

BIG Corp

2019 - 2020

GUIDE DU BIG DATA : 365 JOURS AU CŒUR DE LA DATA ÉCONOMIE

DOSSIER
SPÉCIAL

Biais des algorithmes,
fake data,
failles de sécurité...
Quand les Big Data
se rebiffent !



NOTHING
SOUNDS BETTER
THAN **YOUR BIG DATA**

Dans ta Tech'

Vers l'Infiniment Précis ?

Working Society

Profils Data : le Mouton à 5 Pattes ?

Dessine-moi demain

L'IA, une Affaire d'Hommes ?

CARNET D'ADRESSES

Les **71** partenaires
qui feront vos **stratégies**
Big Data en 2020



EPSILON®

ENEZ NOUS
RENCONTREZ

**STAND
A34**

VOTRE MASTER DATA PARTNER

L'alliance unique de compétences data, technologiques et marketing, qui réunit 750 experts, compte plus de 350 clients actifs et propulse les organisations dans l'ère de la mass personnalisation.

UN ACCOMPAGNEMENT END-TO-END DE VOS AMBITIONS



IDEATE



DESIGN



BUILD



ROLL OUT



RUN

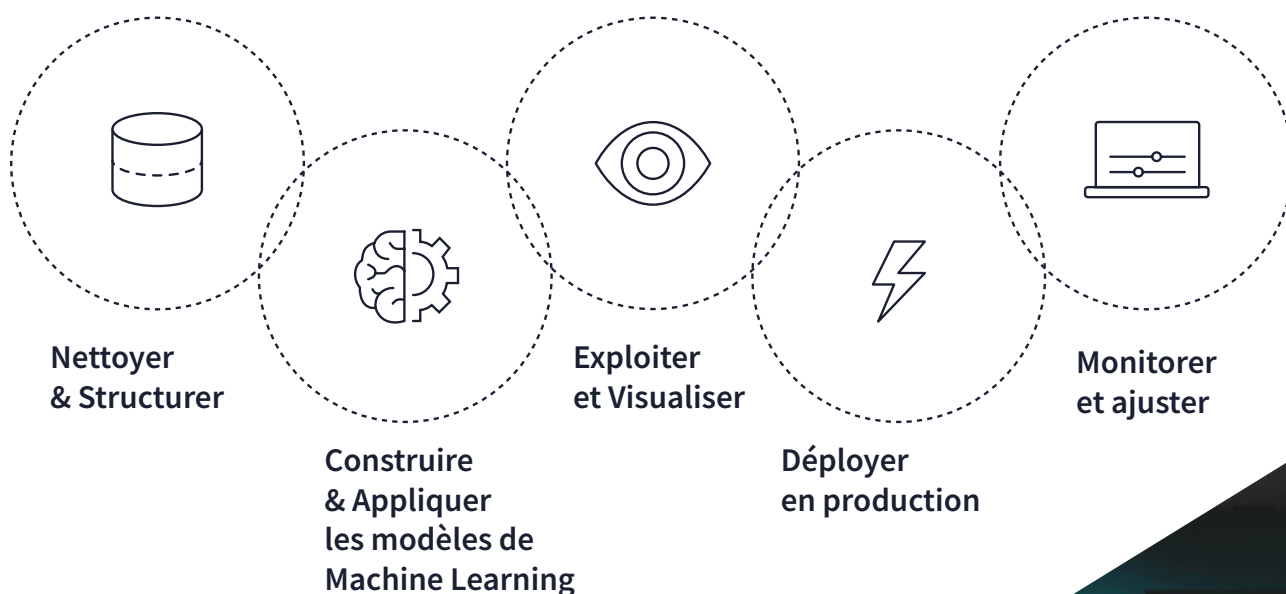


TRAIN

NOUS SOMMES EPSILON
www.epsilon-france.com

En route vers l'Enterprise AI

Dataiku la plateforme d'Analytique et de Data Science centralisée pour les experts data et les métiers.



EDW

SOMMAIRE

EDITO	4
DOSSIER SPÉCIAL : Biais des Algorithmes, Fake Data, Failles de Sécurité... Quand les Big Data se rebiffent	6
RUBRIQUE TECHNOS : Vers l'Infiniment Précis ?	26
RUBRIQUE MARCHÉS : Big Data : une Année dans le Rétro	36
RUBRIQUE RÉGLEMENTATION : RGPD : un An Après	50
RUBRIQUE SOCIÉTÉ : L'IA, une Affaire d'Hommes ?	56
RUBRIQUE CARRIÈRES : Profils Data : le Mouton à 5 Pattes ?	64
RUBRIQUE PROSPECTIVE : La Dataviz en AR/VR	72
RUBRIQUE USE CASES : De Big à Small : la Data dans tous ses Usages	78
CARNET D'ADRESSES : Les 71 partenaires qui feront vos stratégies Big Data en 2020.....	84

EN 2019...

... on a fini par se demander si l'infiniment petit n'était pas la clé de voûte du Big Data. Si après avoir tant vanté les traitements en masses, les serveurs planétaires et l'industrialisation, il ne faudrait pas finalement se pencher sur la valeur du plus fin, l'objectif du plus précis... En 2019, le Cloud a

opté pour des modèles d'applications « en conteneurs » et Hadoop poursuit son évolution vers toujours plus de modulaire, tandis qu'on rapproche stockage et traitement « à la périphérie » et que la Réalité Augmentée permet de zoomer au cœur du Big Data... Jusqu'où ? La décennie 2020 s'ouvrira probablement avec cet espoir de temps réel localisé, dont la voiture autonome semble être le fantasme le plus partagé. Mais alors, que craindre au juste dans ce monde où la data pourrait répondre « au doigt et à l'œil » au plus fin des besoins ?... Peut-être justement que cet infiniment petit n'ait pas dévoilé tous ses travers, entre fake data, insécurité et biais des algorithmes. 2019, l'année de la qualité des données ? En tous cas, on en parle dans BIG !

... on a également fini par se demander si l'Europe allait enfin remettre les GAFA à leur place (en termes de parts de marché bien sûr... car en termes de réglementation, le RGPD s'en est déjà chargé - cf notre rubrique *Réglementation*). Mais que

manque-t-il encore aux licornes européennes pour briller hors du Vieux Continent ? Et à la France, pays des mathématiques-roi, pour placer ses algorithmes en tête des charts ? Tandis que l'Asie commence à déployer ses ailes (*voir notre article sur l'Inde*), et que les Etats-Unis gardent le cap qu'ils ont eux-mêmes tracé, nous ferons le point sur les chances françaises dans un marché qui évolue de plus en plus autour des innovations BI.

... on a enfin fini par se demander à quoi pourrait bien servir le Big Data pour la société... Et à quoi il pourrait bien l'exposer. Big Data for good certes, mais aussi cyber-insécurité, biais algorithmiques et sous-représentation des femmes... Le déluge de données charrie toutes ses menaces et amène à des questions cruciales : l'IA, une affaire d'hommes ? la sécurité, une affaire de GAFA ? C'est aussi cela le Big Data version 2019 : une réflexion éthique qui ne laissera aucun citoyen sur le bas-côté, technophile ou non, corporate ou non. Alors que faire ? Par la tech ou par la loi ? L'éthique aura sûrement besoin des deux pour progresser, tandis que la nouvelle génération de data scientists arrive en entreprise la conscience pleine de tous ces enjeux...

2020 PROMET D'ÊTRE **BIG** ! VOUS EN SEREZ ?

DOSSIER SPECIAL

BIAIS DES ALGORITHMES, FAKE DATA, FAILLES DE SÉCURITÉ...

Pourquoi un dossier spécial ?

Parce que ces problématiques liées au cœur de la donnée sont trop souvent abordées à la périphérie, dans les annexes des publications Big Data.

Alors BIG a voulu replacer le sujet au centre, parce qu'il incarnera probablement le défi majeur des prochaines années, quand infrastructure et analyse auront trouvé leurs propres rythmes de croisière.

Qualité, biais, sécurité : bienvenue dans un triptyque qui pourrait faire frissonner... mais qui confirme qu'aujourd'hui, face aux chercheurs et aux nouveaux outils, les Big Data n'ont qu'à bien se tenir !



...QUAND
LES **BIG DATA**
SE REBIFFENT

QUALITÉ DES DONNÉES, CHANTIER N°1



Au départ, on parlait de doublons ou de fautes d'orthographe. Et puis, au fur et à mesure du déploiement des Big Data, on en est venu à parler « cohérence de l'usage », « fraîcheur du contexte » ou « conformité réglementaire »... Et la qualité des données est devenue une matrice à part entière, avec sa galaxie de critères et ses enjeux de gouvernance.

Quelques années après l'apparition des CDO et un an après celle du DPO et des enjeux de conformité RGPD, la mise en qualité des données semble s'imposer comme le chantier numéro un de la data-driven company. Quitte à achopper encore et toujours sur l'enjeu de la valorisation...

Retour en 5 actes sur LE sujet du moment.

LA SAGA

La Data Quality en 5 Actes

ACTE 1 : L'entreprise l'affirme : « La gouvernance est ma priorité ! »



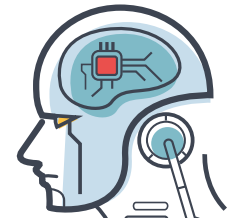
- En mai 2019, une étude rapportée par le cabinet Keyrus¹ souligne que, pour 90% des entreprises, le sujet de la gouvernance est devenu plus important qu'il y a dix ans. Parmi les sondés de cette même étude, 65% déclarent avoir déjà lancé un programme de data gouvernance...
- La même année, Gartner publie un sondage², dans lequel la mise en œuvre d'une « stratégie dédiée à l'analyse et la valorisation des données » est énoncée comme priorité 3 des CDO.

Ils sont d'ailleurs 45% à s'appuyer régulièrement sur des indicateurs de qualité des données (exactitude, exhaustivité, etc).

... Preuves que le sujet est sorti du bois et que les décideurs ont pris conscience du manque à gagner d'une mauvaise gestion de la donnée.

Selon Christophe Bonnefoux (cf interview) : « **Les organisations perdent 20% de leur temps à rechercher la donnée et à la mettre en qualité** ».

ACTE 2 : Le marché répond « Il y a des solutions technologiques ! »



Master Data Management et Metadata Management deviennent dès lors les ciments de la gouvernance d'entreprise : en urbanisant les données de référence, elles créent une traçabilité du Big Data au cœur de l'entreprise.

- Le marché des logiciels s'emballe : de 800 millions d'euros en 2016, il passe à 915 millions d'euros en 2018, avant de venir se positionner autour de 3,75 milliards d'euros en 2025 (estimation), toujours selon Keyrus.
- Cependant des voix discordantes commencent à émerger : Debra Logan, vice-présidente de recherche chez Gartner dé-

clare : « Se concentrer exclusivement sur la gouvernance, même le Master Data Management, ne suffit pas pour réussir en tant que CDO ». Pour elle, les actions menées en matière de qualité doivent avant tout être porteuses de résultats commerciaux chiffrés pour s'établir dans la durée.

En attendant, les CDO s'engagent dans des chantiers d'identification de leurs données et recourent de plus en plus à la priorisation des données stratégiques : **admettre un pourcentage d'erreurs pour des données non essentielles, pourquoi pas... mais les key data, jamais !**

ACTE 3 : Bruxelles s'en mêle...



Le 25 mai 2018, le Règlement Général sur la Protection des Données voit le jour dans la zone UE et, avec lui, le spectre des 4% de CA confisqués en cas de faille majeure. **Panique dans les boards : il faut accélérer sur la gestion de la conformité !**

Cartographie des SI, durée de conservation, gestion du consentement, plans de remédiation : autant de chantiers qui obligent les CDO (et leurs nouveaux parents, les DPO) à se pencher sur leurs data. Selon Keyrus, 36% des entreprises auraient ainsi déclaré que le RGPD les a incités à lancer une démarche de data gouvernance.

¹ Infographie « Les tendances de la data gouvernance », mai 2019, Keyrus
² 4e sondage annuel des CDO effectué en ligne par Gartner entre septembre et décembre 2018. Base : 257 CDO et analystes data

ACTE 4 : Eurêka, il y a l'IA !



Dans la foulée du mouvement impulsé par le RGPD, les progrès réalisés sur le front de l'IA amènent les CDO à s'intéresser à la détection automatisée des données erronées : créer des algorithmes adaptés à chaque culture d'entreprise devient dès lors le leitmotiv.

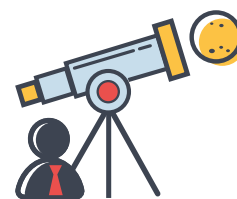
- Selon Gartner, le machine learning est ainsi jugé d'une importance critique pour 76% des CDO interrogés. Et 65% déclarent utiliser ou tester des projets en ce sens.

Au fur et à mesure, l'IA impose son rythme accéléré et sa précision dans la détection... sans toutefois s'affranchir d'une analyse humaine

Et puis c'est un juste retour des choses : **sans données de qualité, pas de modèles d'apprentissage fiables, donc pas de cas d'usage IA.**

Pour survivre dans la durée, l'IA prend donc le problème à sa source... eurêka !

ACTE 5 : Vers l'infini et la valorisation des données ?



Si la détection des « mauvaises données » progresse et si l'enjeu semble avoir fait son chemin dans les esprits, l'impact financier reste encore terra incognita pour les entreprises :

- Selon Gartner, seules 29% des entreprises mesureraient par KPI l'impact de ces données qualitatives sur leurs processus métiers.
- Pire, elles ne seraient que 8% à tenter de mesurer la valeur financière de ces données. Soit la valeur d'actif que celles-ci représentent dans le capital de l'entreprise.

Le manque de compétences, d'outils et de stratégie est souvent avancé pour expliquer l'inertie des Grands Comptes sur cette question.

Mais c'est surtout la notion de standardisation qui coupe l'herbe sous le pied des plus audacieux : à date, seules deux normes ISO³ sont identifiées pour qualifier et certifier des catégories de données. C'est encore trop peu pour permettre aux entreprises d'évaluer leurs « actifs informationnels » et les échanger.

- Selon Gartner, 19% des entreprises déclareraient vendre des données ou concéder des licences d'exploitation via courtiers ou places de marché en ligne.

En 2016 en France, un rapport du CIGREF⁴ consacré à la valorisation du « capital immatériel » appelait à la mise en place de tels référentiels et normes dans l'écosystème français.

Sans suite jusqu'à présent...

|| Selon Gartner, le machine learning est ainsi jugé d'une importance critique pour 76% des CDO interrogés ||

|| Panique dans les boards : il faut accélérer sur la gestion de la conformité ||

LA CHECKLIST

La Data Quality en 5 critères

Vous voulez vérifier que vos données ont un niveau de qualité suffisant pour être exploitables et générer de la valeur ?

Voici une petite checklist de critères qui vous aideront à séparer le bon grain de l'ivraie...

2 Exhaustivité

Les données nécessaires à l'aboutissement d'un résultat métier sont toutes mobilisées : la population statistique est suffisamment diverse.

4 Cohérence

Les données sont cohérentes d'une plateforme à l'autre, d'une mesure à l'autre : la comparaison avec d'autres situations métiers ne modifie pas la validité de la donnée.

1 Exactitude

Les données reflètent la réalité d'une situation ou d'un métier : leurs sources ne sont pas entachées de biais ou d'erreurs pour l'usage qui leur est destiné.

3 Intégrité

Les données ne sont pas corrompues par l'utilisation d'un format, d'une écriture ou par une défaillance technologique.

5 Fraîcheur

Les données sont « à jour » ou clairement associées à un contexte temporel.

En réalité, il existe d'autres critères utilisés sur le marché : par exemple **la conformité** (respect des exigences réglementaires), **l'unicité** (absence de doublons) ou encore **l'accessibilité** (requête et visualisation).

Le DAMA-DMBOK⁵, véritable Bible du Data Management sur le marché, en recense pas moins de 32 !

Cependant les grands principes évoqués ci-dessus devraient déjà vous mettre le pied à l'étrier ! Et si vous paniquez, sachez que la plupart des solutions de Master Data Management intègrent déjà un certain nombre de critères pour vous aider...

⁵ Data Management Body of Knowledge, disponible depuis 2017 sous sa seconde édition

INTERVIEW

Le secteur bancaire est-il à la pointe dès qu'il s'agit de qualité des données ?

Ce n'est pas le seul secteur à se pencher sur cette question mais il réunit un certain nombre de conditions favorables : des données en masse, des bibliothèques d'algorithmes en Open Source, des machines de calcul puissantes et une bonne attractivité auprès des écoles d'ingénieurs. Je suis convaincu que c'est le bon moment pour lancer des projets d'optimisation liés à la qualité des données et le secteur bancaire au sens large l'a compris : les organisations perdent 20% de leur temps à rechercher la donnée et la mettre en qualité, c'est bien trop ! Il y a évidemment du ROI à trouver...

Pourtant cela ne se fait pas sans heurts...

Il y a effectivement quelques challenges : technologiques (la puissance machine notamment) mais aussi et surtout humains. D'abord parce que ce n'est pas évident de trouver des profils experts, à la fois data scientists et financiers, sur ces sujets qui sont assez nouveaux. Ensuite parce que les équipes opérationnelles qui traitent la qualité de la donnée a posteriori vont forcément se trouver impactées. Le métier va changer...

C'est-à-dire ?

L'automatisation des tâches de détection de la non-qualité des données va conduire inéluctablement les analystes à se spécialiser sur la manipulation des résultats d'algorithmes et sur des fonctions de « double check » : par exemple identifier et comprendre comment l'algorithme a détecté certains éléments qui n'étaient pas détectés par des règles métiers. Et ainsi construire la confiance par rapport aux résultats. Cette spécialisation amènera forcément une concentration du nombre d'analystes, il faudra redéployer les équipes vers d'autres besoins.

La gouvernance des données, c'est un sujet en vogue. Comment s'incarne-t-il chez BNP Paribas Asset Management ?

Grâce à trois programmes prioritaires : 1/ le programme Know Your Data qui consiste à identifier des données stratégiques et à déployer une gouvernance simple et efficace, 2/ le programme RGPD qui a permis une meilleure protection des données personnelles via une mise en conformité des systèmes IT et de nos procédures internes, et 3/ le programme Data Hub, la plateforme technologique de mutualisation des données de l'entreprise, qui accroît la valeur d'usage des données et diminuera le temps de déploiement des cas d'usage métier...

Sur la partie « Know your Data », qu'appellez-vous une donnée stratégique ?

Disons que sur les 8 à 10000 types de données que nous possédons approximativement, nous avons choisi d'en classer une centaine que nous appelons les « key data » - nos données stratégiques. Nous avons dû théoriser pour définir en quoi elles sont stratégiques et avons défini plusieurs critères : est-ce qu'elles sont utilisées par un membre du comité exécutif ? est-ce qu'elles sont publiées en externe ou non ? est-ce qu'elles sont contrôlées par un régulateur ? est-ce qu'elles sont assorties d'une amende si leur qualité fait défaut ? L'idée étant de les définir, de les mettre en lumière auprès des responsables opérationnels et de leur en donner la responsabilité : la qualité et la remédiation font désormais partie de leur périmètre.

Le RGPD a-t-il été une étape-clé dans votre gouvernance data ?

Il nous a aidés à mieux gérer les données clients en mettant en avant des sujets clés comme l'anonymisation ou la protection

des données. Nous avons d'ailleurs lancé des réflexions sur le sujet connexe de la pseudonymisation. Maintenant, il reste des chantiers, comme celui de l'automatisation de la maintenance du registre de traitement ; cela fait partie des sujets d'optimisation que je vais chercher à déployer dans les douze prochains mois.

Pour vous, quel est encore le principal défi de la mise en qualité des données ?

Techniquement, c'est probablement celui d'une intervention « au plus tôt » : être capable de mettre en qualité une donnée avant même son entrée dans un datamart ou une application métier. Et humainement, c'est le change management : faire comprendre aux équipes métier qu'elles ne doivent plus perdre de temps à gérer une mauvaise qualité de la donnée. Et travailler avec elles sur le sens et la gouvernance avant de s'équiper d'outils partiels, tactiques et générant de la dette technique.

Et votre souhait pour l'avenir ?

Le graal pour un CDO, c'est probablement que la donnée soit enfin reconnue comme un actif de l'entreprise et apparaisse dans son bilan : je rêve notamment de normes comptables qui permettront de donner une valeur financière aux données. Aujourd'hui, le seul outil juridique, c'est le RGPD, mais il ne valorise rien. Créons des standards européens pour ne pas reposer uniquement sur des standards et méthodes américains. Et recourons davantage à l'IA pour les mettre en qualité : cela permettra d'accélérer la valeur d'usage des données. C'est un vrai sujet aujourd'hui pour l'ensemble de nos partenaires (régulateurs, cabinets d'audit, RegTech...).

Nous avons
besoin d'une
**réglementation
européenne**
concernant
l'intelligence
artificielle qui
permettra de
mieux
valoriser
la **donnée**



Christophe
Bonnefoux

Chief Data Officer
BNP Paribas Asset Management

BIAIS DES DONNÉES PEUR SUR L'ALGO



Vous êtes-vous déjà demandé pourquoi votre moteur Google vous suggérerait certaines associations de mots plutôt que d'autres quand vous tapez votre recherche ? Pourquoi certaines publicités apparaissent beaucoup plus fréquemment sur votre IP que sur l'IP de votre voisin - qui se trouve être du sexe opposé ? et les séries suggérées sur Netflix différentes selon le profil renseigné

(alors que les programmes regardés sont les mêmes) ?

Le terme est connu : ce sont les « biais ». Soit l'introduction de discriminations ou, plus globalement, d'une réalité déformée au cœur des algorithmes. Avec pour conséquence : un résultat déloyal et inéquitable, qui ne représente pas la réalité.

Les biais ont envahi la sphère de pensée autour de l'IA et les enjeux sociaux qu'ils représentent sont désormais clairement identifiés... sans qu'on ne parvienne vraiment à apporter des réponses de long terme. Car derrière le débat sur les biais se cache une autre question, bien plus ardue scientifiquement : comment expliquer les résultats d'un algorithme et le mode opératoire qui lui a permis d'aboutir ?

Co-auteur d'un article⁶ sur le sujet et spécialiste des liens entre algorithmes et sciences sociales, David Bounie de l'Institut Télécom ParisTech a consacré un an à cette problématique ; il nous apporte son éclairage sur les notions à mettre derrière le mot « biais » et les défis (nombreux) qu'il reste à relever.

UN PEU DE THÉORIE

Profiling des Biais*

Les Biais Cognitifs

Ce sont les biais qui reposent sur la vision partielle du développeur par rapport à la réalité : celui-ci va alors modéliser des corrélations en fonction de ses perceptions.

- **Le biais « du mouton de Parnurge »** : le programmeur utilise des corrélations ou des modélisations qui sont populaires sans s'assurer de leur exactitude.
- **Les biais de confirmation** : le programmeur favorise sa propre perception de la réalité sur l'ensemble des données disponibles. Il ne prend en considération que les données qui confirment sa croyance et ignorent celles qui le contredisent.
- **Le biais de corrélations illusoire** : le programmeur décèle des corrélations entre deux événements qui sont par nature indépendants, ou amplifie le poids d'une corrélation dans le résultat. Par exemple : il surestime le lien entre une caractéristique psychologique d'un individu et son appartenance à un groupe social.
- **Le biais de stéréotype** : celui-ci n'est pas le fait du programmeur mais bien de la population statistique : un individu agit en référence au groupe social auquel il s'identifie plutôt que sur ses capacités individuelles. Par exemple : une femme qui se censure sur la réponse à certaines offres d'emploi.

Les Biais Statistiques

Ce sont les biais qui proviennent de « mauvaises » données ou de méthodologies statistiques déséquilibrées.

- **Le biais des données** : les données d'entraînement par exemple ne sont pas fiables et ne représentent pas l'univers de référence.
- **Le biais de variable omise** : une caractéristique sensible (ex : appartenance à un groupe social, origine ethnique ou caractéristique psychologique) est omise dans le modèle et cette variable est fortement corrélée à d'autres variables du modèle. Le résultat de l'algorithme est par définition amputé d'un paramètre clé qui conduit à des estimations biaisées.

- **Le biais de sélection** : l'échantillon de population étudié n'a pas les mêmes caractéristiques que la population en général. Le biais de sélection est parfois difficile à contourner dans la mesure où ce sont souvent ses caractéristiques propres qui ont conduit le statisticien à construire l'échantillon.
- **Le biais d'endogénéité** : l'algorithme ne prend en compte que les données passées mais ne tient pas compte des ajustements que les individus peuvent faire par anticipation. Par exemple : ajuster des investissements boursiers sur un titre stable par anticipation d'un krach, ou ajuster un comportement à risque (santé, finances) par anticipation d'une crise à venir (maladie, non-obtention d'un crédit, etc).

Les Biais Économiques

Ce sont les biais qui reposent sur une analyse coût/bénéfice de l'utilisation de telle donnée ou de tel algorithme.

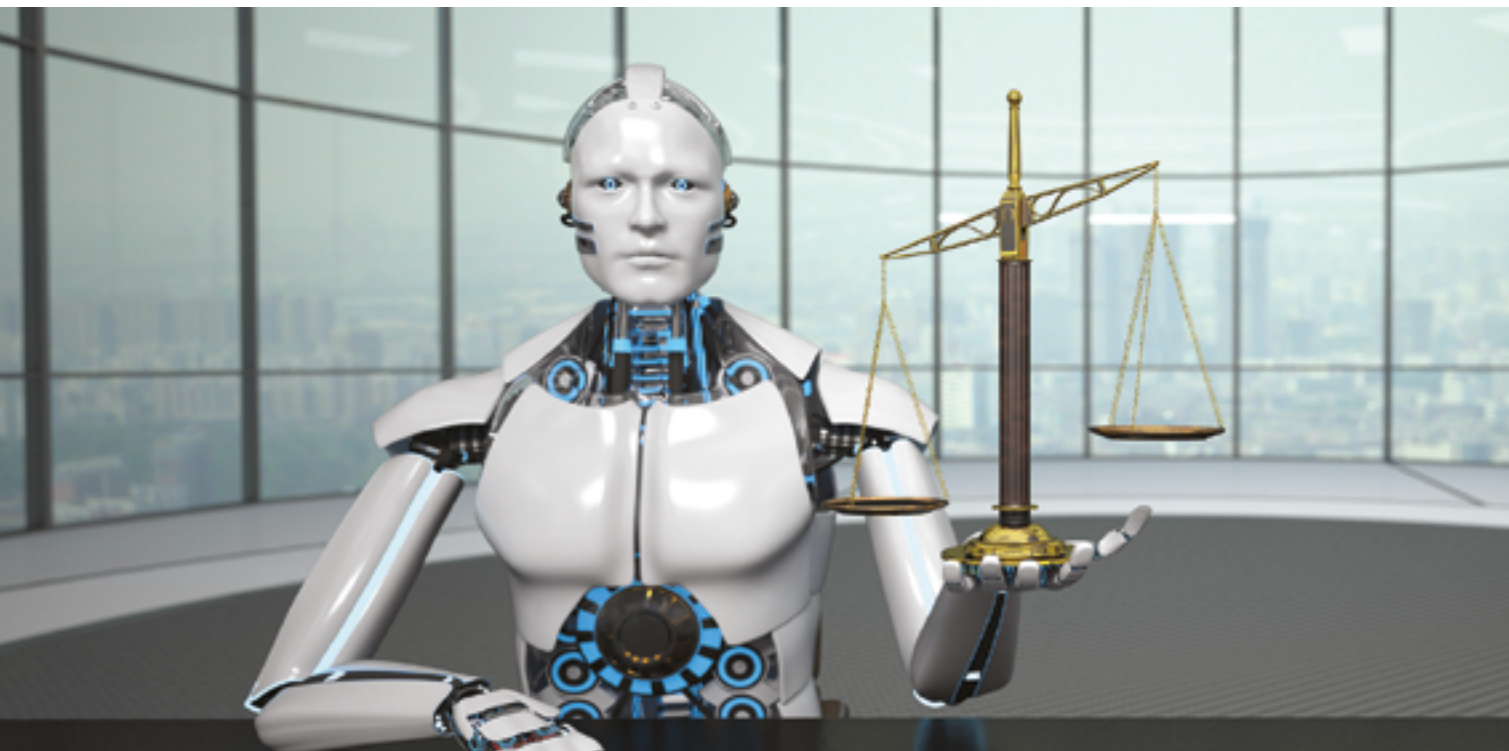
Ils conduisent les décideurs à privilégier les populations statistiques qui sont plus facilement rentables et moins entachées d'incertitudes. Par exemple : pousser un algorithme de publicité ciblée sur des populations qui ont le plus de chance d'acheter... et limiter ainsi la base d'apprentissage à une certaine catégorie d'individus.

* *Ce profiling s'inspire très largement de l'article publié par David Bounie et ses confrères de l'Institut Télécom ParisTech et de l'Université de Paris-Nanterre. Cf référence en bas de page.*

⁶ Article « Algorithmes : biais, discrimination et équité » par Patrice Bertail, David Bounie, Stephan Cléménçon et Patrick Waelbroeck, février 2019

UN BIAIS CÉLÈBRE

L'Algorithme de Risque de Récidive



En 2013-14, l'entreprise Northpointe étudie les données de 7000 personnes arrêtées par la justice dans le comté de Broward en Floride. Elle met alors au point un algorithme (nommé COMPAS) assignant à chaque individu un score de risque de récidive.

Les données prises en compte par COMPAS : le niveau d'éducation, l'employabilité, l'historique judiciaire de l'entourage...

L'algorithme est immédiatement adopté par certaines juridictions d'autres Etats, sans accès à la méthodologie de celui-ci car il constitue un secret commercial.

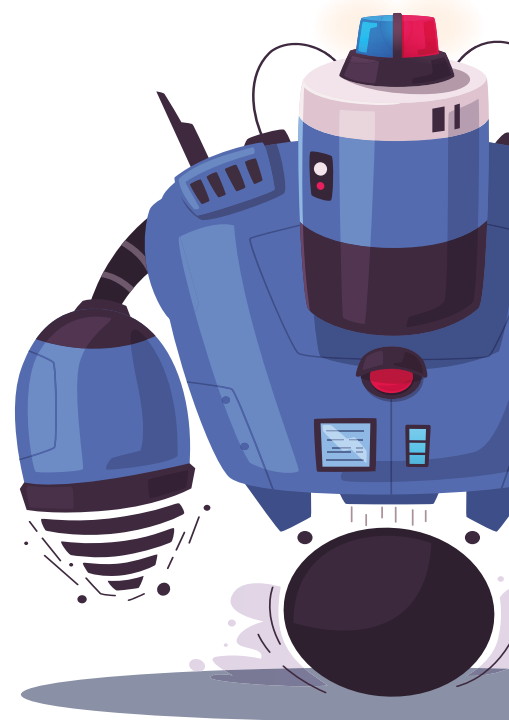
|| Les prévenus noirs seraient deux fois plus notés à tort comme de futurs criminels que les prévenus blancs. Et à l'inverse, les prévenus blancs seraient deux fois plus inscrits à tort en risque faible ||

Or, en mai 2016, des chercheurs⁷ publient un article dans lequel ils mettent en garde la justice américaine contre cet algorithme.

Ils s'appuient sur les cas de récidive réelle observés sur ces populations et concluent que les prévenus noirs y seraient deux fois plus notés à tort comme de futurs criminels que les prévenus blancs. Et à l'inverse, les prévenus blancs seraient deux fois plus inscrits à tort en « risque faible ».

Une justice à deux vitesses qui affole les cours d'assises... et qui amène les prévenus à inscrire l'explicabilité de l'algorithme en ligne de défense.

L'un d'eux, Eric Loomis, lourdement condamné à six ans d'emprisonnement pour participation à une fusillade a ainsi fait appel auprès de la Cour Suprême en pointant le caractère inexplicable de cet algorithme dans son procès... Avec gain de cause.



LE GLOSSAIRE DES MOTS EN -TÉ

Les défis qu'il reste à relever

#INTER-PRETABILITÉ

Être capable de comprendre le fonctionnement interne de certains modèles afin d'en déduire la règle de décision.

#QUANTITÉ

Être capable de fournir suffisamment de données d'apprentissage à la machine pour que sa base d'entraînement soit fiable. Selon Léon Bottou⁸ du Laboratoire d'IA de Facebook, « il faut plus qu'un humain peut voir ou lire en une vie pour entraîner les algorithmes de reconnaissance d'image ou de traduction les plus performants »...

#CAUSALITÉ

Être capable de faire comprendre à des algorithmes le lien de causalité existant entre des paires de variables afin que la corrélation n'aille que dans un sens. Ex : la montée en altitude fait baisser la température et non l'inverse.

#EXPLICABILITÉ

Être capable d'expliquer aux utilisateurs finaux le fonctionnement de l'algorithme pour garantir l'aspect vérifiable d'une décision et sa conformité au droit.

#AUDITABILITÉ

Être capable d'accéder au protocole et aux données de l'algorithme pour en garantir sa fiabilité. Peut être effectué par un organisme de contrôle indépendant pour préserver le secret commercial sur l'algorithme.

#RESPONSABILITÉ

Être capable de fournir une explication ou une compensation au préjudice subi, en cas de litige provoqué par un algorithme qu'on a mis sur le marché.

POUR
ALLER PLUS LOIN

Le LIVRE

- Cathy O'Neil (2016) : « **Algorithmes, la bombe à retardement** »

L'ÉVÉNEMENT

ACM FAT*: ACM Conference on Fairness, Accountability and Transparency

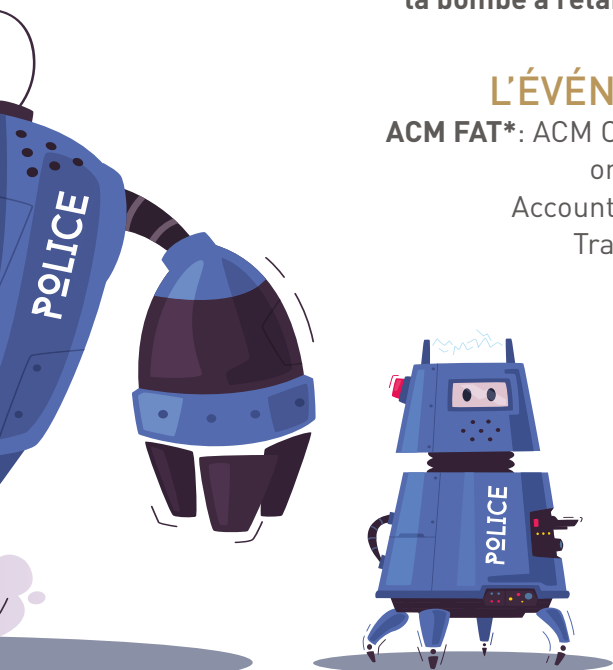
Les THINK TANKS

• Institutionnels

- **The AI Now Institute**, au sein de New York University (NYU)
- **The AI Fund**: fonds financé conjointement par le MIT, Harvard University, the Knight Foundation, Omidyar Network et l'un des fondateurs de LinkedIn.
- **World Economic Forum** : il demande une « Magna Carta » de l'IA.

• Privés

- **Facebook AI Research (FAIR)**: le programme de recherche de Facebook
- **FATE**: le programme de recherche de Microsoft
- **PAIR - People + AI Research Initiative**: le programme de recherche de Google
- **Partnership on AI** : consortium créé par Amazon, Google, Facebook, IBM, Microsoft et rejoint par des ONG et des entreprises comme Salesforce ou EBay.



⁸ Lors d'un exposé le 6 mars 2018 à Sorbonne Université

INTERVIEW

Qu'est-ce qui rend le problème des biais plus critique aujourd'hui que dans le traitement statistique d'hier ?

Principalement la méconnaissance des sources d'apprentissage. Auparavant, on avait des petits échantillons représentatifs de certaines populations statistiques pour lesquels on avait un plan de sondage. On connaissait les probabilités d'inclusion de ces individus dans les échantillons sélectionnés. Avec le Big Data et l'IA, les données sont massivement issues du web et on connaît mal les caractéristiques des individus dont elles sont issues : difficile d'en tirer des conclusions sur ceux qu'elles sont censées représenter !

C'est-à-dire ?

Tout le monde ne s'exprime pas ou n'est pas actif sur internet ! Les réseaux sociaux par exemple sont le fait d'une minorité qui contribue, écrit du texte, poste des commentaires et possède souvent des caractéristiques singulières en termes d'âge ou de milieu social. Extrapoler à partir de ces données est donc risqué pour représenter une population dans son ensemble.

Comment identifie-t-on des biais ?

C'est toute la difficulté justement : car dans l'absolu, il faudrait avoir toute la population à disposition pour prouver les biais, ce qui est impossible évidemment. Du coup il faut utiliser des modèles et opérer de multiples tests sur ces populations afin de montrer les écarts : c'est ce que nous allons faire grandeur nature dans les prochains mois avec des industriels et des partenaires publics. A noter que ce n'est pas évident de trouver des entreprises qui acceptent de se prêter à ces tests car c'est souvent difficile pour elles d'admettre que leurs données peuvent être sujettes à des biais...

Quels sont les outils scientifiques à disposition pour corriger les biais ?

Comme expliqué dans l'article que j'ai co-écrit, il existe plusieurs méthodes telles que le redressement, le ré-échantillonnage ou le fait de compléter l'information. Il s'agit bien souvent de créer des modèles correctifs pour compenser l'absence ou la sur-représentation de certaines données. Cela ne corrige pas entièrement l'algorithme mais cela peut limiter l'impact des biais.

Où en est la France en matière de Recherche sur la question ?

Il y a quelques pools de recherche en France, à Toulouse ou à Paris Saclay notamment, mais l'essentiel des travaux sur la question provient de l'étranger, Etats-Unis et Chine en tête bien sûr, avec une vraie avance des GAFAM. C'est souvent très difficile de trouver des cas concrets, des travaux de recherche appliqués sur la question : l'essentiel des papiers, y compris américains, se concentrent sur des aspects théoriques.

|| Dans l'absolu,
il faudrait
avoir toute
la population
à disposition
pour prouver
les biais,
ce qui est
impossible
évidemment ||

L'explicabilité est pourtant un enjeu de politique publique...

Oui, et c'est pour cela que les travaux que nous menons mêlent à la fois des mathématiques, des computer sciences ET des sciences sociales : parce qu'il y a un impact social derrière l'utilisation de la plupart des algorithmes. Bien sûr, si vous créez un modèle algorithmique pour jouer au jeu de go, ce n'est pas un souci s'il n'est pas totalement explicable : l'enjeu social sera faible et vous viserez plutôt son efficacité. Mais pour un diagnostic médical ou un conseil financier qui ont donc un coût social fort, il faut pouvoir mettre en place des mécanismes de contrôle.

Que suggérez-vous ?

Progresser sur l'explicabilité des algorithmes via la Recherche bien sûr. Mais, à ce stade, on est encore confrontés à la boîte noire pour un certain nombre de modèles : les réseaux de neurones sont très difficiles à expliquer tandis que les arbres de décision deviennent totalement opaques à partir d'un certain nombre de variables. Du coup, il faudrait a minima instaurer une transparence sur les protocoles qui ont mené à la création de l'algorithme : il faudrait créer une sorte de notice d'utilisation de l'algorithme, comme pour les médicaments qui permette de comprendre les éléments constitutifs des résultats (la nature des jeux de données, les modèles utilisés, les mises à jour, etc).

Avec la mise en place d'une autorisation de mise sur le marché ?

Tout à fait. Il faudrait que le régulateur puisse se prononcer sur la possibilité ou non d'utiliser tel ou tel algorithme, tel ou tel jeu de données, etc. Des régulateurs sectoriels - comme l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) pour le secteur financier - pourraient ainsi avoir leur mot à dire sur le recours à certaines données et algorithmes dans l'établissement d'un score de crédit par exemple. Et puis il faudrait pouvoir constamment réévaluer la performance de l'algorithme, comme un « contrôle technique » régulier. En bref, il y a beaucoup de chantiers à lancer... en espérant que les algorithmes n'aillent pas plus vite que nous.



David
Bounie

Head of Economics and Social
Sciences, Telecom ParisTech

Il faudrait créer une **notice d'utilisation des algorithmes**

|| Pour un diagnostic médical ou un conseil financier qui ont donc un coût social fort, il faut pouvoir mettre en place des mécanismes de contrôle. ||

SÉCURITÉ DES DONNÉES, L'HEURE DU DOUTE



Il y avait les pare-feux, les proxy, et on pensait le réseau protégé de l'intérieur.

Et puis il y eut le Cloud, l'explosion des données et l'ouverture sur Internet et on a découvert que les menaces s'étaient démultipliées...

|| On est passé d'une sécurité périmétrique à une sécurité applicative, à l'intérieur même de chaque application ||

... En 2018, 80% des entreprises françaises auraient ainsi subi une cyberattaque avérée (cf encadré), malgré des dispositifs et des solutions de sécurité intégrée. « On est passé d'une sécurité périmétrique à une sécurité applicative, à l'intérieur même de chaque application », explique Laurent Grangeau, Cloud Solutions Architect chez Sogeti.

DevSecOps

Soit l'intégration de process de sécurisation dès la phase amont, au niveau administrateur lors de la création et de la gestion des applications. « Cette approche DevSecOps (comprendre : Développement-Sécurité-Opérations) nous conduit à scanner les applications et leurs potentielles failles de sécurité tout au long du projet ».

Vae Leaktis

Mais alors... les données du Big Data sont-elles effectivement bien protégées ? La prolifération des ransomware (littéralement : « rançongiciel » ou logiciel de rançon) témoigne du nouveau business que se livrent les cyberattaquants : extorquer des données stratégiques à des entreprises pour menacer ensuite de les faire circuler à la concurrence ou sur le Net... à moins de payer une rançon.

« Pour se prémunir contre ces fuites de données, tout le travail se situe autour de la gestion des droits et des accès, d'autant plus que, concernant les données personnelles, le RGPD applique maintenant de lourdes sanctions financières (4% du CA) pour toute fuite détectée » continue Laurent Grangeau.

Kerberos, Knox, Senry

« La difficulté, c'est qu'il peut y avoir plusieurs points d'entrée sur une donnée dans un cluster Hadoop », complète Mohamed Mahdi Benaïssa, Big Data Solutions Architect chez Umanis. « Il faut donc être capable de mettre en place des outils qui gèrent à la fois l'authentification (ex : Knox), puis l'autorisation (ex : Ranger, Senry) et la traçabilité... tout cela sans ralentir le système en termes de per-

formance ou de facilité d'usage ! ». Un sacré challenge qui justifie selon lui la mise en place d'outils de gouvernance globale au niveau d'Hadoop (outils distribués de Kerberos, la solution open source) pour gérer ces différents besoins, y compris dans les applications. « Mais alors, soupire Alain Boullé, président du CESIN, on accepte de confier toute son architecture de sécurité à des solutions du marché non souveraines, et bien souvent packagées... ».

SIEM

Quelle solution entrevoir alors face au risque grandissant de « data leaks » et l'absence d'alternatives aux logiciels des GAFAs ? « L'utilisation des traitements de Machine Learning pour détecter les risques potentiels est pour nous une avancée », poursuit ce dernier (cf interview). La corrélation des données d'événements intervenus sur le réseau, y compris au niveau le plus fin grâce aux logiciels de SIEM (Security Information and Event Management) est d'ailleurs devenue un indispensable du security management. D'autres expérimentations de machine learning « a posteriori » font également florès : en octobre dernier, la start-up CybelAngel annonce ainsi une levée de fonds de 10 millions d'euros pour poursuivre le développement de sa solution, un outil de scan du web à la recherche des données « égarées » par les entreprises...

Rogue AI

Pourtant, l'intelligence artificielle peut, elle aussi, représenter une faille du système, ne serait-ce qu'en introduisant des données litigieuses au cœur du modèle d'apprentissage. En 2016, une équipe américaine découvre ainsi qu'en observant le fonctionnement des réponses d'un algorithme propriétaire, elle était capable de reconstituer celui-ci et de lui soumettre des données frauduleuses... pour le faire dérailler. Machiavélique n'est-ce pas ?

Machine learning, gouvernance, DevSecOps, rôle du RSSI, les enjeux sont donc explosifs pour les années qui s'ouvrent... Comme si cette « heure du doute » avait finalement sonné l'An II de la sécurité des Big Data.



||| L'utilisation des traitements de Machine Learning pour détecter les risques potentiels est pour nous une avancée |||

L'INFOGRAPHIE

Diagnostic des risques cyber en France

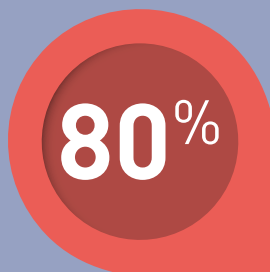


Issu du Baromètre 2019 de la Cybersécurité des entreprises⁹, sondage OpinionWay pour le CESIN

59% des entreprises ont constaté un impact sur leur business

(+10 points par rapport à 2018).

Impacts constatés : ralentissement ou arrêt de la production, indisponibilité du site web, perte de CA, retards sur la livraison...



En 2018, 80% des entreprises françaises ont subi au moins 1 cyberattaque avérée

Les principales attaques

Phishing ou spear-phishing : 73%

Arnaque au Président : 50%

Infection par un malware, cryptolocker ou ransomware : 44%



73%



50%



44%

Les principaux risques



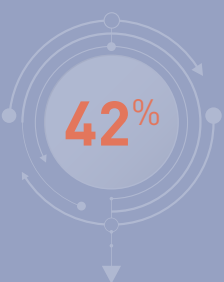
Shadow IT (mise en place d'applications non approuvées)



Vulnérabilités résiduelles permanentes



Erreur de manipulation d'un salarié



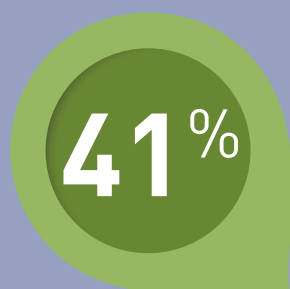
Connexion de postes non approuvée sur le réseau de l'entreprise



Exposition de données sur un système d'un prestataire (défaut de configuration/négligence)



Mise à jour logicielle entraînant des dysfonctionnements ou des altérations de données



Moins d'une personne sur deux considère que son entreprise est préparée

à gérer une cyber-attaque de grande ampleur



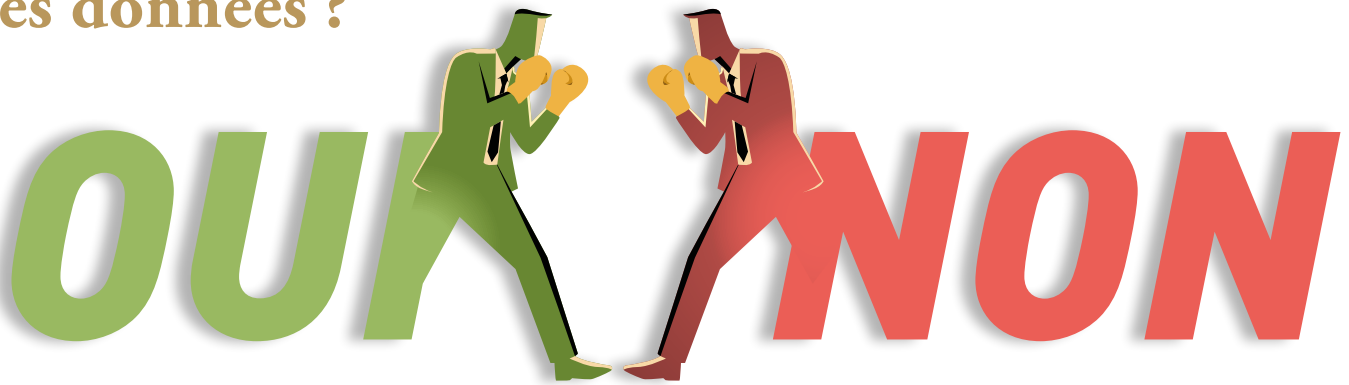
50% des entreprises ont souscrit une cyber-assurance

(+10 points)

⁹ Étude quantitative réalisée par OpinionWay auprès de 174 membres du CESIN, à partir du fichier membre du CESIN (498 contacts).

LE MATCH DES IDÉES

Le Cloud est-il plus risqué pour la sécurité des données ?



|| Par nature, le Cloud augmente la surface d'exposition aux menaces et donne plus de fenêtres d'opportunités aux attaquants. Mais c'est surtout le challenge de la dispersion des données qui m'inquiète : difficile de protéger chaque donnée quand celles-ci sont éparpillées dans les serveurs d'une architecture multi-Cloud. Certes, il existe des solutions de sécurité vendues en package avec ces services multi-Cloud mais sont-ils toujours à la hauteur des enjeux de sécurité des entreprises quand on connaît la complexité du paysage technologique ? ||

|| Avec la force de frappe technologique des Cloud providers et la finesse des services de détection qu'ils sont capables de mettre en place, le Cloud me semble mieux préparé pour répondre à des attaques de grande envergure. Un exemple récent chez l'un de nos clients ? Un serveur virtuel qui s'est trouvé éteint automatiquement par le Cloud provider sans que le DSI n'ait détecté quoi que ce soit de son côté. Après enquête il y avait bien une attaque de mineurs de bitcoin... stoppée immédiatement par le système interne du Cloud. Plutôt rassurant non ? ||



Alain *Bouillé*

Président du CESIN



Laurent *Grangeau*

Cloud Solutions Architect - Sogeti

INTERVIEW

Les risques en matière de cybersécurité ont-ils augmenté dans les entreprises ?

Assurément ! Déjà avec le recours plus massif au Cloud (*cf encadré*). Mais aussi parce que les systèmes d'information des entreprises sont désormais beaucoup plus ouverts sur Internet : la hausse du nombre d'applications par utilisateur en entreprise est devenue effarante. Et, sur toutes ces applications, il y en a des centaines qui sont inconnues de la DSI !

Comment l'IA peut-elle transformer ce risque en opportunité ?

L'intelligence artificielle peut être une menace si elle est l'arme des attaquants... mais une grande opportunité si elle est utilisée pour se défendre ! Au CESIN, nous croyons beaucoup à l'utilisation des traitements IA pour détecter les attaques et les fuites : les outils Big Data devraient permettre, à terme, une plus grande automatisation des processus de surveillance au cœur de l'activité de l'entreprise. Sans s'affranchir bien sûr d'une analyse humaine pour compléter le check-up.

Ne risque-t-on pas d'aller vers trop de surveillance ?

La limite que l'on peut y voir c'est la surdétention : lorsque les algorithmes détecteront des failles mineures qui n'ont pas vocation à être considérées comme des menaces. Il faudra alors affiner le système d'alerte pour éviter que les opérationnels n'adoptent un comportement d'évitement (ex : déclassification de documents) pour échapper aux contrôles...

Et la sécurité « by design », c'est possible ?

Effectivement, je crois beaucoup à l'intégration de la sécurité de manière native, directement dans le code de chaque applicatif ou de chaque objet si on se parle d'IoT (capteur, scanner, caméra de surveillance, voiture autonome...). Là encore, l'IA peut aider : certains outils sont conçus pour détecter les vulnérabilités directement dans le code. Au-delà de la valeur immédiate que cela apporte (lever la menace), c'est un outil de formation inestimable pour les développeurs qui apprennent ainsi « par l'erreur » et seront plus attentifs à l'avenir à ces zones de fragilité.

Vous semblez soucieux quant au niveau de protection des données « stratégiques » des entreprises...

Contrairement aux données personnelles externes qui sont passées sous le filtre du RGPD, les données sensibles à usages internes (secret industriel, données financières, compte-rendu business, etc) sont souvent exposées dans des solutions commerciales non souveraines au même titre que les autres données de l'entreprise. C'est le cas des solutions de messagerie et des outils collaboratifs proposés uniquement par des acteurs étrangers. Nul ne sait véritablement la nature de ce qu'il externalise lorsque qu'il confie sa messagerie à un provider du cloud. Autre menace : les entreprises ont tendance à stocker des don-

nées non sensibles dans leurs data lakes, sans identifier le risque que le traitement de celles-ci peut faire peser sur l'entreprise : si, individuellement, elles n'ont pas de grande valeur stratégique, elles peuvent constituer une information capitale (et donc vulnérable) une fois agglomérées et traitées...

Que peut-on faire pour limiter ces risques ?

Pour moi, il s'agit d'un enjeu d'intelligence économique nationale : on ne peut confier à des solutions non souveraines la protection de ces actifs stratégiques. Il faut donc en appeler à la puissance publique pour réaliser ce qu'elle a déjà accompli côté données personnelles avec le RGPD : passer par la législation au niveau français et européen. Il faut encourager le recours à des logiciels souverains pour sortir des solutions clés en main du marché : créer de véritables « opérateurs de confiance » nationaux.

Vous insistez également sur le rôle du RSSI...

Oui et c'est lié à mon propos précédent : pour imposer ces logiciels qui sortent « des clous », il faudrait pouvoir les faire porter par des RSSI reconnus dans l'organisation. Pour cela, j'appelle à une meilleure implication des RSSI dans la gouvernance de la data d'entreprise, qu'ils soient réellement partenaires des CDO pour intervenir a priori et non, comme souvent encore, a posteriori.

|| Les outils Big Data devraient permettre, à terme, une plus grande automatisation des processus de surveillance ||

|| Nul ne sait
véritablement
la nature de ce
qu'il externalise
lorsque qu'il
confie sa messagerie
à un provider
du Cloud ||



Alain
Bouillé

Président du CESIN¹⁰

On ne peut confier **nos**
data stratégiques
à des **solutions**
non souveraines

¹⁰ Club des Experts de la Sécurité
de l'Information et du Numérique

TECHNOS

Si le début des années 2010 avait fait du Big Data l'apôtre du « toujours plus » et de l'infiniment grand (quitte à surcharger les serveurs Cloud), la fin de la décennie s'oriente, elle, vers une recherche du dimensionnement au plus fin et au plus précis qui pourrait bien transformer le paysage Big Data en architecture « à la Tetris ».

Conteneurisation des applications, approches serverless et orchestration des données, le Big Data se décline désormais en briques interconnectées, réparties sur plusieurs serveurs, qui mutualisent leurs fonctionnalités et leurs data besoin par besoin, en fonction de l'optimum que leur combinaison représente : pas de temps de chargement ou de traitement inutile, pas de ressources gâchées ou inertes, désormais on cherche le temps réel et on rationalise !

Cette modularité qui a commencé à atteindre le Cloud depuis deux ans est en passe de s'imposer aussi dans l'écosystème Hadoop et pourrait bien constituer une remise à niveau majeure pour le framework numéro 1 du Big Data.

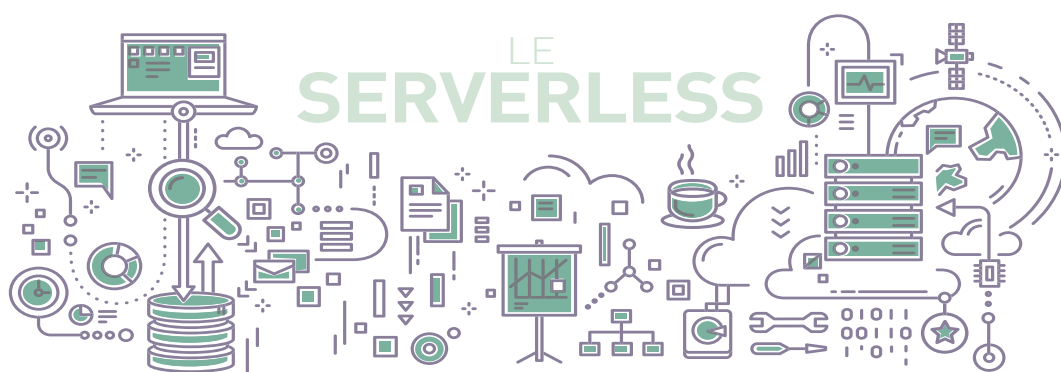
Avec à la clé, une autre révolution qui se profile : l'entrée dans l'IoT et l'ère du Edge computing...

VERS L'INFINIMENT PRÉCIS ?



HADOOP, CLOUD : QUAND LE BIG DATA SE FAIT MODULAIRE

Le Buzzword 2019



Qu'est-ce-que c'est ? Une approche « à la demande » du Cloud computing, dans laquelle les ressources d'infrastructure sont attribuées uniquement en fonction du besoin de l'application : le développeur client n'a que le code à entrer et le cloud provider fournit ponctuellement toute l'infrastructure nécessaire (hébergement, bases de données, puissance de calcul, maintenance...) pour exécuter ce code.

Comment ça marche ? Le serverless repose sur des architectures en containers et en micro-services, directement administrées par le Cloud provider et mobilisées grâce à des données d'événements : pour un déclencheur donné (data x), le cloud provider interroge son orchestrateur de containers qui lui trouvera la place et les ressources nécessaires pour une exécution éphémère.

Pour quel usage data ? Le serverless porte essentiellement sur les applications pilotées par des données d'événements : applications bots (ex : accueil d'un nouveau visiteur sur un site web), chargement d'images, traitement batch sur des usages de monitoring, etc. Et bien sûr : les traitements IoT.

Son surnom : FaaS ou « Function as a Service » ou « Approche applicative à-la-fonction »

Est souvent couplé à : « Architecture serverless », « Approche serverless », « T'as pensé à passer en serverless ? », « What is serverless ? »...

Famille de buzzwords associée : « containers », « microservices », « PaaS », « Kubernetes »

Note sur l'échelle du buzzword : après sa mention dans le Top Ten Trends 2019 de Gartner, il passe le cap des 17/20. D'autant que le cabinet d'études mise sur un potentiel d'adoption de 20% des entreprises de la planète d'ici 2020.

Ses fournisseurs stars : AWS, Microsoft Azure.

Sa petite coquetterie : Contrairement à sa traduction littérale, « serverless » ne signifie pas sans serveur. Il y a toujours un ou plusieurs serveurs pour héberger l'infrastructure mais ceux-ci sont entièrement gérés par le Cloud provider et donc

« invisibles » pour le client. D'où la notion associée de « NoOps » en lieu et place de « DevOps » côté client.

Son « plus » pour les développeurs : Finies les anticipations de charges sur les serveurs, les problématiques de maintenance et de dimensionnement des infrastructures ! Du code, rien que du code !

Son « plus » pour le contrôleur de gestion : Une facturation à l'usage. On ne paie que les ressources que l'on consomme.

Le hic : Par définition le serverless ne supporte pas de charges lourdes et continues, ne serait-ce que pour des raisons économiques. Pour toute exécution complexe ou de long terme, préférer les machines virtuelles (VM).

Le GROS hic : Le Cloud provider obtient toute latitude sur l'administration du réseau. Et sur les questions de sécurité...

Kubernetes et les containers expliqués à ceux qui n'y comprennent rien

Dans le monde physique, un container, c'est une boîte scellée qui contient tous les composants d'un produit pour une livraison.

Dans le monde informatique, c'est le même principe : un container, c'est un objet informatique scellé qui contient toutes les pièces concourant à l'exécution d'une application. On y trouve l'application en elle-même, mais aussi le noyau du système d'exploitation, les fichiers de configuration et tous les autres fichiers et bibliothèques accompagnant l'application (ex : son, image). L'un des modèles de containers les plus connus s'appelle Docker.

Un peu comme on charge une marchandise dans un cargo pour un long voyage, placer les applications dans des containers permet ainsi de les faire migrer vers d'autres systèmes sans abîmer leur intégrité et la multiplicité des composants : à l'arrivée, l'application pourra s'exécuter sur n'importe quel serveur.

A plus grande échelle, si l'on décide de « conteneuriser » l'ensemble des applications de son SI, celui-ci finira donc par ressembler à un gigantesque puzzle agrégeant une multitude de containers. Cette modularité aura alors un grand avantage : trouver de la place pour chaque container dans différents serveurs. Et optimiser ainsi l'espace (et le coût) du SI...

Problème : pour trouver une place à chaque container et garder ensuite en mémoire son emplacement pour le mobiliser quand l'opérateur voudra l'utiliser, il faudra un programme de supervision qui orchestrera les ressources et associera dynamiquement les composants entre eux. On appelle ces programmes des « orchestrateurs » et le plus connu d'entre eux s'appelle... Kubernetes.

Pas si compliqué que
ça finalement, non ?

Pour éblouir le jury de l'oral :

**Kubernetes en grec,
ça signifie « celui qui tient
la barre », soit « le pilote »**

La citation

**« Créer une boîte fermée à clé permet
de résoudre une partie du problème en aval
(vous savez ce que vous avez), mais pas
le problème en amont (vous ne savez
pas de quoi vous dépendez) »,**

**Rob Hirschfeld, CEO de RackN et
membre du conseil d'administration
de la Fondation OpenStack.**

Inspirés ?
Vous avez 4 heures.

INTERVIEW



Laurent
Grangeau

Cloud Solutions Architect
chez Sogeti

|| Aujourd'hui si l'on détecte une faille de sécurité dans un container, personne n'est capable d'endosser la responsabilité du problème ni de la corriger ||

J'encourage au maximum
mes clients
à opter pour
le PaaS et le **serverless**

Vers quels types d'infrastructures orientez-vous vos clients ?

Mon boulot est de tirer parti au maximum de la puissance du Cloud donc, autant que faire se peut, j'essaie de les diriger vers du PaaS et du serverless pour garantir une scalabilité maîtrisée, y compris dans les coûts. Pour l'instant, elles sont majoritairement axées sur des modèles IaaS car leur migration sur ce type d'infrastructures a commencé plus tôt, vers 2015-16, mais certains commencent à explorer le PaaS.

Pourtant, cette complexité dans l'architecture amène de plus en plus d'opacité dans la construction des coûts du Cloud...

... Raison pour laquelle les entreprises acceptent de moins en moins que certains serveurs fonctionnent pour rien en mode IaaS. Certaines embauchent des FinOps pour identifier au plus près leur consommation Cloud et être capables d'éteindre les serveurs quand ceux-ci tournent à vide. Mais c'est plutôt pénible et chronophage...

La sécurité est devenue un enjeu fort...

Effectivement, au niveau du Cloud, on cherche à développer des services de sécurité de plus en plus poussés pour repousser les attaques de type DDoS (Distributed Denial of Services), qui bloquent les infrastructures. Le Machine Learning nous aide de plus en plus à accéder à la donnée ultra fine pour détecter les failles. Mais ce qui m'inquiète un peu, c'est la gestion de la sécurité avec des approches en containers.

C'est-à-dire ?

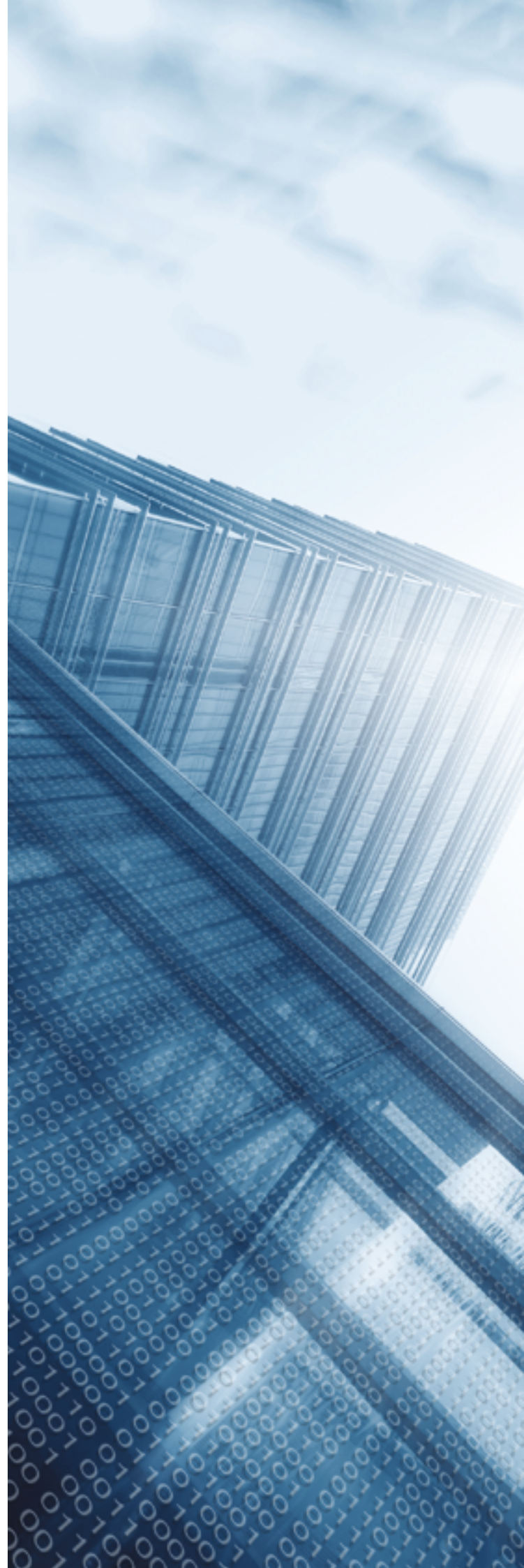
Aujourd'hui si l'on détecte une faille de sécurité dans un container, personne n'est capable d'endosser la responsabilité du problème ni de la corriger. Est-ce le cloud provider ? Est-ce le client utilisateur ? Idem concernant le versioning de chaque application : on ne sait jamais s'il faut ajouter des patches ou pas pour aboutir à la mise à jour du marché. Il y a un vrai problème d'observabilité des containers, qui limite la capacité de consultation et d'intervention sur ceux-ci.

