sacem a choisi la plateforme Pentaho Data Integration and Business Analytics qui assure à la fois l'intégration et l'analyse de son environnement Big Data. La technologie ouverte de Pentaho permet d'accéder à d'importants volumes de données hétérogènes et sa facilité d'utilisation donne aux utilisateurs métier l'autonomie nécessaire pour faire leurs propres analyses. Pentaho peut être intégré aux applications métier opérationnelles telles qu'elles existent, ce qui permet de gagner en délais de mise en œuvre et de réduire les coûts d'intégration des analyses au sein des applications métiers.

« L'ouverture de l'outil pour l'utilisation d'algorithmes de statistiques avancés est un atout différenciant. Il permet de les embarquer tels qu'on les a pensés en phase de projet exploratoire et de

minimiser les coûts lorsque l'on souhaite les industrialiser. » se réjouit Frédéric Falkoff, Directeur Web, Data et Ressources à la Sacem.

Pentaho fournit une cartographie macro regroupant en entrée les données de collecte, en sortie celles des feuillets de répartition des droits d'auteur, et entre les deux des référentiels par type d'utilisation, par famille d'œuvres et de répertoires. À titre d'exemple, on répertorie environ 5,1 milliards de lignes associées à un provider tel que Youtube pour un an. Et 1,5 millions de lignes représentent les données de répartition des droits pour un an (1,4 Milliard de lignes pour dix ans de droits), y compris ceux que la Sacem se charge de collecter et répartir pour les répertoires des membres étrangers exploités en France. La plateforme permet aussi de croiser les données des 161 170 membres de la Sacem avec les œuvres représentées par la Sacem en perspective d'analyses d'usage sur lesquelles repose un nouveau service offert aux sociétaires sur le portail internet.

« Avec Pentaho, nous avons comme objectif de mettre en place un dispositif de données à la demande selon le périmètre et la profondeur d'historique souhaités, afin de rendre nos utilisateurs autonomes dans leur exploration et croisement de diverses sources Big Data, désormais prêtes pour des analyses détaillées et la consolidation automatisée de la collecte et redistribution des droits d'auteur, » déclare F. Falkoff

#### LES BÉNÉFICES

Pentaho crée le lien qu'il manquait pour consolider un environnement Big Data hétérogène, puis se connecter et interroger aisément des bases de données telles que Cassandra, MariaDB, Oracle, MySQL, Vertica ou Elasticsearch. Sa simplification et automatisation des processus accélèrent les analyses sur la collecte et la préparation de diverses sources de données de façon exhaustive et non pas sous forme d'échantillons. Sa gestion automatisée des droits d'authentification donne aux utilisateurs métier en fonction de leur profil un accès sécurisé à une diversité de données qu'ils peuvent désormais croiser à la volée une fois qu'elles sont à disposition sur un portail externe.

Sans besoin de déployer un client lourd sur les postes utilisateurs les données sont disponibles en « libre-service » avec une gestion des droits d'accès pour des analyses détaillées, telles que la répartition des droits d'auteur en cohérence avec les analyses d'usage. Les utilisateurs métier accèdent et croisent à la volée des volumes de données plus importants. Ils bénéficient dans les meilleurs délais de mises en forme graphiques ou de tableaux de bord clairs et automatisés. Les utilisateurs gagnent ainsi en autonomie et la DSI peut se recentrer sur son cœur de métier.

« Il est simple d'embarquer Pentaho au sein d'une application opérationnelle, puis de naviguer dans un environnement analytique sans devoir changer d'outil et se reconnecter. Dans un futur proche, ce dispositif permettra également de mettre des tableaux de bord à disposition sur la tablette des 700 collaborateurs Sacem, présents en région, » conclut F. Falkoff.

### QUEL CONSTAT FAITES-VOUS DE L'ÉVOLUTION DE L'ÉCOSYSTÈME BIG DATA ?

Des niveaux d'innovation très élevés, par exemple, un flux constant de nouveaux projets Apache sur Big Data combinés avec aucun acteur ne possédant une partie significative de l'écosystème. Bien que cela signifie une fragmentation, il existe des normes et des protocoles avec des API complémentaires qui génèrent de la valeur. L'intégration de Pentaho Data Integration consiste à supprimer certaines des frontières les plus difficiles entre différentes technologies. De cette façon, vous pouvez protéger vos investissements et vos compétences existants tout en profitant des nouvelles idées au fur et à mesure.

### QUELS SONT, SELON VOUS, LES NOUVEAUX CHALLENGES TECHNOLOGIQUES ET DE GOUVERNANCE DU BIG DATA ?

Pour la technologie, le principal défi est de suivre le rythme du changement. Notre couche d'exécution adaptative (AEL) permet à un utilisateur de créer une tâche ETL en une seule fois, dans une interface graphique auto-documentée et de l'appliquer à un moteur d'exécution de son choix, Apache Spark ou autre, protégeant ainsi l'utilisateur des changements technologiques. Pour la gouvernance, la capacité à calculer la dispersion des données pour identifier les master data, la maintenance d'un modèle de sécurité de la source à la cible pour les données combinées contribuent toutes à la gouvernance. Les compétences sont l'un des enjeux clés du marché du Big Data. L'AEL permet d'économiser beaucoup de réécriture de code, et l'utilisation de l'interface graphique de PDI contribue fortement à ne pas avoir à écrire des map reduce scripts pour faire des travaux ETL.

### DU PRÉDICTIF AU PRESCRIPTIF : UNE RÉALITÉ BUSINESS ?

De l'origine des données à l'analyse et à l'analyse prédictive, Pentaho associe étroitement l'intégration de données à l'analytique dans une solution Big Data continue pour éliminer la complexité et réduire le temps nécessaire à la création de valeur. Au-delà de la visualisation interactive et de l'exploration de données, Pentaho propose des algorithmes de machine learning et des outils de traitement de données performants et à la pointe de la technologie. Les spécialistes des données et les analystes peuvent découvrir des modèles significatifs et des corrélations autrement cachées avec l'analyse et les rapports standard. Des analyses sophistiquées et avancées, telles que la prévision de séries chronologiques, aident à planifier les résultats futurs en fonction d'une meilleure compréhension de l'histoire antérieure.

### QUELLE EST LA MISSION DE PENTAHO SUR LE MARCHÉ BIG DATA ?

Au sein d'une plate-forme unique, notre solution fournit des outils Big Data pour extraire, préparer et fusionner vos données, ainsi que des visualisations et des analyses qui changeront la façon dont vous dirigez votre entreprise. De Hadoop et Spark à NoSQL, Pentaho vous permet de transformer rapidement vos Big Data en connaissances.

### QUELLES SONT LES NOUVEAUTÉS DE VOS OFFRES DATA ?

- Facilité d'utilisation - pour les ingénieurs de données, les spécialistes des données et les analystes de données
- Data Agnostic - intégrer n'importe quelle source de données
- 100 des étapes prêtes à l'emploi avec la modélisation et la visualisation
- Automatisation pilotée par les métadonnées - intégration de données évolutives, lignage
- Extensibilité linéaire - intégration native de Hadoop, résultats de référence
- Machine Learning - apportez votre propre outil de modélisation prédictive et parcourez facilement
- Data Delivery - livrer des données physiquement ou virtuellement
- Extensibilité - ouverte, architecture pluggable
- Full Suite Analytics - tableaux de bord, rapports, analyses interactives
- Mise à disposition de Cloud et échelle élastique dans le cloud

### DANS LA DERNIÈRE ÉDITION DU GUIDE DU BIG DATA, L'IOT ÉTAIT AU CŒUR DE VOTRE BUSINESS ET STRATÉGIE, EST-CE TOUJOURS LE CAS ?

Tout à fait, comme l'augmentation des volumes, de la variété et de la vitesse des données provenant des capteurs, des machines et d'autres périphériques connectés, entraînent d'importants problèmes d'intégration des données. Pentaho aide les organisations à préparer et fusionner facilement les données des machines et des capteurs avec d'autres actifs, tels que les données de vos systèmes ERP ou CRM, afin de fournir le contexte nécessaire pour générer des résultats commerciaux transformateurs.

### LA DERNIÈRE VERSION DE PENTAHO EST COMPATIBLE À 100% AVEC SPARK, POURQUOI CE CHOIX ?

Pour un traitement encore plus rapide du processus ETL des Big Data. L'avantage potentiel pour les concepteurs ETL est la possibilité de concevoir, tester et régler les travaux ETL dans l'environnement de conception graphique facile à utiliser de PDI, puis de les exécuter à grande échelle sur Spark. Cela réduit considérablement les compétences requises, augmente la productivité et réduit les coûts de maintenance lorsque vous utilisez Spark pour l'intégration Big Data. En intégrant également Pentaho Data Integration à Spark et MLLib, les Data Scientists bénéficieront d'un environnement facile à utiliser (PDI) pour préparer les données à utiliser dans les solutions MLLib. De plus, cette intégration facilitera l'opérationnalisation du travail de l'équipe Data Science en orchestrant l'intégralité du flux de bout en bout, de l'acquisition des données à la préparation des données en passant par l'exécution des tâches MLLib et le partage des résultats, dans un flux de travail PDI simple. AEL mettra votre application à l'épreuve des moteurs émergents. Aujourd'hui, vous pouvez développer visuellement une application pour le moteur Kettle de Pentaho et un cluster Spark. Au fur et à mesure que de nouvelles technologies émergent, l'AEL sera implémenté pour d'autres moteurs afin que les développeurs n'aient pas besoin de réécrire leur application. En résumé, Pentaho Data Integration avec Apache Spark permet aux utilisateurs du moteur de traitement open source d'atteindre plus d'utilisateurs en simplifiant l'orchestration des jobs Spark. Le support natif de PDI pour Spark fournit une interface intuitive pour exécuter des tâches de traitement de données à haute vitesse.



## LES SOLUTIONS SECTORIELLES BIG DATA DE PITNEY BOWES SOFTWARE

Quels que soient les secteurs, les entreprises obtiennent plus rapidement que jamais les réponses dont elles ont besoin. En effet exécuter un logiciel de Big Data de façon native dans Hadoop ou Spark leur permet d'analyser leurs données avec rapidité et précision. Mais pour donner de la valeur à cette analyse, il leur faut intégrer ces informations à leurs flux de travail et processus métier quotidiens. Nos alliances technologiques avec Cloudera, Hortonworks et Amazon Web Services permettent de prendre en charge les environnements sur site, mais aussi SaaS et IaaS dans le Cloud. L'ajout de données géospatiales leur permet de relier un large éventail de données, puis de les agréger et les analyser en vue d'obtenir des informations précieuses pour leurs activités spécifiques. En combinant de puissants SDK à leurs technologies de distribution préférées, ces sociétés sont assurées que leur solution de données spatiales fonctionnera de manière fiable, que ce soit sur site ou dans le Cloud.

Pitney Bowes Software accompagne de nombreuses entreprises pour des cas d'usage extrêmement variés selon les secteurs, nous pouvons notamment souligner :

### POUR L'ASSURANCE & LES SERVICES FINANCIERS

**Vue unique des risques :** Combinez d'énormes volumes de données à des vitesses incroyables afin de créer une vue consolidée des risques assurés potentiels associés à n'importe quelle propriété. Établissez un aperçu de l'exposition et des risques s'appliquant à l'ensemble de votre portefeuille (polices multiples) afin d'équilibrer correctement vos offres et vos tarifs.

**Tarifification selon l'usage et le comportement :** Collectez d'énormes volumes de données télématiques concernant le mode de conduite des automobilistes (vitesse, moment de la journée, jour de la semaine, tendances de freinage, lieux, etc.) pour une prévision des risques et une tarification plus précises.

**Notification des sinistres en temps réel :** Mesurez l'impact d'un sinistre en temps réel, assignez les experts, fournissez un soutien rapide à vos adhérents et comprenez les pertes potentielles.

**Conformité et criminalité financière :** Résolution d'entités pour la lutte contre le blanchiment d'argent, la vérification d'identité (KYC, KYCC, CDD) et le contrôle des sanctions. Garantissez la qualité des données traitées par les systèmes de surveillance des transactions (TMS). Identifiez, reliez et visualisez des relations complexes (non évidentes).

**Évaluation hypothécaire :** Grâce à un grand nombre d'attributs de propriété, les prêteurs peuvent estimer précisément la véritable valeur d'une propriété.

**Opportunités de vente croisée/incitative :** Avec une vue unique et centralisée des clients, vous pouvez obtenir et partager des informations précises sur les comptes afin de créer des opportunités de vente croisée et incitative.

**Analyse de l'emplacement et la performance des agences :** Combinez intelligence géodécisionnelle et données démographiques pour visualiser et analyser le véritable potentiel de vente de diverses agences. Comparez rapidement la performance des agences et définissez des objectifs réalistes afin d'orienter la prestation de services adéquats aux bons emplacements.

### POUR LES OPÉRATEURS DE TÉLÉCOMMUNICATION

**Ce qui prenait des mois prend à présent quelques minutes :**

Un fournisseur de services de communication aux États-Unis était confronté à l'opinion publique selon laquelle son réseau n'était pas aussi fiable que celui de ses concurrents. Afin de changer cette perception, il a voulu élaborer une carte de couverture du réseau détaillant l'état du service dans les petites régions des États-Unis.

**Les jeux de données étaient incroyablement vastes. Malgré tout, à l'aide d'une solution de traitement géospatial s'exécutant de façon native dans Hadoop, ce fournisseur a agrégé pas moins de 19 milliards d'enregistrements téléphoniques (dans 950 millions de polygones) en seulement 31 minutes.**

**Optimisation du réseau :** Comparez les performances prévues et les interactions réseau collectées à partir de milliards de connexions des adhérents. Assurez la planification de la capacité du réseau et une prestation de service cohérente, ainsi qu'un meilleur ciblage et un meilleur service aux clients de grande valeur.

**Monétisation des données :** Organisez et anonymisez les données de déplacement de vos adhérents pour créer de nouveaux modèles commerciaux et services numériques. Établissez des partenariats rentables dans une chaîne de valeur complexe et assurez la satisfaction de vos clients.

**Interactions durant le cycle de vie des adhérents :** Apprenez à cerner les déplacements et l'utilisation du réseau de chaque abonné pour quantifier la valeur client à vie et orchestrer des communications omnicanales cohérentes, contextuelles et pertinentes. Prévenez l'attrition et identifiez la prochaine action à entreprendre en temps réel.

**Acquisition de clientèle :** Créez des cartes de couverture participatives afin de mettre en avant la performance du réseau et acquérir de nouveaux adhérents.



## LE BIG DATA TOUT LE MONDE EN PARLE, MAIS QUELS SONT LES PREMIERS ENJEUX SUR LE SUJET ?

Exécuter un logiciel de Big Data de façon native dans Hadoop ou Spark permet à nos clients d'analyser leurs données avec rapidité et précision. Mais pour donner de la valeur à cette analyse, il leur faut intégrer ces informations à leurs flux de travail et processus métier quotidiens. Nos alliances technologiques avec Cloudera, Hortonworks et Amazon Web Services permettent de prendre en charge les environnements sur site, mais aussi SaaS et IaaS dans le Cloud. L'ajout de données géospatiales leur permet de relier un large éventail de données, puis de les agréger et les analyser en vue d'obtenir des informations précieuses pour leurs activités spécifiques.

On peut notamment penser à trois corollaires du Big Data :

**Accélérez les flux de travail.** Les processus actuels ralentissent et surchargent parfois les systèmes et l'infrastructure informatiques. En allégeant ces processus avec des flux de Big Data plus rapides, les entreprises observent une amélioration considérable de leurs performances.

**Organisez et reliez divers types de données.** Pour une meilleure compréhension, nombre d'organisations collectent les données dans divers formats et sur plusieurs systèmes, et fusionnent les données correspondant à des cas d'utilisation similaires. Pitney Bowes permet aux entreprises de relier des données spatiales, structurées et non structurées dans des formats et des systèmes communs.

**Tirez parti des informations issues des capteurs.** L'Internet des Objets (IdO), en croissance exponentielle, aide les entreprises à mieux comprendre la performance de leurs équipements, de leur personnel et de leurs clients. Grâce à nos solutions, les entreprises peuvent contextualiser les énormes volumes de données générées par les capteurs pour obtenir une vision détaillée.

## QUELS SONT LES ENJEUX DE L'UTILISATION DE SDK DANS L'ENVIRONNEMENT BIG DATA ?

Les SDK sont avant tout la source de résultats rapides. En accédant aux informations offertes par les Big Data, les nouveaux kits de développement logiciel (SDK) sont capables de transformer les entreprises. Les solutions avancées actuelles s'exécutent de façon native dans des environnements de Big Data, tels que Hadoop ou Spark, ce qui permet d'analyser d'énormes volumes de données disparates afin d'obtenir des résultats d'une précision exceptionnelle. Les entreprises peuvent également effectuer des opérations de géocodage, d'analyse spatiale et de gestion de la qualité des adresses en un temps record. En combinant de puissants SDK à leurs technologies de distribution préférées, nos clients sont assurés que leur solution de données spatiales fonctionnera de manière fiable, que ce soit sur site ou dans le Cloud. Les SDK pour les Big Data de Pitney Bowes sont : simples (exécution des processus de façon native dans Hadoop et Spark), puissants (gestion de données toujours plus complexes), rapides (traitement d'énormes volumes de données en un temps record.), flexibles (exécution sur site ou dans le Cloud) et évolutifs (modularité du système selon vos besoins).

## POURQUOI MISER GROS SUR LE BIG DATA ?

La gestion des informations est un défi pour les entreprises de toutes tailles. La plupart manquent de puissance pour extraire de la valeur des données géospatiales à la vitesse requise ou à l'échelle appropriée. En effet, les outils utilisés par certaines entreprises peuvent nécessiter plusieurs jours pour traiter les données nécessaires afin de parvenir à des conclusions exploitables. Celles-ci s'abstiennent de résoudre les problèmes liés aux Big Data pour plusieurs raisons. Certaines les considèrent « trop volumineuses » ou « trop difficiles » à maîtriser. D'autres doivent combattre une culture hostile au changement. Il y a aussi celles qui doutent des promesses des Big Data en matière de rapidité, de sécurité et de développement des activités. Cependant de nouvelles solutions balaient les réticences des plus sceptiques, en changeant la dynamique et en créant des opportunités incroyables pour booster les résultats et générer des informations pertinentes de manière ultra-rapide au sein de votre infrastructure. Exploiter les Big Data à l'aide des logiciels adéquats peut offrir à nos clients des opportunités dont ils n'auraient pas pu bénéficier avec leurs solutions existantes. Avec les bons outils, ils peuvent approfondir leur connaissance des personnes, lieux et objets afin de découvrir des relations dissimulées et de les convertir en opportunités de croissance. Une approche modulaire leur donne la possibilité d'étendre leurs solutions pour répondre à l'évolution de leurs besoins. Ils peuvent ainsi intégrer des fonctions de qualité des données, des analyses de type « point dans un polygone » ou le calcul d'isochrones à leurs processus de Big Data natifs. Fort d'une telle flexibilité, ils ne sont plus limités à une simple solution ponctuelle de géocodage ou de validation des adresses. Désormais, les entreprises peuvent enfin obtenir les résultats qu'elles attendaient et rentabiliser leurs investissements dans le Big Data.



## INTRODUCTION

L'entreprise eXplorance est un leader mondial de solutions de Learning Experience Management pour la gestion des entreprises et des établissements d'enseignement supérieur. eXplorance sert plus de 9 millions d'utilisateurs et de clients à l'échelle internationale. Sa plateforme d'évaluation collecte de grandes quantités de données au moyen de questionnaires, d'enquêtes et d'outils d'évaluation à 360°. Les données ainsi collectées aident les organisations à mesurer les besoins, les attentes, les aptitudes et les compétences des employés et facilitent l'implantation d'un cycle d'amélioration continue.

## DÉFI

Les données massives sont omniprésentes et déstabilisent les entreprises. Maîtriser les flots de données et les traiter convenablement pour en extraire les informations pertinentes constitue un véritable casse-tête. Autant dans une université que dans une entreprise, le challenge consiste à transformer cette immense quantité de données brutes générées quotidiennement en avantage compétitif. Dans la plupart des universités, les étudiants sont invités à la fin de chaque trimestre à produire des évaluations sur la prestation de l'enseignement. Les questionnaires d'évaluation contiennent généralement des questions fermées, des scores et des questions ouvertes pour permettre à l'étudiant de s'exprimer librement. Les questionnaires d'établissement d'enseignement ont besoin d'analyser rapidement cette quantité considérable de données pour évaluer les compétences pédagogiques des professeurs et chargés de cours, et mesurer l'impact et l'efficacité des programmes dans le but d'améliorer la qualité des services et de favoriser l'engagement et la réussite. Bien que très performant et offrant une grande variété d'outils de mesure, le système de collecte de données mis au point par eXplorance ne permet pas d'analyser les commentaires textuels, pourtant si riches en informations. En effet, seules les données provenant des questions fermées et les scores se présentent sous une forme immédiatement exploitable. L'ensemble des informations non structurées sous forme textuelle nécessite un traitement préalable avant d'être analysé. L'usage des méthodes traditionnelles d'analyse de données non structurées devient impossible devant la taille considérable des données récoltées. Pour les clients d'eXplorance confrontés à ce défi, une stratégie efficace d'analyse de ce type de données est essentielle.

## PROJET

Pour garder sa position de leader et répondre au besoin de ses clients, eXplorance a demandé la collaboration de Provalis Research afin de mettre au point une nouvelle plateforme analytique tout-en-un, capable d'automatiser l'analyse de données non structurées collectées à partir des formulaires d'évaluation et de faciliter leur interprétation. La stratégie proposée par Provalis Research a consisté, dans un premier temps, à utiliser le logiciel d'analyse sémantique WordStat pour faire une recherche exploratoire de contenu, en utilisant des techniques de traitement automatique des langues (TAL), de statistiques textuelles et d'apprentissage machine pour extraire rapidement les principaux thèmes et catégories de mots. Pour ce faire, les experts de Provalis Research ont analysé plus de 1,8 million d'évaluations de cours provenant d'étudiants à travers le monde et construit un corpus de 35 millions de mots. La complexité et l'hétérogénéité des données analysées ont nécessité l'utilisation de

techniques de désambiguïsation, de mots-clés en contexte (KWIC) et de correction orthographique. Toutefois, ces opérations ont dû être effectuées avec prudence, étant donné le caractère international des données et les spécificités linguistiques régionales. Par exemple, en Nouvelle-Zélande, le mot « paper » réfère à un cours, alors qu'il peut désigner une épreuve écrite, un document, un article, un journal ou un exposé dans les autres pays anglo-saxons. Par la suite, les fonctionnalités d'aide à la construction de dictionnaires de WordStat ont été mises à contribution pour développer et valider une taxonomie customisée contenant des dizaines de milliers de mots, d'expressions et de règles réparties dans une centaine de catégories. Enfin, la trousse de développement logiciel (SDK) de Provalis Research a été utilisée pour créer Blue Text Analytics (BTA) et traiter automatiquement les questions ouvertes.

## TÉMOIGNAGE

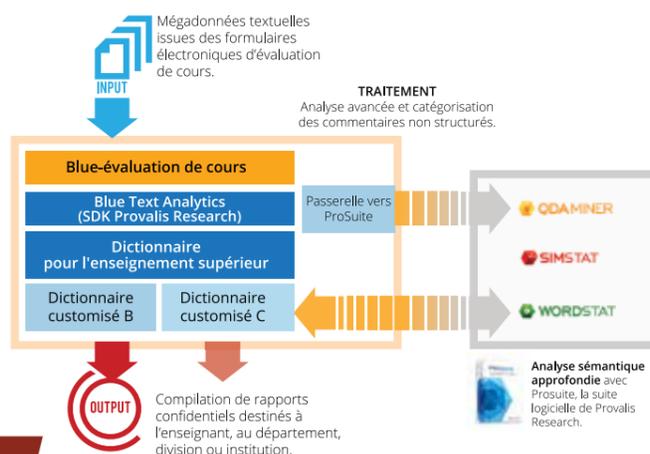
« Provalis a une approche rigoureuse en analyse sémantique [...] L'approche d'analyse de texte par les méthodes mixtes de Provalis a permis une meilleure compréhension de données recueillies par notre système. » eXplorance a des clients partout dans le monde et « l'outil d'analyse sémantique développé par Provalis prend en considération les spécificités linguistiques régionales, le contexte et les erreurs d'orthographe » a déclaré Samer Saab, le fondateur et PDG d'eXplorance.

## CONCLUSION

L'approche proposée par Provalis Research a permis à eXplorance d'implanter Blue Text Analytics (BTA) avec succès. Grâce à cette plateforme analytique tout à fait unique sur le marché, l'entreprise dispose désormais d'un avantage compétitif difficilement imitable. En combinant les techniques de traitement automatique des langues, les statistiques textuelles et l'apprentissage machine, les logiciels d'analyse sémantique de Provalis Research permettent d'extraire rapidement de l'information pertinente à partir d'un grand volume de documents. Ainsi, il suffit d'une fraction de temps pour identifier et extraire les principaux thèmes, des tendances, des opinions ou des sentiments.