Pourtant, les containers ont beaucoup ouvert le marché du Cloud...

Disons qu'ils ont permis de faciliter la migration du on premise vers le Cloud, en favorisant une ouverture des systèmes (le container étant universel, pas de vendor locking). Les entreprises ont pu ainsi développer des modèles agiles sans forcément changer toute leur infrastructure. Cependant, on ne peut pas encore vraiment parler de portabilité multi-cloud dans la mesure où il n'existe pas encore de standards transversaux pour les applications en containers : chaque acteur (AWS, Azure, Google Cloud) dispose de son propre protocole.

Voyez-vous des tendances émerger pour les années à venir ?

Justement, je constate une vraie volonté d'établir des standards autour de Kubernetes et des approches micro-services. Les protocoles de type Service Mesh qui permettent de mailler et faire discuter entre eux les micro-services sont en pleine phase exploratoire : on peut notamment citer Istio qui est à l'étude en open source. Et puis, pour favoriser les déploiements serverless d'un cloud à l'autre (ou du on premise au Cloud), on cherche de plus en plus à établir Kubernetes comme un moteur standard et automatiser ainsi son développement, via des projets comme KNative¹.





LE TOP FAKE

La révision de la data locality va être une vraie révolution pour Hadoop



Mohamed Mahdi Benaïssa

Big Data Solutions Architect et IT Manager chez Umanis

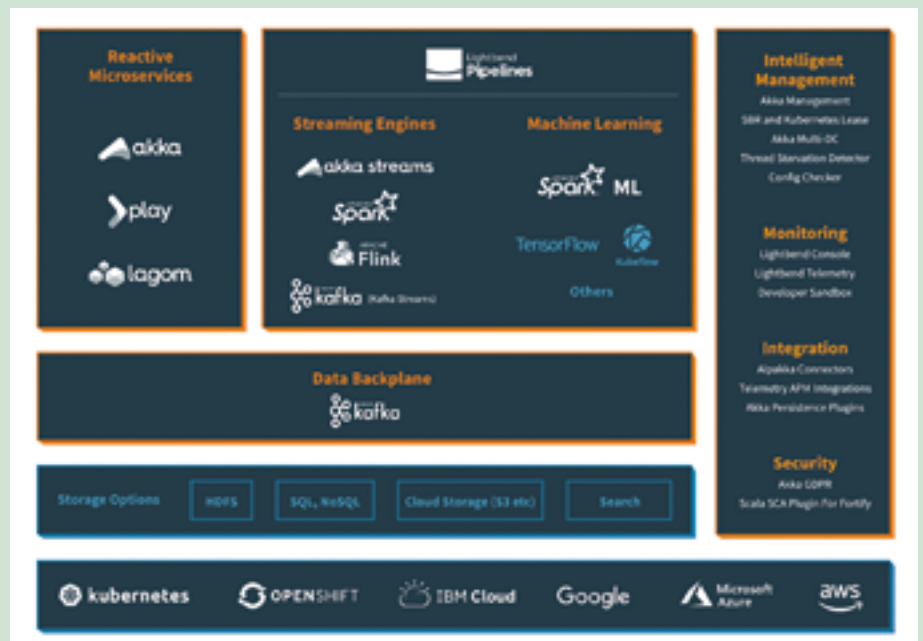


#FAKE 1 : HADOOP EST UN ET INDIVISIBLE

On a toujours tendance à penser Hadoop comme une entité unique. Comme s'il suffisait d'appuyer sur un seul bouton pour faire du Big Data... En réalité, Hadoop est une boîte à outils qui agrège une multiplicité de composants (Hive, Spark, etc), responsables chacun d'une fonctionnalité (stockage, NoSQL, sécurité, streaming, traitement, BI, etc). Quand on installe Hadoop, on installe donc une grande boîte à outils dans le même cluster de machines... ce qui conduit désormais le marché à questionner cette lourdeur et chercher des approches plus modulaires.

#FAKE 2 : HADOOP A LE MONOPOLE DU BIG DATA

L'écosystème Hadoop n'a, à ce jour, pas trouvé de concurrent « aussi global » sur le marché : c'est la seule plateforme qui réunisse toutes les fonctionnalités évoquées pour faire du Big Data. Cependant, les utilisateurs se sont rendus compte qu'ils n'avaient pas forcément besoin de tout le « package » : ils pouvaient faire du Big Data en se concentrant sur un ou plusieurs outils ou frameworks - par exemple Spark pour le streaming, Impala pour le requêtage SQL, etc. Et c'est là que des alternatives ont vu le jour, fonctionnalité par fonctionnalité, surtout dans le Cloud sous forme de « services managés » : par exemple, Azure Data Lake Storage (Microsoft) ou S3 (Amazon Web Services) qui peuvent remplacer HDFS sur la partie pur stockage (sans calcul). Autre type d'alternatives : les boîtes à outils spécialisées qui ciblent des problématiques particulières comme la Fast Data Platform de l'éditeur Lightbend. Hadoop est toujours la plateforme numéro 1 du Big Data mais, désormais, les utilisateurs se sont tournés vers des sous-marchés plus spécialisés.



Graphique d'illustration de la Fast Data platform de Lightbend

#FAKE 3 : HADOOP, C'EST FORCÉMENT LA MEILLEURE SOLUTION

Tout dépend de l'usage que vous souhaitez en faire : si vous avez une architecture grande et complexe, il vaut mieux opter pour Hadoop qui assurera un ensemble cohérent pour adresser toutes vos problématiques. Si votre architecture est petite avec des problématiques spécialisées, le modulaire peut apporter des réponses plus pointues à moindre coût (Hadoop étant par nature un cluster coûteux à mettre en place et maintenir).

#FAKE 4 : HADOOP N'A PAS CHANGÉ

Depuis son apparition au milieu des années 2000 et le projet Apache qui l'a fait naître, Hadoop a connu beaucoup d'évolutions de chacun de ses composants. Mais le principal changement, on le doit à la version 2 en 2013 et l'apparition de YARN (Yet Another Resource Negotiator) qui lui permet d'exécuter d'autres types d'applications que le seul traitement batch de MapReduce : dépassant le modèle stockage / traitement distribué, l'écosystème Hadoop découvre Spark, Storm, le streaming et d'autres enjeux comme la gouvernance ou la sécurité. Avant le prochain big bang de la data locality (cf plus bas) ?

#FAKE 5 : MAPREDUCE EST MORT

Si MapReduce est à la base le nom d'un paradigme de programmation - celui du traitement distribué - il est devenu par extension le nom d'une API (l'API Java MapReduce d'Hadoop). Et si l'API MapReduce historique est moins utilisée (car moins rapide que Spark ou Storm) et a été remplacée par TEZ dans la plupart des composants incontournableables d'Hadoop (Hive, Pig, etc.), le modèle de programmation, lui, perdure : on le retrouve d'ailleurs dans Spark et dans la plupart des outils de l'écosystème Big Data. Quant à l'application historique MapReduce, il ne faut pas l'enterrer non plus : pour des volumes de données massifs, elle reste toujours plus efficace que Spark dont le modèle en mémoire sature plus vite...

#FAKE 6 : SPARK EST LE PLUS PERFORMANT

Spark n'est pas le plus performant des outils de traitement mais c'est sa polyvalence qui lui donne son caractère incontournable. Spark est capable d'accéder à n'importe quel type de données et il concentre une pluralité de modules (pour faire de la BI, pour faire du streaming / temps réel, pour faire du Python et du machine learning, etc) qui lui donnent un véritable avantage comparatif : pour chaque module, on trouvera toujours un outil plus performant (ex : pour le SQL, Impala / pour le streaming, Storm) mais aucun n'est capable de réunir autant de fonctionnalités à ce niveau.

#FAKE 7 : UNE DONNÉE HADOOP NE VOYAGE PAS

Nativement, Hadoop est conçu pour fonctionner en cluster dans un même datacenter : quand on processe des données sur Hadoop, elles sont stockées et traitées dans des machines situées au même endroit. Il n'y a pas de possibilité de diviser l'architecture stockage/traitement entre la France et l'Inde par exemple : c'est ce qu'on appelle la data locality. Or, avec la fusion entre Cloudera et Hortonworks, de nouvelles perspectives voient le jour pour remettre en cause ce paradigme : désormais on envisage de créer des clusters de stockage d'un côté (par exemple en Inde) et des clusters de traitement de l'autre (par exemple en France) pour décorréliser les deux. Donc, à terme, une donnée, même distante, pourra atterrir et être traitée dans le datacenter le plus proche... accélérant ainsi considérablement les perspectives de temps réel.



Graphique d'illustration de la Cloudera Data Platform, résultant de la fusion avec Hortonworks

#FAKE 8 : KUBERNETES ET HADOOP N'ONT RIEN À VOIR

Avec la révision de la data locality (cf ci-dessus), l'affectation des données sur les clusters de stockage et de traitement sera largement dépendante des orchestrateurs en place. Si YARN a jusqu'à présent parfaitement effectué son rôle de resource manager pour affecter les tâches d'exécution à des machines sur un même cluster, il deviendra important d'orchestrer dynamiquement le stockage et le traitement sur plusieurs machines, tout en optimisant la gestion des ressources sur la bande passante... D'où le recours probable à Kubernetes pour remplacer YARN sur cette fonction.

#FAKE 9 : LES CLUSTERS HADOOP SONT SATURÉS

Dans le modèle actuel reposant sur la data locality, les nœuds (master nodes) qui font la coordination entre les nœuds du stockage et les nœuds des traitements constituent le goulet d'étranglement du système : c'est leur saturation qui peut empêcher la scalabilité d'un cluster Hadoop. Avec la décorrélation stockage/traitement, cette coordination n'aura plus à être effectuée en simultané pour un grand nombre de nœuds. Le cluster sera divisé en plusieurs parties et chaque sous-ensemble de nœuds (stockage et calcul) pourra être géré par un master. On pourra alors augmenter le nombre de machines disponibles pour exécuter les tâches.

#FAKE 10 : LE EDGE COMPUTING, C'EST ENCORE LOIN...

Le Hadoop post-datalocality va permettre de traiter une nouvelle génération de cas d'usage en lien avec l'IoT : si le cluster est éparpillé sur plusieurs datacenters, la donnée choisira le plus proche pour être traitée et renvoyée ensuite au device, autorisant ainsi un traitement edge. Dans le cas de la voiture autonome par exemple, on peut donc s'attendre à une prise de décision beaucoup plus rapide, sans latence. Cela peut expliquer pourquoi Cloudera communique désormais sur un slogan « From the Edge to AI »...

INTERVIEW

LES DÉFIS DU **EDGE** **INDUSTRIEL**

On l'a vu précédemment : la révision de la data locality pourrait donner un sérieux coup d'accélérateur au Edge Computing côté stockage et traitement. Mais que se passera-t-il quand Cloud, Fog et Edge se trouveront à cohabiter au sein des mêmes usines ? Et plus globalement, au-delà du classique cas d'usage de la voiture connectée, que peut-on attendre du Edge Computing dans un environnement industriel où les data tournent par nature en circuit fermé, avec un haut degré d'automatisation ?

Le Cloud

été gagné par les **GAFA**,

le **Edge** reste une

opportunité pour la

France et l'Europe



Alexandre
Lacour

Référent Industrie 4.0 au
HUB Institute

Côté face, Alexandre Lacour est Directeur exécutif de Someflu, une entreprise qui conçoit et fabrique des pompes centrifuges pour transférer des fluides dans l'industrie. Côté pile, il intervient au sein du HUB Institute pour mobiliser les entreprises industrielles, notamment sur la thématique du Edge. Le lien entre les deux ? Someflu compte prochainement équiper ses pompes de traitement embarqué pour mieux monitorer les data d'usage et améliorer le coût total de possession... Energie, coût opérationnel, mise en route, fin de vie du produit, interopérabilité : retour sur les défis du Edge !

Concrètement, que peut apporter le Edge Computing par rapport à un traitement Cloud ?

Probablement d'abord le « temps réel » et continu pour extraire de la valeur depuis les données... Par exemple, si vous souhaitez un système d'alerte en continu sur vos équipements, le Edge Computing pourra vous permettre de réaliser des analyses sur un échantillonnage de données plus important que les objets connectés au Cloud, qui fonctionnent selon un principe d'analyse séquentielle et qui sont - dans le cas de capteurs dit « IoT » - dépendants de la batterie. De plus, le Edge évite de transférer des masses de données à haute fréquence dans un cloud pour demain renvoyer l'information sur la prise de décision : c'est plus économique, écologique, résilient. Et enfin, il y a la question de la propriété industrielle : en gardant le traitement en local, les données ne risquent pas d'être captées dans un environnement ouvert comme le Cloud.

Mais le Edge n'est-il pas coûteux à maintenir ?

Disons que c'est énergivore en raison de la puissance de calcul à apporter. Cette question agite d'ailleurs beaucoup la sphère des fournisseurs : de nombreux travaux sont menés pour augmenter la puissance des puces, sans répercussion sur le coût en énergie. Récemment, le constructeur SOITEC a ainsi annoncé la préparation de puces de plusieurs téraflops pour une consommation énergétique ultra-réduite.

Cette limitation du champ des ressources, n'est-ce pas ce qui rend le Edge plus complexe ?

Effectivement, dans cette optique, le Cloud est plus « simple » pour faire de l'IA : le Edge exige un vrai effort en matière d'électronique, en plus de tout le travail sur l'algorithme et l'entraînement des données. C'est d'ailleurs pour cette raison que l'écosystème Edge commence à converger

(protocoles IoT et OT, adoption d'architecture de référence au sein du consortium OpenFog) : il faut passer à une échelle industrielle et construire ensemble cet environnement tech et métier !

À quel horizon ?

On peut aller vite : deux-trois ans peut-être pour différentes offres industrialisées. Les entreprises sont d'ailleurs à la croisée des chemins quand il s'agit de leur transformation : certaines préfèrent aller vite et se tournent vers le Cloud, d'autres préfèrent temporiser et favoriser la maîtrise de leur donnée en optant pour le Edge.

L'interopérabilité, c'est pour vous le plus gros défi de l'usine 4.0 ?

Tout à fait ! Une machine ou un équipement doit être interopérable dans son environnement, dans une usine connectée. Pendant longtemps, les usines ont fonctionné sur la base de systèmes automatisés (« OT ») qui reposaient sur des réseaux et protocoles dédiés. Un important travail de standardisation avait alors été fait... et puis sont venus les objets connectés avec leur connectivité internet : tout à coup, c'est le monde de l'IT qui s'est immiscé dans l'usine. Et puis maintenant, il y a le Edge Computing qui propose de stocker et traiter directement sur l'objet... Il faut faire communiquer tous ces protocoles entre eux !

Quelles sont les solutions envisageables ?

Les professionnels du secteur commencent à se mobiliser pour travailler sur des standards communs : un premier pas a été franchi avec la création du standard OPC UA issu de la discussion de près de 650 entreprises du monde de l'OT. C'est une belle évolution mais on peut aller plus loin en utilisant notamment des middlewares (API).

C'est-à-dire ?

Les API sont des atouts capitalistiques importants pour les entreprises : c'est un peu le coffre-fort numérique de chacune. Dans le monde de l'IT, il est devenu usuel de les rendre accessibles pour fluidifier les échanges et s'ouvrir à d'autres marchés. Mais dans le monde industriel qui fonctionne en circuit fermé, ce n'est pas dans la culture... ce qui complexifie en amont le choix d'une solution plutôt qu'une autre. Je crois que les prochaines années vont bouleverser tout cela et forcer les industriels à réfléchir à l'ouverture de leurs API : tout simplement parce que le risque de se fermer des opportunités de marché face aux concurrents deviendra trop fort. Néanmoins, je n'ignore pas l'enjeu primordial que représente la sécurité pour les industriels...

La France est-elle compétitive en matière de Edge Computing ?

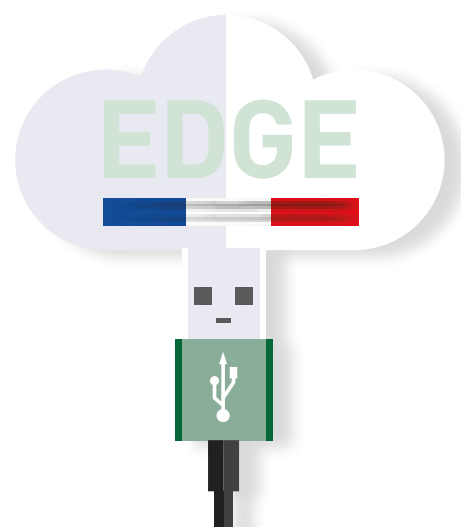
Le Cloud a été gagné par les GAFAs, mais il y a de belles opportunités pour la France et l'Europe sur le Edge. Tout va beaucoup reposer sur les leaders industriels comme Schneider, Siemens, STMicroelectronics qui tirent la transformation de leur secteur et ont été pionniers en matière d'automatisation : ce sont un peu les GAFAs de l'OT ! Aujourd'hui, elles sont à la manœuvre sur le Edge Computing et leurs évolutions ainsi que ceux de l'ensemble de la chaîne de valeur apportent de belles opportunités pour les uns mais aussi des risques pour les autres.

Va-t-on vers des architectures 100% Edge, 100% Cloud ou un mix entre les deux ?

Ce sont les besoins métiers qui détermineront le choix de telle ou telle technologie et les architectures à construire. On n'a pas forcément besoin de Cloud ou de Edge sur toutes les applications donc on peut faire tourner les deux technologies en parallèle. Ou favoriser l'une par rapport à l'autre. C'est un curseur à trouver en termes de coûts, d'enjeux de stockage et de distribution de données, de sécurité, de disponibilité... A ce titre, l'apparition de la 5G et la réduction de la latence associée va probablement encore renforcer le débat...

Dernière question : le Fog Computing est-il un bon compromis pour se lancer ?

Effectivement, en termes de mutualisation des équipements et de puissance de calcul, c'est une bonne architecture intermédiaire à mettre en place, certainement plus accessible. Le Edge Computing nécessitant encore quelques évolutions technologiques, proches du hardware, cela peut constituer une première approche...



MARCHÉS

Côté pile, il y a une demande qui ne cesse de s'amplifier pour les technologies Big Data / IA et des besoins qui semblent évoluer de l'infiniment grand (Cloud, IaaS, data lakes) à l'infiniment précis (analytics appliqué, logiciels SaaS, etc).

Et côté face, il y a un marché de fournisseurs qui poursuit sa transformation vers une offre de plus en plus concentrée, applicative et user-centric. Avec un biais renforcé vers les plateformes d'analytics ...

« Pendant de nombreuses années, nous étions dans une phase de découverte où il fallait identifier les cas d'usages qui justifiaient la mise en place de technologies Big Data et IA. Désormais, on ne cherche plus à prouver que ceux-ci fonctionnent : on veut qu'ils fonctionnent, et très vite » explique Emmanuel Lartigau, Senior Consultant chez CXP-teknowlogy Group.

Alors en 2019, le Big Data poursuit sa mue vers davantage d'efficacité et laisse au marché français un territoire d'expression encore à conquérir, entre soutien affiché du gouvernement, excellence algorithmique et... manque de données.



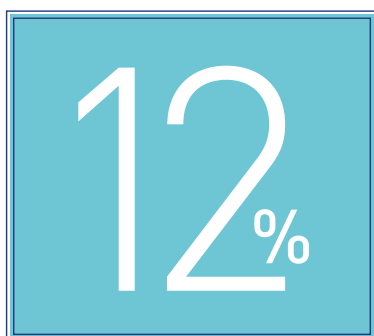
BIG DATA : UNE ANNÉE DANS LE RÉTRO



UNE ANNÉE DANS LE RÉTRO,
POUR REVOIR 2019 EN **CHIFFRES**,
EN **IMAGES...** ET EN **VOYAGE**

L'INFOGRAPHIE UTILISATEURS

Un besoin end-user toujours important



Le Chiffre 2019

C'est le taux de croissance anticipée du marché Big Data et Analytique en 2019¹

NB : Ce chiffre devrait se confirmer sur les années suivantes, l'institut IDC prévoyant un taux de croissance annuel de 13,2% tout au long de la période 2018-2022.



Le Revenu 2019

C'est le total estimé en valeur des investissements mondiaux dans les domaines du big data et de l'analytique en 2019²

NB : En 2022, toujours selon IDC, ce chiffre devrait atteindre 274 milliards de dollars... soit 85 milliards de plus en trois ans !



La Valeur montante 2019-2022

C'est le taux de croissance annuel estimé des dépenses consacrées par les entreprises aux Systèmes Cognitifs et d'Intelligence Artificielle sur la période 2017-2022

NB : En 2022, selon IDC, les dépenses consacrées à l'IA devraient atteindre 78 milliards de dollars, soit un triplement par rapport à 2018 !

Le Sondage 2019


Interrogées en décembre 2018 par le média MagIT/TechTarget, 150 entreprises ETI et Grands Comptes ont bien voulu dévoiler leurs besoins d'investissement pour l'année 2019³.

Leurs réponses confirment la prééminence forte des infrastructures Big Data et de l'orientation data-driven de la plupart des transformations engagées.

Extrait de l'enquête
Priorités IT, TechTarget

ENSEIGNEMENT

1




La data occupe une place de choix dans les **priorités d'implémentations**

- *Lesquelles de ces initiatives seront particulièrement implémentées par votre entreprise en 2019 ?*
 1. BI/Analytics : cité par 35% de répondants
 2. Business Applications : 33%
 3. IT automation : 31%
 4. Intelligence artificielle/Machine learning : n°4 ex, 30%
 5. Plateforme Big Data/data warehouse/data lake : n°4 ex, 30%

ENSEIGNEMENT

2

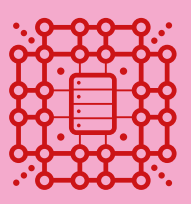


Les solutions data (et Big Data) au **top des applications SaaS**

- *Lesquelles de ces applications envisagez-vous de déployer en mode SaaS en 2019 ?*
 1. CRM/Sales Management : 35%
 2. BI/Analytics : 34%
 3. Plateforme Big Data/data warehouse/data lake : 25%
 4. Plateforme Machine Learning: 23%

ENSEIGNEMENT

3




L'**urgence** côté data : mettre en place l'**infrastructure**

- *Lesquels de ces outils consacrés aux data seront déployés en 2019 ?*
 1. Plateforme Big Data/data warehouse/data lake : 48%
 2. Data governance: 44%
 3. Data quality: 42%
 4. Data visualization: 41%

ENSEIGNEMENT

4




Le **Big Data** s'envisage de plus en plus en format « **as a service** »

- *Quels services managés votre entreprises prévoit-elle d'implémenter en 2019 ?*
 1. Stratégie, architecture et planification IT : 31%
 2. **Big Data as a service** : 25%
 3. IT service management: 24%

ENSEIGNEMENT

5



La transformation **digitale**, un quasi **incontournable** des entreprises

- *Engagement dans la transformation digitale*
 1. Transformation digitale radicale: 46%
 2. Transformation digitale encore balbutiante: 37%
 3. **Transformation digitale non entamée**: 17%

VERS UN MARCHÉ DE PLUS EN PLUS CONCENTRÉ ?



Quand on l'interroge sur son approche du marché des fournisseurs BI/Big Data, Emmanuel Lartigue, Senior Consultant chez CXP-teknowlogy Group assume une approche résolument originale : « La plupart des experts segmentent le marché entre la partie plateforme et la partie applicative, ou bien entre les différents composants de la chaîne de valeur. De mon côté, je trouve plus pertinent d'envisager le marché Big Data sous deux angles : les besoins en Front Office et ceux en Back Office ».

Que suggère-t-il derrière ces termes ?

- Le FRONT OFFICE correspondrait à toutes les applications ayant trait aux clients des entreprises : Relation client, Avant-vente, Fraude clients, etc
- Le BACK OFFICE serait quant à lui concentré sur les applications liées aux process internes de l'entreprises et comment les améliorer (supply chain, détection de pannes, etc).

Avec cette nouvelle grille d'analyse, Emmanuel Lartigue décrypte le marché des fournisseurs et met en lumière les raisons qui ont poussé certains éditeurs à lancer des vagues de rachats et de concentrations.

⁴ cf Initiative TALAGRAN : Traitement Automatique du Lan-gage au service du GRAND Débat National. Outil d'analyse textuelle développé par la société Quam pour traiter les contributions ouvertes (verbatim) des citoyens lors du Grand Débat National.

⁵ Quand un paramètre est changé dans la base de données, le système capte l'événement

Côté Front Office

Développer des outils intégrés et utiliser l'IA

Le constat d'une diversité d'outils : « Le marketing digital à l'ère du Big Data a permis de faire éclore une grande diversité d'outils, qu'il s'agisse de solutions applicatives type Data Management Platforms (DMP) ou de plateformes technologiques comme les Operational Database, les bases de données de graphe ou encore les Cloud Database (comme Snowflake par exemple). »

Vers une intégration horizontale : « La difficulté principale pour ces solutions est désormais de les faire discuter entre elles afin que le CRM, le site internet ou encore la chaîne Youtube puissent bénéficier des mêmes outils, ce qui facilitera l'exploitation et permettra de faire apparaître de nouveaux insights ».

Et une intégration verticale : « Mais d'un autre côté, les entreprises souhaitent également bénéficier d'outils décisionnels directement intégrés dans les applications métiers, au plus près des opérations. Par exemple via des popups de graphes décisionnels qui apparaîtraient sur les données clients en temps réel. C'est ce que les anglo-saxons appellent l'Operational BI ».

Avec pour résultat : une vague de rachats. « Ce besoin d'intégration horizontale et verticale a mené les éditeurs à se pencher sur leurs offres existantes et voir comment les faire communiquer et les compléter. Cela s'est souvent traduit par une recherche de solutions externes, notamment sur l'Operational BI : Workday a racheté Adaptive Insight (1,5 Mds \$) et Salesforce a racheté Tableau (15,7 Mds \$) pour lancer son offre Salesforce Lightning Platform. Google s'est emparé de Looker (2,6 milliards \$) tandis que Birst était racheté par Infor, Actuate par OpenText ou encore Pentaho par Hitachi. Ce besoin d'intégration s'illustre également en interne chez Microsoft avec la Power Platform, réunissant Power BI, Power Apps et Flow ».

L'entrée dans l'IA pour simplifier l'intégration. « La difficulté de ces solutions intégrées réside dans la relative complexité qu'elles sous-tendent, surtout avec des données volumineuses : dès lors, permettre aux opérationnels du Front Office d'interagir avec les résultats devient un enjeu clé et c'est ici que les innovations apportées par l'IA entrent en piste : NLP, chatbots à travers des fonctionnalités comme Ask Data (Tableau)... Le souci du marché français est qu'il est très dépendant des solutions anglo-saxonnes comme Microsoft Language Understanding Intelligent Service (LUIS), IBM Watson Tone Analyzer ou encore Amazon Comprehend... ce qui signifie que le langage naturel par défaut de l'algorithme est l'anglais (le français n'émanant que d'une traduction). Il va falloir remédier à cela rapidement pour éviter de se faire distancer⁴ ! »

La concentration du marché est-elle terminée ? « Absolument pas ; le RGPD a reporté cette concentration (si non les GAFA auraient tout envahi !). Il reste encore de la place pour des solutions techniques favorisant l'intégration, notamment sur le marché des API. Le mot d'ordre : simplification ! »

Côté Back Office

Gagner la bataille du temps réel en accentuant l'approche Big Data

Un enjeu de temps réel : « Les entreprises le savent : le temps s'est désormais accéléré et les délais de plusieurs jours ne sont plus acceptés par le marché quand Amazon livre en moins d'une journée à l'autre bout du pays. »

Le Big Data comme seul recours : « La conséquence pour les solutions de Back Office sera le recours systématique à des outils Big Data axés sur le temps réel, c'est-à-dire des solutions de capture, stockage, consolidation et analytics encore plus rapides. Voilà pourquoi les outils les plus recherchés sont probablement ceux qui ont trait à l'évent processing, au data streaming ou encore au « change data capture⁵ : les acteurs clés du marché l'ont bien compris et s'équipent pour proposer une offre pertinente... ce qui explique le rachat d'Attunity par Qlik et celui de Mulesoft par Google. »

Une transformation dans le Cloud ? « En termes de stockage et de traitement, le besoin de temps réel s'appuie forcément sur des bases de données de type Big Data (utilisant Spark ou Sqoop). Pour l'instant, ce sont les approches data lake qui sont privilégiées, plutôt on-Premise, mais le Back Office pourrait tout à fait monter dans le Cloud, comme en témoigne l'utilisation de plus en plus massive de Cloud database... »

L'Analytique avancée, un marché ouvert : « Mais s'il est un phénomène qui a marqué l'année 2019, c'est probablement celui de l'Advanced Analytics développé par des acteurs comme Microstrategy, SAS Institut, Information Builders ou encore TIBCO Software – soit la création d'outils BI transversaux permettant d'analyser en temps réel une multitude de données, notamment d'événements, via des plateformes BI complètes. Par exemple : un goulot d'étranglement dans une chaîne de livraison. C'est ici que le marché se montre probablement le plus innovant et conduit à une diversification des acteurs : on le voit avec des plateformes comme Dataiku, Report One, Toucan Toco, Dig Dash, Coheris, ou encore le concept d'hyperintelligence développé par Microstrategy. Cela promet une vraie évolution dans la manière dont l'information est rendue opérationnelle auprès des équipes métiers du Back Office : sur ce segment, je crois que le marché a encore beaucoup de choses à raconter... »

La difficulté de ces solutions intégrées réside dans la relative complexité qu'elles sous-tendent



ILS ONT DIT EN 2019



Elias Baltassis

■ ■ Il existe un risque pour l'Europe si elle n'investit pas, ce n'est rien de moins que sortir de l'Histoire. ■ ■

Cédric O, Ministre de l'Economie Numérique, en ouverture du congrès AI PARIS

■ ■ Nous réunissons le numéro 1 mondial du CRM avec le numéro 1 des plateformes analytiques ■ ■

Marc Benioff, CEO de salesforce.com, à propos de la fusion avec Tableau Software

■ ■ En France on a un problème en analytics et en qualité de données - probablement les deux problèmes seraient liés. Et c'est assez grave parce que ça, ce sont les deux drivers du futur. ■ ■

Elias Baltassis, Directeur Data et Analytics au BCG, présentant son étude du marché Big Data français lors de BIG DATA PARIS

■ ■ Nos résultats du premier trimestre ont été non seulement décevants, mais également inattendus. ■ ■

John Schroeder, CEO et fondateur de MapR Technologies lors de l'annonce de la procédure de liquidation, le 3 juin 2019

■ ■ Nous sommes ravis de commencer l'année 2019 avec un chiffre d'affaires record de 57,8 millions de dollars au premier trimestre ■ ■

Mike Tuchen, CEO de Talend, le 9 mai 2019, lors de la présentation des résultats Q1



Luc Julia



Mounir Mahjoubi



Cédric O

■ ■ Entre le laisser-faire et l'interventionnisme maximum, il existe une troisième voie à l'européenne et à la française. La France, l'Europe, toutes les démocraties doivent aujourd'hui reprendre en main leur avenir numérique. ■ ■

Mounir Mahjoubi, ancien Ministre de l'Economie Numérique, lors du congrès BIG DATA PARIS

■ ■ Les excès dans la manipulation des peurs ou des espoirs associés à l'IA peuvent faire naître le risque d'un nouvel hiver de l'IA. ■ ■

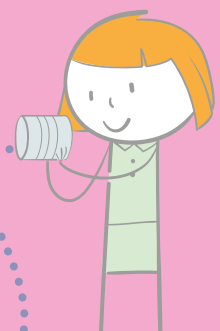
Luc Julia, VP Innovation de Samsung Electronics, lors du congrès AI PARIS

■ ■ Un nouveau chapitre s'ouvre. ■ ■

Tom Reilly, CEO de Cloudera, en octobre 2018 lors de la fusion avec Hortonworks

■ ■ J'ai déterminé que le moment était venu pour moi de prendre ma retraite et de faire la transition dans la direction de la société, alors qu'elle entre dans le prochain chapitre de sa croissance ■ ■

Tom Reilly, ex-CEO de Cloudera, le 5 juin 2019, à l'annonce de son départ



LA FEUILLE DE ROUTE

Financer le Big Data

Le Plan Deeptech de Bpifrance

Contexte du Plan : Avec des investissements multipliés par trois entre 2013 et 2017, la France est placée à la 2e place européenne (derrière le Royaume-Uni) en matière d'investissements deeptech.

Pilote du programme : Bpifrance

Etape 1 : Le 30 janvier 2019, présentation devant les ministres F. Vidal et A. Pannier-Runacher

Scope du programme : Les DeepTech ou « innovations de rupture ».

- **Quels sont les domaines représentés ?** Intelligence artificielle et Big Data, nouveaux matériaux, informatique quantique, robotique, production et stockage de l'énergie, biotechnologies, nanotechnologies...

- **Quels sont les critères pour identifier une innovation « de rupture » ?** 1/ Un fort lien avec la Recherche, 2/ Une capacité à lever des verrous technologiques, 3/ La création de barrières à l'entrée dans le secteur, 4/ Un «go-to-market» long et complexe, donc capitalistique.

Objectif du programme : Soutenir financièrement les initiatives liées aux DeepTech et créer des passerelles entre monde académique, monde entrepreneurial et investisseurs.

- **1 500 startup** devraient être soutenues à l'horizon 2023.

Montant total du programme : 1,35 milliards d'euros sur cinq ans

- **dont 800 millions d'euros d'aides nouvelles** (« Génération Deep Tech »). Issus notamment du Fond pour l'Innovation et l'Industrie (FII) et du Fonds FrenchTech Seed

- **dont 550 millions d'aides existantes** via les Sociétés d'Accélération du transfert des technologies (SATT) qui passent sous bannière Bpifrance et l'aide de l'ANR (Agence Nationale pour la Recherche).

Le Plan Intelligence Artificielle du Gouvernement

Contexte du Plan : Le 30 mars 2018, lors d'un discours au Collège de France, le président Emmanuel Macron déclare souhaiter investir 1,5 milliards d'euros sur 5 ans (dont 500 millions d'investissements privés) pour faire de la France un acteur majeur dans le domaine de l'IA.

Pilotes : Le Ministère de l'Enseignement supérieur et de la recherche, le Ministère de l'Economie numérique.

Etape 1 : Première réunion du comité de pilotage à Toulouse le 28 novembre 2018, en présence des ministres F. Vidal et M. Mahjoubi.

Scope du programme : Toutes les activités liées à la recherche fondamentale et la recherche appliquée en matière d'IA. Une attention prioritaire est portée sur les thématiques **Santé, Environnement, Biais**.

Objectifs du programme : Renforcer les moyens de la Recherche française en matière d'IA, attirer et retenir les talents, développer la coopération internationale... Au total six chantiers sont proposés :

- **La création de 4 Instituts Interdisciplinaires en IA (3IA)** pilotés par l'INRIA: les centres de Grenoble, Nice-Sophia Antipolis, Paris et Toulouse.

- **La création de 40 chaires** d'enseignement supplémentaires dès 2019, et la **formation de 300 nouveaux doctorants** par an d'ici 2022.

- **Le renforcement des projets de recherche en IA** de l'Agence Nationale de la Recherche.

- **L'installation d'un supercalculateur** de 16 pétaflops sur le site du CNRS (Saclay) accessible à toute la communauté scientifique.

- **Pour encourager les partenariats de recherche privés-publics**, le renforcement des moyens des Instituts de recherche technologique (IRT).

- **La création d'un programme bilatéral** avec l'Allemagne.

Montant total du programme : environ 1 milliard d'euro d'ici 2022 - dont 665 millions issu du budget de l'Etat



INTERVIEW

Quand on lui demande si Cap Digital est un incubateur ou un accélérateur, Simon Guichard préfère corriger : « Cap Digital c'est un pôle de compétitivité, c'est-à-dire qu'il a pour rôle d'animer l'écosystème français du digital : mettre en relation des entreprises, aider à trouver des financements européens ou des subventions, accompagner sur une levée de fonds, conseiller, promouvoir... Mais en tant qu'association loi 1901, il ne réalise aucun profit. Et il n'a pas vocation à héberger des start-up ou à les financer directement ».

Présentation faite, **Simon Guichard** peut maintenant nous parler de son expérience de six mois au sein du pôle, « une mission qui m'a été confiée suite à un projet entrepreneurial que j'avais sur la Blockchain ». Passionné par les sujets Data et IA (une thématique transversale au sein du pôle), il a pu ainsi accompagner un nombre important d'entreprises (près de 120)... et confirme que la marche est encore haute pour rivaliser avec les Etats-Unis ou la Chine.

Quelle proportion représentent les sujets data/IA au sein d'un pôle comme Cap Digital ?

C'est une grande majorité, environ 60 à 80% des entreprises du pôle travaillent sur ces sujets. Et les 2/3 des projets qui demandent des financements sont orientés IA et Data, avec une part de plus en plus importante de sous-marchés comme le Traitement Automatisé du Langage. Cette tendance ne devrait pas s'inverser dans les années qui viennent, bien au contraire : l'intérêt des venture capital (VCs) est quasiment exclusivement concentré sur ces secteurs...

|| L'intérêt des venture capital (VCs) est quasiment exclusivement concentré sur ces secteurs ||

Ce sont surtout des start-up ?

Il y a de tout : grands groupes, PME/ETI, start-up, avec un vrai enjeu de mise en relation pour permettre aux petits de répondre aux besoins des plus grands... Le pôle a ainsi accueilli plusieurs entreprises devenues célèbres comme Snips, Dataiku, ou encore OVH : certaines sont encore là, d'autres sont parties à l'étranger. Mais ce qui est intéressant, c'est que l'on observe de plus en plus de start-up adopter la stratégie de niche...

C'est-à-dire ?

Ce sont des start-up qui se déploient d'abord sur des cas d'usage précis pour bien valider leur marché, puis qui descendent sur d'autres verticales ou répliquent leur modèle sur d'autres marchés. Par exemple, des start-up qui se concentrent sur la détection de fraude appliquée à un cas d'usage et qui le transposent ensuite dans un autre contexte. C'est un phénomène que l'on observe depuis deux-trois ans.

Les entreprises que vous accompagnez sur ces sujets trouvent-elles des financements ?

Sur la partie R&D, on a un taux d'acceptation vraiment important : 52% des projets labellisés Cap Digital passent ensuite les étapes de sélection et de financement... c'est plus qu'encourageant ! Pour ce qui est

des levées de fonds sur un délai 12-18 mois, on est également sur un taux de 50%. Mais il ne faut pas se voiler la face : même si les entreprises françaises ont le talent nécessaire et proposent des algorithmes vraiment compétitifs, cela leur est très difficile de passer la série C auprès d'un VC. On a encore un fossé en termes de moyens avec les entreprises américaines et chinoises.

Et au niveau européen ?

A cet échelon-là, c'est davantage une question de leadership qui se pose : les entreprises françaises sont très sollicitées pour participer à des consortiums, mais elles en sont rarement le pilote. Et la lourdeur administrative constitue un frein réel pour des petites structures qui n'ont pas forcément de temps à y consacrer. On remarque qu'à ce jeu-là, les italiens et les espagnols s'en sortent bien mieux que les consortiums 100% français...

Que faut-il faire pour développer davantage l'écosystème IA français ?

On aimerait que les grands groupes investissent davantage dans les start-up, mais en format « prestation de services », et non via un rachat. Trop souvent, les grands groupes y voient l'opportunité d'internaliser des technologies et de s'offrir un avantage compétitif sur le marché ; sauf qu'ils cassent la logique d'innovation en plein élan... Et puis il y a la question des bases de données : les start-up françaises ont des algorithmes vraiment puissants mais pas de bases de

52%

des appels à projets
R&D IA/Data
obtiennent des
financements

données pour entraîner leurs codes. C'est dommage de se dire qu'il leur faut puiser dans les datacenters des entreprises américaines pour accéder à du Big Data.

Justement, en parlant de Big Data, quel serait pour vous l'enjeu numéro dans les années à venir ?

J'ai envie de parler d'environnement parce qu'on omet souvent cet aspect dans les discussions autour de la data. Concrètement, si on considère qu'il faut 10 kilowatts pour qu'une IA batte un humain au jeu de go, je vous laisse imaginer l'énergie qu'il faudra générer pour une utilisation généralisée et quotidienne des IA (et donc des datacenters derrière). Il va falloir qu'on se penche sérieusement sur la question de la sur-utilisation des IA : c'est un enjeu à la fois éthique et environnemental.

|| Les start-up françaises ont des algorithmes vraiment puissants mais pas de bases de données pour entraîner leurs codes ||



Simon
Guichard

**Référent Data et IA
pour CAP DIGITAL**



EN INDE, UN MARCHÉ ENCORE BALBUTIAN

« Le Big Data en Inde, ce n'est pas encore tout à fait un marché de end-user, mais on commence à observer des taux d'adoption encourageants ». Avec CORP, Harrish Murugesan s'est établi à Bangalore en 2017 pour créer l'antenne locale de l'agence événementielle, et ses premières recherches sur les marchés porteurs l'ont naturellement mené vers le secteur du Big Data et de l'IA : « C'est un marché qui a le vent en poupe avec un taux de croissance impressionnant : 33,5%⁶ ! Mais on est encore beaucoup sur un marché de fournisseurs, avec des géants de l'IT qui s'efforcent de créer des solutions appliquées pour les revendre ensuite à l'étranger »...

... L'IT en héritage

Ces géants de l'IT, tout le monde en Inde les connaît : Tata Consultancy Services (TCS), Wipro, Infosys, TechMahindra, HCL Technology. Ils ont prospéré dans les années 2000 sur un schéma d'organisation mondialisée qui est devenu le paradigme de l'IT indienne : l'outsourcing. Ou en termes plus techniques : le BPO (Business Process Outsourcing) et le KPO (Knowledge Process Outsourcing). Soit une offre de services supports, « back end », destinés aux multinationales américaines ou européennes : « Aujourd'hui, ces géants packagent des offres Big Data et Analytics couplées avec leurs services BPO et KPO, ce qui peut leur permettre de dégager encore davantage de ROI », explique Harrish Murugesan.

Mais ils ne sont pas les seuls à se lancer sur le marché prometteur du Big Data : quelques 700 entreprises, pour la plupart pure players analytics, s'engagent à leur tour dans la mêlée. Et parmi elles : plusieurs centaines de start-up. « L'économie indienne s'appuie de plus en plus sur les start-up, qui sont extrêmement valorisées au niveau post-graduate : on parle du second écosystème de start-up au monde, avec pas moins de douze licornes recensées et plus de 50 millions de dollars engagés par les venture capitals ». Même s'il n'a pas encore réussi à décrocher sa propre licorne en Inde, le Big Data fait également partie de la dynamique, avec un nombre multiplié par quatre entre 2013 et 2017, fruit des nombreux programmes de soutien lancés par les gouvernements (dont StartUp India, lancé en 2016 pour faciliter l'accès aux investissements bancaires).

Une fabrique d'ingénieurs

Comme pour les géants de l'informatique, la plupart de ces start-up viennent puiser dans le vivier des IIT (comprendre : Indian Institute of Technology, l'équivalent des écoles d'ingénieurs en Inde) pour se fournir en main d'œuvre qualifiée. « Le parcours universitaire traditionnel d'un étudiant indien dure environ quatre ans et, à la sortie, les stages de fin d'études au sein des grandes entreprises de l'IT vont souvent

|| C'est un marché qui a le vent en poupe avec un taux de croissance impressionnant :

33,5% ||

Infosys®
Navigate your next

TATA
TATA CONSULTANCY SERVICES

wipro

Quelques grands noms de l'IT Made in India

amener les jeunes diplômés à intégrer des fonctions de software architecture ou data analytics. Parfois, certains choisissent d'aller compléter leur expérience à l'étranger (aux USA ou au Canada) avant de revenir : c'est véritablement après ces quelques années qu'ils intègrent des fonctions de data science appliquées ». Même si les IIT indiens sont très bien pourvus, la NASSCOM⁷ (puissante fédération des industries de l'IT en Inde) ne cesse d'alerter sur le déséquilibre offre-demande en matière de talents : selon elle, les 78 000 étudiants formés en 2018 n'ont pas réussi à combler les attentes des quelques 140 000 offres d'emploi proposés dans le secteur. Et ce phénomène devrait s'accroître en 2021, avec un « gap » estimé à 140 000 (soit 90 000 étudiants formés pour 230 000 offres d'emploi estimées).

Au Sud : la Big Data Belt

Raison pour laquelle la plupart des entreprises posent leurs bagages près des universités, le plus souvent au Sud de l'Inde. Des hubs comme Bangalore, Hyderabad ou Chennai attirent les sièges et filiales des entreprises IT face aux traditionnelles Bombay et Delhi, davantage tournées vers l'appliquatif (finance, telecom, retail). Selon une étude de 2013, si Bangalore reste le pool le plus connu, Hyderabad serait celui qui aurait vécu la plus forte croissance lors des dernières années : depuis 2011 la ville accueille notamment l'INSOFE (International School of Engineering) qui délivre des formations universitaires reconnues en matière de Data science et de Big Data analytics (le magazine CIO.com a ainsi classé le diplôme de l'ISOFE parmi les 16 formations Big Data qui comptent le plus entre 2013 et 2016).



Sur le campus tech et moderne de l'IIT d'Hyderabad

|| On parle du second écosystème de start-up au monde ||



Le gouvernement local de l'Etat d'Andhra Pradesh (où se trouve Hyderabad) est d'ailleurs très offensif pour attirer les entreprises et développer des programmes d'incitation : la création de clusters industriels fait ainsi partie de son agenda économique et l'annonce en 2018 de l'ouverture de 4 nouveaux clusters IT semble conforter ces ambitions. « En Inde, ce sont les gouvernements locaux qui sont en première ligne sur les politiques sectorielles : dans le Big Data, on pourra retrouver beaucoup d'incitations du côté des Etats de l'Andhra Pradesh, du Maharastra ou encore du Karnataka », explique Harrish Murugesan.

L'Etat indien à la manœuvre

Mais l'Etat central n'est pas en reste et affiche un soutien appuyé au secteur technologique : lancé par le Premier Ministre Nahendra Modri en 2015, le programme Digital India vise ainsi à généraliser l'emploi des technologies au cœur de l'appareil d'Etat. En clair : dématérialiser de nombreuses procédures administratives en attribuant un identifiant unique (le « aadhaar ») à chaque citoyen indien (soit 1,3 milliards de aadhaar attribués). Une initiative qui cherche notamment à désenclaver certains territoires isolés, mais qui reste parfois décriée pour son manque de sécurité liée aux données. L'Inde ne bénéficie certes pas (encore ?) de politique de protection de données personnelles « à la RGPD », mais les signaux semblent commencer à converger : en 2011, l'Information Technology Act, pierre angulaire de la régulation des données

numériques, intègre un amendement de dédommagement des citoyens en cas de fuite de données.



Le président indien Nahendra Modri

« Nous avons besoin de faire de l'intelligence artificielle en Inde et aussi de la rendre utile pour l'Inde » : cette déclaration d'intention de Nahendra Modri en février 2018 lors du lancement d'une fondation « IA for good » par deux milliardaires indiens Romesh et Sunil Wadhvani témoigne bien de l'ambivalence indienne. D'une part, un besoin immense pour une population de 1,3 milliards d'habitants (dont 500 millions d'internautes quotidiens) et d'autre part, des technologies et des capitaux encore concentrés par des élites. Des initiatives dans le domaine de la santé (identification des pôles de tuberculose, détection des bébés en sous-poids) ou de l'agriculture (détection des risques de sécheresse ou de parasitisme) commencent à voir le jour, sous l'impulsion de fonds privés ou de l'organisme central de transformation, le Niti Aayog.

Le momentum est proche côté end-user

Reste que le Big Data est encore une affaire de capitaux privés et que sa progression au sein des entreprises reste moins rapide que dans les autres pays : sur le seul secteur du Cloud, les dépenses incriminées ne représenteraient que 6% du total des dépenses informatiques, là où la moyenne mondiale s'établirait plutôt autour de 8%, voire 12% dans les pays occidentaux⁶. Cependant, sur cette question, les chiffres de croissance semblent converger vers un

proche « momentum », avec 30% de croissance annoncée par NASSCOM pour les prochaines années.

« Les organisations les plus matures commencent à envisager l'IA/ML non plus comme une spécialisation mais véritablement comme un moyen de résoudre des problèmes », déclarait récemment Dakshinamurthy V Kolluru, le fondateur et président de l'INSOFE. Une affirmation renforcée côté analytics par l'étude que son université a publiée conjointement avec le magazine Analytics India en novembre 2018, et qui établit le revenu du secteur de l'analytics, du Big Data et de la data science à 2,7 milliards de dollars (il représentait environ 1 milliard en 2013 selon NASSCOM). « L'adoption est en train de progresser, confirme Harrish Murugesan, mais elle est très variable en fonction des secteurs : finance, telecom et e-commerce sont bien sûr en pointe. Mais l'industrie, l'énergie, les infrastructures ont encore du mal à se convertir ». A ce titre, les investissements générés par le programme Make in India⁹ pourraient bien accélérer la transformation du secteur industriel...

« Le business du Big Data est encore en phase de construction, conclut Harrish Murugesan. Mais l'Inde connaît une phase d'accélération globale qui se traduit dans les indicateurs : on parle du 3e marché mondial en 2025 et d'un taux de croissance de 7% ». Alors, investir ou ne pas investir dans le Big Data en Inde ? En 2019, le classement Doing business a placé le pays à la 77e place sur 190, lui faisant ainsi gagner... 23 rangs. Peut mieux faire, certes ; mais la trajectoire semble toutefois prometteuse...

POUR EN SAVOIR PLUS :

- La presse spécialisée : Siliconindia, Analytics India Magazine
- Les événements : AIR Summit, Bangalore / Data Science congress, Bombay
- Le site de référence : www.nasscom.in (site de la NASSCOM)

⁶ Selon l'étude "State of Analytics At Domestic Firms In India 2018" réalisée par Analytics India Magazine et INSOFE

⁷ National Association of Software and Services Companies

⁸ Etude NASSCOM

⁹ Programme gouvernemental lancé en 2014 pour encourager les entreprises mondiales à installer leurs usines de production en Inde.

Conférence & Exposition

9 & 10 mars 2020

Palais des Congrès • Paris

BIG DATA corp

PARIS

18 000 + Participants
300 + Partenaires
100 + Intervenants

This is
BIG

Innovations radicales, applications inédites,
créativité débridée... Plus que jamais l'édition 2020
de Big Data Paris s'annonce **BIG** !

Rejoignez l'événement Big Data de la scène européenne sur

www.bigdataparis.com

RÉGLEMENT

On prédisait le Big Bang et finalement...

Finalement le passage des entreprises en format RGPD se sera fait plutôt en douceur. Du moins, par paliers progressifs. Car le RGPD ne s'est pas imposé sur un terrain vierge : dans les grands groupes, les réglementations Informatique et Libertés étaient déjà à la manœuvre pour encadrer le traitement des données personnelles et pousser les entreprises à rester vigilantes. Et si la marche à gravir a semblé plus haute pour les petites entreprises et les start-up, notamment dans le digital, le challenge de la protection des données a surtout été une occasion de se pencher une bonne fois pour toutes sur le sujet clé de la gestion des données : démarches qualité, MDM, cartographies... Avec le RGPD, l'entreprise entre dans une nouvelle phase de sa transformation data.

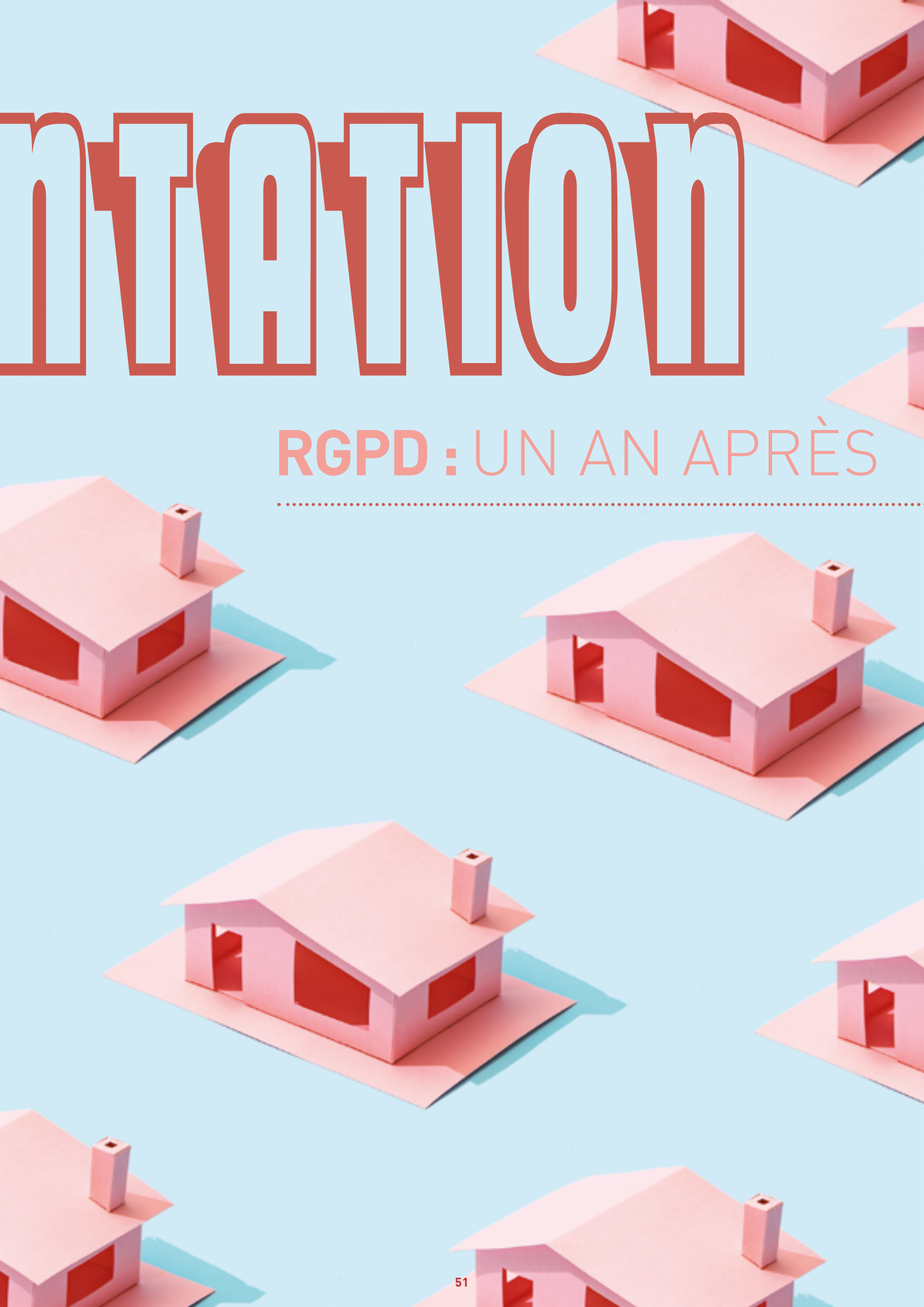
Alors outre le fait de ne pas se faire sanctionner (ce qui peut lui coûter cher... cf encadré), que peut-elle espérer en tirer ? Le ROI et la valeur du RGPD seront probablement des mesures difficiles à établir (et des enjeux politique pour les prochaines années, à l'heure où certains dénoncent la perte de compétitivité de l'Europe sur cette question), mais son coût, lui, commence à être connu : selon Wavestone, 4,5 millions d'euros en moyenne auraient été investis dans les grandes organisations sur la question de la vie privée numérique, dont une grande partie pour la conformité RGPD.

Pour avoir le son de cloche du « gendarme », nous sommes allés voir du côté de la CNIL, et des chiffres qu'elle a publiés elle aussi, « un an après ».

Petite revue de data en simplifié. Pour prendre le pouls d'une société post-RGPD.

INTRODUCTION

RGPD : UN AN APRÈS



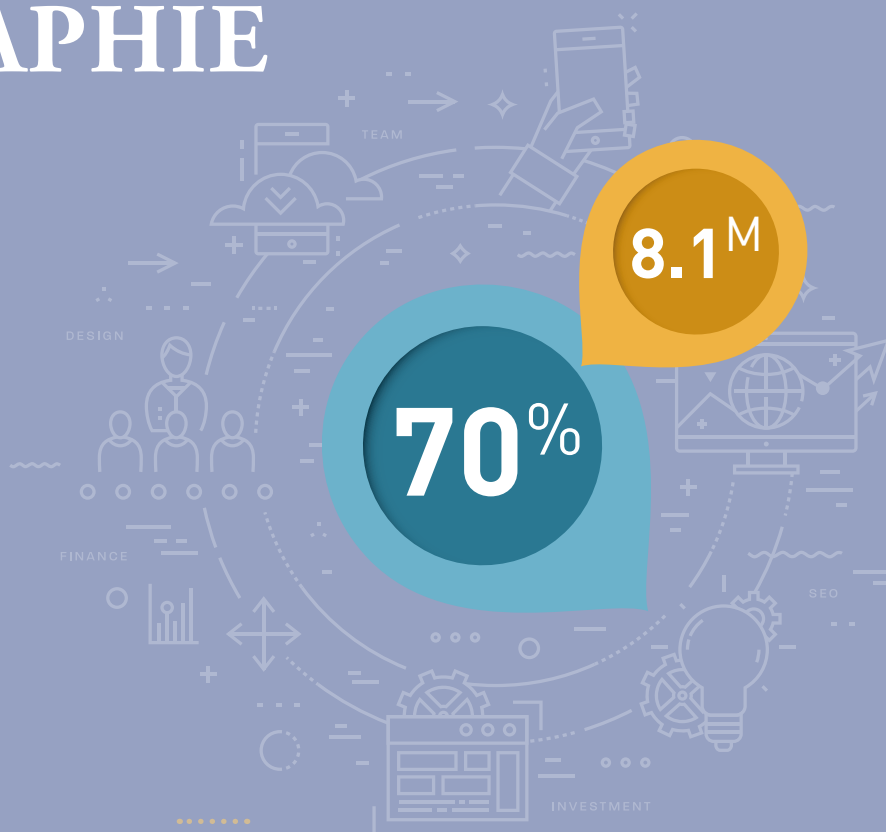
L'INFOGRAPHIE

Un an de RGPD

Des citoyens plus concernés...

70% des français se disent aujourd'hui plus sensibles aux problématiques de protection des données¹

8,1 million de visites en cumulé sur le site de la CNIL²



... mais aussi plus inquiets³ :

32% des citoyens déclarent avoir moins confiance dans les entreprises pour le traitement de leurs données par rapport à l'année précédente.



32%



25%

25% d'entre eux assurent même avoir cessé d'utiliser certains services pour mieux protéger leur vie privée.

D'où une recrudescence des procédures engagées⁴

Près de 12 000 plaintes en France et près de 145 000 au niveau européen. Ces plaintes portent sur :



La diffusion de données sur internet avec demandes de déréférencement

Le secteur marketing/commerce

Les ressources humaines

La banque et le crédit

Le secteur santé/social

¹ Selon un sondage IFOP réalisé en avril 2019 sur 1000 personnes

² Chiffres CNIL de mai 2018 à mai 2019

³ Selon Wavestone

⁴ Chiffres CNIL de mai 2018 à mai 2019

Et des sanctions qui ne tardent pas ⁵



+2000 notifications de violations de données en France (près de 90 000 au niveau européen). La majorité due à des atteintes à la confidentialité des données

310 contrôles dont :

- 204 contrôles sur place (dont 20 portant sur des dispositifs vidéo)
- 51 contrôles en ligne

49 mises en demeure

Deux secteurs plus particulièrement concernés :

- Assurance : 5 décisions
- Ciblage publicitaire technologique (SDK) sur mobile : 4 décisions

11 sanctions

- 10 sanctions pécuniaires
- 1 avertissement
- 1 non-lieu

Au total : **61 millions** d'euros de sanctions en Europe⁶

61^M

Le RGPD, une opportunité pour les entreprises ?

Les entreprises qui investissent dans la protection des données pourraient espérer 10% de revenus additionnels⁷.

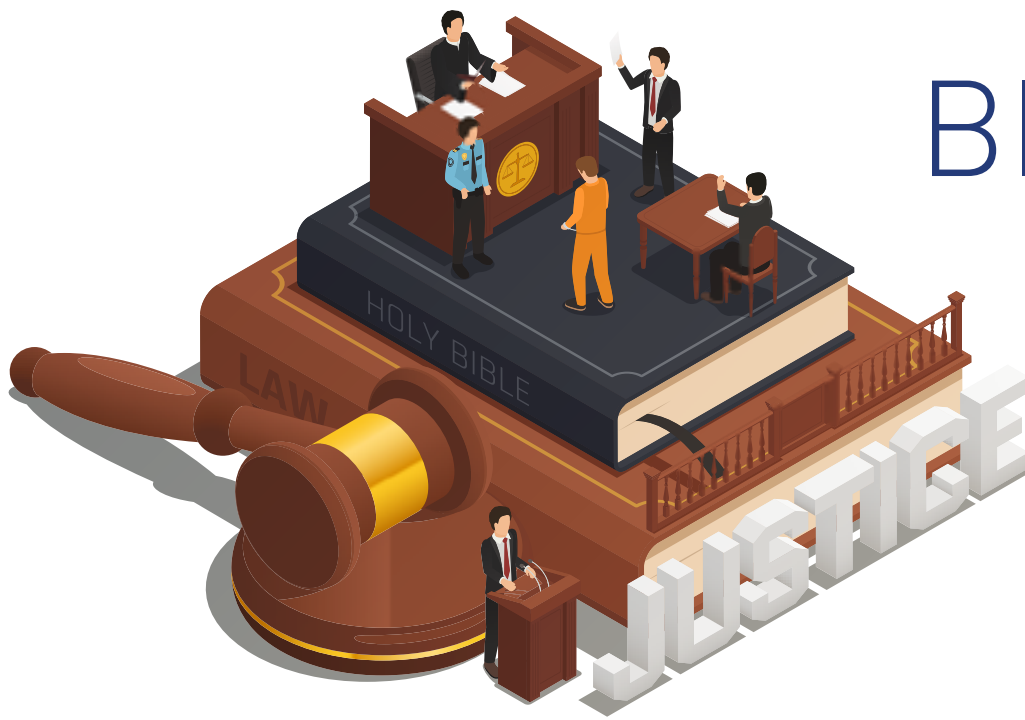
10%



⁵ Chiffres CNIL de mai 2018 à mai 2019

⁶ Selon TNP Consultants

⁷ Selon Gartner



BIG DATA STORY

Tout commence quand...

... l'association None Of Your Business (« NOYB ») et l'association La Quadrature du Net (« LQDN ») déposent des plaintes auprès de la CNIL les 25 et 28 mai 2018... soit en pleine entrée en vigueur du RGPD.

Ce qui est reproché

D'après ces associations, Google traiterait les données personnelles de ses utilisateurs à des fins de personnalisation publicitaire, sans avoir la base juridique suffisante pour le faire.

La procédure judiciaire

La première question cruciale à régler est la détermination de l'autorité compétente pour instruire la plainte, le siège de Google étant situé en Irlande. Après trois mois de discussions entre juin et septembre, c'est finalement la CNIL française qui obtient le traitement du dossier, en coopération avec les CNIL européennes.

Le contrôle

En septembre 2018, la CNIL effectue un contrôle en ligne sur les procédures de Google : elle analyse ainsi le parcours d'un utilisateur qui créerait un compte Google lors de la configuration de son portable sous Android.

Ce que la CNIL découvre

1. Les informations concernant les traitements et l'utilisation des données personnelles sont **difficilement accessibles** : il faut effectuer plusieurs actions, cliquer sur plusieurs boutons ou pages pour y accéder.
2. Ces informations sont **peu claires et peu compréhensibles** au regard de l'ampleur des traitements mis en place par la société et au regard de l'intrusion représentée.



LA SANCTION GOOGLE

3. **Le consentement de l'utilisateur est souvent recueilli à son insu** : l'utilisateur doit faire la démarche de cliquer sur « plus d'options » pour accéder au paramétrage et, une fois cette démarche effectuée, l'affichage d'annonces personnalisées est pré-coché par défaut. De plus, lors de la création de son compte, il doit cocher les pavés « j'accepte les conditions d'utilisation de Google » et « j'accepte que mes informations soient utilisées (...) » pour pouvoir finaliser la procédure de création... ce qui le conduit à accepter en bloc toutes les personnalisations prévues par Google.

Ce que la CNIL conclut

Sur la base de ces découvertes, la CNIL conclut donc que le consentement de l'utilisateur ne peut être éclairé (manque d'informations) ni spécifique (consentement en bloc) ni univoque (paramétrage par défaut).

Le 21 janvier 2019, elle décide donc de condamner Google pour manque de transparence, information insatisfaisante et absence de consentement valable pour la personnalisation de la publicité.

La sanction

Google est condamnée à une amende de 50 millions d'euros et cette sanction est rendue publique.

La CNIL justifie l'importance du montant et la publicité donnée à l'amende par « la gravité des manquements constatés » et le fait que ceux-ci perdurent à ce jour. Elle déclare également que la situation de quasi-monopole de l'entreprise lui confère certains devoirs et certaines responsabilités auprès de la société qui justifient en partie l'ampleur de cette sanction.

Tout commence quand...

... en août 2018, un utilisateur du site internet de la société SERGIC formule une plainte auprès de la CNIL. Spécialisée dans la gestion immobilière (achat, location, vente, etc.), SERGIC propose une page sur son site www.sergic.com qui permet aux candidats à la location d'uploader directement leurs pièces justificatives pour constituer leur dossier.

Ce qui est reproché

L'utilisateur déclare que, depuis son espace personnel, il aurait pu accéder à des documents enregistrés par d'autres utilisateurs en modifiant légèrement l'URL affichée dans le navigateur.

Le contrôle

Le 7 septembre 2018, la CNIL effectue un contrôle en ligne en se rendant sur le site internet de SERGIC : elle vérifie alors si les documents mentionnés sont librement accessibles ou non.

Elle double la procédure par un contrôle sur place, dans les locaux de l'entreprise, quelques jours plus tard.

Ce que la CNIL découvre

1. Les documents uploadés sur le site – parmi lesquels : cartes d'identité, carte Vitale, avis d'imposition, RIB, relevés de compte, etc – sont effectivement **librement accessibles, sans authentification préalable**.

2. La société SERGIC avait connaissance de cette vulnérabilité sur son site depuis le mois de mars 2018. Elle avait entamé des développements informatiques pour le corriger mais **ce n'est que suite au contrôle en ligne de la CNIL que la correction totale est devenue effective** (le 17 septembre 2018).



3. De plus, la société conservait tous ces documents personnels **sans limitation de durée**, y compris pour les candidats n'ayant pas accédé à la location.

Ce que la CNIL conclut

Sur la base de ces découvertes, la CNIL conclut donc que la société SERGIC a manqué à deux obligations :

- l'obligation de préserver **la sécurité des données personnelles** des utilisateurs du site via une procédure d'authentification
- l'obligation de déterminer **une durée de conservation des données** personnelles en fonction d'une finalité de traitement (ex : la gestion de candidatures pour une location).

La sanction

La société SERGIC est condamnée à verser une amende de 400 000 euros et cette sanction est rendue publique.

La CNIL justifie ce montant par la gravité du manquement et de ses conséquences (les documents accessibles révélant des aspects intimes de la vie des personnes) et par le manque de diligence de la société dans la correction de la vulnérabilité.

La taille de la société et sa surface financière ont toutefois également été pris en compte pour établir le montant de l'amende.

The screenshot shows the SERGIC website interface. At the top, there is a navigation bar with links for 'Le groupe', 'Les agences', 'Nous rejoindre', 'Nous contacter', and 'ESPACE CLIENT'. Below this is the SERGIC logo and a menu with options: 'LOUER | ACHETER | VENDRE | GESTION | COPROPRIÉTÉ'. The main content area features a search bar with fields for 'Je veux' (set to 'Louer'), 'Quoi?' (empty), 'Où?' (set to 'Vie, CP'), and 'Budget max.' (set to '2000+€'). There are also buttons for 'Recherche par référence', 'Nombre de chambres', and 'Dans un rayon de...'. A 'RECHERCHER' button is on the right, next to a notification '742 biens disponibles'. Below the search bar, there is a banner for 'Retrouvez tous les avis des clients / sergic sur immoAdvisor' with a 'JE DÉCOUVRE LES AVIS' button. At the bottom, there is a footer with 'Les sites du groupe' and a list of URLs: 'sergic-residences.com | sergic-entreprises.com | sergic-vacances.com | syndic-one.com | viva-syndic.com | esis-syndic.com'. Social media icons for Facebook, LinkedIn, and a blog link are also present.

SOCIÉTÉ

Les chiffres parlent d'eux-mêmes : en 2019 à échelle mondiale, les femmes ne représenteraient pas plus de 20% des effectifs de programmation en IA. Et certains spécialistes situeraient même plutôt ce ratio autour de 10%.

En France, la Fondation Femmes@numérique place, elle, ce chiffre à 15% pour l'ensemble des fonctions techniques (programmation, production, développement). Bataille de chiffres donc, mais au bout, toujours le même constat : les femmes manquent à l'appel sur le terrain technologique. Et plus cruellement sur celui de l'IA.

On pourrait feindre d'ignorer le phénomène mais ce serait compter sans la dégradation rapide d'une situation que rien ne semble endiguer (de 30% de femmes ingénieures dans les années 80, elles sont passées à seulement 15%) ; et sans les ravages sociétaux que ce déséquilibre ne semble plus en mesure d'empêcher : biais des algorithmes, isolement des développeuses en poste, enracinement culturel, exclusion des futures générations... Le verdict semble sans appel pour un secteur qui s'inquiète déjà d'autres débordements de ses technologies sur le terrain de l'éthique...

|| De 30% de femmes ingénieures dans les années 80, elles sont passées à seulement 15% ||

L'IA, UNE AFFAIRE D'HOMMES ?



... L'éducation, pierre angulaire

Alors que peut-on faire ? Les associations françaises qui ont pris le sujet à bras le corps (Duchess France, Women in AI...) le clament haut et fort : l'éducation est la pierre angulaire d'une IA de long terme. Apprendre aux jeunes filles à coder, lever les stéréotypes du « geek » masculin et créer des « role models » féminins sont probablement les premiers moteurs du changement, ceux qui permettront de déconstruire le fameux modèle d'une informatique « pour les hommes et par les hommes » héritée des années 80. « Nous avons un programme éducatif Wai2GO qui vise à apprendre l'IA aux femmes de tous âges et de tous horizons professionnels, avec un focus particulier sur les jeunes filles ; et nous utilisons de la robotique dans des ateliers dédiés pour créer une expérience engageante, leur montrer qu'elles peuvent elles aussi manipuler des robots sur le terrain », explique Hanan Salam, co-fondatrice de Women in AI.



Mais ce n'est pas suffisant : si l'on regarde les chiffres actuels, les classes de Terminale S seraient occupées à 46% par des filles... soit un équilibre relatif. C'est pendant les études universitaires puis lors de l'entrée sur le marché du travail que la désertion s'accroîtrait. La faute à des environnements parfois trop masculins qui ne laissent pas suffisamment de terrain d'expression et d'affirmation aux dernières aspirantes. « En entretien avec des candidates, cela m'est arrivé d'avoir cette question : 'Y a-t-il d'autres femmes dans l'équipe ?' », raconte Milena Le Mançq, recruteuse chez Kapten. Selon elle, ce serait surtout sur les fonctions data engineer et data scientist que le fossé serait le plus important, davantage que sur les fonctions de data analyst.

Se réunir et s'encourager

D'où le besoin de se réunir via des associations comme Women in AI ou Duchess France : « On s'est rendu compte qu'il y avait beaucoup de femmes qui se sentaient

seules », témoigne Amira Lakhal, l'un des membres dirigeants de Duchess France. « On cherche à orienter nos ateliers sur le partage de connaissances parce que cela nous permet d'échanger sur les sujets de fond tout en créant un sentiment d'appartenance à un réseau ». Côté Women in AI, la valorisation passe par la création d'un trophée célébrant les fondatrices de sociétés d'IA en partenariat avec Cap Gemini, avec un total de trois prix remis. Valorisation, mentorat, networking, coaching en prise de parole... autant d'initiatives qui s'efforcent de pallier une situation endémique de déséquilibre : « Pour cette génération, difficile d'aller à l'encontre des chiffres, il faudra probablement attendre la prochaine génération pour tenter d'aboutir à une parité. Cependant, on peut encourager des femmes issues d'autres environnements professionnels à se former au code. Il n'est jamais trop tard ! » poursuit Amira Lakhal.

Pas de discrimination positive

En attendant d'accéder à ce pool de futurs talents féminins, les entreprises commencent à se mobiliser sur le sujet, à la fois dans leurs services RSE mais aussi au sein des ressources humaines ou des équipes métiers. Le cercle InterElles réunissant des femmes de 14 grands groupes s'est ainsi lancé dans une étude pour savoir si les responsables d'équipes métiers IA avaient bien en tête cet objectif de revalorisation des effectifs féminins : selon les résultats, 55% des personnes interrogées ignoraient l'existence d'une telle démarche... preuve que le chemin est encore long. Quant aux services de ressources humaines ou aux cabinets de recrutement, ils sont encore hésitants sur la démarche à adopter, entre discrimination positive assumée et attentisme : « Je ne fais pas de ciblage spécifique « femmes » mais je note que certaines entreprises peuvent être sensibles à l'argument », explique Maité Allain, responsable du cabinet Upward Data.

Reste que le quota est encore considéré comme un terrain miné et que les associations s'y refusent : « Trop blessant d'être remise en cause sur sa légitimité », pointe Amira Lakhal. Pourtant, en Norvège, l'Université de Sciences et Technologies de Trondheim a, elle, franchi le pas et proposé de lancer un recrutement sous quotas... avec pour résultat un bond de 7 à 40% de ses effectifs féminins.

Alors, peut-on vraiment aller contre la tendance ? « L'Intelligence Artificielle, pas sans elles ! » semblent crier Aude Bernheim et Flora Vincent dans leur dernier livre paru en mars 2019. Elles y reviennent sur le déséquilibre de genre observé dans l'IA et alertent sur les risques de biais encourus. Un livre préfacé par un certain... Cédric Villani. Preuve, peut-être, qu'à défaut de résultats chiffrés, le message, lui, semble bien passé.

|| Je ne fais pas de ciblage spécifique « femmes » mais je note que certaines entreprises peuvent être sensibles à l'argument ||

POUR EN SAVOIR PLUS :

Quelques initiatives

- Women in Tech
- Women in AI
- Duchess France
- #SheMeansBusiness
- Girls in Tech
- Les Hackeuses







Hanan
Salam

Co-fondatrice Women in AI
et chercheuse sur les biais

La sur-représentation des hommes dans l'IA multiplie le risque de voir apparaître des biais

|| Il n'y a pas
de consensus sur
la façon de mesurer
l'équité des
algorithmes ||

**Pourquoi est-il important pour l'IA et pour la société en général
d'assurer la présence de femmes dans les équipes de développement ?**

Tout simplement parce que la sur-représentation d'hommes dans ces équipes multiplie le risque de voir apparaître des biais et préjugés au cœur des algorithmes : si l'on considère que l'IA est le reflet de l'intelligence du concepteur dans la machine, alors il est probable que ces équipes masculines façonnent leur modèle selon leurs propres modes de pensées, leur propre partialité. Et quand on considère la force transformatrice des algorithmes d'IA pour les années à venir, il y a urgence à agir...

Avez-vous un exemple de biais à nous donner ?

Prenez Google translate « turc/anglais » et tapez "O bir dockor" côté turc. Pour info, en turc, le pronom « O » est neutre, donc sans attribution de genre. Eh bien, Google translate côté anglais vous proposera « He's a doctor ». Maintenant essayez le même exercice avec « O bir hemĐire », qui parle cette fois-ci du métier d'infirmier. La traduction anglaise sera... « She's a nurse ».

Un état de l'art réalisé en 2014 a cité 422 papiers de recherche sur le sujet - dont 50% entre 2008 et 2012

Que peut faire la Recherche sur le sujet ?

Concrètement il y a deux terrains d'exploration dans la recherche actuelle : l'identification des discriminations d'une part, et la prévention de celles-ci d'autre part. Le challenge, c'est l'aspect pluridisciplinaire de la problématique, car elle nécessite à la fois des données sociologiques, des définitions légales et des techniques statistiques et informatiques. Ce qui peut expliquer les difficultés avec les approches actuelles.

C'est-à-dire ?

Il n'y a pas de consensus sur la façon de mesurer l'équité des algorithmes : des dizaines de mesures, parfois contradictoires, coexistent, ce qui rend les correctifs proposés très difficiles à généraliser. D'autre part, chaque solution est généralement adaptée à un contexte et une situation spécifique, ce qui rend là encore la généralisation difficile, y compris à d'autres types de discrimination.

Les efforts de Recherche se sont-ils accentués ?

Cela fait de nombreuses années que l'on s'intéresse au sujet (depuis le début des années 2000 en fait) mais il y a un vrai phénomène d'accélération : un état de l'art réalisé en 2014 a cité 422 papiers de recherche sur le sujet - dont 50% entre 2008 et 2012. Au niveau mondial, de nombreuses initiatives émergent pour lutter contre les inégalités hommes-femmes dans le domaine de l'IA : je citerais notamment l'ONG AI4all fondée par Fei-Fei Li, professeur à Stanford.

Et en France ? en Europe ?

En France, l'ambition se manifeste surtout dans le rapport Villani qui consacre une partie de ses recommandations à la thématique de la mixité. Mais le volontarisme s'affiche aussi au niveau européen où une Commission spéciale a été mise en place pour suivre les questions d'éthique et de diversité : notre association Women in AI fait ainsi partie d'AI4EU, le projet phare de l'Union Européenne en matière d'IA pour partager des ressources en IA et faire de l'Europe un acteur d'envergure mondiale sur le sujet. Nous cherchons vraiment à sensibiliser le public à cette problématique et attirer davantage de femmes dans l'industrie.

Our Programs

Shape the future by learning about AI today. Our programs are open to high schoolers around the US and around the world.

Summer Programs

Learn on-campus from top AI researchers at universities around North America with peers who care about solving real-world problems.

[EXPLORE SUMMER PROGRAMS](#)

Open Learning

Bring AI education to where you are, and learn to build your own AI project for social good.

[EXPLORE OPEN LEARNING](#)

DANS LA PEAU DE...

Amira *Lakhal*

Développeuse Hadoop/Cassandra/Spark
et membre du bureau Duchess France

En tant que **femme**, il faut apprendre à justifier sa solution



FORMATION Je suis venue au code par le jeu vidéo, parce qu'à l'époque c'était le début des systèmes d'exploitation et la plupart des ingénieurs en informatique en passaient par là. Et puis j'ai décroché mon diplôme d'ingénieur et je me suis spécialisée sur java. Aujourd'hui, je me forme beaucoup toute seule via l'open source ou des ateliers de fournisseurs : il faut aimer découvrir les nouvelles solutions que le marché amène, en tous cas toujours rester en veille...



MIXITÉ Quand j'ai eu mon diplôme, on était trois filles sur une quarantaine de garçons. Cela ne me dérangeait pas trop car il y avait une vraie bienveillance du côté des garçons et j'étais focalisée sur mes études. C'est vraiment avec l'entrée sur le marché du travail que les choses ont changé de façon choquante : on me demandait tout le temps si j'étais RH ou commerciale, il fallait se battre dans les réunions... forcément, cela amène quelques frustrations.



EQUIPE J'aime toujours beaucoup l'aspect créatif du travail en équipe : le fait de trouver ensemble une solution aux problématiques du client. Souvent, on part de rien donc c'est d'autant plus challengeant. Et, en tant que femme, quand on a une idée, il faut apprendre à justifier sa solution devant les autres ou intervenir quand on pense que l'équipe va droit dans le mur. Il faut « tenir tête » et l'effort demandé à une femme là-dessus est peut-être plus important que celui demandé à un homme qui, culturellement, a moins de mal à se vendre.



DUCHESS Duchess est née il y a neuf ans et j'ai rejoint l'association un an après sa création : l'objectif était vraiment d'apporter un espace d'échanges et de networking via des ateliers d'expertise (on fait beaucoup d'ateliers de formation sur des nouveaux outils). Au début c'était spécialisé sur les programmatrices java parce qu'on se rendait compte qu'elles étaient seules dans leurs équipes et n'allaient pas trop aux soirées de networking. Devant le succès et les besoins, on s'est ouvert à d'autres langages de programmation et d'autres spécialités de l'IT.



HOMMES Duchess est ouverte aux hommes et c'est un principe que nous défendons : à l'origine, certaines femmes souhaitaient rester sur une association 100% féminine mais nous avons voté pour la mixité. Certains hommes viennent aux réunions et échangent avec nous, ce qui apporte un enrichissement mutuel : ils nous disent souvent qu'ils ne se rendent pas compte que l'environnement ne met pas les femmes à l'aise. Participer à Duchess les rend plus sensibles à la problématique et à l'envie de recruter plus de femmes.



CONFIANCE Chez Duchess, nous proposons beaucoup d'ateliers de prise de parole pour aider les femmes à se lancer en public ou à se défendre lors d'entretiens de recrutement. Je crois qu'il faut lutter contre le syndrome de l'imposteur que beaucoup de femmes ont : si elles ne prennent la parole que lorsqu'elles sont 100% sûres de leur réponse, alors il y a toutes les chances qu'un homme ait parlé avant. Pourtant elles sont souvent très calées sur les technologies et la solution à adopter...



DATA ENGINEER Sur la mixité des postes data il y a une gradation : les rôles de data analysts sont plutôt bien pourvus en femmes, puis il y a les data scientists qui commencent à se féminiser... mais les data architects et data engineers restent encore très masculins. Cette gradation correspond à une proximité relative avec le client et le business : plus on s'en approche, plus on trouve de femmes, plus on se rapproche de la technologie et du produit, plus les équipes sont masculines. On peut lutter contre cette segmentation !



QUOTA Chez Duchess, nous sommes opposés aux quotas parce qu'ils ne valorisent pas le potentiel d'une personne. J'ai déjà eu à subir des sous-entendus sur le fait que j'avais obtenu mon poste parce que j'étais une femme, c'est très blessant. Nous préférons plutôt soutenir l'initiative du CV anonyme qui met davantage en avant les compétences et la motivation.

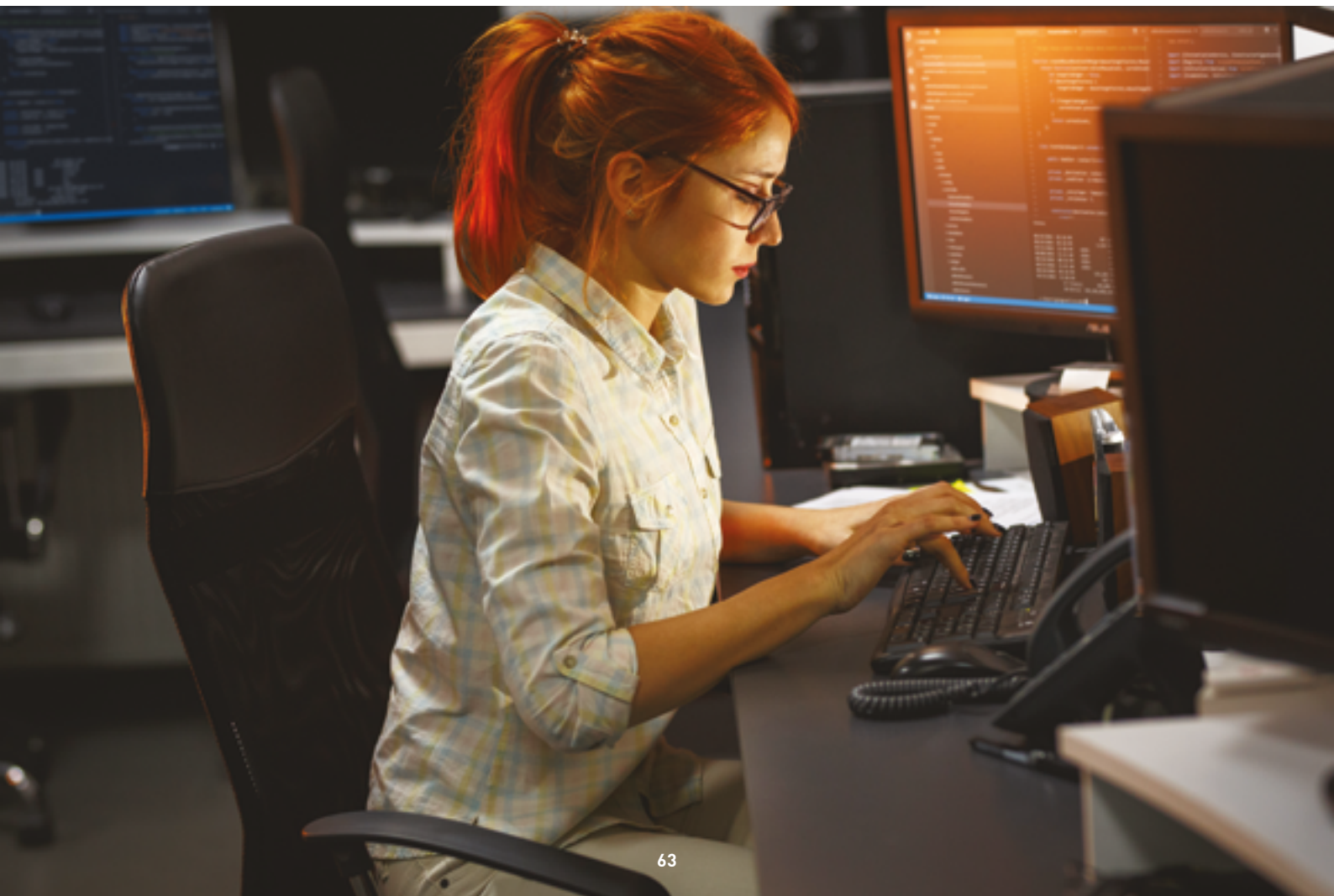


RELÈVE Dans les écoles, quand on participe à des forums d'orientation ou des ateliers de code¹, on se rend compte que les jeunes filles sont curieuses de notre métier. Cela les étonne de voir des femmes maquillées qui viennent leur parler de code, parce qu'elles ont encore le préjugé du geek masculin en tête. On met alors en avant l'aspect social du métier, le côté « team working » et « innovation »... au final la programmation ne leur fait pas forcément peur.



RÉORIENTATION On s'est rendu compte qu'il y avait de plus en plus de femmes issues d'autres métiers (communication, RH) qui souhaitent se réorienter vers le développement et créer leur propre projet. C'est pour cela qu'on a créé un programme de marrainat : pour leur proposer l'assistance et les conseils d'une développeuse confirmée. Il n'est jamais trop tard pour se former !

¹ Devoox for Kids

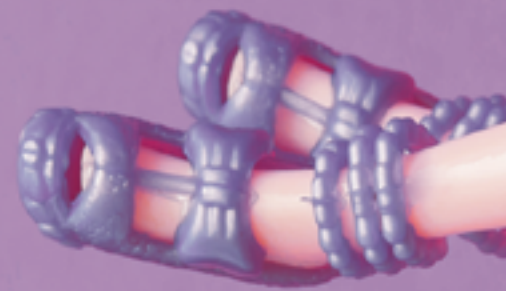


CARRIÈRES

On les espère comme les prochains messies de la data et on s'agace quand ils n'arrivent pas assez vite : depuis dix ans, les data scientists ont débarqué dans les fiches de poste des organisations et la pénurie est encore là : « on manque de candidats au global sur les fonctions data », témoigne Milena Le Mancq, recruteuse chez Kapten, quand on la questionne sur sa stratégie « women in tech ». Le message est clair : d'abord trouver les profils, ensuite catégoriser par critères.

Pourtant, en 2019, la dynamique pénurique semble progressivement s'atténuer, sous l'effet conjugué des formations d'enseignement supérieur et des efforts de mobilité des entreprises. « Chez Kapten, on cherche de plus en plus à créer des mappings de carrière sur les profils techniques pour les orienter soit vers une branche d'expertise, soit vers une branche de management », confirme Milena Le Mancq. Et les entreprises rivalisent d'excellence sur leur marque employeur pour attirer et retenir les candidats.

|| Sur les fonctions « tech » pures comme les data engineers ou les data architects, les pools de recrutement et d'enseignements semblent encore réduits ||



PROFILS DATA : LE MOUTON À 5 PATTES ?



... Reste que sur les fonctions « tech » pures comme les data engineers ou les data architects, les pools de recrutement et d'enseignements semblent encore réduits : « Plusieurs candidats nous ont dit : je me suis formé par moi-même » témoigne Maïté Allain, responsable du cabinet de recrutement Upward Data. Plusieurs années d'efforts au sein des écoles d'ingénieurs semblent donc encore nécessaires... à moins que la technologie n'apporte d'ici là quelques solutions de « patch »?

Démocratiser la data science pour combler la pénurie de talents, c'est en effet la recette conseillée par le cabinet Deloitte dans son article éponyme de décembre 2018¹ ; le cabinet y met en lumière les nombreuses innovations en data science qui pourraient permettre aux profils métiers de s'immerger dans la data : Machine Learning automatisé, développement sans codage, algorithmes d'IA pré-entraînés... Un premier pas qui ne convainc pas encore tout à fait : « Certaines plateformes et bibliothèques mettent à disposition des outils qui permettent de démocratiser un pan de l'analyse de données mais les tâches des data scientists restent très spécifiques et nécessitent des compétences techniques poussées », commente Maïté Allain.

Ironie du sort : chez Kaptén, l'un des cas d'usage « data » que l'on expérimente est justement... le recrutement de data analysts ! Monitoring des candidats qui abandonnent au milieu des tests techniques, scoring des hard skills et soft skills... le process de recrutement est désormais totalement data-centric : « On souhaite de plus en plus objectiver le choix d'un candidat par rapport à un autre », explique Milena Le Mancq, qui a lancé le processus.

La data au secours d'elle-même ? Le Big Data ne pouvait imaginer meilleur scénario...



Milena
Le Mancq

Recruteuse chez Kaptén



¹ "Democratizing data science to bridge the talent gap", article de Deloitte USA paru le 13 décembre 2018

Conférence & Exposition

AIPARIS **X**corp

BY

4^e
ÉDITION

22-23 JUIN 2020

Palais des Congrès de Paris

Venez à la **rencontre** de
l'**écosystème IA** lors du
plus grand événement
B2B français.

6 000

VISITEURS
ATTENDUS

150

INTERVENANTS

200

PARTENAIRES



Plus d'informations sur www.aiparis.com

INTERVIEW

BINAIRE

01101
0010



Guillaume
Chevillon

Co-directeur académique / professeur
Master in Data Sciences & Business
Analytics
ESSEC & CentraleSupélec



Former les profils **management** à la **data** (et vice-versa)

Cela fait seulement cinq ans qu'il a monté le master ESSEC/Centrale-Supélec sur le Big Data et, depuis, tous les indicateurs sont au vert : « Cette année on a obtenu une 4e place au classement QS² et, en parallèle, le Ministère de l'Enseignement Supérieur et de la Recherche nous a donné l'accréditation - ce qui, à l'ESSEC, était réservé aux deux seuls master Grande Ecole et master Finance et constitue une première pour un master conjoint entre deux Ecoles, d'ingénieur et de commerce. Cela nous confirme qu'il faut poursuivre et renforcer le programme : c'est un Master d'avenir ! ».

Si Guillaume Chevillon est aussi enthousiaste, c'est aussi parce que le défi était de taille : mixer des cours de Machine Learning et de coding SQL/NoSQL avec des enseignements business (finance, marketing, supply chain), c'était un pari et les profils concernés ne semblaient pas légion. Mais l'engouement autour du Big Data et les attentes fortes des entreprises ont donné l'impulsion...

Guillaume Chevillon nous donne les clés de « son » Master. En format binaire.

² Le classement des Masters Business Analytics et Data Sciences issu de la revue académique *Quarterly Science*, section Business Analytics

DATA SCIENTIST OU BUSINESS ANALYST ?

Les deux bien sûr ! L'objectif du master, c'est justement de lier les deux compétences parce qu'on s'est rendu compte que les entreprises avaient un double besoin : il faut une vraie compétence d'ingénieur pour assurer une bonne interaction avec le client et vice-versa, une bonne capacité business pour trouver les modèles algorithmiques adéquats. C'est ici que le Master cherche à apporter de la valeur : sur l'interface entre les données à analyser et les besoins du client.

THÉORIE OU APPLICATION ?

Là encore : les deux ! Et c'est d'ailleurs tout le challenge pédagogique : réussir à mixer ces deux aspects au sein de chaque cours (et non par phases calendaires comme c'est parfois le cas pour les formations concurrentes). Au-delà des enseignements, on s'efforce vraiment de faire intervenir le monde de l'entreprise, que ce soit en Masterclass ou lors d'études de cas. Et puis il y a le projet de recherche appliquée où les étudiants. Et puis il y a le projet de recherche appliquée où les étudiants travaillent sur des données issues de l'entreprise pour apporter des réponses à une problématique business...

« FROM SCRATCH » OU HÉRITAGE ?

Je dirais « Héritage ». Historiquement, on retrouvait des formations orientées data dans les écoles de commerce, mais c'était plus spécifique et avec un vrai prisme « économétrie et statistique » en actuariat, finance ou marketing quantitatif par exemple... Ce qui est original ici, c'est l'hybridité : on forme véritablement des ingénieurs-managers ! Et là-dessus, l'héritage de l'ESSEC et de CentraleSupélec est prégnant puisque ce sont les deux premières écoles à avoir permis à leurs étudiants respectifs issus de classes préparatoires d'obtenir leurs deux diplômes : ingénieur et manager, manager et ingénieur.



COURS À CENTRALE OU COURS À L'ESSEC ?

Les deux ! Comme c'est un diplôme commun, les élèves sont amenés à évoluer sur les deux campus, que ce soit pour les cours fondamentaux (maths appliquées, informatique, predictive analytics) ou pour les cours électifs. Les enseignements de Data Sciences comme « Advanced Deep learning » ou « Natural Language Processing » auront lieu à CentraleSupélec. Ceux de Business Analytics (comme Quantitative Risk Management, Business Intelligence ou Marketing Analytics) auront lieu à l'ESSEC. Au minimum, les élèves passent trois mois sur chaque campus.

FACILE OU DIFFICILE ?

Si je suis honnête, je dirais Difficile. Les étudiants travaillent beaucoup en un an, il y a beaucoup de compétences à acquérir ou développer, et pas mal de travaux de groupe. Pour les profils issus du management, nous validons lors de la phase de sélection leur appétence pour la technicité et les mathématiques, c'est indispensable. Mais, au final, on constate que nos étudiants sont presque 50/50 entre les profils issus du management et les profils ingénieurs. Et c'est un programme intensif, certes, mais il permet à chacun de se spécialiser en fonction de ses goûts et objectifs professionnels.

FRANÇAIS OU ÉTRANGERS ?

Plutôt étrangers ! 70% de nos étudiants sont non-francophones et proviennent d'universités étrangères comme la Bocconi, la LSE ou Yale. D'ailleurs tous les cours sont enseignés en anglais. En tout, on recense une quinzaine de nationalités, ce qui n'est pas étonnant car la compétition entre recruteurs sur ces profils se joue à l'échelle internationale. Et on voit d'ailleurs d'autres masters concurrents se développer (par exemple en Australie ou à Hong-Kong), ce qui montre l'intensité des besoins.

EN 1 OU 2 ANS ?

Les deux ! Au départ, le programme se déroulait sur une année, pour des étudiants avec un bagage professionnel ou un Bac+4. Et on prévoyait une petite période entre août et septembre pour rattraper l'éventuel retard. Mais on s'est rendu compte que, pour certains, il valait mieux augmenter la période de remise à niveau ; et que l'on pouvait également ouvrir les enseignements aux étudiants non professionnalisés pourvu qu'ils compensent par une année : c'est pour ces profils-là qu'on a fait évoluer notre Master en deux ans (même s'il est toujours possible de le faire en un an). On compte aujourd'hui une quarantaine d'étudiants en « première année ».

JUNIORS OU SENIORS ?

Stricto sensu, tous les étudiants sont plutôt juniors (dans la mesure où ils ont moins de 35 ans). Mais on se rend compte qu'il y a de plus en plus de profils qui s'inscrivent après plusieurs années d'expérience professionnelle pour se réorienter, parce que la data leur offre beaucoup de perspectives : ils voient dans ce master une formation accélérée. A noter : pour les profils plus juniors, qui sortent d'école, nous sommes particulièrement vigilants à ce qu'ils effectuent leur stage dans un environnement « coaché », en présence de seniors, et non dans des pools de data scientists juniors où ils risqueraient d'être livrés à eux-mêmes, sans bagage de formation.

GRANDS GROUPES OU START-UP ?

Là-dessus, je dirais les deux et même plus encore... L'éventail de débouchés qu'on a observés est vaste : bien sûr, les filières classiques (banque, conseil, marketing, GAFAM...) recrutent volontiers chez nous, mais il y a une forte appétence chez les étudiants pour le monde des start-up et des structures plus légères (même si un bon nombre d'étudiants privilégie l'embauche deux-trois ans dans un grand groupe avant de basculer côté entrepreneuriat). Ce qui est plus inattendu, ce sont des filières comme la parfumerie ou l'œnologie qui témoignent que l'analyse quantitative s'est imposée même dans des secteurs de niche ! Et je n'ai pas mentionné ceux qui partent faire un PhD ou une thèse CIFRE...

FRANCE OU INTERNATIONAL ?

Pour cette année, la réponse est France. Mais le développement à l'international fait partie de nos ambitions : déjà on offre aux étudiants la possibilité d'effectuer leur troisième trimestre sur le campus de Singapour. Et nous sommes en discussion pour des accords d'échanges avec d'autres universités, notamment avec l'objectif de renforcer les spécialisations appliquées : par exemple dans les domaines agrotech, fintech ou dans la santé...

INTERVIEW

PORTRAIT-ROBOT



Maité
Allain

Responsable d'Upward Data

Depuis 2015 et le lancement d'Upward Data² dont elle est responsable, Maité Allain est sur tous les fronts pour recruter les talents data : analytics, data science, architecture, project management... Ce n'est pas moins d'une centaine de profils qu'elle a rencontrés. Pas moins d'une centaine avec qui elle a pu échanger pour mieux comprendre les parcours et aspirations et proposer des

opportunités adaptées au sein de cabinets de conseil, entreprises du web, grands groupes et start-up.

Aspirations, craintes, salaires, exigences... Maité Allain partage avec BIG sa connaissance du candidat-type dans la data.

² Business Unit data du cabinet de conseil en management Upward Consulting

ASPIRATION

Ça recherche quoi un candidat data ?

Ça recherche d'abord une vraie stratégie data, un projet qui mette la data au cœur du produit et au cœur du business. Les premières générations de candidats ont parfois été mises en marge, dans des pools data qui ne les valorisaient pas ; du coup les générations suivantes sont plus méfiantes. Aujourd'hui, je dis aux entreprises : « Soyez très transparents sur les fiches de poste et sur les enjeux auxquels la data correspond pour vous. Sinon, vu le niveau de concurrence du marché, les candidats partiront rapidement ».

PARCOURS

Ça vient d'où un candidat data ?

Assez classiquement, ça vient des écoles d'ingénieurs, des formations mathématiques et informatique ; mais on recense quand même un petit vivier d'écoles de commerce. Et puis des post-doc et sorties de thèses qui se reconvertissent... En ce qui nous concerne, les candidats ont déjà trois ans d'expériences minimum, et c'est généralement dans les cabinets de conseil ou dans les services data des grands groupes qu'on a pu les dénicher (même si les territoires de chasse sont très spécifiques par applicatif).

CRAINTES

Ça craint quoi un candidat data ?

Par-dessus tout : le manque d'agilité des entreprises, les validations qui prennent un mois et l'aspect siloté des services. Un jour, j'ai un candidat qui m'a dit : « J'avais besoin d'un outil pour travailler qui coûtait trente euros. La validation prenait tellement de temps que j'ai fini par l'acheter en notes de frais pour pouvoir me mettre au travail ». Par essence, les profils data ne comprennent pas non plus les logiques de rétention de données : c'est ce qui les pousse souvent à préférer la start-up au grand groupe.

PÉNURIE

C'est rare un candidat data ?

Cela dépend du poste : les data engineers et data architects sont effectivement les profils les plus pénuriques, comme tous les profils IT en général. En revanche, sur la filière data scientists, je trouve que le vivier s'est bien élargi : les écoles ont pris le pli, on manque moins de profils qu'avant. Après, il peut y avoir des déséquilibres sur certains critères, comme celui du genre par exemple : là-dessus, je ne fais pas de ciblage spécifique « femmes » mais je note que certaines entreprises peuvent être sensibles à l'argument.

SALAIRE

Ça gagne combien un candidat data ?

Bonne question. Le marché est très disparate, difficile d'établir des grilles et ce n'est pas faute d'avoir essayé ! Pour un poste de Data Scientist en sortie d'école, les rémunérations fixes varient de 35k à 55k... Et dès trois ans, on peut atteindre 70-75k. Tout dépend des parcours, des opportunités et des exigences - notamment école - des entreprises.

RECRUTEMENT

Ça intéresse qui les candidats data ?

Les cabinets de conseil, les entreprises du web, les grands groupes, les start-up... la liste est longue. Mais, si je dois synthétiser, je dirais qu'il y a deux approches de recrutement data : ceux qui ont des besoins urgents à combler, et ceux qui restent en veille sur des profils exceptionnels. Les entreprises 100% digital et pure player data sont notamment les plus susceptibles de rentrer dans la deuxième catégorie.

AVENIR

Ça évolue vers quel profil un candidat data ?

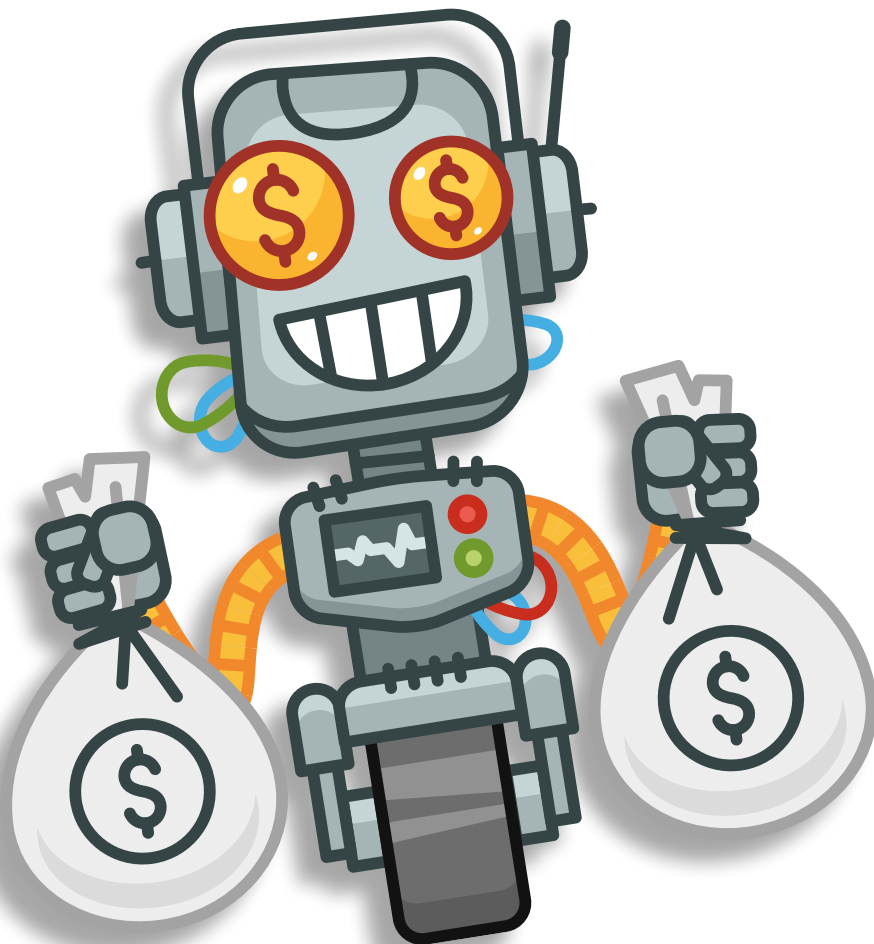
Pour moi, il y a deux tendances qui opèrent des mouvements inverses : d'une part, on voit émerger des postes très spécifiques techniquement comme le « Machine Learning engineer » qui met en production les algorithmes de data science, et d'autre part on assiste à l'ouverture des profils

data vers les fonctions plus généralistes comme le marketing, le produit ou la stratégie. Je ne serais pas étonnée que certains candidats qui sont « head of data » aujourd'hui deviennent « head of marketing » ou « head of strategy » demain : la data a essaimé sur toutes les fonctions clés de l'entreprise.

EXIGENCE

Ça demande quoi un candidat data ?

Ça demande des équipes de haut niveau technique, des problématiques Data au cœur du business, des environnements agiles, et... de plus en plus de sens dans les missions - ce qui explique pourquoi je vois davantage de candidats se diriger vers les secteurs Santé, Energies renouvelables ou Agrotech, plutôt que Banque ou Adtech. Après, je leur conseille quand même la patience : on ne peut pas avoir tout tout de suite et les secteurs n'ont pas tous le même niveau de maturité technique. Ceux qui passent leur temps à papillonner de poste en poste peuvent faire peur aux entreprises. Pour faire le bon choix, il faut se renseigner en amont sur l'équipe Data, son fonctionnement et les enjeux qu'elle porte.



PROSPECT

ET SI **L'AVENIR** DE LA DATAVISUALISATION PASSAIT PAR **L'IMMERSION 3D** ?

Encore embryonnaires en cette fin des années 2010, la Réalité Augmentée et sa fausse jumelle la Réalité Virtuelle laissent entrevoir des usages dans de multiples environnements professionnels (industrie, médecine, agriculture, etc). Et le secteur de la data n'y échappe pas.

Avec ses besoins en matière de représentation des données et de démocratisation de l'accès aux métiers, nul doute que les Big Data y trouvent ainsi un remède contre leur complexité. Pour naviguer au cœur des data et analyser plus rapidement les résultats de leurs traitements temps réel, quel meilleur moyen que de les incarner physiquement ?



WAVE

LA DATAVIZ EN AR/VR





Pourtant, ce pari un peu fou, peu de start-up ou grands comptes l'ont encore relevé. En cause : le prix encore trop prohibitif des dispositifs d'équipements, et la lourdeur de l'ergonomie associée. Les quelques exemples identifiés proposent souvent cette offre en complément d'un service d'analytics ou de dataviz 2D : Virtualitics (USA), DataKlub (France), DataView VR (USA), 3Data (USA) ou, dans un domaine plus appliqué : Energiency (France, spécialiste Big Data de la performance énergétique).

ALORS...

AR OR NOT AR ?

Si vous hésitez encore, la citation d'Edward Tufte pourrait vous inspirer : « L'excellence en matière de graphisme statistique consiste à communiquer des idées complexes avec clarté, précision et efficacité »²

LIMPIDE N'EST-CE-PAS ?



A moyen terme, il semble cependant que le marché AR/VR pour les visualisations Big Data soit appelé à un bel avenir. Des études menées par plusieurs groupes de recherches au sein de Caltech University ou de l'université de Rio Grande pointent les apports opérationnels importants de ces dispositifs par rapport à la 2D ou la 3D sur ordinateur. Outre une meilleure identification des distances et des aberrations statistiques, ces technologies proposent une vraie valeur ajoutée sur la manipulation des objets. Avec en ligne de mire : un meilleur engagement des métiers.



Mais dès que l'on évoque le ROI potentiel de ces technologies, il reste encore difficile d'avancer des chiffres... Dans une étude récente¹ qui concerne l'ensemble des applications AR/VR en entreprise (et pas seulement les Big Data), Cap Gemini indique que 82% des entreprises ayant déployé des technologies AR ou VR ont obtenu des résultats qui atteignent ou dépassent leurs attentes... De quoi y voir un signe encourageant, d'autant que la France est placée par le cabinet d'études dans le tiercé de tête des pays leaders en la matière, derrière (sans surprise) les Etats-Unis et la Chine. Mieux : l'étude souligne la volonté des « early adopters » de continuer à investir pour déployer ces solutions à grande échelle...

Vous êtes plutôt...



... smartphone ?

La Réalité Augmentée est tout à fait compatible avec les systèmes d'exploitation Android et iOS. Il suffit d'intégrer les API AR Core (pour Android) et AR Kit (pour iOS) pour visualiser à travers l'écran de son smartphone l'environnement enrichi d'objets 3D, et surtout apprendre à interagir avec les éléments via l'aspect tactile de l'écran.

Avantage : La simplicité ! Pas de branchement, pas d'équipement additionnel, juste une application !

Inconvénient : Les capacités du système sont limitées tout comme le nombre d'interactions. À proprement parler, il ne s'agit pas de Réalité Augmentée dans la mesure où l'écran opère un filtre avec l'environnement.



... casque ?

Dispositif-roi de la Réalité Augmentée (voir la Réalité Mixte, cf encadré 2), le casque HoloLens de Microsoft propose une immersion dans l'environnement grâce à une projection par hologramme sur les lentilles du casque. Totalement autonome, il permet à l'utilisateur de se déplacer parmi les objets grâce aux capteurs de mouvement. Son seul concurrent à date : Magic Leap.

Avantage : Une immersion complète dans la modélisation 3D qui ouvre la voie à de nombreuses applications industrielles.

Inconvénient : Son prix. 3299€ TTC pour le HoloLens 1, 3073€ TTC pour le HoloLens 2....



... lunettes ?

Conçues pour apporter légèreté et ergonomie à l'utilisateur, les lunettes d'AR proposent de projeter des informations (GPS, météo, etc) ou des objets 3D directement devant les yeux de l'utilisateur via les deux verres de la monture. Après l'échec des Google Glass considérées comme peu esthétiques et trop intrusives (la caméra était intégrée), le marché s'enrichit de nouveaux produits - comparables en apparence aux lunettes de vue.

Avantage : Comme dans Iron Man, vous avez directement accès à l'information sans écran interposé. Et ça c'est cool.

Inconvénient : Euh... vous aimez vraiment porter des lunettes ?

¹ CapGemini Research Institute "Augmented and Virtual Reality Analysis", Septembre 2018

² in « The Visual Display of Quantitative Information », 1983

INTERVIEW



Tony
Alibelli

CEO et CTO de DataKlub

Nous sommes entrés dans l'ère de l'Immersive Analytics

|| Je suis persuadé que le lancement de l'Hololens 2 va nous permettre d'enrichir les contenus proposés et les modes de représentation ||

Tony Alibelli a créé DataKlub en 2014 et, depuis, l'évangélisation autour de la Réalité Augmentée est devenue un leitmotiv. Parmi ses services : un outil AR/VR agnostique, misant sur un requêtage NLP propre et couplé avec Splunk pour la partie analytics. La solution pourrait bien faire grand bruit en novembre lors de son lancement...

Tony Alibelli fait le point pour nous sur les enjeux de la révolution AR/VR.

La Réalité Augmentée, la Réalité Virtuelle, cela fait rêver mais on reste encore sur notre faim quant à leur déploiement...

C'est normal car nous en sommes encore aux prémices ! Des cabinets comme Forrester ou Gartner le disent : le déploiement va prendre encore un peu de temps car on manque de cas d'usage. Mais la maturité n'est plus si loin : de grandes entreprises comme Airbus font appel à ces technologies, et plus globalement tout le secteur de la Smart Industrie s'équipe. Sans compter les progrès des équipements : on est désormais capable de détecter l'ouïe, l'odorat...

Concrètement, qu'est-ce que ces technologies apportent à l'analyse des Big Data ?

Pour moi, il y a trois « valeurs ajoutées » très fortes : 1/ un meilleur accès à la complexité de la donnée, 2/ un meilleur pilotage des paramètres grâce aux gestes et à la voix, et 3/ une meilleure collaboration temps réel entre ceux qui interviennent sur la donnée.

Revenons d'abord sur l'accès à la donnée...

Le grand « plus » de la Réalité Augmentée, c'est qu'elle fait appel aux sens, en plus du rationnel : elle exploite l'intuition, la reconnaissance naturelle que l'on a des objets grâce à la représentation 3D. Beaucoup d'informations nous échappent en 2D alors qu'elles apparaissent évidentes en 3D, d'autant plus avec la possibilité de zoomer au cœur de celles-ci.

Par exemple ?

Par exemple les données d'interactions : on a voulu représenter des menaces de cybersécurité par des graphiques de liaisons entre plusieurs adresses IP. En passant en 3D, les connexions entre les différents nœuds apparaissent de façon beaucoup plus naturelle, tout comme les catégories de flux matérialisées par des traits et des textures...

Et puis, on peut aussi travailler sur la représentation métaphorique de tout cela : on peut très bien imaginer qu'une menace de cybersécurité fasse changer les couleurs de l'environnement, en passant d'un ciel très clair à un ciel brumeux voire orageux pour alerter sur l'intensité de la menace.

En termes d'interactions avec les données, ce serait une nouvelle forme d'UX Design ?

Oui car la Réalité Augmentée permet de pousser des interactions fluides et naturelles grâce aux gestes et à la voix. C'est d'ailleurs dans ce

cadre que le Deep Learning est mis à contribution, grâce au NLP qui permet de requêter l'outil d'analyse sans passer par une commande textuelle. Avec à terme, un pilotage plus réactif de la phase d'analyse (j'aime d'ailleurs parler « d'Immersive Analytics »).

Est-ce le seul emploi de l'Intelligence Artificielle ?

Non, on s'oriente également vers davantage de reconnaissance d'images pour recréer en temps réel les environnements d'interactions : une usine, une chaîne de montage, des produits pourraient ainsi être plus facilement identifiés...

Il y a un autre apport important que vous avez évoqué tout à l'heure : la collaboration...

Effectivement, l'environnement recréé par la Réalité Augmentée ou la Réalité Virtuelle peut permettre à plusieurs individus de se connecter et interagir avec un objet en temps réel depuis différents postes géographiques. C'est une forme de couplage entre représentation des données d'une part et mode collaboratif d'autre part.

Cependant, derrière tous ces apports, il y a des best practices à préciser...

Disons que l'AR/VR pour l'analytics représente une immersion ponctuelle, il ne faut pas considérer que cela remplace le poste de travail. C'est un complément pour mieux comprendre les données, essentiellement celles qui concernent les flux ou les logs.

Et puis il faut bien définir en amont ce que l'on souhaite faire d'un point de vue UX (filtrage, représentation) pour coller aux besoins métiers.

A date, quels sont les secteurs qui s'intéressent à votre solution ?

Nous travaillons beaucoup avec le secteur bancaire : la Société Générale nous a ainsi demandé un graphe de liaison entre les 700 applications de la banque de détail, tandis que BNP Paribas a travaillé avec nous sur le monitoring de certaines transactions financières. Au-delà, nous avons d'autres projets dans le domaine de la « remote assistance³ » pour l'industrie, ou encore dans le secteur pharmaceutique.

Existe-t-il des différences d'apport entre l'AR et la VR pour représenter les données ?

Pour certains clients, cela peut avoir un intérêt d'aller plus loin en termes d'immersion en recréant virtuellement tout l'univers. Cependant, il faut être attentif à la performance de l'ensemble : plus on ajoute d'objets et d'interactions - d'autant plus en device autonome - plus on met de pression sur la performance des cartes graphiques et des outils dynamiques.

Justement qu'attendez-vous des prochaines évolutions sur le plan technologique ?

À partir de septembre, Microsoft livrera ses premiers dispositifs HoloLens 2, annoncés comme beaucoup plus puissants en termes de CPU et de mémoire que le HoloLens 1. Cela va probablement nous permettre d'enrichir les contenus proposés et les modes de représentation, jusqu'ici limités en nombre de polygones 3D. Je suis persuadé que ces évolutions et la multiplication des cas d'usage vont permettre d'accélérer la maturité du marché !

sur l'équipe Data, son fonctionnement et les enjeux qu'elle porte.



³ Remote Assistance = Service d'assistance à distance permettant à un expert d'aider un opérateur portant un casque holographique dans un contexte industriel (ex : dessiner en 2D par-dessus le champ de vision de l'opérateur, partager la documentation et des modèles 3D des pièces à maintenir, etc.)

LE WIKI

Réalité Virtuelle OU Réalité Augmentée OU Réalité Mixte ?

#RÉALITÉ VIRTUELLE

Un environnement virtuel est créé à 100% (fondé sur la réplique d'environnements réels ou la conception d'environnements imaginaires) et l'immersion de l'individu dans cet univers est formalisée par l'intégration temps réel de ses mouvements dans ce champ virtuel, ainsi que sa capacité à visualiser l'environnement et ses interactions en format 360°.

Exemples : The Climb VR (gaming), Exposition Léonard de Vinci en VR (Le Louvre), Henry (film)...

#RÉALITÉ, AUGMENTÉE

Dans la réalité augmentée, l'environnement de l'utilisateur reste le monde réel mais la technologie vient superposer en temps réel des éléments graphiques ou sonores : l'écran d'un smartphone ou d'une tablette vient souvent servir d'interface entre les deux « mondes » (le réel et les objets virtuels) et c'est sur cet écran que s'opèrent les interactions entre l'utilisateur et l'image ainsi proposée.

Exemples : Pokemon Go (gaming), Snapchat (application mobile)...pérature et non l'inverse.

#RÉALITÉ MIXTE

Très proche de la Réalité Augmentée, la Réalité Mixte est un concept apparu lors du lancement de l'Hololens de Microsoft pour souligner la disparition de l'interface « écran » - comme c'est le cas pour l'AR - et son remplacement par un dispositif holographique. En réalité mixte, l'individu est donc entouré d'objets virtuels superposés à son environnement réel et ses mouvements sont captés, lui permettant ainsi de se déplacer parmi les objets et d'interagir à 360° avec eux.

Exemples : les expérimentations industrielles de l'Hololens



USAGE CASÉS

On a parlé de marketing, on a parlé d'industrie, on a parlé de finance... et au bout de dix ans, on avait presque l'impression d'avoir fait le tour des usages data. Et puis, voilà, on est tombés sur ces deux-là : small et big.

Alors on s'est dit, comme en conclusion, qu'en 2020 décidément, la data n'aurait pas fini de nous surprendre...

A L'ANNÉE PROCHAINE !

DE BIG À SMALL : LA DATA DANS TOUS SES USAGES



Mélanie *Rochoux*

chercheuse au CERFACS

Utiliser les Big Data pour **prévoir** les propagations d'incendies ou de polluants

A l'origine, il y a un constat : la variabilité importante du climat depuis trente ans et un impact humain de plus en plus fort.

Et puis il y a des symptômes : des incendies estivaux de plus en plus incontrôlables, des épisodes de pollution de plus en plus récurrents, des crues printanières...

Au sein du CERFACS (le Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, basé à Toulouse), ces phénomènes sont constamment observés par l'équipe Climat et Environnement¹.

Mieux : ils sont simulés - c'est-à-dire recréés virtuellement pour permettre l'analyse et la prédiction.

A l'aide de super calculateurs qui englobent des téraoctets de données, les chercheurs de l'équipe sont ainsi capables de simuler mathématiquement le comportement d'un feu ou d'un polluant atmosphérique pour estimer ensuite les surfaces atteintes.

Et livrer ainsi leurs analyses aux décideurs en période de crise.

Observation physique vs Simulation

Mélanie Rochoux est l'un des membres de cette équipe² et son expérience l'a portée plus spécifiquement sur la propagation des incendies, avant qu'elle ne démarre des travaux sur la dispersion des polluants atmosphériques.

Pour elle, il est important de décloisonner les expertises purement mathématiques et les expertises physiques : « Sur ces sujets, j'ai besoin de la vision d'un expert en mécanique des fluides : le traitement statistique ne suffit pas ». Car son travail repose beaucoup sur un croisement entre l'information issue des modèles d'observation physiques (les incendies réels) et l'information produite par les simulations statistiques : le but final étant bien sûr d'affiner au maximum l'algorithme de prédiction et de quantifier l'incertitude.

Le Deep Learning au secours des données d'images

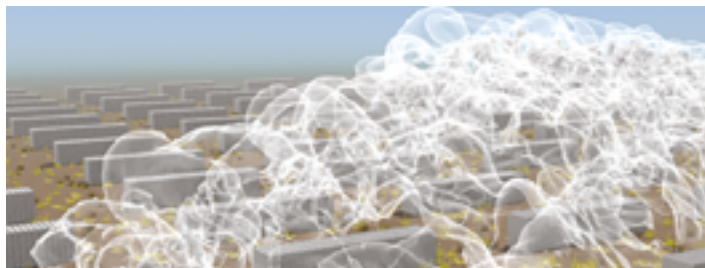
Concrètement, comment se passe une analyse ? A la base, il y a une abondance de données variées : données de topographie (relief, végétation, etc), données météorologiques (qui peuvent atteindre 106 variables), mais aussi données directionnelles. « Ce qui change désormais, ce sont les données non structurées que l'on récupère : celles issues des drones, des caméras infra-rouge ou des capteurs. Elles sont plus complexes à traiter ».

Et c'est là que les premières difficultés apparaissent : « Le problème de ces données, c'est qu'elles sont parfois bruitées, c'est-à-dire peu lisibles ou peu fiables. Par exemple, si on passe en hélicoptère au-dessus d'un feu et que l'on récupère des données pixels caméra au poing, il est fort probable que l'image soit floutée et qu'on ait un doute sur certains pixels : feu ou pas feu ? ».

Tout le challenge réside donc d'abord dans la lecture de ces données : « Ici, le Deep Learning peut nous apporter des solutions : on apprend à un algorithme à lire les images issues des hélicoptères et, après plusieurs apprentissages, il arrivera à détecter le front de feu grâce notamment à la corrélation entre plusieurs pixels qui

peut l'aider à déterminer des structures cohérentes - et donc une zone de feu ». Mais le passage d'une vision physique 3D à une image en 2D est encore un biais que les chercheurs ont du mal à contourner...

L'autre difficulté de ces données d'images réside dans leur structure : par nature, elles sont non structurées et recouvrent des formes très différentes ; elles intègrent également des informations de géolocalisation. Le cœur des travaux de Mélanie repose donc sur ce qu'elle nomme « l'assimilation de données », soit le fait de rendre ces données physiques non structurées combinables avec les données statistiques du modèle de simulation. « À terme, l'objectif est de réduire l'incertitude qui entoure l'amplitude de la propagation ».



Visualisation (avec Blender) d'une simulation de dispersion de polluants sur une expérience de terrain correspondant à un quartier urbain idéalisé - les conteneurs représentant les infrastructures du quartier idéalisé.

Vers l'infiniment petit de la modélisation physique ?

Le Big Data est donc à la manœuvre à la fois dans sa dimension volume (les données de simulation) et sa dimension variété (les données physiques). Mais les enjeux économiques se mêlent à l'analyse : « Bien sûr que le simulateur peut agir sur des téraoctets de données : il en a la puissance de stockage et de calcul. Mais il ne faut pas oublier que chaque simulation va représenter un coût financier et un temps investi non négligeable : chaque simulation réclame de coupler une quantité de modèles algorithmiques (par exemple : modèle atmosphérique, modèle de propagation), ce



Simulations haute fidélité d'un incendie de végétation (brûlage dirigé FIREFLUX réalisé au Houston Coastal Center, USA, en 2006) couplant modèle atmosphérique et modèle d'incendie (réalisées conjointement par les équipes du CNRM (Météo-France, CNRS) et du CECI (CERFACS, CNRS))

|| Ce qui change désormais, ce sont les données non structurées que l'on récupère : celles issues des drones, des caméras infra-rouge ou des capteurs ||

qui prend beaucoup de temps. On ne peut pas se permettre de faire mille simulations... donc il faut donc trouver des parades ».

Et ces parades, ce sont une nouvelle fois les algorithmes de Machine Learning qui les apportent : « Il est possible de ne simuler que quelques scénarii de progression du feu correspondant à un panel de conditions météorologiques et environnementales déterminées. Puis d'utiliser un modèle de machine learning pour interpoler entre chaque scénario ».

Des traitements statistiques extrêmement complexes qui rendent compte des défis engendrés par l'approche Big Data : « Le fait d'avoir augmenté la puissance de calcul et le nombre de données a conduit à un raffinement extrême de la modélisation physique : on entre dans l'infiniment petit, ce qui conduit inévitablement à devoir prendre en compte des phénomènes qu'on aurait ignorés auparavant... ». D'autant que l'introduction des données haute résolution (drones, satellites, etc) complique encore la tâche : « Il faut avoir une analyse statistique préalable sur ces données-là sinon on croule dessous et on ne peut les intégrer dans les algorithmes ».

Plusieurs applications envisagées

Une complexité qui peut expliquer le statut encore très « Recherche » de ce projet : « Actuellement on est davantage en train de développer un démonstrateur qu'un outil industrialisé. À terme, on aimerait proposer un outil qui permette d'identifier les quartiers impactés par des dispersions de polluants lors d'accidents ». Et plusieurs autres applications sont envisagées, pour la détection de risques notamment... mais Mélanie Rochoux estime travailler surtout sur la partie amont de tout cela.

« Mon défi, c'est d'arriver à réunir tous ces traitements (l'observation, la simulation) et de ne perdre aucune information. Et ensuite de combiner toutes ces données pour savoir ce que j'en fais ». Pas si évident selon elle : le Big Data côté recherche physique est encore un modèle à construire...

¹ Équipe CECI – Climat Environnement Couplages et Incertitudes. Unité mixte de recherche CERFACS/CNRS
² NB : d'autres équipes au CERFACS travaillent sur des problématiques industrielles, notamment celle de la mécanique des fluides (ex : simulation du fonctionnement d'un moteur d'avion en milieu atmosphérique variable)

CÔTÉ SMALL



Xavier *Perrin*

Directeur Immobilier,
Microsoft



Jean-Christophe *Buvat*

Directeur Transformation et
Environnement Collaborateur,
La Française des Jeux

Utiliser les Data pour améliorer les espaces de travail

A l'origine, il y a une question : « Dans une journée de travail, quels sont tous les besoins qu'un collaborateur va rencontrer ? » ; et une conviction : « Echanges, concentration, détente, documentation, collaboration... Le poste de travail propriétaire et sédentaire n'est plus en mesure de répondre à cet éventail-là. Et la data peut aider à faire bouger les lignes... »

Ce constat, c'est Jean-Christophe Buvat, Directeur de la Transformation de La Française des Jeux qui l'a fait il y a quelques années.

A l'époque, la Française des Jeux entame un grand mouvement de transformation de ses organisations pour se mettre au diapason des changements du marché (privatisation de l'entreprise, concurrence, nouveaux services...). La stratégie « workplace » fait partie des chantiers-clé : « L'objectif, explique Jean-Christophe Buvat, c'était vraiment de libérer l'autonomie des collaborateurs : leur permettre de prendre davantage de décisions et d'enclencher des initiatives sans se laisser enfermer dans une vision hiérarchique. On s'est rendu compte que le traditionnel espace de travail qui séparait les fonctions de décision des fonctions d'exécution n'était plus en adéquation avec cette philosophie ».