**Pour obtenir plus d'informations sur nos logiciels ou pour voir une Démo Web sur vos propres données, contactez-nous à : [info@provalisresearch.com](mailto:info@provalisresearch.com)**

Tableau 1: Diagramme de Blue Text Analytics



# FAITES PARLER VOS DATA TEXTUELLES

Grâce aux techniques d'apprentissage machine et de traitement automatique des langues, vous pouvez extraire automatiquement les sujets les plus saillants, explorer rapidement les relations entre mots ou concepts et catégoriser automatiquement vos données textuelles. Nos solutions d'analyse sémantique vous permettent également d'explorer les relations entre les données structurées et non structurées.



Pour une Web-démo ou pour plus d'information sur nos produits, contactez-nous au (+1) 514 899-1672 ou à [sales@provalisresearch.com](mailto:sales@provalisresearch.com), [provalisresearch.com](http://provalisresearch.com)



**DÉCOUVREZ TOUTE L'HISTOIRE QUI VIT DANS VOS DONNÉES**

A travers une approche unique, Qlik répond aux besoins croissants des utilisateurs et des entreprises : du reporting à l'analyse visuelle en libre-service, en passant par l'analyse guidée, intégrée et personnalisée des données. En 2017, Qlik se classe pour la 7e année consécutive parmi les leaders du Magic Quadrant de Gartner pour les plateformes de BI & Analytics.

La technologie associative innovante de Qlik libère des contraintes des outils basés sur les requêtes. Le moteur associatif Qlik associe toutes les données et permet de les explorer librement quel que soit l'axe d'analyse, et de façon exhaustive. C'est ce qui fait l'Associative Difference™. Grâce à cette vue complète de leur activité, les entreprises peuvent mieux planifier, mieux gérer la complexité et définir des plans d'actions efficaces.

**RICARD MET L'ANALYSE VISUELLE AU CŒUR DE LA CONNAISSANCE CLIENT AVEC QLIK**

Les sociétés Pernod et Ricard utilisent Qlik depuis plusieurs années pour améliorer le pilotage opérationnel, dans les domaines des ventes et de la supply chain notamment. Début 2017, l'équipe BI de la DSI de Ricard et Pernod a souhaité mettre en place un nouveau projet Qlik Sense pour améliorer la connaissance clients en analysant les gros volumes de données de son site « Place Ricard », le programme de fidélité de la marque.

Pour Ricard, la puissance d'analyse visuelle de Qlik Sense facilite la communication avec un public de non-statisticiens, en leur permettant de visualiser les modèles et les tendances. C'est ce qui permet de faire adhérer les collaborateurs, grâce à une meilleure compréhension de concepts qui sont parfois complexes.

**ALSTOM UTILISE QLIK DANS TOUTES LES ÉTAPES DU CYCLE DE VIE DE SES TRAINS**

Alstom utilise Qlik dans tout le cycle de vie des trains, depuis la réponse aux appels d'offres jusqu'à la maintenance, en passant par la conception et la fabrication. L'entreprise analyse toutes les

données techniques remontées par des capteurs pour analyser la vie du train et optimiser la maintenance. Ce type d'indicateurs apporte une forte valeur ajoutée et peut être commercialisé ensuite aux partenaires, clients et fournisseurs d'Alstom. Après avoir développé un grand nombre d'applications sur QlikView, l'entreprise a choisi de développer de nouvelles applications avec Qlik Sense, car la solution apporte des avancées en termes de mobilité, avec la possibilité, grâce au responsive design, de créer une seule application qui peut être affichée aussi bien sur un grand écran tactile dans une usine, que sur le smartphone utilisé par les managers. Qlik Sense apporte également une réponse aux utilisateurs qui souhaitent créer eux-mêmes leurs applications en libre-service.

**REXEL ALLIE PUISSANCE ANALYTIQUE ET INTUITIVITÉ POUR TRAITER DE TRÈS GRANDS VOLUMES DE DONNÉES**

Les analyses prédictives font partie du processus décisionnel de Rexel dans de nombreux domaines (optimisation des stocks, des tarifs, anticipation du churn...) mais elles ne seraient pas utiles si les scénarios déduits ne pouvaient être joués sur les données réelles historiques. Pour visualiser clairement ces mesures d'impact et donner le pouvoir aux utilisateurs, Rexel recherchait une solution à la fois puissante et simple d'utilisation, et qui puisse gérer de gros volumes de données : 1 million de références produits, 500 millions de lignes de stocks, et des analyses sur un grand nombre d'axes.

Qlik Sense répond à la fois aux besoins des utilisateurs métiers et des data scientists, qui utilisent Qlik Sense pour créer des mini-applications pour les présenter en comité de projet et manipuler les données « en live ».

**LE MONDE DES ANALYTICS ÉVOLUE BEAUCOUP, OÙ EN EST QLIK AUJOURD'HUI ?**

**Angélica Reyes :** Qlik a beaucoup évolué depuis sa création, mais la disruption technologique qui fait toute la différence par rapport aux autres solutions du marché reste le moteur associatif, qui est le socle de notre plateforme. Cette technologie permet d'associer toutes vos données et de les explorer librement sur n'importe quel axe d'analyse, sans les contraintes de hiérarchies prédéfinies que l'on trouve dans les outils traditionnels. C'est ce qui a ouvert le marché de la data discovery. Sur la base de notre moteur, nous avons fait évoluer notre plateforme pour répondre aux nouveaux enjeux : analyse en libre-service, plateforme totalement web et responsive, intégration des fonctionnalités de mapping et geoanalytics, intégration de données externes avec notre Data Market, et toujours plus d'intuitivité pour les utilisateurs avec des fonctionnalités comme le smart search, par exemple.

Qlik a été choisi par plus de 45 000 clients dans le monde pour toutes ces raisons. Notre solution répond aux enjeux majeurs des entreprises aujourd'hui : données en silos, issues de sources différentes, gouvernance de la donnée pour procurer une seule version de la réalité, rapidité d'implémentation, intuitivité, une puissance d'analyse inégalée. Nous sommes également très fiers d'être reconnus par les cabinets d'analystes : Qlik est le seul acteur positionné comme leader sur le marché de la Data et des Analytics dans toutes les études les plus récentes (Gartner, Barc, Forrester).

**QUELS SONT LES SUJETS SUR LESQUELS VOUS ÊTES LES PLUS PRÉSENTS ?**

**Angélica Reyes :** Qlik est présent partout, dans tous les métiers et tous les secteurs, et de plus en plus sur les projets Big Data. En effet, pour que ces projets aient du sens, il faut de vrais cas d'usages, pour lesquels les utilisateurs métiers doivent prendre des décisions. Ce que nous disent nos clients, comme Ricard, Rexel, ou encore Auchan:Direct, c'est que la visualisation des données est un pilier essentiel d'une stratégie Big Data : pour faire collaborer les data scientists et les décideurs métiers, Qlik permet de rendre concrets les algorithmes en visualisant les modèles et leurs impacts.

**QUE PRÉPARE QLIK POUR L'AVENIR ?**

**Angélica Reyes :** Qlik investit sur 4 sujets majeurs :

- **Laisser les données où elles sont :** les sources de données à analyser se multiplient, dans l'entreprise mais aussi à l'extérieur, on-premise et dans le cloud. Pour simplifier leur exploitation, nous proposons d'amener Qlik aux données et non pas d'amener les données à Qlik. Avec le Big Data, par exemple, Qlik, en plus de la génération d'applications à la demande, travaille sur l'indexation des big data, pour bénéficier de la puissance du moteur associatif, tout en laissant les données là où elles se trouvent, que ce soit dans des data lakes ou hadoop.
- **Intelligence Augmentée :** Nous pensons que combiner les capacités cognitives des humains avec l'intelligence artificielle est la bonne formule pour démultiplier le potentiel d'exploitation des données. Dans la lignée de cette stratégie, nous travaillons sur l'intégration de l'intelligence artificielle à chaque étape du traitement de la donnée, et sur un moteur de règles cognitives, qui permettra de générer automatiquement des analyses et des indicateurs clés. Couplé au machine learning, le système deviendra plus intelligent dans ses suggestions, au fur et à mesure qu'il captera les interactions avec les utilisateurs.
- **Cloud Hybride :** le cloud se développe rapidement, Gartner prédit qu'une entreprise sur deux aura déployé ou programmé une solution dans le cloud cette année. Mais ce qui fait la différence de Qlik, c'est que nous donnons le choix à nos clients. Nous n'imposons pas un modèle, mais nous proposons un cloud hybride, qui permet de répartir les traitements, les données et la publication sur différentes plateformes, on-premise ou dans le cloud, qu'il soit public ou privé. Cette architecture Cloud Hybride, basée sur les micro-services, garantit la même expérience au travers des infrastructures. Le client pourra choisir selon ses besoins, son organisation et ses contraintes, l'organisation qui lui procurera le plus de valeur et qui correspond le mieux à ses cas d'usages.

**45 000 clients satisfaits**

*Qlik est une brique essentielle de l'analyse Big Data, il nous permet d'allier Data Science et analyse visuelle pour améliorer la connaissance clients.*

**RICARD**  
DEPUIS 1932

*Ce qui fait la différence de Qlik, c'est la puissance de traitement, la facilité à analyser des sources de données disparates, et le moteur associatif.*

**ALSTOM**

*La dataviz offerte par Qlik Sense est un élément essentiel du processus prédictif : elle permet de redonner le pouvoir à l'utilisateur.*

**REXEL**

*Qlik repousse les limites de la BI traditionnelle grâce à la puissance de son moteur associatif et à la rapidité des développements.*

**THALES**



# Saagie®



CONTACT

Dans un monde où 80% des projets de Data Science ne voient jamais la production\*, les entreprises peinent à créer de la valeur à partir de leurs initiatives de Big Data et d'IA pour chacune de leurs équipes et dans des délais acceptables.

### COMMENT LES ENTREPRISES RÉINVENTENT LEURS MÉTIERS AVEC SAAGIE

Saagie fournit une plateforme qui permet de déployer rapidement des projets Big Data en production et d'embarquer de l'IA dans les applications des entreprises, au service de leurs métiers. Elles peuvent extraire, préparer et travailler la donnée de la manière la plus sécurisée possible avec la gouvernance des données appropriées, en pleine conformité avec les réglementations en place. C'est une plateforme d'industrialisation de la data science fondée sur une architecture distribuée et sur un socle open-source qui permet de rassembler les technologies et les compétences (data scientists, data analysts, équipes IT) pour mener à bien leurs projets. Prête à déployer, entièrement managée et maintenue à jour, elle leur permet de se concentrer sur ce qui est important pour elles : la collaboration des équipes au service de leurs métiers et de leur transformation.

### SAAGIE - SMART DATA PLATFORM

Grâce à la Saagie Smart Data Platform, les entreprises peuvent exploiter toutes les données à leur disposition pour concevoir des applications prédictives innovantes. Elle inclut 2 composants principaux :

**Data Fabric** : véritable ligne d'assemblage permettant l'extraction de la donnée, les traitements et la création d'algorithmes de Machine Learning avec le meilleur des technologies open-source Big Data & IA. Elle comprend également des fonctionnalités clés telles que le versionnement, le rollback, les logs et des data pipelines pour industrialiser la Data Science.

**Data & Analytics Governance** : portail d'accès à la donnée permettant de documenter et comprenant des outils d'Habilitation, d'Audit et de Data Lineage pour assurer la sécurité, la conformité et la qualité de la donnée.

Et fédère tous les utilisateurs de données en répondant à leurs exigences :

**Data Analysts** : en utilisant des outils graphiques performants pour travailler avec la donnée.

**Data Scientists** : en tirant profit de l'écosystème open-source des technologies de Data Science.

**Développeurs** : en bénéficiant d'une agilité technologique importante tout au long des processus de développement et de livraison continus.

**IT / Ops** : en assurant la mise en production, la sécurité et la gouvernance de la donnée.

\*Source : Gartner

### LES CAS D'USAGE :

**Saagie Smart Platform** est la plateforme unique permettant de couvrir tous les usages métiers dans le secteur de la Banque et l'Assurance, de manière sécurisée et conforme :

**Détection de fraude** par l'analyse en temps réel des flux et la modélisation des comportements atypiques et frauduleux

**Churn** par l'analyse des données des clients, à la fois personnelles (âge, CSP, etc.) et transactionnelles (nombre de transactions par mois, montant des transactions, le type de paiement utilisé, etc.) afin de construire des algorithmes prédictifs. Le but est de détecter les moments de vie des clients afin de leur proposer des offres plus adaptées à leur profil.

**Automatisation des processus métiers (BPA)** en utilisant l'Intelligence Artificielle pour faciliter l'onboarding des clients, en mettant en place des processus de KYC et d'aide au traitement de mails.

**Prévision** de soldes, de ventes de produits bancaires pour l'aide à la gestion personnalisée de trésorerie.

### Saagie et ses clients Banques & Assurances



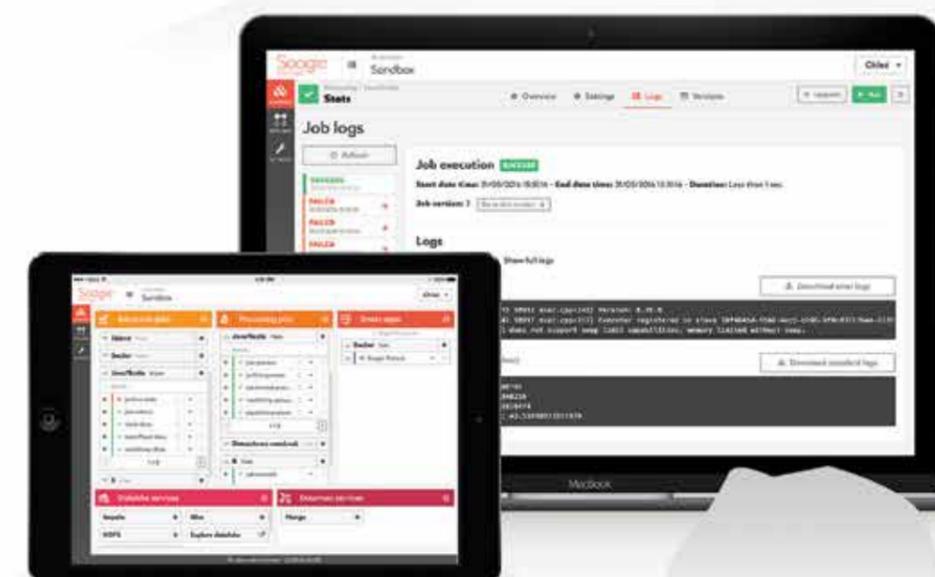
# Saagie®

## Réinventez votre business avec Saagie

Embarquez l'Intelligence Artificielle dans vos applications métiers avec le meilleur des technologies Big Data

### Ce que vous avez à y gagner ?

- ✓ Le meilleur des technos IA intégré à votre infrastructure existante
- ✓ Une plateforme unifiée pour tous vos collaborateurs : les métiers, l'IT et les équipes Data
- ✓ Le stockage et le traitement des données de manière sécurisée et conforme à vos standards



### Exemple de cas d'usages



### En savoir plus...





## UNIFIER LES SILOS DE DONNÉES DU BIG DATA



**Christophe BOGGERO**  
Pre-sales Technology Manager  
christophe.boggero@sap.com



**Olivier ENAULT**  
Pre-sales Information Excellence  
olivier.enault@sap.com

Après avoir rationalisé les données des mondes transactionnel et du datawarehouse, SAP s'attaque à la gouvernance des Big Data. Un chantier prioritaire tant ces silos de données se sont développés de façon anarchique. Les explications de Christophe Boggero, pre-sales Technology Manager et Olivier Enault, pre-sales for Information Excellence chez SAP France.

La stratégie de SAP dans le Big Data se présente autour du portefeuille de solutions Leonardo. Quelle approche adoptez-vous avec cette offre ?

**Christophe Boggero** : SAP Leonardo était à l'origine une offre logicielle centrée sur l'Internet des Objets. Aujourd'hui, en plus de s'enrichir de solutions pour le Big Data, le Machine Learning ou la Blockchain, SAP Leonardo concrétise avant tout une démarche d'innovation. Car, ce n'est pas tant la question du comment qui compte, mais plutôt celle du pourquoi. Autrement dit, les outils ou solutions avec lesquels on va récupérer et manipuler la donnée ont moins d'importance que ce qu'on va en faire. C'est tout l'enjeu de notre démarche basée sur la co-innovation avec nos clients et sur notre méthodologie de design thinking, qui aide les organisations à bien identifier le problème à résoudre.

Cela ne signifie pas que SAP Leonardo ne résolve pas aussi la question du comment, avec les solutions qu'il renferme ou celles qui lui sont connexes. Citons par exemple le cas de MII (Manufacturing Integration and Intelligence), un module qui préexistait et sert à optimiser les chaînes de production. En couplant Leonardo à MII, on peut imaginer pousser plus loin le concept de l'usine étendue, en intégrant des données issues du marketing ou des ventes afin d'ajuster la production en temps réel. SAP Leonardo est un tout permettant de capter l'information et de la transformer en intelligence immédiatement exploitable par l'organisation.

**Les applications Big Data se sont largement déployées sur des technologies Open Source comme Hadoop. Quel rôle peut jouer SAP dans ce paysage ?**

**C.B.** : On assiste aujourd'hui à un morcellement des points de stockage de données dans les organisations. Avec Vora, outil permettant d'intégrer données structurées et données non structurées des environnements Hadoop, SAP a fait un premier pas dans l'unification de ces informations. Avec le lancement de Data Hub, point focal de toutes les données de l'entreprise, nous allons désormais plus loin. Cette offre permet aux métiers de tirer directement de la valeur de l'information stockée, en bâtissant une vision logique des données. Ainsi, on peut par exemple mixer une analyse de sentiments sur les réseaux sociaux, avec des prévisions de

vente et rapprocher le tout de la gestion des stocks afin d'optimiser ces derniers et ajuster la production.

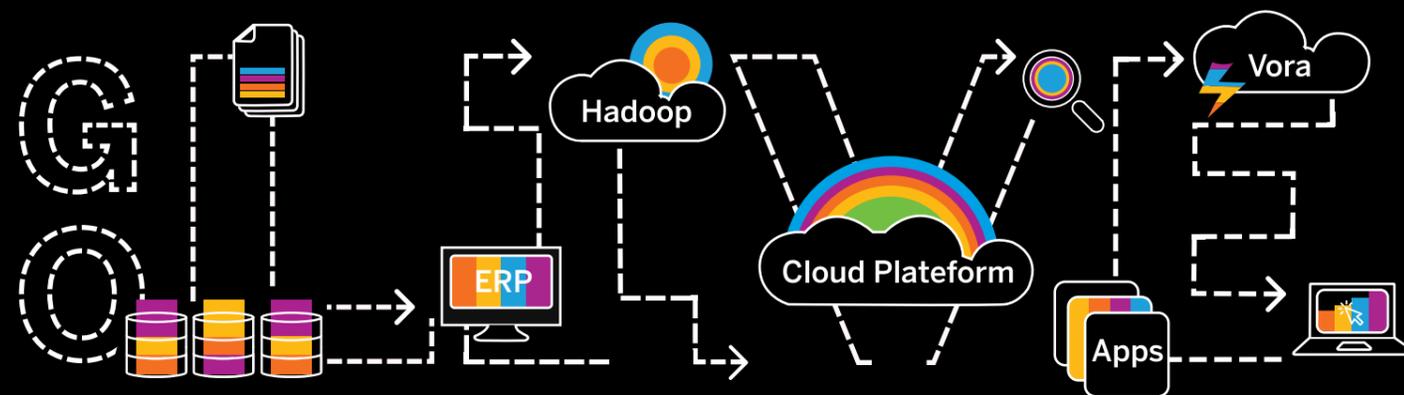
**Olivier Enault** : C'est un prolongement de l'effort de rationalisation des mondes du transactionnel et du datawarehouse que nous menons depuis plusieurs années avec notre architecture de gestion de données In-Memory, SAP HANA. Ces deux univers renferment certes un ensemble d'informations central dans la vie des organisations, mais qui ne représente in fine qu'environ 5 % des données que détient une entreprise. Si on additionne les environnements Hadoop, le stockage objet fourni par les géants du Cloud et les stockages intermédiaires, il existe aujourd'hui de multiples silos créés pour les applications Big Data. Dans ces environnements, les organisations ont besoin de gouverner la donnée - d'où provient-elle ? ; qui et comment y accède-t-on ? -, pour mettre en place des traitements permettant par exemple des croisements avec les données présentes dans les applications historiques. Data Hub amène cette couche de gouvernance de la donnée, tout en s'appuyant sur les capacités de traitement des environnements Hadoop. On peut le comparer à la tour de contrôle d'un aéroport, sachant profiler la donnée, en répertorier les flux, tout en étant capable de la traiter directement dans le monde Hadoop, sans déplacement inutile, via ce qu'on appelle des data pipelines

**Les entreprises ont accumulé des données dans les data lakes sans intégrer les exigences du RGPD, Règlement européen sur la protection des données personnelles qui entre en vigueur le 25 mai. SAP a-t-il une approche pour la mise en conformité de ces systèmes ?**

**C.B.** : Concernant les nouvelles données qu'on injecte dans les silos de stockage, SAP est capable de fournir l'outillage nécessaire pour se conformer aux exigences du RGPD. Le problème est plus épineux avec le patrimoine de données : aller fouiller ces couches d'information, pour anonymiser certaines données personnelles ou en effacer d'autres, s'avère très complexe. SAP, pas plus qu'un quelconque autre fournisseur, ne dispose pas d'une baguette magique pour y parvenir. Par contre, notre offre SAP Data Hub peut apporter sa contribution, en facilitant l'identification des données personnelles.

Notons d'ailleurs que SAP s'engage pleinement dans l'amélioration du respect des données personnelles. Nous avons notamment racheté Gigya, un éditeur spécialisé dans la gestion des profils de clients dont la solution est appelée à jouer un rôle central dans la stratégie de SAP. En commençant par unifier la gestion des clients dans les différentes offre Hybris (CRM, gestion commerciale, marketing). Cette centralisation facilitera la vie des clients, en autorisant leur identification via des profils sociaux par exemple, fiabilisera les campagnes marketing ou les applications analytiques, en fournissant une vision unique sur les clients, tout en amenant des solutions assurant la conformité au règlement européen. En concentrant les profils clients, Gigya facilite toute opération sur les données personnelles, en particulier l'exercice du droit de retrait par les individus.

## TRANSFORMER LES BIG DATA EN INSIGHTS ET STIMULER L'INNOVATION POUR FONDER L'ENTREPRISE LIVE



**SERGE BOULET**  
Directeur marketing

01 60 62 11 11  
comsas@sas.com  
Domaine de Grégy - Grégy-sur-Yerres  
77257 Brie Comte Robert Cedex France  
www.sas.com/france



**HELGE ROSEBØ**

Architecte et Responsable Visualisation des Données



## MUNICH RE MET EN PLACE UNE PLATE-FORME BIG DATA AVEC SAS ET HORTONWORKS

Le plus grand réassureur mondial, a retenu SAS® Analytics et Hortonworks Data Platform pour déployer une plate-forme innovante capable d'analyser des volumes gigantesques de données.

Pour Munich Re, la donnée est une composante essentielle de sa stratégie pour prendre en compte activement l'évolution mondiale du risque. Lancée en février 2016, la plate-forme permet aux différents départements d'explorer de nouvelles idées, de développer de nouveaux domaines commerciaux et d'améliorer le service à la clientèle.

L'analyse des données semi-structurées et non structurées, de données météorologiques ou de capteurs sur des véhicules, machines et autres dispositifs en réseau, font émerger des idées de produits, services et de nouveaux processus.

La plate-forme basée sur une approche ouverte fournit aux data scientists de Munich Re le plus haut degré de liberté et de créativité. Un lab leur permet de tester les idées sans affecter les opérations quotidiennes. Environ 500 utilisateurs travaillent déjà avec le système, qui est en cours d'extension par étapes.

## LE FEU A DÉCUPLÉ L'INTELLIGENCE HUMAINE. COMMENT S'EN INSPIRER ?

Nous consacrons une grande part de notre existence à faire des choix. En essayant de faire les meilleurs possibles. Baser les décisions sur des faits importe plus que le type et la qualité des données qui l'alimentent. En ancrant cette approche dans votre culture d'entreprise, vous saurez tirer parti du vaste potentiel qu'offrent les données. **Par Helge Rosebø, Architecte et responsable visualisation des données, SAS Norvège.**

La ressemblance entre cet état d'esprit et l'évolution de l'intelligence humaine me paraît frappante. En utilisant le feu comme technologie, nous avons décuplé notre intelligence. De quelle manière votre entreprise pourrait-elle développer la sienne ?

### Le feu nous a rendus plus intelligents

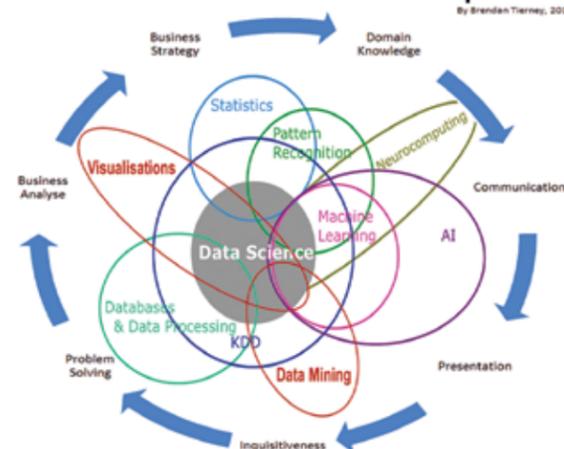
Le feu a eu un énorme impact sur le mode de vie de nos ancêtres, en leur apportant du confort et en leur permettant de cuire leurs aliments. La viande est plus digeste cuite que crue. L'organisme utilise donc moins d'énergie à assimiler un steak qu'un morceau de viande crue. Moins d'énergie pour digérer signifie également plus de vitalité pour d'autres fonctions corporelles. Autrement dit, la cuisson a contribué à développer notre intelligence, et le reste appartient à l'histoire. Nous sommes aujourd'hui arrivés à un stade où les nouvelles technologies pourraient bien surpasser nos **capacités cognitives**. Auparavant, le feu était la technologie charnière et la viande était une source d'énergie. La combinaison de la viande (données), du feu (technologie) et de la cuisson (processus) a dopé notre intelligence.

## DANS LE PAYSAGE NUMÉRIQUE ACTUEL, COMMENT DEVENIR PLUS INTELLIGENT QUE NOS CONCURRENTS ?

....Grâce à la science des données !

Les applications de la data science ont pour finalité de traduire les nouvelles connaissances en croissance économique. Le diagramme de Brendan Tierney offre un aperçu des compétences et du processus qui composent cette science. Il me sert d'aide-mémoire, et vous pouvez vous aussi l'utiliser pour replacer ce phénomène ultra-médiatisé dans le contexte de vos tâches et compétences actuelles.

### Data Science Is Multidisciplinary



S'il n'existe encore aucune technique d'intelligence artificielle ou de machine learning capable de faire disparaître tous nos problèmes d'un coup de baguette magique, elles sauront en revanche augmenter votre efficacité. Et celle-ci atteint son apogée lorsque l'on associe un panel de compétences pour répondre à des problématiques métier en constante évolution. Si l'on sait quelles compétences appliquer à une problématique donnée, c'est gagné !

### « Faire des concessions sur la technologie impacterait l'efficacité du processus. »

L'application de la science des données peut impliquer différentes fonctions (développeur, statisticien, analyste métier, expert en entretêts de données, chef de projet, etc.), ce qui rend l'accès à la réussite plus complexe. L'équipe de data science doit être réactive, souple et rapide, et ses travaux doivent pouvoir être aisément partagés à d'autres branches de l'entreprise. Elle a besoin d'une technologie qui valorise efficacement ses différentes contributions et compétences, et qui combinerait par exemple des solutions Open Source et propriétaires. Quoi qu'il en soit, les équipes ont besoin d'un environnement technique propice au bon déroulement du processus, du développement à la production, en passant par les tests et l'évaluation qualité. Faire des concessions sur la technologie impacterait l'efficacité du processus. Mais comme toujours, une excellente technologie ne peut que libérer le potentiel, le reste relève de notre responsabilité et de nos processus.

*L'intelligence artificielle comprend de nombreux sous-domaines :*

**L'apprentissage automatique (machine learning)** automatise le modèle analytique. Il utilise des méthodes à partir de réseaux de neurones, de statistiques, de recherche opérationnelle et de physique pour trouver des informations cachées dans les données, sans être programmé pour savoir où chercher et quoi conclure.

**Un réseau neuronal** est une sorte de machine learning inspiré du fonctionnement du cerveau humain. C'est un système informatique composé d'unités interconnectées (comme les neurones) qui traite l'information en répondant aux entrées externes, en relayant l'information entre chaque unité. Le processus nécessite plusieurs passages aux données pour trouver des connexions et dériver des données non définies.

**L'apprentissage profond (deep learning)** utilise d'énormes réseaux de neurones avec de nombreuses couches de traitement, profitant des progrès de la puissance de calcul et de l'amélioration des techniques pour l'apprentissage de modèles complexes sur de grandes quantités de données. Les applications courantes incluent la reconnaissance de l'image et de la parole.

**La vision par ordinateur** repose sur la reconnaissance des formes et l'apprentissage profond pour reconnaître ce qui se trouve dans une image ou une vidéo. Lorsque les machines peuvent traiter, analyser et comprendre des images, elles peuvent capturer des images ou des vidéos en temps réel et interpréter leur environnement.

**Le traitement du langage naturel** est la capacité des ordinateurs à analyser, à comprendre et à générer du langage humain. La prochaine étape de cette science est l'interaction du langage naturel, qui permet aux humains de communiquer avec les ordinateurs en utilisant une langue normale et quotidienne pour effectuer des tâches.

Struggling to develop and deploy data analytics?

LET'S CHANGE THAT™





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 DIFFÉRENCE : ASSOCIER DATA ET MARKETING

Notre mission est d'accompagner nos clients dans la définition et la mise en œuvre de stratégies de développement de leur business, sur l'ensemble des drivers : 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, les données non structurées (notamment websemanitique), les données publiques...

Nous croyons fondamentalement aux vertus de la data mais ne concevons pas notre métier sans y ajouter une dimension forte de stratégie et de marketing.

### DÉTAIL DE NOS PRESTATIONS

#### Stratégie Business :

- Mettre les nouvelles technologies, les nouvelles formes de communication et l'intelligence des data au service du développement du business des entreprises.

#### Optimiser la performance du commerce connecté :

- Exploiter la donnée pertinente pour concevoir et optimiser les parcours cross-device & cross-canaux : audit de performance des interfaces & parcours; UX/UI design ; plan d'optimisation continue (A/B tests) ; audit de la performance des stratégies pluri-media pour optimiser les budgets investis.
- Intégrer la digitalisation des magasins dans la réalité du business de chaque client

#### Capter une donnée de qualité et assurer sa bonne exploitation :

- Mettre en place des systèmes de collecte de données (ex : plan de tag) et de stockage (DMP) afin d'en extraire l'intelligence (KPIs, algorithmes) en vue d'améliorer le pilotage des activités par les équipes métiers
- Concevoir et éditer de rapports multi-sources personnalisés et automatisés.

#### Réinventer les conversations et le CRM :

- Prendre soin de chaque client, prospect, ami pour développer la préférence durable à la marque et un bouche-à-oreille positif.
- Concevoir et mettre en place des stratégies autour du commerce conversationnel.

#### Faire du Big Data un facteur de transformation de l'entreprise :

- Appréhender le Big Data comme un levier de transformation durable de l'entreprise, ses métiers, son offre, ses prévisions de vente, sa relation à ses cibles et non comme un sujet technologique.

#### NOS VALEURS : TRANSPARENCE - PRAGMATISME - CURIOSITÉ

- **Transparence** = accompagner nos clients dans leur montée en compétence.
- **Pragmatisme** = évaluer nos recommandations sur un ratio complexité-coût / apport de valeur et ne pas mener des projets simplement intellectuellement satisfaisants
- **Curiosité** = exploiter au mieux les technologies innovantes existant à travers le monde grâce notamment au foisonnement de startup

#### SECTEURS CIBLES

Chez Scinetik, nous travaillons pour tous types de secteurs d'activités. Toutefois, notre histoire fait que nous travaillons majoritairement pour de grands groupes internationaux dans les domaines du retail, de la grande consommation, de la mode, de la décoration et du sport.

## FAIRE DU BIG DATA UN FACTEUR DE TRANSFORMATION DE VOTRE ENTREPRISE



Cadrage stratégique



Sources & dataquality



Modélisation statistique



Application métiers & partage

# LES BEST SELLERS SCINETIK



### 1 Comment prédire le succès d'un lancement de produit ?

Identifier les principales variables explicatives de la réussite ou de l'échec d'un lancement de produit pour détecter les signaux pertinents. Définir des seuils qui permettront de déclencher les actions correctives au plus tôt.



### 2 À quel moment déclencher des actions commerciales et avec quelle démarque ?

Croiser les données internes (stock, web analytics, CRM, ticket de caisse ,.../...) avec les facteurs explicatifs externes (météo, indice du moral des ménages, calendaires, géolocalisation, concurrence, etc.) pour décider de la période optimale pour lancer les opérations de promotion.



### 3 Comment l'écoute du web permet-elle de comprendre la réalité de la perception des marques ?

Utiliser les conversations du web social (forum, blogs, sites spécialisés, réseaux sociaux) pour mesurer les co-occurrences entre la marque et ses concurrents et évaluer les items d'image positifs et négatifs et leurs évolutions dans le temps.



### 4 En quoi les nouvelles sources de données permettent-elles de concevoir des services innovants ?

Enrichir la donnée client existante avec des données type circulation des individus dans les espaces de vente (et autres lieux de passage) pour créer des profils avancés. Élaborer de nouveaux services et offres personnalisés à partir de ce profiling.



### 5 Quelle modélisation du mix-media pour optimiser la rentabilité des investissements futurs ?

Auditer les campagnes pluri-media afin de construire un modèle d'attribution permettant d'améliorer l'efficacité des stratégies de communication.



### 6 Comment utiliser l'Open data pour mieux réinventer le magasin dans son contexte local ?

Réaliser une cartographie des zones étudiées en visualisant les performances des magasins et les facteurs explicatifs liés aux caractéristiques locales. Intégrer les outputs de l'analyse pour reconcevoir des magasins plus adaptés à leur environnement.



## LE COGNITIVE SEARCH AU CŒUR DES PROJETS DE TRANSFORMATION DIGITALE

Transmettre l'information appropriée à la bonne personne au bon moment : telle est la finalité du cognitive search. Plus précisément, il s'agit de fournir des informations pertinentes et complètes à des personnes selon leur contexte professionnel, de sorte qu'elles puissent prendre des décisions plus avisées, renforcer leur efficacité, mieux répondre aux besoins de leurs clients et connaître une plus grande satisfaction au travail. Pour ce faire, le cognitive search associe différentes méthodes d'extraction d'informations sur des données structurées et non structurées, et sur le comportement des utilisateurs.

### Cognitive Search = Search+ NLP + AI/ML

Dans cette équation, par le terme « search », nous n'entendons pas la recherche de mots clés telle que nous l'avons connue dans le passé, mais une recherche haute performance qui intègre différents types d'analyse. Le NLP (TALN - traitement automatique du langage naturel) désigne, en plus du traitement statistique des langues, une analyse linguistique et sémantique approfondie. Quant à l'IA, elle ne se limite pas ici à définir un élément qui ne serait que « saupoudré » sur une structure de recherche classique. Elle fait partie intégrante d'une architecture de bout en bout, intégrée et évolutive.

### L'intelligence artificielle, vorace en données

Pour que les algorithmes d'IA et de Machine Learning (ML) fonctionnent de façon optimale, vous devez les alimenter avec toutes les données qu'il vous est possible de récupérer. Une plateforme de recherche cognitive doit accéder à la majeure partie des sources de données d'une entreprise : données internes et externes de tous types, sur site et dans le cloud.

### Un enrichissement continu

La recherche cognitive permet d'accumuler des connaissances sur des données structurées et non structurées, ainsi que sur les préférences et le comportement d'une personne. C'est ainsi que les utilisateurs peuvent obtenir des informations de plus en plus pertinentes par rapport à leur contexte de travail. Pour accumuler des connaissances, une plateforme de recherche cognitive a besoin d'un référentiel, qui est en principe un lieu de stockage. Ce référentiel s'appelle un entrepôt de données logique. Il contient des informations sur des données, et est continuellement enrichi par le TALN et par les résultats du clustering et des calculs des similarités, etc. Par conséquent, les algorithmes de Machine Learning et autres analyses intelligentes doivent ingérer les données de cet entrepôt et l'alimenter avec leurs résultats.

### Combiner différentes méthodes d'extraction d'informations

Pour produire les meilleurs résultats possibles, les différentes méthodes analytiques doivent être combinées, et non exécutées séparément les unes des autres. Par exemple, si vous exécutez des algorithmes de Machine Learning sur des données textuelles d'où des concepts et des liens entre les concepts ont déjà été extraits via des analyses linguistiques et sémantiques, vous obtiendrez des résultats beaucoup plus intéressants, plus vite.

### Sécurité/Droits d'accès

Une plateforme de recherche cognitive ne doit pas permettre à des utilisateurs individuels de contourner le contrôle d'accès. Même à la fin d'une session élaborée de Machine Learning ou autre processus d'enrichissement, aucun utilisateur ne doit pouvoir voir les informations auxquelles il n'a pas le droit d'accéder.

### Exemples et cas d'usage

La vision à 360° et l'accès à des réseaux d'experts sont toujours les deux cas d'utilisation les plus répandus. De fait, de nombreux cas d'utilisation peuvent être présentés en tant que vision à 360° sur un sujet spécifique, que ce sujet soit le client, un produit, un sujet de recherche, etc. Cela dit, examinons plus en détail deux cas d'utilisation spécifiques.

#### • Conformité réglementaire

Dans ce cas d'utilisation, les entreprises passent au crible les dernières mises à jour des agences de régulation du monde entier, découvrent en quoi consistent les nouvelles réglementations et déterminent si elles concernent l'un de leurs produits. Si c'est le cas, elles transmettent les informations pertinentes concernant ces nouvelles réglementations aux personnes concernées au sein de l'entreprise et chez les fournisseurs. Un traitement automatique approfondi du langage naturel, des algorithmes Machine Learning servent à trouver les nouveaux règlements qui concernent un produit de l'entreprise et les experts internes à être alertés. Un autre aspect de ce cas d'utilisation est : « écouter le client », c'est-à-dire voir ce que les gens disent de vos produits sur les réseaux sociaux. Pour les entreprises du secteur pharmaceutique, cette action peut vouloir dire détecter de nouveaux effets secondaires dans les discussions entre patients. Des effets ainsi détectés doivent être signalés aux agences de régulation dans les plus brefs délais.

#### • Maintenance et réparation de systèmes complexes

Par « systèmes complexes », nous entendons les avions ou les hélicoptères, les trains, les centrales électriques, etc., tout ce qui intègre d'innombrables composants, souvent associés de manière unique, qui rendent ces systèmes très particuliers. Sur ce genre de systèmes, les rapports de problème peuvent être générés par voie humaine ou provenir des données de capteurs, ou les deux. Dans tous les cas, il s'agit d'orienter un opérateur humain afin qu'il fasse le nécessaire, dans les temps impartis. Cette étape implique une extraction rapide d'informations utiles d'une documentation produit complexe et volumineuse, des procédures de maintenance/réparation, des réglementations et des rapports. Elle nécessite également de tenir compte de l'interdépendance entre les différents composants. Les rapports sur des cas similaires, pour lesquels les solutions apportées ont fonctionné, peuvent considérablement accélérer le processus et apporter une aide précieuse aux opérateurs. Les utilisateurs doivent être alertés qu'ils ne peuvent pas changer une pièce sans en changer d'autres et qu'ils ne peuvent pas utiliser certaines pièces dans certaines configurations. Ils doivent également savoir quelles pièces sont interchangeables avec d'autres, et quels tests sont obligatoires après certains remplacements/réparations. Ils doivent savoir quand ils doivent signaler un dysfonctionnement ou même quand alerter les propriétaires de systèmes possédant une configuration similaire.

La recherche et l'analyse cognitives peuvent générer un ROI significatif en permettant aux équipes de maintenance de traiter un plus grand nombre d'interventions.

## A LEADING COGNITIVE SEARCH PLATFORM FOR INFORMATION-DRIVEN ORGANIZATIONS

- Make informed decision
- Accelerate innovation
- Increase efficiency to enhance competitiveness
- Improve customer service



@Sinequa

www.linkedin.com/company/sinequa

www.sinequa.com



## RETOURS D'EXPERIENCES PROJETS DE MACHINE LEARNING OU BIG DATA

### Offre d'analyse automatique d'échanges clients

L'offre d'analyse automatique de textes proposée par Skapánè vous permet d'améliorer l'efficacité de vos contacts client grâce à l'extraction automatique d'information pertinente de tous les messages provenant de tous vos canaux de contact client.

Ce robot numérique propose des services de catégorisation automatique et d'aide à la décision tels que la préconisation de réponse et la qualification automatique de contact s'appuyant sur un DataLake de données multi-sources provenant de votre système d'information interne (informations du compte client, niveau de fidélisation, ...) et/ou externe (catégorie socio-professionnelle, zone d'habitation, ...).

L'utilisation des outils du machine Learning, utilisés, selon les cas, en apprentissage supervisé (Random Forest, NaiveBayes, ...) ou non supervisé (Kmeans, ...), couplée aux meilleures techniques de Text Mining des messages garantit un haut niveau de pertinence de catégorisation et de qualification. Grâce à ces technologies, le système est capable de s'adapter à tous les types de phraséologie et à tous les niveaux de langage utilisés par vos clients.

Grâce à un apprentissage permanent, le système est aussi capable d'évoluer et de s'améliorer au fur et à mesure de son approvisionnement en messages, et de toujours mieux « comprendre » les messages formulés en langage naturel.

Pour vous permettre de mesurer, de comprendre, et d'optimiser la gestion de vos contacts, l'offre vous propose des outils statistiques, des indicateurs de suivi, et des capacités d'exploration dynamique des données de contact, exposés via une interface web. Et pour vous offrir une vision unifiée à 360° des canaux de contact client, ce produit propose en option l'intégration de tous vos messages papier, qui seront interprétés par la machine et traités de manière unifiée grâce à l'utilisation d'outils de reconnaissance de caractères et d'écriture manuscrite, ainsi que des messages vocaux, grâce à l'intégration d'un moteur de reconnaissance vocale. Cette offre est fournie à un acteur de la banque et permet de fixer automatiquement la priorité de traitement d'un message client sur la base de détection du sujet et de la tonalité (« l'humeur ») du client.



### Maintenance prédictive

Le monde de l'industrie dispose de processus de maintenance préventive permettant de diminuer le risque d'incidents sur des critères établis à la conception des systèmes. Le complément idéal de ces processus consiste en la capacité à prévoir des incidents potentiels sur la base d'équipements en production et de sites industriels en fonctionnement. Ce nouveau type d'outils, la maintenance prédictive, permet d'ajouter des capacités de prévision d'incidents avant qu'ils n'apparaissent sur des environnements industriels opérationnels.

Les capacités de traitement nécessaires au calcul de modélisation sur des centaines de milliards d'événements et de mesures remontés par l'ensemble de capteurs et systèmes d'environnements industriels complexes (raffinerie, centre de tri, ...) ont été mobilisés par Skapánè pour créer des algorithmes prédictifs sur la base de Support Vector Machines incluant des algorithmes de gestion des séries temporelles. La technologie de Skapánè permet l'utilisation d'événements asynchrones (alertes) et des événements synchrones (mesures régulières) dans les mêmes algorithmes, maximisant le pouvoir prédictif des modèles utilisés en production. Ces solutions prévoient correctement 95% des incidents réels avant qu'ils n'apparaissent sur une période de plus de 10 ans de la vie de l'installation.

Les industriels ont maintenant la capacité de prévoir et donc d'éviter des incidents qui apparaissent malgré la maintenance préventive, économisant ainsi des coûts élevés liés à l'arrêt complet possible des équipements suivant le type d'incidents. L'utilisation d'outils de suivi en temps réel fournis par Skapánè permet une intégration complète dans les outils industriels de pilotage.

Cette solution a été validée sur la base de l'historique de mesures d'une raffinerie sur une période de 10 années et a permis de démontrer la capacité à prévoir les incidents techniques avant leur survenance.



### DITES-NOUS QUI EST SKAPÁNÈ ?

Depuis notre création en Mars 2015, nos locaux se situent au sein du pôle d'excellence et d'innovation Euratechnologies à Lille. D'abord incubés, nous sommes aujourd'hui accélérés. Le projet de Skapánè repose sur l'innovation apportée par les nouvelles solutions de traitement de la donnée et d'apprentissage automatique, notamment pour le temps réel : recommandation de produits, maintenance prédictive, lutte contre la fraude... Il s'agit de mettre en œuvre des algorithmes mathématiques poussés et puissants existants sur le marché afin de dépasser l'usage traditionnel de la donnée en tant que moyen d'analyse, pour en faire un outil de prédiction, et être capable de l'appliquer on-line directement sur les applications de traitement, interactives. La culture du résultat et la vertu du modèle Cloud, permettent aux clients de Skapánè de bénéficier de l'expertise d'acteurs spécialisés. Pour financer son développement, Skapánè a réalisé un premier tour de table au dernier trimestre 2017.

### QUELS SONT LES NOUVEAUTÉS À VENIR CHEZ SKAPÁNÈ ?

La grande nouveauté chez Skapánè est la mise en place d'une plateforme de service client sous forme d'un logiciel SaaS facilement accessible et utilisable par les clients. Skapánè fournit toute l'infrastructure mais le client est totalement autonome concernant l'utilisation de l'application. L'interface proposera plusieurs fonctions métiers tel que la détection de fraude, et plus généralement la prédiction de comportements clients : achats, mode de livraison, moyen de paiement, ...

### SUR QUELS TYPES DE PROJETS TRAVAILLEZ-VOUS ?

Nous travaillons sur des projets variés tel que la lutte contre la fraude en apprentissage automatique, ici il s'agit d'analyser le comportement des clients via des algorithmes de Machine Learning mis en place par nos soins, ce qui va permettre de définir un score de risque de fraude et donc d'autoriser le paiement en un clic pour les clients fiables et le 3DS pour les clients potentiellement frauduleux. Parmi les axes de développement, l'analyse automatique d'échanges clients, le traitement d'images issues du monde médical ou le pilotage intelligent de bâtiments connectés sont des priorités pour les mois qui viennent. Nous formons beaucoup de data scientists au traitement des données volumineuses avec Hadoop & Spark par exemple chez, Cofidis, MonaBanq ou Oney Banque Accord. Nos formations sont très appréciées car nous y intégrons beaucoup de pratique, les stagiaires manipulent leurs propres données et nous leur mettons ainsi le pied à l'étrier pour intégrer un projet de datascience. Aussi, nous formons ou plutôt coachons également des membres de comités de direction et des managers qui ont besoin d'être accompagnés à la transformation numérique sur la dimension de la gestion de la donnée, et notamment de la donnée volumineuse ou non-structurée.

### HÉBERGEZ-VOUS LES DONNÉES DE VOS CLIENTS ?

Nous utilisons les services d'un fournisseur de data center hautement sécurisé, dans lequel nous avons monté et opérons notre propre plateforme Cloud. L'offre en mode Cloud et SaaS est une offre que nous souhaitons développer car elle permet à nos clients de lancer des pilotes, des expérimentations, très rapidement .... Afin de tester des usages et de démontrer les apports de ces technologies dans leur business. Par exemple, sur cette plateforme, nous réalisons des projets de connaissance clients: analyse de la navigation des clients et prospects sur le site web, analyse des messages reçus et tri automatique par nature, urgence, traitement des messages vocaux, ...

### SUR QUELLES TECHNOLOGIES SKAPÁNÈ S'APPUIE-T-ELLE ?

La vague de développement de logiciels d'infrastructure Open Source est particulièrement réusie sur le monde du Big Data et du Machine Learning. Plusieurs solutions de référence de grande qualité sont disponibles. Nous avons choisi Hadoop et Spark pour le stockage distribué sécurisé et les traitements distribués. Nous utilisons les bases de données NoSQL du marché : principalement HBase. Spark, particulièrement depuis la version 2.0, propose une librairie très complète sur les algorithmes de Machine Learning modernes. Nos plateformes en production nous ont permis de démontrer la qualité, la production sécurisée et le coût de possession attractif de ces solutions.

### COMMENT VOYEZ-VOUS SKAPÁNÈ DANS LES PROCHAINES ANNÉES ?

Nous souhaitons faire rayonner le Big Data et le Machine Learning en France et en Europe et ainsi développer les savoir-faire et l'expertise. L'extension de notre offre vers du logiciel en mode SaaS contribuera à l'adoption de ces technologies. Notre ambition est de transformer le minerai brut des données en valeur raffinée pour le métier de nos clients, ayant une valeur directe sur leur activité. Nous espérons contribuer à la formation d'ingénieurs experts, de Datascientists et à l'introduction de ces technologies dans tous types de projets d'entreprises.

**IVAN SMETS**  
VP South Europe

ivan.smets@snowflake.net

Siège: - San Mateo  
www.snowflake.net



## KEY CAPABILITIES

### Data Warehouse as a Service

Snowflake allows you to focus on collecting, unifying and using your data—we take care of the rest. Snowflake's data warehouse as a service eliminates the need to deploy and manage hardware or software. There's built-in enterprise-class availability, end-to-end encryption and data protection, and no manual knobs to turn. With Snowflake, managing the data layout, maintaining indexes, tuning the system, and worrying about security are things of the past.

### Multidimensional Elasticity

Snowflake's unique architecture enables it to support any scale of storage, compute, and users. Scale storage to match any volume of data, and adjust your compute up or down on the fly – even automatically – without disruption or data redistribution. Easily create additional compute clusters to support as many workloads or users as you need without having to move or copy data. With Snowflake you can deliver the perfect amount of resources at the exact time they are needed.

### All Business Data

Snowflake automatically optimizes the storage and processing of structured and semi-structured data in a single system. Directly load semi-structured data such as JSON or Avro without transformation or a mandatory fixed schema. Query that data using SQL, even combining structured and semi-structured data in a single query. Snowflake brings the benefits of database optimization to semi-structured data for great performance.

### Performance On Demand

Our high-performance columnar database engine uses advanced optimizations to deliver differentiated performance on diverse data. Combine that with our multidimensional elasticity to scale up and down exactly when needed, and you'll see compelling performance at low cost.