Optimiser et rendre intelligents les espaces de travail

Changement de décor de l'autre côté de la Seine... mais même constat : chez Microsoft, la question de l'« intelligent workplace » est

prise en compte depuis 10 ans déjà à l'échelle du groupe, mais l'utilisation de la data offre de nouvelles perspectives. « Nous sommes passés du paradigme dans lequel l'utilisateur avait un espace de travail dédié (son bureau) à celui dans lequel l'utilisateur consomme des espaces de travail en fonction de ses besoins, explique Xavier Perrin, Directeur Immobilier de Microsoft. La data permet alors de mesurer de façon objective ces besoins utilisateurs à différents moments de la journée afin de dimensionner les bureaux quantitativement et qualitativement dans le plus strict respect de la privacy. L'utilisation de la technologie facilite ainsi l'accompagnement au changement des nouvelles formes de travail en supprimant la subjectivité managériale sur les besoins du collaborateur ».

Sur le site d'Issy les Moulineaux, la transformation s'est engagée dès 2015 et elle a déjà permis de libérer 15% de mètres carrés en plus. « Ce sont des mètres carrés supplémentaires pour pouvoir investir ensuite de manière qualitative sur le bien-être des collaborateurs, appuie Xavier Perrin. Outre les objectifs d'agilité,

d'échanges, de productivité, il y a d'autres enjeux qui sont sous-tendus, notamment sur l'écologie ou l'équilibre vie pro/vie perso : c'est un vrai changement de paradigme ! »

Une analyse small data

Alors concrètement, de quelles data parle-t-on ? Chez Microsoft, les données utilisées sont les données de logs sur les différents points de captation : badgeuses, points wifi... Mais aussi les données issues des SIRH de l'entreprise – par exemple les organigrammes d'équipes en ligne. « Aujourd'hui, on ne se sert pas de données de reconnaissances d'images par exemple, mais on envisage de renforcer la partie IoT sur certains espaces comme les salles de réunion. Par exemple, pour signaler qu'une salle de réunion s'est libérée ou qu'elle doit être nettoyée : la data peut nous aider à améliorer le service et la gestion sur ces espaces ». À proprement parler, pas de déluge de données pour ce cas d'usage (« sur un immeuble de 1500 collaborateurs, difficile de parler de Big Data »), mais une analyse temps réel basée sur le Machine Learning pour savoir en permanence combien de personnes sont dans l'immeuble et quels espaces et objets leur affecter. « On a travaillé sur des populations cibles aux besoins de consommation différenciés pour bien valider les actifs immobiliers affectés à chacun » confirme Xavier Perrin.

Retour à la Française des Jeux : en mai 2018, avec l'aide de Microsoft justement, Jean-Christophe Buvat lance son projet d'espace dynamique. Son premier chantier consiste à calculer le nombre de postes dont l'organisation a véritablement besoin : « Sur les 1300 collaborateurs du site, on s'est rendu compte qu'au pic de fréquentation, on n'était qu'à 850 personnes présentes en simultané dans l'immeuble ». Déplacements professionnels, congés, missions à l'étranger, absences... autant de raisons qui justifient ce différentiel et qui poussent Jean-Christophe Buvat à n'installer que 950 postes. « Le reste a été utilisé pour créer un incubateur, un restaurant, une conciergerie, un lab d'innovation... »

Deuxième étape : avec l'appui de 200 collaborateurs représentatifs de la société, il détermine cinq grands clusters d'activités correspondant aux besoins-type d'une journée de travail - collaboration et échanges, repos et cocooning, travail individuel, activités sportives, espaces pratiques... Et affecte un taux d'occupation d'espace pour chacun (« on n'a pas encore une photographie temps réel des occupations d'espace mais on compte s'équiper de capteurs sur les différentes zones »). En résumé : « On peut mettre son costume à la conciergerie et aller faire du sport entre midi et deux ».

Un changement culturel

Au cœur du dispositif en termes de changement culturel : la suppression du bureau individuel pour tous les collaborateurs (y compris le top management) et son remplacement par une flotte limitée de bureaux clos (1 pour 10 collaborateurs) : « Au total, il y a 109 bureaux fermés : ils ne se réservent pas, seule une petite lumière extérieure indique si le bureau est occupé ou non. On pensait que les gens allaient se ruer dessus le matin et on a été surpris de voir que ce n'était pas le cas... »

D'autant que le projet prévoit aussi de garder une zone d'attache pour



Dans les locaux de la Française des Jeux

|| Nous sommes passés du paradigme dans lequel l'utilisateur avait un espace de travail dédié (son bureau) à celui dans lequel l'utilisateur consomme des espaces de travail en fonction de ses besoins ||

chaque équipe, un « open space » à l'ancienne pour ceux qui veulent retrouver leurs collègues réguliers. L'analyse effectuée chaque soir sur les données de remplissage témoigne ainsi de la sédentarité ou non d'une équipe. « Si on se rend compte que certaines équipes restent très attachées à leur espace commun quitte à créer un engorgement, on peut envisager de lancer une réflexion avec elles sur la meilleure manière d'adapter l'espace et de créer de nouveaux environnements. En terme financier, tester de nouveaux agencements coûte beaucoup moins cher qu'avec les précédentes configurations ».

Alors qu'en pensent les équipes justement ? Lors de son enquête interne effectuée au bout de six mois, Jean-Christophe Buvat a pu constater que 75% des répondants ne souhaitaient pas revenir en arrière... Satisfait ? « Cela signifie qu'il y en a encore 25% qui n'ont pas encore adopté ce schéma, relève-t-il. Mais ce sera progressif... » Il compte d'ailleurs refaire une enquête dans les prochains mois pour mesurer l'évolution du taux de satisfaction.

Chez Microsoft, Xavier Perrin insiste sur cette partie « change management » qui constitue pour lui le plus gros enjeu : « C'est indispensable d'être transparent avec les collaborateurs, de bien leur communiquer ce qui va se produire et pourquoi on le fait. La data permet d'ailleurs souvent de communiquer une mesure objective pour rassurer ».

Côté Française des Jeux, même état d'esprit : la donnée de remplissage est partagée au sein des équipes et les analyses librement accessibles. « Plus on aura de la data fine et actualisée, plus on pourra lui faire dire des choses intéressantes pour les équipes », appuie Jean-Christophe Buvat. Une affirmation qui montre bien que le besoin en data devrait encore s'amplifier sur le terrain du « workplace »... et que, sur ces actifs-là, la création de valeur (financière, sociale, environnementale) ne devrait pas démeriter. Small data, small profit ? Pas si sûrs...



Dans les locaux de Microsoft








CARNET D'ADRESSES










NOTHING
SOUNDS BETTER
THAN **YOUR BIG DATA**



L'Annuaire

	 Analytique	 API	 Applications	 Consultants/SSII Intégrateurs	 Dataviz	 Infrastructure	 Institutionnels
AB INITIO	★	★	★			★	
ADVANCED SCHEMA				★			
ALTERYX	★						
ASG TECHNOLOGIES			★				
ASKR.AI	★				★		
ATACCAMA			★				
BIRST AN INFOR COMPANY	★				★		
BLACK TIGER		★	★			★	
CLARANET			★			★	
CLOUDERA	★		★			★	
COHERIS	★				★		
COLLIBRA	★		★			★	★
CONFLUENT						★	
COUCHBASE						★	
DATAFAB	★			★		★	
DATAIKU	★	★	★		★		
DATAROBOT	★						
DATASATX	★		★		★		
DATAVALUE CONSULTING				★			
DENODO	★	★	★			★	
DIGDASH	★		★		★		
EBIZNEXT				★			
EKIMETRICS	★	★	★		★	★	★
ELITECH	★		★	★	★		
EPSILON	★		★	★	★		
EULIDIA	★	★		★	★	★	★
EXPERT SYSTEM	★	★	★				
FINAXYS				★			
GOOGLE CLOUD						★	
GROUPE ESTIA	★	★	★	★	★		
HEWLETT PACKARD ENTERPRISE	★		★	★		★	
HURENCE	★		★			★	
IDEATRANS	★		★	★	★		
INDEXIMA	★				★		
INFORMATICA	★	★					★
IRI	★	★	★		★	★	

Entreprises & Labels

	 Analytique	 API	 Applications	 Consultants/SSII Intégrateurs	 Dataviz	 Infrastructure	 Institutionnels
IVIDATA	★			★	★		
JEMS				★	★	★	
KEYRUS	★		★	★	★		
LANSROD	★			★			
LINCOLN	★			★	★	★	
MARIADB CORPORATION AB	★						
MATHWORKS	★		★		★		
MICROPOLE	★	★	★	★	★		
MICROSOFT	★	★	★	★	★	★	
MICROSTRATEGY	★		★		★		
MONGODB	★	★	★		★	★	
NE04J	★				★	★	
OPENDATASOFT	★	★			★		
ORACLE	★	★	★	★	★	★	
PwC FRANCE	★		★	★	★		
QLIK	★				★		
REEPART	★	★		★	★		
SAAGIE	★		★		★	★	
SAS	★		★		★		
SEGMENT	★	★				★	
SEMARCHY	★		★		★		
SINEQUA	★						
SKAPÁNÊ	★	★	★	★	★	★	
SNOWFLAKE	★					★	
STIBO SYSTEMS	★		★				
SUADEO	★	★	★	★	★	★	★
TABLEAU	★		★		★		
TALEND						★	
TERADATA FRANCE	★		★	★			
TIBCO SOFTWARE	★	★			★		
TIGERGRAPH	★				★	★	
TIMI	★		★		★		
TRIFACTA	★						
VERTICA	★		★				
YSANCE	★	★		★	★	★	



Ab Initio software is a powerful, flexible platform for building data processing applications and managing enterprise metadata. It offers an integrated architecture for processing data from many sources: files, database tables, HDFS, message queues, web services, and cloud object storage. It processes the data in parallel across multiple servers located on premises or in a cloud environment. Ab Initio applications run in batch or real-time, and they support low-latency event processing. Their distributed architecture allows them to scale easily — for example, to take advantage of elasticity in the cloud.

The world's largest companies choose Ab Initio for their key applications.

Telecom

One of the world's largest phone companies switched to Ab Initio to build a real-time marketing system. Their new system, fully integrated with powerful machine learning technologies, handles tens of millions of subscribers generating billions of events per day and fulfills tens of millions of offers each month. Their competition was left far behind.

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 work together seamlessly, 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 with Ab Initio, and the stock exchange has opened new lines of business providing trading behavior data for stocks.

Why companies choose Ab Initio

Not only is Ab Initio software intuitive and easy to use, but it also stands up to the most complex application logic and to massive quantities of data. And it does this with the unique combination of high performance and remarkable robustness. That's why many of the world's largest companies have chosen Ab Initio to help them make the most out of their data.

Ab Initio offers a complete solution for reading, writing, processing, and querying "big data," wherever it is stored. 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 that is integrated with agile development processes.
- The ability to run applications in parallel, with complete scalability, on heterogeneous clusters of all major platforms; in containerized environments such as Kubernetes, OpenShift, and Docker; and in cloud computing environments, such as Amazon Web Services (AWS), Google Cloud, and Microsoft Azure.
- High performance for both batch and real-time processing
- Support for service-oriented architectures and low-latency event processing.
- Support for and integration with enterprise metadata.

Ab Initio and Big Data

Ab Initio can handle big data with capabilities far beyond those available from other vendors:

- Ab Initio integrates easily with multiple data storage technologies. From Hadoop clusters to traditional databases to cloud storage, Ab Initio reads, processes, and writes data — fast.
- Ab Initio applications can be deployed across multiple platforms from different vendors. From containerized environments such as Kubernetes, OpenShift, and Docker to cloud vendors such as AWS, Google Cloud, and Microsoft Azure, Ab Initio lets you choose the mix of vendors and platforms that best suits your needs.
- In a clustered environment, an Ab Initio application can run as many ways parallel as the data is distributed, maintaining data locality for optimal processing performance.
- Ab Initio integrates with most popular data sources and formats out of the box: Parquet, ORC, Avro, Amazon S3, Google PubSub, Kafka, JMS, MQ, and more.
- Ab Initio jobs can run under the control of resource managers such as YARN or Kubernetes and can be fully monitored.

Federated Queries

Ab Initio software lets you issue federated queries that can include all the databases and files that Ab Initio can access — including Hadoop and cloud data.



Operational Management / Data Monitoring / Data Profiling / Data Quality /
Data Matching / Metadata Management / Data Governance /
Data Masking / Alerting

Business Rules / Entity Management /
Data Visualization / Application Metadata

Batch Processing / Request Response / Complex Event Processing /
Asynchronous Real-Time Processing

In-Memory Processing / Parallel Processing /
Parallel File System Distributed Processing

One System. One Architecture.
Many Applications.

ADVANCED SCHEMA

HERVÉ SEBAN

Président

+33 (0)6 03 22 89 15

hseban@advanced-schema.com

19 rue Galilée 75116 Paris FRANCE

www.advanced-schema.com



advanced
schema



ADVANCED SCHEMA est un incubateur d'offres et de solutions technologiques innovantes dans les domaines de l'Analytique, du Big Data, de l'Intelligence Artificielle et de la Réalité Augmentée ou Virtuelle.

Ainsi, et depuis plus de 17 ans, ADVANCED SCHEMA ne cesse d'accroître et d'enrichir ses expertises au service de ses clients, en se positionnant comme une « end-to-end company », intervenant du conseil à l'hébergement, en passant par le développement d'applications, l'intégration de projets 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 par de forts investissements en R&D, par l'acquisition des sociétés KERNEL 42 et ARFORIA spécialisées dans les domaines du Digital et des nouveaux modes de représentation 3D.

Afin d'apporter souplesse et flexibilité, 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 d'agilité de ses clients,
- Le mode « Tailored » pour un développement à façon d'applications analytiques et Big Data.

Filiale du Groupe ADVANCED SCHEMA, la société NODATA propose **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.



FOCUS SUR LES FILIALES DU GROUPE **ADVANCED SCHEMA**

nodata

Kernel 42

Réinventez l'exploitation de la donnée avec **nodata**, filiale du groupe, entièrement dédiée à l'édition de logiciel.

La suite **Nodata (Not Only Data)** offre la possibilité d'allier innovation technologique, performance et expérience utilisateur pour des coûts maîtrisés. Cette solution complète et totalement modulable couvre aussi bien la transformation de la donnée, la modélisation que la data visualisation...

Kernel 42 étend l'expertise d'ADVANCED SCHEMA aux projets digitaux : La combinaison de ses expertises Web, IA et Data permet de concevoir des **dispositifs digitaux intelligents** et parfaitement intégrés au Système d'Information.



connexin

ARFORIA est expert de **la visualisation 3D** interactive et immersive en réalité augmentée et réalité virtuelle. Au sein d'ADVANCED SCHEMA, la conjugaison de l'exploitation de la donnée et des nouvelles solutions 3D va apporter souplesse et intuitivité dans l'exploration des informations.

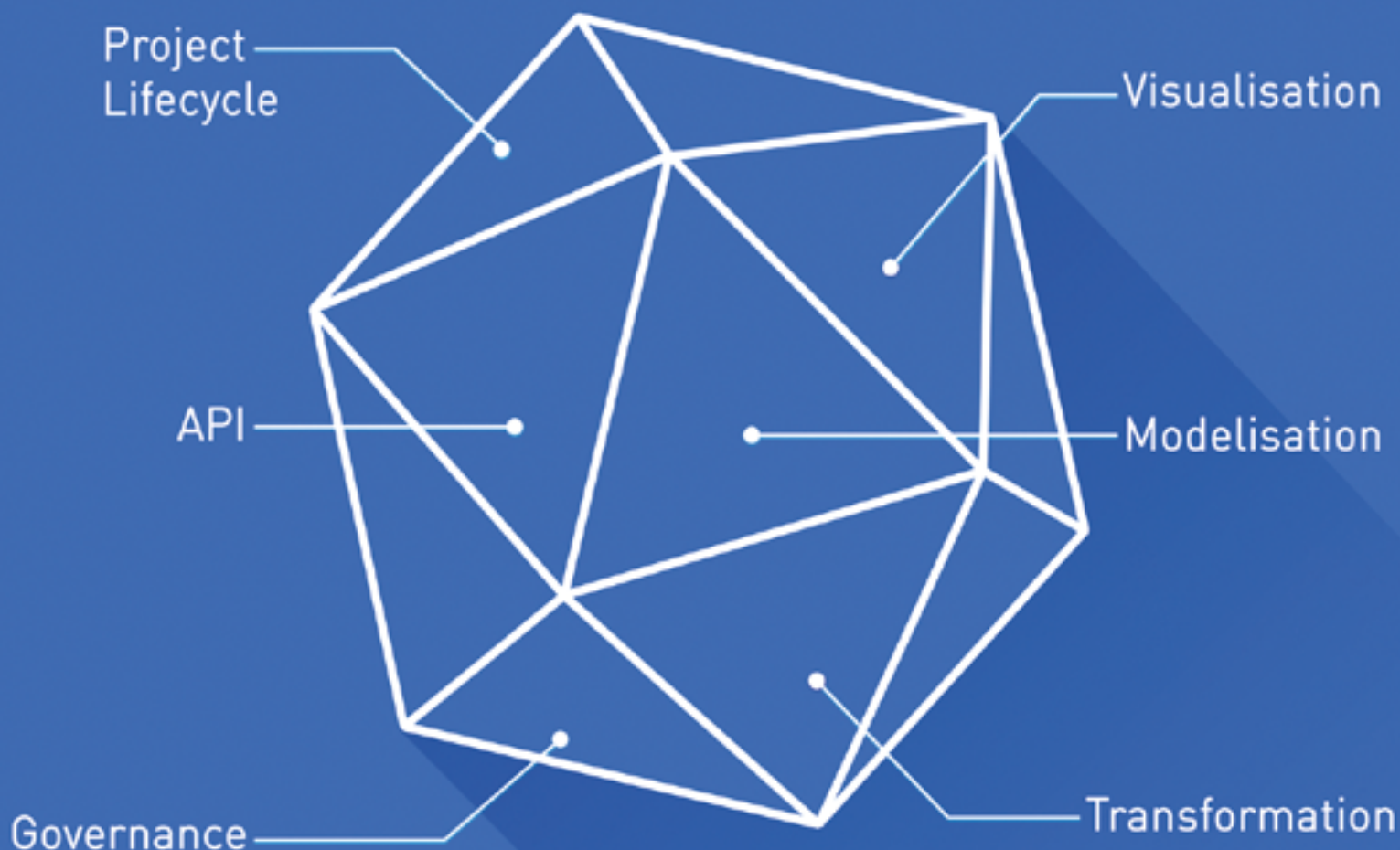
La Data Visualisation Immersive va permettre de simplifier toutes sortes d'opérations complexes dans un objectif de productivité, de limitation des risques d'erreurs, d'apprentissage ou de transmission des connaissances.

Connexin est la filiale « digital Healthcare » du groupe ADVANCED SCHEMA, spécialisée dans le déploiement d'applications connectées en santé et dans la mise en œuvre de socle de données médicales et cliniques.

Connexin permet de **faciliter la communication dans le lien ville-hôpital** entre professionnels de santé en utilisant les **canaux digitaux**.

Révolutionnez l'exploitation de vos données

{not only data}



nodata

AUGMENTED DATA PLATFORM

ALTERYX

RAPHAEL SAVY

Directeur Europe du Sud

+33 6 10 92 18 59

rsavy@alteryx.com

WeWork, 40 rue du Colisée 75008 Paris FRANCE

www.alteryx.com/fr



alteryx



ALTERYX EVERYTHING

Glissez, déposez, terminé

alteryx

The Thrill
of Solving

Venez sur notre stand
A53 pour en savoir plus



Grâce à l'essor technologique et l'augmentation du volume de données générées, tous les secteurs d'activités ont été transformé par le Big Data. : telcos, retail, fitness, banque ou encore énergie. Le Big Data permet d'analyser des données structurées ou non structurées, afin d'améliorer ses services et augmenter ses revenus.

Afin de répondre aux besoins de leurs clients, les entreprises doivent optimiser leurs stratégies et tirer profit des données dont elles disposent. Pour y parvenir, elles optent désormais pour des stratégies data driven, où la qualité de la donnée est aussi importante que la quantité : le Big Data doit devenir le Smart Data.

DATA SCIENTIST : LE MÉTIER DU 21^e SIÈCLE ?

Pour cela, elles recherchent désespérément des experts afin d'évaluer et analyser cet amas inépuisable d'informations. Les offres d'emploi de data scientist fleurissent depuis quelques années. La plateforme Joblift a dévoilé une étude sur le marché de l'emploi des data scientists en France. Après avoir analysé plus de 14 millions d'offres d'emploi sur 2 ans, elle a constaté que près de 5 824 d'entre elles concernaient une recherche de data scientist, soit une hausse mensuelle moyenne de 5 %.

Quelle que soit la quantité de données que possède une entreprise, ces données sont complètement inutiles si personne ne peut les traiter ou les évaluer. Puis, il faut interpréter les résultats obtenus afin de faciliter la prise de décision. Enfin, ces résultats doivent être partagés au reste des équipes. Tel est le rôle du data scientist, d'où son importance grandissante. Pour preuve, le magazine Harvard Business Review l'a reconnu, en 2012, comme étant « l'emploi le plus sexy du 21^e siècle ».

Néanmoins, le travail quotidien de ces scientifiques de la data est loin d'être glamour. Leur journée se résume à des réunions pour : identifier et indiquer quelles données sont pertinentes pour l'entreprise ; sélectionner quels outils utiliser ; déterminer où et comment stocker les données et enfin, concevoir des algorithmes. Enfin, l'entrée en vigueur du RGPD complique encore la donne en donnant le droit aux particuliers de partager, ou non, leurs informations personnelles.

DU BIG DATA VERS LA SMART DATA ?

Le Big Data implique la collecte de données complexes sur 5 dimensions : le volume, la variété, la vélocité, la véracité et enfin la valeur commerciale. Ces dimensions décrivent la quantité et le type de données recueillies, à quel moment, ainsi que leur importance et pertinence par rapport aux enjeux de l'entreprise.

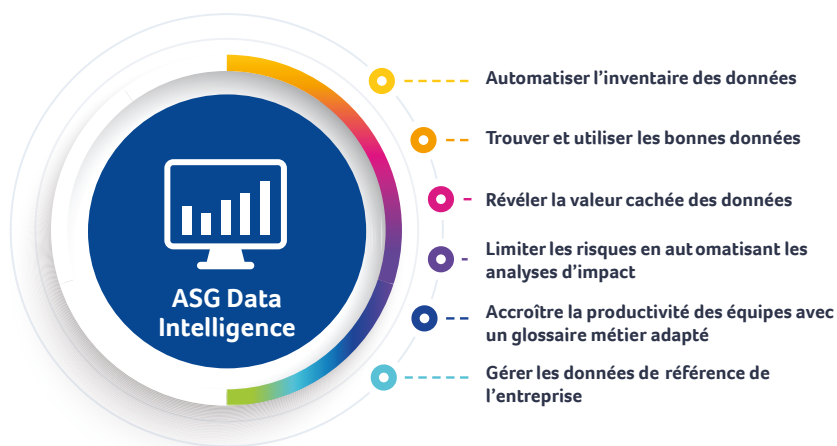
En revanche, le Smart Data va un cran plus loin et consiste à extraire de l'ensemble des données collectées, les informations les plus pertinentes pour l'entreprise afin de pouvoir faire des prédictions. Pour résumer la Smart Data = Big data + sémantique + qualité des données + sécurité + protection des données. Cette formule le fait que le Big data est la matière première du Smart data.

Prenons l'exemple d'une gare. Toutes les données collectées à l'intérieur et autour de la station telles que les informations sur les heures d'arrivée et de départ des trains, les travaux d'entretien ou encore la météo représentent le Big Data. La Smart Data se traduira, par le fait d'anticiper les annulations de trains dues aux conditions météorologiques. Concrètement, cela reviendrait à savoir combien de passagers seraient concernés si un train est annulé à cause de la météo et comment la compagnie ferroviaire peut y faire face. Les data scientists peuvent ainsi déterminer, à l'avance, quelle pourrait être la meilleure solution pour chaque problème.

Une étude du MIT indique que le Big Data confère un avantage compétitif aux grandes entreprises par rapport aux PME/TPE. Ce fossé continuera à se creuser, notamment grâce aux nouvelles technologies, qui permettront aux grands groupes de drainer plus de données. Ils pourront ainsi créer des offres personnalisées, améliorer l'expérience client, proposer une meilleure qualité de service, lutter contre les fraudes ou encore résoudre certains problèmes efficacement et rapidement. Heureusement, le constat n'est pas aussi fataliste dans les faits. Des entreprises telles que Airbnb, Uber ou encore Amazon (start-up à leur début), ont su, grâce à l'innovation technologique et une offre adaptée aux besoins des utilisateurs, se démarquer face aux acteurs traditionnels. À travers ces exemples, on se rend compte que de fantôme il y a encore quelques années, le Big Data est donc devenu un enjeu crucial pour toutes les entreprises.



Êtes-vous en mesure de fournir la bonne information à la bonne personne au bon moment ? Avez-vous confiance dans vos données ? Garantissez-vous la conformité de vos données ?



ASG Data Intelligence propose un ensemble de modules paramétrables pour gérer la conformité des données assurant aux entreprises une agilité sans compromis et une base solide pour rechercher, comprendre, partager, gouverner et analyser les données en toute confiance dans l'ensemble du patrimoine de l'entreprise.

Data Intelligence : Conformité, Reporting, Analytique et Gouvernance des Données

ASG Data Intelligence supporte la conformité des données ainsi que les initiatives de gouvernance tout en améliorant le reporting et l'analyse des données. Avec la solution Data Intelligence, les entreprises peuvent utiliser leurs données en toute confiance et prendre des décisions avisées plus rapidement, les équipes en charge de la conformité peuvent également réduire les risques et garantir la mise en conformité.

Rendez-vous sur le stand A68 pour découvrir comment ASG Data Intelligence permet aux entreprises d'inventorier, de tracer, de gérer et de gouverner leurs données stratégiques tout en restant conformes.

Venez découvrir des cas clients et des bonnes pratiques lors de la présentation ASG prévue le lundi 9 Mars de 13h30 à 14h dans la Salle E.

À PROPOS D'ASG TECHNOLOGIES

ASG Technologies est un éditeur de logiciel d'envergure mondiale qui fournit la seule plateforme intégrée et solution flexible de bout en bout pour permettre aux entreprises de réussir dans l'économie de l'information. Seul fournisseur de solutions destinées à la fois à la gestion de l'information et aux systèmes informatiques, ASG compte plus de 3 500 clients dans le monde entier. Pour en savoir plus, consultez le site www.asg.com.

ASG TECHNOLOGIES RECONNU LEADER DANS LE MAGIC QUADRANT POUR LES SOLUTIONS DE GESTION DES MÉTADONNÉES (OCTOBRE 2019)

Gartner
asg.com/gartnermqdi

GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally, and is used herein with permission. All rights reserved.

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

STAND
A68

#dataintelligence

Know Your Data, Know Your Business



technologies®

www.asg.com/dataintelligence



Chat with your data!

Quelles sont les dépenses par an sur ces fournisseurs ?

Voici l'analyse de la somme des montants des achats par date (année) et par fournisseur.





ASKR.AI, EST UN NOUVEL ACTEUR DISRUPTIF DANS LE DOMAINE DE LA BI/ DATA - POUVEZ-VOUS NOUS EN DIRE PLUS ?

Avec mes deux co-fondateurs, Alexandre Schneider et Joël Célaries, nous évoluons dans le monde de la business intelligence depuis plus de 20 ans maintenant. Nous avons créé à l'époque, la première solution de tableaux de bord 100% web. Il y a trois ans, nous avons fait un constat alarmant qui nous a poussé à vouloir révolutionner les usages.

Je ne vous l'apprends pas : le volume des données en entreprise explose, la donnée devient un véritable "asset" et des millions sont investis chaque année par les organisations pour sa collecte, sa fiabilisation et sa valorisation. Qu'en est-il de la mise à disposition de ces données aux personnes dans l'entreprise qui en ont besoin ? Le réflexe traditionnel est de se tourner vers les outils Business Intelligence ou de Data Discovery. Mais, depuis plus de 10 ans, le taux d'adoption de ces outils de BI plafonne à 22% (30% pour les plus optimistes !). Résultat : l'investissement sur la fiabilisation des données ne profite qu'à un ¼ des personnes qui en ont besoin pour créer de la valeur.

Autre problème : certains collaborateurs, n'ayant qu'un besoin occasionnel de consommation de données, n'ont pas accès à la BI car elle serait complètement sur-dimensionnée pour leur usage.

Avec mes associés, nous nous sommes posés les questions suivantes : Comment adresser tous ces utilisateurs restants ? Comment multiplier par trois ou quatre son ROI sur la donnée ? Et comment soulager tous les experts data, contraints de répondre à des demandes de requêtes métiers "simples" ?

Et si... il suffisait à l'utilisateur de poser tout simplement sa question en langage naturel à un data assistant virtuel, disponible 24/7 et capable de répondre en quelques secondes seulement ? C'est là qu'est né askR.ai.

askR.ai est le premier data assistant doté d'une intelligence artificielle qui permet à des utilisateurs non techniques d'interroger leurs bases de données en langage naturel, à l'écrit ou à l'oral. askR.ai analyse et comprend la question et y répond en quelques secondes en générant à la volée le bon graphique, le bon tableau, ou le bon chiffre. Pas de qualification requise si ce n'est de savoir s'exprimer en langage naturel.

Le data assistant askR.ai est donc une alternative aux outils de business intelligence ?

Je dirais plutôt que c'est une nouvelle façon de faire de la business intelligence pour 80% des utilisateurs pour qui les outils mis en place sont trop complexes.

Depuis deux-trois ans, la maîtrise d'une nouvelle vague technologique (IA, NLP, Machine Learning, Conversational analytics...) ouvre à nouveau le champ des possibles pour rendre la business intelligence encore plus accessible aux utilisateurs métiers et faciliter la vie des experts data. Gartner a baptisé cette nouvelle génération d'acteurs "Augmented Analytics" : un accès à la donnée et aux analyses presque sans effort et immédiat grâce à l'IA, au machine learning ou encore au NLP. Résultat : c'est la donnée qui vient à l'utilisateur et non l'inverse.

Loin d'être au stade d'une simple utopie, ce sont des outils qui sont déjà commercialisés et adoptés par de grandes entreprises (askR.ai, Data Robot...)

Notre ambition : devenir LA plateforme d'Augmented Analytics conversationnelle des utilisateurs métiers.

Qui sont vos clients ?

askR.ai s'adresse aux grands comptes ou ETI pour qui la démocratisation de la data est un enjeu majeur et qui sont prêts à passer à l'Augmented Analytics ! De multiples cas d'usages d'askR.ai sont possibles et notre portefeuille de clients est très varié. On peut citer par exemple : Renault, Technicolor, TF1, Décathlon...

Un exemple d'application à nous donner ?

Nous travaillons par exemple avec le groupe Renault, qui a équipé d'askR.ai son réseau de concessionnaires B2B. Véritable data assistant des chefs des ventes, askR.ai leur permet d'accéder en moins de 3 secondes aux données stratégiques sur leur territoire dans un mode "conversationnel" et ainsi d'optimiser leur prospection. Un succès également pour l'équipe projet : un accès à la donnée multiplié par 7 et un lancement applaudi par les utilisateurs métiers !

Le mot de la fin ?

It's time to chat with your data !



ataccama ONE

Case Study

US Payment Processor

One of the largest US payment processing companies improved their data quality, built a reliable big data reservoir, and uncovered millions of dollars owed to them in unpaid balances.

BUSINESS OBJECTIVE & PROJECT REQUIREMENTS

Due to its strategy of growth by acquisition, the financial services company was composed of several different payment networks (acquired companies), each with their own technical stack and sales processes. The processor aimed to reach a stage where all of its payment networks could access the same data and use standard practices, sales strategies, and technology.

- **Move from data lake to reservoir:** Transition away from data being stored, maintained, accessed, and utilized by individual payment networks, to data being collected, accessible, and usable in a single place. Facilitate and accelerate data quality initiatives.
- **Be able to instantly work with data,** perform analytics, and detect real time patterns.
- **Detect payment fraud:** Although fraud is also handled by banks and payment technology companies like Visa and Mastercard, the company wanted to expand their product offer with services such as fraud detection.

PROJECT PHASES

1. Proof of concept and early work in big data and reference data

Ataccama gave meaning to Hadoop by defining reference values such as payment and error codes, **saving the company \$1 million at the start of cooperation.**

2. Standardization & Enrichment

Data was not consistently organized across various platforms, so Ataccama worked to standardize all company data.

Ataccama also helped the company enrich their data with information that was not originally available, such as addresses, postal codes, and transaction types.

3. Distribution and business user empowerment

Ataccama allowed all company analysts to access their standardized and enriched data in Hadoop. We also enabled them to build small databases on demand, which proved especially beneficial to their marketing department.

SOLUTION & BENEFITS

Ataccama migrated data from former individual companies to a single location with Hadoop, and implemented a data quality solution so that data could be better analyzed, understood, and utilized.

- **Hadoop implementation:** In the first stage, Ataccama helped the company migrate their data from individual companies to a Hadoop cluster.
- **Data profiling and cleansing:** Ataccama made the company's data more useful by implementing a comprehensive data quality solution, using profiling to allow the organization to better understand the contents of its data lake, and cleansing to improve and standardize their data.

Transform your business with AI-Powered Data Curation

Ataccama makes your digital transformation easy. We deliver **AI-powered data curation** with Ataccama ONE. It's a fully integrated yet modular platform for any data, user, domain, or deployment. It gives you the option to **start with what you need** and **seamlessly extend** as your business requires.

Enjoy a single interface to analyze, process, manage, monitor, and provide data.

CLIENT REVIEWS

GARTNER PEER INSIGHTS

Read about our clients' experience with our tools.

Review our products.

→ [GARTNER.COM/REVIEWS](https://www.gartner.com/reviews/ataccama) (ATACCAMA)

GET STARTED

ASK FOR A DEMO

See what we can do for your data project.
Get in touch to see a full demo of Ataccama ONE.

→ [INFO@ATACCAMA.COM](mailto:info@ataccama.com)

READ CLIENT CASE STUDIES

Discover relevant case studies in your vertical.

→ [ATACCAMA.COM/CUSTOMERS](https://ataccama.com/customers)

LOOK FOR US IN THE MAGIC QUADRANTS

Gartner Magic Quadrant for Data Quality Tools

Gartner Magic Quadrant for MDM Solutions

READ HOW MACHINE LEARNING TRANSFORMS DATA MANAGEMENT

→ [ATACCAMA.COM/FREEMIUM](https://ataccama.com/freemium)

ONE-CLICK DATA PROFILING powered by

ataccama **ONE**

Join **55,000+** active users and turn your data into insights.

Free account. No commitments.

→ [ONE.ATACCAMA.COM](https://one.ataccama.com)



ATACCAMA ONE: AI-POWERED DATA CURATION FOR DIGITAL TRANSFORMATION



Data Discovery
& Profiling



Data Catalog
& Business Glossary



Data Quality
Management



Master & Reference
Data Management



Big Data Processing
& Data Integration

Discover more platform modules at ataccama.com

BIRST AN INFOR COMPANY

FRANCOISE CHAUMETTE

Marketing Manager

+33 (0)1 86 99 81 29

francoise.chaumette@infor.com

3 rue Joseph Monier 92500 Rueil-Malmaison FRANCE

www.infor.com



 **birst**
an Infor® company



L'AVENIR CONNECTÉ DE L'ENTREPRISE

Les puissantes évolutions technologiques qui nous touchent en tant que consommateurs modifient fondamentalement la façon dont les personnes et les entreprises interagissent, créant ainsi une période de transformation digitale. Les exemples ne manquent pas. Regardez comment la musique est désormais consommée (Spotify®, Pandora®, Sonos), comment nous gérons nos maisons (Amazon® Alexa, Google® Home, Apple® HomePod), comment les enfants utilisent leurs ordinateurs, tablettes ou smartphone (en guise de passerelle vers l'information, plutôt que comme un contenant pour l'information elle-même), ou comment nous interagissons avec nos amis et les membres de notre famille. Maintenant, imaginez si nous pouvions utiliser ces technologies de façon transparente pour révolutionner notre façon de travailler. C'est ce que la transformation digitale signifie pour les entreprises. C'est l'occasion de définir une stratégie digitale, d'éliminer les obstacles qui limitent l'impact des nouvelles technologies numériques et, en fin de compte, d'offrir une plus grande valeur ajoutée. Nous appelons cela une opportunité. La réalité est que les entreprises ont de moins en moins le choix. L'économie connectée d'aujourd'hui signifie passer au digital ou devenir obsolète, indépendamment de la taille de votre entreprise, secteur d'activité ou localisation. Examinons ensemble comment, grâce à l'innovation, cela impacte les usages Analytics/BI dans l'Entreprise. Le modèle de BI connectée identifie 7 exigences associées aux programmes d'analyse moderne, considérées comme vitales pour leur réussite. Ces exigences réunissent les meilleurs aspects de l'analyse centralisée et décentralisée. Les entreprises qui évaluent les solutions BI de dernières générations doivent rechercher des offres capables de répondre à ces exigences.

1. Unifier les données à partir de multiples systèmes sources ou structures complexes. Les solutions qui prennent en charge un modèle de BI connectée doivent pouvoir unifier rapidement les données à travers une multitude de sources de données hétérogènes et les présenter comme si elles provenaient d'une source unique. Les entreprises doivent comprendre si la solution de BI combine des sources de données physiquement, en déplaçant les données à partir de systèmes sources vers un système cible, ou logiquement, en utilisant des requêtes fédérées pour accéder aux données là où elles se trouvent et livrer des résultats à l'utilisateur comme une source unique.

2. Analyser des processus métier ou des modèles opérationnels complexes. Un modèle de BI connectée reconnaît que différentes personnes peuvent avoir besoin de comprendre différentes parties de l'entreprise en fonction de leur position. Cependant elles doivent être alignées via des instances virtuelles et pas physiques. La solution de BI doit pouvoir fournir des informations sur des processus métier complexes qui couvrent de multiples fonctions à l'intérieur et à l'extérieur de l'entreprise.

3. Mélanger des données centralisées et décentralisées et du contenu analytique. Une BI connectée doit absolument présenter un scope complet de l'entreprise mélangeant données centralisées (globales) et décentralisées (locales). Cela permet l'analyse de BU pour lesquelles les équipes ont besoin de données locales, mais peuvent également exploiter des données globales.

4. Créer et tenir à jour des indicateurs et des définitions réutilisables
La solution est de mettre en place un ensemble de définitions métiers uniques. La BI connectée doit fournir une couche sémantique métier qui lève toute ambiguïté sur la signification d'une dimension métier. Les utilisateurs doivent pouvoir se fier à leurs données et se consacrer à l'analyse au lieu de débattre de leur pertinence.

5. Fournir aux utilisateurs des outils analytiques spécifiques à un poste
La BI connectée rompt avec les paradigmes de la BI traditionnelle de « constructeurs » et « visionneurs » en transformant tous les consommateurs

d'informations en producteurs d'informations. Une solution de BI connectée doit diminuer le gap entre tableaux de bord et découverte, en créant une expérience utilisateur unique adaptée aux modes de travail contemporains. Les utilisateurs verront alors les mêmes données que ce soit, sur un ordinateur, un PC portable ou un smartphone, y compris en mode déconnecté. De plus, une solution de BI connectée doit intégrer les outils que les utilisateurs adorent et connaissent bien comme Excel. Cette approche favorise l'adoption utilisateurs, augmente le ROI des outils tiers et supprime le besoin du département IT jouant le rôle « d'usine à rapports ».

6. Permettre un retour sur investissement rapide avec un déploiement itératif et agile. Le contexte actuel exige un déploiement rapide des projets, suivi d'un processus agile et itératif qui implique les utilisateurs finaux. La BI connectée permet aux entreprises d'être plus agiles en accélérant considérablement le temps d'implémentation et en raccourcissant le temps requis pour apporter de la valeur à l'entreprise. En supprimant un grand nombre de tâches associées aux déploiements traditionnels, la BI connectée donne aux entreprises la possibilité de faire des projets en prenant beaucoup moins de risques et de réagir aux changements plus rapidement.

7. Redimensionner les utilisateurs et les volumes de données selon une logique économique et en fonction de la demande. La mise en place de fonctionnalités analytiques dans l'entreprise nécessite la prise en charge de volumes de données importants ainsi que leur accès simultané. Une solution de BI connectée construite sur une architecture cloud mutualisée (Multi-tenant) doit permettre une évolutivité immédiate dans un environnement complexe. Ainsi, contrairement aux solutions traditionnelles, les performances sont très bonnes. Le temps et les coûts requis pour ajouter de nouveaux utilisateurs et des données au système sont considérablement réduits.

COMMENT SE LANÇER DANS LES ANALYTIQUES ET LA BI CONNECTÉE

La nature agile d'une approche de BI connectée supprime de fait les projets de développement « en cascade » longs et coûteux. Parce que le réseau analytique connecté peut être étendu sans réplique physique, il est tout à fait possible de commencer une implémentation avec un seul cas d'usage. Une fois que ce cas d'usage initial est achevé et qu'il s'avère performant, des cas d'usages supplémentaires peuvent être ajoutés au réseau en exploitant les données et les objets sémantiques existants. Cette approche limite les risques, réduit les coûts et accélère le ROI.

La première étape pour les entreprises qui envisagent une approche de BI connectée consiste à identifier un cas d'usage spécifique qui sert de point de départ. Voici quelques exemples de cas d'usage courants qui peuvent aider les entreprises à démarrer : Analyses commerciales, marketing et financières : Avoir une plus grande visibilité sur les opérations de base (rapidité des opportunités, attribution des campagnes, taux de flux de trésorerie) aide les dirigeants à comprendre les performances de leur entreprise, des ventes jusqu'aux résultats financiers. - Adopter une stratégie « Cloud First » : Alors que les solutions On Prem traditionnelles ne répondent plus aux besoins, les Entreprises se tournent de plus en plus vers des solutions Analytiques déployées dans le Cloud pour bénéficier de la mutualisation. Cette alternative moderne et agile augmente le ROI significativement et réduit considérablement le coût de possession de la solution. - Migrer à partir d'une BI traditionnelle : Le coût de maintenance d'une BI traditionnelle est supérieur au coût induit par un changement. Les dsi recherchent des solutions modernes afin de supprimer les goulets d'étranglement des rapports centralisés, de libérer des ressources et d'offrir aux utilisateurs finaux décentralisés, une autonomie véritable. - Intégrer Hadoop dans votre stratégie de données : Les entreprises qui exploitent la flexibilité de Hadoop ont besoin d'une architecture BI qui sépare le stockage de données et le traitement des données.



Smart BI & analytics for the business user



Connected



Smart



Trusted



Enterprise

Meet us at the stand A11
Try Birst today: birst.com/trial



BLACK TIGER

QUENTIN LEIGNÉ

Responsable Marketing

+33 (0)1 46 12 24 99

leigne.quentin@blacktiger.tech

86 rue Henri Farman 92130 Issy-Les-Moulineaux FRANCE

www.blacktiger.tech



Black Tiger



DÉPLOYER UNE STRATÉGIE DATA-DRIVEN

Si les entreprises sont conscientes de l'importance croissante de privilégier un modèle data-driven, le déploiement sur le terrain est souvent plus complexe.

Présentez-nous Black Tiger

Nous sommes l'éditeur de 2i, l'unique plateforme Big Data, end-to-end, spécialisée dans la gestion des données personnelles, garantissant une conformité RGPD continue, et intégrant les expertises métiers de nos clients. Nous travaillons avec les entreprises pour qui la donnée personnelle est centrale dans leurs activités. Notre plateforme leur garantit la certification des données en continu ; la conformité au RGPD et la mise en œuvre d'un modèle data-driven adapté à leur métier. En d'autres termes, elle permet de centraliser la donnée, de la nettoyer, de la certifier, de la structurer et de l'exploiter pour enrichir le business.

Aujourd'hui, les entreprises peinent à mettre en place une stratégie leur permettant de suivre les transformations numériques de leurs secteurs économiques respectifs, dans des contextes d'évolution rapide et face une concurrence de pure player, toujours plus présents.

Pourquoi est-ce si difficile selon vous ?

La présence de pure-players numériques dans les métiers traditionnels oblige les marques à se réinventer toujours plus vite. Le modèle data-driven s'impose alors comme le pivot de ces transformations numériques. Au cours des cinq dernières années, pour y parvenir, certaines entreprises ont expérimenté les Data Labs avec des technologies open-source du Big Data. Cela a conduit à mettre en place des environnements fragmentés, toujours plus complexes à maintenir et à exploiter, utilisant de nombreuses technologies indépendantes et d'importantes ressources.

Il est clair aujourd'hui que le ROI attendu des Data Lakes n'est pas au rendez-vous. Ces entreprises se retrouvent aujourd'hui face à des impasses de déploiement pour industrialisation. De plus, ces technologies adressent essentiellement les dimensions Volume et Vitesse du BigData, laissant de côté la Variété et la Véracité des données, qui sont critiques pour l'exploitation des données personnelles. Ainsi, elles ne peuvent atteindre le 6ème V, la Valeur.

À quel niveau intervenez-vous donc ?

Black Tiger intervient dans toute entreprise dont la maturité à la transformation numérique est assez élevée. Comme je l'ai dit précédemment, beaucoup d'entreprises ont lancé leur transformation numérique à partir de mises à disposition technologiques des Data Labs. Deux ans plus tard les résultats d'industrialisation de ces chantiers ne sont pas atteints et ne pourront pas l'être. Nous intervenons pour industrialiser et globaliser cette transformation numérique.

Contextualisons un peu. Lors d'un déploiement, nous sommes confrontés à des contextes économiques et applicatifs existants. Dans toute entreprise, il a y déjà une multitude de solutions récurrentes du marché dans des verticales métiers (CRM, ERP, marketing...). Notre constat est que ces applicatifs ont été pensés pour gérer la donnée et non la traiter. En amont de cela, nous trouvons une multitude de sources de données éparses (applications internes, externes, fournisseurs...) qui finissent dans ces appli-

catifs métiers. Il faut prendre conscience que nous ne pouvons prétendre à construire une stratégie data-driven et donc réaliser sa transformation numérique sans globaliser l'ensemble des flux de l'entreprise autour des 3 convictions suivantes: qualité de la donnée, ethic-data et adaptabilité. 2i est la seule solution actuelle du marché adressant l'exploitation de tous les flux de données d'une entreprise en une seule techno intégrant trois savoir-faire cruciaux, Data Quality, RGPD et Règles Métiers.



Scalabilité, adaptabilité, quelle différence faites-vous ?

La scalabilité est un enjeu technologique. L'adaptabilité est un enjeu de modèle économique.

Les technologies que nous avons évoquées via les Data Labs ciblent principalement des problématiques de scalabilité, avec des packages de CPU ou de RAM qui devaient permettre aux entreprises de « s'adapter ».

Là où le bât blesse, c'est que cet enjeu occulte la véritable problématique, celle de disposer de plateformes et d'infrastructures toujours plus agiles dans leurs capacités à répondre rapidement aux nouveaux besoins des clients. La transformation numérique vous oblige à vous adapter constamment et ce, dans des délais de plus en plus restreints. Si votre plateforme n'est pas performante sur ce point, vous ne répondez pas aux enjeux du marché du numérique actuel.

Le deuxième volet de l'adaptabilité, c'est notre capacité à implémenter des modules métier ad hoc pour tous les secteurs et métiers.



Quelle solution pragmatique apportez-vous donc à ce besoin d'adaptabilité ?

De façon pragmatique, le déploiement de la plateforme 2i se fonde sur la mise en place d'un modèle conceptuel de données, transversal à l'entreprise, qui soit adapté aux métiers et aux enjeux économiques du secteur de nos clients. Chacun de nos clients bénéficie donc de ses modèles conceptuels et physiques propres. Ces MCD et MPD sont construits en mettant en leur centre la donnée personnelle de l'entreprise.

C'est cette construction qui, à elle seule, est garante d'une entreprise data-driven.

Les entreprises du marché cherchent à évoluer vers cette orientation data-driven.

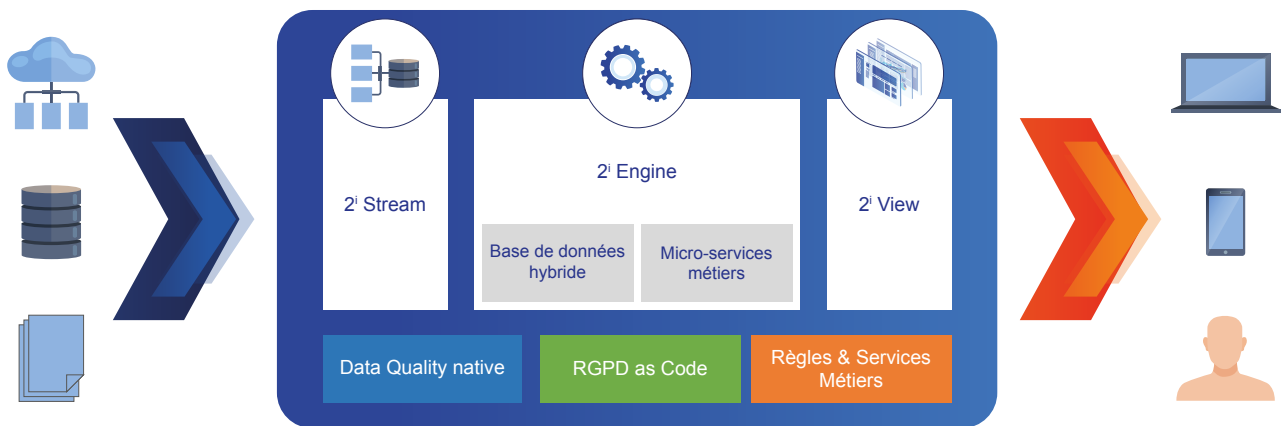
Cela suggère une entreprise qui définit sa stratégie, son organisation et ses process internes en corrélant systématiquement ses services et produits, avec l'ensemble des données personnelles de ses clients et prospects. Nous pourrions remplacer cette expression par « organisation et culture d'entreprise centrées autour du couple produit-individu ou service-individu ».

Et d'un point de vue technologique ?

Notre job, c'est de donner les moyens à nos clients de valoriser leur data-asset via une base centralisée, ouverte et customer-centric. 2i unifie la totalité des données d'une entreprise et les rend exploitables par tout service pour tout objectif métier, grâce à nos trois modules technologiques qui sont : Data Quality, règles métiers et RGPD.

Ils garantissent la certification de la donnée, une adaptabilité totale au modèle économique de l'entreprise ainsi qu'à ses enjeux business, et finalement, une conformité en continu au RGPD dans toute la chaîne d'exploitation.

La plateforme est le reflet de nos convictions et porte notre vision de ce que doivent être les environnements Big Data de demain.



Que retenir-vous de l'édition 2019 du Salon Big Data Paris ?

Particulièrement enrichissante, cette édition initie un véritable virage vers une industrialisation du Big Data en entreprise. Le salon a été, pour nous, l'occasion de rencontrer des entreprises qui doivent faire face aux problématiques que nous adressons. Il y a aujourd'hui une prise de conscience des problèmes liés aux Data Labs et au Cloud public (dérives de planning des projets exploratoires, apports insuffisants aux business units, effet de locking par les providers de Cloud, ou encore faille de confidentialité liée au Cloud Act américain).

Les entreprises veulent maintenant passer à une étape d'industrialisation dans l'exploitation des données personnelles.

J'en retiens aussi l'omniprésence de l'IA qui s'alimente des data-assets des entreprises. Cela représente pour nous une opportunité immense, car l'IA suppose l'automatisation et la fiabilisation des données en amont, ce qui est notre cœur de métier.

CLARANET

CLÉMENT MARY

Directeur des opérations

+33 (0)1 44 78 63 65

clement.mary@fr.clara.net

134 rue des Templiers 59000 Lille FRANCE

www.claranet.fr

Analytique
API
Applications
Consultants/SSII
Intégrateurs
Dataviz
Infrastructure
Institutionnels

claranet

helping **our customers**
do amazing things



claranet | BIGDATA

Infogérance, gouvernance et sécurité de votre plateforme



Experts DataOps



Experts DevOps



Experts SecOps



COMMENT CONCILIER L'EXTERNALISATION DE PLATEFORME BIG DATA DANS LE CLOUD AVEC LA SÉCURITÉ ET LA CONFORMITÉ ?

La sécurité constitue un des critères principaux à la réussite d'un projet big data. Sans elle, des incidents désastreux peuvent impacter les entreprises et leurs données.

Claranet propose une approche d'externalisation des plateformes vers le cloud en assurant un niveau de sécurité adapté à la spécificité des données.

Externaliser sa plateforme big data dans le cloud répond aux enjeux techniques de disponibilité, de performance et d'évolutivité des entreprises. La data se place effet au cœur de leurs enjeux, car elle représente un levier essentiel pour réaliser des analyses cognitives, prédictives et approfondies.

La donnée devient le facteur déterminant de la plupart des projets big data des entreprises.

Plus de 80 % des projets actuellement gérés par Claranet comportent de la donnée personnelle. Les plateformes Big Data doivent donc répondre à des normes de sécurité dans le cadre de la réglementation RGPD.

La typologie de données influent également sur les critères de sécurité des plateformes. C'est notamment le cas pour les données bancaires et la norme PCI DSS, ou les données de santé avec l'agrément HDS.

Qu'elle entre dans une typologie spécifique ou qu'elle soit fondamentale pour la stratégie des groupes, la data nécessite un niveau optimal de sécurité.

Comment mettre en place une stratégie de sécurité ?

Bien trop souvent les entreprises lancent des projets de big data en mode POC ou test en interne, sans placer la sécurité au cœur du dispositif.

Autre constat : le risque de vouloir intégrer le cloud trop vite dans ces projets peut avoir des impacts cruciaux pour les entreprises. Elles doivent donc définir en amont une stratégie précise autour de la data et la sécurité associée, portée par des profils comme le CDO (Chief Data Officer), le CIL (Correspondant Informatique et Libertés) ou encore le CISO (Chief Information Security Officer).

Car maîtriser la sécurité des données ne s'improvise pas. Pour obtenir les certifications nécessaires pour les héberger, les processus sont lourds et nécessitent d'être organisés, d'avoir l'expertise et l'expérience, ce que les entreprises n'ont pas forcément, leur cœur de métier étant ailleurs.

Dans ce contexte, les équipes de Claranet proposent la mise en place d'une approche d'accompagnement, basée sur l'externalisation des plateformes big data vers le cloud, avec un niveau de sécurité adéquat et adapté à la nature des données à préserver. Le tout en respectant évidemment les politiques de sécurité des entreprises.

Quelles sont les étapes de la sécurisation ?

Cette démarche se concrétise en trois étapes : au niveau de la plateforme cloud, du cluster et des outils applicatifs, et des services d'accompagnement.

Il convient d'abord de s'assurer de la sécurisation des plateformes chez les cloud providers : interconnexion sécurisée et filtrage réseau, durcissement des systèmes d'exploitation (OS), stratégie d'authentification forte, mises à jour système et applicatives régulières, veille en vulnérabilités...

Au niveau applicatif, ensuite, il faut travailler avec des solutions matures, fournissant toutes les fonctionnalités pour s'assurer de la sécurité de la donnée. La distribution Cloudera, solution partenaire et privilégiée de Claranet pour son offre big data, propose notamment des mécanismes de chiffrement de la donnée, des outils de traçabilité et des réponses aux normes et certifications.

Enfin, un accompagnement par des experts de la sécurité et un ensemble de services de sécurité pour la gestion à long terme du dossier client, comme des scans de vulnérabilité, des tests de pénétration ou un processus strict de veille en vulnérabilités et de patch management (gestion des correctifs logiciels), sont proposés par les équipes Claranet.



BYLINE PAR ROMAIN PICARD, REGIONAL VICE PRESIDENT, EMEA SOUTH



Concevoir la bonne infrastructure pour le Big Data

Qui dit digitalisation, dit données... données et Big Data. Bien que la plupart des entreprises exploitent déjà le potentiel d'analyse du Big Data, la technologie et l'infrastructure restent souvent sources de difficultés. Une complexité renforcée par la multiplicité des types de données, qui peuvent être structurées ou non, générées soit par des humains, soit par des machines ou des capteurs. Pour pouvoir correctement exploiter ces différentes données afin de prendre des décisions éclairées, les entreprises ont besoin d'une architecture Big Data qui leur fournisse des informations fiables.

L'importance des données n'a cessé de croître au cours des 20 dernières années. Les problématiques de stockage, de gestion, d'analyse et d'exploitation des données sont ainsi devenues de plus en plus prioritaires, non seulement pour les services informatiques, mais également pour les équipes dirigeantes, qui ont compris leur importance. Cette nécessité de collecte permanente a entraîné une croissance exponentielle des volumes, avec des architectures construites essentiellement autour de vastes lacs de données. Tout a changé avec l'essor du cloud et de ses services toujours plus fiables et plus puissants : les données ne sont plus collectées de façon centralisée mais de plus en plus éparpillées à des emplacements différents. Du fait de cette dispersion sur plusieurs sites (et parfois même entre entreprises), ces données sont désormais constamment en mouvement, que ce soit à l'intérieur ou à l'extérieur des data centers, ou même dans le cloud.

Structurées, non structurées et réparties... des données de différents types

Autre difficulté, différents types de données (structurées ou non structurées, issues de capteurs, de fournisseurs, de clients ou d'employés) peuvent se trouver dans chacun des clusters de données. Et à l'inverse, des données associées peuvent être réparties et stockées à des emplace-

ments différents. Un commerçant peut par exemple stocker les grilles tarifaires de ses produits dans un data center à Bangkok et les informations de ses clients dans un second à Dublin. Par ailleurs, certaines formes de données provenant de capteurs IoT sont plus faciles à stocker sur un système cloud tel qu'Amazon Web Services (AWS) ou Microsoft Azure. Ce type de données pourrait également être conservé « sur site » dans les systèmes de stockage internes à la société. Ces dernières ont besoin d'une plateforme de données complète intégrant toutes les données des centres et des environnements publics, privés, hybrides et multi-clouds. Cette plateforme pourrait également proposer d'autres services, tels que des consignes de protection et de conformité concernant ces dernières, à différents endroits.

L'essence des données

D'un point de vue technique, le choix d'une plateforme appropriée représente une étape importante dans l'analyse des données collectées par les entreprises, mais cela est loin d'être suffisant. Selon Gartner, « 87% des entreprises ont une faible maturité en matière d'analyse de données ». Il est en effet important que l'entreprise sélectionne les experts appropriés et compétents pour examiner ses données à différentes étapes de leur cycle de traitement. Ces derniers doivent pouvoir appliquer les règles à tous les data centers. Pour que les analyses basées sur les données entrantes aient du sens, il est impératif que la qualité de ces informations reste toujours élevée à toutes les étapes de l'analyse.

Enterprise data cloud

Tout cela doit être pris en compte lors de la conception d'une plateforme Big Data, à savoir un enterprise data cloud. Les organisations en transformation digitale ont impérativement besoin d'un système analytique moderne qui couvre les environnements publics, privés, hybrides et multi-clouds. En effet, elles ont besoin de l'agilité, de la flexibilité et de la facilité d'utilisation fournies par les infrastructures cloud, mais souhaitent également exécuter des charges de traitement analytiques où bon leur semble, quel que soit le lieu de stockage de leurs données. Elles veulent également des architectures ouvertes et la possibilité de déplacer ces flux de travail vers différents environnements cloud (publics ou privés). Elles veulent également des architectures ouvertes et la possibilité de déplacer ces flux de travail vers différents environnements cloud (publics ou privés). Enfin, elles souhaitent avoir la possibilité d'exécuter plusieurs fonctions d'analyse sur le même ensemble de données dans un cadre de sécurité et de gouvernance commune garantissant la confidentialité des données et la conformité aux diverses réglementations.

A retenir : 4 caractéristiques essentielles d'un enterprise data cloud

- 1) Hybride et multi-cloud :** prenant en charge l'ensemble des principaux clouds publics et privés
- 2) Multifonction :** pour rationaliser les difficultés majeures en matière de données et d'analyse
- 3) Sécurité et gouvernance :** pour satisfaire les exigences de sécurité et de conformité
- 4) Ouvert :** impliquant l'emploi de logiciels open source, d'architectures de calcul ouvertes et de banques de données comme Amazon S3 et Azure Data Lake Storage. Cela aide à bénéficier de cycles d'innovation plus courts et d'un net avantage concurrentiel



**WE LOVE WHEN
CLOUDS PLAY
NICE TOGETHER.**

WITH LOVE, CLOUDERA

We love data. Everything about it. Its chaos and complexity. Its structure and scale. Every packet and petabyte. We love making sense of it all. And making sure the data in your private, public, hybrid, and multi-clouds work together. Securely. We're not just open source. We're also open for business anytime you need us. So let's get started.
With Love, Cloudera

[CLOUDERA.COM/WITHLOVE](https://cloudera.com/withlove)

CLOUDERA
The Enterprise Data Cloud Company



Une intelligence artificielle pour optimiser le traitement des messages entrants

Suite à l'essor des technologies connectées et la multiplication des canaux de communication, les entreprises doivent aujourd'hui faire face à l'explosion du nombre de données générées, collectées et exploitées.

Pour s'adapter à ce nouveau contexte et tirer profit du Big Data, elles doivent faire évoluer leurs services. Cela s'applique notamment à la gestion des messages entrants. Qu'il s'agisse de demandes de contact par email, de réclamations ou encore de messages sur les réseaux sociaux, il est nécessaire de pouvoir analyser et traiter ces messages, quelle que soit leur volumétrie.

De plus, un type de message pouvant passer virtuellement par n'importe quel canal, l'enjeu n'est plus uniquement de gérer les différents canaux en silos mais également de trouver la bonne orientation du message, c'est-à-dire, de détecter l'urgence et la nature de la demande.

Face à ce nouveau paradigme, l'humain ne suffit plus à traiter la quantité des messages entrants. Text Mining et Machine Learning deviennent donc des atouts dans le traitement automatisé des messages.

Optimiser le traitement des messages entrants grâce au Text Mining et au Machine Learning

Afin de répondre à cette problématique, Coheris s'est appuyé sur le Text Mining et le Machine learning afin d'optimiser le traitement d'importants volumes de messages entrants dans un CRM. Comment cela fonctionne ?

D'une part, le Text Mining va permettre de rendre le texte intelligible par les algorithmes de classement qui pourront, par la suite, associer une catégorie à un texte à partir de caractéristiques structurées. C'est-à-dire, qu'ils pourront classer les messages pour les envoyer aux bons services afin que ces derniers puissent adopter la réponse appropriée en fonction de la nature de la demande.

D'autre part, grâce au Machine Learning, il sera possible de programmer des règles pour apprendre à classer. Cela veut dire que l'algorithme se fondera sur des exemples et apprendra de lui-même.

La classification automatique

Pour ce faire, la machine reçoit, tout d'abord, les premiers messages qui ont été préalablement catégorisés et qui lui permettent d'apprendre les règles de catégorisation. Par la suite, dès l'entrée de nouveaux messages, la machine sera en mesure de classer, elle-même, les messages entrants selon les catégories définies dans les règles d'apprentissage.

La classification automatique par Coheris, dans des catégories préalablement déterminées, s'intègre directement dans le CRM et facilite la gestion des messages par les différents services. En effet, elle permet d'accélérer le traitement des demandes clients qui sont remontées dans le CRM, qu'il s'agisse d'emails ou de messages sur les réseaux sociaux comme Facebook ou Twitter. Cette classification automatique permet de pré-renseigner le CRM. Les messages issus de Facebook y seront intégrés et classés par l'algorithme selon différentes catégories définies en amont comme « Réclamation » ; « Demande » ; « Satisfaction » par exemple. De la même manière, les emails entrants seront automatiquement classés selon ces catégories et également intégrés dans le CRM.

De plus, le classement de chaque message pourra être validé ou modifié par l'humain. La nouvelle catégorisation humaine sera par la suite absorbée par la machine pour être intégrée à sa règle de classement. C'est à l'humain qu'il reviendra de valider les classements afin de faire évoluer la machine, le but étant de constamment faire progresser son modèle grâce à une véritable collaboration humain-machine. Le classement automatique des messages dans la bonne catégorie, permet ainsi aux services clients de répondre plus rapidement aux clients et d'accroître leur satisfaction.