## KEY BENEFITS AND BUSINESS IMPACT

### The Snowflake Elastic Data Warehouse provides:

**Maintenance free, self-service data for anyone:** Snowflake transparently takes care of infrastructure, availability, optimization, and tuning so that your analysts can focus on getting value from your data.

**“Load and go” ease of use:** Native support for structured and semi-structured data in an SQL data warehouse means you can simply load data and start analyzing it, no additional transformation required.

**Access to data at any time, for any need:** The architecture makes it possible to run diverse workloads simultaneously without contention or performance impact. Load, transform, query, and even share data at the same time.

The Snowflake Elastic Data Warehouse is priced based on the amount of data you store and the compute hours you use. To learn more and to request a trial, visit us at [snowflake.net](http://snowflake.net).



SNOWFLAKE'S UNIQUE ELASTICITY SUPPORTS ANY SCALE OF DATA, PROCESSING, AND WORKLOADS



# Snowflake Elastic Data Warehouse

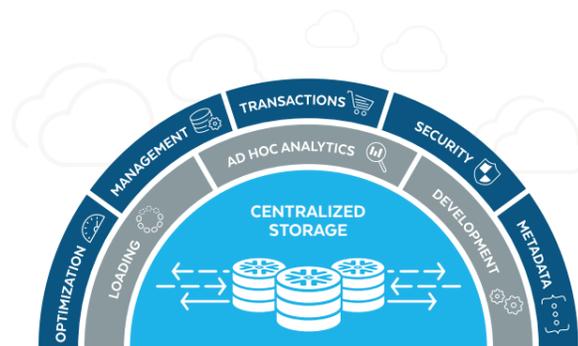
DIVERSE AND ACCESSIBLE DATA FOR EVERYONE



## Reinventing the data warehouse

Data today comes from diverse sources in diverse forms and needs to be analyzed by a growing number of people as quickly as possible. Those demands are stressing the limitations of traditional data warehouses and platforms.

Snowflake has reinvented the data warehouse, making it possible to bring all your business data together in a unified system that can support all your users and workloads. Built from the cloud up as a software service, Snowflake eliminates the cost, complexity, and inflexibility of existing solutions while allowing you to use the tools, skills and people you already have.



BUILT FOR THE CLOUD, SNOWFLAKE'S UNIQUE ARCHITECTURE PHYSICALLY SEPARATES AND LOGICALLY INTEGRATES COMPUTE AND STORAGE

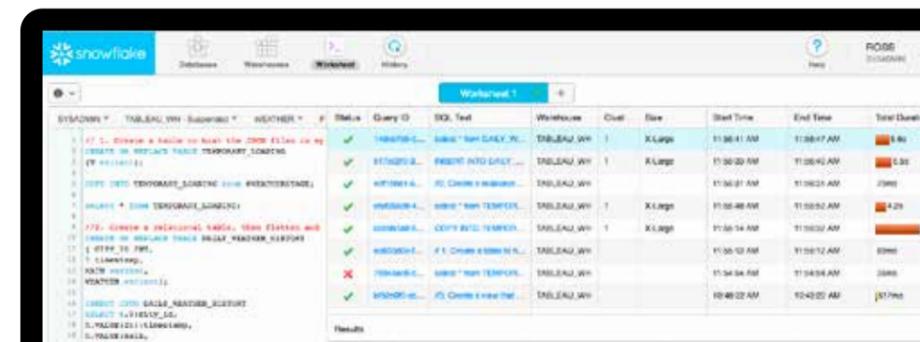
## Snowflake Elastic Data Warehouse

The Snowflake Elastic Data Warehouse is the first SQL data warehouse built for the cloud. Its patent-pending multi-cluster shared data architecture physically separates storage and compute, taking full advantage of the flexibility of cloud infrastructure in ways that are impossible with traditional data warehousing and big data solutions.

Snowflake is a data warehouse as a service that makes it possible for you to:

- Bring all of your business data together in one place for analysis.
- Allow all your analysts, applications — or even your customers — shared access to data, without contention or the need to move or copy data.
- Focus on analyzing your data without worrying about hardware, software, or database tuning.
- Utilize the SQL skills you already have, and leverage your existing ecosystem of tools — from ETL to business intelligence and more.

SNOWFLAKE CAN STORE AND PROCESS DIVERSE DATA WITH COMPELLING PERFORMANCE



“With Snowflake’s speed, we can now explore this information map at the speed of thought, and move from data, to information, to a decision, 10 times faster.”

— Chris Frederick, Business Intelligence Manager, University of Notre Dame

**SANDRA GOMES**

Directrice Commerce et Marketing

01 73 00 55 00

contact@softcomputing.com

55 quai de Grenelle  
75015 Paris France

www.softcomputing.com



soft.computing



soft.computing

**SOFT COMPUTING ACCOMPAGNE SES CLIENTS SUR LE DESIGN, LE SETUP ET L'OUTSOURCING DE SOLUTIONS BIG DATA POUR LE DATA-DRIVEN DIGITAL MARKETING. SES CONSULTANTS EXPERTS CONSEillent, METTENT EN ŒUVRE ET OPÈRENT DES DISPOSITIFS SUR MESURE DE CUSTOMER DATA PLATFORM (CDP), DIGITAL MARKETING HUB (DMH), DATA MANAGEMENT PLATFORM (DMP), MARKETING AUTOMATION (EMA), ET D'ARTIFICIAL INTELLIGENCE FOR MARKETING (AIM).**

### Au cœur de la révolution digitale

Nous sommes convaincus que nous vivons actuellement à l'échelle de la planète une révolution de même ampleur que celles liées à l'apparition de l'écriture puis à l'invention de l'imprimerie. En cinq ans, l'explosion des mobiles, des réseaux sociaux, des objets connectés et du cloud a bouleversé notre quotidien. De nouveaux business models ont émergé à des vitesses vertigineuses tandis que des pans entiers de notre économie disparaissaient tout aussi vite.

### Le déluge de données

Ces mutations profondes ne sont à notre sens que l'écume d'une vague beaucoup plus puissante liée à l'explosion des données. La plupart des ruptures que nous constatons sont à la fois la cause et la conséquence d'une explosion des données captées, stockées, analysées et échangées. Notre conviction profonde est que l'avenir appartient à ceux qui sauront dompter et valoriser ce déluge de données pour innover et se transformer.

### La transformation digitale, data-driven et customer-centric

Notre mission est d'aider nos clients à prendre ce virage pour tirer profit de l'opportunité que présente cette révolution numérique et devenir digitaux, data-driven et customer-centric dans tous les aspects de leurs relations avec leurs propres clients ; nos prestations s'inscrivent au croisement du marketing, des technologies et des sciences de la donnée autour de deux promesses :

- Imaginer, bâtir et opérer des programmes de conquête et de fidélisation client cross-canaux, rentables et innovants ;
- Concevoir, développer et déployer des solutions de Data-driven Digital Marketing & CRM performantes, pragmatiques et adaptées.

### Artificial Intelligence et Data Science : nouveau gisement de ROI

Ces trois dernières années étaient focalisées sur des Proof-of-Concepts et la généralisation de use-cases Big Data probants. Nous avons la conviction que c'est dorénavant dans la valorisation de la donnée que se situe le gros des gisements de valeur ajoutée. Nous avons constitué un pôle d'expertise unique en France

de près de 100 Data Scientists spécialisés dans les techniques statistiques, Data Mining, d'Artificial Intelligence et de Machine Learning appliquées aux problématiques marketing : segmentation, look-alike, prédictifs comportementaux, sentiment analysis, image processing, recommendation engine, optimisation mix-media, attribution et campagnes...

### Big Data : un marketing à maturité

A ce titre, nous considérons le Big Data comme de simples technologies, certes disruptives, qui élargissent l'horizon des possibles en matière de marketing et d'expérience client. Pour aider nos clients à en tirer pleinement profit, nous avons développé des compétences uniques en matière de maîtrise des processus et des technologies de collecte, d'enrichissement, de gestion, d'analyse, de valorisation et de restitution des données volumineuses, structurées ou non, en temps quasi-réel et dans le respect d'un juste équilibre entre industrialisation et agilité.

### Think, build, run MarTech and AdTech Data-driven Digital Marketing platforms

Notre offre de services couvre la définition d'architectures, la conception ou la sélection de solutions techniques, l'intégration et l'exploitation technique et business de l'ensemble des composants d'une solution Big Data pour le data-driven digital marketing :

- Customer Data Platform,
- Digital Marketing Hub,
- Data Management Platform,
- Marketing Automation,
- Artificial Intelligence for Marketing,
- Data Science for Marketing,
- Social Listening and Engagement,
- Data-driven Customer Experience,
- Media Activation,
- Data Quality,
- Data Governance,
- GDPR Compliance
- Data (1st, 2nd, 3rd party, CRM) on boarding,
- Data Visualization.

### Data Gouvernance, Data Quality, GDPR compliance : pré-requis à la généralisation

Les technologies et les usages Big Data arrivent à maturité. Comme telles, elles requièrent un cadre de règles et de processus pour en assurer le contrôle et en vérifier la bonne utilisation. Le resserrement des réglementations, l'hétérogénéité géographique des principes de respect de la vie privée, la simplicité d'ingestion de nouvelles données, la transversalité des usages dans l'entreprise, la multiplication des touchpoints et des sources de données, rendent d'autant plus critique la mise en place d'une gouvernance solide de la data et une clarification des responsabilités et des rôles autour des processus de collecte et de transformation de la data.

Exploiter tout le potentiel de la data.  
Créer des expériences Client sans couture.  
Démultiplier la performance du marketing digital.



Digital



Marketing



Big Data

Profitez de l'opportunité Big Data et développez la valeur de vos données

Data Science



Le Big Data au service de l'innovation et du marketing prédictif : Moteur de Recommandation, Machine Learning, Intelligence Artificielle ...

Big Data as a Service



Vos use cases industrialisés en organisation agile

Gouvernance des données



Vos données maîtrisées sur tout leur cycle de vie et en conformité réglementaire

### La confiance des leaders



www.softcomputing.com | contact@softcomputing.com



# Sopra Steria & le big data



## COMPOSANTE MAJEURE DU DIGITAL TRANSFORMATION OFFICE DE SOPRA STERIA, FORTE DE PLUS DE 1200 COLLABORATEURS EN EUROPE, LA PRACTICE ANALYTICS & BIG DATA :

- Fédère et anime les équipes d'experts et architectes sur les projets Big Data, Data Science, MDM et décisionnels du Groupe
- Crée les parcours de formations pour nos collaborateurs (250 formés par an sur Hadoop, Data Science, Architectures Big Data)
- Capitalise les bonnes pratiques (gouvernance des données des Data Lake intelligents, industrialisation des projets Hadoop, architecture à l'état de l'art, ...)
- Met à disposition et administre les environnements Big Data Ready pour nos clients.

L'offre Analytics et Big Data de Sopra Steria se déploie principalement sur les thématiques suivantes :

- Valorisation des données clients (parcours omni-canal, données Internet, données d'usage des produits/services, ...) pour améliorer la relation client
- Analyse des logs des capteurs et de l'IoT pour la maintenance préventive
- Analyse sémantique des données non structurées (email, voix du client, CV, ...) pour en tirer le meilleur parti
- Passage du mode réactif au mode prédictif à l'aide des nouveaux outils algorithmiques et analytiques
- Exploitation des nouveaux modes de visualisation des données pour mieux les valoriser
- Intégration des nouvelles technologies Hadoop, Appliances, In Memory, ... pour servir les enjeux et cas d'usage métier.

## AVIS D'EXPERT : BIG DATA – COMMENT INDUSTRIALISER UN DATALAKE HADOOP ?



**Eric LESTIENNE**  
Practice Data Management  
« Un Datalake se construit d'abord par une réflexion sur les services à rendre et non pas simplement par une implémentation Hadoop »

### Une réflexion orientée services plus que technologies

Nos clients mettent en œuvre Hadoop pour tester la technologie et traiter les premiers cas d'usage exprimés par les métiers. Cette approche agile est bonne, à condition de l'associer à une réflexion orientée services. En effet on peut se contenter de simplement déposer des flux de données brutes sur un Datalake, comme on peut y associer des services de traitement de données (contrôles, mise en qualité, rapprochements, dédoublonnage, calculs, agrégations, ...), services analytiques, et évidemment services de traçabilité. Cette approche services est essentielle et met en évidence que la quantité de flux de données traitée requiert une approche industrielle pour garantir la pérennité du Datalake.

### Un framework applicatif pour compléter Hadoop.

La très grande quantité de flux de données va générer une encore plus grande quantité de traitements, avec des outils très variés comme Sqoop, Hive, Spark, Kafka, Oozie... Ainsi la plateforme va orchestrer des milliers de programmes. C'est pour cela qu'il faut absolument dépasser le cadre de l'usage standard des différentes briques Hadoop pour constituer un framework applicatif qui va industrialiser le développement et faciliter la maintenance.

Ce framework doit traiter l'ingestion des données sur la plateforme, et peut être implémenté à plusieurs niveaux complémentaires : avec à minima des normes de développement et nommage des composants logiciels, des templates de développement selon les différents types de flux et technologies, un générateur de code des traitements selon un paramétrage défini, et enfin un code dynamique qui exécute N traitements via table de paramétrage.

### Des référentiels spécifiques pour compléter Hadoop

Pour être efficace, le framework applicatif doit s'appuyer sur des référentiels spécifiques peu encore aujourd'hui gérés par l'écosystème Hadoop. Selon notre expérience des grandes plateformes de gestion de données nous nous focalisons sur :

- La gestion des flux avec le référentiel des flux de la plateforme d'ingestion, la définition des contrôles opérés sur les flux et enfin la trace des contrôles effectués.
- La gestion des traitements avec le référentiel des traitements qui décrit l'ensemble des traitements gérés par la plateforme et l'agrégation de toutes les traces d'exécution issues des nombreux outils/langages qu'il est nécessaire d'assembler sous un format unifié.
- La gestion des espaces de données avec le référentiel des espaces de données, pour HDFS mais pas seulement.
- La gestion des sécurités d'accès aux données avec le référentiel des entités / utilisateurs et de leurs droits.

### Un Datalake qui assure qualité et gouvernance des données

La gestion de la qualité de données s'opère à plusieurs niveaux sur un Datalake; au-delà des contrôles sur les flux déjà évoqués, pour éviter que les contrôles soient noyés dans des centaines de programmes, il faut envisager :

- La définition des règles fonctionnelles des contrôles à opérer, en allant jusqu'à proposer un IHM aux métiers pour qu'ils paramètrent eux-mêmes les règles.
- Le reporting qualité, qui exécute des requêtes paramétrées pour publier des tableaux de bord de mesure de la qualité des données ingérées, qui seront exploités par les data stewards.

En synthèse sur le volet de la gouvernance, il faut soit se satisfaire des briques Hadoop toujours en maturation, soit les compléter par un développement spécifique afin de prétendre à l'industrialisation ; le choix reste délicat car les composants Hadoop concernés vont évidemment évoluer rapidement.

40 000  
collaborateurs

3,7 Mds €  
CA 2016

+ 20 PAYS  
EN EUROPE ET DANS LE MONDE

+ 45 ANS  
D'EXPERTISE

### DES SECTEURS CIBLÉS

Banque & Assurance -  
Secteur Public - Aéronautique,  
Défense & Sécurité - Transport -  
Télécoms & Médias - Énergie -  
Distribution

**Sopra Steria, leader européen de la transformation numérique, propose l'un des portefeuilles d'offres les plus complets du marché : conseil, intégration de systèmes, édition de solutions métier et Business Process Services.**

Sopra Steria apporte une réponse globale aux enjeux de développement et de compétitivité des grandes entreprises et organisations, notamment en relevant les défis du Big Data :

- comprendre et analyser les comportements possibles avec l'internet des objets, les réseaux sociaux...
- identifier de nouveaux usages à valeur ajoutée,
- vous différencier par de nouveaux services.

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





### HISTORIQUE

- 2006 : Splunk 1, Splunk 2 : 150 clients
- 2007 : Splunk 3 : 350 clients
- 2008 : Expansion APAC et EMEA : 900 clients
- 2009 : Splunk 4 : 1 000 clients
- 2010 : Splunk 4.1 : Première conf. internationale des utilisateurs ; 2 000 clients
- 2011 : Splunk 4.2 : Ouverture du département de recherche et dev. de Seattle ; plus de 3 000 clients
- 2012 : Splunk Enterprise 5 : plus de 4 800 clients
- 2013 : Splunk Enterprise 6 : Splunk Cloud ; Hunk ; plus de 7 000 clients • 2014 : Splunk Enterprise 6.2 : Splunk MINT ; Application Splunk for Stream ; plus de 9 000 clients
- 2015 : Splunk Enterprise 6.3 : Splunk ITSI ; Splunk UBA ; plus de 10 000 clients
- 2016 : Splunk Enterprise 6.5 : plus de 12 000 clients
- 2017 : Splunk Enterprise 7.0 : plus de 13 000 clients

### PRINCIPAUX SECTEURS D'ACTIVITÉS

- Communications
- Énergie et Infrastructures
- Finances
- Santé
- Enseignement supérieur
- Industrie
- Organisations à but non lucratif
- Services en ligne
- Secteur Public
- Distribution

### SPLUNK EN CHIFFRES

- Nombre de clients = Plus de 13 000
- Applications Splunk = Plus de 1 400
- Clients du Fortune 100 = Plus de 85
- Siège = San Francisco Employés = Plus de 3 000 dans le monde
- Cotation = NASDAQ - SPLK - investors.splunk.com

### RENDRE LES DONNÉES MACHINE ACCESSIBLES, UTILES ET EXPLOITABLES PAR TOUS.

Qu'est-ce que Splunk ? Grâce à Splunk, vos données machine prennent du sens, que ce soit sur le plan informatique, de la sécurité, commercial, ou métier. Les solutions Splunk permettent à votre département IT et aussi à toute votre entreprise d'obtenir une meilleure visibilité de vos activités et de découvrir de nouvelles perspectives. Splunk a été fondé sur la base d'une vision radicalement innovante : rendre les données machine accessibles, utiles et exploitables par tous.

### DONNÉES MACHINE

Les données machines sont l'un des segments les plus porteurs et les plus complexes des « big data ». Splunk transforme les données machines en renseignements précieux quel que soit votre secteur. C'est ce que nous appelons l'Intelligence Opérationnelle.

### LA PLATEFORME SPLUNK INTELLIGENCE OPÉRATIONNELLE

L'Intelligence Opérationnelle vous offre une vision en temps réel de ce qui se passe sur l'ensemble de vos systèmes informatiques et infrastructures technologiques, afin que vous puissiez prendre des décisions éclairées, pour vous bénéficier à l'ensemble de l'entreprise.

### NOS SOLUTIONS

Infrastructure et opérations informatiques : Détectez, analysez et visualisez les problèmes de réseau, de serveur, de stockage, de virtualisation ou d'infrastructure cloud, et mettez-les en corrélation avec les problèmes utilisateurs ou applicatifs. Vous pouvez ainsi profiter d'une surveillance en temps réel, d'alertes préventives et d'une visibilité sur la santé des services informatiques de tous les niveaux de votre organisation.

### GESTION DES APPLICATIONS

Bénéficiez d'une vision en temps réel de la pile applicative dans son intégralité - Splunk fournit des vues d'ensemble sur la performance, les transactions et les activités d'utilisateur. Avec les logiciels Splunk, les DevOps publient plus fréquemment de nouvelles versions, les équipes opérationnelles réduisent leur temps moyen de résolution, et les équipes d'ingénierie optimisent la qualité, les performances et le coût des applications.

### SÉCURITÉ ET CONFORMITÉ

Grâce à la surveillance des menaces de sécurité, l'analyse des historiques et la visualisation des ensembles de données massifs, bénéficiez d'une meilleure intelligence sécuritaire, tant pour les menaces connues qu'inconnues. L'exploration en temps réel des données d'incidents étant grandement facilitée, il est possible de mener des enquêtes complètes, d'assurer une protection préventive, et de créer des rapports ad hoc en quelques minutes.

### BUSINESS ANALYTICS

Les données machines renferment de précieux renseignements commerciaux pertinents et utiles pouvant conférer un avantage compétitif considérable. Bénéficiez d'une visibilité sur les processus métier complexes, le comportement des clients, l'utilisation des produits et les campagnes marketing digitales, le tout en temps réel.

### INTERNET DES OBJETS ET DONNÉES INDUSTRIELLES

Exploitez les données générées par les appareils connectés, les systèmes de contrôle, les détecteurs, les systèmes SCADA et autres. Surveillez les opérations, analysez les usages et intégrez ces renseignements pour produire une image globale de vos opérations métier.

### NOS CLIENTS

Plus de 13 000 clients dans 110 pays utilisent Splunk pour améliorer leurs niveaux de service, réduire les coûts opérationnels et les risques de sécurité, favoriser la conformité et créer de nouveaux produits et de nouvelles offres de service. Grâce à Splunk, elles peuvent dépasser les modes traditionnels de stockage de données, élever leur Intelligence Opérationnelle à un autre niveau, et obtenir des perspectives qui profitent à toute l'entreprise.



“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

#### 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 reverse proxy Varnish
- Logs applicatifs Tomcat

#### L'ENTREPRISE

Fondée en 2001, Karavel est le plus grand vendeur en ligne de formules de vacances en France. Karavel gère un total de 15 marques. Parmi celles-ci se trouvent notamment Promovacances, le principal site de voyages de l'entreprise, ABCroisière, agence de voyages en ligne pour les vacances en croisière, Partir Pas Cher, proposant des vacances à bon prix, Un Monde à Deux, spécialisé au contraire dans les séjours haut de gamme, et ClubPrivéVacances, qui propose des offres exclusives réservées à ses membres. Elle connaît une croissance rapide, avec en moyenne une fusion-acquisition par an. Son site principal, Promovacances, reçoit quatre millions de visiteurs uniques par mois ; en saison haute, le trafic peut atteindre 250 000 visiteurs uniques par jour. Promovacances offre plus de 10 000 séjours vers 120 destinations, et emploie une équipe de 700 personnes.

#### 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, ce qui pouvait provoquer des « abandons de panier », et des pertes de recettes.

#### 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 reverse proxy 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.



### VOUS AVEZ UNE APPROCHE CLIENT-CENTRIC ? SOYEZ CONTEXT-CENTRIC !

En questionnant ses clients au sujet de leur appréhension du Big Data, SQLI a constaté que la complexité principale n'était pas de savoir comment y parvenir, mais pour quels gains l'entreprendre. Ce challenge de poids trouve son explication dans la difficulté grandissante à maîtriser le parcours client et ses variables d'influence. La démultiplication des points de contact dans un écosystème digital toujours plus étendu a l'avantage d'accroître la résonance d'une marque mais le désavantage de complexifier la compréhension des comportements client.

**Mais alors, comment atteindre cet idéal, ô combien recherché par le commun des directions marketing, qui consiste à « adresser le bon message, au bon contact, au bon moment » ?**

A ce stade, il n'est plus possible d'enfermer les clients dans des segments statiques, comme il était d'usage de le faire. Cette ère est révolue. Les besoins et les exigences des consommateurs, qu'ils soient professionnels ou particuliers, varient en fonction du contexte dans lequel ils se trouvent.

Le transport en zone urbaine est un exemple frappant. Face à un grand nombre d'offres de déplacement, le résident fait ses choix selon plusieurs critères contextuels : suis-je pressé ? quelle heure est-il ? y a-t-il des embouteillages ? quelle est la météo ? Ces facteurs influencent le choix à l'instant T. En définitif, rien n'empêche un résident de prendre le Vélip' le matin et de revenir en taxi le soir.

**L'exploitation du Big Data permet de tirer parti de cette nouvelle donne, le challenge consistant à structurer son organisation autour d'une approche résolument data-centric. Cette transformation est incontournable pour les marques qui visent une stratégie de marketing individualisée génératrice de valeur.**

### COMMENCEZ PAR PRENDRE DU RECUL

Avant même de démarrer un projet data, il s'agit de décrypter l'écosystème dans lequel l'entreprise se trouve. La nature des données utiles, leur provenance, l'instant où elles sont générées sont autant d'éléments à analyser et structurer pour initier la création de valeur.

Un client peut tout à fait débiter ses recherches sur l'application mobile d'une marque, se rendre dans un magasin pour tester le produit avant de finaliser son achat sur le site marchand. Il contacte ensuite le service client pour faire une réclamation puis partage son expérience personnelle sur les réseaux sociaux. L'expérience promise à chaque étape du parcours doit se révéler engageante, lisse et sans couture.

La confrontation de ces informations avec des jeux de données externes propriétaires (données issues d'écosystèmes partenaires) ou publiques (météo, circulation...) permet de saisir le contexte. **Le challenge consiste à identifier précisément les données sensibles et leur degré d'implication dans le business.**

### DEVEZ-VOUS ÊTRE DATA-DRIVEN, AU CŒUR MEME DE L'ORGANISATION

Le Big Data n'est pas un outil, c'est une philosophie qui bouscule les habitudes et implique un alignement métier autour d'un même enjeu. Au-delà d'une expertise technique et fonctionnelle, le rôle de SQLI consiste à accompagner la transformation. Pour cela, il est essentiel d'appliquer une méthodologie sur-mesure, avant, pendant et après la mise en place du projet.

La gouvernance des données est une étape-clé qui permet aux entreprises de définir des insights précis et fiables et de bâtir une stratégie qui va générer de la valeur. Les consultants en marketing digital de SQLI aident à exploiter les données au travers des bons référentiels afin de faciliter les prises de décision et leur déclinaison en actions concrètes.

### MAIS POUR QUELS GAINS ?

La notion de « temps réel » que procure le Big Data permet d'atteindre un niveau de réactivité et de pertinence optimal. On note que 30 % des données des entreprises deviennent obsolètes après un an d'exploitation. A ce titre, les gains du Big Data sont considérables pour les entreprises qui souhaitent faire la différence en proposant une expérience unique et inédite, qu'il s'agisse d'image, de conquête client ou de fidélisation.

Plus votre connaissance client s'affine, mieux vous savez adresser vos clients. Un projet data n'est pas figé. Il doit être enrichi et développé pour entrer dans un cercle vertueux générateur de business.

### ET SQLI DANS TOUT ÇA ?

En tant que partenaire des grandes marques qui font le choix du digital et de l'expérience client/utilisateur, SQLI aide ses clients à adopter une vision context-centric en s'appuyant sur des compétences en conseil marketing, analytics, big data, dataviz, business intelligence et data science.

Son positionnement unique au confluent du marketing et de la technologie lui permet de répondre de façon globale aux enjeux de développement des ventes et de notoriété (marketing digital & social, parcours client, commerce connecté, data intelligence...) ainsi qu'aux enjeux de productivité et d'efficacité interne (digitalisation des opérations, entreprise collaborative, mobilité et objets connectés, CRM...).



Identification des données sensibles pour votre business



Collecte des données issues du monde réel ou digital



Construction de l'architecture en fonction de votre contexte



Transformation des données en insight via des dashboards



Prédiction des événements avec des modèles



Plan d'actions dans une logique ROIste

# VOS DONNÉES ONT DES CHOSES À VOUS DIRE

SQLI, leader de l'expérience connectée et partenaire de référence des entreprises et des marques dans la transformation digitale, allie compétences technologiques, expertise sectorielle et expériences concrètes pour vous aider à transformer vos données en insight précis et contextuels.

#### EXPERTISE × METIERS

Marketing, agent conversationnel, mobile, e-commerce, IoT...

#### EXPERTISE × SECTORIELLE

Banque, assurance, énergie/services, distribution, luxe, aérospatial...

#### EXPERTISE × TECHNOLOGIQUE

Data scientists, architectes et développeurs Big Data, Data analysts...



### SYNALTIC ACCOMPAGNE LE POINT DANS LE DÉPLOIEMENT DE SON DATAHUB, PLATE-FORME DE CENTRALISATION DE DONNÉES.

Le Point, acteur majeur de la presse hebdomadaire, souhaite disposer d'une plate-forme de centralisation de ses données afin d'avoir la possibilité de disposer d'analyses transverses, voire à 360°, de son activité.

Synaltic accompagne l'hebdomadaire pour la mise en place de ce « DataHub », s'appuyant à la fois sur des technologies Big Data (Hortonworks) et classiques (Tableau, MariaDB Columnstore, Talend). Le Point a fait appel à Synaltic à la fois pour l'accompagnement technique mais également pour sa capacité de conseil sur les technologies à envisager.

Afin de mener ce projet à bien, Synaltic a constitué une équipe pluridisciplinaire afin de traiter l'ensemble de la donnée, de la récupération des sources jusqu'à la restitution en passant par l'intégration ou l'architecture. Cette équipe « Data » a ainsi été en mesure de mettre en place une architecture basée sur un cluster Hadoop, lui-même alimenté par des jobs Talend et le gestionnaire de workflow Airflow, ainsi qu'une base de données MariaDB Columnstore, permettant des restitutions analytiques, via Tableau.

C'est lors d'une phase de Proof of Concept, qui aura duré deux mois au total, que le choix de l'architecture à mettre en place a été effectué avec les équipes du Point. Ce « POC » a également permis à l'équipe projet de Synaltic d'évoluer en fonction des besoins et de l'avancement du développement, sans diminution d'efficacité. Par exemple, sur l'architecture, trois personnes se sont succédées sans subir de latences ni de pertes, grâce aux outils utilisés, facilitant des passages de témoin rapides et efficaces. Synaltic possède désormais plusieurs personnes en connaissance du projet, pouvant intervenir pour accompagner Le Point.

#### UNE PREMIÈRE SOURCE PEUT EN CACHER UNE AUTRE

La démarche proposée par Synaltic a permis de traiter la première source de données en quelques semaines. De la découverte de la source à la modélisation décisionnelle, en passant par l'ingestion des futures données, l'équipe de Synaltic a pris soin de travailler conjointement avec celle du Point. A la fin du projet, Le Point a pu bénéficier d'un accès à des données issues de différents systèmes.

Ce "désilotage" donne au Point la possibilité d'analyser de manière transverse ses données. Les barrières entre les différents systèmes existants sont retirées pour offrir aux utilisateurs finaux une vue globale de leur activité. Les corrélations possibles permettront au Point de fournir des analyses enrichies.

Basée sur ce « DataHub », la première analyse réalisée a croisé des données provenant d'une base MariaDB Columnstore pour les agrégations et du cluster Hadoop pour les données de détails. La volonté finale du Point était d'une part, de pouvoir disposer de données agrégées et rapidement interrogeables pour fournir des analyses et restitutions aux utilisateurs métiers et d'autre part, d'avoir à portée de main toutes les données historisées pour créer des analyses ad-hoc plus poussées. L'architecture mise en place par Synaltic a permis de répondre à ces deux besoins.

A terme, le but est de mettre à disposition des utilisateurs des données, issues de plusieurs systèmes internes, qui seront traitées et analysées avec des outils de visualisation type Tableau. Ce « DataHub » deviendra alors le point d'accès unique aux données du Point.



### COMMENT VOYEZ-VOUS L'ÉVOLUTION DU BIG DATA ? QUEL EST LE PROCHAIN «BUZZWORD» SELON VOUS ?

Je pense que le big data est en train d'évoluer vers de plus en plus d'outils «Platform». Je peux citer l'exemple de Confluent, «The Apache Kafka Company», qui propose une solution dont l'objectif consiste à séduire un spectre d'utilisateurs de plus en plus large et de moins en moins technique. Autour de l'analytique, nombreuses sont les solutions dont l'offre s'étend de l'ingestion jusqu'à la diffusion des données, entre autres, sous forme de visualisations. C'est un peu le cas de Tableau, qui entre lui aussi dans cette « plateformes », avec Hyper et Maestro, la firme ouvrant son offre en alliant la data preparation à la visualisation.

La tendance du self-service se confirme. On le voit avec bon nombre d'éditeurs qui ont sorti leurs solutions à destination d'utilisateurs moins experts (Hortonworks avec SAM, Cloudera avec Data Science Workbench, Trifacta, Talend Dataprep, Talend DSC, Streamset, Alation...). Aujourd'hui, il est surtout question de former les personnes déjà en poste. On ne peut pas attendre que tout le monde devienne informaticien !

### QUELLE EST LA MATURITÉ DES ENTREPRISES FRANÇAISES ?

Au salon Viva Tech, j'ai pu voir des startups dans l'IOT et des nouvelles solutions propulsées par les données. Toutefois, peu d'entre elles ont déjà une approche «scalable» ! En effet l'architecture de leur solution ne repose pas toujours sur les briques déjà tolérantes à la panne, capables d'exister dans des PME et passant à l'échelle au sein de grands groupes. D'un autre côté, les petites entreprises n'imaginent toujours pas avoir accès à de telles technologies, alors que la mutualisation pourrait primer pour que, toutes ensembles, elles servent globalement mieux leurs clients. C'est un peu ce qu'il s'est passé dans la presse française qui s'est unie pour coconstruire un nouvel opérateur de publicité en ligne.

La notion même de big data progresse : la presse en débat, les radios consacrent de plus en plus d'émissions avec des invités de qualité. Chez Synaltic, nous avons noté une perception nouvelle de la part des DSI, qui franchissent de plus en plus le pas vers le Big Data. Elles font appel à Synaltic pour s'informer, se former, se faire une nouvelle idée de ce phénomène. Ainsi, de nouvelles solutions fleurissent au sein de ces organisations : par exemple, Trifacta facilite très clairement l'accès à la donnée dans HDFS, Talend sait l'y amener, la traiter et l'extraire, quand Tableau la sublime !

### QUELS SECTEURS SONT SELON VOUS LES PLUS AVANCÉS DANS LEUR UTILISATION DU BIG DATA ?

Les sociétés ayant déjà un pied dans le web font face désormais aux problématiques d'historisation des données. Elles sont dans l'obligation de penser ou repenser leur architecture pour rester compétitives. Quant aux grands groupes qui se sont lancés dans la digitalisation, ils embrassent désormais le Big Data ou les Data Lakes sous différentes formes. La principale problématique : avoir une vision à 360 degrés du client, suivre étroitement tous les contacts qui peuvent exister avec, en ligne de mire, la volonté de le servir toujours mieux. A cela s'ajoute l'envie féroce d'anticiper les besoins de leurs clients pour être le premier à vendre.

Il est de notoriété publique que le secteur des Télécom génère beaucoup de données. Mais, c'est aussi le cas de l'énergie : l'exemple le plus important étant Linky, sans compter toutes les données liées à la surveillance du réseau !

Mais à partir de mai 2018, tout le monde risque d'être concerné par le big data, avec l'arrivée de la GDPR qui met toutes les organisations dans l'obligation de surveiller et gouverner la gestion de leurs données.

### VOTRE CELLULE R&D EST EN VEILLE SUR LES TENDANCES TECHNOLOGIQUES, POUVEZ-VOUS NOUS PARLER DE TECHNOS QUI VONT METTRE LE SECTEUR EN ÉMOI DANS L'ANNÉE À VENIR ?

Cette année nous avons surtout mis l'accent sur les métadonnées. Nous avons cherché à ce que tous nos traitements intègrent les sondes nécessaires, afin de pouvoir suivre pas à pas ce que devenaient chaque résultat de chaque traitement. Bien entendu, les solutions open source existent pour cela ! Il y a par exemple Apache Atlas. Les métadonnées prennent un rôle très important car elles permettent de garantir que les traitements dans leurs détails répondent aux spécifications. Au-delà de capacité technique c'est surtout une nouvelle facilité de dialogue avec les utilisateurs, à qui il est plus simple d'expliquer tant les traitements que les données elles-mêmes.



### TRAVEL PLANET INTÈGRE TABLEAU POUR FAIRE ÉVOLUER LE MÉTIER DE TRAVEL MANAGER ET GÉRER 2000 VOYAGES PAR JOUR, GÉNÉRANT 4 MILLIONS DE DONNÉES

Travel Planet est l'une des entreprises les plus dynamiques de la gestion des voyages d'affaires en ligne, ou TMC (Travel Management Company). Elle gère près de 70 millions d'euros de déplacements professionnels par an et 2000 voyages par jour, générant 4 millions de données au total. Face à cette grande volumétrie de données, la TMC est donc confrontée à trois enjeux majeurs : rassembler l'offre, connaître exactement ce que consomment les voyageurs et informer en temps réel ses utilisateurs. En intégrant les tableaux de bord Tableau à ses services, Travel Planet permet à ses clients de visualiser et d'interpréter facilement toutes les données relatives à leurs déplacements en temps réel et de mesurer ainsi les dépenses. Travel Planet enregistre près de 300 data-visualisations régulières. Forte de ces résultats, l'entreprise adopte Tableau en interne, et facilite ainsi le traitement des données financières et commerciales par ses équipes.

### JUSQU'À 4 MILLIONS DE DONNÉES DE VOYAGE RENDUES ACCESSIBLES

Sur le marché hautement concurrentiel des TMC, il est primordial de définir un facteur de différenciation fort. Pour se démarquer, Travel Planet a choisi de concentrer ses efforts sur la gestion des profils clients et de favoriser leur lisibilité. Avec un système de réservation qui génère 3 à 4 millions de données chaque jour, l'entreprise cherchait un moyen de permettre à ses clients de consulter et de comprendre plus facilement les informations liées à leur profil. « Ce volume était l'un des problèmes que nous rencontrons car nous voulions assurer une restitution de l'information aux personnes chargées de gérer les réservations de voyages au sein de nos entreprises clientes », explique le directeur technologie et acquisition des marchés de Travel Planet, Tristan Dessain-Gelinet. En utilisant Tableau, Travel Planet supprime pour ses clients un des obstacles liés aux données ; celui des présentations parfois incompréhensibles ou difficiles à déchiffrer et marque ainsi davantage la différence : « Tableau est un outil d'analyse et de visualisation extrêmement facile à manipuler et ludique. Il a notamment la capacité de gérer rapidement un volume de données très important, le tout en temps réel. Tableau a totalement bouleversé la façon dont nous faisons notre travail au quotidien. », poursuit Tristan Dessain-Gelinet.

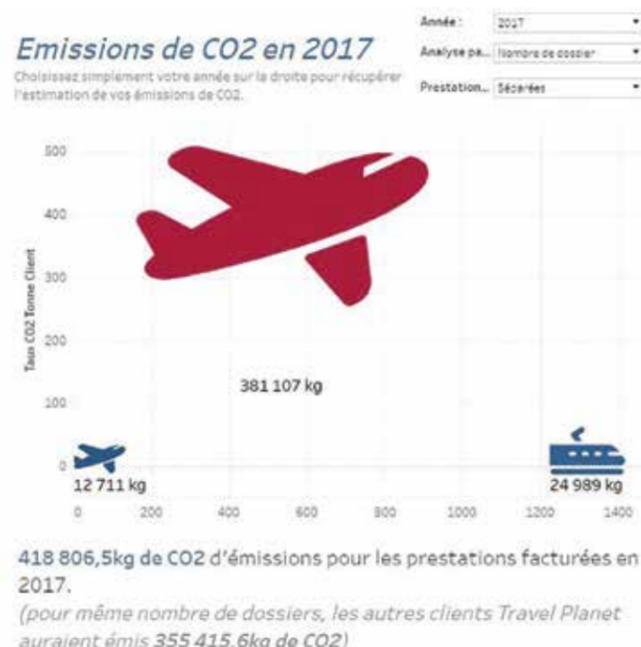
### UNE RÉVOLUTION DANS LE MONDE DU VOYAGE D'AFFAIRES AVEC 160 000 COMPTES ACTIFS

Travel Planet pense modifier profondément le secteur des voyages d'affaires où les données occupent une place prépondérante. Les 160 000 comptes enregistrés sur le tableau de bord Tableau génèrent 250 à 300 vues chaque jour. Chaque voyageur accède, via un portail, à des données filtrées selon son profil. Il choisit ensuite le type de données qu'il souhaite visualiser : consommations par mois, par type, etc. La facilité d'analyse permet aux clients d'avoir une meilleure visibilité sur les dépenses, problématique centrale dans de nombreuses organisations. Mais Travel Planet veut aller plus loin en-

core dans cette direction en se servant de Tableau pour élaborer des prédictions sur l'année : « Nous travaillons avec nos clients pour simuler les comportements à venir. En travaillant sur l'historique de la dépense 2016, on peut ainsi redéfinir la politique de voyages en 2017. Par exemple en jouant sur des leviers comme le délai d'anticipation des réservations ou la souscription d'abonnements SNCF. », affirme Tristan Dessain-Gelinet.

### UN OUTIL INTERNE POLYVALENT POUR LES 70 COLLABORATEURS DE TRAVEL PLANET

Travel Planet utilise aujourd'hui Tableau en interne. Les 70 collaborateurs de la TMC basée à Lille se servent du logiciel pour l'analyse visuelle et le pilotage des données liées à l'activité, aux marges et à la production quotidienne. L'outil est donc présent dans les différents départements, de la direction au département financier. Pour les commerciaux de l'entreprise, Tableau facilite la gestion des mailings, des factures et des relances clients.



### L'IDENTIFICATION DU POTENTIEL DES INNOVATIONS ANALYTIQUES

A l'heure de la transformation numérique des entreprises, les solutions d'analyse métier figurent toujours au premier plan sur le marché des logiciels professionnels. Cependant, le stade de maturité, qui les rendrait exploitables, n'a pas encore été franchi par nombre d'innovations technologiques.

Il est devenu primordial de différencier les tendances arrivées à maturité et qui peuvent être appliquées à des scénarios professionnels réels, et celles dont les démonstrations sont très prometteuses, mais qui sont encore à perfectionner.

### QUELLES SONT CES TENDANCES D'APRÈS VOUS ?

D'après moi, ces tendances se divisent en plusieurs catégories sur le marché, suivant leur degré de maturité : l'intelligence artificielle, le traitement du langage naturel et les analyses intégrées. Aujourd'hui l'intelligence artificielle n'a-t-elle pas déjà remplacé le travail des analystes ?

Le buzz créé par l'intelligence artificielle, qui englobe le machine learning et le deep learning, dépasse celui que connaît actuellement le Big Data. L'idée selon laquelle l'intelligence artificielle remplacerait et automatiserait les tâches analytiques manuelles n'a pas encore d'application concrète dans la plupart des scénarios réels.

### DONC DANS LE FUTUR CELA SERA POSSIBLE ?

En réalité, aujourd'hui comme demain, l'automatisation totale des workflows analytiques ne doit pas être considérée comme un but en soi. Le terme « intelligence d'assistance » serait plus approprié et rassurerait les analystes qui considéreraient l'automatisation comme une menace. La notion d'« intelligence d'assistance », qui indique que les fonctionnalités analytiques sophistiquées et intégrées et les algorithmes de machine learning viennent renforcer les compétences des analystes et des utilisateurs, séduit un nombre croissant d'acteurs sur le marché. Ces technologies intelligentes se sont révélées utiles dans l'aide à la préparation et à l'intégration des données, ou encore dans le cadre des processus analytiques comme la détection de tendances, de corrélations, de données atypiques ou d'anomalies.

### POUR CE QUI EST DE LA DEUXIÈME TENDANCE QU'ENTENDEZ-VOUS PAR LANGAGE NATUREL ? EST-CE LA MÊME CHOSE QUE LA GÉNÉRATION AUTOMATIQUE DE TEXTES ?

Le traitement du langage naturel et la génération automatique de textes sont des notions utilisées de manière interchangeable, même si elles ont des objectifs totalement différents. Bien qu'elles permettent toutes les deux des interactions naturelles avec les plates-formes analytiques, le traitement du langage naturel se charge de poser la question, alors que la génération automatique de textes s'occupe du rendu des résultats et des informations sous forme de langage naturel. Le traitement du langage naturel est plus facilement reconnaissable, dans la mesure où les interfaces de langage naturel comme Siri, Cortana, Alexa ou Google Home sont de plus en plus répandues.

### LE LANGAGE NATUREL EST DONC RÉPANDU DANS LES SOLUTIONS ANALYTIQUES D'AUJOURD'HUI ?

Les prestataires de solutions analytiques ajoutent des fonctionnalités de traitement du langage naturel à leurs produits et services pour surfer sur cette vague, et ainsi toucher un plus grand nombre d'utilisateurs, qui pourraient trouver ce genre d'interfaces plus attrayantes que les outils d'analyse traditionnels. Le traitement du langage naturel va inévitablement occuper une place centrale dans les plates-formes analytiques, mais n'est pas exploité par un panel d'utilisateurs ou d'applications suffisamment diversifié pour qu'on puisse le considérer comme une tendance majeure du marché actuel.

### QU'EN EST-IL DE LA GÉNÉRATION AUTOMATIQUE DE TEXTES ?

La génération automatique de textes est présente sur le marché depuis plusieurs années, mais ce n'est que récemment qu'elle a été intégrée dans les principaux outils analytiques pour améliorer la représentation visuelle des données. Qu'il s'agisse d'événements sportifs, de statistiques de joueurs ou des performances réalisées par des fonds de placement, de nombreux comptes rendus sont créés automatiquement à l'aide de cette technologie. La génération automatique de textes est également de plus en plus utilisée comme moyen de publication pour faciliter l'utilisation des résultats générés par des processus d'intelligence artificielle.

Les prestataires de solutions analytiques ont récemment commencé à établir des partenariats avec des fournisseurs de solutions de génération automatique de textes pour ajouter une nouvelle dimension à la visualisation de données, notamment pour identifier automatiquement les informations clés et les exprimer sous forme de récit basé sur la langue naturelle à associer à la visualisation. Bien que la combinaison des analyses métier et de la génération automatique de textes soit relativement récente, cette tendance a le vent en poupe sur le marché et ouvre la porte à de nouvelles utilisations.

### ENFIN QU'APPORTE L'ANALYTIQUE INTÉGRÉE AU MARCHÉ ?

L'analytique ne prend tout son sens que lorsque les découvertes permettent de prendre des décisions éclairées et d'améliorer les résultats. Intégrer l'analytique dans les applications et les systèmes pour qu'elle se fonde dans les activités quotidiennes des décideurs favorise l'adoption et la prise immédiate de décision.

Les prestataires de solutions analytiques modernes permettent aux entreprises d'appliquer très facilement des stratégies d'intégration et donc de démocratiser l'accès aux données dans des services où cela était auparavant impossible avec des solutions traditionnelles. Les entreprises proposent désormais des fonctionnalités similaires à leurs utilisateurs, leurs partenaires ou leurs fournisseurs, pour mieux se démarquer de la concurrence et même, dans certains cas, créer de nouvelles sources de revenus en monétisant leurs données et leurs applications analytiques.



**Industrie :** Services  
**Information :** Siège social : France • 51-200 employés  
**Cas d'usage :** Internet des objets (IoT)  
**Problématique :** Traiter en temps réel des données volumineuses  
**Produit Talend utilisé :** Talend Real-Time Big Data  
**Bénéfices :** 2,3 millions de capteurs intelligents répartis dans 3000 villes  
• 6 millions de données clients par jour • 140 millions de messages par semaine • 10 000 usagers disposant d'un accès aux données via un portail web opéré par m2ocity

« Avec Talend, nous collectons toutes sortes de données – eau, température, électricité, pollution, bruit, etc. et les analysons pour développer des services publics innovants au sein des villes connectées. » Xavier Diab, Directeur Informatique, m2ocity

### DÉVELOPPER LES SMART CITIES EN TENANT COMPTE DU BIEN-ÊTRE DES HABITANTS L'EAU EST UNE RESSOURCE ESSENTIELLE À LA VIE POUR CHAQUE HABITANT. IL EST NÉCESSAIRE DE LA PRÉSERVER.

Chaque année en France, 1 300 milliards de litres d'eau potable (soit 25 % de l'alimentation totale en eau) sont perdus du fait de fuites dans les réseaux hydrauliques des villes. Ces pertes représentent un coût pour la collectivité et donc un coût pour le citoyen. L'expertise de m2ocity répond à l'ambition des collectivités et plus généralement de l'Europe de préserver la ressource eau en réduisant ces pertes, notamment au travers de solution IoT permettant une gestion intelligente de l'eau.

m2ocity est une filiale de Veolia, créée en 2011 dans le double objectif d'industrialiser l'activité de télérelevé des compteurs d'eau et de capitaliser sur ce savoir-faire pour développer des offres répondant aux enjeux des collectivités locales et des industriels en matière d'objets connectés et de digitalisation. Devenu le 1er acteur de l'IoT en France, m2ocity intervient sur 3 grandes thématiques des objets connectés : les réseaux urbains intelligents, l'efficacité énergétique et la ville intelligente.

Qu'il s'agisse de suivi des consommations énergétiques, qualité de l'air, température, éclairage public, bruit, gestion des déchets, etc., la ville intelligente c'est l'organisation des services publics autour de la ville connectée. Cela passe par l'accès à des données volumineuses, qui doivent être structurées et exploitées au regard des usages à rendre.

M2ocity dispose de 2,3 millions de capteurs intelligents répartis dans 3 000 villes. 140 millions de messages transitent chaque semaine par le réseau de données de m2ocity.



### TRAITER EN TEMPS RÉEL DES DONNÉES VOLUMINEUSES

m2ocity a fait le choix de Talend Real-Time Big Data et a bâti une infrastructure technique dédiée reposant sur un cloud privé Microsoft Azure (garantissant des temps de traitement stables quelle que soit la montée en charge, tout en réduisant les coûts), Hadoop et SQL Server.

« Les besoins client ont évolué rapidement vers des services en temps réel. Nous devons donc être en mesure d'adapter, de faire évoluer rapidement nos traitements de données et nos modes de diffusion, » explique Xavier Diab, Directeur Informatique de m2ocity. « Nous avons besoin d'une solution Big Data capable de traiter des volumes importants de données, en temps réel ou périodiquement. Ces données alimentent des services ou portails web selon les exigences de chaque client. Le tout dans un contexte de croissance constante de la volumétrie (→30% tous les ans) et donc d'une exigence de performance et de fiabilité des solutions mises en oeuvre. Talend Real-Time Big Data a répondu à tous ces points. »

### POURQUOI TALEND ?

« Talend Real-Time Big Data offre beaucoup de souplesse pour, d'une part, gérer la donnée qui ne cesse d'évoluer (format, volume, fréquence de collecte, etc.), et d'autre part, créer de nouveaux services (maintenance prédictive, alerte temps réel), » souligne Christophe Roussel, Responsable Exploitation Infrastructure. « Si, au début, nous traitons les données en mode asynchrone, désormais nous travaillons en temps réel. Ce qui a permis de créer des services à valeur ajoutée. De plus, Talend fournit, de manière native, tous les connecteurs Cloud nécessaires, qui permettent de manière très simple et très rapide de se connecter aux sources de données.