Enfin, bien qu'il soit de plus en plus difficile pour l'humain de maîtriser les messages face à la multiplication des canaux de contact, l'intelligence artificielle ne supprime pas l'intervention humaine. En effet, l'intelligence artificielle, et plus particulièrement le Machine Learning, sont de réels atouts dans l'optimisation du traitement des messages entrants grâce à l'automatisation de leur classement par les algorithmes de Machine Learning. Un des éléments différenciant de cette méthode est l'implémentation de règles évolutives qui permettent à l'algorithme d'apprendre seul. Les entreprises sont ainsi plus à même de traiter d'importants volumes de données et les différents services gagnent du temps dans le traitement des messages.

A propos de Coheris

Editeur de référence de solutions **CRM, d'Intelligence Artificielle et Analytics**, à la fois sur site ou en cloud, Coheris place la donnée au cœur de la stratégie des entreprises et les accompagne dans les domaines de la Gestion de la Relation Client, de la Data Intelligence et de la Gouvernance des Données.

Pour répondre à chaque besoin, Coheris propose des modules packagés orientés métier à valeur ajoutée « IA & Analytics » : **RGPD Text-Control**, gouvernance et classification automatique des messages entrants, moteur de recommandation en temps réel.

De la PME à la multinationale, plus de 1 000 entreprises et collectivités font confiance à Coheris et à ses solutions multilingues dans plus de 80 pays. S'appuyant sur un réseau de partenaires étendu et sur ses propres experts, Coheris apporte une proposition de valeur unique sur le marché, permettant de capitaliser sur une connaissance client précise, de disposer des bons indicateurs pour décider et piloter une stratégie performante.

Pour plus d'informations : www.coheris.com ; www.ia-data--analytcs.fr ; www.donnees-rgpd.fr

COLLIBRA

NATASHA LAUER

Director EMEA Marketing

+33 (0)1 55 68 11 33

natasha.lauer@collibra.com

Spaces La Défense, Le Belvédère 1-7 cours Valmy Puteaux
92800 Paris FRANCE

www.collibra.com

★

Analytique

★

API

★

Applications

★

Consultants/SSII
Intégrateurs

★

DataViz

★

Infrastructure

★

Institutionnels



Collibra



Data Intelligence : **Permet** l'Autonomie des Acteurs , **Transforme** les Processus, **Impulse** le Progrès.



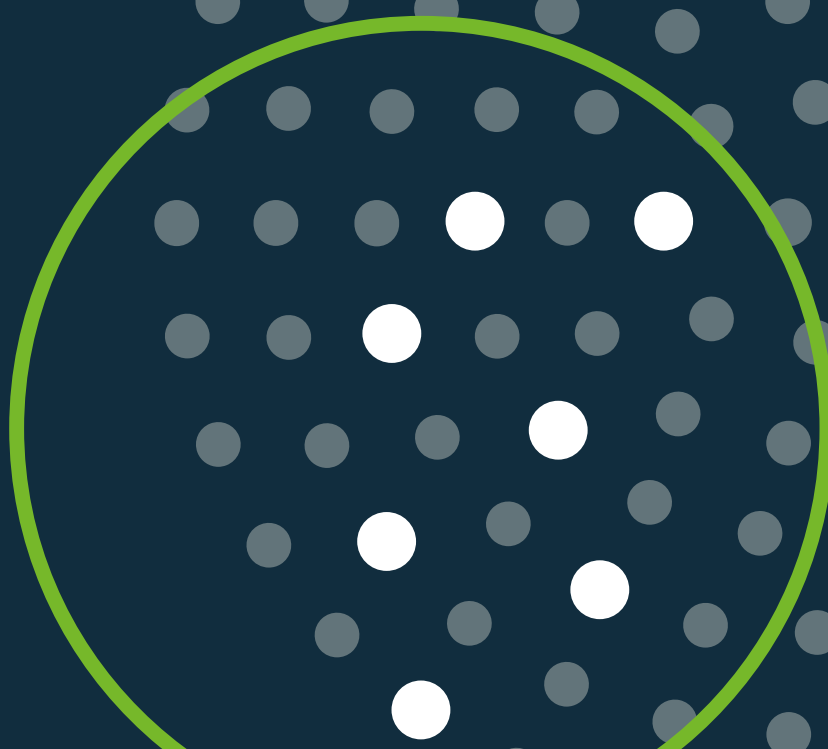
Collibra

Rendez-vous sur Collibra.com

Collibra est le spécialiste de la Data Intelligence. Les données, informations et algorithmes sont mis à la portée de tous les acteurs afin d'accélérer l'impact commercial. Grâce à notre plateforme Cloud, l'IT et le métier peuvent collaborer pour renforcer une culture Data qui favorise l'entreprise digitale. Des organisations du monde entier ont choisi Collibra pour libérer le potentiel de leurs données et en faire un avantage concurrentiel stratégique.

Donnez du sens à vos **données.**

Venez rencontrer Collibra sur
le **stand A19** pour en savoir plus.





Euronext mise sur Confluent pour propulser sa plateforme de trading événementielle

VUE D'ENSEMBLE

Première bourse paneuropéenne, Euronext exploite des marchés réglementés de valeurs mobilières et de produits dérivés à Amsterdam, Bruxelles, Lisbonne et Paris, ainsi qu'un marché réglementé des titres en Irlande et au Royaume-Uni. Euronext a récemment développé une nouvelle plateforme de négociation événementielle, Optiq®, qui décuple la capacité de négociation et offre une latence moyenne de performance de 15 microsecondes pour les ordres aller-retour et les données du marché.

DÉFI

Dans l'architecture et la construction d'Optiq, l'équipe d'Euronext a mis l'accent sur la fiabilité et l'évolutivité, mais également la performance. "Nous avons besoin de la plateforme de streaming pour pouvoir ingérer jusqu'à un million de commandes par seconde, avec une latence en temps réel de l'ordre de la milliseconde. C'est pourquoi nous avons commencé investiguer Kafka", ajoute Pujalte.

SOLUTION

Après avoir évalué plusieurs alternatives, Euronext a choisi Confluent Platform et Apache Kafka® pour la couche de persistance de la plateforme de trading multi-marchés Optiq.

L'équipe de développement d'Euronext a commencé par un prototype de validation de principe, qui a démontré sa capacité à répondre aux exigences globales de haute performance de la plateforme.

"Grâce à notre premier benchmark avec Kafka, nous avons constaté que la capacité à ingérer des messages jusqu'à un taux d'un million par seconde était facilement atteinte", dit Pujalte. Forte de ce succès initial, l'équipe a entrepris un projet en plusieurs phases afin de construire la plateforme de production. Première étape importante: le remplacement de la couche passerelle de données de marché de l'infrastructure de négociation existante par une nouvelle version basée sur Kafka. Cette passerelle traite plusieurs milliards de messages par jour, envoyant des données de marché aux vendeurs, ainsi qu'aux membres d'Euronext qui utilisent ces informations dans leurs stratégies de courtage.

"La nouvelle passerelle de données du marché fait passer une quantité massive de données par Kafka", note M. Pujalte. "Ce projet a confirmé que Kafka était le bon choix et était notre première

étape pour apprendre à utiliser Kafka dans des conditions réelles de production."

Optiq s'appuie sur une couche de persistance construite à l'aide de Confluent Platform, qui lui fournit une infrastructure de streaming fiable et évolutive, prenant en charge les latences en millisecondes sans perte de messages. "Lorsque nous avons commencé, Kafka était pour nous une nouvelle technologie que nous avons décidé d'utiliser pour une application très critique de notre système", déclare Philippe Planchon, Architecte et Directeur des Solutions de Trading Innovantes chez Euronext. "Avec Confluent, nous nous sommes sentis soutenus dans notre décision et nous savions que nous avions le bon niveau d'expertise pour nous préparer et nous aider si nous rencontrions des problèmes. C'était un élément clé de notre succès."

Euronext étant au cœur de l'économie européenne, la fiabilité était impérative. "Nous construisions une plate-forme stratégique pour soutenir la capitalisation boursière de six pays européens", déclare Alain Courbebaisse, Chief Information Officer d'Euronext. "En tant qu'organisation, nous sommes également très ouverts d'esprit, c'était donc l'un de nos critères pour ce projet également. Et avec tous nos efforts en faveur de l'open source, nous voulons nous assurer d'avoir le soutien de l'entreprise."

Durant cette mise en service, l'équipe développait déjà d'autres composants, y compris le moteur de production correspondant et les passerelles membres. À cette époque, l'équipe a commencé à travailler avec les ingénieurs de Confluent pour peaufiner certains aspects de leur implémentation et assurer une performance constante sous une charge extrêmement lourde. "Avoir le soutien de Confluent a été un réel avantage pour nous. Confluent nous a aidés à comprendre comment peaufiner notre stack et améliorer nos performances", explique Pujalte.

En continuant d'étendre la couche de persistance, l'équipe a développé de nombreuses applications clientes Kafka, y compris des applications qui surveillent la latence du marché, transfèrent les données vers une base de données SQL en support des opérations du marché, effectuent la réplication des données pour la reprise après incident et entreposent les enregistrements dans un datacenter conformément aux exigences réglementaires.

Euronext exploite désormais la plateforme de négociation Optiq, avec sa couche de persistance construite sur Confluent Platform, en production sur tous ses marchés.

RÉSULTATS

- Fiabilité des opérations 24 heures sur 24, 5 jours sur 7, réalisées et maintenues.
- Exigences de performance dépassées.
- Un soutien spécialisé et dédié reçu.

DÉCOUVREZ EURONEXT
WWW.EURONEXT.COM/



Complete Event Streaming Platform

Powered by

Apache Kafka[®]

Available on:





Marriott: Creating a modern customer experience for the digital age

OVERVIEW

To maintain its competitive edge in the digital economy, Marriott wanted to create personalized customer experiences, improve online reliability, and release new apps faster. After an evaluation against MongoDB™ and Cassandra, Marriott chose Couchbase to replace its legacy infrastructure. Couchbase had already proven itself in the industry, and Marriott's solutions architects were impressed by Couchbase's built-in cache, ease and flexibility for moving and adding cluster nodes, and easy disaster recovery.

"For many many years, we said, 'Wouldn't it be nice to have a data store where we could go from the Java object right into the database and back without a big translation and lots of overhead?' Well, this is it." says Thomas Vidnovic, Solutions Architect at Marriott International.

About Couchbase

Unlike other NoSQL databases, Couchbase provides an enterprise-class, multicloud to edge database that offers the robust capabilities required for business-critical applications on a highly scalable and available platform. Couchbase is built on open standards, combining the best of NoSQL with the power and familiarity of SQL, to simplify the transition from mainframe and relational databases.

Couchbase has become pervasive in our everyday lives; our customers include industry leaders Amadeus, American Express, Carrefour, Cisco, Comcast/Sky, Disney, eBay, LinkedIn, Marriott, Tesco, Tommy Hilfiger, United, Verizon, as well as hundreds of other household names. For more information, visit www.couchbase.com.

PR Contact: Agence Ballou // Léonard Facchino - Mickael Barreateau
+33 1 42 22 24 10 // leonard.facchino@balloupr.com - mickael@balloupr.com

CHALLENGES

- On a single central mainframe, an outage could cost millions in lost booking revenue
- Maintaining the reservation system and updating the mainframe was very costly
- Wanted to move to a cloud-forward strategy, use commodity hardware, and lower licensing costs
- Needed to deploy new applications faster and more reliably

OUTCOMES

- Data replicated to multiple geographic areas optimizes response times and improves availability
- A scalable, flexible cloud-based model reduces application development costs and improves speed
- Developers use SQL for JSON to deliver personalized customer experiences
- Couchbase supports 30 million documents, accessed at 4,000 transactions per second

ENTERPRISE-CLASS MULTICLOUD TO EDGE NoSQL DATABASE

The only database that combines the best of NoSQL with the power and familiarity of SQL, all in a single, elegant platform spanning from any cloud to the edge.



Couchbase

NOEQUAL



DATAFAB : LA FABRIQUE DE VOS PROJETS DATA

NOTRE POSITIONNEMENT

DataFab conseille et accompagne la révolution digitale au travers d'une approche centrée sur les données.

Elle développe son expertise pour accompagner les directions techniques et métiers dans la mise en œuvre de projets stratégiques, qui mettent la donnée au cœur de l'organisation et des business models.

CHIFFRES
CLÉS



50 collaborateurs



4M€ de CA



Paris & Luxembourg

DataFab est une équipe de "doers" (faiseurs) qui fait bénéficier ses clients de son expérience projets data à grande échelle et industrialisés. Elle active les données pour améliorer les services et accélérer la transformation numérique de ses clients.

NOS PARTENAIRES



CLOUDERA



DATABRICKS

NOTRE ESPRIT & NOS VALEURS

L'organisation agile de DataFab, basée sur le développement de l'autonomie et de l'intelligence collective, permet à chacun d'atteindre son plein potentiel pour accomplir ensemble les objectifs communs, dans le respect des valeurs qui nous rassemblent :



CHALLENGE & CRÉATIVITÉ

Parce qu'on adore les nouvelles aventures, inventer notre propre futur, et que **c'est l'essence de l'existence de tenter de nouvelles expériences** et d'apprendre aux travers d'elles.



LIBERTÉ

D'entreprendre, d'apprendre, de progresser. Chez Datafab **chacun est libre d'approfondir son expertise dans la direction qu'il souhaite** contribuer à l'évolution de l'entreprise et comme il le souhaite.



ÉTHIQUE

La data mais pas n'importe comment, **dans le respect des règles** avec une part de notre activité consacrée à des projets éthiques et/ou à but non lucratifs.

DATAFAB

CETTE ANNÉE PLUS DE 20 NOUVEAUX FAB SE JOINDRONT À NOTRE AVENTURE, EN FEREZ-VOUS PARTIE ?

NOS MÉTIERS



DATA SCIENTIST



DATA ARCHITECT



DATA ENGINEER



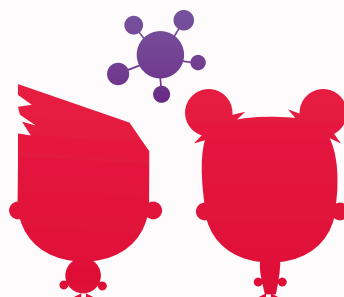
DEVOPS & CLOUD



ARTIFICIAL INTELLIGENCE



NOS TECHNOLOGIES



#WEAREDOERS

Rejoignez-nous
recrutement@datafab.io

datafab.io





De la donnée brute à l'impact business

Dataiku développe une plateforme collaborative qui facilite et accélère l'analyse de données et le développement de services prédictifs en environnement Big Data. Elle permet l'adoption massive des analytics à l'échelle d'une entreprise en donnant accès au self-services analytics et à l'opérationnalisation. Notre collaboration avec les entreprises à travers le monde valide le principe de cette approche, grâce à l'impact business dont elle témoigne. En effet, l'adoption massive de nouvelles technologies avancées est plus puissante que leurs adoptions précoces par une faible minorité.

Cette plateforme unique réunit à la fois les experts data et les métiers, fournit des référentiels des meilleurs modèles et offre des outils innovants pour la gestion et le déploiement du Machine Learning et de l'IA. Avec cet environnement centralisé et contrôlé, Dataiku est un véritable catalyseur pour les entreprises qui veulent se repenser autour de la donnée.

Les entreprises du Retail, E-commerce, Santé, Finance, Transport, Secteur Public, Industrie, pharmaceutique utilisent Dataiku pour fournir à leurs équipes des outils de dernière génération d'analytique et d'analyse prédictive et garantir les conditions de déploiement des modèles les plus matures. En favorisant la collaboration, Dataiku offre une multitude de possibilités pour à la fois créer des modèles ayant un impact visible et rapide mais aussi construire les expertises Data de demain.

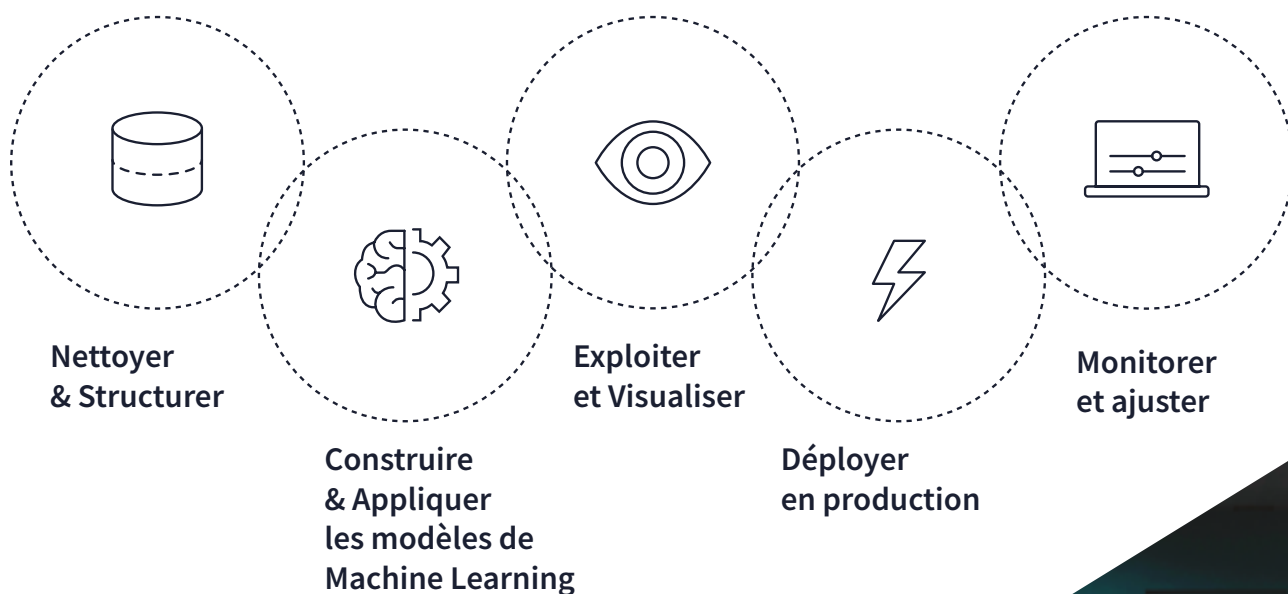
Dataiku fournit la technologie qui permet aux entreprises de:

- Découvrir et adopter rapidement les analytics, la data science, le machine learning et l'IA à l'échelle de votre entreprise
- D'intégrer et de travailler avec les dernières technologies pour un travail productif de la donnée à l'échelle de l'entreprise
- Développer et cloner rapidement vos projets data tout en mettant en production le bon modèle (mieux qu'une multitude d'excellents modèles dans une sandbox)

Aujourd'hui, Dataiku est l'acteur incontournable des plus grandes entreprises data-driven de ce monde, apparaissant comme une plateforme centralisée pour les analytics et le déploiement des modèles de machine learning. En 2017, Dataiku a doublé de taille et triplé ses revenus conduisant à une levée de fond en Septembre 2017 de 28 millions de dollars en série B mené par Battery Ventures. En décembre 2018, Dataiku annonce une quatrième levée de fond en Série C de 101 millions de dollars mené par Iconiq Capital.

En route vers l'Enterprise AI

Dataiku la plateforme d'Analytique et de Data Science centralisée pour les experts data et les métiers.





Devenez une entreprise AI-Driven avec le Machine Learning Automatisé

Devenez une entreprise AI-Driven en implémentant des modèles de Machine Learning Automatisé afin d'adresser davantage de cas d'usage, identifier de nouvelles opportunités commerciales et améliorer considérablement la productivité.

Avec DataRobot, les entreprises peuvent :

- Accélérer la prise de décision — la résolution de problématiques urgentes se fait plus rapidement
- Augmenter la productivité des Data Scientists — réaliser plus de projets en moins de temps
- Démocratiser la Data Science — renforcer votre équipe existante en déployant l'IA

Pour de plus amples informations, veuillez consulter datarobot.com



DataRobot

Découvrez la plateforme DataRobot en action !



"DataRobot est intégré à l'ensemble de notre processus de prise de décision et de nos opérations, fournissant des prévisions et des résultats précis."

— Directeur Innovation et Données,
Carrefour

Rejoignez-nous sur **le stand #A56 à Big Data Paris** pour découvrir comment des entreprises comme Carrefour ont adopté l'intelligence artificielle et tirent parti de la plateforme DataRobot.



DataRobot

Réservez dès à présent
votre rendez-vous !



UNE BASE DE DONNEES OPERATIONNELLE MODERNE

Construite sur Apache Cassandra™ et conçue dès le départ pour fonctionner dans tous les environnements, dans tous types de clouds, dans tous types de data centers avec toutes les combinaisons possibles.

DataStax fournit la base de données cloud hybride et multi-cloud ultime. DataStax aide les entreprises à faire face à la concurrence dans un monde en mutation rapide, où les attentes sont élevées et où de nouvelles innovations apparaissent chaque jour. DataStax est un partenaire expérimenté dans les déploiements d'applications on premise, dans le cloud hybride et multi-cloud qui propose une suite de solutions de gestion distribuée des données et de services cloud. Nous aidons les entreprises à proposer des applications révolutionnaires pour devancer votre concurrence.

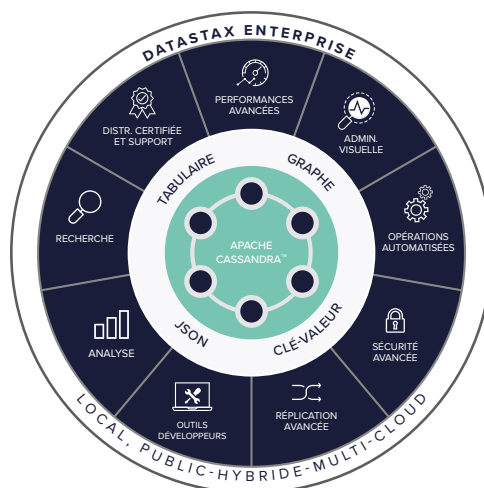
400 des plus grandes marques mondiales - Bouygues Telecom, Capital One, Cisco, Comcast, Delta Airlines, eBay, Kapsh, Macy's, Metro, McDonald's, RCI Banque, Safeway, Sony, Walmart ... - utilisent des solutions DataStax pour créer des applications modernes pouvant être déployées dans tous types de clouds.

ADAPTEE A VOS TOPS INITIATIVES

Chaque entreprise est mise au défi avec un ou plusieurs des projets suivants.

- Architectures Cloud et Cloud Hybride
- Modernisation de Base de Données
- Développement d'Applications On Line

DataStax peut vous aider à faire évoluer votre stratégie de base de données dans chacune d'entre elles afin de vous assurer que vous utilisez la technologie de base de données moderne pour répondre aux exigences des applications actuelles.



A PROPOS DE DATASTAX

DataStax, permet de profiter d'une base de données cloud distribuée et toujours disponible, reposant sur Apache Cassandra™ et conçue pour le cloud hybride. Le déploiement d'applications personnalisées, en temps réel et à grande échelle repose sur notre produit phare DataStax Enterprise. Il permet aux entreprises d'exploiter plus facilement les environnements de multi cloud et cloud hybride et confère également aux entreprises une autonomie totale en matière de données, ce qui leur permet de garder le contrôle et la propriété stratégique de leur actif le plus précieux dans un monde cloud hybride.

DataStax, dont le siège social est établi à San Mateo en Californie, est soutenue par les plus grands investisseurs du secteur notamment Lightspeed Venture Partners, Meritech Capital et Crosslink Capital.

Pour plus d'informations, visitez le site DataStax.com

Join DataStax at

Big Data Paris 2020

Palais Des Congrès

Mars 9-10

Hall Passy • Zone A • Stand 51





Data Specialist par passion

#CONSEIL #DIGITAL #EPM #BUSINESS INTELLIGENCE #BIG DATA

DataValue Consulting vous apporte plus qu'une expertise : nous vous délivrons un accompagnement personnalisé sur toute la chaîne de valeur de la data.

+200
consultants

12,5%
du CA investis
en R&D chaque
année

16,5M€
de CA 2019



75
certifications

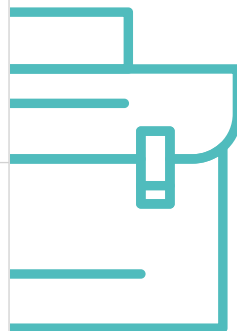


100%
de croissance
chaque année
depuis la création

un
Lab Innovation



3
filiales



54%
des projets
réalisés
au forfait

Ils nous font confiance



...

NOUS TRANSFORMONS VOTRE
DATA EN PERFORMANCE PURE,

AVEC **TALENT**
ET **PASSION**

DataValue Consulting est un acteur indépendant du Conseil,
spécialisé dans la chaîne de valeur des données.

Data Specialist par passion.

www.datavalue-consulting.com



DataValue[®]
CONSULTING

STAND
C13



Migration vers le cloud : pourquoi s'appuyer sur la virtualisation des données ?

Selon une récente étude¹, à l'heure où 36% des entreprises déclarent avoir amorcé leur migration cloud, 20% seulement estiment avoir atteint les phases avancées de leur implémentation. Si l'adoption du cloud continue de prendre de la vitesse, elle reste soumise à un certain nombre de défis comme la sécurité, la gestion et le suivi des dépenses cloud, ou encore le manque de compétences cloud. Les entreprises éprouvent en effet des difficultés à migrer leurs ressources d'information critique compte tenu de la complexité du Big Data.

La virtualisation des données peut permettre aux entreprises de surmonter cet obstacle en développant une architecture de services de données qui les aidera à tirer le maximum de valeur de leurs données et à tirer pleinement parti de la modernisation du cloud, de l'analytique et des configurations hybrides. En associant différents types et sources de données cloud et sur site, de nombreuses organisations optent pour des solutions de virtualisation des données pour bénéficier de l'agilité et de la flexibilité caractéristiques du cloud tout en permettant à leurs équipes d'exploiter les données issues de ces environnements hybrides.

Migration des données

Lors d'une migration vers le cloud, passer d'une infrastructure physique à une infrastructure virtuelle représente un défi pour les DSI. Comment doivent-ils s'y prendre pour migrer les informations, faire fonctionner leurs systèmes on-premise aussi efficacement et s'assurer que les différents systèmes et applications qui se connectent à des API s'y connecteront toujours une fois la migration vers le cloud effectuée ?

La virtualisation des données leur permet de répondre à ces problématiques. Parce qu'elle génère une vue d'ensemble logique sur toutes les données d'entreprise, sans avoir à enregistrer les informations dans un référentiel physique, elle permet de gérer la connectivité entre les différentes sources de données. Pour simplifier, la virtualisation des données gère la couche de virtualisation ainsi que la performance des différentes sources. Par exemple, si une couche de virtualisation des données est déplacée dans le cloud, il sera inutile de reconfigurer les applications. Il suffira simplement de reconfigurer la couche virtuelle. La virtualisation des données permet donc aux entreprises de gagner du temps, de réaliser des économies et de moins solliciter leurs ressources.

Réconcilier le business avec l'IT

Pour les entreprises, parvenir à retrouver les données dont elles ont besoin au cœur du Big Data peut être un véritable parcours du combattant. Il existe aujourd'hui trois façons d'adresser la donnée : au travers d'un moteur de requête, au travers de services de données, de services web ou d'API, ou

encore au travers d'un Data Catalog, le « supermarché » des sources de données et des vues métiers pour les utilisateurs et les data scientists.

Du point de vue de la sécurité, puisqu'elle facilite la visualisation de tous les consommateurs et la protection des sources de données, la virtualisation des données permet de mettre en place des règles de gouvernance tout en continuant de servir les consommateurs.

Flux et traitement des données

Une fois les données obtenues, comment s'assurer de leur fiabilité et en tirer le meilleur parti ? Plusieurs étapes sont à prévoir : la Data Quality, qui permet de gérer et de comparer des données entre différents systèmes d'information ou bases de données, la Data Science, qui permet de tirer des enseignements des ensembles de données, le Machine Learning, qui produit les résultats des traitements, et enfin la Data Visualisation, qui permet de représenter visuellement les données brutes.

Cette étape fonctionnelle peut être très difficile à gérer mais peut être simplifiée grâce à la virtualisation des données. La couche de virtualisation des données permet en effet de se connecter aux différentes sources de données pour mieux les organiser, les contrôler et maintenir un niveau de performance adéquat. Enfin, elle apporte les capacités de processing aux sources de données qui n'en ont pas (un fichier Excel, par exemple).

La virtualisation des données permet aux DSI de simplifier la migration vers le cloud. Elle apporte une touche de collaboration qui permet de simplifier la mise en place des infrastructures en fournissant une infrastructure virtuelle. Elle facilite par ailleurs le traitement des données tout en maintenant la performance et en ne dupliquant pas les données. Peu coûteuse, elle tire le meilleur parti des investissements existants et permet de mieux les utiliser.

À propos de Denodo

Denodo est le leader de la virtualisation des données - une technologie agile et de haute performance d'intégration, d'abstraction et de services de données en temps réel. La virtualisation des données permet un accès unifié à un large éventail de sources de données (d'entreprise, big data, cloud et non structurées), à un coût inférieur de moitié à celui de l'intégration de données traditionnelle. Les clients de Denodo issus de tous secteurs ont gagné en agilité métier et augmenté leur retour sur investissement de manière significative en permettant un accès simplifié, plus rapide et unifié aux données d'entreprise pour les besoins en BI agile, big data analytics, intégration web et cloud, applications de vue unique et services de données d'entreprise. Fondée en 1999, Denodo est une société privée, solidement financée et rentable.

Pour plus d'informations, visitez www.denodo.com/fr

Une plateforme d'intégration et de livraison de données en temps réel sans réplication



Réduction des coûts de **80%**



Réduction du temps d'intégration de **85%**



ROI atteint en **3 mois**



Répond aux exigences de **sécurité/ confidentialité**

Single View of Customer



Logical Data Warehouse / Lake



Cloud Modernization



Data Services



Agile BI



Testez Denodo sur le Cloud

Avec Denodo Test Drive, faites vous-même l'expérience des bénéfices de la Data Virtualization pour la BI agile et l'analytique.

www.denodo.com/TestDrive

Pour en savoir plus, visitez denodo.com ou contactez-nous au +33 (1) 73 06 26 75





ÉLARGISSEZ VOS HORIZONS GRÂCE À VOS DONNÉES



Fondée en 2006 par des experts en BI ayant fait leurs classes chez Business Objects dans la Silicon Valley, DigDash est aujourd'hui la PME française qui défie les géants américains du tableau de bord et de l'analytique.

Soucieux d'offrir à ses clients l'un des **meilleurs outils décisionnels du marché**, DigDash a créé un logiciel de tableaux de bord analytiques à l'image des équipes : adaptable, rapide, intuitif et puissant.

AGILE, EASY and STRONG sont d'ailleurs les maîtres-mots de ce spécialiste de l'Analytics et de la Dataviz.



LA SOLUTION DIGDASH ENTERPRISE RÉPOND À 3 BESOINS FONDAMENTAUX DES ENTREPRISES



PILOTER : en permettant aux décideurs d'accéder à tous les indicateurs de performance de manière synthétique et intelligible.



EXPLORER : les utilisateurs peuvent parcourir les données de leur métier pour en tirer des indicateurs pertinents.



COMMUNIQUER : grâce aux fonctionnalités de Data Storytelling, d'export ou de partage, la communication et la circulation des informations est optimisée.

Intégrable dans l'infrastructure existante, DigDash Enterprise est particulièrement adapté aux milieux sensibles de la banque ou de l'industrie.

La solution s'est également imposée **dans le secteur public**, et compte aujourd'hui parmi ses clients français **plus de 15 départements, 6 ministères, 4 grandes villes..**

Le retail, la logistique, les télécommunications et le secteur de la santé plébiscitent également DigDash comme outil décisionnel.



LE MOT D'ANTOINE BUAT, CEO DE DIGDASH

« Notre valeur ajoutée vient de notre capacité à faire **évoluer notre logiciel** vers une solution toujours centrée sur les besoins de nos clients, en particulier sur les besoins des utilisateurs métiers pour piloter leur activité. Ceci passe à travers un **accompagnement personnalisé** et un processus de **formation continue** à nos outils. »

NOTRE EXPERTISE SECTORIELLE

RETAIL



Du fournisseur au client final... La BI au service du Retail

Générez des **tableaux de bord analytiques pour piloter les KPI**, au niveau groupe jusqu'au niveau du magasin (fidélité/satisfaction client, stocks, ventes/paniers), pour optimiser toute la **chaîne de valeur Retail**.

« Les tableaux de bord générés aujourd'hui nous donnent une visibilité sur le chiffre d'affaires, la marge, les statistiques par magasin, par marque, par vendeur, le nombre d'unités vendues... Tous ces indicateurs permettent d'affiner le pilotage de l'activité de l'entreprise et d'analyser les résultats. »

SECTEUR PUBLIC



Centraliser les outils pour mieux piloter les budgets des collectivités territoriales

Grâce à ses tableaux de bords dynamiques, DigDash permet aux administrations de **maîtriser leur budget**, de **piloter l'allocation des ressources** financières et humaines, et de **monitorer la performance** des services et des politiques publiques.

« Ce qui plaît aux utilisateurs ce sont ses multiples fonctionnalités : drill down, vision globale, exploration... Il permet d'aller de la vision globale au détail. Le fait de pouvoir réaliser des exports PowerPoint, Excel ou de programmer des emails ... »

INDUSTRIE



Améliorer les performances de la chaîne logistique pour maximiser la productivité

Avec DigDash, disposez **en temps réel** d'analyses pertinentes vous permettant **d'anticiper et prendre des décisions éclairées dans les temps et les coûts impartis**.

« DigDash Enterprise peut nativement analyser des données structurées et non structurées et gérer des volumes de données très importants : dans ce cas, un traitement de 5 milliards de lignes avec des restitutions quasi instantanées. »

FINANCE



Acquérir une vision synthétique pour répondre aux exigences métier

Générez intuitivement des **tableaux de bord clients** (profils, services personnalisés), optimisez vos activités (réduction des files d'attente, amélioration des actions marketing en agence...) et produisez un **ensemble de reporting** liés à ce secteur réglementé.

« DigDash a su nous démontrer à la fois la simplicité d'intégration des données dans un système d'information déjà en place. Ainsi que la simplicité de mise en œuvre car les délais étaient courts. En effet, la notion de tableaux de bord simples et dynamiques se fait ressentir dans tous les départements de l'entreprise. »



EBIZNEXT, ACTEUR DE RÉFÉRENCE DANS LES ARCHITECTURES ON-PREMISE, CLOUD HYBRIDE ET MULTI-CLOUD

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 » sur le Cloud, 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 socles technologiques à la fois modernes et éprouvés : Spark / Kafka / Elasticsearch / GCP / AWS / Azure.

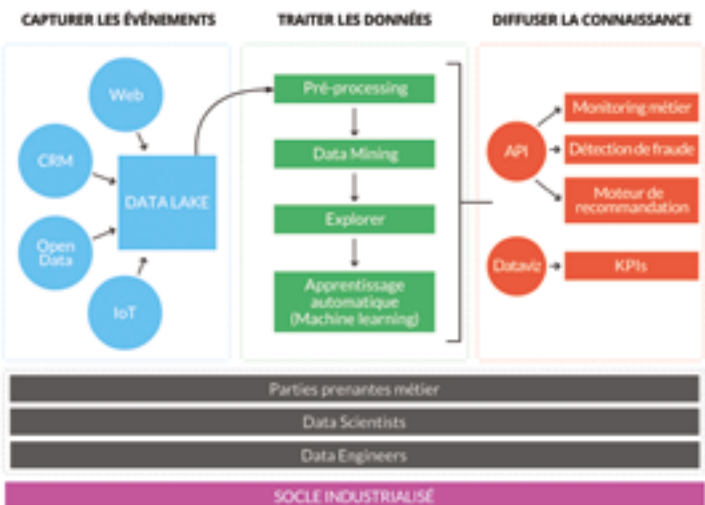
À titre d'exemple : ebiznext a conçu et mis en place des Datalake industrialisés aptes à répondre à des cas d'usages nécessitant des données capturées, traitées et diffusées en temps réel.

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.

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 / GCS / S3.

Diffusion de la connaissance (Portail utilisateurs)

Les informations métiers sont ensuite diffusées via une API de présentation ou un outil de DataViz.

Sécurité des données :

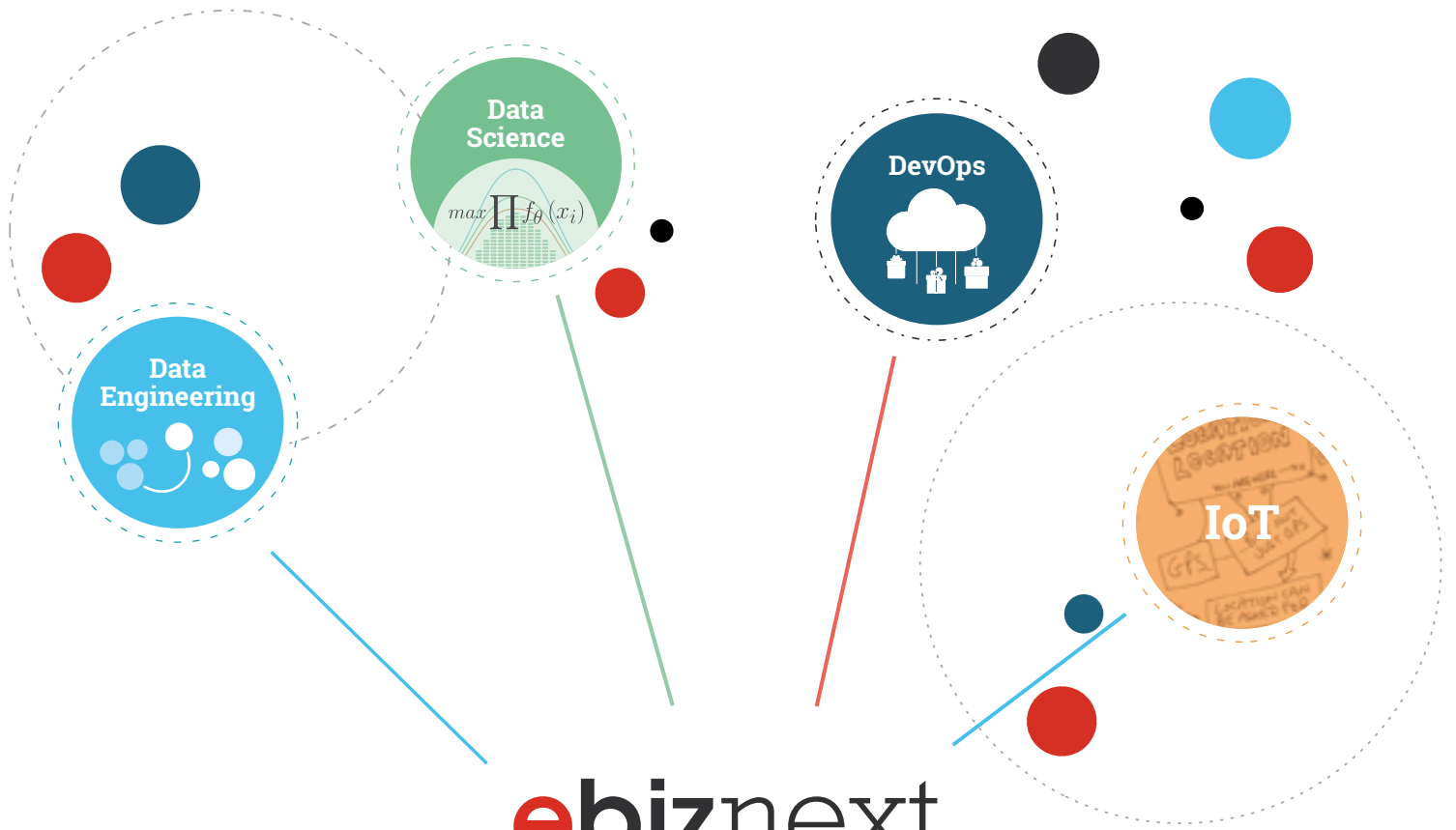
Mise en place de sécurité périmétrique et sélective dans une approche On-premise, Cloud hybride et MultiCloud grâce à nos experts certifiés sur les plateformes Cloud GCP / AWS et Azure..

CONTRIBUTIONS EBIZNEXT

Parallèlement à la réalisation de projets fast data, les consultants ebiznext animent différents types d'événements :

- talks et workshops (ScalaIO, Devox, PSUG, JUG, Data Marketing Paris...),
- meetups intra-entreprises et hackathons (thématiques : IoT, Machine Learning...),
- formations qualifiantes (Centrale Paris, DSTI, ENSAI, ESGI...) et certifiantes (Lightbend...).

ebiznext apporte par ailleurs des contributions majeures sur blog.ebiznext.com et github.com/ebiznext.



ebiznext

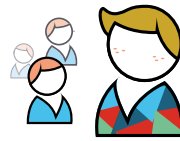
Votre Datalake à la carte



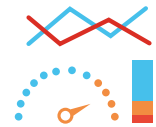
Fast & Smart Data



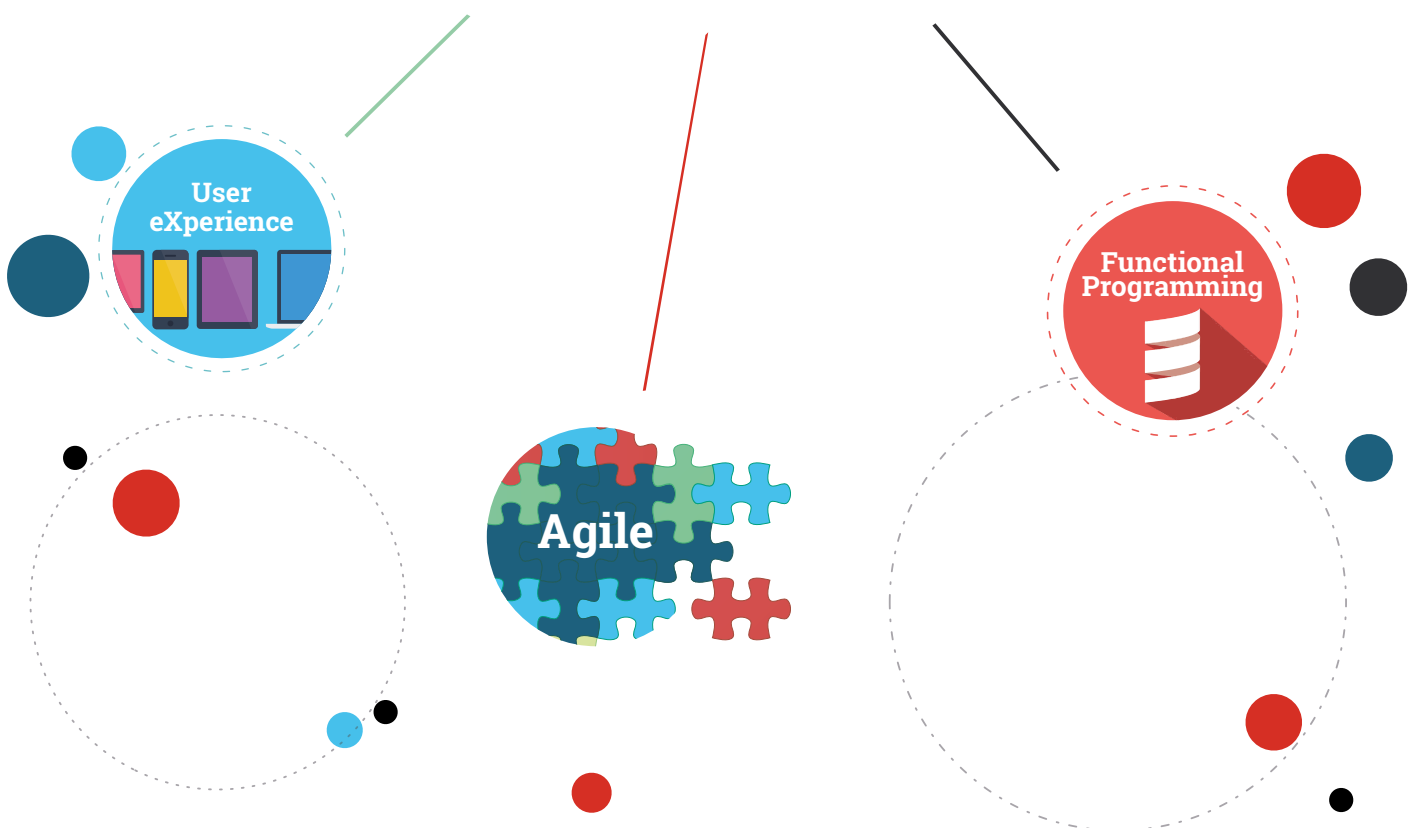
On Premise & Cloud



Marketing comportemental



Monitoring métier





NOTRE METIER : VOUS ACCOMPAGNER DANS LA CONSTRUCTION DE NOUVELLES CAPACITES DATA & ANALYTIQUES

Ekimetrics est une société de conseil internationale fondée en 2006, pionnière et leader européen en Data Science.

Avec +280 consultants data scientists basés à Paris, Londres, New York, Hong Kong et Dubaï, nous menons des projets dans +50 pays, dans des secteurs aussi variés que Banque et Assurance, Automobile, Cosmétiques, Tourisme, Luxe, Retail, Telecom, etc.

L'ambition d'Ekimetrics est d'aider les entreprises à faire de la Data Science le moteur de leur stratégie. Nous vous accompagnons dans la construction de nouvelles capacités data et analytiques, pour vous permettre de démultiplier la performance de votre stratégie de revenu, ou de faire évoluer votre business model, grâce à vos données. Pour ce faire, nous utilisons des approches statistiques et algorithmes sophistiqués (Machine Learning, Intelligence Artificielle, etc.), que nous mettons systématiquement au service des métiers.

TRANSFORMEZ VOS DONNEES EN VALEUR METIER

Notre objectif est de transformer vos données en nouvel actif d'entreprise pérenne, qui soutiendra votre stratégie à court et à long terme. C'est pourquoi nous avons structuré notre offre en 2 piliers, pouvant être traités séparément, ou conjointement comme un tout, au bénéfice de vos objectifs business.

• DOPEZ VOTRE STRATEGIE DE REVENU

Transformez votre stratégie marketing et commerciale en nouvel avantage concurrentiel

Pour les CMOs, CEOs, Directions Expérience Client et CRM

- Modélisation du marketing mix (MMM)
- Multi-Touch Attribution (MTA)
- Customer Analytics et promesse relationnelle data-driven
- Stratégie Trade, Prix, Optimisation de la Promotion
- Suivi et mesure d'impact sur la marque

• DEVELOPPEZ VOS CAPACITES ANALYTIQUES

Faites évoluer votre business model grâce à vos données

Pour les CEOs, COOs et CFOs

- Transformation et stratégie data
- Roadmap data
- Process, gouvernance data et infrastructure IT
- Implémentation et acculturation

POURQUOI EKIMETRICS ?

• **Une approche intégrée de la performance et une volonté de réaffirmer la valeur de la compréhension des métiers**, au cœur selon nous, de la transformation des données en nouvel actif business. Nous considérons vos projets bien au-delà des aspects data : les impacts financiers, l'image de marque, votre culture d'entreprise, vos processus de décision,

votre maturité data... Nous accompagnons la conduite du changement imposée par la data, et aidons votre organisation à s'emparer de ces nouveaux enjeux sur le plan technique, mais aussi humain et organisationnel.

• **Une librairie de benchmarks et frameworks** : forts de nos 13 années d'expérience en Data Science, avec plus de 350 clients dans tous les secteurs d'activité, nous avons cumulé des bonnes pratiques et méthodes propriétaires qui nous permettent de développer des solutions en quelques mois : frameworks propriétaires, comparatifs sectoriels, grilles de ROI et canevas stratégiques qui nous rendent ultra efficaces de la stratégie à la mise en œuvre, tout en restant réalistes.

• **Une flexibilité adaptable à votre secteur** : nous ne croyons pas à la standardisation systématique, aux approches uniques pour tous les clients. Notre force réside dans notre capacité à saisir les nuances de votre business model et à construire les modèles statistiques basés sur les spécificités de votre entreprise : c'est comme cela qu'on obtient les meilleurs résultats en quelques mois. Nos plateformes de productivité nous permettent d'aller vite, notre approche métier d'être bien plus spécifiques.

NOS CONSULTANTS... SONT TOUS DATA SCIENTISTS. ET VICE VERSA.

Il faut des profils d'exception pour résoudre les problèmes les plus complexes. C'est pourquoi nous tenons à ce que nos consultants développent à la fois des compétences avancées en conseil stratégique, ainsi qu'une excellence analytique. Nous recrutons des candidats issus des plus grandes écoles (X, Telecom, HEC...), et formés selon des méthodes propriétaires (plus de 70 formations internes).

Nous avons la conviction que cette double casquette consultant / data scientist, unique sur le marché, est le secret des résultats les plus impactants. Concrètement, cela signifie que les personnes interagissant avec les dirigeants, seront aussi celles qui développent les modèles statistiques. Cela permet d'intégrer beaucoup de finesse dans les modèles, et donc plus d'exactitude, de fiabilité et de mieux piloter le risque dans les recommandations stratégiques.

Dans le cerveau d'un Data Scientist





Ekimetrics.

**Démultipliez la performance
de votre stratégie de revenu
ou de votre business model,
grâce à la Data Science.**

**Fondée en 2006, Ekimetrics
est une société de conseil internationale,
pionnière et leader européen en Data Science.**

- Modélisation du marketing mix (MMM)
- Customer Analytics
- Mesure d'impact sur la marque
- Transformation et stratégie data
- Roadmap data
- Process, gouvernance data et infrastructure IT

ELITECH

SLIM BAKLOUTI

+33 (0)1 84 25 81 55
contact@elitech.education
41 Rue Périer 92120 Montrouge FRANCE

www.elitech.education

Analytique
API
Applications
Consultants/SSII
Intégrateurs
Dataviz
Infrastructure
Institutionnels



FORMATION BIG DATA DIPLOMANTE

En présentiel ou à distance sur ELITECH365

INITIAL OU ALTERNANCE

BTS

Niveau III
BAC+2

LICENCE

Niveau II
BAC+3

MASTÈRE

Niveau I
BAC+5

INFORMATIQUE

COMMERCE

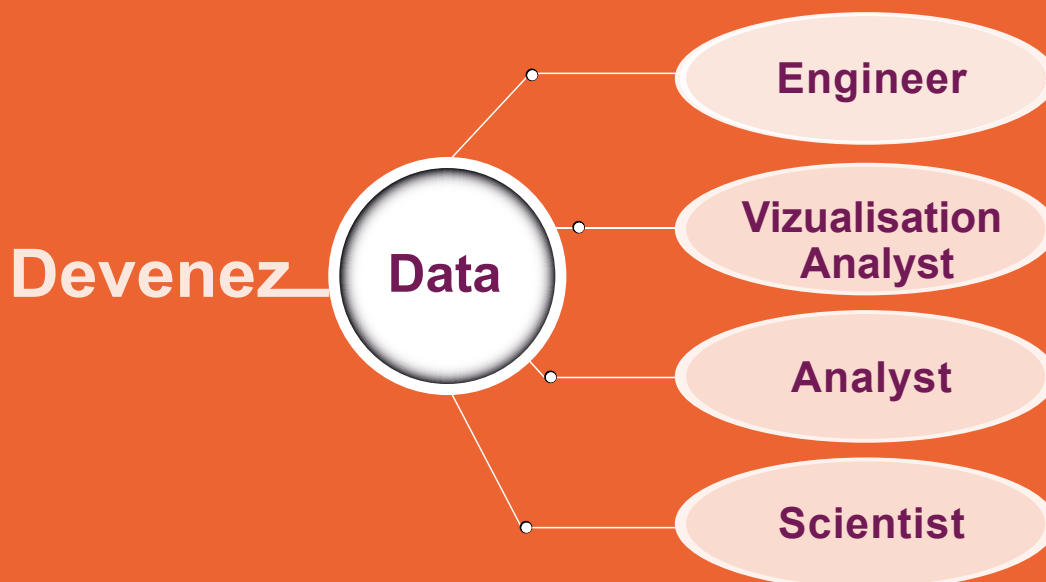


Inscrivez-vous
admission@elitech.education

www.elitech.education

FORMATION BIG DATA CERTIFIANTE

En présentiel ou à distance sur ELITECH365



Inscrivez-vous
learning@elitech.education



www.elitech.education



A PROPOS D'EPSILON FRANCE

EPSILON France est un alliage unique de compétences data, technologiques et marketing, combinant 3 actifs clés de Publicis Groupe : Soft Computing, Publicis ETO et Data Sciences de Publicis Media. Ce nouvel ensemble réunit 750 experts adtech et martech, compte plus de 350 clients actifs et pilote quotidiennement 80 plateformes data (DMP, CDP, RCU, CRM, programme relationnel, etc.).

EPSILON France appartient au réseau EPSILON acquis par Publicis Groupe en juillet 2019. Leader mondial du datamarketing, du CRM et de la personnalisation à grande échelle, EPSILON opère pour plus de 3 000 clients avec 10 000 collaborateurs présents dans 40 bureaux, dont 3 700 data scientists et 2 000 ingénieurs.

NOTRE VISION

Au modèle mass marketing qui prévaut aujourd'hui va se substituer très rapidement un modèle de mass personnalisation. Un prérequis à cette bascule : la data, sans laquelle aucune personnalisation à grande échelle n'est possible. Or, les marques sont face à des données éparpillées, des partenaires multiples et des empilements technologiques. EPSILON France est le master partner des entreprises dans leur stratégie data.

Nous intervenons à toutes les étapes-clés d'un projet datamarketing : de la définition de la stratégie à l'exécution opérationnelle, en passant par le design et la mise en œuvre de la solution.

Nous proposons des modes d'intervention agiles s'ajustant au plus près des besoins de nos clients, avec des solutions externalisées, in-team ou mixtes. Nous avons également construit une offre de formation pour accélérer la transmission des compétences. Enfin, l'ensemble des solutions phares du marché est complètement maîtrisé (Dataiku, Google, IBM, Microsoft, Oracle, Salesforce, SAS, Talend...) à côté de solutions propriétaires telles qu'EPSILON People Cloud.

NOTRE PLATEFORME PROPRIÉTAIRE

EPSILON PeopleCloud est la plateforme propriétaire de Publicis Groupe. Alliant Data, Technologie et Talents, elle permet aux marques d'accéder à une meilleure connaissance de leurs consommateurs.

Les enjeux et bénéfices sont nombreux : identifier les opportunités de croissance business, les segments de population qui y correspondent, pouvoir générer des insights qui nourrissent la créativité publicitaire, activer le bon message à la bonne personne sur tous les canaux, optimiser le mix de points de contact et enfin mesurer l'impact réel et la contribution de chaque action à la performance.

DIAGNOSTIC & STRATÉGIE

Qualifier le niveau de maturité data de l'entreprise au travers d'un indice data 360 intégrant l'ensemble des dimensions (cas d'usage réalisés, compétences, technologies et architectures, processus métiers et points de rupture, qualité de la donnée, budgets, organisation et gouvernance...). Définir la cible d'organisation data-driven et poser une trajectoire adaptée et son playbook pour valoriser progressivement le patrimoine de données par la mise en place de cas d'usage leviers de croissance.

GOVERNANCE & COMPLIANCE

Assurer une trajectoire d'entreprise data-driven pour démultiplier les cas d'usage autour de la donnée, par le déploiement de processus et d'outils, l'identification d'une communauté data, l'évangélisation et l'acculturation data de cette communauté, ainsi que la définition des rôles et d'une organisation opérationnelle autour de la donnée, dans un cadre cohérent et en conformité avec les réglementations nationales et internationales.

ARCHITECTURE

Définir le schéma directeur data de l'entreprise pour dé-siloter les systèmes d'information, gagner en agilité, se réapproprier ses data, bénéficier des avantages du cloud et construire une trajectoire de mise en œuvre ambitieuse mais cohérente. Forts de ses nombreux retours d'expérience, nous élaborons avec les équipes métier l'ensemble des dimensions d'une architecture data répondant aux nouveaux enjeux des organisations.

INTELLIGENCE ARTIFICIELLE

Mettre en œuvre tout type d'application marketing : scores d'appétences, de churn ou de ré-achat, segmentation clients, analyse des publications sociales, optimisation des parcours clients, moteur de recommandation ou de substitution produits, prévision de ventes, optimisation du yield, pricing dynamique. 100 Data Scientists en France, experts en Machine Learning, IA, analyse du langage naturel, traitement d'images, prévisions temporelles, traitement du signal.

NOTRE MISSION

Devenir le master partner des entreprises dans la constitution d'un patrimoine data, sa réconciliation, son enrichissement et son exploitation marketing afin de les propulser au plus vite dans l'ère de la mass personnalisation.

1

STRATÉGIE **DATA TRANSFORMATION**

Diagnostic & stratégie

Architecture

Gouvernance & compliance

Application IA & voice

2

MARKETING **BUSINESS EFFICIENCY**

Connaissance & insight

Parcours & expérience

Mesure & optimisation

Valorisation & monétisation

3

OPÉRATIONNEL **MARKET ACCELERATOR**

Process & compétences

Innovation & R&D

Plateforme & outils

Accompagnement & formation

EPSILON[®]

SUIVEZ-NOUS

in EPSILON France **🐦** @epsilondatamktg **f** epsilondatamktg **📷** epsilon_france
www.epsilon-france.com



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 sur 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





WE ARE

BUSINESS PERFORMANCE PARTNERS

eulidia

BUSINESS PERFORMANCE PARTNER



**FROM DATA
TO BUSINESS
PERFORMANCE**

Mixing Business Intelligence & Data Science

INNOVATION | DATA FACTORY | ADVANCED ANALYTICS | BUSINESS INSIGHTS | DATAVIZ

www.eulidia.com

EXPERT SYSTEM

ALAIN BIANCARDI

Vice President Sales & Marketing

+33 (0)6 14 43 44 94

abiancardi@expertsystem.com

207 rue de Bercy 75012 Paris FRANCE

www.expertsystem.com



INTELLIGENCE ARTIFICIELLE POUR LA COMPRÉHENSION DU LANGAGE NATUREL

L'intelligence artificielle cognitive constitue un levier de création de valeur par sa capacité à comprendre et rendre exploitables toutes les données, qu'elles soient structurées, c'est à dire représentées ou stockées dans un format prédéfini ou non structurées comme les textes rédigés en langage naturel.

Ces données sont toujours destinées aux humains. Typiquement, elles sont constituées de documents texte et peuvent contenir des faits, dates, montants, noms, instructions ou interrogations. Les technologies d'IA jouent un rôle clé dans la transformation de ces données en informations prêtes à l'emploi pour les processus de gestion ou pour enrichir les bases de connaissances.

L'automatisation des processus s'oriente vers l'intelligence artificielle

L'automatisation des processus séduit de plus en plus. En particulier le RPA (Robotic Process Automatisation) qui cible les activités à forte intensité de main d'œuvre, répétitifs et propices aux erreurs, donc coûteuses. Dans une publication de 2018, Gartner indique que d'ici 2020, au moins 30% des entreprises auront recours à l'intelligence artificielle pour renforcer un ou plusieurs de leurs processus principaux. Nous sommes déjà en 2020, donc il est grand temps de passer de la théorie à la pratique.

L'automatisation des processus s'avère très efficace pour gagner en productivité, améliorer les prises de décision, fidéliser les clients, garantir une parfaite conformité, rester compétitif... La liste est sans fin. Toutefois, de nombreuses tâches se basent sur la compréhension précise de données textuelles fournies dans un format qui nécessite encore une intervention manuelle et donc humaine.

Comment rendre l'automatisation intelligente ?

En comprenant le langage naturel de façon précise, l'intelligence artificielle cognitive convertit les données non structurées textuelles en formats lisibles par des machines pour chaque tâche. L'automatisation peut donc être utilisée pour des activités plus complexes jusque-là réservées aux êtres humains.

De ce fait, le champ d'intervention des outils RPA peut s'étendre à l'intégration de l'utilisation des informations non structurées issues de l'extraction et de la catégorisation automatiques et précises des données provenant de textes.

Optimisation des interactions clients grâce à l'IA

« Dans les premières années de mise en œuvre de l'IA, l'expérience client est la première source de valeur perçue par les entreprises, qui voient l'intérêt d'utiliser des techniques d'IA pour améliorer chaque interaction client et ainsi stimuler l'acquisition et la fidélisation client. L'expérience client est suivie de près par la réduction des coûts, car les entreprises cherchent à utiliser l'IA pour renforcer l'efficacité des processus, optimiser les prises de décisions et automatiser plus de tâches », a expliqué J.D. Lovelock, vice-président et analyste chez Gartner.

Une gestion efficace de l'expérience client requiert une vision à 360° des interactions clients, incluant les canaux traditionnels (conversation téléphonique, échange d'emails ou réunion en face-à-face), les messages provenant du site Web, des chats en direct avec les conseillers, des discussions avec un agent conversationnel (chatbot) ou encore les échanges sur les médias sociaux.

Au fil des interactions, l'ensemble de ces canaux génère un important volume de texte donc de données non structurées.

Les entreprises doivent être en capacité de les gérer efficacement, pour extraire rapidement et précisément les informations utiles et les intégrer aux autres données client existantes. Associer différentes données non structurées, issues de sources diverses parfois jusque-là inexplorées, peut s'avérer difficile. C'est là que l'intelligence artificielle cognitive intervient. L'IA permet d'analyser et d'organiser toutes les informations client afin de proposer des services personnalisés aux clients et réduire les coûts de la relation client.

Gestion automatisée des e-mails

L'intégration d'une IA à la gestion des emails améliore la qualité du service rendu et l'expérience client. L'intelligence artificielle permet de comprendre ce que le client a formulé avec ses propres mots, en extraire des informations, tout en prenant en compte ses préférences et ses besoins. L'intelligence artificielle garantit une compréhension complète du contenu, procède à une classification multi-sujets et extrait les données contenues dans les e-mails et les pièces jointes. Ainsi elle peut automatiser des opérations de support récurrentes afin d'accroître la productivité des agents tout en réduisant le temps et les coûts de la relation client.

Interactions client omnicanales en langage naturel

Plus les questions traitées dans leur contexte en langage naturel (par opposition à la recherche par mots-clés) sont comprises et adressées précisément plus le client est facilement et rapidement satisfait, donc fidélisé, à travers l'utilisation de tous les canaux de communication mis à sa disposition.

Intégrer l'IA aux chatbots, à la section FAQ, à la barre de recherche du site Web ou aux applications mobiles assure une compréhension précise des conversations même complexes et apporte des réponses automatiques dans un langage naturel.

Automatisation du back office

L'automatisation et la standardisation des tâches manuelles chronophages et répétitives liées à l'extraction de données issues de documents tels que commandes, formulaires ou rapports est synonyme d'une meilleure efficacité pour les activités de back office.

En conclusion

L'IA Cognitive appliquée à la compréhension du langage naturel offre de redoutables avantages concurrentiels aux entreprises qui ont compris les implications et les bénéfices de sa mise en application au sein des processus automatisés et les interactions clients.

IA COGNITIVE

L'Intelligence Artificielle qui transforme les données textuelles en informations utiles.

Comprendre le texte dans son contexte



Automatisation
des processus
métiers



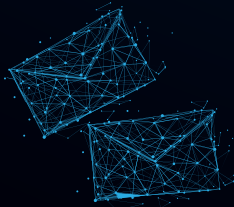
Optimisation des
interactions client



Analyse de
réglementations



Veille
stratégique
(KYC, KYS)



Catégorisation
d'emails



**RENDEZ-VOUS
AU STAND A46**



FINAXYS est un cabinet de conseil & d'intégration de solutions IT pour les acteurs de la finance, leader du marché Data for Finance.

Sans le savoir, vous nous faites confiance

FINAXYS

partenaire privilégié
du secteur bancaire Français

DES SUCCÈS DATA



Refonte Big Data du SI
Risques de marché



Formations avancées
de 300 développeurs



Gestion de la plateforme
de streaming



Construction d'un SI Data
Centric dans le Cloud



RÉGLEMENTATIONS



RISQUES



RÉFÉRENTIELS



DÉTECTION DE FRAUDE



MARKETING

NOS MÉTIERS

- ▶ MOA Finance
- ▶ Architecture
- ▶ Étude et développement
- ▶ Opérations et production
- ▶ Gestion de projet

350 CONSULTANTS



Notre histoire commence en 2008 avec 5 collaborateurs, nous sommes fiers d'être à présent 350 à porter les couleurs de Finaxys.

TOP 5 FOURNISSEURS



Notre savoir faire unique allié à notre connaissance des métiers de la finance sont des atouts reconnus de nos clients banques d'investissement.