« Tout en masquant la complexité à l'utilisateur final, Talend répond à des objectifs de performance ambitieux, en toute simplicité, » conclut Christophe Roussel.

### UNE LOCALISATION PROACTIVE DES FUITES ET UN LARGE ÉVENTAIL DE NOUVEAUX SERVICES

Aujourd'hui, 10 000 responsables de la gestion des services dans les municipalités, les entreprises et les institutions accèdent facilement aux données via un portail web opéré par m2ocity. Ces données sont restituées à la demande, le client choisissant la fréquence, le format des données et le périmètre d'analyse.

« A Nice, par exemple, on surveille la consommation d'eau de personnes âgées seules pour repérer d'éventuels problèmes :

si elle s'interrompt, on peut en déduire que cette personne a un problème, surtout en période de canicule, » explique Xavier Diab.

Au-delà des télé-relevés et de l'optimisation du réseau, m2ocity compte aussi sur les Big Data pour alerter les clients d'éventuelles menaces (fuite ou autre) et régler le problème de manière proactive. La société prévoit ainsi de s'appuyer sur des technologies de Machine Learning pour une analyse prédictive plus poussée et plus fine, notamment pour éviter les pannes. Grâce à Talend Real-Time Big Data, m2ocity est désormais en mesure de repérer des fuites avec une précision d'un mètre, ce qui évite de paralyser un quartier entier et d'engager des travaux d'urbanisme trop lourds, tout en réduisant les pertes. A Lyon, près de 500 fuites ont été localisées en 2016.

Au-delà du domaine de l'eau, d'autres services ont été imaginés par m2ocity : par exemple, enregistrer le niveau sonore dans une rue, signaler une place libre en temps réel dans un parking, ou optimiser les tournées de collecte afin d'économiser 20 % par an sur la gestion des déchets. La société a également développé une expertise dans la gestion énergétique des bâtiments : optimisation des coûts (lumière, chauffage la nuit, par ex.), bien-être des collaborateurs (rationalisation du chauffage ou de la climatisation, et contrôle de la qualité de l'air). La palette de nouveaux services potentiels paraît infinie pour matérialiser les innombrables promesses des villes intelligentes.



01 45 81 81 00  
masters@telecom-paristech.fr  
46 rue Barrault  
75634 Paris Cedex 13 France  
www.telecom-paristech.fr/bigdata



## LE BIG DATA, UNE EXPERTISE AU CŒUR DE TÉLÉCOM PARISTECH

Première grande école française d'ingénieurs dans le domaine des Technologies et des Usages du Numérique, Télécom ParisTech forme ses diplômés à innover et entreprendre dans un monde désormais numérique. Près de 1500 étudiants la choisissent chaque année dont 55 % d'internationaux. Toutes les disciplines fondamentales des Technologies et Usages du Numérique (y compris les Sciences Economiques et Sociales), sont couvertes par ses enseignements et sa recherche.

Télécom ParisTech a une forte prise sur le monde socio-économique et le pourvoir en ingénieurs, docteurs et professionnels, formés pour accompagner la transformation numérique des entreprises, de l'industrie et de la société. Télécom ParisTech transfère à l'industrie les résultats de sa recherche de très haut niveau, via des brevets, des licences et des start-ups, notamment au sein de son incubateur, ParisTech Entrepreneurs. École de l'IMT, membre fondateur de ParisTech et de l'Université Paris-Saclay, Télécom ParisTech se définit comme le Collège de l'innovation par le numérique de Paris-Saclay, dont l'ambition est de devenir l'un des premiers pôles d'innovation mondiaux.

## UNE RECHERCHE PLURIDISCIPLINAIRE UNIQUE

Portée par 170 enseignants-chercheurs permanents et 300 doctorants, sa recherche, de réputation internationale est évaluée A+ par l'AERES et est couronnée chaque année par des prix prestigieux (Médaille d'Argent CNRS, ERC, Google Awards, etc.). Ses laboratoires allient des expertises disciplinaires de très haut niveau dans les sciences et technologies de l'information et une capacité rare de mobilisation transversale interdisciplinaire.

Le dynamisme de la filière Big Data à Télécom ParisTech découle d'une recherche pluridisciplinaire autour des mathématiques appliquées, de l'informatique et des sciences humaines et sociales. Elle réunit une cinquantaine d'enseignants-chercheurs et autant de doctorants. En partenariat avec une quinzaine d'entreprises, l'école s'investit dans deux chaires de recherche dans le domaine du Big Data, ainsi qu'une chaire d'enseignement. La chaire Machine Learning for Big Data conduit ses recherches à l'interface des mathématiques et de l'informatique. Le machine learning vise à élaborer des algorithmes permettant aux machines d'apprendre automatiquement à partir des données et ainsi d'améliorer leurs performances. La chaire Valeurs et Politiques des Informations Personnelles traite des aspects juridiques, techniques, économiques et philosophiques qui concernent la collecte, l'utilisation et le partage des informations personnelles.

## UN ENVIRONNEMENT D'ENSEIGNEMENT EXCEPTIONNEL

Télécom ParisTech a bâti un écosystème unique, basé sur une forte interaction entre formation, recherche et innovation et qui se caractérise par l'accueil régulier de conférences et de colloques internationaux, un incubateur qui héberge plusieurs start-up issues du Big Data, des partenariats avec les entreprises clé du secteur... Cet environnement a permis au Mastère Spécialisé® « Big Data » d'être le premier du genre en France et de faire figure de référence dans le domaine. Par la suite, l'école a mis en place une large gamme de formations pour tous les besoins, notamment dans le cadre de Télécom Evolution. Télécom ParisTech entretient

des liens étroits avec l'industrie : une dizaine d'entreprises sont membres du comité de veille et de perfectionnement des formations et plusieurs dizaines participent aux enseignements, études de cas, mises en situation professionnelle, tables rondes et séminaires. Elles proposent également stages, thèses professionnelles, projets de groupe mais aussi données réelles anonymisées. En 2017, l'école a créé, en partenariat avec des entreprises, deux nouvelles chaires d'enseignement sur la science des données.

## PRIORITÉ À L'INNOVATION

L'incubateur ParisTech Entrepreneurs a accueilli en 15 ans plus de 300 start-up innovantes du numérique. Beaucoup d'entre elles utilisent les technologies du Big Data, que ce soit dans l'analyse et le traitement des données, le marketing et le e-commerce, l'industrie, les transports, la santé, la high-tech, les services Internet, la finance... L'entrepreneuriat est en effet fortement encouragé au sein des différentes filières de l'école. Télécom ParisTech participe également à la plateforme souveraine Teralab qui a pour but de lever la barrière de l'adoption des technologies Big Data. Infrastructure puissante, flexible et de confiance, Teralab accompagne ses utilisateurs sur les plans technologie et juridique. Elle héberge des projets portés par les entreprises, les start-up ou les organismes de recherche et de formation.

## TÉLÉCOM ÉVOLUTION, DES COMPÉTENCES TOUJOURS A LA POINTE

Télécom Evolution est le centre de formation continue commun à IMT Atlantique, Télécom ParisTech et Télécom SudParis. Il accompagne les entreprises et les organisations dans la montée en compétences de leurs collaborateurs autour des grandes problématiques du numérique : Internet des objets, Big Data, cybersécurité, transports intelligents... Télécom Evolution propose une gamme complète de formations depuis l'initiation jusqu'à l'expertise, dans une grande variété de modalités pédagogiques, du MOOC à la formation certifiante, dispensés par les meilleurs experts et des professionnels reconnus.



Les tables rondes entreprises font salle comble



Gilles Babinet, Digital Champion de la France auprès de la Commission Européenne

## MASTÈRE SPÉCIALISÉ® BIG DATA :

### GESTION ET ANALYSE DES DONNÉES MASSIVES



Un Bac+6 pour des diplômés en poursuite d'études ou en reconversion, accrédité par la Conférence des Grandes Écoles. Plus de 600 heures sur 9 mois de cours, travaux pratiques, séminaires, et un projet de groupe « fil rouge » proposé par une entreprise.

Un stage et une thèse professionnelle en entreprise sur 4 à 6 mois concluent un programme conçu et enseigné par des spécialistes du domaine, en forte adéquation avec les besoins des entreprises.

Le MS Big Data combine un large socle de connaissances techniques et une approche complète des aspects juridiques, économiques, business.

**91 % des diplômés** ont trouvé un travail facilement ou très facilement, et **75% avant la fin de leur stage\***.

\*Étude Télécom ParisTech, novembre 2016



## CES DATA SCIENTIST

### (CERTIFICAT D'ÉTUDES SPÉCIALISÉES)

**Pour les professionnels en poste** qui souhaitent une certification en science des données (collecte, stockage et calcul distribués, analyse statistique, visualisation).

**Très opérationnel**, le CES permet la maîtrise des techniques de gestion et d'analyse des Big Data et des principaux algorithmes du machine learning.

**12 sessions de 2 jours sur 10 mois** : cours, travaux dirigés, travaux pratiques, témoignages de professionnels et une plateforme d'e-learning.



## MOOC : FONDAMENTAUX POUR LE BIG DATA

**Gratuit, flexible et indispensable** pour valider une inscription au **Mastère Spécialisé®** ou au **CES**.

**En 6 semaines**, révisez les prérequis essentiels dans les domaines de l'analyse, de l'algèbre, des probabilités, des statistiques, de la programmation Python et des bases de données.

**7 modules**, des vidéos, des quizz et un projet final, et plus de **7 000** inscrits par session !



## FORMATIONS INTER-ENTREPRISES

- Big Data : enjeux stratégiques et défis technologiques
- Big Data : panorama des infrastructures et architectures distribuées
- Big Data : premiers succès et retours d'expérience
- Data Science : introduction au Machine Learning
- Data Science avec Python **[NOUVEAU]**
- Data Science avec R **[NOUVEAU]**
- Data Science dans le Cloud : Big Data, statistiques et Machine Learning
- Droit, RGPD et protection des données dans le big Data **[NOUVEAU]**
- Extraction d'informations du Web
- Opinion mining : e-reputation et recommandation **[NOUVEAU]**
- Text-Mining
- Visualisation d'information (InfoVis)



Toutes nos formations sur : [www.telecom-paristech.fr/bigdata](http://www.telecom-paristech.fr/bigdata)

01 81 89 14 08  
 Teradata.France@teradata.com  
 Antony Parc 1 - 2-6 place du Général de Gaulle  
 92160 ANTONY France  
 www.teradata.com

# TERADATA



Teradata aide les entreprises à créer plus de valeur avec leurs données grâce à une stratégie « data et analytic centric ». Nos solutions analytiques Big Data, nos services, nos experts métiers et architectes aident votre société à obtenir des avantages compétitifs durables grâce à l'exploitation agile et industrielle des données internes et externes pour connaître leurs clients et leurs processus métiers tout en se concentrant sur les bénéfices.

### LUFTHANSA S'ENVOLE POUR L'EXCELLENCE OPÉRATIONNELLE

Lufthansa a pour mission de « s'imposer comme le premier choix des passagers, employés, actionnaires et partenaires dans le transport aérien » en misant sur les données et leur analyse, pour améliorer l'efficacité opérationnelle de ses nombreuses marques. Face à une farouche concurrence, une concentration des acteurs, le transport aérien a vu sa complexité s'accroître, rendant plus laborieux les gains d'efficacité et l'identification de nouvelles opportunités. De plus, les décisions opérationnelles doivent être prises en temps quasi réel, avec un marché volatile. Les décideurs arbitrent ainsi entre risque-coût-perse pour qu'infrastructure opérationnelle et stratégie restent en parfaite adéquation.

### ZOOM SUR LES FACTEURS CLÉS DE RÉUSSITE

Pour atteindre l'excellence opérationnelle, Lufthansa s'efforce de maximiser le potentiel de son personnel et de ses équipements apportant des améliorations continues tout en adaptant l'exploitation aux besoins changeants. Afin de trouver un équilibre entre exploitation d'informations pour maximiser les bénéfices et réduction des coûts sans sacrifier le service client, Lufthansa se concentre quotidiennement sur trois indicateurs de performance (KPI) stratégiques : Maximisation des recettes • Minimisation des coûts • Maintien du niveau de satisfaction. Pour les mesurer, Lufthansa a entrepris la tâche colossale de créer un référentiel de données commun à l'ensemble de ses activités en faisant voler en éclats les silos pour intégrer les données dans un entrepôt de données d'entreprise Teradata Data Warehouse, autorisant l'extraction et l'analyse d'informations exploitables pour soutenir ses processus décisionnels. Cela a permis de mieux comprendre les secteurs d'activité transverses, notamment : Recettes • Réservations • Marketing • Efficacité des équipages • Liaisons aériennes • Calendriers des vols • Kérosène. En analysant les données des tarifs, itinéraires, avions et segments de clientèle, Lufthansa peut évaluer sa « juste part » de réservations sur un marché très évolutif. Ainsi, si le marché progresse de 5 %, la compagnie souhaite vérifier que le groupe voyageurs d'affaires a également progressé de 5 %, ou plus, ou moins ?

### L'ENVOI DES OFFRES SPÉCIFIQUES

Analyser des données intégrées permet de proposer des offres spécifiques sur l'ensemble de ses marques, lignes et départements. Cette vue globale de sa relation commerciale avec chaque client, conjuguée à l'octroi optimal des avantages, aide à orienter les ventes et les stratégies associées. Source unique d'information de l'entreprise, l'entrepôt de données fait partie d'un écosystème analytique utile à de multiples domaines d'activité. Chaque département utilise les informations pour répondre aux besoins de centaines de marchés locaux, identifier les préférences des clients et déterminer la valeur

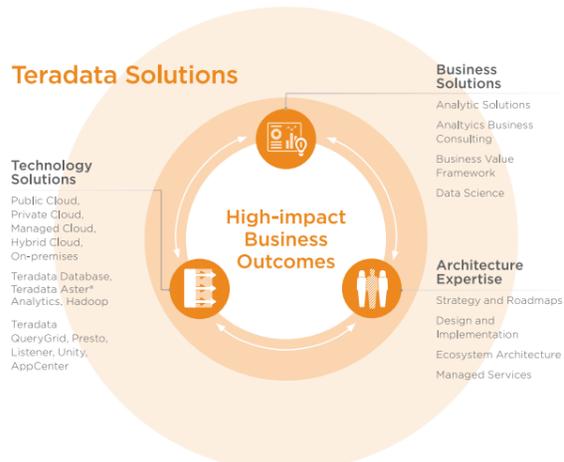
en devises locales. Ainsi, Lufthansa a mis en œuvre un processus pour dialoguer de manière proactive avec les clients victimes d'un « dark flight » - vol où, les passagers ont été privés de films. Un dédommagement automatique est offert aux passagers concernés — avant toute réclamation. L'analytique des données joue ainsi un rôle plus important pour garantir la cohérence des échanges avec tous les points de contact, toutes marques confondues.

### ATTEINDRE DE NOUVEAUX SOMMETS GRÂCE À L'ANALYTIQUE

Avec ces moyens analytiques, Lufthansa peut relever les défis économiques pour atteindre de nouveaux niveaux d'excellence. L'entrepôt de données fait table rase des silos en intégrant les données de toutes les entités puis met ces données à la disposition de toutes les parties prenantes. Cela offre une meilleure gestion des marques, met en lumière les gains d'efficacité et un accès à des informations concurrentielles qui encouragent la réflexion stratégique. L'analytique a joué un rôle crucial dans le remplissage des avions tout en réduisant efficacement les coûts. Alors que de nouveaux circuits de distribution émergent supplantant les systèmes de réservation traditionnels, il est désormais possible de vérifier le meilleur prix et les détails des vols sur Internet avant de réserver un billet. Pour valoriser ses prestations complémentaires, Lufthansa compte sur l'analytique pour créer des offres répondant aux besoins individuels des clients, en interagissant avec eux sur les réseaux sociaux. En termes d'analytique des données Lufthansa a lancé un programme stratégique composé de sept « champs d'action », applicables aussi bien à chaque segment d'activité qu'à l'ensemble de la compagnie : • Priorité au client • Concepts de croissance • Innovation et numérisation • Organisation rationnelle • Culture et leadership • Pilotage par la valeur • Amélioration de l'efficacité

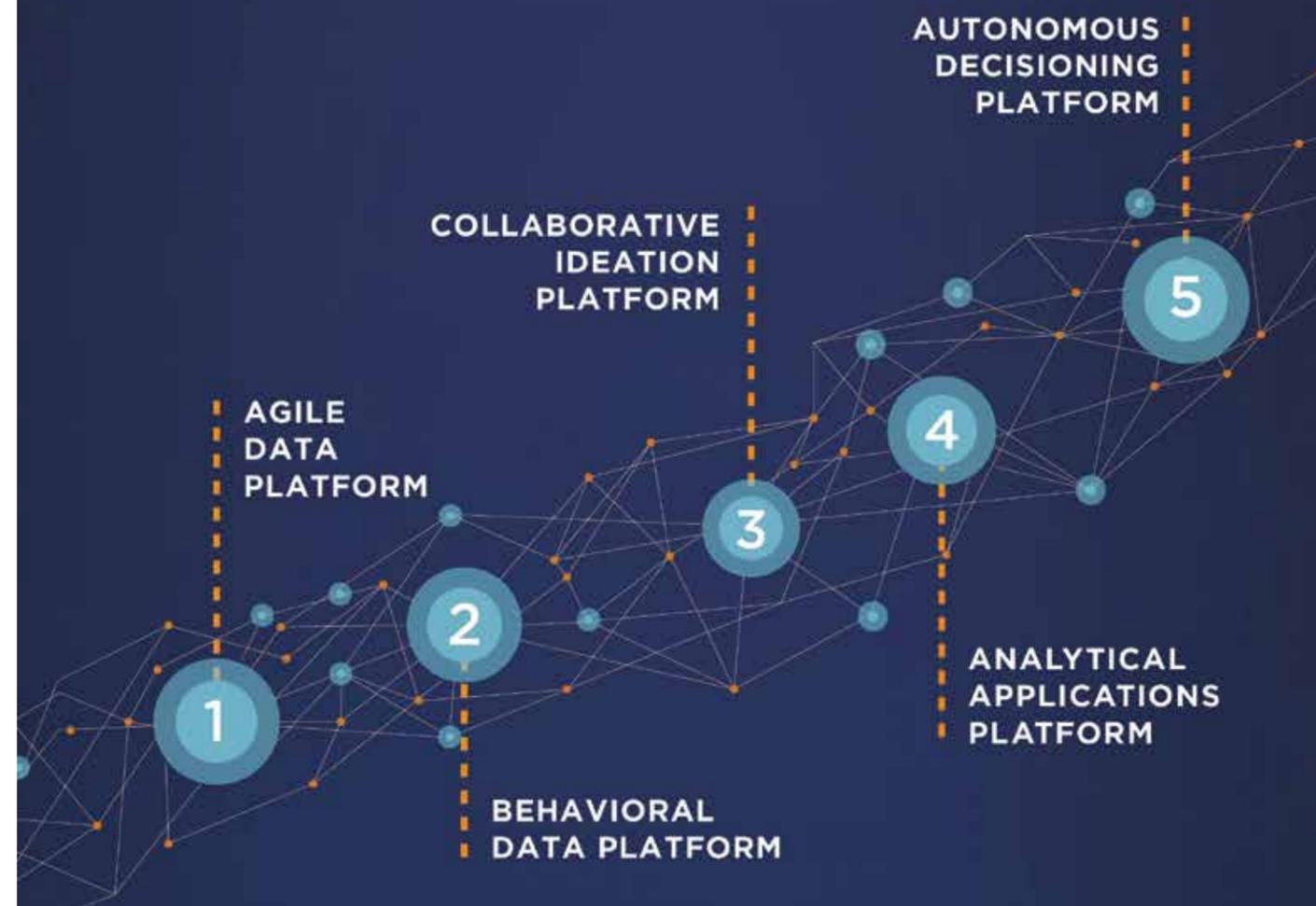
### VISER L'EXCELLENCE OPÉRATIONNELLE

Atteindre un tel niveau d'efficacité exige une excellence opérationnelle, induisant l'utilisation d'analyses dynamiques réalisées sur un large éventail de données offrant une visibilité sur tous les aspects des opérations de l'entreprise. Cela requiert également une visualisation avertie pour impacter les stratégies commerciales, apporter des améliorations continues et adapter les modèles. L'excellence opérationnelle libère de l'approche en silo des données pour impliquer toutes les parties prenantes. Pour Lufthansa, l'excellence opérationnelle l'aide à atteindre de solides résultats financiers, favoriser une croissance positive et consolider sa position stratégique dans une industrie ultra concurrentielle.



# THE SENTIENT ENTERPRISE

## A REVOLUTIONARY APPROACH TO AGILE DECISION MAKING



### 1 UNE PLATEFORME DE DONNÉES AGILE

Les fondations pour offrir les capacités et processus analytiques. Il est primordial de passer à un framework équilibré et décentralisé en intégrant des architectures de cloud hybrides, des vues métiers virtuelles et les outils associés pour créer une plate-forme technologique agile évolutive.

### 2 UNE PLATEFORME DE DONNÉES COMPORTEMENTALES

En plus des transactions, capturer des quantités massives de données comportementales clients et machines pour comprendre la raison de certains comportements.

### 3 UNE PLATEFORME D'INNOVATION COLLABORATIVE

La collaboration, la « gamification » et les connexions sociales au sein de l'entreprise peuvent être exploitées conjointement pour relier les acteurs humains et les données de manière rapide et autonome qui surpassent les méthodes traditionnelles de métadonnées centralisées.

### 4 UNE PLATEFORME D'APPLICATIONS ANALYTIQUES

Déployer des fonctionnalités analytiques et des applications en self-service à destination des acteurs métiers. Favoriser et démocratiser l'utilisation des données pour augmenter « l'écoute » de l'écosystème de l'entreprise et actionner les changements sur la base de connaissances factuelles.

### 5 UNE PLATEFORME DE DECISION AUTONOME

Aller au-delà des technologies prédictives et augmenter le déploiement des algorithmes, du machine learning, de l'intelligence artificielle pour analyser toutes les données afin de détecter avec peu ou pas d'intervention humaine les tendances, les modèles et les cas particuliers ou anomalies.

# TERADATA

[www.teradata.com/sentient-enterprise](http://www.teradata.com/sentient-enterprise)



## TIBCO SOFTWARE, PARTENAIRE DE L'ENTREPRISE DIGITALE

Avec TIBCO Connected Intelligence Cloud, l'entreprise intègre toutes les données de son écosystème, optimise et accélère prises de décisions et actions. Les technologies TIBCO interconnectent individus, systèmes, équipements et API ; elles capturent les données en temps réel et augmentent l'intelligence numérique des entreprises à travers le monde font confiance à TIBCO pour offrir des expériences convaincantes à leurs clients, dynamiser leurs opérations et stimuler l'innovation.

## INTELLIGENCE AUGMENTÉE

La plateforme analytique TIBCO Spotfire permet aux analystes et aux utilisateurs opérationnels d'avoir une vision temps réel sur les données de l'entreprise et de son écosystème. En 2017, TIBCO a acquis Statistica qui enrichit la plateforme en permettant aux data scientists d'exploiter facilement la puissance du machine learning notamment pour l'analyse de données issues de l'IoT.

## TOUT INTERCONNECTER

La plateforme TIBCO, associant solutions d'intégration hybride, BPM-Case Management et API Management, constitue le socle de l'entreprise digitale reliant systèmes, processus, collaborateurs et objets, dans le cloud ou on-premise.



## CONNECTED INTELLIGENCE

## INTELLIGENCE CONNECTÉE TRANSFORMEZ VOS DONNÉES EN ACTIONS

Avec TIBCO Connected Intelligence, l'entreprise renforce sa stratégie numérique à partir de données, de nouveaux services et de la capacité de comprendre et d'agir rapidement. La transformation numérique modifie les activités de l'entreprise comme elle change la vie privée. Nous nous connectons à des sources de données pour prendre des décisions d'achat ou gérer notre santé. En habilitant des collaborateurs à comprendre, décider et agir sur les données, la différence se fait à tous les niveaux de l'entreprise.

### LES BÉNÉFICES

#### • CONNAÎTRE CHAQUE CLIENT

Les clients attendent des expériences personnalisées, engageantes et pertinentes. Lorsque le système d'information interconnecte tout, vous êtes en mesure de les accompagner dans des expériences convaincantes sur un ou plusieurs canaux. Augmenter votre intelligence par des données contextuelles vous permet de personnaliser l'expérience client et de renforcer vos relations avec chacun.

#### • INNOVER AVEC DE NOUVEAUX BUSINESS MODELS

En interconnectant les composants d'un système d'information, votre entreprise assemble ses services, ses données et ses canaux pour élargir ou redéfinir ses business models. Vous enrichissez votre palette de services et captez de nouvelles audiences. L'intelligence augmentée vous aide à identifier les meilleures opportunités et à accroître l'impact de votre stratégie.

#### • EXCELLENCE OPERATIONNELLE

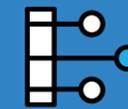
Utilisez des analyses prédictives et des visualisations pour accroître votre efficacité opérationnelle en apprenant plus rapidement et en innovant sur la base de données temps réel notamment dans les domaines de la maintenance prédictive ou de l'optimisation d'actifs.