30 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

Couchbase

confluent

ververica



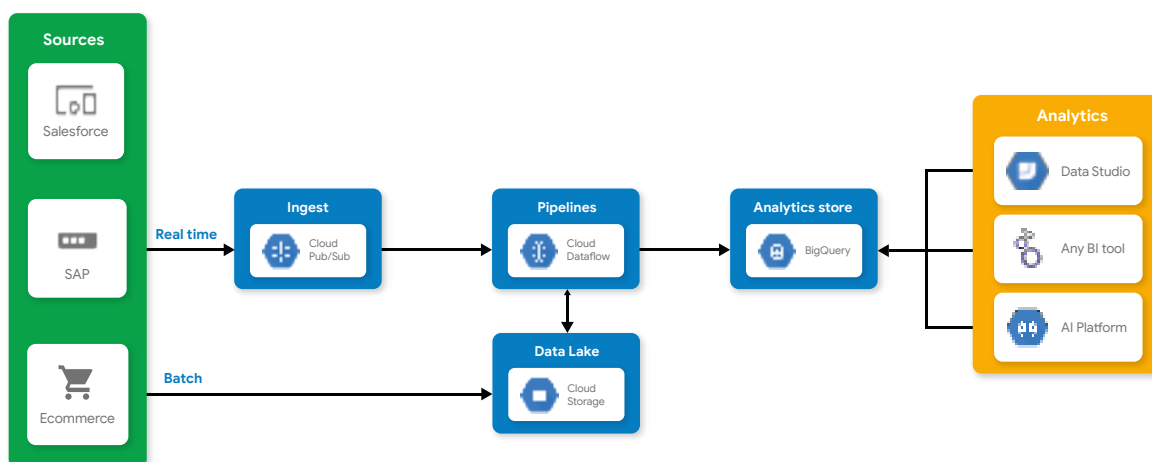
Les grandes tendances du Big Data selon Google Cloud

Une Plateforme Data Serverless

Concentrez vous sur l'exploitation de vos données, Google Cloud se charge de l'administration.

L'approche entièrement serverless de Google Cloud Platform sur ses services Data libère l'esprit de vos équipes des problématiques de mises à jour, scalabilité, infrastructure, disponibilité, sécurité ... occupant la majorité de leur temps, leur permettant de se concentrer sur la meilleure façon d'extraire de la valeur de vos données. Et grâce à un modèle de tarification intuitif, vous ne payez que les ressources que vous utilisez, souvent à la seconde de traitement ou à la requête.

Par exemple, BigQuery vous permet d'analyser jusqu'à des pétaoctets en utilisant le langage familier SQL, sans que vous n'ayez à penser à l'infrastructure sous-jacente. Pub/Sub et DataFlow vous permettent de récupérer les données en temps réel quelle que soit la volumétrie entrante pour l'acheminer jusqu'à BigQuery et Cloud Storage où elle pourra être persistée, puis visualisée avec Data Studio ainsi que nos solutions BI partenaires.



Transformez l'essai sur vos initiatives Hadoop

En utilisant Google Cloud Platform vous pouvez facilement maîtriser vos coûts, vos efforts et votre temps, comparé à une solution Hadoop on-premise.

Avec GCP, il est facile de concevoir une migration de vos traitements Hadoop et données HDFS vers des solutions scalables, plus performantes et moins difficiles à maintenir notamment en utilisant :

Cloud Storage comme base de stockage hautement résiliente

Cloud BigQuery pour toutes vos analyses type OLAP / SQL

Cloud BigTable qui supporte aisément vos usages NoSQL type HBase

Cloud Dataproc, environnement Hadoop et Spark managé pour exécuter la plupart de vos tâches existantes avec un minimum de modifications, et en bénéficiant de la flexibilité et la tarification à l'usage au plus proche de vos workloads.

57%

Moins cher que les solutions Hadoop on-premises

32%

Moins cher que AWS EMR



...les clients ont également fait état d'**avantages substantiels dans la valeur stratégique** qu'ils ont pu retirer des données hébergées dans Google Cloud.

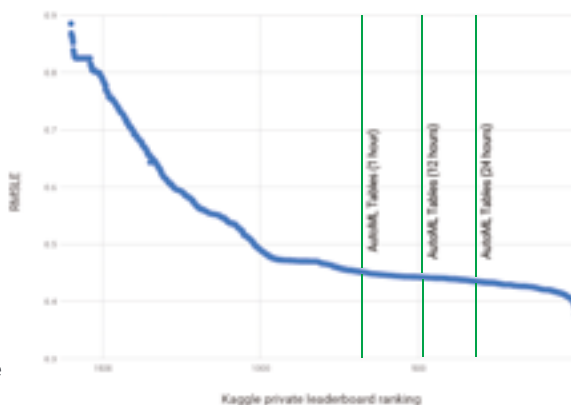
D'après une étude réalisée par Enterprise Strategy Group



Le machine learning automatisé par Google Cloud

Cloud AutoML est une suite de produits de machine learning qui permet à toute personne maîtrisant un périmètre de données de construire et de déployer des modèles de Machine Learning **automatiquement et en toute simplicité, sans une seule ligne de code.**

Cloud AutoML permet de travailler avec différents types de données : images (AutoML Vision), vidéo (AutoML Video intelligence), textuelles (AutoML Natural Language et AutoML Translation) ou encore structurées (AutoML Tables) et vous permet de tirer parti des dix ans d'innovation et de recherche des laboratoires de Deep Learning de Google sur vos données et pour vos usages.



Ce graphique permet de voir qu'en une heure d'entraînement, Cloud AutoML est déjà plus performant que la moitié des participants sur cette compétition de machine learning.





NOTRE HISTOIRE

Le Groupe Estia est un cabinet de conseil spécialisé dans la transformation des données en actifs, travaillant sur l'intégralité du cycle de valorisation de la data.

Le cabinet, fondé en 2010 à Paris, compte plus de 110 collaborateurs experts de la data (data scientists, data analysts, data stewards, data engineers, experts métier : CRM, Digital, Risque, ...) pour un CA de 11m€.

En 2017, la société accélère son développement en s'appuyant sur un nouvel actionnaire de référence, le Groupe Consort, intégrateur et opérateur de solutions et d'infrastructures (2 400 collaborateurs, 172m€ de CA, 7 pays). Cette alliance permet de proposer au marché un continuum de services composé de briques technologiques, d'expertises en infrastructure et en algorithmie au service de la performance opérationnelle.

LA DATA, NOTRE CŒUR D'ACTIVITE

Le Groupe Estia couvre l'ensemble de la chaîne de valeur data : de la transformation des données en asset au travers de leur exploitation opérationnelle par les métiers (pilotage, communication omnicanale, prévision...) à la conception d'outils, d'applications analytiques, d'algorithmes & modèles en passant par la spécification et le développement d'environnements performants de stockage...

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 et de l'intelligence artificielle (de la conception à la mise en opération).

NOS OFFRES DE SERVICES

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 métiers : pilotage des activités, connaissance clients, communication omnicanale, modélisation à visée prédictive, transformation vers une organisation data centric, gestion de projets data... et de pratiques sectorielles : Banque, Assurance, Energie, Distribution, Médias, Télécoms...

COUVERTURE GEOGRAPHIQUE

Le Groupe Estia dispose d'une implantation en France (Paris, Lille) et en Amérique du Nord (Montréal). Il bénéficie du réseau du Groupe Consort en France (Lyon, Sophia Antipolis, Toulouse, Nantes, Rennes) comme à l'international (Munich, Stuttgart, Nuremberg, Francfort, Bruxelles, Luxembourg, Rabat, Casa, Genève).

ILLUSTRATIONS DE REALISATIONS



Secteur assurantiel : Calcul d'algorithmes permettant de déterminer le coût d'acquisition omnicanal d'un nouveau client



Secteur bancaire : Développement de modèles prédictifs garantissant l'alignement avec la réglementation émise par la BCE



Secteur de l'énergie : Construction de crawlers pour capter et utiliser des contenus digitaux à des fins de veille concurrentielle



Secteur de la distribution : Création et mise en œuvre de segmentations prédictives de la valeur future à des fins d'optimisation de l'allocation du budget marketing



Secteur des médias : Accompagnement à l'implémentation d'un outil de CRM multicanal et opération du plan d'actions marketing



Secteur du service : Conception d'algorithmes de deep learning pour monétiser une base de données

DIFFERENTIEATEURS

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 nous exonérant de tout effet 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.

Estia est, par ailleurs, totalement indépendant des éditeurs de solutions logicielles, ce qui nous permet de proposer les solutions les plus efficaces et de faire preuve d'un haut niveau d'impartialité dans nos recommandations.

Le Groupe Estia consacre, depuis sa création, des moyens conséquents à l'innovation (label Jeune Entreprise innovante obtenu en 2013, programme de R&D annuel, organisation orientée sur l'émergence de nouvelles méthodes, réalisation de kaggles internes tout au long de l'année) et propose à ses clients de s'associer à des programmes de recherche pour identifier ensemble de nouveaux leviers de croissance !

- Sophie est avocate
- au barreau de Lyon
- depuis 9 ans.



Et vous, qu'avez-vous à apprendre sur vos clients ?

Une expertise unique pour faire de la data un levier de performance.

WWW.GROUPE-ESTIA.FR

 **estia**
Turn data into business value
by Consort[®]

Paris - Lyon - Sophia Antipolis - Toulouse - Nantes - Rennes - Lille - Montréal - Munich - Stuttgart - Francfort - Cologne - Bruxelles - Luxembourg - Rabat



HPE accélère l'innovation en Intelligence Artificielle avec une solution d'entreprise pour gérer l'ensemble du cycle de machine Learning

Le nombre d'entreprises qui ont adopté l'Intelligence Artificielle a plus que doublé ces quatre dernières années et celles-ci continuent d'investir beaucoup de temps et de ressources dans la création de modèles de machine learning et de deep learning. Cela afin de proposer un large éventail d'applications de l'IA, allant de la détection de fraude à la médecine personnalisée ou à l'analyse prédictive des comportements clients. Cependant, **le plus grand défi auquel sont confrontés les professionnels réside dans la mise en œuvre du machine learning**, c'est-à-dire l'étape opérationnelle, en vue de réussir le déploiement et la gestion de ces modèles et d'en dégager de la valeur pour l'entreprise. Selon Gartner, d'ici à 2021, **au moins 50 % des projets de machine learning ne seront pas totalement déployés en raison de lacunes sur le plan opérationnel.**

Pour apporter à cette problématique, **Hewlett Packard Enterprise (HPE)** a introduit début Septembre **HPE ML Ops**, une solution logicielle à base de conteneurs destinée à supporter l'ensemble du cycle de vie des modèles de machine learning dans les environnements sur site, de cloud public ou de cloud hybride. La nouvelle solution inaugure un **processus de type DevOps afin de normaliser les workflows de machine learning et ramener les délais de déploiement de l'Intelligence Artificielle (IA) de plusieurs mois à quelques jours.**

La solution HPE ML Ops diminue le délai de déploiement de l'Intelligence Artificielle de plusieurs mois à quelques jours

La nouvelle solution **HPE ML Ops** étend les capacités de la plateforme logicielle HPE BlueData EPIC™ pour les conteneurs, et donne aux équipes de data scientists un accès à la demande aux environnements conteneurisés pour les applications en IA, machine learning et analytiques distribuées. **HPE a acquis BlueData en novembre 2018 afin de renforcer ses offres dans les domaines de l'IA, de l'analytique et des conteneurs**, en complément de ses propres solutions d'informatique hybride et de ses services Pointnext pour les déploiements IA en entreprise.

HPE ML Ops transforme les projets d'IA expérimentaux et pilotes en opérations de production en couvrant l'ensemble du cycle de vie du machine learning, depuis la préparation des données et la création des modèles jusqu'à leur apprentissage, leur déploiement, leur supervision et leur utilisation collaborative.

Olivier TANT, Sales Category Manager « Scale-Out / Big Data / AI Platforms » au sein de la division HPE Hybrid IT France : « *la création de valeur au sein des entreprises nécessitent des plateformes agiles, standardisée et des modèles de Machine Learning opérationnels. Aux technologies BlueData et à cette nouvelle offre HPE MLOps, nous pouvons désormais associer les technologies MAPR – acquises cet été 2019 – au sein d'une vision 'Container Data Platform' unique. Nous apportons à l'analytique Big Data et à l'AI la rapidité et l'agilité de DevOps, permettant ainsi d'accélérer leur opérationnalisation et leur rentabilité, avec une approche de bout en bout pour gérer le cycle de vie des environnements déployés sur site ou dans le cloud hybride.* »

Avec **HPE ML Ops**, les data scientists participant à la création et au déploiement des modèles de machine learning bénéficient en effet de la **solution la plus complète du marché pour la gestion du cycle de vie de l'IA en entreprise:**

- Création de modèles : environnements de machine learning et de data science prépackagés, en libre-service
- Entraînement des modèles : environnements d'entraînement évolutifs avec accès sécurisé aux données.
- Déploiement des modèles : flexibilité, rapidité et reproductibilité des déploiements.
- Suivi des modèles : visibilité de bout en bout sur le cycle de vie des modèles de machine learning.
- Collaboration : possibilité de workflows CI/CD avec référentiels de code, de modèles et de projets.
- Sécurité et contrôle : Multi-location sécurisée avec intégration aux mécanismes d'authentification de l'entreprise.
- Déploiement hybride : prise en charge des environnements sur site, de cloud public ou de cloud hybride.

La solution HPE ML Ops fonctionne avec un large éventail de frameworks de machine learning et de deep learning en open source, notamment Keras, MXNet, PyTorch et TensorFlow, ainsi que les applications de machine learning de nos partenaires, tels que Dataiku et H2O.ai.



**Hewlett Packard
Enterprise**

FASTER TIME-TO-VALUE FOR ENTERPRISE AI



**Accelerate Innovation
with Machine Learning
and Data Analytics**

[HPE.com/info/BlueData](https://www.hpe.com/info/BlueData)

HURENCE

HUBERT LAURENCE

CEO

+33 (0)6 50 89 86 27

Laurence.Hubert@hurence.com

400 chemin des longs prés 38660 FRANCE

www.hurence.com

Analytique

API

Applications

Consultants/SSII
Intégrateurs

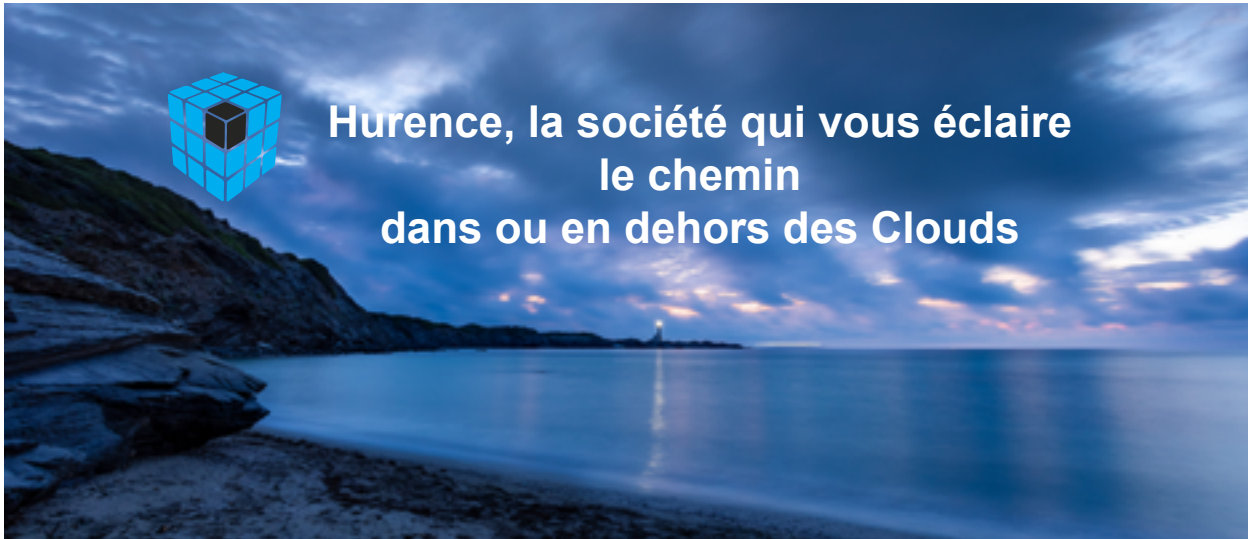
Dataviz

Infrastructure

Institutionnels



hurence



Assurez votre indépendance dans le Cloud comme On Premise

avec l'**accompagnement Big Data / IA** de Hurence

et en option les outils de la plateforme open source LogIsland



Le traitement d'événements temps réel
on-premise et dans le cloud

Web Analytics

- Recommandations
- Up-selling, cross-selling
- Gestion avancée du Panier
- Génération de leads
- Prédications de stocks, ventes
- Réutilisation de plans de taggag

Industrie 4.0

- Connecteurs sur robots
- Connecteurs sur serveurs OPC
- Supervision de la fab
- Smart alerting
- Détection d'anomalies
- Détection de défauts



Pour en savoir plus: contact@hurence.com

LAURENCE, PEUX TU NOUS FAIRE UN RETOUR SUR LES ÉVÉNEMENTS DE L'ANNÉE ÉCOULÉE ?

Le Guide du Big Data est en effet le média idéal pour faire à ses lecteurs, dont beaucoup sont des clients de Hurence, un retour d'expérience sur l'année écoulée. Force est de constater que la période a été riche en rebondissements et suspenses avec des fusions et acquisitions qui nous ont tous fait nous interroger sur nos stratégies techniques et leur pérennité.

Le monde du Big Data a été secoué par la fusion et acquisition des deux distributions Hadoop: Hortonworks HDP et Cloudera CDH. Rien de moins. Après un suspense insoutenable de plusieurs mois, l'annonce tombe: la nouvelle offre issue de la fusion, baptisée CDP sera bien open source mais payante avec une tarification au noeud conséquente. Dans le même temps, le rachat par HPE de MapR, en difficulté financière, constitue une autre incertitude pour un autre lot de clients. Seule certitude, nous revenons vers un modèle standard de licensing sans plus aucune alternative open source, packagée, gratuite mais néanmoins supportée, pour tous les outils que nous avons adoptés.

Hurence se faisait toujours le chantre de solutions open source de qualité et gratuites, bien packagées, parce que sa clientèle est aussi composée de PME ou petites ETIs qui pouvaient dès lors, elles aussi, bénéficier de ce monde Big Data et de ses apports innovants. Adieu donc notre HDP...

Heureusement nous étions préparés et avons commencé à faire notre deuil d'une distribution Hadoop complète, aussi pour des raisons de simplicité et coût de mise en oeuvre... Nous avons déjà embarqué certains de nos clients sur des solutions alternatives, assemblées avec nos coups de coeur, récents (Kubernetes, OpenFaas) ou anciens (Spark, Mesos, Elasticsearch-SolR, Postgresql cluster...), plus légères.

En parallèle l'offre Cloud s'est étoffée chez les 3 premiers grands acteurs et nombre de nos clients ont souhaité transférer une partie de leurs développements vers quelque chose qui leur semblait plus pérenne et plus agile en terme d'évolutivité: le cloud!

D'ailleurs la concomitance nous a parfois surpris, à tel point que nous nous sommes souvent demandés si les choses n'avaient pas été orchestrées intelligemment pour que tous nous finissions "dans le Cloud" de guerre lasse de trouver la bonne distribution Big Data pérenne pour au moins les 10 prochaines années. N'oublions pas que les actionnaires des startups sont souvent les géants de la tech!

Alors nous avons pris le taureau par les cornes et commencé à concevoir nos déploiements "lift and shift" - pour les novices du cloud cela veut dire répliquer à l'identique une solution avant de remplacer progressivement des éléments et d'adopter des outils éventuellement plus «propriétaires». On a donc éliminé HDFS au profit d'un S3 ou d'un Azure ou autre Filestore - souvent gardé notre Spark mais étendu nos possibilités d'utiliser d'autres engines - et nous avons commencé à abstraire nos couches d'événements et nos stockages de telle sorte de s'adapter à l'un ou l'autre et à redevenir indépendants.

Rendre indépendants nos clients tout en leur offrant le meilleur de l'analytique temps réel à moindre coût grâce au logiciel libre et à nos savoirs-faire, c'est un peu notre marque de fabrique à Hurence. A l'aune du cloud, notre spécialité est d'assurer cette indépendance qui peut se révéler vitale. Nos technologies ou nos projets clients savent maintenant vivre dans le cloud, utiliser les technologies qu'elles offrent, mais elles savent en quitter un pour en rejoindre un autre, ou les quitter tous, sans dommages collatéraux.

Nos petits clients nous en remercient mais nos grands clients ont encore trop le réflexe de se sourcer à ces entreprises qui brillent de mille feux pour disparaître en quelques années sous nos yeux interloqués. Peut-être pourraient elles profiter des enseignements de cette année riche, pour tisser des liens plus forts sur des acteurs qui s'inscrivent dans la durée, et se faisant les renforcer, car la financiarisation à outrance de l'open source et plus généralement des startups a des effets aussi rapides que brutaux.

Cette année nous avons dit adieu à notre distribution Hadoop. Il y a une vie après Hadoop... Nous le remercions néanmoins pour ce chemin : il nous a emmené vers le Big Data, l'IA avec ses forces et ses faiblesses et nous a exposés à des technologies sur lesquelles nous pouvons continuer, isolément, à bâtir.

Est-ce qu'il aura ressuscité l'an prochain dans une autre enveloppe? Toute réincarnation est possible... Mais gardons en tête la jolie maxime de Thomas Baitel, notre CTO: la donnée c'est comme de l'eau elle suit le chemin le plus facile.



IDEATRANS, AGENCE DE TRANSFORMATION DATA DIGITALE & INNOVATION ! PARIS – BRUXELLES

iDeaTrans : Transformer & Transmettre ! iDeaTrans vous aide à développer et transformer vos organisations, vos métiers et votre business par l'univers de la Data (Open, Big, Small, Smart, IA & ML...), le Digital et l'Innovation, avec une méthode adressant les principaux éléments de Valeur suivants :

- Mesurer votre maturité Data (-Digital), auditer
- Révéler, faire émerger, clarifier, libérer
- Sourcer, sécuriser, intégrer
- Gouverner, piloter, pivoter, gestion de projet
- Analyser, valoriser, transformer la Data-Digital en éléments de différenciation métiers, en création de valeur et de confiance - Monétisation
- Schéma directeur, meilleur efficacité, nouveaux business, voire de nouveaux paradigmes

Clients : EVOLUTION xy (plateforme Data et nouveaux services pour une coopérative de bovins), SSI - AIRBUS (initialiser un projet Data pour améliorer la maintenance et la gestion des pannes moteurs d'avion, 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), 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,)), de l'idée au marché), AREAS Elior, DANONE, SIEMENS, CLUB MED, H5AUDIT, 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 (création de Plateformes SIRH, ERP, ETL, MDM & Quality, 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'évènements Data - Digital - Innovation (ex : www.lanuit-dubigdata.com), etc...

Contributions : AFNOR, AFDEL TECHIN, IDC, Gartner, CIO, CIOnet, IT for Business, 01, IT Expert, Mag IT, CIGREF, IBM, Oracle, Informatica, QLIK, Business Objects SAP, AFAI, AFCDP, Finance Innovation, EBG, Alliance Big Data, Big Data Paris, Cap Digital, ESSEC, Novancia, AFM, CMIIT, Distributive, Printemps des Etudes, G9+, Total, Micropole, Neurones, B&D, Linkbynet, etc...

Organisation innovante et collaborative : iDeaTrans réunit des acteurs pertinents et des experts métiers (indépendants) depuis +15 ans, organisation ouverte. iDeaTrans se nourrit de méthode et démarche d'innovation s'appuyant sur l'intelligence collective pour poser des visions, révéler, faire émerger les idées et clarifier les besoins, les défis à relever, 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 président de l'Alliance Big Data et associé dans dans plusieurs startups comme « Ange d'Affaires » depuis 1997 (StartupWay radio classique éco, Integra, Moneytis, Privowny, Stample, Keakr, Moichef, ...) ainsi que dans plusieurs organisations comme, ITIIP (plateforme pour le commerce inclusif par l'innovation et l'investissement, Nouméa), etc. Il anime des ateliers « de l'idée au marché ».

Ses 25 années d'expérience internationale dans le Business de la Donnée (360° métiers, Marketing, Marché, Distribution, Comportements, Traces, Plateformes, Intégration, Valorisation, Protection (CIL), Open & Shared Data, etc) font de lui un expert en matière de (Big) Data (de l'usage à la création de valeur Data, jusqu'à la monétisation) et de Transformation (numérique) de l'entreprise (startup et grands groupes). Fabrice est également Commissaire aux Données pour de grands projets territoriaux et sociétaux (HIA www.hackinadour.com (2017), www.c3challenge.com COP21 C3 (2015)).

Intra-entrepreneur innovant, Fabrice a été Directeur des Systèmes d'Information et Opérations (DOSI - CIO), en charge des Projets Métiers, membre du COMEX de GfK IFR Global (60 pays), et CIL.

Animé par ses valeurs, Fabrice publie, enseigne, prend régulièrement la parole pour des conférences, organise des événements et tables rondes sur ses sujets, lui permettant de partager et d'enrichir ses connaissances.



TRANSFORMER

Initier & accompagner

DATA

Usages
Valeur

INNOVATION

Idée > Marché
Startups

DIGITAL

Agilité
Efficacité

*Auditer
Révéler
Faire émerger
Sourcer
Sécuriser*

*Gouverner
Analyser
Valoriser
Différencier
Créer de la valeur
& de la confiance*

iDEATTRANS
Innovation Data Digital

TRANSMETTRE

Créer les conditions

*Traces
Informations
Expériences
Parcours*

*Expertises
Connaissances
Savoir-faire
Excellence*

iDeaTrans

Agence de transformation data, digitale & innovation
Paris • Bruxelles • fabrice.benaut@ideatrans.net
www.ideatrans.net



1^{ère} communauté
du Big Data



RETOUR D'EXPÉRIENCE — MEDIA

"A LA CLÉ, UNE CONNAISSANCE CLIENT PLUS FINE ET PERTINENTE"

LA SOLUTION POUR ATTEINDRE UNE ANALYSE INSTANTANÉE ET PRÉCISE DES BIG DATA

LES BIG DATA, SOCLES DE LA CONNAISSANCE CLIENT

L'UN DES GROUPES DE **MÉDIAS AUDIOVISUELS LEADER EN FRANCE**, AVEC 15 MILLIONS D'ABONNÉS DANS LE MONDE ENTIER, A SU SE DIVERSIFIER EN PROPOSANT DE NOMBREUSES OFFRES CAPABLES DE RÉPONDRE AUX BESOINS SPÉCIFIQUES DE SES CLIENTS SOIT UN ENSEMBLE DE SERVICES AUTOUR D'UNE EXPÉRIENCE UTILISATEUR MULTICANAL.

A L'ORIGINE, LES PROBLÉMATIQUES DE CHARGE ONT POUSSÉ L'ÉQUIPE DATA À **MIGRER LES DONNÉES SUR UN DATA LAKE EN MODE CLOUD**. UN ENVIRONNEMENT ADAPTABLE QUI PERMET DE RÉALISER DES CROISEMENTS ET DES CALCULS MULTIPLES SUR DEUX SEGMENTS CLÉS : LA VOD POUR ANALYSER LES ACHATS À L'ACTE ET LE REPLAY POUR ANALYSER LA CONSOMMATION GLOBALE SUR LES DIFFÉRENTES PLATEFORMES.

MAIS, PRÉCALCULÉES **AVEC DES DATAMARTS, CES OPÉRATIONS SE MONTRENT ALORS PARTICULIÈREMENT CHRONOPHAGES** : POUR UNE JOURNÉE DONNÉE, LE TEMPS DE CALCUL POUVAIT S'ÉLEVER À 12H. RÉSULTAT : LES ÉQUIPES MÉTIERS RECEVAIENT LES CHIFFRES CONSOLIDÉS 24 À 36 H PLUS TARD. « NOUS NE FAISONS QUE DU COUNT DISTINCT SUR L'ENSEMBLE DE NOS PLATEFORMES : COMBIEN D'UTILISATEURS UNIQUES ONT CONSOMMÉ CE PROGRAMME, COMBIEN DE COUNT DISTINCTS PAR PLATEFORME, COMBIEN PAR JOUR... OR LE DIGITAL IMPLIQUE DE L'INSTANTANÉ ! » EXPLIQUE LE RESPONSABLE DE L'ENTITÉ DATALAB DE L'ENTREPRISE. OBJECTIF : EFFECTUER DES REPORTINGS EN MOINS D'1 H !

OUTRE LES FAIBLES PERFORMANCES ET LA LENTEUR AFFICHÉES, L'ÉQUIPE DATA DOIT FAIRE FACE À UN **AUTRE PROBLÈME LIÉ À LA PROFONDEUR DE SON HISTORIQUE** : EN EFFET, LES MÉTIERS AIMERAIENT POUVOIR COMPARER LA CONSOMMATION ET L'ÉVOLUTION DES ACHATS À L'ACTE D'UNE ANNÉE À L'AUTRE. UNE ANALYSE QUI IMPLIQUE LÀ AUSSI UN TEMPS D'INTERROGATION DES DONNÉES TRÈS LONG.

L'ANALYTIQUE EN INSTANTANÉ SUR L'ENSEMBLE DES DATA ET LES CAPACITÉS RELATIONNELLES, POINTS FORTS D'INDEXIMA

UN BENCHMARK EST ALORS RÉALISÉ DES SOLUTIONS CAPABLES À LA FOIS D'ADRESSER UNE VOLUMÉTRIE IMPORTANTE DE DONNÉES AVEC UN HAUT NIVEAU DE PERFORMANCE, ET DE RÉPONDRE À CERTAINES CONTRAINTES IMPOSÉES PAR L'INFRASTRUCTURE CLOUD EN PLACE DANS AWS S3 ET REDSHIFT. EN EFFET, L'ÉQUIPE SOUHAITAIT POUVOIR CONNECTER LA SOLUTION AVEC LES OUTILS DU GROUPE, À SAVOIR MICROSOFT OFFICE 365 ET POWER BI, POUR **NE PAS AVOIR À DÉVELOPPER UNE NOUVELLE INTERFACE ET FAVORISER UN ACCÈS SÉCURISÉ AUX DONNÉES « ANYTIME, ANYWHERE »**.

« NOUS ÉTIONS EN RELATION AVEC INDEXIMA DEPUIS 2 ANS ET NOUS AVONS FORTEMENT APPRÉCIÉ LEUR CAPACITÉ D'ÉCOUTE ET LEURS CONSEILS. DE PLUS, LEUR SOLUTION RÉPONDAIT EN TOUT POINT À NOS ATTENTES : LEUR MOTEUR RELATIONNEL NOUS PERMET D'EFFECTUER DES JOINTURES À LA VOLÉE CE QUI NOUS ÉVITE DE RETRAITER LA DONNÉE A POSTERIORI POUR L'ENRICHIR, ET LEUR SOLUTION EST INTERROGEABLE SOUS SQL STATIC ! »

VERS DES ANALYSES PREDICTIVES

DÉSORMAIS, L'ENSEMBLE DES INFORMATIONS SONT COLLECTÉES EN TEMPS RÉEL. **TOUTS LES CALCULS SONT RÉALISÉS PAR LE DATA HUB D'INDEXIMA AVEC UN TEMPS DE TRAITEMENT INFÉRIEUR À 3 MINUTES SUR UN HISTORIQUE DE DONNÉES DE 24 MOIS**. DES PERFORMANCES QUI OUVRONT DE NOUVELLES PERSPECTIVES AUX ÉQUIPES DATA MAIS AUSSI MÉTIERS.

« LES UTILISATEURS NOUS DEMANDENT D'ORES ET DÉJÀ DE NOUVEAUX INDICATEURS POUR UNE ANALYSE PLUS COMPLEXE DU COMPORTEMENT DES CONSOMMATEURS ET DE NOUVEAUX CROISEMENTS DE DONNÉES PLUS POUSSÉS. NOUS POURRONS PAR EXEMPLE PRÉCISER LE STATUT DE L'UTILISATEUR AU MOMENT DE L'ACHAT OU ENCORE TRAITER DES VOLUMES BEAUCOUP PLUS IMPORTANTS. » À LA CLÉ : UNE CONNAISSANCE CLIENT PLUS FINE ET PERTINENTE.

DES TESTS DE PERFORMANCES SONT ÉGALEMENT ENVISAGÉS SUR DES VOLUMES CENT FOIS PLUS IMPORTANTS SUR LES PÉRIMÈTRES LIÉS AUX USAGES ET À LA NAVIGATION. PROCHAINE ÉTAPE : DÉPLOYER LA DATA SCIENCE POUR EFFECTUER DES CIBLAGES PLUS PRÉCIS ET DU CALCUL DE SCORE PRÉDICTIF.



SELON VOUS, QUEL EST L'IMPACT DE LA MIGRATION GLOBALE DES DATA DANS LE CLOUD POUR LA BI DANS LES ENTREPRISES ?

Ces dernières années, nous avons assisté à une tendance généralisée où les entreprises sont passées d'une infrastructure informatique rigide au cloud, plus flexible et évolutif. Nous savons qu'il ne s'agit pas d'une mode mais d'une formidable opportunité pour les experts de la donnée, dans la lignée de la « BI self-service ». Plus besoin aujourd'hui pour les experts métiers de confier leurs données à des spécialistes de la data en interne pour l'analyse. Désormais, il est possible pour les équipes de gagner un temps considérable en faisant directement l'analyse.

Mais le cloud n'est pas une solution miracle, il reste de vrais enjeux à régler sur deux piliers : la performance et le coût. On voit deux tendances émerger : les datawarehouses et les datalakes dans le Cloud. Certaines entreprises construisent leurs datalakes, sur S3 dans AWS par exemple, et rencontrent rapidement des problèmes de performance. C'est-à-dire qu'elles peuvent passer des minutes voire des heures à récupérer leurs données pour les corrélérer et les voir s'afficher dans leur outil de BI, ce qui n'est plus acceptable à l'heure de l'instantané. D'autres mettent en œuvre des solutions plus rapides avec les Datawarehouses de type Snowflake ou BigQuery mais alors les prix s'envolent avec l'augmentation de la donnée utilisée. Et, c'est là que nous intervenons avec Indexima afin de les aider à relever ces deux défis avec succès.

COMMENT INDEXIMA ACCOMPAGNE LES ENTREPRISES DANS CETTE NOUVELLE ÈRE ?

Chez Indexima, nous proposons un Data hub permettant aux entreprises d'accéder à leurs données pour simplifier et accélérer l'analyse, rendant possible l'analytique en instantané, quelles que soient les sources de données, leur volumétrie et les applications de restitution. Indexima est né il y a 6 ans dans ce contexte de migration vers le cloud, ce qui nous a permis d'être un spécialiste de la BI dans le cloud dès notre création.

Pourquoi Indexima est-il performant ? Le moteur va piocher la donnée qui se trouve dans le cloud puis, grâce au machine learning, apprendre automatiquement à partir de l'utilisation faite. Il n'y a plus besoin de préparer la donnée soi-même car le moteur apprend à la place de l'utilisateur, libérant un temps énorme aux équipes. Aujourd'hui, nous rendons possibles les promesses du big data. C'est le one-click cloud !

Mais ce que nos clients apprécient le plus, c'est la simplicité de mise en œuvre et la facilité à utiliser Indexima. Grâce notamment à nos partenariats technologiques avec les grands fournisseurs d'infrastructure comme Microsoft, Amazon, Google et Cloudera nous permettent d'assurer une mise en œuvre d'Indexima en quelques heures quelle que soit l'architecture de nos clients, on-premise ou dans le cloud mais également avec les principaux outils de restitution.

COMMENT INDEXIMA REPOND-ELLE AUX PROMESSES DE LA DATA SCIENCE ET DES DIFFERENTS USAGES DATA POUR LES ENTREPRISES ?

Notre volonté chez Indexima est de simplifier la vie de nos clients et de faciliter l'accès à la data science. L'enjeu pour nos clients, c'est de réussir à interroger la donnée et à lui donner du sens. Ils veulent comprendre ce qui se passe sans avoir besoin d'orienter leurs recherches au préalable.

Tout comme la BI self-service, il est possible pour les équipes métier d'accéder directement à la data science. Avec Indexima, la donnée est déjà dans la chaîne BI, la data n'a pas besoin d'être déplacée ni d'être analysée par des data scientists qui nécessiteraient un temps et des coûts considérables. Le moteur d'Indexima est plus flexible grâce au SQL, permettant de simplifier la data science tout en la rendant compréhensible et accessible.

Mais plus largement, notre ambition est de garantir une capacité à simplifier et accélérer l'ensemble des usages data de nos clients : BI, data science, data discovery, data preparation.

INFORMATICA

KARINE BUISSON-GENTLHOMME

Field Marketing manager

+33 (0)1 42 04 89 25

kbuisson-gentilhomme@informatica.com

Tour CB 21, 16 place de l'Iris 92040 FRANCE

www.informatica.com/fr



Informatica®



Informatica®

Qui sommes-nous ?

Informatica est l'unique leader dans la gestion des données Cloud d'entreprise qui permet d'accélérer la transformation digitale axée sur les données. Leader dans la gestion des données Cloud d'entreprise, Informatica est le partenaire stratégique des entreprises accélérant leur transformation digitale à partir des données. Informatica propose des solutions qui permettent aux entreprises d'exploiter tout le potentiel de leurs données afin de stimuler l'innovation, de gagner en agilité et de concrétiser de nouvelles opportunités de croissance, avec à la clé des évolutions positives du marché. Depuis plus de 25 ans, Informatica accompagne plus de 9 000 clients à libérer le potentiel de leurs données dans le monde entier.

Awards et récompenses

Informatica est nommé leader par Gartner dans tous les segments de marché couverts par Informatica, notamment dans :

- Magic Quadrant pour les outils d'intégration des données (août 2019)
- Magic Quadrant for Enterprise Integration Platform as a Service (iPaaS) (Avril 2019)
- Magic Quadrant pour les solutions de Master Data Management (Décembre 2018)
- Magic Quadrant pour les outils de qualité des données (Mars 2019)
- Magic Quadrant pour les solutions de gestion des Métadonnées (Août 2018)

Innovation

100

brevets

8

centres de R&D

196M\$

d'investissement en R&D par an

La plateforme Informatica Intelligent Data Platform™ est la solution la plus complète et modulable du marché, construite sur une architecture de micro services, permettant aux entreprises de libérer le potentiel et la valeur de leurs données. La plateforme basée sur l'IA couvre le Big Data, le multi-cloud ou on premises, afin que la donnée soit fiable, sécurisée, gouvernée, accessible, en temps souhaité, pertinente et exploitable. Cela permet aux entreprises les plus évolutives de mener leur transformation digitale à partir de leurs données.

Contacts : Tel : 01 42 04 89 00

Email : info_fr@informatica.com | Site web : www.informatica.com/fr

Solutions

- Gouvernance et sécurité des données
- Cloud / Cloud hybride
- Engagement 360
- Big Data / Next Generation Analytics

Chiffres clés

- Société fondée en 1993
- Siège social à Redwood City, Californie
- Anil Chakravarthy, CEO
- En 2015, le fonds d'investissement Permira et l'Office d'investissement du régime de pensions du Canada acquièrent Informatica pour 5,3 milliards de \$
- Plus de 9000 clients dans 82 pays depuis 25 ans
- Parmi les clients, 85 des 100 sociétés du classement Fortune
- Plus d'1 milliards de \$ de revenu annuel
- Plus de 4200 employés
- Plus de 8 trillions de transactions cloud par mois à travers tous les écosystèmes

© Copyright 2019 Informatica LLC.



Informatica®

AI driven data
illuminated new
possibilities





IRi
Growth delivered.



IRI, leader mondial dans les domaines du big data des analyses prédictives et des études prospectives, guide dans leur croissance les entreprises du secteur des Produits de Grande Consommation, dans les différents circuits de distribution (on line et off line), de la parapharmacie et de la Consommation Hors Domicile.

IRI accompagne plus de 5 000 clients à travers le monde, dans leurs prises de décisions stratégiques. Sa division Analytics Centre of Excellence (ACE), le plus grand centre d'analyse des produits de consommation au monde, traite chaque année plus de 20 billions de données.

L'INTELLIGENCE ARTIFICIELLE POUR FACILITER VOS PRISES DE DECISIONS

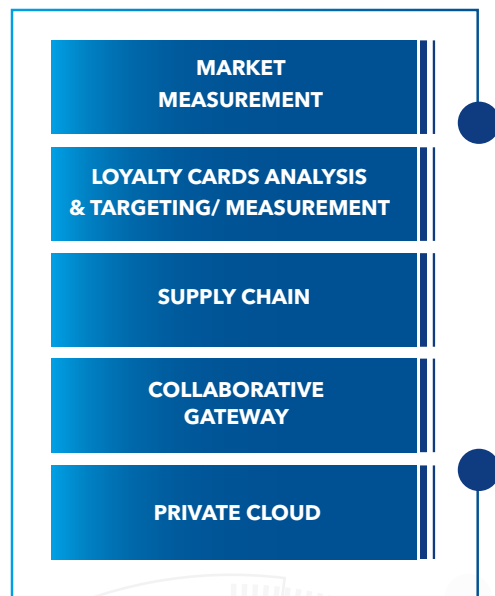
Obtenir des insights en temps réel, pour des prises de décisions rapides et avisées est devenu une priorité afin de gagner en croissance dans un environnement complexe et en perpétuelle évolution.

Nous avons initié la transformation du monde des études grâce à l'Intelligence Artificielle et le Machine Learning pour donner à nos clients, industriels et distributeurs, le pouvoir de maîtriser ce bouleversement en utilisant des processus décisionnels automatisés.

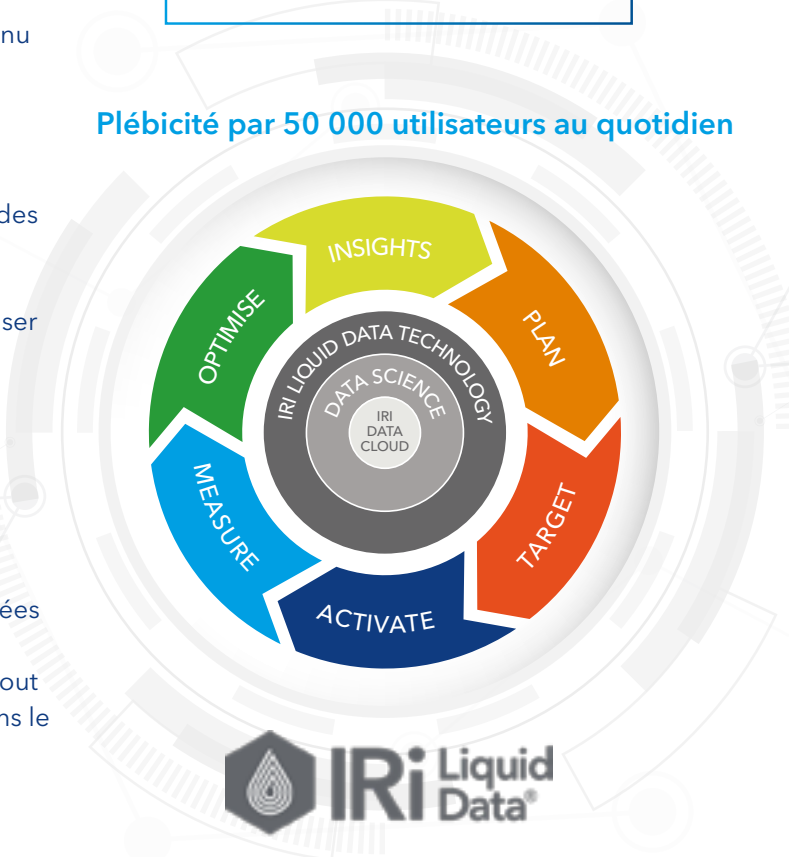
IRI a investi 1 milliard de dollars dans sa technologie de pointe pour accompagner ses clients dans ce changement.

Nous disposons de la plus large base de données intégrant les ventes des distributeurs, les promotions associées, les cartes de fidélité le tout intégré dans notre plateforme multi-source dans le Cloud, IRI Liquid Data®, unique sur le marché.

Technologie et Consulting Supportés par IRI Liquid Data® Au service de :



Plébiscité par 50 000 utilisateurs au quotidien





Automated Insights Augmented Decisions Automated Execution

Boost your growth with **AUTOMATED INSIGHTS** and **AUGMENTED DECISION MAKING** solutions powered by IRI artificial intelligence, machine learning and the most advanced technology as of today.

Learn more at IRIworldwide.com



IRi

Growth delivered.



IVIDATA – An Artificial Intelligence Company

IVIDATA est un Groupe de consulting spécialisé dans le BIG DATA et l'Intelligence Artificielle. Il accompagne les entreprises dans leur transformation digitale en proposant une large gamme de services couvrant l'ensemble de la chaîne de valeur du Big Data, de l'élaboration de la stratégie des données jusqu'à l'implémentation technique, en passant par la Data Science (science des données) et la Data Privacy (confidentialité des données).

Créé en 2013, le Groupe emploie 450 collaborateurs et réalise un chiffre d'affaires 2019 de 45 millions d'euros.

Une expertise 360...

L'ensemble des services proposés en one-stop-shop par IVIDATA est structuré autour de 3 domaines d'expertise, qui s'articule comme un cluster multi-disciplinaire unique :

- La Stratégie : ce pôle regroupe la partie Data Strategy et Data Science ;
- Le Consulting BIG DATA : ceci inclue l'ensemble de notre offre BIG DATA (Data Engineer, Architecte Big Data, Consultant Cloud...);
- Le département vertical Life Sciences : accompagnement en biométrie, biostatistique et essais cliniques pour les professionnels de l'industrie pharmaceutique, cosmétique et agro-alimentaire.

IVIDATA propose ainsi une offre BIG DATA 360, de la stratégie à l'utilisateur final.

... Sur des secteurs ayant des problématiques fortes en matière de data

Le portefeuille de clients d'IVIDATA est particulièrement diversifié. Le Groupe accompagne notamment plus de 100 sociétés de premier plan, et entretient des relations historiques avec de grands laboratoires pharmaceutiques.

L'expertise IVIDATA s'adresse aux organisations de toutes tailles évoluant principalement sur les secteurs stratégiques où les enjeux liés aux données sont cruciaux : santé, pharmacie, banque et assurance, finances, grande distribution, énergie, services...

Des consultants experts

L'une des forces d'IVIDATA réside dans le savoir-faire, l'agilité, les connaissances et l'expertise de ses consultants. Le Groupe concentre en effet plusieurs communautés d'experts et propose régulièrement des meetups autour de thématiques métiers afin de continuellement être à la pointe de l'excellence. Il s'est doté de sa IviAcademy, un espace dédié à la formation pour que ses collaborateurs puissent sans cesse monter en compétence et accompagner les clients dans tous les aspects de leur transformation.

Le bien-être au travail, une priorité pour IVIDATA

De nos jours, les collaborateurs ont davantage d'opportunités, leurs besoins et environnements ont évolué, les sphères de la vie personnelle et professionnelles sont de plus en plus proches. Arriver sur le marché du travail dans la Tech ou le Web, c'est avoir de nombreux endroits où l'on peut possiblement s'épanouir.

C'est pourquoi IVIDATA a créé le label IviWell dédié au bien-être de ses salariés. Ce label matérialise l'engagement Wellness du Groupe vis-à-vis de ses collaborateurs. A travers celui-ci, IVIDATA organise des événements où leurs collaborateurs peuvent se ressourcer autour de différentes activités telles que le yoga ou encore des cours de cuisine... IVIDATA vise ainsi à fidéliser ses collaborateurs et à construire une équipe de consultants soudée et attachée à leur entreprise.

Un fondateur doté d'un esprit d'entrepreneuriat

Diplômé d'HEC et de l'Ecole Centrale Paris, Étienne Aboulker est le co-fondateur d'IVIDATA.

Après avoir débuté sa carrière dans la finance et le trading, il se lance dans l'entrepreneuriat en 2007 et fonde EABI Consulting. Cette société de services spécialisée dans les domaines de la Finance de Marchés et des Risques deviendra en 2012 une filiale à 100 % de ManpowerGroup.

Passionné par les technologies et les problématiques SI, il co-fonde IVIDATA en 2013 avec pour ambition de faire de ce Groupe le premier cabinet de conseil dédié au Big Data et à l'Intelligence Artificielle en Europe.



Comment voyez-vous l'évolution du marché du Big Data ?

Le BIG DATA est le cœur de notre métier depuis la création d'IVIDATA il y a 6 ans. Au démarrage, les besoins clients autour du BIG DATA étaient très sporadiques ; il y a ensuite eu une accélération avec la création des Data Lakes, notamment dans le domaine bancaire. Depuis maintenant 3 ans, toutes les entreprises traitent de sujets Big Data, avec des niveaux de maturité hétérogènes en fonction des secteurs et des clients. Le Big Data regroupe plusieurs sous-segments, que ce soit la Data Science, l'Architecture IT, ou la Data Privacy.

Dans ce domaine, quels sont les principaux défis auxquels les entreprises et organisations sont confrontées ?

Il y a deux types de défis principaux. Le premier est d'ordre organisationnel : du fait que les projets Big Data sont souvent transverses, cela oblige les entreprises à repenser l'organisation de leurs équipes. Ces dernières ne sont plus dédiées à un seul domaine de compétence, mais sont devenues multi-compétences ; on peut y retrouver aussi bien un profil IT, un profil plus scientifique, et un profil marketing.

Le deuxième enjeu est celui du recrutement de nouveaux talents autour des sujets du Big Data et de l'Intelligence Artificielle. Le process pour recruter de tels profils doit être très agile et court, ce qui rend la tâche plus difficile pour certaines entreprises d'une certaine taille. Enfin, il y a une dimension innovation forte ; il faut pouvoir suivre l'évolution technologique pour être à même d'amener les projets à bon port.

A quels enjeux stratégiques et opérationnels répond IVIDATA pour les aider à relever ces défis ?

IVIDATA propose une offre unique. Nous avons monté un véritable cluster multidisciplinaire autour du BIG DATA. Notre force est de proposer des expertises fortes, et qui ont la capacité d'interagir.

Cette offre BIG DATA 360 traite les sujets depuis la Data Strategy jusqu'au end-user, en passant par toutes les étapes IT (Architecte Big Data, Data Engineer...), scientifique (Data Science, DataViz...) et Business (Product Owner...).

IVIDATA entend devenir le leader européen du Big Data... comment comptez-vous y parvenir ?

A ce jour, il n'existe pas de véritable leader dans le Consulting Big Data au niveau européen ; il y a soit des acteurs de niche, soit de grandes sociétés technologiques qui ne sont pas natives du Big Data.

Nous devons relever 3 défis pour accéder à cette ambition de leader européen : une dimension de taille (nous devons à minima faire un chiffre d'affaires de 100 millions d'euros, soit doubler de taille), une dimension People en mettant en place tous les éléments différenciateurs nécessaires aux recrutements de collaborateurs très recherchés, et enfin une dimension internationale en ayant une empreinte européenne. Tous ces enjeux seront traités avec de la croissance organique, ou par des acquisitions stratégiques.



JEMS est une société de conseil opérationnelle, intégrateur de solutions Cloud et Big Data. Le groupe a développé une position de leader autour du cycle de vie de la donnée.

Créé en 2002, JEMS est un pionnier de l'industrie 4.0. Avec plus de 500 experts, plus de 80 projets Big Data ont ainsi été réalisés. JEMS travaille pour les grands comptes du CAC 40 sur l'ensemble des verticales métier : finance, industrie, énergie, retail, luxe, automobile, agriculture ou santé. Son siège est à Neuilly-Sur-Seine.

QUE PROPOSE JEMS ?

La 4ème révolution industrielle, celle de la donnée et de l'économie collaborative oblige les entreprises à accélérer radicalement leur transformation digitale. La mission de JEMS est de répondre à ces grands enjeux en valorisant le patrimoine donnée et l'expérience client. La donnée est comme un minerai. Notre métier consiste à l'extraire, la mettre en forme et à la transformer pour créer de nouveaux services. Pour cela nous avons mis en place une vision en 4D :

#4D : Data : pour organiser, extraire et transformer vos données

#4D : Digital pour faciliter et développer les cas d'usages

#4D : Devops, SysOps et Agile pour accélérer vos projets

#4D : Design Thinking pour savoir que faire de vos données

JEMS ET LA COMMUNAUTE CLOUD

Le Cloud Computing est devenu une plateforme essentielle pour l'innovation. JEMS l'a bien compris en développant non seulement une expertise sur la majorité des fournisseurs de solutions, Google, Microsoft et Amazon mais aussi en lançant en 2019 sa première communauté Cloud.

Espace de partage et de collaboration, cette communauté a créé un espace unique où chacun des membres contribue à la formation et à la certification des autres.

POURQUOI CHOISIR JEMS ?

Parce que nous savons de quoi nous parlons. JEMS existe depuis 17 ans et a évolué au rythme des nouvelles technologies : B.I, technologies web, Big Data, Design Thinking. Nous avons mené plus de 80 projets Big Data d'envergure.

Parce que nous sommes un partenaire sûr. Nous avons établi des partenariats de long terme avec nos clients car ils nous ont fait confiance. JEMS fait également partie des 0,2% des ESN certifiées ISO27001 sur la sécurité des systèmes d'information.

Parce que nous nous adaptons en permanence. L'agilité est une des valeurs de l'entreprise sur laquelle nous évaluons nos salariés. Nous sommes en permanence en train de les former afin d'offrir les meilleures expertises en Data, Cloud Computing ou Expérience Client.

Solutions Data intégrées

Cloud AWS | Cloud GCP | Cloud Azure | Talend | Snowflake | Confluent | Tibco | Datastack | Elastic | Toucan Toco
Dataiku | Tableau | MAPR | Cloudera | Qlik | Sense | Alteryx | Outscale | Attunity | NetApp

#Data

#Digital

#Devops

#Design
Thinking

Vous avez les données,
nous avons le savoir-faire



RÉCUPÉRER
LA DONNÉE



RÉPONDRE
RAPIDEMENT



OPTIMISER
LA DONNÉE



CRÉER LES
CAS D'USAGE

Data
DevOps
Digital
Design Thinking



KEYRUS, CRÉATEUR DE VALEUR À L'ÈRE DE LA DATA & 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 et de compétitivité.

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
Transformation Digitale
Pilotage de la Performance
Accompagnement des Projets

DATA INTELLIGENCE
Data Science
Intelligence Artificielle
Big Data & Cloud Analytics
Business Intelligence
EIM
CPM/EPM

DIGITAL EXPERIENCE
Innovation & Stratégie Digitale
Marketing Digital
DMP & CRM
Commerce Digital
Performance Digitale
User Experience

NOS PARTENAIRES

Big Data & NoSQL



Cloud



Data Science



CHIFFRES CLÉS



273,2 M€
CA 2018



+ 3 300
collaborateurs
dans le monde



18
pays



+ 350
experts dédiés
à la data en France

Penser chaque projet de valorisation des données comme un projet transversal : collecte des données, enrichissement & exploitation par un processus opérationnel métier.

Constituer une équipe de « Business partner » pour accompagner les métiers sur toutes les phases des projets de valorisation des données.

Construire un catalogue de services « data » pour assister les métiers lors de la construction des use cases.

Se doter des compétences éventuellement manquantes : Data Ingénieur, Data Scientist & spécialiste de la gouvernance des données.

Faciliter l'accès pour les métiers à la donnée, en mettant en place : un catalogue et un dictionnaire des données & des capacités de data visualisation.

Organiser & gouverner la « gestion » des données (maîtrise de la sécurité, la disponibilité et la qualité des données).

Urbaniser les référentiels.

Penser & moderniser l'architecture data dans son ensemble : transport et exposition des données, stockage et consommation (data visualisation et datalab).

**DEVENEZ
BUSINESS
DATA-DRIVEN**

ENTREZ DANS LA DIMENSION DATA

Data For Business



BUSINESS DATA-DRIVEN COMPANY

Élaborer votre stratégie data :

- ▶ De votre vision stratégique métier jusqu'au portefeuille dimensionné des projets à réaliser
- ▶ Définition de votre architecture data cible



GOVERNANCE DE LA DONNÉE

- ▶ Définir la gestion de vos données maîtrisées
- ▶ Garantir la qualité, sécurité et disponibilité de vos données
- ▶ Déployer l'organisation, les compétences et les outils associés



VALORISATION DES DONNÉES MÉTIER

- ▶ Concevoir des tactiques d'extraction de la valeur issue de vos données
- ▶ Valeur marchande, valeur d'optimisation opérationnelle, valeur d'actif stratégique

Data IS Transformation



CONSTRUCTION ARCHITECTURE DATA

- ▶ Mettre en place une architecture data complète
- ▶ Maîtrise des référentiels & silos de données
- ▶ Transport & exposition de vos données : stockage, enrichissement, consommation et visualisation



TRANSFORMATION MODEL DELIVERY

- ▶ Sortir de la relation client/fournisseur appliquée en BI
- ▶ Mettre en place un dispositif de Business Partner avec un Delivery Data Agile

Artificial Intelligence Speeding-up



SOLUTIONS IA MÉTIER DÉPLOYABLES

- ▶ Création des calculs avancés de valorisation de la donnée
- ▶ Industrialisation & déploiement de ces solutions métier



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 blocs 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 fichiers 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 couplé à une pénurie de data engineer entraine 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 projets 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 intégrés 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@lansrod.com / +33 6 58 78 10 95

LINCOLN

MAXIME BOURSIN

Directeur du développement

+33 (0)1 46 99 36 81

maxime.boursin@mel.lincoln.fr

4 rue d'Anjou 92100 FRANCE

www.lincoln.fr

Analytique

API

Applications

Consultants/SSII
Intégrateurs

Dataviz

Infrastructure

Institutionnels

LINCOLN
LIGHT UP YOUR DATA



250 PASSIONNÉS DE LA DATA

+85 ENTREPRISES NOUS FONT
CONFIANCE

+16% DE CA EN 2019

VENEZ DÉCOUVRIR NOS NOUVELLES OFFRES
DATA DIGITAL & IA

LINCOLN
LIGHT UP YOUR DATA

INTELLIGENCE ARTIFICIELLE

BI – BIG DATA

DATA DIGITAL

DATA SCIENCE

LA SOCIÉTÉ DE CONSEIL QUI VOUS ACCOMPAGNE DANS TOUS
VOS PROJETS DATA DRIVEN

www.lincoln.fr

Suivez les actualités LINCOLN sur nos réseaux





L'EXPERTISE IA ET DATA DIGITAL AU SERVICE DE NOS CLIENTS

Quelles sont les ambitions de LINCOLN ?

Dans un contexte de transition digitale, les organisations cherchent à se transformer.

Pour faire face à ces nouveaux usages et la valorisation de ces multiples données, LINCOLN a développé depuis maintenant 2 ans, une feuille de route Data Digitale à l'échelle de l'entreprise.

Ce plan « Light Up 2020 » nous a permis de poser les bases de notre stratégie, pour devenir un cabinet de Conseil Data Digital.

Notre objectif est de faire le lien entre nos expertises Data (Big Data, Data Sciences, Analytique) avec les évolutions que nous offrent le Digital.

Avec des convictions, apporter à nos clients du delivery, du conseil, de l'innovation avec un ROI à la clé.

L'apport de valeur de LINCOLN pour ses clients ?

Notre engagement vis-à-vis de nos clients est de construire des partenariats durables et évolutifs avec eux. Permettant ainsi de relever leurs défis opérationnels et de garantir des engagements de qualité et des gains associés.

Nous avons donc mis en place une Direction Conseil qui va accompagner nos clients dans leurs enjeux et une Direction de l'innovation pour développer des cas d'usages notamment autour de l'Intelligence Artificielle.

Une excellence de nos consultants que nous accompagnons avec un parcours de formation adapté et des certifications vous garantissant un haut niveau de services.

Votre vision du marché de la data ?

La Data représente un actif essentiel de l'entreprise aujourd'hui. La transformation digitale et les avancées technologiques ont fait évoluer les habitudes des utilisateurs dans la gestion des données.

Ils souhaitent à présent collecter, manipuler, exploiter, visualiser la Data avec une volonté très forte de produire de la valeur. C'est dans cette trajectoire que nous affirmons notre positionnement Data Digital.

MARIADB CORPORATION AB

LAURENT VIDAL

Regional Sales Director EMEA

+33 (0)6 10 25 92 06

laurent.vidal@mariadb.com

Tekniikantie 12, 2150 Espo FINLAND

mariadb.com/fr



MariaDB



LES ENTREPRISES FONT CONFIANCE À MARIADB

Libérez-vous des coûts, des contraintes et de la complexité des bases de données propriétaires et réinvestissez dans l'essentiel : développer le plus rapidement possible des applications et des services innovants. Bénéficiez des mêmes avantages que des clients tels que Deutsche Bank, DBS Bank, Nasdaq, Red Hat, ServiceNow, Verizon et Walgreens. Ces chefs de file de l'industrie font confiance à MariaDB pour offrir une agilité opérationnelle inégalée, fournir une fiabilité d'entreprise, favoriser une innovation collaborative et font confiance à MariaDB, qui satisfait aux mêmes exigences essentielles que les bases de données propriétaires, mais pour une fraction du coût.

POURQUOI MARIADB ?

MariaDB utilise des moteurs de stockage spécifiques enchassables, pour prendre en charge des charges de travail qui nécessitaient auparavant diverses bases de données spécialisées. La complexité et les contraintes étant éliminées, les entreprises peuvent désormais compter sur une seule base de données complète pour tous leurs besoins, tant sur un environnement sur site ou dans le Cloud de choix. Déployée en quelques minutes pour des cas d'utilisation transactionnels ou analytiques, MariaDB offre une agilité opérationnelle inégalée sans sacrifier les fonctionnalités clés de l'entreprise, notamment la conformité ACID et le SQL. MariaDB répond aux mêmes exigences essentielles que les bases de données propriétaires pour un coût nettement inférieur. Pas étonnant qu'elle soit la base de données open source à la croissance la plus rapide du marché.

LA VALEUR D'UN ABONNEMENT D'ENTREPRISE À LA PLATEFORME MARIADB

Les entreprises ont besoin d'une alternative complète aux modèles onéreux et restrictifs de gestion des licences proposés par Oracle, Microsoft et IBM. MariaDB est la base de données open source à la croissance la plus rapide du marché et est soutenue par une communauté d'utilisateurs dynamique. MariaDB Platform, disponible uniquement par abonnement, est une extension du système de base de données open source populaire MariaDB Server, avec une gamme complète de fonctionnalités et de services d'entreprise qui vous offrent la fiabilité requise pour les déploiements critiques.

Conçue pour des entreprises, MariaDB Platform complète MariaDB Server avec la haute disponibilité, les performances, la sécurité et l'analyse auxquelles les entreprises modernes font confiance. Lorsque vous gérez une entreprise sur MariaDB Platform, les personnes qui ont conçu le logiciel sont présentes pour garantir votre succès.

Pour les applications transactionnelles plus exigeantes en matière d'analyse, MariaDB Platform peut être déployée dans une configuration hybride en choisissant à la fois les stockages en ligne et en colonne.

Un abonnement à MariaDB Platform fournit tout ce dont vous avez besoin pour exécuter vos applications critiques. Mais cela va bien au-delà du logiciel d'entreprise. Vous avez également accès à des mises à jour régulières, à une communauté dynamique et à des services de support pour l'ensemble de votre infrastructure d'applications, de votre architecture et du cycle de vie.