## INTERCONNECT EVERYTHING

## AUGMENT INTELLIGENCE



Data & Systems



API Management



People & Processes



Reporting & Dashboards



Visual Analytics



Streaming Analytics

### Avec TIBCO Spotfire, la MAIF donne le pouvoir de la donnée à ses utilisateurs

« Adopter TIBCO Software, c'est donner à nos utilisateurs métiers un accès facile aux données, la possibilité de les manipuler aisément et d'en tirer le meilleur dans le cadre de leur activité quotidienne. »

« Vous imaginez bien qu'une organisation de notre taille exploite depuis longtemps des outils de business intelligence. Le problème avec les solutions habituelles, c'est leur technicité. Tant qu'il s'agit de produire du reporting chiffré, des tableaux, ces solutions sont adaptées. En revanche, quand l'objectif consiste à élaborer des solutions analytiques plus visuelles sans imposer de prérequis techniques aux utilisateurs, les difficultés surgissent. »

Stéphane Renoux, *Responsable de projet, MAIF, 1er assureur du secteur associatif*

### Vestas confère un nouveau souffle à l'éolien avec TIBCO Spotfire

« Notre relation avec TIBCO a commencé lorsqu'il est devenu manifeste que nous avions besoin d'un partenaire possédant l'expérience nécessaire pour traiter le potentiel du volume de données contenu dans notre supercalculateur. TIBCO, c'est bien plus que Spotfire, c'est aussi les données en streaming, le bus d'entreprise, la plateforme Fast Data. Tout cela revêt une grande valeur pour nous. »

S.J. Knudsen, *Sr Data Scientist, Vestas Wind Systems, leader mondial de l'énergie éolienne*





Nous créons le système analytique nommé TIMi. TIMi couvre tous les besoins analytiques des organisations, depuis les besoins élémentaires (tels la collecte de données, la création de KPI et de dashboards) jusqu'à la création de modèles prédictifs de haute précision et le Big Data. TIMi élimine les peines généralement associées aux systèmes analytiques et les remplace par une interface autonome, réactive, interactive, visuelle, conviviale et en self-service (sans code). L'exploration de vos données n'a jamais été aussi facile et ludique! Avec TIMi, les entreprises peuvent capitaliser sur leurs données pour développer de nouvelles idées, prendre des décisions commerciales critiques plus rapidement et plus facilement que jamais ("TIMi: Faster Predictions, Better Decisions"). Notre mission est de : «Fournir à nos clients des outils conviviaux, rapides et productifs pour l'analyse prédictive, le BI et le Big Data pour résoudre les problèmes réels des entreprises en quelques jours et pas en quelques mois». «To boldly go where no dataminer has gone before».

#### LES PRINCIPAUX OBSTACLES RENCONTRÉS LORS DE PROJETS BIG DATA/ADVANCED ANALYTICS SONT :

**1. Coût d'Infrastructure Prohibitifs : Sur un simple laptop, TIMi analyse des bases de données contenant des milliards d'enregistrements et des dizaines de milliers de variables/colonnes.**

Nos clients, MTN et Vodaphone, sont les 2 plus grosses multinationales dans le domaine des Telecoms en Afrique. Leurs opérations s'étendent sur plus de 20 pays chacun. L'analytique de presque toutes ces opérations est basée sur TIMi. C'est, à partir des données brutes (logs des Call-Data-Record-CDR), TIMi (re)crée automatiquement, chaque jour, des «vues clients à 360°», des datasets analytiques (avec plus de 2000 variables, dont des variables SNA) et des modèles prédictifs. Chaque jour, dans chaque pays, TIMi manipule plusieurs dizaines de milliards de records (CDR). Les populations analysées varient entre 6 (Tanzanie) et 50 (Kenya) millions d'abonnés. Nos concurrents (S?S, I?M, etc.) tentent vainement de reproduire nos résultats. Typiquement, il leur faut un cluster de plus de 100 machines, et deux mois de prestations/calculs pour délivrer des résultats inférieurs à un ou deux laptops équipés de TIMi pendant quelques heures (de 6 à 18 heures, dépendant de la population analysée).

**2. Complexité très élevée : Les projets «advanced analytics» ou «big data» impliquent l'écriture de dizaines de milliers de lignes de code dans des langages (Scala, R) que seuls des spécialistes introuvables et onéreux comprennent.** Créer des modèles prédictifs de qualité nécessite aussi de faire appel à des profils introuvables (PhD en Stats). Par opposition, avec TIMi, tout est en self-service «à la souris». Plus précisément, les travaux de data management, la création de dashboards et de modèles prédictifs, sont 100% sans code et accessibles à un utilisateur avec un esprit analytique mais sans formation en IT ou en Statistique. Comme le démontre nos résultats exceptionnels lors de compétitions de «Machine Learning» (KDD Cups, Kaggle), nos modèles prédictifs sont parmi les meilleurs de la planète. Nos utilisateurs nous disent que la complexité d'utilisation de TIMi est similaire à celle de Excel. Notre client, AXA Assurance, souhaitait créer un Data Lake basé sur Hadoop. Ils ont été confrontés à deux obstacles. Il était très difficile (voire impossible) de trouver des spécialistes Spark (Spark est le seul outil plus ou moins performant dans Hadoop pour réaliser les tâches de data management) et il était très difficile d'intégrer dans Hadoop leurs données de type «Legacy» (par ex.: fichiers Mainframe AS/400, fichiers .sas7bdat). TIMi apporte

une solution optimale à ces deux problématiques : Avec TIMi, en quelques clics de souris, un utilisateur business crée une transformation de données complexe 100% native à Hadoop (TIMi est la seule solution de l'écosystème hadoop qui est codée 100% en code natif avec un accès natif à HDFS et parquet) qui s'exécute à une vitesse de (typiquement) 10 à 100 fois supérieure à Spark (sur un hardware réduit ou équivalent). Les coûts de maintenance ont aussi été réduits à néant (ou presque).

**3. Temps d'implémentation déraisonnable : Quelle est la réactivité de vos processus IT par rapport à un environnement business en perpétuelle mutation ?** TIMi vous donne la possibilité d'investir facilement vos données pour y trouver la perle rare. Vos collaborateurs ont déjà pleins d'idées, potentiellement très lucratives. Maintenant, ils peuvent les implémenter/tester pratiquement instantanément avec TIMi, sans impact sur votre infrastructure et votre DSI. Notre Client, VOO/BeTV, a créé en 24 mois plus de 6000 transformations de données «Ad Hoc» complexes, pour réaliser des simulations (scénario «what if...»), pour automatiser des flux de données gérant des fournisseurs, gérant leur parc hardware (décodeurs, antennes, capteurs IoT), etc. Ces transformations de données représentent plus de 70 années-hommes de développement avec tout autre technologie que TIMi. Elles ont été réalisées par des personnes «business» sans background en IT (en «self-service»). Des témoignages de VOO : «We initiated with TIMi a data driven business revolution with 15% sales cost-cutting», «Before TIMi, we were maybe using 10% of our data. Now, we use 99%», «TIMi helped us reduce by 10% the churn on the customer-segment with the highest Churn-Rate». Parmi ces 6000 transformations, un petit millier de transformations sont maintenant en production journalière (industrialisation en un clic!) et délivrent une connaissance client étendue et une grande valeur à VOO/BeTV.

**4. Capacité de stockage et de traitement des données limitées : Combien payez-vous pour avoir un accès rapide à vos données ?** Quelle fraction de vos données pouvez-vous réellement manipuler ? Combien de Cluster Hadoop avez-vous ? Quelle est votre facture Amazone/Azure ? Cessez de vous poser ces questions : TIMi possède un format de stockage propriétaire fortement compressé qui résout tous vos problèmes de (coût de) stockage et de traitement des données.

Par exemple, deux années de données à l'échelle la plus fine (c'est à dire une ligne de donnée contient 17 attributs : C'est l'achat d'un produit précis par un client dans un magasin d'un acteur de la grande distribution (Colruyt) avec 1 millions de clients récurrents et 251 supermarchés pèsent seulement 24GB avec TIMi (contre 0.5 TB avec S?S et 3TB avec Ter?d?t?). N'importe quel laptop peut stocker (et manipuler) ce volume.



#### TIMi en résumé

- ▶ Sièges sociaux à Bruxelles et Bogota.
- ▶ Supportée par des firmes de consultance reconnues telles que Deloitte, PwC, EY, Real-Impact, etc.
- ▶ Utilisée par de nombreuses companies à travers le globe (les données de plus de 300 millions de personnes manipulées tous les jours avec TIMi).
- ▶ Régulièrement parmi les équipes gagnantes de compétitions internationales de datamining (KDD cup, Kaggle)

#### Turbo-Chargez vos Analystes !

TIMi est une solution innovante pour les sociétés qui souhaitent utiliser leurs données comme un avantage concurrentiel et capturer la valeur cachée dans leur base de données. TIMi révolutionne les domaines de l'Analytics, de l'Analyse Predictive et du *Advanced Analytics* en supprimant les quatre obstacles principaux à l'adoption de telles technologies :

- **Rapide & précis** : En quelques secondes, TIMi calcule des milliers de modèles, sélectionne et délivre automatiquement le modèle le plus robuste et le plus précis. Anatella effectue des transformations complexes sur plusieurs Tera de données en quelques minutes.
- **Apprentissage court & aisé** : Un business user avisé peut aisément construire des modèles de qualité après une simple formation et cela sans qu'un doctorat en «Data Science» ne soit nécessaire.
- **Coût d'infrastructure négligeable** : À l'aide d'un simple PC standard de bureau, TIMi analyse et manipule des bases de données contenant des milliards d'enregistrements et des milliers de variables.
- **Stockage réduit** : Grâce à des algorithmes de compression originaux, TIMi manipule, par exemple, un volume de données de 5 TB en BDD en utilisant seulement 100GB d'espace disque.

#### TIMi : Quatre outils pour révolutionner l'advanced analytics

To boldly go where no data-miner has gone before !



##### TIMi modeler : Des prédictions, plus vite

Transformez les coûts liés à la gestion de vos données en des investissements à fort ROI.

En un simple click de souris, TIMi modeler génère des modèles prédictifs plus rapides, plus performants et plus actionnables que tout autre solution pour l'optimisation des ventes, des inventaires, de la loyauté client, etc.



##### Anatella : l'ETL facile

Anatella simplifie l'accès à l'information en rendant le processus d'extraction et de transformation de vos données plus facile et plus rapide.

À l'aide d'Anatella, connectez vous à toute source de données (par ex. HADOOP) et préparez vos données en vue d'analyses (simples ou «advanced») et/ou de tout reporting/dashboarding.



##### Stardust : Exploration Multivariée

Une nouvelle façon de découvrir vos données : Avec l'outil d'exploration de données 3D et l'outil de segmentation multivariée, identifiez les groupes de clients clés, les valeurs aberrantes, et analysez facilement les transactions (par ex. pour y découvrir des comportements aberrants).



##### Kibella : Self-service Dashboarding

Nombre illimité de designers, de servers et de consommateurs de dashboards.

Outil de design léger & intégré au navigateur web.

Haute sécurité et déploiement aisé de part l'intégration directe dans Apache.

En conjonction avec Anatella, pas de limite sur le volume de données visualisé.

#### Quelques références



**DEBORAH BARON**

VP OF MARKETING

+1 510-208-4370

marketing@tomsawyer.com

1997 EL DORADO AVENUE  
BERKELEY, CALIFORNIA 94707

www.tomsawyer.com



## L'APPLICATION GAIA D'AIRBUS: UNE SOLUTION MULTI-VUES PERMETTANT UNE APPROCHE INGÉNIEURIE SYSTÈME MBSE

**CONTEXTE :** Airbus compte plus de 9000 avions en opération. Le groupe a une activité de conception, de fabrication et d'assemblage d'avions, principalement en Europe. Une hausse de 20% du trafic aérien est attendue sur les vingt prochaines années. Airbus a déjà enregistré des commandes représentant une dizaine d'années de production et une augmentation de près de 75% de la flotte actuelle.



**BIG PRODUCTS – BIG DATA :** Un Airbus A380 compte approximativement : 3 millions de pièces, 500 kilomètres de câbles, 100 millions de lignes de codes de logiciels et systèmes embarqués.

Pour répondre à la demande et aux exigences en termes de sécurité et de confort des passagers, Airbus devait :

- augmenter les cadences de productions,
- réduire les durées de conception et de développement,
- atteindre une fiabilité opérationnelle maximale pour la sécurité des passagers.

Pour réaliser la conception et l'entretien de leurs avions, des milliers d'outils et de modèles de données sont nécessaires. Les informations d'Airbus sont stockées dans 15 ensembles de données différents, eux-mêmes associés à de nombreuses applications autonomes.

**UNE NOUVELLE APPROCHE BASÉE SUR LES MODÈLES :** A l'ère du Big Data, Airbus est confronté aux mêmes challenges que bien d'autres organisations : un environnement utilisateurs et systèmes hétérogène, énormément de données disparates et de grandes difficultés à les rassembler et à les synchroniser pour prendre des décisions en temps voulu. Pour répondre à ses nouveaux besoins, Airbus a adopté l'approche ingénierie systèmes basée sur les modèles (MBSE). Ils ont choisi Tom Sawyer Perspectives et notre équipe Solutions Engineering pour ensemble conceptualiser et créer une application d'analyse visuelle collaborative intégrant toutes leurs données : GAIA (Global Aircraft Integration & Analysis).

**UNE IMAGE VAUT MILLE MOTS :** GAIA est une solution multi-vues permettant de suivre toutes les étapes de la conception à la fabrication d'un avion ainsi que tout système industriel. L'application utilise Tom Sawyer Perspectives pour intégrer des ensembles de données isolés et disparates, dans un environnement visuel unique et entièrement synchronisé. Pour la première fois, tous les composants et assemblages à bord de l'appareil sont représentés ensemble, dans leur contexte.



Un Architecte d'Airbus en train de naviguer sur un écran tactile «multi-touch» pour accéder à des visualisations de données et des graphes dans GAIA, grâce à Tom Sawyer Perspectives.

### RÉSOLURE DES PROBLÈMES DES ENTREPRISES FACE AU BIG DATA

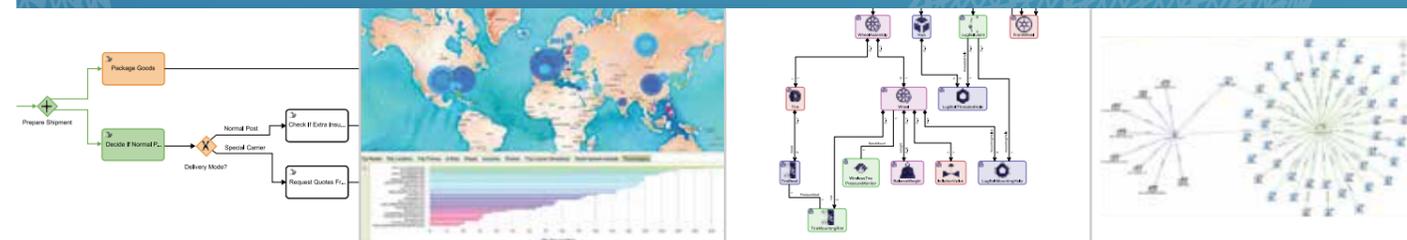
GAIA permet à Airbus de répondre à des besoins fondamentaux tels que : rendre possible l'intégration de données « prises en otage » dans des systèmes autonomes, utiliser et consolider les applications et l'expertise existantes, faciliter la communication et la collaboration entre les différents départements, valider et intégrer le concept dans son ensemble en assurant une cohérence fonctionnelle, avoir une solution adaptée à un large spectre d'utilisateurs. GAIA offre la possibilité aux architectes et aux autres parties prenantes dans la création d'un avion d'avoir une vue globale intégrant toutes les disciplines (structure, électricité, sûreté et sécurité, consommation de carburant, prix, etc.).

**POURQUOI TOM SAWYER PERSPECTIVES :** GAIA est le résultat d'une collaboration fructueuse avec Airbus qui a choisi notre plateforme technologique leur permettant : l'intégration de toutes leurs données disponibles, des vues interactives automatiquement synchronisées, un environnement flexible et évolutif, une approche ingénierie systèmes MBSE, une meilleure qualité de données, un affichage et des fonctions zoom efficaces, une navigation intuitive.

**EN QUELQUES MOTS :** L'application GAIA d'Airbus exploite la puissance du Big Data et permet de visualiser des systèmes, de synchroniser des vues et de rationaliser l'analyse et la fabrication d'avions de façon collaborative. Tom Sawyer Perspectives permet à GAIA de gérer des environnements complexes en utilisant des outils puissants de visualisation de graphes et de données dans une interface unique et synchronisée. La stratégie d'Airbus consistant à augmenter les cadences de productions des futurs avions tout en répondant aux critères de fiabilité, de sécurité et de confort des passagers, a pu décoller grâce à GAIA. En fait, la plus part des systèmes industriels pourraient être plus efficaces s'ils fonctionnaient avec des vues synchronisées et interactives telles que celles générées et rendues possibles par GAIA.

**LES CLIENTS DE TOM SAWYER SOFTWARE :** Aujourd'hui, nos produits et services sont utilisés dans le monde entier. Parmi nos clients figurent des organisations spécialisées dans l'aérospatiale, l'automobile, la défense et le renseignement, les services financiers, la santé, la fabrication et les télécommunications.

# Voir les connexions. Gagnez.



Nous sommes un fournisseur pionnier dans la création de logiciels et de services, permettant aux organisations de créer des applications intégratives de visualisation de graphes et de données, hautement modulables et évolutives.

Rencontrez notre équipe Tom Sawyer Software sur le stand A32



www.tomsawyer.com

Copyright © 2018 Tom Sawyer Software. Tous droits réservés.



## DATA PREPARATION

### Prepare Raw & Diverse Data Faster

#### COMMENT NORDEA RÉPOND AUX CONTRAINTES RÉGLEMENTAIRES GRÂCE À TRIFACTA.

##### A propos de Nordea

Avec plus de 10 million de clients dans les pays nordiques, Nordea Bank est une des plus grandes banques d'Europe. Outre leur siège à Stockholm, Nordea dispose de succursales, de filiales et de bureaux commerciaux dans 19 pays à travers le monde. Se consacrant à un service client exceptionnel, Nordea est une des rares banques internationales qui a augmenté sa clientèle malgré les fluctuations des marchés financiers.

##### Hausse des exigences réglementaires

En tant que banque mondiale d'importance systémique (Global Systemically Important Bank - GSIB), Nordea doit obligatoirement respecter les réglementations et est soumise au plus haut niveau de surveillance parmi les entreprises du secteur financier.

Ces dix dernières années, Nordea Bank, comme toutes les banques, a dû faire face à la forte hausse des exigences réglementaires. Les banques doivent non seulement fournir plus de données pour répondre à des demandes réglementaires nouvelles et de plus en plus fines, mais aussi satisfaire à des standards de qualité de données plus élevés. Les agences de contrôle exigent également que cette information soit délivrée plus rapidement, certaines réglementations obligeant même l'information à être reportée aussitôt qu'un événement particulier se produit. Face à l'amplification massive des demandes réglementaires, son impact sur les coûts opérationnels et la pression sur la rentabilité, le sujet est devenu d'une importance capitale pour Nordea. La banque s'est définie des objectifs ambitieux pour réduire les coûts de production des rapports réglementaires, ce qui a amené Nordea à repenser entièrement son architecture et ses processus existants. Ce ne fut pas une initiative aisée à accomplir pour une banque plus que centenaire, issue de plusieurs fusions et acquisitions et ayant à gérer l'héritage compliqué de nombreux systèmes anciens.

##### Des processus ETL (Extract-Transform-Load) dépassés

Lorsque Nordea a voulu construire une nouvelle plateforme de données, une grande partie de ce déploiement impliquait une transition de l'approche des processus d'extraction, transformation et de chargement (ETL). Ceux-ci n'étaient plus adaptés aux exigences de coûts et de rapidité nécessaire pour satisfaire aux impératifs réglementaires.

Historiquement, le département informatique de Nordea était responsable du développement des logiques de transformation de données d'après les spécifications soumises par le métier. Cependant, les processus et systèmes existants n'étaient plus capables de tenir la charge pour délivrer dans les délais les réponses aux exigences croissantes des réglementations. Avec un processus ETL rigide, Nordea avait peu de flexibilité pour rapidement prototyper et remédier aux problèmes systémiques d'incohérence, de conformité et de complétude des données. La complexité du traitement des données conduit souvent à des postulats qui mutent ou disparaissent entre les équipes informatiques et le métier, amenant souvent à des solutions incomplètes, lentes et non-conformes. De plus, le processus ETL ne fournissait pas

une traçabilité complète, et avec l'avènement de réglementations comme BCBS 239, le suivi de bout en bout des données devenait primordial. Toutes ces limitations, menaçaient non seulement la capacité de Nordea à respecter les réglementations mais aussi son image face aux clients et au secteur financier.

##### La solution : le data wrangling en self service

La banque Nordea a adopté la technologie Data Wrangling de Trifacta en conjonction avec la plateforme de données Cloudera Hadoop et la solution de visualisation de Arcadia Data, dans le but d'améliorer radicalement son reporting réglementaire. Avec Trifacta, Nordea peut maintenant produire ses rapports de 5 à 15 fois plus rapidement. Par exemple, la réponse aux demandes réglementaires des moyens de paiement SEPA est réalisée en une seule journée, au lieu de 15 jours préalablement. L'interface visuelle et intuitive de Trifacta a permis à Nordea de travailler interactivement avec des données ayant des structures hétérogènes et de qualité variée, de gagner en précision pour produire de meilleurs résultats. Mieux encore, ce travail a été entièrement délégué aux utilisateurs métiers (experts en réglementation) au lieu de dépendre fortement du service informatique pour délivrer en fonction des spécifications des utilisateurs. Concernant la sécurité et les normes techniques, l'équipe informatique de Nordea a pu parfaitement intégrer Trifacta grâce à ses fonctionnalités de sécurité robustes mais aussi de gouvernance collaborative et la visibilité fournie sur la traçabilité des données. La qualité du reporting réglementaire a été fortement améliorée et les coûts opérationnels sensiblement réduits.

*"L'ETL est en phase de transition du département informatique vers le métier. Nous pensons que les données ne devraient pas appartenir à l'informatique"* Alasdair Anderson, directeur Data Engineering chez Nordea Bank.

##### Les avantages en quelques points

- Produire 5 à 10 fois plus rapidement des rapports réglementaires.
- Identification des problèmes de qualité de données et de leur nettoyage en amont pour chaque source de données.
- Capacité à répondre à plus de demandes dans de meilleurs délais et avec des résultats de meilleure qualité.
- Transfert de la responsabilité de la transformation des données du département informatique vers les experts métiers.
- Amélioration de la flexibilité pour répondre rapidement aux changements des impératifs de réglementation.

##### A propos de Trifacta :

Trifacta est le leader mondial du Data Wrangling. Trifacta tire profit de décennies de recherches innovantes dans l'interaction homme-machine, de la gestion de données à grande échelle et d'apprentissage automatique (machine learning) afin de rendre le processus de préparation de données plus rapide et plus intuitif. Partout dans le monde, des milliers d'utilisateurs dans plus de 8 000 entreprises leaders, telles que Deutsche Boerse, Google, Kaiser Permanente, New York Life ou PepsiCo peuvent exploiter le potentiel de leurs données grâce à Trifacta, la solution de data wrangling la plus avancée du marché.

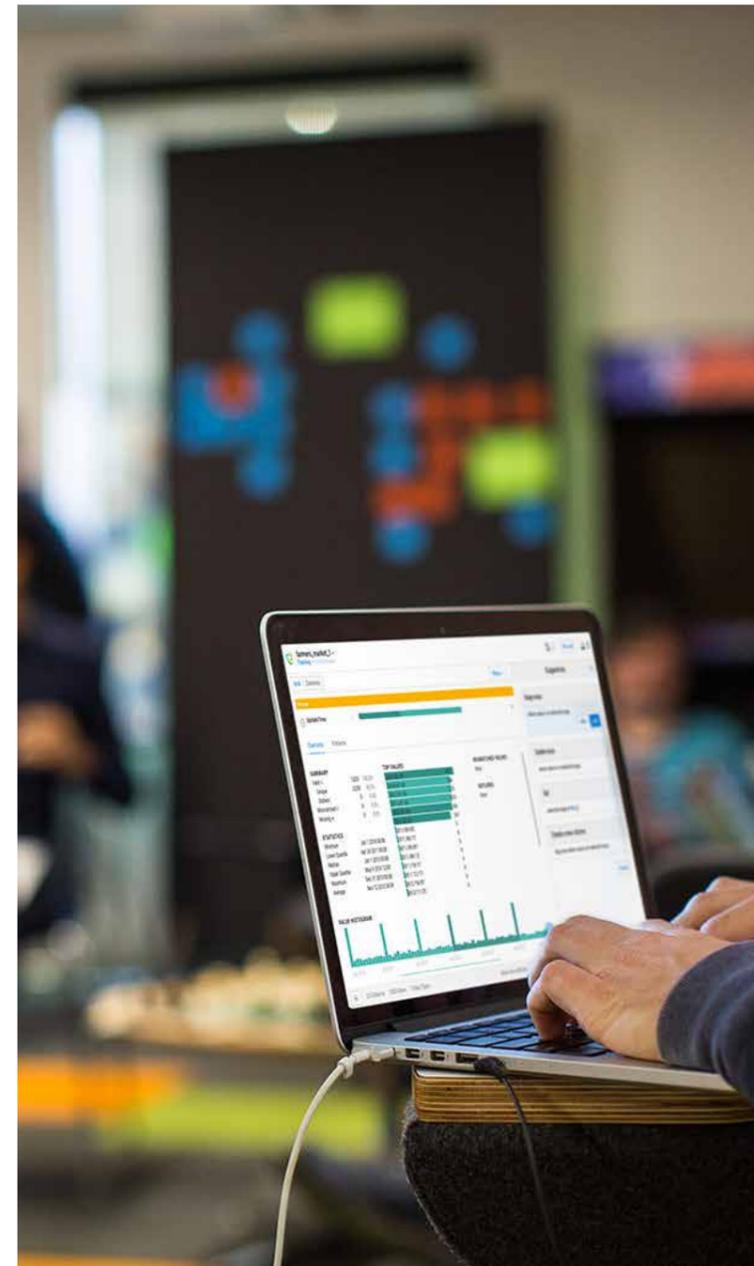
## Trifacta is the Global Leader in Data Preparation

Trifacta leverages decades of innovative research in human-computer interaction, scalable data management and machine learning to make the process of preparing data faster and more intuitive.