Déverrouillez toutes vos ressources hybrides avec MariaDB Platform X3

Traitement transactionnel et analytique hybride (HTAP)

Déployez MariaDB Platform X3 sur une infrastructure de cloud privé et public pour assurer votre future évolutivité. Utilisez-la pour gérer aussi bien les transactions que les analyses, et créez des modèles de données relationnels/JSON hybrides, vous garantissant une grande flexibilité.

mariadb.com/fr



Baker Hughes, une société de GE (BHGE), développe un logiciel de maintenance prédictive 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

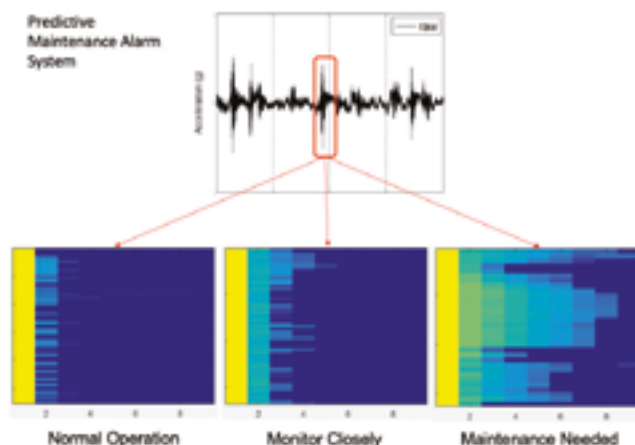
En travaillant avec MATLAB®, les ingénieurs de BHGE ont développé un logiciel de surveillance du bon fonctionnement des pompes qui applique des techniques de machine learning en temps réel pour prédire l'instant idéal pour effectuer la maintenance. Ils ont traité et analysé jusqu'à un téraoctet de données 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 réduit 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 devions 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 ».



DEEP LEARNING ET INTELLIGENCE ARTIFICIELLE AVEC MATLAB®

“ La vision à long terme des produits MathWorks a toujours consisté à faire de l’IA un domaine accessible pour tous, à mettre à disposition des produits interopérables et à créer des outils simples à déployer, aussi bien dans le cloud que dans d’autres environnements informatiques. ”

Prashant Rao, Technical Manager, MathWorks India

```

Connect to the camera
camera = webcam;

Load the neural network Alexnet
nnet = alexnet;

Make prediction in live
while true
    img = camera_snapshot;
    img = imresize(img,[227,227]);
    label = classify(nnet, img);
    image(img);
    title(char(label));
    drawnow;
end
    
```

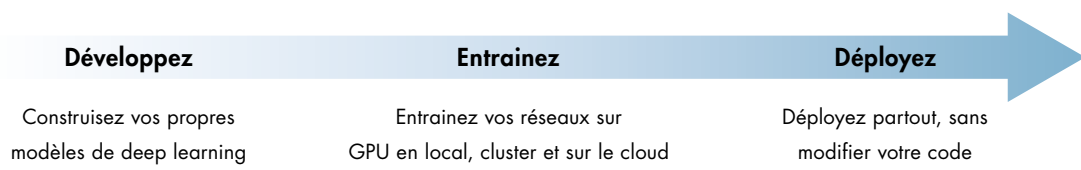
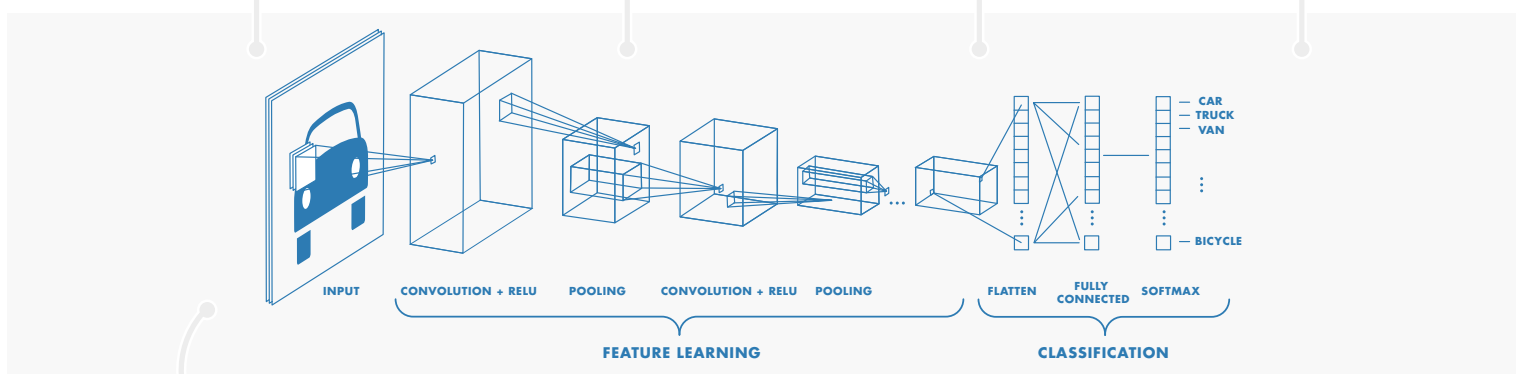
Deep learning en 10 lignes de code

Construisez et modifiez facilement vos réseaux de neurones de manière interactive avec l’App Deep Network Designer

Importez les derniers modèles pré-entraînés, comme GoogLeNet ou ResNet-101

Intégrez des réseaux complexes, comme les LSTM ou les DAG

Etiquetage automatique des données



MICROPOLE

EMMANUEL MANCEAU

Directeur Data Intelligence et Big Data

+33 (0)1 74 18 74 18

emanceau@micropole.com

91 - 95 rue Carnot 92300 FRANCE

www.micropole.com



MICROPOLE



MICROPOLE, GROUPE INTERNATIONAL SPÉCIALISÉ EN CONSEIL ET TECHNOLOGIES INNOVANTES : DATA & DIGITAL.

Trois offres complémentaires au service de la performance des directions métiers :



DIGITAL EXPERIENCE

Les équipes digitales Micropole & WIDE allient expertise conseil auprès des CMO et CDO, expérience client, excellence technologique et design.

Nos 1250 consultants, designers, architectes, dataengineers et data-scientists vous aident à générer et capter les données qui vous sont utiles, à les maîtriser et les contrôler, et à en tirer de plus en plus de valeur pour les métiers et les instances de direction.



DATA INTELLIGENCE & PERFORMANCE

Spécialiste data et analytique, Micropole rend les entreprises data intelligentes, et nos 250 projets réalisés font de nous les leaders reconnus de la performance financière.

Grâce aux réseaux planétaires ou locaux, le Big Data est la résultante naturelle de la transformation numérique de toutes les interactions et de tous les processus. C'est un atout.



DATA GOVERNANCE & ARCHITECTURE

Nous bâtissons le socle nécessaire à la mise en œuvre des transformations digitales et organisationnelles des entreprises. Un acteur majeur en Europe avec plus de 70 projets réalisés.

La technologie rend (presque) tout possible, mais c'est la qualité du conseil, adossée à sa capacité d'opérer et d'industrialiser, qui garantissent la pertinence économique et éthique de l'utilisation de vos projets Data. Quand les systèmes ne suffisent plus : c'est la complémentarité et l'agilité des équipes qui fait la différence.

**UN SPÉCIALISTE
DATA & DIGITAL
AU POSITIONNEMENT
UNIQUE**

Cabinet de conseil
Agence digitale
ESN spécialisée

CUSTOMER
CENTRIC

**#INNOVATIVE
PEOPLE
CREATING
VALUE**

Groupe international spécialisé
en conseil et technologies innovantes

MICROPOLE

COMMENT ADRESSER

SEREINEMENT / EFFICACEMENT / DURABLEMENT

LES

USAGES

ACTUELS

FUTURS

DES

DATA

INTERNES

EXTERNES

DE VOTRE ENTREPRISE ?

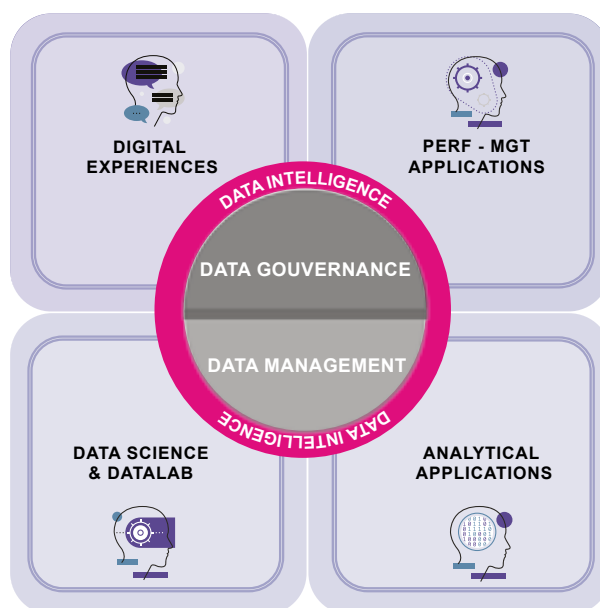
La « matière data » a la côte ! Et c'est tant mieux ! Chez Micropole nous aidons nos clients à la valoriser depuis plus de 30 ans.

Autant dire que nous en étions convaincus ... et aussi un brin désespérés de constater que nous semblions les seuls à militer pour une organisation durable de l'entreprise autour du patrimoine data, bien au-delà des querelles et débats technologiques.

PASSEZ EN MODE DATA INTELLIGENCE !

L'entreprise « data intelligente » considère les data comme un asset clairement partagé entre les métiers, demandeurs de valeur, et les informaticiens, organisateurs des données et moteurs de leurs échanges applicatifs. Les rôles sont distribués et les responsabilités clairement établies. Le rôle d'un Chief Data Officer est porté au niveau des organes de direction, et tous les métiers sont impliqués dans la maîtrise des données qui les concernent.

Le portefeuille des usages de la data est géré comme un chantier d'entreprise commun tiré par l'amélioration continue de la performance ou le contrôle des risques.



Vous vous sentez concernés ?

Ces colonnes sont trop étroites pour exposer toutes nos convictions et nos savoir-faire. Le Salon Big Data de Paris est devenu l'endroit privilégié pour parler Data, tout simplement. Qui peut le plus, peut le moins : le Big Data a mis en exergue la nécessité pour les organisations de se mettre en capacité de traiter TOUTE la data sous TOUTES ses formes. Donc, d'un point de vue technique, il est devenu évident que tout est possible ... mais que voulez-vous vraiment achever dans votre chantier data ?

**VOUS SOUHAITEZ EN SAVOIR BEAUCOUP PLUS ?
VENEZ RENCONTRER NOS EXPERTS !
ou prenez contact : info@micropole.com**



TF1 optimise la connaissance et l'expérience de ses utilisateurs grâce à la data et à l'IA

Retour sur l'interview de Philippe Toublant, directeur adjoint data chez TF1

Contenus personnalisés, publicités ciblées, prévention du churn... Au sein du groupe TF1, une équipe « danse avec les data » pour améliorer la connaissance de ses clients, leur proposer une expérience optimisée... et accroître la consommation.

La problématique de TF1

Comment mieux connaître les utilisateurs et leurs habitudes de consommation pour leur proposer des contenus qui répondent à leurs attentes ?

TF1 est la première chaîne de télévision française généraliste. Aujourd'hui, le groupe TF1 compte de nombreuses chaînes (TMC, LCI, TFX...), des sites Internet (Aufeminin, Marmiton, Doctissimo...), des plateformes de vente à distance (Téléshopping), une régie publicitaire, etc.

Avec myTF1, une plateforme qui permet de regarder des vidéos en replay ou à la demande, le groupe compte bien s'adapter à la demande croissante de consommation personnalisée de sa clientèle.



Comment TF1 optimise sa connaissance de ses utilisateurs grâce à Microsoft Azure ?

Historiquement, la chaîne a toujours reçu chaque jour les chiffres d'audience de Médiamétrie, des données qui lui permettent d'adapter sa grille aux goûts de ses spectateurs. Mais, depuis 2016, TF1 est passé à la vitesse supérieure, en constituant une équipe d'experts de la data, capables de la traiter, de l'analyser et d'en tirer des enseignements précieux.



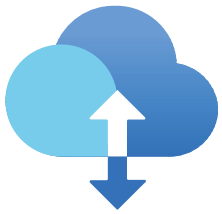
Une équipe d'experts dédiée à la data



Chez TF1, plusieurs profils experts exploitent les 1 à 2 millions de données collectées chaque jour :

- 5 data engineers conçoivent et maintiennent les architectures
- 4 data scientists transforment la donnée brute en information exploitable en lui donnant du sens
- 2 data analysts classent les données dans différentes segmentations, avec une vision plus marketing

Ces trois métiers sont complémentaires et permettent de traiter la donnée tout au long de son cycle de vie, et d'établir des modèles prévisionnels à l'aide du machine learning.



Une plateforme 100 % cloud et un écosystème responsive

TF1 ne voulait pas être coincé par le type de format des données, ni par la capacité de stockage. C'est pourquoi le groupe a opté pour une plateforme 100% cloud : Microsoft Azure, qui permet de traiter et d'avoir accès à l'information à n'importe quel moment.



Azure peut absorber des données collectées depuis de multiples sources : les différentes entités de TF1, les annonceurs du groupe, ou encore les éditeurs partenaires. Ouvert sur l'écosystème, Azure permet à TF1 de travailler avec de nombreuses technos autour de la datascience (Apache Spark, Databricks, Hadoop...). Une fois valorisée, la donnée peut ensuite être activée auprès d'éditeurs, de régies ou encore de partenaires e-commerce.

Philippe Toublant



Répondre aux attentes de ses clients, choisir les bons sponsors, diffuser de la publicité ciblée...

Les données sont centralisées sur une plateforme de gestion (DMP, data management platform). D'un point de vue structurel, l'écosystème est responsive : la donnée est récupérée sur la DMP et il est possible de faire des backflows vers la DMP pour renvoyer des informations. Azure allume et éteint les clusters, et déclenche les scripts de storage.

Les bénéfices de l'exploitation de la data sont nombreux pour TF1. Pourtant, la première année, cette activité n'était pas rentable pour le groupe. La direction a cependant fait confiance aux spécialistes de la data, et désormais c'est une très belle croissance qui s'annonce.



Il est impossible de réussir à traiter la donnée sans savoir investir, note Philippe Toublant. Notre équipe a coûté de l'argent la première année. Mais notre direction a cru en nous et nous a donné la chance de continuer. Nous sommes actuellement en progression constante et nous apportons des bénéfices importants. Nous allons étendre le projet aux émissions e-commerce comme Téléshopping.

Philippe Toublant



Microsoft Azure a ainsi permis à TF1 de stocker et analyser d'immenses volumes de données en temps réel pour mieux connaître ses utilisateurs et optimiser l'expérience client.

Découvrir nos solutions Azure sur : aka.ms/azurefr

 Microsoft Azure



GOUVERNER SA DONNÉE : LE PROCESSUS ET LA PRATIQUE

D'après Roxane Edjlali,

Senior Director Product Management chez MicroStrategy

Gouverner sa donnée est une nécessité réglementaire certes mais aussi économique. Pour que cela ne devienne pas un sacerdoce, il existe un processus et une pratique à mettre en œuvre. Sachons adapter notre gouvernance aux usages des collaborateurs pour garantir l'exploitation de la donnée en toute confiance.

Conflit de loyauté

Avec la prolifération des données recueillies ou produites en entreprise, stockées dans les bases de données décisionnelles, les applications ou le Cloud, l'on s'aperçoit que les opportunités d'analyses mais également les besoins de traitement émis par les métiers et les organisations ne font que croître. Alors que les utilisateurs métier réclament toujours plus d'autonomie dans l'exploration et l'usage de la data, un conflit entre les attentes et les questions cruciales de gouvernance se dessine.

D'un côté, une forte volonté de renforcer ses qualités data-driven et de se différencier, des espoirs d'innovation qui reposent tout à la fois sur les analyses à créer et les atouts de l'Intelligence Artificielle. Et de l'autre, une réalité, contrainte principalement par les réglementations, quelque peu différente des demandes, voire des exigences des utilisateurs métier.

L'information et la donnée deviennent des enjeux géopolitiques, réglementés par des États soucieux de garder le contrôle dans un monde numérique en évolution permanente. Au cœur du cyclone réglementaire, les entreprises, parfois paralysées, oscillent entre respect des textes et opportunités business et doivent gérer une tension grandissante au sein de leurs services.

Pourtant, il n'y a pas lieu d'alimenter un quelconque conflit. Au contraire, marier gouvernance et exploitation de la donnée permet à la fois de s'assurer de sa conformité réglementaire et d'une exploitation utile de son information disponible. Comment ? En adaptant correctement son modèle de gouvernance d'une part aux données dont on est dépositaire et d'autre part aux utilisateurs.

Les usages à la loupe

Mais qu'est-ce qu'est exactement la gouvernance de la donnée ? C'est un processus constant, évolutif et quantifiable, qui vise à déterminer avec précision quelles données bénéficient de quel niveau de gouvernance et de certification pour un usage défini dans un contexte particulier.

Voici pour une définition, laquelle laisse entrevoir la démarche à entreprendre. L'entreprise pilotant ses objectifs à l'aune de ses données devra inévitablement s'intéresser de près à ses utilisateurs. Certains, peu formés, sont essentiellement des consommateurs d'information. D'autres, plus autonomes et plus au fait des usages acceptables de la data, disposeront de plus de latitude.

On le comprend aisément, le niveau d'autonomie et de gouvernance sont complémentaires et inversement proportionnels. Ainsi, toute donnée exposée à des tiers via le site internet ou une API est gérée sous le niveau de gouvernance le plus élevé, puisqu'on ne peut supposer a priori que ces tiers sauront convenablement traiter l'information. A l'inverse, en interne par exemple, les données propres à un même

département et non destinées à quitter le périmètre, ne relèvent pas, a priori, de règle de gouvernance particulière.

Sous l'égide d'un Chief Data Officer, en charge de la vision et de la stratégie Data de l'entreprise, se réunissent ainsi tous ceux et celles qui ont des attentes et des rôles fort différents, afin de mettre sur pied un modèle de gouvernance adaptable aux usages et aux besoins applicatifs. Il ne s'agit pas, évidemment, de décider d'une gouvernance globale, couvrant l'ensemble de l'information détenue, par précaution. C'est une approche collaborative et itérative qu'il faut privilégier, en adaptant le niveau de gouvernance aux usages et aux compétences de chacun au fil de la production et du temps.

Le processus et la pratique

La démarche fait émerger un processus qui permet à chacun de contribuer à la découverte des données susceptibles d'être gouvernées et d'effectuer un tri en fonction des réglementations.

Ce n'est qu'ensuite et autour de ce processus que s'inscrit la pratique supportée par la technologie. Dans ce cadre, on recommande généralement de laisser les utilisateurs avertis accéder à l'ensemble de l'information quel que soit son niveau de gouvernance. Ainsi ils débutent leurs analyses et créent leurs premiers rapports. Le monitoring de l'usage de la donnée permet alors d'identifier celles sujettes à gouvernance, puis à certification quand elles ont été complétées. Un monitoring réussi est un monitoring alimenté par une cartographie des données, des usages et des sources. De ce fait, cette collecte des usages de la donnée dans l'entreprise permet d'enrichir considérablement l'offre de services. Il devient possible par exemple de faire de la recommandation de données certifiées, à l'instar des sites marchands recommandant des produits en fonction des pages visitées. La proposition évite en outre aux utilisateurs, avertis, de réinventer sans cesse la roue.

Le grand intérêt de la pratique est sa déclinaison en self-service des données certifiées pour les consommateurs d'information, les données non certifiées étant réservées aux utilisateurs plus aguerris. Progressivement, la pratique de la gouvernance s'étoffe de niveaux. Local pour des analystes, départemental et réservé à un petit nombre, d'entreprise pour des données susceptibles d'être partagées par tous voire avec des tiers.

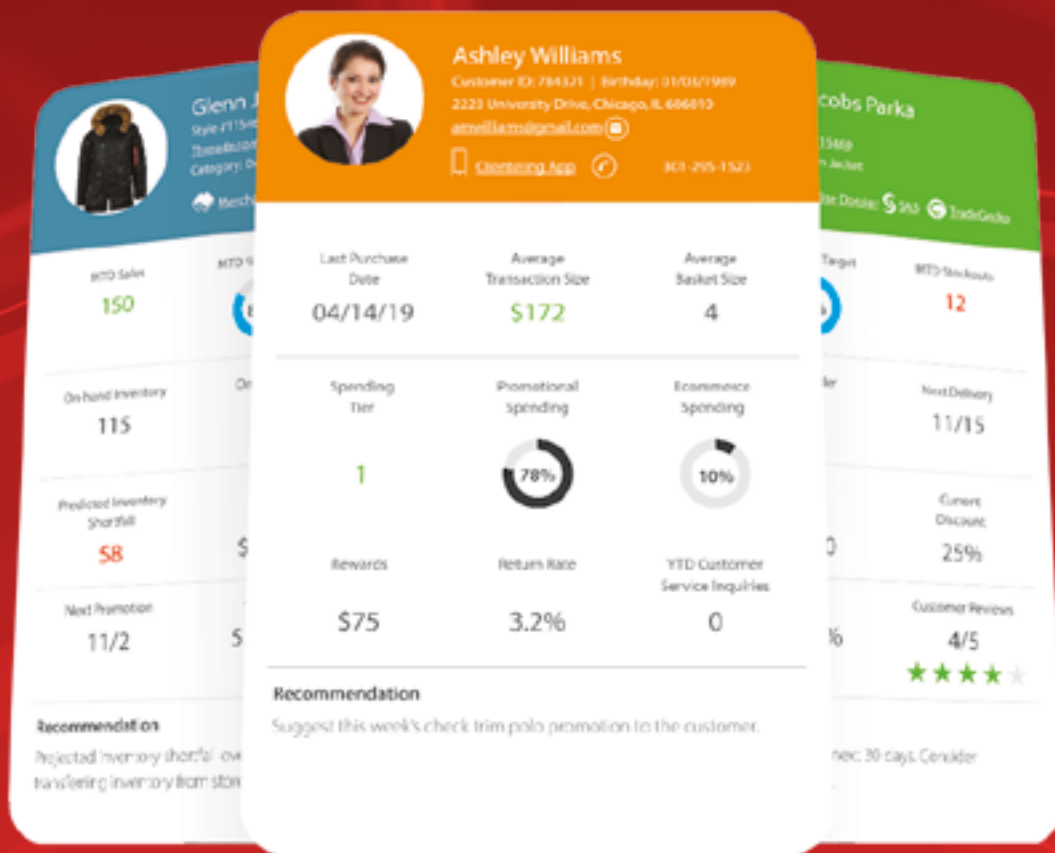
Le patrimoine de données de confiance s'enrichit et s'étend peu à peu. Parallèlement à cela, tous les utilisateurs gagnent en expertise et en autonomie car ils sont plus à même d'évaluer les usages acceptables de la donnée.

Conseils de dernière minute

A ce jour, beaucoup d'entreprises, notamment nord-américaines et ouest-européennes, ont déjà réalisé une part importante de ce travail de gouvernance et disposent donc d'un important capital de données gouvernées. Mais l'affluence des données de capteurs et de celles issues du click stream pour ne citer qu'elles contribuent à étendre le champ d'analyse, à dépasser la gouvernance traditionnelle et à faire preuve de créativité. C'est pourquoi la data-stratégie choisie doit éviter la rigidité de ses règles. Par ailleurs, en avançant pas à pas, le processus permet de juguler l'aspect parfois titanesque de la tâche. Enfin, évitons de contraindre les utilisateurs, c'est une méthode qui n'a jamais fait ses preuves. Le paysage technologique des entreprises s'est beaucoup diversifié et bien souvent les outils d'analyse sont choisis par les métiers eux-mêmes. Tant qu'une donnée est dûment certifiée, il est tout à fait possible de la mettre à disposition, de la promouvoir et de l'exploiter sur une solution tierce.

TRANSFORM YOUR ENTERPRISE IN ZERO CLICKS.

INTRODUCING HYPERINTELLIGENCE™



Smarter Analytics | Better Collaboration | Richer CRM
Insightful Websites | Intuitive Applications | Intelligent Email



SEGA HARDLIGHT MIGRE VERS MONGODB ATLAS POUR SIMPLIFIER SES OPÉRATIONS ET AMÉLIORER L'EXPÉRIENCE DE SES MILLIONS DE GAMERS MOBILES.

C'est en juin 91 que Sonic, le plus célèbre des hérissons, se lançait dans sa première course-poursuite sur nos écrans 2D pixélisés. Des consoles statiques de 1991 aux plateformes mobiles et de réalité virtuelle d'aujourd'hui, en passant par les jeux PC en ligne des années 2000, le gaming a parcouru un long chemin. Or, chose étonnante, l'infrastructure sous-jacente n'a pour sa part pas vraiment évolué au cours de ces 25 dernières années. Concrètement, elle s'articule toujours autour de bases de données relationnelles. Mais, face au besoin toujours croissant de scalabilité, de flexibilité et de créativité, les choses sont en train de changer rapidement dans l'univers du gaming. En précurseur de ce changement, SEGA HARDlight s'est converti à la culture DevOps et mise sur le service Cloud MongoDB Atlas pour offrir à ses clients une expérience de gaming à la hauteur de ses ambitions.

Des icônes sur mobiles

SEGA HARDlight est le studio de développement des jeux mobiles de SEGA – un nom qui vous dit certainement quelque chose. Basée à Royal Leamington Spa, en Angleterre, SEGA HARDlight est connue pour avoir ressuscité le petit hérisson bleu sur écran mobile. Parallèlement aux jeux Sonic, HARDlight gère le développement et l'exploitation d'autres titres célèbres comme « Crazy Taxi: City Rush » et « Kingdom Conquest: Dark Empire ».

Sonic Dash

Les premières versions des jeux mobiles comme Sonic Jump et Sonic Dash n'exigeaient aucune connexion Internet et n'intégraient aucune fonctionnalité serveur. Leur architecture relativement statique permettait aux développeurs d'exploiter une « stack » technologique interne basée sur Java et MySQL, le tout hébergé dans le data center de SEGA HARDlight. À chaque nouvelle version, les serveurs faisaient l'objet de stress tests jusqu'à leur point de rupture. Les ressources étaient ensuite provisionnées en fonction du point de défaillance acceptable. Or, non seulement une telle approche limitait les fonctionnalités applicatives, mais elle entraînait aussi des pannes à répétition lorsque les seuils de saturation étaient atteints. Alors que la « sonicmania » semblait à nouveau s'installer et que les nouvelles fonctionnalités en ligne se multipliaient, la « stack » traditionnelle commençait à craqueler.

Une adoption massive, des pics de trafic

Les jeux mobiles suivent un modèle très intéressant. En général, les joueurs affluent en masse à la sortie d'une nouvelle version. Ils se comptent même parfois par millions en quelques jours, voire quelques heures, pour les jeux blockbusters. Ensuite, après la ruée initiale, les affluences retombent vite pour se stabiliser aux seuls vrais fans du jeu. Côté infrastructure, ces pics de trafic sont un vrai casse-tête.

Une scalabilité supersonique

Pour SEGA HARDlight, il était clair que les fonctionnalités en ligne étaient appelées à se multiplier et les pics de trafic à s'accroître. Côté architecture de la base, la structure des données spécifiques aux comptes joueurs entraînait mal dans le schéma lignes-colonnes des bases de données relationnelles. Pour remédier à ces problèmes, l'équipe s'est mise en quête d'alternatives. Elle s'est d'abord penchée sur les bases de données Cassandra et Couchbase. Or, l'une comme l'autre se révélaient trop complexes à gérer et n'offraient pas le support nécessaire pour permettre à l'entreprise d'honorer

ses engagements SLA. Les ingénieurs HARDlight se sont donc tournés vers MongoDB Atlas, le service DBaaS de MongoDB.

S'en est alors suivi une batterie de tests sur différents critères (coûts, maintenance, monitoring, sauvegarde, etc.). Pour l'équipe HARDlight, il ne faisait aucun doute que la solution MongoDB offrait la scalabilité et la flexibilité nécessaires au traitement non-stop de gros volumes de données. Toutefois, elle devait aussi s'assurer de son efficacité au niveau opérationnel.

Grâce aux outils MongoDB Atlas, un seul binôme DevOps a suffi à traiter et exécuter des jeux pour des millions de gamers. Pour la maintenance, les mises à jour et les sauvegardes, MongoDB s'occupe de tout. La fonction de sauvegarde et de restauration PITR, qui permet de revenir à un point antérieur précis en un seul clic, a d'ailleurs été un facteur de choix décisif pour HARDlight. Avec MongoDB Atlas sous AWS, SEGA HARDlight avait donc toutes les cartes en main pour offrir aux gamers la scalabilité nécessaire aux niveaux de jeu les plus élevés.

« Chez HARDlight, nous avons à cœur de trouver les outils taillés pour nos exigences. Nous savions qu'une base de données orientée document et à scalabilité horizontale constituait la solution idéale pour les jeux basés sur des comptes joueurs, » explique Yordan Gyurchev.

« La capacité de créer une solution hautement scalable pour un trafic ultra-dense repose sur la maîtrise des plus petits détails. Pour y parvenir, les ingénieurs doivent normalement se concentrer sur divers éléments. Avec MongoDB Atlas et le support MongoDB, nous prenons un gros raccourci. Si nous devons gérer cela en interne, nous serions obligés de nous en remettre entièrement aux seules compétences de notre expert en bases de données. Or, désormais nous avons accès à une mine de connaissances, d'expertise et de technologies de pointe. »

Sonic Forces

Kingdom Conquest: Dark Empire, a été le premier jeu HARDlight lancé sur MongoDB. Pour cette première expérience, tout s'est déroulé sans accroc. Fin 2017, ce fut au tour de Sonic Forces: Speed Battle de débarquer sur mobile. Très gourmande en ressources, cette application « always-on » permet une connexion non-stop à Internet et aux classements en direct. En arrière-plan, Sonic Forces dispose d'un cluster à 3 shards sur MongoDB Atlas, capable de monter rapidement en capacité pour traiter les charges complexes de millions de gamers engagés dans la course. La base de données s'est montrée remarquablement stable, avec des temps de latence très bas et aucune interruption de service. Au final, ce lancement s'est opéré en toute sérénité, avec une équipe DevOps aux anges et des gamers on ne peut plus enthousiastes.

Et Yordan Gyurchev de conclure : « Avec MySQL, le back-end des bases de données n'était réellement au point qu'au bout de plusieurs versions de jeux. Avec MongoDB Atlas, nous avons réussi du premier coup. Ce n'est pas un mince exploit. »

Tout comme les plateformes de jeu se sont transformées au fil du temps, les bases de données se devaient elles aussi d'évoluer et de s'adapter. Aujourd'hui, SEGA HARDlight fait confiance à MongoDB Atlas pour tous ses nouveaux jeux lancés en ligne. Opérations, gestion, montées en charge... MongoDB Atlas s'occupe de tout pour permettre aux équipes HARDlight de recentrer leurs efforts sur le développement et l'exploitation des jeux les plus emblématiques de la planète gaming – en toute sérénité.



mongoDB® Atlas

Un service MongoDB conçu et piloté par les créateurs de la base de données. Rapide, sécurisé, hautement disponible et adapté à toutes les tailles d'environnement.

mongodb.com/atlas

NEO4J

CÉDRIC FAUVET

Business Développement

+33 (0)8 05 08 03 44

ventes@neo4j.com

Nordenskiöldsgatan 24 211 19 SWEDEN

www.neo4j.com

Analytique

API

Applications

Consultants/SSII
Intégrateurs

Dataviz

Infrastructure

Institutionnels



La plateforme N°1 pour les données connectées

Étude de cas



Grâce au traitement du langage naturel à grande échelle par Neo4j, Caterpillar gagne en efficacité dans la réparation de son matériel

SECTEUR

Industriel

CAS D'USAGE

Traitement du langage naturel / intelligence artificielle

OBJECTIF

Créer un outil de TLN permettant d'extraire du sens à partir de documents à grande échelle.

DÉFI

La base de données relationnelle renvoyait des chaînes de texte non analysées qui manquaient de sens.

SOLUTION

Création d'un outil de TLN évolutif avec Neo4j pour améliorer l'efficacité des réparations et de la maintenance.

RÉSULTATS

- Recherches en temps réel parmi des millions de documents
- Possibilité de relier la cause à l'effet et d'identifier des questions à un niveau plus approfondi

Caterpillar, Inc. détient plus de 27 millions de documents de suivi de réparation et de maintenance de véhicules. En utilisant Neo4j pour effectuer un traitement du langage naturel (TLN), l'entreprise peut désormais faire des recherches à grande échelle pour détecter les tendances et les problèmes issus des réparations, prescrire des solutions et faire de précieuses prédictions. Le tout afin d'améliorer l'efficacité des réparations et de la maintenance des véhicules dans toute l'entreprise.

Caterpillar, Inc., premier fabricant mondial d'équipement de construction et d'exploitation minière est sollicité par ses clients pour développer leurs actifs dans les infrastructures, l'énergie et les ressources naturelles.

Dans le cadre de cette mission, à chaque fois qu'une machine de Caterpillar arrive en réparation ou en maintenance, le technicien crée un document de garantie qui enregistre la demande, l'analyse du problème et sa solution. Cette base documentaire constitue un volume de 27 millions de documents. Pour gérer cette masse énorme de données, l'entreprise a lancé la création d'un outil de TLN afin de mettre à jour les tendances et les relations non-visibles. S'est toutefois posée la question des moyens de traitements et d'utilisation de cette base.

La solution de graphes de Neo4j, la clé du succès

La principale raison ayant orienté Caterpillar vers une solution à base de graphes : leur grande flexibilité les rend parfaitement adaptés au traitement du langage et au machine learning. De plus, le traitement du langage est généralement fait sur la base de l'analyse des liens entre le verbe et les autres mots qui s'y réfèrent ou par l'utilisation d'un arbre syntaxique. Or ces 2 structures sont des graphes.

Caterpillar s'est alors tourné vers Neo4j pour les structures de données de graphes afin de créer un modèle logique de connaissances. La solution a ainsi permis d'établir des ontologies et de faire des déductions. Pour passer du langage naturel aux résultats d'une requête de graphes, l'équipe a mis au point une architecture de données qui intègre le texte via une boîte à outils open-source de TLN, qui utilise Python pour combiner les phrases en chaînes, ajuster les limites et ne pas tenir compte du texte inutile. Les données sont également importées depuis les systèmes ERP SAP et non SAP. WordNet est aussi utilisé comme dictionnaire lexicographique pour donner une définition aux mots, Neo4j identifie les motifs et les relations, établit les hiérarchies et ajoute les ontologies. Le tout combiné permet aux utilisateurs de faire des recherches pertinentes à partir de simples requêtes Cypher.

Des résultats qui dépassent les attentes de Caterpillar

« Les solutions de ce genre vont beaucoup plus loin que ce que l'on peut faire avec une IA standard » déclare Morgan Vawter, directrice de l'analytique chez Caterpillar, Inc. « Elles incarnent l'état d'esprit de l'entreprise, son domaine de connaissances et sont ainsi le résultat d'une traduction minutieuse de l'humain à la machine. »

neo4j.com



Bonjour Cédric Fauvet, nous souhaitons aborder avec vous le sujet des graphes et de l'IA. Tout d'abord, de façon générale, en quoi les graphes sont adaptés pour les projets d'IA ?

Pour rester générique, je vous dirai que la technologie des graphes, en reliant les données et en définissant les relations, apporte à l'IA des éléments de contexte précis, permettant de favoriser efficacement le développement d'applications d'IA sophistiquées.

Effectivement, le contexte est important dans le cadre d'un apprentissage machine, mais en quoi les graphes apportent ce contexte nécessaire ?

Les graphes offrent une plus grande efficacité de traitement des données connectées et, par conséquent, l'apprentissage automatique s'en trouve accéléré... Pour obtenir un contexte efficace il est nécessaire d'analyser les données à proximité de la donnée sur laquelle on souhaite travailler, ce qui induit souvent beaucoup de jointures et donc des temps de traitements longs et des requêtes difficiles à construire avec une base de données non-graphes. Avec une base de données de graphes native il ne s'agit que de suivre quelques relations, c'est rapide et facile à écrire.

Mais encore ?

Les graphes de connaissances offrent un moyen de rationaliser les flux de travail, d'automatiser les réponses, d'améliorer la productivité et de prendre des décisions intelligentes à grande échelle. Par ailleurs, les graphes de connaissances sont des ensembles de faits interdépendants qui décrivent des entités, des faits ou des choses du monde réel et leurs interrelations sous une forme compréhensible pour l'homme. La Nasa a d'ailleurs utilisé ces graphes de connaissance avec succès, je vous invite à découvrir le cas d'usage sur Neo4j.com

L'un des défis de l'adoption de l'IA consiste à comprendre comment une solution d'IA a pris une décision particulière, en quoi les graphes apportent cette « explicabilité » ?

Cette capacité est cruciale pour l'adoption à long terme de l'IA parce que, dans de nombreux secteurs, comme les soins de santé, l'évaluation du risque de crédit et la justice pénale, nous devons être capables d'expliquer comment et pourquoi les décisions sont prises par l'IA. C'est là que les graphes peuvent ajouter un contexte à la crédibilité.

Les graphes abordent assez facilement la question des données explicables, en utilisant les méthodes de généalogie des données employées par la plupart des principales institutions financières d'aujourd'hui. Pour ce faire, nous devons stocker nos données sous forme de graphe, mais nous avons la possibilité de suivre comment les données sont modifiées, où les données sont utilisées et qui a utilisé quelles données.

Enfin, des algorithmes explicables nous permettent de comprendre quelles couches et quels seuils conduisent à une prédiction.

Un dernier mot ?

L'IA et l'apprentissage machine ont un grand potentiel. Les graphes libèrent ce potentiel. C'est ainsi que l'on peut dire que la technologie des graphes, parce qu'elle intègre le contexte et les connexions, rend l'IA plus largement applicable et la fait entrer dans une nouvelle ère. De plus, il n'a été question jusqu'ici que d'enrichir les méthodes traditionnelles de faire de l'IA avec les graphes, alors qu'il est possible d'utiliser la théorie des graphes pour faire l'IA directement dans les graphes.

En d'autres termes, les graphes donnent tout leur sens au i de l'IA.



opendatasoft

Leading solution for data sharing

Opendatasoft développe la solution phare en Europe pour le partage et la réutilisation des données des entreprises et administrations. Grâce à Opendatasoft, organiser, partager et visualiser tout type de données n'a jamais été aussi accessible pour tout professionnel qui cherche à booster l'innovation et le développement de son organisation.

Fondée en 2011, la plateforme Opendatasoft a gagné la confiance de plus de 250 clients dans le monde: en Europe, en Amérique du Nord, en Australie ainsi qu'au Moyen-Orient.

En France, nous travaillons avec de nombreuses villes et collectivités locales (dont les villes de Paris, Issy-les-Moulineaux ou encore Lille), les Ministères, et de grandes entreprises telles que Schneider Electric, Kering, Enedis, Saint-Gobain, SNCF, Euler Hermes, SFR, Total ou Indigo Group.

Il est temps de faire briller vos données!

#MakeYourDataBright



Pourriez-vous nous en dire plus sur la genèse du projet de partage de données au sein du groupe BPCE ? Avec quels objectifs de base êtes-vous partis ?

Nous avons décidé de lancer notre portail de données en open data en droite ligne de notre stratégie de transformation numérique. Celle-ci vise à rendre les choses simples pour nos clients, pour nos collaborateurs et pour nos partenaires.

Ce projet est né d'une belle rencontre. La rencontre entre notre volonté d'ouverture et de simplicité vis-à-vis de nos communautés de partenaires et la proposition de valeur de la plateforme de publication de données développée par Opendatasoft. Cette initiative repose sur deux axes de travail : un axe interne et un axe externe. L'axe interne nous permet d'acculturer les différentes entités du Groupe sur l'open data, d'échanger sur ce que cela regroupe, ses enjeux et ses opportunités.

L'axe externe nous permet de nous rapprocher de l'écosystème open data en France qui, on peut le noter, est particulièrement dynamique et ouvert à la co-création et au partage d'expériences. Beaucoup de collectivités ainsi que de nombreuses entreprises dans les transports, l'industrie, la météo ou l'énergie, par exemple, sont déjà très actives dans ce mouvement de l'open data. Nous venons nous joindre à ces acteurs pour permettre à tous de travailler plus vite et d'inventer ensemble de nouveaux modèles.

Quelle est la nature des données que vous partagez ? Le type de jeux de données publiés a-t-il beaucoup évolué depuis le début de votre démarche ?

Il n'y a bien entendu ni de données personnelles, ni commerciales sur le portail open data du Groupe. Nous rendons uniquement accessibles les données publiques des entreprises du Groupe BPCE dans des formats facilement réutilisables par tous. Le catalogue de jeux de données est dynamique et constitué de données structurées, fiables et centralisées, mises à jour par des collaborateurs en charge des données et de leur fiabilité. Les données sont disponibles sous forme de tableaux chiffrés mais aussi sous forme graphiques, sous forme de cartographies interactives et même de calendriers.

Très récemment, la Fédération Nationale des Banques Populaires et Natixis Payment ont publié en open data une étude sur les paiements de proximité réalisés par les voyageurs étrangers dans les régions et départements français. C'est une première en France qui offre une nouvelle clé de lecture sur le dynamisme des régions.

J'imagine que la gouvernance des données est un vrai sujet dans votre secteur. Comment gérez-vous cela au sein du groupe BPCE ?

La gouvernance des données est un des enjeux clés de la direction en charge du numérique et data pilotée par Yves Tyrode, directeur général du Groupe BPCE. Nous venons d'ailleurs d'accueillir au sein de son comité de direction, Ludovic Favarette en tant que Directeur Gouvernance et démocratisation de la Data, en charge de l'animation et du déploiement de la gouvernance et de la culture data au sein du groupe.

Notre métier, vous le savez, est « Privacy by design ». Les données bancaires de nos clients sont leur propriété et ne nous appartiennent pas. Nos clients nous en confient leur sécurité et leur gestion dans le cadre d'une relation de confiance et de secret bancaire. En matière d'open data, nous ne partageons donc que les données publiques propres à nos entreprises ainsi que des études. Toutes les entités du groupe qui le souhaitent peuvent publier leurs données publiques sur le portail du groupe.

Quels sont les principaux bénéficiaires du projet de partage de données du Groupe BPCE ?

Les principaux utilisateurs des données que nous mettons à disposition sont des analystes, des journalistes, des chercheurs, des enseignants, des étudiants, les autres acteurs de l'open data, des collectivités locales, des chambres de commerces, des acteurs du transport, des startups ... et même des collaborateurs du groupe qui trouvent intérêt à accéder par ce canal à certaines données publiques. Vous le voyez, les premiers bénéficiaires sont déjà nombreux.

Si vous deviez convaincre un acteur de votre secteur à entamer une démarche open data ou partage de données, que lui diriez-vous ?

Je lui dirais que selon moi, c'est une démarche qu'il est urgent d'entreprendre car elle est dans le sens de l'histoire, dans le sens de mettre à disposition les données publiques de son entreprise pour permettre au plus grand nombre de pouvoir y accéder et de créer de la valeur. Toujours selon moi, tous les secteurs d'activité devront tôt ou tard adopter l'open data soit de manière volontariste, soit parce que le régulateur l'exigera. Alors autant s'y préparer au plus tôt, on gagne toujours à innover.



LE BIG DATA AU SERVICE DE LA TRANQUILLITÉ PUBLIQUE : une nouvelle façon de gérer la ville de Marseille

Marseille lance sa plateforme “big data” pour gérer et sécuriser l’espace public

Marseille vient de mettre en service sa plateforme de données “big data” qui permet de détecter en temps réel, tous services confondus, les incompatibilités entre les différents événements sportifs, culturels, sociaux... organisés dans la ville et d’assurer la sécurité des habitants.

La Ville a décidé, en 2015, de développer une plateforme municipale de données “big data” au service du citoyen. Le projet mis en service au mois d’août, repose sur une plateforme ultra performante de stockage, d’analyse et de planification qui permet de surveiller la ville en temps réel et de signaler ou prévoir les dysfonctionnements. Dans le cadre de la protection des données personnelles, les données collectées devront respecter les règles de la CNIL.



Améliorer la transversalité des données entre les services

Le dispositif Oracle a été mis en place par Ineo digital, filiale d’Engie Ineo. Il regroupe les données de tous les services municipaux de ces dix dernières années et celles collectées en temps réel et facilite donc les prises de décision.

“ La Ville de Marseille a décidé, en 2015, de développer une plateforme municipale de données “big data” au service du citoyen.

On y retrouve les informations de la police municipale et du centre de supervision urbain, l’agenda culturel, les emplacements des marchés, les transports publics, les informations des marins pompiers, de l’AP-HM, la gestion des parcs et jardins de la ville, les chantiers en cours et à venir, les arrêtés de stationnement et de circulation, les permis de construire, les signalements des habitants, les tournages audiovisuels... La plateforme permet de repérer si un objet est abandonné dans la rue ou dans un lieu public ou encore si le seuil de densité d’un lieu public est atteint.

QU'EST-CE QUE LA SOLUTION Oracle Autonomous Database Cloud ?

Oracle dévoile le premier service Cloud de base de données autonome au monde, conçu à partir d'Oracle Database 19c, la nouvelle génération de bases de données Oracle leaders sur le marché.

Oracle Autonomous Database Cloud élimine la complexité, l'administration manuelle et les erreurs humaines pour atteindre une sécurité, une fiabilité et des performances opérationnelles accrues à un coût nettement plus bas.



SELF-DRIVING

Automatise la gestion de la database et de son infrastructure, la surveillance et le tuning



SELF-SECURING

Protège contre les attaques externes et les utilisateurs internes malveillants



SELF-REPAIRING

Protège contre tous les temps d'arrêt, y compris la maintenance planifiée

2 offres gratuites pour essayer le Cloud Oracle

Always free, accédez en illimité à :

- ✓ 2 instances Oracle Autonomous Database et outils associés tels que Oracle Application Express (APEX) et Oracle SQL Developer
- ✓ 2 instances Oracle Cloud Infrastructure Compute VMs ; Block, Object, and Archive Storage ; Load Balancer et data egress ; Monitoring et Notifications

Crédits gratuits 300\$ pendant 30 jours sur :

- ✓ les services Oracle Cloud Database, Analytics, Compute et Container Engine for Kubernetes
- ✓ jusqu'à 8 instances sur tous les services disponibles
- ✓ jusqu'à 5 TB de storage

Offrir une nouvelle vision des données, obtenir de nouvelles perspectives, vous ouvrir des possibilités illimités.

Rencontrez un expert sur le stand n°A9



Présentation

Cabinet de conseil et d'audit, PwC est un réseau international de 251 000 collaborateurs dans le monde dont 6 400 en France et en Afrique francophone qui partagent solutions, expertises et perspectives innovantes au bénéfice de leurs clients et partenaires. PwC continue sa transformation pour renforcer la confiance sur le marché et accompagner les acteurs de l'économie dans leur réussite, à travers des missions de conseil (stratégie, management, transactions, cybersécurité, data & analytics, juridique et fiscal), d'audit et d'expertise comptable.

Les experts data de PwC France

L'activité Data Intelligence de PwC France compte plus de 300 experts (dont 50 data scientists) qui accompagnent les entreprises, de toutes tailles et tous secteurs d'activités confondus, dans leur transformation « data driven ». Ils les aident ainsi à repenser leur business avec créativité, leur mode d'organisation et leur sécurité dans le respect des réglementations en vigueur et dans un contexte où la donnée est au cœur de leur performance.

Les principales activités data de PwC France

- Transformation des modes de travail en tirant parti des nouvelles opportunités offertes par la technologie (automatisation, rapprochement de données systématiques, anticipation/projection, etc.)
- Accompagnement à la création de nouveaux produits et services innovants « value driven » (IoT, recommandation clients, maintenance prédictive, digitalisation de processus, etc.)
- Conception d'architectures et eco-systèmes data, conception et développement de plateformes de nouvelles générations (microservice, containerisation, managed datalake, etc.)
- Définition d'operating models, accompagnement à la structuration d'une gouvernance de données et des processus métiers pour garantir le passage à l'échelle
- Diagnostic de performance des processus métiers (achats, supply chain, RH, etc.) et recommandation personnalisée sur une analyse de data (process mining/intelligence)
- Accompagnement et conseil sur les grands sujets réglementaires (dont la data privacy), de conformité (ex. RGPD) et de maîtrise des risques
- Transformation « data driven » de la fonction d'audit et du contrôle interne ainsi que de l'ensemble des fonctions supports et des opérations (finance, RH, achats, etc.)
- Appui à l'analyse juridique et réglementaire de contrats via des solutions de Contract Analytics
- Définition et mise en place de solutions de Robotic Process Automation (RPA) intégrées aux projets data
- Appui à la transformation des ventes et de l'expérience client : maîtrise des données client, déploiement d'une culture centrée sur le client, autonomie et performance des organisations front-office
- Audit des algorithmes (performance, explicabilité/transparence, réglementation, IA responsable)

Des espaces innovants dédiés à la data

PwC développe en France des espaces à la pointe de la technologie digitale comme la Delta Room, un espace de travail collaboratif pensé pour favoriser l'immersion et la prise de décisions qui feront la différence pour l'activité des entreprises qui l'utilisent. Cet espace intensifie le partage et le travail en mode collaboratif grâce à une technologie intuitive. Le DataLab de PwC est par ailleurs un outil interne permettant de réaliser des projets data, mis à disposition des clients souhaitant accélérer un passage de phase pilote à l'industrialisation.

Intelligence Artificielle : les experts PwC récompensés lors des « Airbus AIGYM Challenges »

L'équipe Data Intelligence, composée de data scientists spécialistes du Machine Learning et du Deep Learning, a atteint au printemps 2019 la deuxième place du classement parmi une vingtaine d'équipes en compétition.

PwC au cœur de vos enjeux data

300 experts de la data, dont **50** data scientists,
en France à vos côtés de la définition de votre
stratégie data au déploiement de l'Intelligence
Artificielle à l'échelle

Un réseau international présent dans **158** pays

Une connaissance sectorielle approfondie

Des solutions adaptées à vos différents métiers

Parce que la donnée est au cœur de la croissance de l'entreprise, l'exploiter et la protéger est essentiel pour innover, gagner des parts de marché, développer sa compétitivité et créer de la valeur.

PwC vous aide à en tirer le meilleur parti en utilisant notamment les technologies de Machine Learning et d'Intelligence Artificielle.

Nos équipes pluridisciplinaires vous accompagnent sur de nombreux sujets clés comme la connaissance client, l'optimisation des processus, la maintenance prédictive, la détection de fraudes et vous conseillent sur les sujets réglementaires, comptables, de conformité, de maîtrise des risques et de gouvernance des données.

Et si la donnée créait aujourd'hui votre valeur de demain ?

Pour en savoir plus :
www.pwc.fr/data-intelligence



Qlik libère le potentiel de vos données

Qlik s'est donné pour mission la création d'un monde « data-literate » où chacun peut utiliser les données pour relever ses plus grands défis. Seule la plateforme d'analyse et de gestion des données de bout en bout de Qlik réussit à rassembler toutes les données d'une organisation, quelle que soit leur source, et permet à tous les collaborateurs d'exercer leur curiosité et de découvrir de nouveaux enseignements.

Les entreprises utilisent Qlik pour approfondir leur connaissance client, réinventer leurs processus métiers, découvrir de nouvelles sources de revenus, et équilibrer le rapport bénéfice-risque. Voici quelques cas d'usages de Qlik en France et dans le monde :



Pernod Ricard analyse ses big data pour suivre sa performance marketing et commerciale

L'entité Data des sociétés Ricard et Pernod, filiales françaises du groupe Pernod-Ricard, a choisi le cloud pour accélérer et simplifier l'exploitation de ses données. Après avoir testé la plateforme Snowflake, partenaire de Qlik, sur des données de panel de ventes en grande distribution représentant une volumétrie conséquente (*de l'ordre du To*), l'équipe Data et les Métiers des Ventes ont développé avec Qlik Sense une application d'analyse du marché CHR (*cafés, hôtels, restaurants*).

L'application combine des données externes (*fournisseur spécialisé, réseaux sociaux...*) et internes (*CRM et ERP*), et associe datascience et dataviz pour affiner la connaissance client, appliquer un modèle de scoring et obtenir une segmentation efficace de ce marché afin de cibler des zones d'actions prioritaires. L'application inclut également une représentation géographique des données clients et prospects qui permet d'analyser la data en naviguant directement sur les cartes. Cette approche a pour avantage de valoriser l'analyse métier tirée de ces données.



Mazars transforme ses métiers avec l'analyse visuelle

Mazars, acteur international d'audit et de conseil, recueille, exploite et analyse les données de ses clients pour mener à bien ses missions d'audit et de contrôle. Dans les grands groupes, ce sont des volumes très importants de données qui sont manipulés, lorsque l'auditeur financier souhaite par exemple analyser un processus d'achats, où des centaines de milliers de commandes et factures sont traitées. Avec Qlik Sense, le métier s'est totalement transformé : les experts en data analytics de Mazars mettent à la disposition de l'auditeur une application analytique qui lui permet de visualiser l'intégralité des données et des flux, de les analyser pour effectuer ses contrôles et d'en tirer des conclusions.

Si la force de l'outil proposé par Qlik réside dans sa capacité à traiter et visualiser toutes les données, elle est aussi dans l'interactivité qui permet de répondre immédiatement à toutes les questions. La capacité de réflexion et d'investigation du cerveau humain peut être reproduite dans le processus d'analyse, grâce au moteur associatif de Qlik. Ainsi, l'auditeur peut observer un phénomène général et investiguer divers points de façon autonome pour comprendre ce phénomène en ayant une réponse immédiate à toutes ses questions. Cela n'était pas possible avec les outils de génération antérieure qui donnaient un résultat statique, et nécessitaient de programmer des requêtes longues et fastidieuses à chaque nouvelle question.



CACEIS diffuse la donnée à ses clients en temps réel

Filiale du groupe Crédit Agricole spécialisée dans la fourniture de services financiers aux investisseurs institutionnels, CACEIS a mis en place une plateforme big data pour tirer pleinement profit de ses données et repenser son expérience client, dans une approche data-centric. Pour simplifier et fiabiliser l'alimentation du datalake Hadoop, CACEIS a fait le choix d'Attunity Replicate, qui lui assure, en temps réel, l'ingestion de ses données issues de sources hétérogènes.

Cette nouvelle architecture permet d'offrir aux clients de CACEIS un accès à des analyses de leurs données via le portail Olis, avec des données en quasi-temps réel, rafraîchies toutes les 10 secondes. Grâce à ce partage des données, CACEIS offre à ses clients un point d'entrée unique pour tous les services fournis et leur permet de suivre au quotidien leurs fonds et comptes bancaires. Lancé en 2019, Olis génère déjà 5000 connexions par jour, avec 10 000 utilisateurs dans le monde entier.



Comment voyez-vous évoluer le marché de la data ?

Il est intéressant d'observer les courants qui prennent naissance au niveau du consommateur. Celui-ci devient de plus en plus méfiant vis-à-vis de l'exploitation de ses données, et les affaires concernant les fake news ou les vols de données ne font qu'amplifier le phénomène. Mais il ne faudrait pas que cela suscite un climat de peur autour de l'analyse des données, sans quoi nous risquons de passer à côté de la véritable valeur de ces dernières. Il appartient à la fois aux professionnels des données et au grand public d'identifier les entreprises responsables de telles pratiques et d'exiger d'elles le respect d'une approche raisonnable et éthique de la collecte et de l'utilisation des données, afin de garantir leur utilisation à de bonnes fins.

Comment regagner la confiance du consommateur ?

1. Un nouveau contrat social numérique : nous disposons déjà du RGPD au niveau européen, ainsi que du projet « Contract for the Web » lancé fin 2018 par Sir Tim Berners-Lee, inventeur du web. Le « Contract for the Web » vise à définir le rôle des gouvernements, des entreprises et des utilisateurs dans l'utilisation du web et des données mais il faut aller encore plus loin. Ce dont la société a besoin désormais, c'est un cadre commun à tous qui pourra redonner aux individus la confiance qu'ils attendent.

2. La data literacy ou dataphabétisation : de la même manière que la démocratie requiert un électorat bien informé pour fonctionner, l'ère des données implique que cet électorat possède des connaissances en matière de données, à savoir la capacité à lire, à comprendre, à remettre en question et à communiquer avec les données. C'est un prérequis pour prendre des décisions informées de façon autonome en matière de politique ou sur tout autre sujet ayant un impact sur le quotidien.

3. Enfin, la dernière étape consiste à créer une plateforme ouverte, suffisamment solide pour assurer la gestion et le traitement des données et des contenus et pour fonctionner de manière distribuée, connectée, transparente, éthique et authentifiée. Cet objectif peut dépasser le stade de l'utopie. Les fondations mêmes de notre société numérique reposent sur la création de valeur à partir du partage ouvert de données et d'informations.

Et dans l'entreprise, comment faire pour développer l'usage de l'analytics ?

Cela passe par la technologie, mais aussi par l'humain. Les collaborateurs présentent des lacunes en matière de Data Literacy, 70% des employés français se sentent submergés par les données dans leur travail, et seuls 19% considèrent qu'ils ont les compétences nécessaires en data literacy.

Les données sont la base de l'économie numérique et chaque utilisateur métier doit savoir les utiliser. Sans cela, les données resteront inexploitées, la plus grande partie de leur valeur sera ignorée, et les entreprises ne seront pas compétitives. La perspective d'une visualisation simple et rapide des données avec les outils de BI de deuxième génération a atteint ses limites. Il faut aller plus loin pour développer un usage plus « démocratique » de la donnée.

La technologie permet-elle d'y parvenir ?

L'IA, les Big Data, les chatbots sont des technologies qui attirent toutes les entreprises, et qui peuvent être mises à contribution pour développer la data literacy. Chez Qlik, nous proposons l'intelligence augmentée au profit de l'analytics.

L'intelligence artificielle s'accompagne d'une promesse extraordinaire et les entreprises cherchent les moyens d'exploiter cette intelligence. Mais en dehors d'une niche restreinte d'applications pouvant s'appuyer uniquement sur l'automatisation machine, la plupart des problèmes métier complexes nécessitent une perspective humaine. C'est pourquoi nous appelons l'IA « Intelligence Augmentée » : l'intelligence et l'intuition de l'homme sont complétées par l'apprentissage machine.

Notre approche de l'intelligence augmentée s'appuie sur notre technologie unique d'indexation associative, qui associe les interactions humaines aux schémas générés par la machine. Lorsque les utilisateurs explorent leurs données, l'application leur suggère de nouvelles manières de les envisager, en mettant en avant les informations à approfondir. Ceci revient à ajouter une vision « rayons X », qui montre à l'utilisateur des connexions qu'il n'aurait normalement pas vues.

Chez Qlik, nous avons toujours eu une vision démocratique de l'analytics. Notre entreprise a été fondée sur l'idée que la véritable promesse de la donnée ne sera tenue que si chaque utilisateur métier peut interagir avec elle et faire des découvertes générant des résultats qui apportent un avantage concurrentiel.

REEPOR

MATTHIEU MEDRINAL

Head of Sales

+33 (0)1 84 13 72 38
matthieu.medrinal@reepor.io
10 Rue Henner 75009 Paris FRANCE

www.reepor.io

Analytique
API
Applications
Consultants/SSII
Intégrateurs
Dataviz
Infrastructure
Institutionnels



www.reepor.io

Automatisez votre reporting digital



Digital Reporting Automation



Selon Reeport, quels ont été les sujets data les plus notables cette année ?

Thibault de Broissia : l'année dernière, je disais ici que les entreprises étaient entrées dans un âge de raison par rapport au Big data. J'entendais par là que les entreprises ont aujourd'hui des attentes très pragmatiques et un très haut niveau d'exigence par rapport aux retombées qu'elles attendent du Big Data. Depuis l'année dernière ce constat n'a fait que se renforcer. Les entreprises veulent utiliser le potentiel du Big data pour répondre à des enjeux pragmatiques et métiers de plus en plus granulaires, comme par exemple la mesure de la performance de leurs investissements médias sur le digital.

Pourquoi le Big Data appliqué à la performance digitale est-il un sujet majeur pour les entreprises ?

Thibault de Broissia : C'est l'un des sujets majeurs pour les annonceurs en tous cas. Il traduit cependant très bien le besoin de mieux analyser pour mieux comprendre. Dans le domaine de la mesure de la performance les entreprises ont aujourd'hui besoin de s'affranchir de plateformes publicitaires qui fournissent en même temps l'espace publicitaire et la solution de mesure de ces performances. On le comprend ici, les entreprises ont besoin d'indépendance et d'autonomie en s'appuyant sur leur propre solution d'analytics.

C'est parfaitement légitime pour un annonceur confronté au quotidien à la pluralité des canaux et des dispositifs marketing : face à la multiplication des points de contacts et de leurs spécificités il a besoin de collecter lui-même la data, afin de l'analyser et la comprendre avec le maximum d'indépendance possible. Si l'annonceur ne se base que sur la data fournie par les plateformes sa vision est silotée dès le départ.

Cette vision s'est-elle traduite par une évolution du produit Reeport ?

Thibault de Broissia : Notre solution était d'ores et déjà prête à répondre à ces nouveaux besoins. Nous avons donc conçu Reeport pour être un compagnon qui simplifie la collecte, l'analyse et le partage de la donnée des performances digitales dans l'entreprise. Concrètement, Reeport diffuse dans l'entreprise une donnée personnalisée pour chaque équipe, donc une information actionnable pour chacun. C'est cette association de puissance et de simplicité qui fait toute la différence par rapport aux autres solutions du marché.

Dans le cas de la mesure de la performance média, il faut comprendre qu'un annonceur qui souhaite gagner en indépendance risque d'être confronté rapidement à une problématique de collecte de la data. Cet annonceur devra alors déployer une architecture de traitement des données adaptée précisément à sa problématique. Reeport lui évite tout simplement d'avoir à se lancer dans un projet lourd tout en lui garantissant d'obtenir l'indépendance attendue.

Reeport est donc une solution standardisée mais déployée sur-mesure ?

Thibault de Broissia : La force de Reeport est sa capacité à bien comprendre le quotidien de l'entreprise afin de s'adapter aux spécificités de chaque organisation. Les équipes, les systèmes, les process, la culture data sont autant d'éléments uniques qu'il faut prendre en compte pour que l'outil soit naturellement adopté et qu'il joue un rôle fédérateur et stimulant dans l'entreprise.

C'est cette compréhension de la réalité quotidienne des équipes qui nous permet de proposer une plateforme puissante et fiable qui fait gagner un temps précieux. Par exemple, les équipes ne sont plus noyées dans des tâches fastidieuses comme la collecte, l'agrégation ou le formatage des données. Ces équipes peuvent se concentrer sur l'analyse et la compréhension de ce qui se passe et cela réduit très considérablement les délais de réaction.

Reeport n'est donc pas qu'une techno robuste. Nous accompagnons les directions et le management dans leur volonté de tirer parti de la donnée, c'est-à-dire d'évoluer vers une organisation Data Informed dans laquelle l'information circule là où elle doit, quand elle le doit, et auprès de collaborateurs réactifs qui ont toutes les cartes en main pour prendre les bonnes décisions.

SAAGIE

RODOLPHE TIBERGHIE

Sales Director - Europe

+33 (0)2 72 88 31 69

contact@saagie.com

92 avenues des Champs Elysées 75008 Paris FRANCE

www.saagie.fr



Saagie®



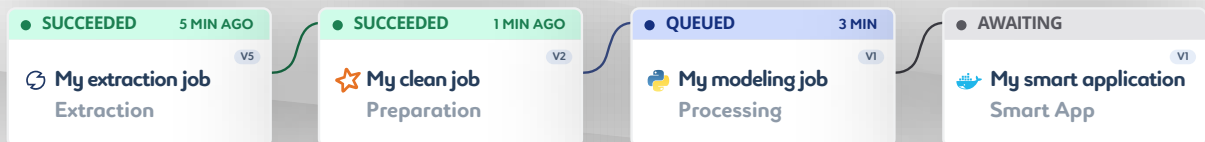
Saagie®

Plug & Play Orchestrator for DataOps

Empower IT, Analytics & Business teams

Integrate best-in-class AI technologies

Accelerate data projects delivery



FIND US AT BOOTH A18



Quels défis rencontre-t-on aujourd'hui quand on veut se lancer dans un projet Big Data ?

Le succès des projets analytics (Big Data et IA) vient de la capacité à amener rapidement de la valeur aux métiers en valorisant leurs données et d'arriver à maintenir cet avantage dans le temps. Or, nous faisons aujourd'hui face à une situation paradoxale : si les promesses du big data et de l'IA ont bien permis de déclencher de nombreuses expérimentations, les cabinets tels que Gartner constatent que celles-ci peinent clairement à arriver en production (on parle tout de même d'un taux d'échec allant jusqu'à 85% selon les analystes). En plus de ne pas apporter la valeur escomptée en bout de chaîne, cette situation va parfois jusqu'à remettre en question la capacité de la DSI à accompagner les métiers sur ces nouveaux sujets.