Start wrangling today!

FREE DOWNLOAD:

<https://www.trifacta.com/wrangling-events/>



TRUSTED BY  
8,000 companies





Xebia est un cabinet de conseil IT agile spécialisé dans les technologies Data, Web, Cloud, les architectures réactives et la Mobilité. Depuis 2004, Xebia produit des logiciels de haute qualité. Crafts(wo)men passionné(e)s, les 150 Xebians se mobilisent autour d'un seul mot d'ordre «Software Development Done Right». Nous comptons plus de 300 clients, aussi bien des entreprises du CAC 40 que des start-ups, dans tous les domaines d'activité. En 2010, nous avons créé la Xebia Data Factory qui accompagne nos clients sur leurs problématiques Data.

**L'AMBITION DE NOS CLIENTS EST NOTRE TERRAIN DE JEUX**

Notre ADN agile allié à notre expérience auprès de clients ambitieux nous permettent d'innover. C'est ainsi que nous avons créé une nouvelle offre «Data Science Design Sprint» et lancé la start-up Zeenea. Voici quelques-unes de nos références clients dans le domaine de la Data :

**Air France KLM:** Les innovations digitales et l'amélioration de l'expérience client sont au cœur de la stratégie d'Air France KLM. Offrir de nouveaux services aux clients demande une adaptation permanente des métiers et de l'IT pour tirer pleinement partie des opportunités du Big Data. Nous les accompagnons sur la phase d'industrialisation de leurs cas d'usage Big Data.

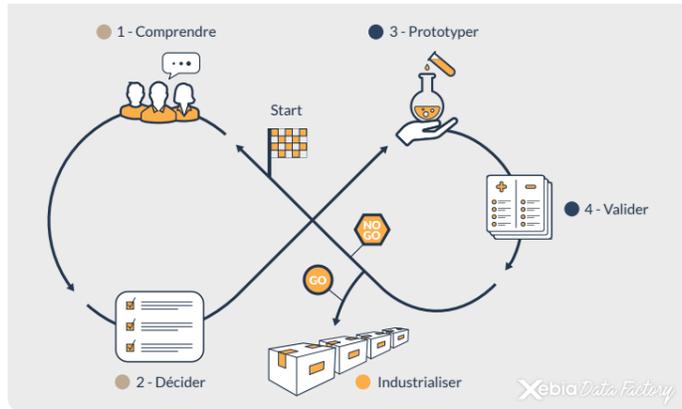
**Société Générale :** Au sein du Lab du Business Solution Center de la Société Générale, nous avons mis en œuvre des applications innovantes (Blockchain, Pepper, etc.) pour répondre aux besoins des différents métiers du Groupe en s'appuyant sur Lean startup, Scrum et DevOps. Nous sommes intervenus sur la conception et la réalisation d'un chatbot.

**PMU :** Avec une démarche agile, nous avons accompagné PMU dans la mise en place d'un Data Lake comme solution de gestion et de gouvernance des métadonnées afin de mieux traiter les données, améliorer la connaissance client et gérer les campagnes marketing.

**DATA SCIENCE DESIGN SPRINT OU COMMENT S'ASSURER QUE LE TEMPS CONSACRÉ À LA RÉALISATION D'UN USE CASE EST OPTIMISÉ ?**

Par leur nature intrinsèquement exploratoire, les Use Cases liés à l'analyse de la donnée sont souvent difficiles à appréhender et à estimer. L'une des difficultés souvent rencontrée est la mauvaise définition du besoin en amont de la réalisation d'un Use Case qui peut parfois mener à un produit peu exploitable ou sous-utilisé.

Le Data Science Design Sprint répond à l'objectif d'accélérer et de faciliter le prototypage d'un Use Case afin de s'assurer au plus vite que votre besoin final soit respecté et que tous les pré-requis soit à disposition pour le mener à bien.



Le Data Science Design Sprint s'adresse à toutes les directions métiers souhaitant :

- challenger une approche business actuelle avec des approches Data Science,
  - cadrer, tester et prototyper rapidement une idée innovante porteuse de valeur pour le métier,
  - avoir une connaissance rapide et factuelle de la faisabilité ainsi que de l'intérêt business d'une idée innovante par la donnée.
- Plus d'informations sur [www.xebia.fr](http://www.xebia.fr) ou contactez [sprint-datascience@xebia.fr](mailto:sprint-datascience@xebia.fr)

**EXPLOITEZ ENFIN VOTRE DATA LAKE AVEC ZEENEA**

Les entreprises, les plus avancées dans la révolution Big Data, souhaitent installer une véritable culture autour de la donnée. Toutefois pour devenir une organisation dite "data-driven", vos collaborateurs ont besoin d'un accès simple et rapide à des données organisées, de bonne qualité et sécurisées.

**Une solution née de l'expérience terrain**

Afin d'accompagner cette évolution, et forts d'expériences réussies chez nos clients ces 7 dernières années dans des missions Big Data, nous avons développé une solution autour de l'exploitation de Data Lakes. Cette solution a aujourd'hui pris son indépendance pour devenir Zeenea !

**Vers un smart Data Catalog**

Connecté au Data Lake des entreprises, Zeenea simplifie, automatise et monitorise l'intégration et la préparation des données afin de délivrer un catalogue de données en libre-service et collaboratif pour l'ensemble des employés d'une organisation.

Notre approche : **Time-saver** : Zeenea automatise et simplifie la préparation et l'intégration des données de vos Data Engineers.

**GDPR compliance** : Zeenea permet de manager vos Data Sets et de les cartographier. **Easy-to-use-interface** : L'interface de Zeenea a été pensée pour que les équipes IT et Business puissent piocher aisément dans ce véritable self-service de la donnée.

**Collaborative** : Un ensemble de fonctionnalités ont été développées pour collaborer autour d'un jeu de données.

Soutenue par Xebia, Zeenea a déjà été adoptée par nos clients.

Plus d'informations sur [www.zeenea.com](http://www.zeenea.com)

# DataScience Design Sprint



Notre méthodologie pour **valoriser vos données en 10 jours** efficacement et durablement.

FAITES COMME NOS CLIENTS, **CHALLENGEZ-NOUS**



## QU'EST-CE QUE LA DATA SCIENCE ET QUELLE EST SA PLACE DANS L'ENTREPRISE ?

Plus aucune activité n'échappe à la dimension digitale et toute activité laisse désormais des traces, des mesures, des impacts, des indices, des informations, bref, des datas. Le Big Data est omniprésent. Les données l'ont toujours été même si leur volume est exponentiel. Mais un changement radical de paradigme s'est opéré : nous sommes aujourd'hui en capacité d'exploiter pleinement le volume toujours plus grandissant de données grâce aux technologies qui offrent des potentiels inédits.

Le monde tel que nous l'avons connu n'est plus. Il faut s'y faire. Mais il faut surtout s'y fier. Et c'est là que la Data Science apparaît, parce que se dessine le rôle d'une nouvelle discipline qui cherche à transformer l'énorme masse d'informations disponibles en intelligence, connaissance, prédiction, prévision. La Data Science est la discipline qui nous permet de garder emprise et contrôle dans ce monde si différent qui se profile et se façonne dès à présent. Dit plus simplement, la **Data Science** ou science des données est une panoplie de mesures consistant à extraire de l'information d'un ou plusieurs ensembles de données et de l'activer.

La Data Science, c'est d'abord des data scientists. Leur travail consiste à nettoyer et à explorer les jeux de données à travers une méthodologie scientifique de travail, à trouver des corrélations significatives entre les différents jeux de data, et aussi à restituer les résultats obtenus de façon simple et claire, ce qui peut notamment passer par de la visualisation. Mais les activités des data scientists ne s'arrêtent pas à l'utilisation d'algorithmes d'apprentissage automatique avancés. Le data scientist sert aussi d'interface humaine. Il doit s'appropriier les enjeux métiers de l'entreprise et être capable de présenter des arguments dans un jargon métier.

L'entreprise, évidemment, est à l'avant-poste de cette révolution pour deux raisons essentiellement. D'abord, parce que ce sont les entreprises les premières qui ont compris la valeur et les avantages concurrentiels que l'on peut tirer de l'intelligence. Il suffit d'évoquer le monde du décisionnel et son histoire pour comprendre que l'essor de la data est éternellement endetté auprès de l'entreprise. La seconde raison est plus terre-à-terre. Ce sont les entreprises qui possèdent la data, elles ont cependant besoin de sociétés expertes pour mettre en évidence des patterns jusqu'à méconnus, effectuer des analyses look-alike sur de nouveaux critères issus de phase de features engineering ou encore faire de la prédiction.

La Data Science est l'approche raisonnée et rationnelle dont se dotent les entreprises data centric qui font de la data un pilier central de leur transformation digitale.

## QUELLE EST L'INTÉRÊT POUR UNE SOCIÉTÉ COMME YSANCE DE PROPOSER UNE OFFRE DE DATA SCIENCE ?

Proposer une offre de Data Science permet à Ysance de se tenir au plus près des préoccupations et des besoins de ses clients, même si ces besoins ne sont pas toujours consciemment manifestes ou précisément verbalisés ; nous aidons nos clients à « faire » - certes - mais nous remontons également dans la chaîne de valeur de la data en aidant nos clients à comprendre, ou plus exactement, à mieux appréhender ce qu'ils comprennent mais qui paraît abrupte et abscons au premier abord.

Ysance est un acteur reconnu dans le secteur de par ses partenariats multiples avec les sociétés du **Big Data**, de la **Data Integration**, de la **Data Architecture** et de la **Visualisation** et est donc confronté à un grand nombre de cas d'usages Big Data. En proposant une offre de Data Science, Ysance accompagne désormais ses clients sur l'intégralité de leur parcours Data.

## QUEL TYPE DE CAS D'USAGE EST FRÉQUEMMENT DEMANDÉ PAR VOS CLIENTS ?

Les cas d'usages abordés par les data scientists d'Ysance sont très variés - de la **détection d'outliers**, à l'**étude de comportement**, en passant par de la **prédiction temporelle** ou de **segmentation non supervisée** - de même que les **algorithmes d'intelligence artificielle** qu'ils doivent utiliser. Certains clients souhaitent répondre à des cas d'usages précis alors que d'autres veulent savoir s'il est possible de discerner des schémas au sein de leurs données. Néanmoins, certains besoins sont récurrents.

Nous sommes fréquemment sollicités, par exemple, pour étudier le comportement des utilisateurs à moyen et long terme à partir d'un historique de leurs parcours. Pour établir des prédictions on utilise des algorithmes d'intelligence artificielle d'apprentissage supervisé. On utilise les données historiques des clients (achats en ligne ou en magasin, données de navigations, ...) et on calcule des coefficients de similitude d'un nouveau consommateur avec des consommateurs existants. Grâce à cette démarche nous sommes en mesure de prédire les futures étapes probables du parcours de chaque consommateur et d'apporter à nos clients de la proactivité, de l'agilité. Cela peut se traduire par des recommandations personnalisées qui vont favoriser la rétention et/ou la transformation du consommateur ou anticiper le taux d'attrition. Ces algorithmes étant auto-apprenants, il est ensuite possible de déceler des changements de comportements des consommateurs. Tout cela bien entendu dans le respect de la vie privée.

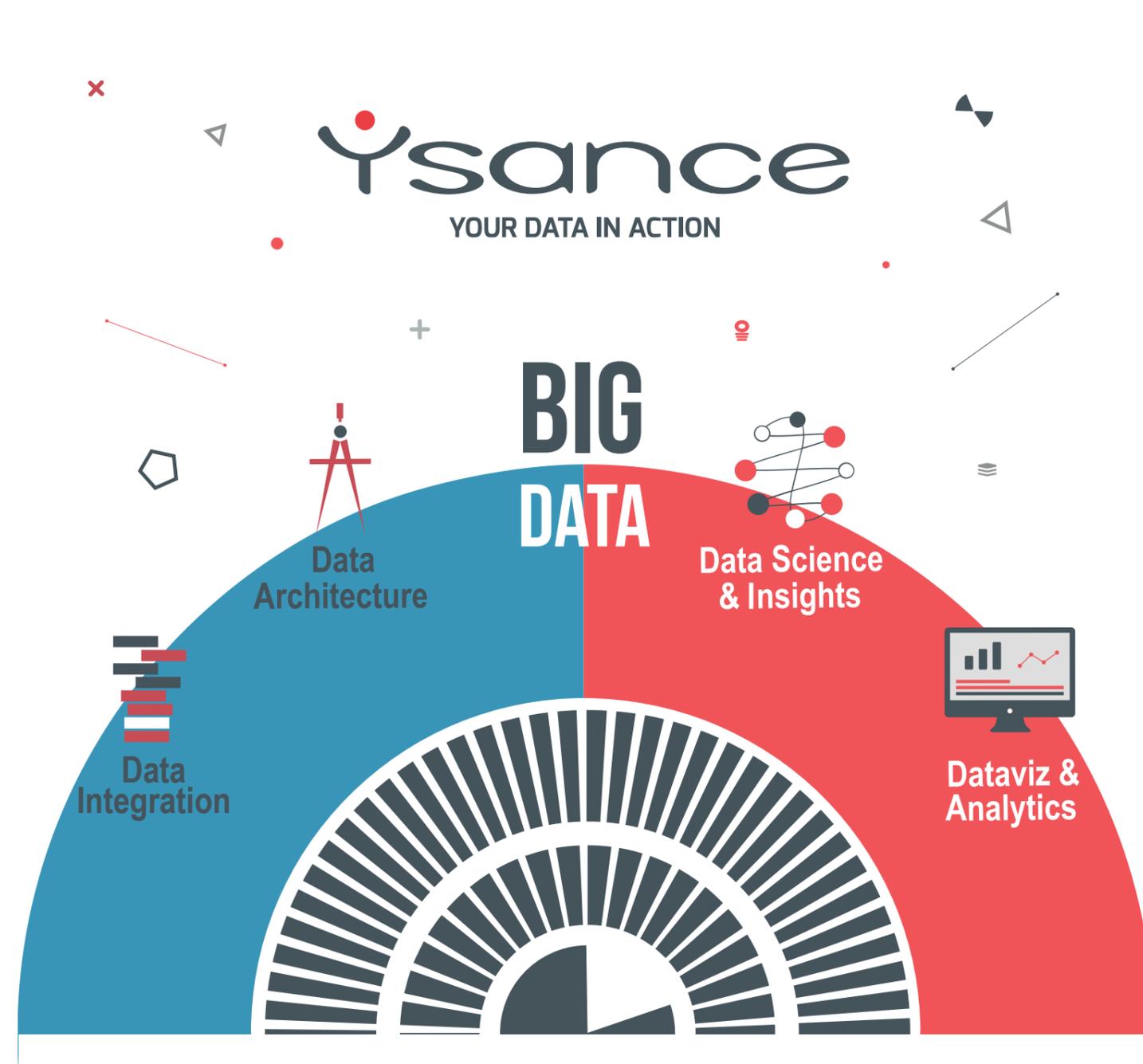
## POUVEZ-VOUS CITER UN EXEMPLE D'ANALYSE DANS LE DOMAINE MARKETING ?

Les problématiques retail les plus importantes du moment touchent toutes de près ou de loin le parcours d'achat des consommateurs. La **mesure et l'analyse du phénomène ROPO** (Research Online, Purchase Offline) permet à nos clients d'adapter leur stratégie marketing digitale online et pour les magasins.

Alors que l'analyse des parcours online seuls est relativement standard, la détermination du canal d'acquisition ou le facteur déclenchant un achat magasin se révèle plus délicat, complexe et difficile. Mais l'avènement des technologies Big Data et des méthodes de Data Science change la donne. Via la **mise en commun de l'intégralité des sources d'informations d'une entreprise** (données de ventes, de navigations, base CRM, ...) dans un **data lake unique**, il est désormais possible de croiser efficacement des données issues de sources indépendantes et de tirer des enseignements des données ainsi croisées.

La Data Science, via des analyses temporelles sur la navigation en ligne précédant un achat magasin, permet désormais d'estimer raisonnablement le bénéfice réel d'une action marketing digitale et son ROI.

Découvrez toute l'expertise Ysance sur notre site : [www.ysance.com](http://www.ysance.com).



## LA CHAÎNE DE VALEUR DATA



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



zenika

<animés par la passion>



<MAKE DATA MATTER>

### METTEZ LA DONNEE AU CŒUR DE VOS ENJEUX

Cabinet d'innovation, de conseil, de réalisation en open-source et de formation, Zenika accompagne les entreprises dans la mise en valeur de leurs données, des données qui sont désormais au cœur de tous les enjeux. Interview croisée de Carl Azoury, PDG et co-fondateur de Zenika, et Florian Hussonnois, Responsable du Pôle Big Data UNKIPPLE de Zenika.

### QUELLE REGARD PORTE ZENIKA SUR L'ÉMERGENCE DU BIG DATA ?

CA : Nous vivons une période disruptive sans précédent. L'internet de l'objet, la digitalisation des entreprises, les approches collaboratives, les Smart Nations ainsi que les réseaux sociaux, transforment en profondeur les modèles économiques des « corporates » qui sont condamnées à évoluer pour ne pas se faire devancer par des start-ups. Les sources de données sont de plus en plus nombreuses ! Les stocker n'a jamais été aussi peu coûteux. Il est devenu nécessaire de les visualiser et de les structurer pour la prise de décision, pour mieux connaître ses clients. Le terme Big Data lui-même recouvre des concepts très (trop) divers.

### COMMENT ZENIKA APPROCHE LE MARCHÉ DU BIG DATA ?

CA : Notre culture est hautement technologique et agile, cela nous permet d'accélérer la mise en place de stratégies adaptées au Big Data. Présent depuis 2006, nous comptons aujourd'hui 300 collaborateurs dans huit agences en France, à Paris, Rennes, Lille, Nantes, Lyon, Bordeaux, ainsi qu'à Singapour et Montréal. La création d'une agence Big Data UNKIPPLE et d'une Digital Agency AYA, spécialisée dans le web et la mobilité, nous permettent d'adresser une offre complète, de l'analyse des données à la création d'une expérience de Data Visualisation.

### QUELLE EST LA PARTICULARITÉ DE CETTE OFFRE DÉDIÉE AU MARCHÉ DU BIG DATA ?

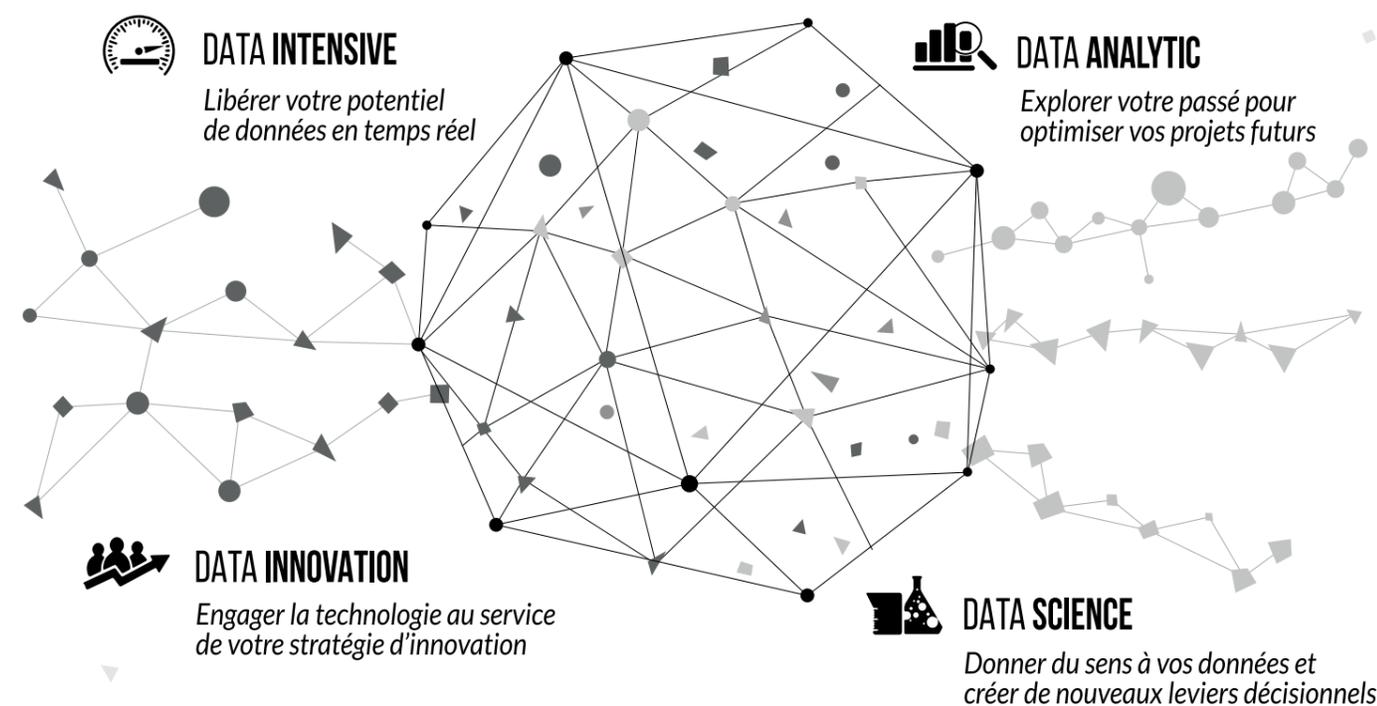
FH : Nous accompagnons les entreprises dans la valorisation de leur patrimoine de données. Un patrimoine encore peu considéré aujourd'hui, qui pourtant possède déjà une véritable valeur marchande. Il y a différentes approches : regrouper un maximum de données pour dégager de nouveaux enseignements métier ou concevoir le produit qui tirera parti du potentiel encore inexploité des données. Notre rôle est d'accompagner de l'idée à la mise en production des projets Data, qu'il s'agisse de projets en B to B ou de produits orientés métier. Nous travaillons aussi bien pour les secteurs de la Finance, des Télécoms et de l'Énergie.

### A QUELLES PROBLÉMATIQUES PEUVENT RÉPONDRE VOS PROJETS ?

FH : La data peut avoir de multiples usages. Les problématiques récurrentes sont celles de l'analyse de données clients, pour améliorer la relation client ou éviter l'évasion. La data peut également permettre de rester compétitif face à des modèles disruptifs grâce à l'analyse en temps réel, d'explorer le passé afin d'optimiser les futurs projets, d'appuyer une stratégie d'innovation ou de faire émerger de nouveaux leviers décisionnels. Nous créons de la valeur rapidement grâce à l'analyse et au traitement des données de l'entreprise.

### ZENIKA A ÉGALEMENT LA PARTICULARITÉ DE TRAVAILLER EN OPEN-SOURCE, POUR QUELLES RAISONS ?

FH : En effet, nous travaillons en partenariat avec des éditeurs de solutions open-source, tels que Confluent, Cloudera, Couchbase ou DataStax qui sont leaders sur le marché de la Data. Nous sommes également centre de formation certifié pour un certain nombre d'entres-eux. Cela contribue à faire monter en compétences nos équipes et à faire de Zenika une société de référence dans ce domaine. A chaque fois, notre objectif est d'accompagner avec le meilleur niveau d'expertise les entreprises qui souhaitent utiliser ces technologies. Aussi, nous partageons notre savoir-faire et notre passion pour faire grandir la communauté open-source. CA : Collaborer avec des éditeurs est un gage de qualité et une contrainte supplémentaire que l'on se donne afin d'améliorer l'expérience client et de créer un lien avec l'éditeur.



www.unkipple.com



zenika

<animés par la passion>

Créé en 2006, Zenika est un cabinet d'innovation technologique, managériale et organisationnelle, qui accompagne les entreprises dans leur transformation numérique.

Nous sommes avant tout une entreprise qui a la capacité de s'adapter et de se positionner sur de nouveaux créneaux technologiques. Nous regroupons ainsi des profils variés sur les domaines d'expertise suivants : BigData, Web, Mobilité, IoT, DevOps, Craftmanship, Agilité, sécurité et transformation des organisations.

50 à 500 salariés



**LEAD** The 4<sup>th</sup>  
DIMENSION

100 Speakers  
200 Marques  
15 000 Participants





## **Savez-vous combien la data contribue à la performance de votre entreprise ?**

Chez Altares, nous replaçons l'information au cœur de vos réalités économiques afin de vous apporter des solutions réellement adaptées à votre business.