Est-ce pour cela que le DataOps est considéré comme le nouveau sujet tendance de l'analytique ?

En effet, car maîtriser le cycle de vie complet de la donnée, c'est réduire les coûts de production d'un projet Data et maintenir son efficacité dans le temps.

Le DataOps est la nouvelle tendance de la Hype Analytics du Gartner. Héritée des pratiques agiles et DevOps, il vise à accélérer la mise en production d'un projet analytics, en gérant mieux son cycle de vie de bout en bout : de l'origine de l'idée (idéation), à la connexion des flux de données, leur stockage, leur préparation, modélisation et exposition aux utilisateurs. Les "data pipelines" deviennent une ligne d'assemblage dans laquelle la qualité, la gouvernance, l'opérabilité, la sécurité et la traçabilité sont des éléments déterminants. C'est le pendant « opérations » des projets analytiques.

Qu'est ce qui fait la particularité du DataOps par rapport aux autres approches ?

L'approche du DataOps est itérative et doit être agile, collaborative et innovante. Les data pipelines qui matérialisent la chaîne de valeur de ces projets doivent pouvoir en permanence être modifiées et améliorées pour tenir compte d'impératifs business ou de nouveaux changements technologiques. Intégration continue, déploiement continu : la valeur est dans le mouvement et le processus, plus encore que dans le produit fini.

Le DataOps impacte donc le choix des outils et les processus existants, mais qu'en est-il de l'aspect humain ?

Chaque profil impliqué dans un projet data doit pouvoir accéder et utiliser ses outils de prédilection.

La DSI veut du contrôle, de la maîtrise, de la robustesse et de la traçabilité sur toute la chaîne de valeur. Le Data Lab souhaite pouvoir explorer facilement ses données, accéder aux dernières versions des technologies de Data Science et ne pas se soucier des contraintes liées à l'infrastructure (stockage et capacité de calcul). Les métiers quant à eux ne doivent pas se soucier de la plomberie pour travailler sur leurs données, et pouvoir toujours garder leur objectif business en tête.

Quel rôle a Saagie dans cette méthodologie DataOps ?

Le projet Data est un enchaînement de processus, supportés par des technologies diverses et impliquant des personnes aux compétences différentes. Saagie fournit l'outil qui supporte la méthodologie DataOps au sein du projet. Notre « orchestrateur DataOps » permet de tirer profit du meilleur des technologies Open Source et propriétaires et d'ainsi gérer le cycle de vie complet d'un projet Analytics. Les Data pipelines sont précisément au cœur de notre offre. Data engineers & data scientists deviennent autonomes dans la mise en place de traitements visant à collecter, stocker, préparer, modéliser ou visualiser les données avec les outils les plus adaptés à chacun.

Il devient facile d'assembler, opérer, monitorer et ordonnancer leurs travaux analytiques. Notre rôle est de mettre de l'huile dans ces rouages pour permettre plus de vélocité et d'agilité dans les processus de production des projets Data.

Quels avantages vos clients en retirent-ils ?

Quel que soit le mode de déploiement (on-premise, cloud ou hybride), nous permettons à nos clients de livrer leur premier cas d'usage data en 10 semaines - contre 12 à 24 mois en moyenne sans orchestrateur - et faire évoluer leurs cas d'usage et data pipelines dans le temps. Par ailleurs, les clients ayant déjà intégré des pratiques telles que l'automatisation des déploiements applicatifs propre à une démarche DevOps peuvent directement capitaliser sur leur expérience et outillage existants. Comment ? Simplement en y liant Saagie grâce au plugin Gradle. Le résultat : la production de la valeur métier à partir de leurs données, avec la même aisance et robustesse que celles mises en œuvre dans le monde applicatif.



Le traitement du cancer entre dans une nouvelle ère avec l'intelligence artificielle

L'Hôpital Universitaire d'Amsterdam (AUMC) utilise les solutions SAS d'Intelligence Artificielle pour augmenter la vitesse et la précision de diagnostic de certains cancers.

L'application de l'intelligence artificielle (IA) gagne du terrain dans le domaine des soins oncologiques. L'Hôpital Universitaire d'Amsterdam utilise le traitement d'imagerie médicale par ordinateur et l'analyse prédictive pour mieux identifier les patients atteints de cancer qui nécessitent une intervention chirurgicale.

L'Hôpital Universitaire d'Amsterdam dispose de l'un des plus grands centres d'oncologie d'Europe, et capitalise sur les informations relatives aux soins de chaque patient pour améliorer la prise en charge des patients suivants. Pour cela l'hôpital collecte de nombreuses données individuelles, y compris des biomarqueurs, de l'ADN et des données génomiques.

«L'IA nous aide à traiter nos volumes de données sans cesse croissants», déclare le Dr Kazemier, professeur de chirurgie et directeur du département d'oncologie chirurgicale à l'Hôpital Universitaire d'Amsterdam.

Le Dr Kazemier recherchait une plate-forme d'analyse robuste et s'est tourné vers SAS, initiant un partenariat qui a fait progresser l'utilisation de l'IA pour évaluer les tumeurs du foie avant et après la thérapie systémique.

Par ailleurs, la plateforme SAS permet à des milliers de chercheurs en oncologie de l'Hôpital Universitaire d'Amsterdam d'accéder à des analyses de pointe pour améliorer la recherche et la collaboration.

Selon le Dr Kazemier, l'IA peut révolutionner les soins de santé à condition d'être transparente et ouverte. «Si vous créez des algorithmes pour aider les médecins à prendre des décisions, il doit être possible d'expliquer ce que fait réellement cet algorithme», dit-il. «Imaginez qu'un algorithme propose un soin se révélant erroné et que le médecin suive cette recommandation. Quelle en serait la conséquence ? L'erreur n'est pas seulement humaine.»

L'établissement d'un processus de bout en bout vérifiable et transparent pour la gestion des décisions est un facteur crucial dans le déploiement de l'IA en matière de soins de santé.

La plate-forme SAS rend compte de la façon dont chaque analyse est effectuée, facilitant ainsi le suivi des modèles et des algorithmes par les médecins. Cette collaboration entre l'homme et l'algorithme consolide la confiance dans l'intelligence artificielle.

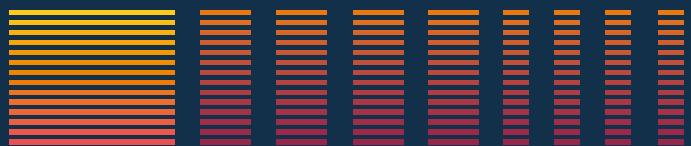
Pour en savoir plus, visionnez ce témoignage vidéo de l'AUMC :



sas.com/en_us/customers/amsterdam-umc.html

DID YOU KNOW?

SAS® IS THE NO. 1 AI AND ADVANCED ANALYTICS SOFTWARE PLATFORM.



Today, our customers fight fraud with machine learning, improve cancer survival rates with computer vision and increase customer understanding with natural language processing. Thanks to the collaboration with our customers and partners, we're proud to be No. 1 globally in AI and advanced analytics software platform combined market share and revenue. Our growth rate in AI alone is more than 3x faster than the overall AI software platforms market.*

sas.com/discover



POWERFUL ANALYTICS. REAL RESULTS.

* Source: IDC, Worldwide Artificial Intelligence Software Platforms Market Shares 2018, June 2019, IDC#US45262419
IDC, Worldwide Advanced and Predictive Analytics Software Market Shares 2018, July 2019, IDC#US45366918.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. © 2019 SAS Institute Inc. All rights reserved. G114794.0919



La prochaine offensive des GAFAs (et des BATX, leurs homologues chinois) visera les banques et le secteur marchand B2C. La bataille qui s'annonce pourrait changer radicalement notre rapport à la consommation et à l'épargne voire déstabiliser nos économies.

Pourriez-vous décrire un scénario de services bancaires assistés par IA ?

Lundi 28 février 2022, 13h30. Depuis votre bureau, vous consultez votre smartphone. Votre assistant personnel vous interpelle : « Votre salaire vient d'être versé. Voulez-vous consacrer 2 minutes à la gestion de vos comptes ? ».

Cet assistant est une intelligence artificielle qui vous connaît bien : il bâtit ses recommandations selon vos centres d'intérêt, vos habitudes de consommation, vos trajets, mais aussi vos dépenses contraintes, votre propension à épargner, appétence aux risques, capacité d'emprunt... Il peut vous proposer de réaliser des projets : épargne pour les études des enfants ou prochaines vacances au soleil... Naturellement il met en concurrence vos fournisseurs de services et ceux qui aimeraient le devenir.

Vous acceptez et il vous propose différentes options de pré-allocations de vos revenus. Il vous conseille d'ouvrir un livret « boosté » à la Banque B et d'y transférer une partie de votre épargne de la Banque A, puis de saisir une promotion de dernière minute pour un week-end au ski, payable en 3 fois sans frais... C'est simple, rapide, pertinent.

Même s'il est technologiquement réaliste, ce scénario est pour le moment fictif. Est-ce de la science-fiction ?

Non. Pour deux raisons.

D'une part, le service rendu par l'assistant digitalise et démocratise les services à forte valeur ajoutée des family offices – ces structures de conseil et d'assistance actuellement réservées aux grandes fortunes. Votre assistant personnel assure une veille continue et vous propose l'offre adaptée à vos critères de qualité, de prix, et à vos contraintes. Son IA affine sa connaissance à chacune de vos interactions. Avec une interface simple, épurée, cet agent rend un service de grande qualité aux consommateurs et « offre » aux fournisseurs de service un ciblage très pertinent.

D'autre part, ce scénario est viable. La réglementation européenne DSP2 permet aux établissements de paiement de proposer des agrégateurs bancaires pour jongler avec ses comptes et initier des paiements. Les GAFAs ont aussi lancé leur offensive bancaire : Facebook et sa monnaie « Libra », Apple Pay en 2014, Android/Google Pay en 2015, Amazon et sa carte Visa gratuite en 2018.

L'intelligence artificielle permet d'imaginer ce « Family Office for the Masses ». La technologie existe et est accessible. Votre assistant est omniprésent : smartphone, enceinte connectée, télévision, ordinateur... Compte tenu de la connaissance client pointue des GAFAs et de l'attente d'une partie du public, le modèle économique est à l'évidence viable.

Dans ce scénario, quel rôle occuperaient les banques ?

Les banques sont ici réduites à leur plus simple expression : producteur de produits financiers. Fini le conseil humain et l'agence. Les produits bancaires sont devenus des produits de consommation comme les autres, rangés entre l'abonnement pour les lames de rasoir et les capsules Nespresso et comparés entre eux tels des hôtels sur TripAdvisor. L'offre et la demande de produits financiers seront désormais automatisées sur une place de marché virtuelle peuplée d'agents logiciels intelligents.

Mais les GAFAs ne reprennent pas toutes les missions des banques. On attend d'un conseiller – réel ou virtuel – qu'il veille sur les intérêts de ses utilisateurs. La relation avec les clients reste donc critique pour expliquer et adapter les recommandations de l'IA, en priorisant le devoir de conseil. Enfin, qui pour gérer la conformité et la sécurité ? Les GAFAs ont-ils vocation à lutter contre le blanchiment, les fraudes, le financement du terrorisme, respecter la réglementation, contribuer à la sécurité des territoires, des places de marché et de vos données personnelles ?

Au-delà du service bancaire, quel sera l'impact économique du déploiement de l'IA ?

Cette réflexion illustre la puissance transformatrice de l'Intelligence Artificielle, mais aussi son impact sur les équilibres macro-économiques. En effet, une IA pourrait orienter les choix de consommation ou d'épargne, et donc avoir un poids politique sans gouvernance étatique. Cette révolution est inéluctable. La question n'est pas de savoir si elle se produira mais quand. Tous les acteurs de la consommation ainsi que nos autorités politiques devraient donc se demander comment se positionner pour préserver leurs souverainetés, tout en profitant des avancées technologiques poussées par les GAFAs.



Reliable infrastructure for your customer data

Collect, clean, and control your customer data at scale with Segment's Customer Data Infrastructure. Segment provides a platform that puts privacy first and enables businesses to generate more valuable and trusted experiences for their customers using first-party data.

CONNECTIONS

Collect your data once. Supercharge your tech stack.

Capture and send every customer interaction—both offline and digital—to the tools your teams use for analytics, growth, marketing, and more.

Collect: Build your foundation for customer data

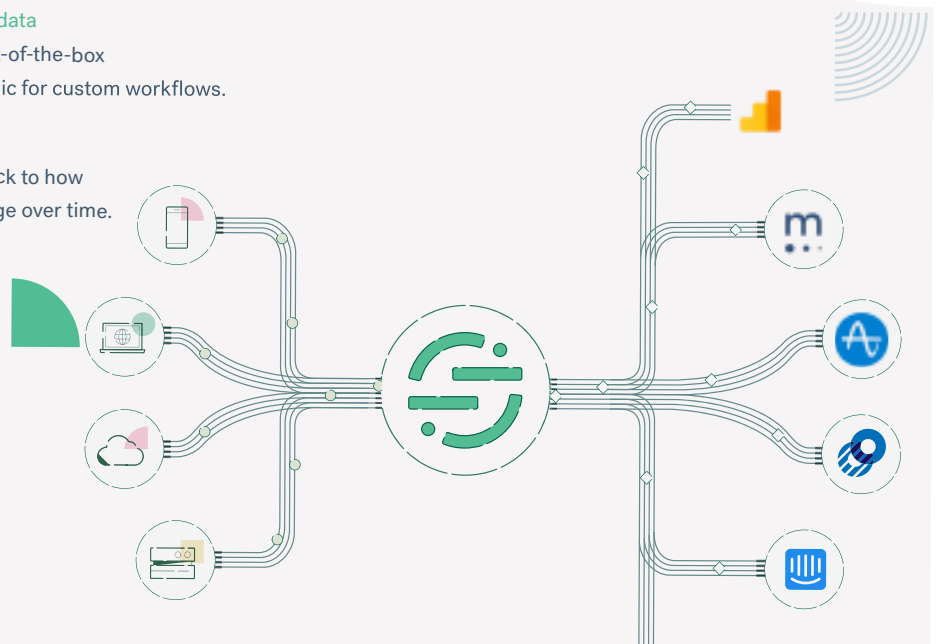
Streamline customer data collection that's trusted and actionable across every team in your organization.

Integrate: Integrate any tool for your customer data

Connect your data to hundreds of tools with out-of-the-box integrations in our catalog or build your own logic for custom workflows.

Scale: Future proof your customer-tech stack

Give your team the ability to adapt your tool stack to how your business and customer expectations change over time.



PROTOCOLS

Say goodbye to bad data.

**Protect the integrity of your data and your decisions.
Be confident your data is accurate in all the tools you use.**

Align: Standardize data collection throughout your organization

Keep your teammates in sync on what your data means with a shared and enforceable data dictionary.

Validate: Diagnose data quality issues before they impact production

Automate the QA process with real-time data validation and violation reporting to ensure perfect implementations.

Enforce: Keep the data in your marketing and analytics tools and warehouses clean

Quarantine any data that's not up to spec and resolve errors during implementation, so you can trust the data that runs your business.

PERSONAS

Personalize every customer interaction.

Power consistent customer profiles and audiences across your marketing stack. Discover what's most important to each user from your raw data and use it to personalize every interaction.

Unify: Consolidate customer identities into a single view

Unify user interactions across devices and channels—from web to mobile to email to in-store purchases—into one comprehensive user or account profile.

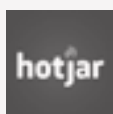
Segment: Use all of your customer data to create any audience imaginable

Combine historical user activity data, demographic data, and transactional data to build granular audiences and sync them across your marketing tools in real-time.

Personalize: Cater your marketing to each customer

Build personalized experiences by adapting your apps and websites to showcase specific products, categories, or content depending on your customer's most recent interactions.

Trusted by startups and the world's largest companies



trivago



BBVA

SPORTS
DIRECT



ClearScore



SEMARCHY

PASCAL BLANCHARD

Vice President of Sales EMEA

+33 (0)4 78 83 05 30

pascal.blanchard@semarchy.com

14 Chemin du Professeur Deperet 69160 FRANCE

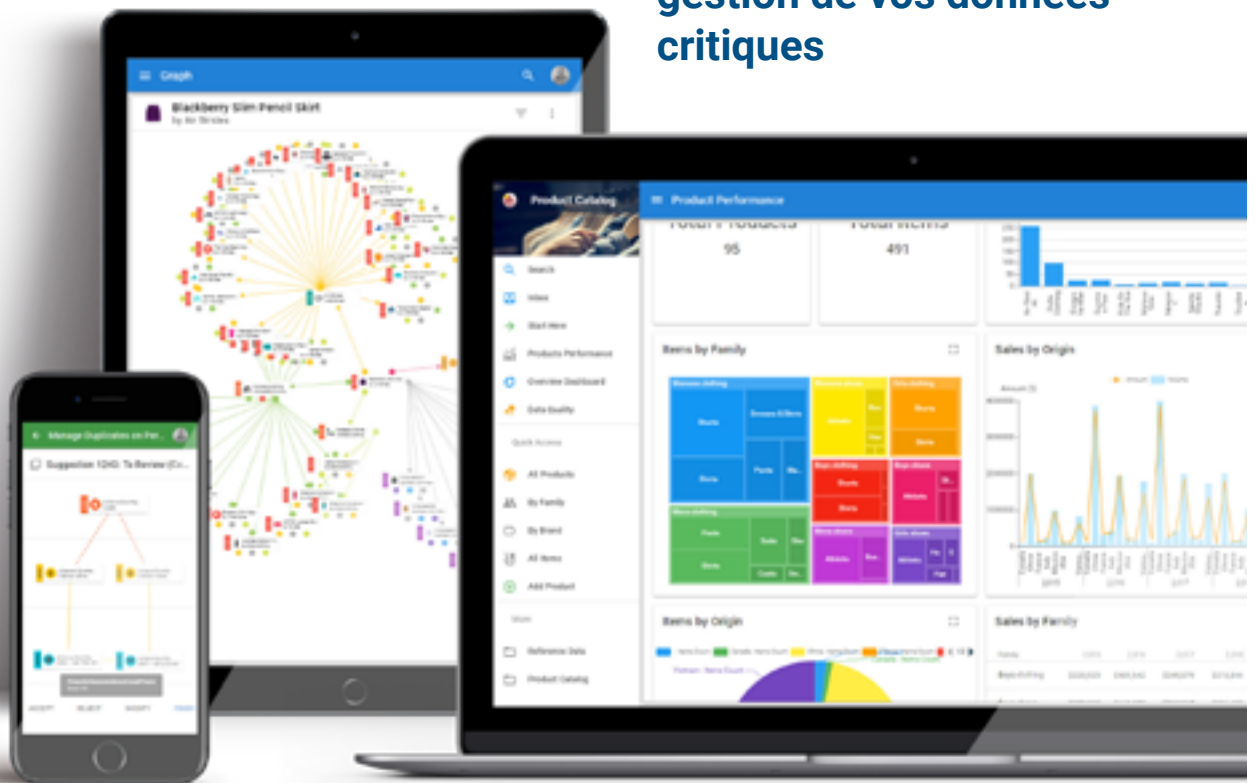
www.semarchy.com



Semarchy



Une seule et unique plateforme pour la gestion de vos données critiques



Découvrez

Profilez vos données sources et construisez vos catalogues.



Gouvernez

Construisez vos glossaires métier, vos règles et processus.



Intégrez

Connectez vos applications et données externes en APIs REST.



Mesurez

Analysez vos métriques grâce aux tableaux de bord.



Gérez

Déployez des apps pour les Data Champions et les métiers.

RDV STAND A12

**Reprenez le
contrôle de vos
données.**

STOP

**Intelligent
Data Hub™**

semarchy.com



Sinequa transforme la donnée brute en une information éclairée et contextualisée pour une meilleure prise de décision.

Sinequa est spécialisé dans la recherche et l'analyse de données basées sur les technologies du machine learning, du deep learning et du traitement automatique du langage naturel (NLP). Pouvez-vous nous en dire davantage sur votre cœur d'expertise ?

Nous permettons aux entreprises d'accéder de manière automatisée et systématisée à un capital de données intégrant des centaines de millions de documents écrits en langages naturels, et ce, dans plus de 300 langues dont 22 en traitement avancé. À travers nos outils de traitement du langage naturel (Natural Language Processing), nous offrons ainsi aux entreprises un accès unifié et contextualisé aux informations extraites de leurs données structurées et non structurées.

Auparavant, les entreprises prenaient leurs décisions en se basant uniquement sur les données structurées, soit 20 % seulement du volume total de leurs données disponibles. Grâce à notre technologie, les entreprises disposent enfin de la capacité à utiliser les 80 % de volume de données dite 'non structurées', non exploitées jusqu'à maintenant. Aujourd'hui, en enrichissant la donnée structurée avec des contenus non structurés et inversement, nous permettons aux entreprises d'optimiser la totalité de leur capital de données. Concrètement, nous transformons la donnée brute en une donnée éclairée et contextualisée, autrement appelée Information. C'est la garantie de prendre ses décisions dans les meilleures conditions possibles.

L'analyse combinée de données structurées et non structurées constitue un levier majeur d'économies. Qu'en est-il ?

Depuis les années 70, les spécialistes du business intelligence, comme SAP avec l'acquisition notamment de « Business Objects », ont énormément travaillé sur l'analyse de la donnée structurée. Depuis, au fil des années, les questions business ont gagné en complexité. En effet, les entreprises génèrent de plus en plus des contenus non structurés (emails, documents, vidéos, chats, contenus des réseaux sociaux...) où il y a plus de données textuelles que de données structurées - typiquement gérées dans des bases de données alimentées par des logiciels d'entreprise (ERP, CRM, BI, etc.). Bien qu'elles soient de plus en plus conscientes de l'importance des contenus non structurés, la quasi-totalité des entreprises se sont focalisées principalement sur l'exploitation des données structurées.

Sinequa a fait le choix de concentrer ses efforts sur la partie la plus difficile : l'analyse de ces données non structurées (texte, image, vidéo, audio), car ces données constituent un véritable trésor en termes de valeur et de levier économique pour les entreprises. Par exemple, dans le domaine financier, les algorithmes développés sur la base des données structurées ne sont plus différenciants pour une évaluation pertinente

des cours en bourses. Les données non structurées sont donc indispensables pour créer des informations analysables et donc capitales pour les décisions d'investissement.

Dans le secteur pharmaceutique, les recherches qui se basent sur l'analyse d'essais pré-cliniques extrêmement coûteux permettent aux laboratoires pharmaceutiques de raccourcir considérablement la mise sur le marché de nouveaux médicaments. Il est donc primordial de comprendre et de croiser d'importants volumes de données structurées et non structurées, et nos technologies, incluant des mécanismes de Machine Learning, le permettent.

Dans quelle mesure vos solutions, appliquées à des usages métiers, permettront-elles de répondre aux problématiques « data » de vos clients ?

Nos solutions sont très utiles pour la lutte contre la fraude et contre le blanchiment d'argent, alors que les fraudeurs ont recours à des pratiques difficilement détectables à partir des données structurées. Le potentiel du Machine Learning réside donc dans la vitesse de détection des transactions frauduleuses ou la « pattern recognition » dans le domaine des investigations, par exemple. Dans ce cadre, nous travaillons avec un organisme européen sur la détection des collusions illégales entre entreprises. Cela consiste à analyser le contenu des échanges entre les entreprises et à détecter les signaux faibles. Nous collaborons également avec une institution étatique qui analyse les transactions bancaires en Europe, afin de détecter des transactions bancaires frauduleuses, y compris les opérations de financement du terrorisme.

A propos de Sinequa

Sinequa est un éditeur de logiciel indépendant spécialisé dans la recherche et l'analyse de données, basées sur les technologies de l'IA, pour les entreprises du Global 2000 et les grandes administrations.

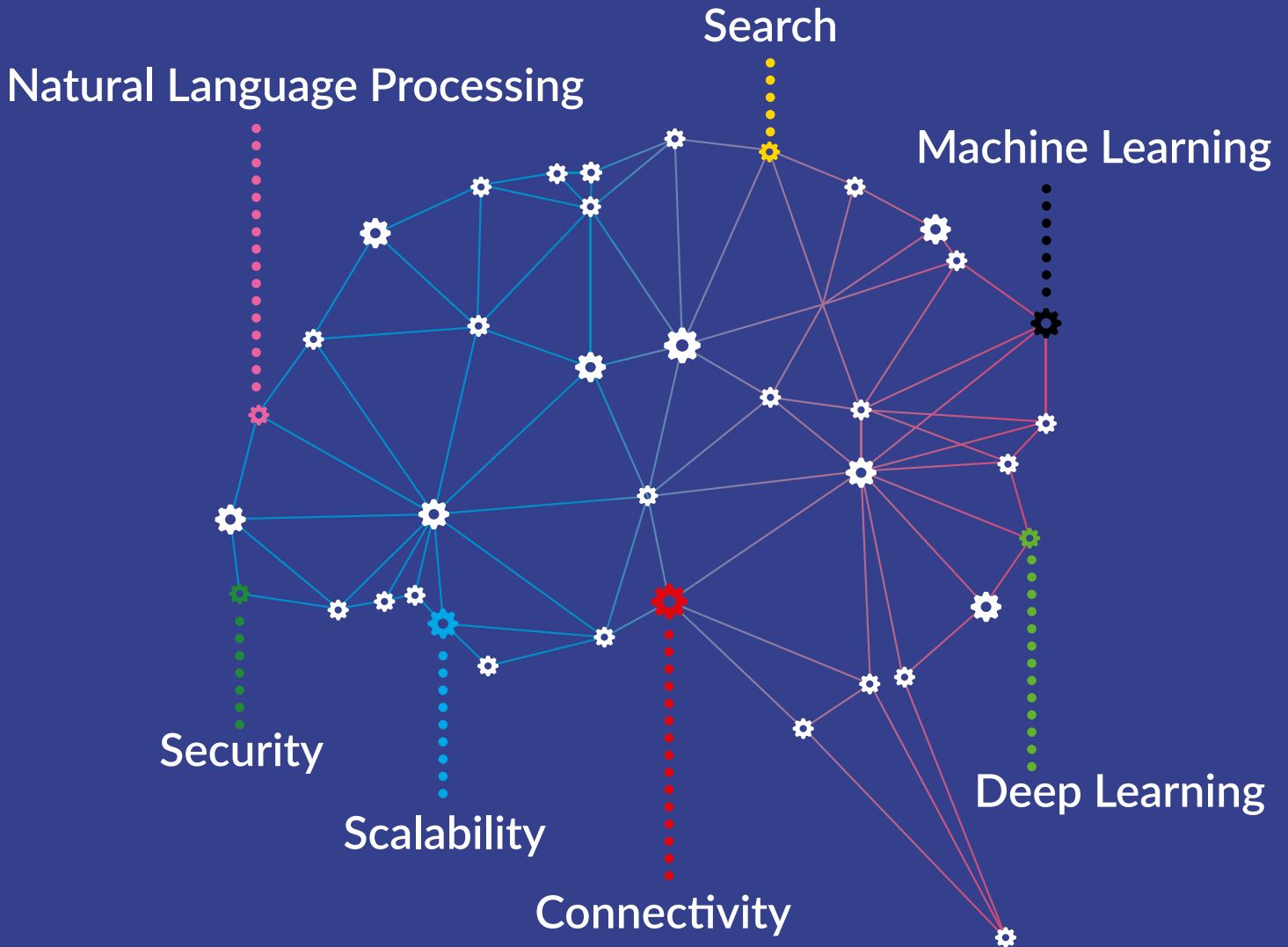
Grâce à ses capacités de machine learning, deep learning et de traitement automatique du langage naturel (NLP), Sinequa transforme les données brutes de ses clients en informations pertinentes, leur permettant ainsi de prendre des décisions éclairées.

Reconnu comme un leader mondial du marché dans le rapport « 2018 Magic Quadrant for Insight Engines » de Gartner et le rapport « Forrester Wave™ : Cognitive Search & Knowledge Discovery Solutions Q2 2019 », Sinequa développe son expertise et son activité à travers le monde grâce à un vaste réseau de partenaires et d'alliances technologiques.

SINEQUA

Become Information-Driven

BECOME INFORMATION-DRIVEN WITH THE LEADING AI-POWERED SEARCH PLATFORM



Follow us



SKAPÂNÊ, FOURNISSEUR DE PLATEFORMES BIG DATA

QUI EST SKAPÂNÊ ?

Skapâne base sa valeur sur sa maîtrise de la technologie du Big Data et du Machine Learning déployés en temps réel. Grâce à sa plateforme Hadoop / Spark disponible dans un cloud privé français, Skapâne facilite l'accès à ces technologies aux entreprises de toute taille.

Créée en 2015, Skapâne est composée de 19 personnes, dont 6 expérimentés cumulant une centaine d'années d'expérience. Rompue aux environnements techniques et aux exigences de divers SI, notre équipe est capable d'appréhender les problématiques d'intégration, de sécurité et de confidentialité de nos solutions avec vos systèmes.

Fort de son expertise et de son savoir-faire, Skapâne vous donne les outils pour appréhender, connaître et visualiser vos données. Le Big Data et l'IA sont à votre portée grâce à notre plateforme orientée traitements Big Data, Machine Learning et IoT. **Demandez-nous la création de votre compte et vous voilà maître de vos datas !**

DATALAB AS A SERVICE...

Skapâne vous propose de disposer d'une plateforme Big Data et Machine Learning 'as a Service', sans projet de déploiement chez vous. Notre Datalab permettra à vos analystes, data scientists de découvrir, de manipuler et de représenter vos données. **Se lancer dans un projet de data science sera chose aisée pour vos équipes.**

Des fonctions sont nativement intégrées à notre datalake ; nous proposons entre autres :

- L'import / export de données de sources diverses ;
- Des fonctions d'anonymisation, d'allonymisation ;
- La création automatique de Tables HIVE/HBase.

Dans le cadre du RGPD, nous avons une politique stricte concernant vos données. Nous disposons de notre propre cloud privé et **vos données sont hébergées en France.**

... UTILISÉ PAR LA BANQUE ET LA FINANCE

Notre Datalab vient en soutien aux équipes Risques, Conformité, Clients dans le secteur bancaire et la Finance. Nos capacités de traitements permettent à chacun de s'approprier ses données. Nos équipes de data scientists les assistent sur des projets de data science tels que la détection de transactions frauduleuses, le scoring de risque, ...

Dans le cadre du KYC, Skapâne propose sa solution de Smart Process Automation pour faciliter le traitement des dossiers client et/ou prospect. Entre autres, elle permet :

- De reconnaître/ catégoriser/ typer des documents ;
- D'extraire de l'information (nom, prénom, adresse, ...) ;
- De vérifier les informations intra-documents / extra-documents.

Les retours de nos traitements sont facilement intégrables dans votre SI.

UN DATALAB ADAPTABLE À L'IOT

Conscients de l'avènement des objets connectés, nous avons intégré à notre plateforme les composants permettant de répondre aux besoins des acteurs de l'IoT. Nous sommes en mesure de collecter vos messages, de les analyser en temps réel, de remonter des alertes et de les stocker dans notre datalake.

LA DATAVIZ

Tout traitement de données nécessite une visualisation de celles-ci. À chaque métier, à chaque division, une visualisation des mêmes données sera différente. Le volume de données à traiter est grandissant et les outils de visualisation traditionnels n'offrent pas toujours la possibilité de les visualiser sur une même représentation graphique. Notre solution de visualisation Big Data, Apache Superset, rend celle-ci possible. Toutes les données analysées peuvent être visualisées, projetées sur un même visuel ; **sans besoin de précalculer des agrégats.**

NOS TECHNOLOGIES

Skapâne se différencie par son choix de l'autonomie. Nous n'utilisons que des logiciels en Open Source et notamment ceux directement issus de la fondation Apache (Hadoop, Spark, Superset, ...).



NOTRE PLATEFORME EN CHIFFRES

2 Data centers 960 CPUs 8 To de RAM
90 serveurs 1 Po de disques

Strengthen your business with our data engine!



SKAPÁNĚ

#BigData

#DeepLearning

#IoT

#AI

#DataViz

#MachineLearning

Stand
A16



ACCOR CHOISIT SNOWFLAKE POUR CENTRALISER ET VALORISER SES DONNÉES.

Avec plus de 4 900 hôtels, Accor a récemment réalisé un important virage stratégique en cédant la propriété des murs des hôtels qu'il opère et en investissant dans ses capacités digitales. Le leader mondial de l'hospitalité a ainsi choisi Snowflake pour sa nouvelle plateforme de donnée. Alimentée par un hub et un data lake, ce système central a pour objectif de permettre une meilleure valorisation des données collectées par l'entreprise et de porter son ambition d'hospitalité augmentée.

« Le groupe a compris qu'il avait besoin de refonder son SI data s'il voulait véritablement faire de la donnée un des piliers de sa stratégie. » explique Michael d'Eboli, Senior Vice President Engineering & Delivery du groupe, qui ajoute : « Le secteur de l'hôtellerie ne sera jamais complètement dématérialisé – nos clients finaux paieront toujours pour une chambre et un lit –, mais le digital a pris une place prépondérante au cours des dernières années et nous nous devons d'être au niveau des pureplayers. »

Des données au service d'une expérience client améliorée et plus personnalisée

Accor souhaite donner corps à sa promesse d'hospitalité augmentée et propose à présent une offre étoffée en matière, d'événementiels, de restauration, de coworking et d'activités diverses. Le digital permet de distribuer ces différents services de manière complémentaire, mais également de multiplier les points de contact avec ses clients. Au cours des prochains mois, le nouveau programme de fidélité – baptisé All – proposera un point d'entrée unique vers ce nouvel écosystème et multipliera ainsi le volume de données collectées par le groupe.

« Qu'elles soient physiques ou digitales, chacune des interactions que nous avons avec nos clients génère de la donnée. Si nous nous dotons du savoir-faire et des outils nécessaires pour les analyser, nous serons en mesure de proposer une expérience personnalisée à nos clients. C'est un enjeu de compliance important pour le groupe, mais surtout une opportunité. » nous explique Michael d'Eboli.

Unifier et centraliser une infrastructure de données vieillissante

Le groupe a donc choisi de faire évoluer son SI data – constitué de nombreuses solutions développées en interne et reposant sur trois datawarehouses distincts, alimentant trois solutions de business intelligence elles aussi distinctes – vers une architecture moderne et unifiée permettant d'atteindre les objectifs de qualité, performance, et sécurité au sein des différentes entités et régions du groupe.

« On se doit de créer un cadre de confiance. Nous utilisons des dizaines de To de données, si un Ko de données est faux, alors c'est tout un système qui s'écroule. Le SI précédent ne pouvait pas garantir cette qualité de données », ajoute Michael d'Eboli.

Fiabilité, robustesse et scalabilité pour maîtres mots

La nouvelle solution mise en place répond donc à trois impératifs : fiabilité, robustesse et scalabilité. Au centre de cette nouvelle organisation, un data lake, hébergé dans le cloud, vient ainsi alimenter un datawarehouse – porté par la technologie Snowflake – qui incarne la vérité unique de la donnée pour l'ensemble du groupe.

« Plusieurs choses nous ont convaincus chez Snowflake. Il s'agit d'une solution SaaS dans laquelle nous pouvons avoir confiance. De plus, le fait qu'elle soit basée sur une approche SQL, une technologie que tout le monde maîtrise et apprécie, nous évite d'avoir à reformer les équipes et nous permet d'être rapidement opérationnel », détaille Michael d'Eboli. « Nous allons pouvoir ouvrir l'accès de manière sécurisée à tout ou partie de nos jeux de données pour les différentes entités géographiques et métiers ainsi qu'à nos prestataires, ce qui facilitera maximisera l'utilisation et la valorisation des données ».

« Nous sommes actuellement en cours de déploiement et versons tous les jours 300 Go de données dans notre data lake », précise Michael D'Eboli. Avec l'essor combiné de l'IoT et de la 5G, Accor prévoit une augmentation exponentielle des volumes de données à assimiler et analyser et espère à présent disposer de tous les outils nécessaires pour lui permettre d'améliorer l'expérience de ses clients en proposant des services plus divers et personnalisés.

«Plusieurs choses nous ont convaincus chez Snowflake (...) une solution SaaS dans laquelle nous pouvons avoir confiance (...) basée sur une approche SQL qui nous permet d'être rapidement opérationnel.»



MICHAEL D'EBOLI
SENIOR VICE PRESIDENT ENGINEERING
& DELIVERY, ACCOR

L'ENTREPÔT DE DONNÉES CONÇU POUR LE CLOUD

Prenez de meilleures décisions, plus rapidement

Pour faire croître votre entreprise, offrez un accès à l'ensemble des indicateurs analytiques issus de l'intégralité de vos données à tous vos utilisateurs, lorsqu'ils en ont besoin.

L'évolution de votre entreprise découle de l'évolution de vos analyses

Minimisez vos coûts grâce à une solution zéro gestion avec facturation à la seconde, qui évolue à l'infini instantanément.

Des Clouds multiples

Choisissez le fournisseur de votre infrastructure, conservez vos données et vos applications à leur emplacement, mettez en place une stratégie multicloud et assurez la souveraineté des données.

CHANGEZ D'ENTREPÔT DE DONNÉES SI CELUI-CI NE VOUS PERMET PAS D'AVANCER

Libérez-vous des systèmes qui inhibent vos capacités à innover, à évoluer et à faire croître votre entreprise.

TÉLÉCHARGEZ NOTRE DUMMIES BOOK



snowflake.com

STIBO SYSTEMS

JOSEPH COHEN

EMEA Marketing Manager

+33 (0) 6 83 60 31 96

jodc@stibosystems.com

43 Bd Haussmann, 75009 Paris, France

www.stibosystems.com/fr



StiboSystems

MASTER DATA MANAGEMENT



Stibo Systems est leader indépendant mondial dans les solutions PIM et de gestion de référentiel de données multi-domaines : Produit, client, fournisseurs, logistiques et autres...

La plateforme PIM & MDM Stibo Systems permet aux sociétés de disposer d'un référentiel unique de données puissant et souple, partagé par les utilisateurs des différents métiers.

Depuis 30 ans, Stibo Systems aide ses clients à développer une source d'information stratégique unique et fiable mais aussi à gérer et améliorer la qualité de leurs informations de référence.

Afin de mettre en place une cohérence cross-canal, des grands noms du marché ont fait confiance à Stibo Systems pour intégrer

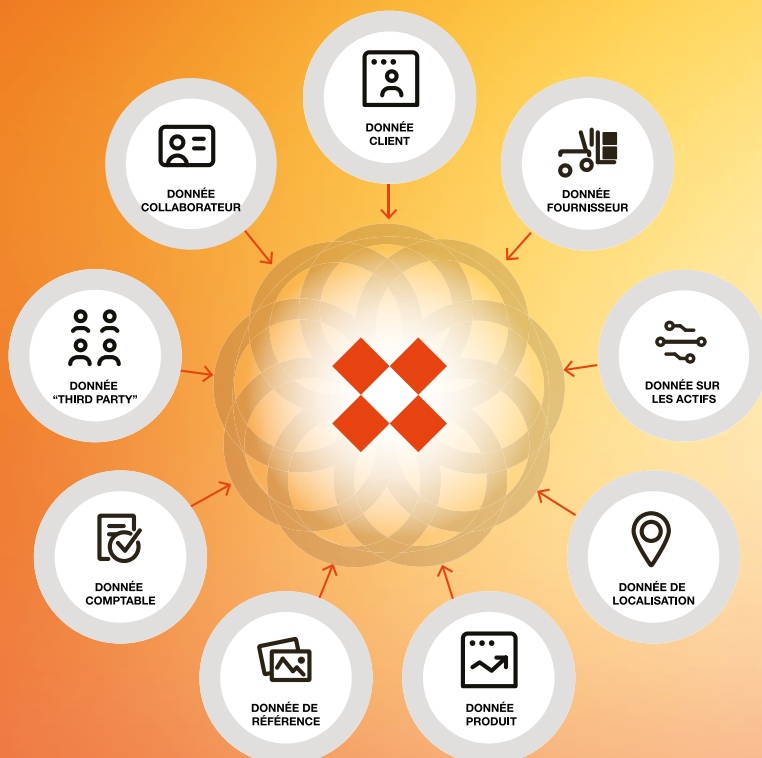
leurs informations produit, client, fournisseur et d'autres données organisationnelles. Cela leur permet d'être plus pertinents dans leur prise de décisions, d'optimiser leurs activités commerciales et de créer de la valeur.

Stibo Systems est reconnu par les analystes comme le « Leader indépendant du PIM (Product Information Management) au niveau mondial » (Gartner) et le « N°1 en termes de satisfaction client » (Ventana).

Stibo Systems est présent dans le monde avec 20 filiales (Europe, Amérique du Nord et Asie-Pacifique) et appartient au groupe privé Stibo A/S, fondé en 1794 et basé à Aarhus, au Danemark.

Gérez les données avec une solution unique, qui connecte les données de plusieurs domaines.

Connectez, partagez et exploitez des données provenant de sources isolées et disparates, de sources tierces et de partenaires.



Une approche stratégique pour acquérir, gérer et partager des informations et du contenu avec vos clients, collaborateurs, fournisseurs, partenaires et autres...

Acquérir

Consolidez les informations issues de multiples sources cloisonnées, internes et externes.

Gérer

Contrôlez et enrichissez la hiérarchie des enregistrements, la structure, les validations, les approbations, la gestion des versions, les attributs, les descriptions, etc.

Partager

Garantissez un accès interne fiable et cohérent et syndiquez les données pour les clients, les fournisseurs, les distributeurs et les partenaires de la chaîne de valeur.

Des solutions axées sur la priorité de votre entreprise

Gestion des données Multi-Domaine

Une solution unique pour créer une synergie entre les données de plusieurs domaines et fournir des informations exploitables qui contribuent au succès de l'entreprise.

Gestion des données Produits (PIM)

Acquérir, gérer et partager des données produit pour fournir aux collaborateurs, aux partenaires et aux systèmes des informations cohérentes, exactes et actualisées.

Gestion des données Client

Utilisez des données client de haute qualité pour vos initiatives d'entreprise et améliorez le marketing, les ventes, l'expérience client, et bien plus.

Gestion du cycle de vie du produit

La solution PLM met de l'ordre dans l'entreprise grâce à une planification, une conception et un développement intégrés. Elle permet de mettre des produits « gagnants » plus rapidement sur le marché.

Syndication de données produit

PDS est une solution simple, rapide et économique qui permet aux fabricants et aux marques de mettre à jour et de partager des données produit avec les détaillants et les fournisseurs de services de contenu (CSP).

La Plateforme STEP

Plateforme de pointe offrant une agilité permettant de connecter et de tirer des informations dans plusieurs domaines (comprenant également les ressources, les fournisseurs, les sites, les employés, etc.).

Les bénéfices des solutions Stibo Systems



Transformation digitale

Mettez en œuvre et accélérez les changements axés sur les données pour stimuler l'innovation et la différenciation concurrentielle



Qualité des données

Des données cohérentes et précises assurent fiabilité, confiance et décisions métier judicieuses



Rentabilisation

Accélérez votre retour sur investissement avec de nouveaux éclairages basés sur les données qui augmentent les ventes incitatives, les ventes croisées et favorisent la croissance



Orientation client

Offrez des expériences personnalisées et attrayantes qui incitent vos clients à revenir.



Conformité optimisée

Assurez la gouvernance et la conformité et réduisez les risques en éliminant les processus sujets aux erreurs



Mise sur le marché accélérée

Lancez et intégrez les produits plus rapidement et favorisez l'agilité collaborative sur toute la chaîne de valeur

“ Stibo Systems convient parfaitement aux entreprises recherchant une excellente plateforme de gestion de données multi-domaine, dotée d'un portail fournisseur et d'un modèle de gouvernance de donnée exceptionnels.

— The Forrester Wave™ : Product Information Management Solutions, Q2 2018

FORRESTER®





Les Métiers revendiquent l'autonomie et la réactivité de la Self BI

- La BI actuelle est incapable de répondre à une question du Comité de Direction qui commence par « Pourquoi ? » dans un laps de temps proche du fameux « temps réel ».
- Les outils de préparation des données étant complexes et techniques, les Métiers ne peuvent pas les utiliser pour chercher directement les bonnes données et n'ont d'autres choix que d'expliquer leurs besoins à l'IT.
- Ils doivent de surcroît s'adapter aux contraintes du projet informatique.

Ce mode opératoire génère un nombre d'itérations qui, régulièrement, tue la spontanéité de l'action et finit par user les acteurs.

La Self BI révolutionne ce mode opératoire en déportant au niveau des Métiers la capacité de construire leurs analyses et en recentrant le service IT sur son rôle de fournisseurs de données propres et sécurisées.

La DataViz ne fait pas la Self BI

Elle offre la possibilité aux Métiers de visualiser les données de l'entreprise en construisant eux même leurs restitutions. Mais les Métiers ne sont toujours pas autonomes dans la préparation des données qui alimentent ces restitutions.

Ainsi la DataViz n'évite pas les itérations liées aux phases de préparation de données.

Le mythe de l'entrepôt de données complet

La DataViz pourrait répondre aux attentes de la Self BI si les données mises à disposition répondaient à toutes les options d'exploration des Métiers.

En réalité, chaque jour, les gisements de données ne correspondent pas exactement à ce qui est attendu et doivent être redéveloppés sans cesse par l'IT.

L'entrepôt de données complet est un mythe, qui n'existera jamais. Pire, il alimente au fil des années le sixième continent de données déchets que chaque entreprise continue de nourrir.

La Self BI Suadeo, une réalité qui transforme et valorise les rôles des Métiers et de l'IT.

Grâce à sa nouvelle technologie «User Data» (Donnée d'Usage), les Métiers accèdent de manière sécurisée aux Catalogues et aux structures des données. Dès lors, autonomes, ils peuvent à la volée créer ou modifier les rapprochements de fichiers sans toucher à l'intégrité des données élémentaires, créer leurs propres indicateurs et en visualiser les résultats avec les outils DataViz de Suadeo.

L'IT devient fournisseur de services à valeur ajoutée autour de la donnée.

Au niveau des restitutions, il n'itère plus avec les Métiers. Il se recentre sur la mise à disposition et la qualité des données de base de l'entreprise, les documente, les sécurise et les rend accessible rapidement.

Avec la Self BI Suadeo, l'IT devient partenaire des Métiers.

Cerise sur le gâteau, la Self BI Suadeo met fin aux espaces de stockage de données énormes et toujours grandissants et répond ainsi aux objectifs de réduction des coûts de nombreuses entreprises.



Prenez 5 années d'avance en découvrant notre
plateforme logicielle de Business Intelligence Agile

DATA VIZ

Self BI

Ne vous contentez pas de regarder

ORCHESTREZ VOS DONNÉES

Diffusez une nouvelle culture de pilotage, collaborative
et transverse **pour aller au-delà de la data.**

**Offrez à vos collaborateurs une complète autonomie,
et une réactivité réelle dans la construction de leurs
analyses.**

DATA MANAGEMENT

TABLEAU

VÉLINA COUBES

Head of marketing France & Southern Europe

+33 1 70 91 88 94

southerneuropemarketing@tableau.com

19 boulevard Maesherves, 75008 Paris France

www.tableau.com



Oney Data intègre Tableau dans le but d'optimiser la gestion de données auprès de 1000 utilisateurs

Filiale de Oney Bank, Oney Data s'appuie sur de grands volumes de données pour optimiser leurs décisions commerciales. Grâce à Tableau, Oney Data centralise la gestion des données, en simplifie la lisibilité et facilite ainsi le travail quotidien des collaborateurs.

Monétiser la donnée grâce à Tableau

Oney Data a pour mission d'aider les distributeurs à analyser leurs données et à les monétiser. À ce titre, Oney Data dispose d'un énorme volume de données très hétérogènes : des données tickets de caisse et magasin, des nomenclatures et attributs produits, des cartes de fidélité ou encore des éléments géomarketings. Une fois collectées, ces données n'ont de réelle valeur que si elles peuvent être interprétées. En utilisant la solution Dataviz de Tableau, Oney Data extrait de l'information de ces données qu'elle commercialise ensuite auprès de ses clients.

« Notre plateforme est ainsi devenue le point central de nos activités, que ce soit pour des dashboards disponibles à près de 1000 utilisateurs, des outils d'intelligence augmentée, d'optimisation d'assortiment et de promotion ou des analyses adhoc créées afin de répondre à un besoin précis, et que nous pouvons mettre à disposition sous une forme extrêmement interactive, intuitive et actionnable », explique Elodie Perrigaud, Chief Data Officer chez Oney Data.

Optimiser la gestion des données

Dès les premiers rapports effectués avec Tableau, Oney Data a pu considérablement améliorer sa gestion du flux d'informations. Avant le déploiement de Tableau, il fallait jongler entre six rapports contre un seul aujourd'hui qui centralise toute l'information. Les collaborateurs de Oney Data disposent alors d'une vision globale de la performance. Intégré dans la plateforme d'insight sharing, Tableau a complètement modifié la perception des fournisseurs. Auparavant, ils utilisaient les données comme une donnée brute qu'ils retraient dans Excel. Aujourd'hui, ils peuvent s'appuyer sur les insights créés par sa visualisation et les algorithmes comme outil de collaboration.

« Le meilleur exemple ? Depuis la mise en production de la nouvelle version, nous avons augmenté de près de 20% notre nombre d'utilisateurs chez les fournisseurs internationaux ».

Un outil adapté aux acteurs du retail

En charge de l'analyse et la monétisation des données de l'activité du groupe Auchan, la connaissance et le parcours client sont au coeur de l'activité de Oney Data. Or, la gestion de données et la production de rapports ralentissaient le traitement ainsi que l'activité commerciale des partenaires. Grâce à l'interactivité et la variété de visualisations disponibles, Oney Data parvient à trouver le meilleur moyen pour analyser la donnée et la transformer directement en plans d'actions pour l'entreprise.

Ces données actionnables permettent ainsi aux fournisseurs (toutes les grandes marques, mais aussi des marques locales, accèdent régulièrement à la plateforme) de multiplier les ventes en ciblant précisément leurs clients et en assortissant mieux leurs produits. De cette façon, fournisseurs et acheteurs peuvent échanger sur une base d'informations fiables et accessibles en temps réel. Habituellement, les entreprises, lorsqu'elles réfléchissent à un projet de big data, pensent très souvent à la partie technique sans mettre l'accent sur la façon dont les datas vont être interprétées ni sur leur utilité finale. C'est ce que Tableau a permis de réaliser chez Oney Data : des données analysées à travers une présentation visuelle interactive et intuitive.

Une grande amélioration des performances

L'adhésion des équipes a été immédiate et les collaborateurs travaillent en autonomie pour effectuer le suivi opérationnel ou budgétaire. De plus, grâce à Tableau, ils peuvent désormais consacrer du temps à analyser et interpréter les données plutôt que de les chercher, les partager, les extraire, etc.

Pour les utilisateurs : une performance passée de 1 minute au premier chargement à moins de 5 secondes et une richesse et flexibilité largement accrue grâce à l'intégration de nouvelles sources de données (client anonymisé, géomarketing, navigation web, marché...).

Grâce à cette solution visuelle, customisable et scalable, Oney Data est passé de 120 à 200 fournisseurs (internationaux et locaux) et de 500 utilisateurs de Tableau à près de 1000 en un an seulement.

Avec Tableau, vos données peuvent faire la différence !

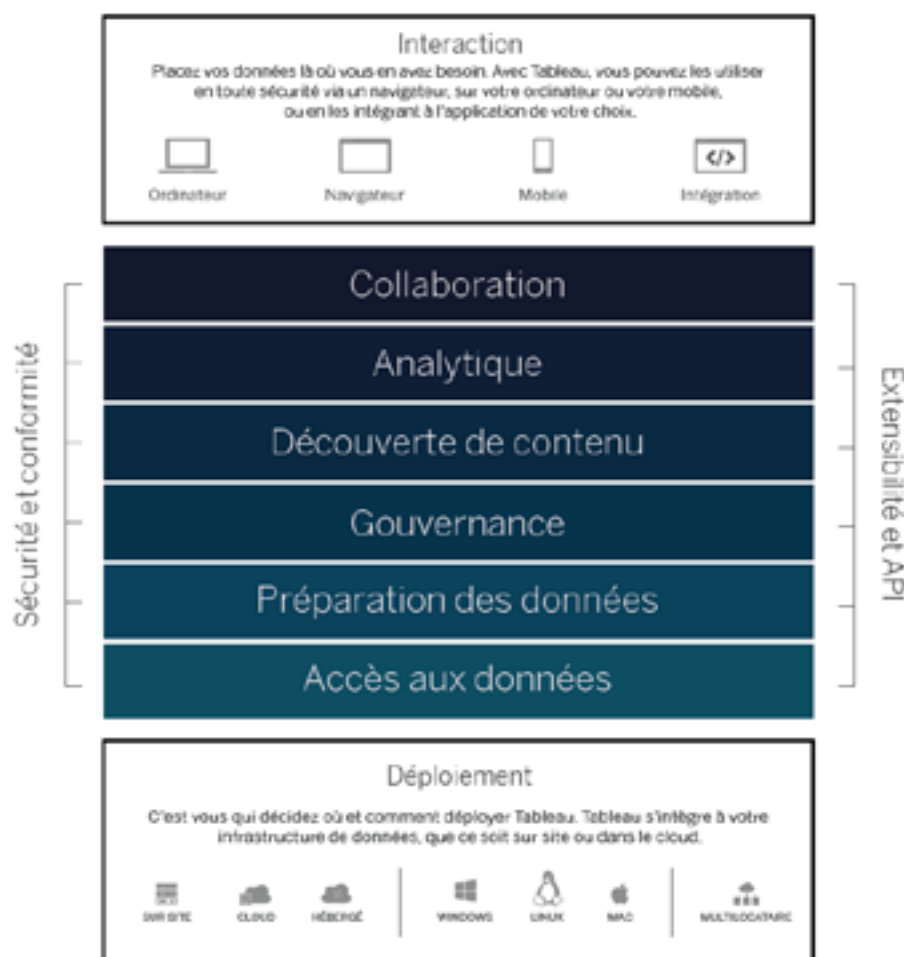
Tableau software aide les plus grandes entreprises du monde à libérer la puissance de leurs ressources les plus précieuses : leurs données et leurs collaborateurs.

Nous proposons une plateforme analytique intuitive permettant de répondre à des questions dont vous ne soupçonniez pas l'existence et de faire des découvertes inédites. En quelques clics, Tableau analyse de grands volumes de *datas* provenant de sources diverses et les publie dans des tableaux de bords et des outils de visualisations. Afin de créer une vue d'ensemble et des reportings BI visuels, ces données analysées peuvent être combinées à des graphiques, des diagrammes et partagés avec des collaborateurs.

Grâce à Tableau, les utilisateurs transforment leurs données en informations exploitables leur permettant ainsi d'être plus réactifs et plus performants dans leurs prises de décisions.

Des grandes multinationales aux startups en passant par les TPE, la souplesse et la richesse de la solution lui permettent de s'adapter à tout type d'entreprises, tout en répondant aux exigences les plus strictes en matière de gouvernance et de sécurité de la donnée.

Découvrez la plate-forme :



Conçue pour l'utilisateur individuel, adaptée à l'entreprise

La plate-forme analytique Tableau est extrêmement complète, puissante, sécurisée et flexible. Fournissez à vos collaborateurs un accès intelligent, visuel et direct à leurs données à chaque étape du processus analytique.

Notre interface intuitive facilite l'apprentissage et rend l'utilisation passionnante. La flexibilité de notre architecture assure une parfaite adéquation avec votre environnement et vos données.

Lorsqu'il s'agit de mettre la puissance des données à la portée des utilisateurs, seul Tableau donne la priorité à la façon dont ils les voient et les comprennent. Notre plate-forme robuste et évolutive se révèle indispensable, même pour les plus grandes entreprises du monde.



« Il n'y a pas une seule application de la Gendarmerie et de la Police qui ne consomme ou qui ne produise de la donnée sans passer par Talend. »

Commandant Christophe Tedesco, Chef de Section au STSISI,
Gendarmerie Nationale

Industrie : Secteur Public

Cas d'usage : Sécurité intérieure

Problématique : Croiser de nombreux fichiers sécuritaires de l'État

Produits Talend utilisés : Talend Data Fabric : *Talend Data*

Integration - Talend Application Integration (ESB) - Talend Real-Time Big Data

Information : Siège social : France - 10,001+ employés

Bénéfices :

- 3 To de données par jour

- De 300 000 identités criblées par an à un million par mois

- De 1 à 30 secondes pour restituer l'information en provenance d'une douzaine de fichiers et centaines d'applications

FACILITER LES ENQUÊTES AU NOM DE LA SÛRETÉ DE L'ÉTAT



Force humaine de près de 100 000 hommes et femmes placée sous l'autorité du Ministère de l'Intérieur, la Gendarmerie Nationale est une institution militaire garante de la sécurité et de la paix des concitoyens, et de la protection de leurs biens. Elle assure des missions de police judiciaire, d'assistance à personnes, de maintien de l'ordre et participe à la défense de la Nation, y compris sur les théâtres d'opérations extérieures.

En 2010, la France est désignée pays hôte de l'Euro 2016 par l'UEFA. Un an avant l'ouverture de cet événement majeur, qui va intéresser la planète entière, le terrorisme frappe le pays de plein fouet. « C'est une grande prise de conscience », se souvient le Commandant Tedesco, Chef de Section au STSISI (Service des Technologies et des Systèmes d'Information de la Sécurité Intérieure). « Nous nous rendons compte que des gens allaient accéder à des endroits très sensibles - comme des fans zones - et nous devions être capables d'identifier les individus dangereux. »

La décision est alors prise de créer une solution pour croiser les informations de sécurité qui étaient jusque-là stockées dans des fichiers indépendants. C'est alors que Talend entre en jeu.

Un autre événement d'envergure internationale et aux enjeux sécuritaires importants a propulsé Talend au cœur du dispositif. Fin 2015, Paris accueille la 21^e conférence sur les changements climatiques (COP 21). Les attentats du 13 novembre 2015 à Paris et du 14 juillet 2016 à Nice, poussent le Ministère de l'Intérieur à renforcer encore le dispositif de traitement automatisé de données personnelles dans le cadre de la lutte anti-terroriste. Baptisé ACCReD (Automatisation de la Consultation Centralisée de Renseignements de Données), ce dispositif de police permet la « consultation automatique

et simultanée » de fichiers déjà existants, notamment celui des personnes recherchées, des fichés S, des enquêtes administratives liées à la sécurité publique, pour faciliter la délivrance et le suivi des autorisations d'accès aux installations nucléaires ou d'importance vitale.

Croiser de nombreux fichiers sécuritaires de l'État

Les fichiers en France sont constitués de fichiers nationaux d'État, auxquels s'ajoutent des fichiers au niveau européen accessibles par les mêmes organismes d'État, et parfois par des organismes privés. La CNIL (Commission Nationale de l'Informatique et des Libertés) en contrôle l'usage.

Pour réaliser sa mission, la Gendarmerie Nationale dispose ainsi de centaines d'applications et de douzaines de fichiers en provenance de la Police, des Ministères de l'Intérieur, des Armées, de la Justice, de l'Environnement. Le TAJ, par exemple, est un fichier de traitement des antécédents judiciaires. CAS-SIOPEE sert au suivi des procédures côté justice. La rédaction des pièces de procédures se fait au sein du Logiciel de Rédaction des Procédures (LRP) en police et en gendarmerie. Enfin, les citoyens de l'Union européenne circulant librement au sein de l'Espace Schengen, le Système d'information Schengen (SIS), permet de suivre des individus signalés, des véhicules volés, tout comme le Fichier FOVeS pour « Fichier des objets et véhicules signalés. »

« Notre problème », se souvient le Commandant Tedesco, « c'est que tous ces fichiers étaient indépendants. Pour accéder à l'information, il fallait les consulter un par un. Il nous fallait parfois taper 4 à 5 fois le même nom, le même prénom pour accéder à l'information, ce qui prenait un temps infini. »

« Talend est devenu le chef d'orchestre de tous les échanges inter-applicatifs entre nos fichiers sécuritaires et un élément critique de notre Système d'Information. Si nous ne transmettons pas les informations dans les temps, nous sommes appelés dans la minute. La solution est stratégique. »



Afin de permettre aux utilisateurs l'accès à l'information recherchée, Talend joue un rôle clef en gérant, de manière unifiée et contrôlée, toutes les requêtes aux différents fichiers. Le Commandant Tedesco explique :

« Nous récupérons tous ces fichiers dans le Système d'Interfaces et de Référentiels (SIR) et assurons les échanges inter-applicatifs avec Talend (Data Integration et Real-Time Big Data). Il n'y a pas de données centralisées. ACCReD reçoit une liste d'identités à cribler. Talend va consulter des douzaines de fichiers déjà existants et restituer l'information. »

La Gendarmerie Nationale a également mis en place des référentiels pour gérer toutes ces données. Quand des recherches sont effectuées au sein des systèmes, qu'il s'agisse d'une commune, d'une marque de voiture, d'un stupéfiant, d'un type de délit, il faut en effet être sûr que l'information est bien la même entre les systèmes et qu'elle est orthographiée de la même façon. « Talend gère ces référentiels au travers des systèmes », se félicite le Commandant Tedesco.

Aujourd'hui, tous les échanges inter-applicatifs de la Gendarmerie et de la Police sont centralisés par Talend et le Système d'Interfaces et de Référentiels (SIR) du ST(SI)2 qui ne gèrent pas moins de 3 To de données par jour et 600 jobs. Sans oublier le partage d'informations avec les différents Ministères et l'alimentation des statistiques, sous forme de données anonymisées, ou semi-anonymisées.



Pourquoi Talend ?

L'histoire entre la Gendarmerie Nationale et Talend n'en était cependant pas à ses débuts. Le premier déploiement de Talend au ST(SI)2 date du milieu des années 2000. A cette époque, la Gendarmerie décide de passer ses postes de travail sous Linux. Logiquement, son centre de production suit. Problème : son ETL est purement Windowsien.

Après une analyse des alternatives, le ST(SI)2 trouve une oreille particulièrement attentive chez un des éditeurs du marché. « Nous avons trouvé chez Talend un esprit dynamique, avec une vraie volonté de faire avancer le produit », se souvient le Commandant Tedesco.

Talend a depuis sorti des connecteurs pour SAP HANA ou le Big Data dont avait besoin la Gendarmerie. « Talend s'adapte continuellement aux évolutions technologiques. On n'attend pas dix ans que la solution soit mise à jour », se réjouit le Commandant. « Et lors d'une migration vers SAP HANA, il a suffi au ST(SI)2 de changer le connecteur. Il était présent, fonctionnel, simple à utiliser. »

Le fait que Talend soit une entreprise créée en France a été un atout supplémentaire pour cette institution nationale. Tout comme sa dimension open source d'ailleurs. Ses équipes apprécient de pouvoir trouver dans des délais très brefs, grâce à la communauté, des réponses à la plupart de leurs problèmes.

Au cœur des applications de la sécurité intérieure

« Cette automatisation de la consultation de tous les fichiers de sécurité est particulièrement sensible », tient à souligner le Commandant

Tedesco. « On peut figurer au TAJ, pour n'importe quelle affaire, conflit de voisinage ou bien plus grave. De même, apparaissent au sein du fichier des personnes recherchées, les personnes n'ayant pas payé leur pension alimentaire ou leurs impôts, ou encore celles qui ne se sont pas présentées pour un retrait de permis. Cela ne fait pas forcément de ces personnes des criminels ou des terroristes ».

C'est pourquoi, le gouvernement, sur les recommandations de la CNIL, a dressé une liste restreinte d'utilisateurs autorisés à consulter des fichiers de renseignement. Et toutes les nuits, Talend construit, à partir de sources diverses, un « référentiel pivot RH » qui prend en compte certains éléments, comme les mutations ou changements d'affectations, permettant la restriction de consultation aux seules personnes autorisées. De lui dépendent en effet tous les accès aux fichiers, en lien avec le LDAP (Lightweight Directory Access Protocol).

Enfin, la décision finale sur la qualification des individus, elle, reste humaine. « Talend aide au traitement de consultations multiples. Mais le retour global est soumis à validation d'Officiers de Police Judiciaire (OPJ) », rappelle le Commandant.

Le Commandant Tedesco se souvient : « Nous avons débuté en analysant 300 000 identités criblées par an. Nous sommes aujourd'hui montés en charge, sans coup férir, à un million d'identités par mois. » Cette croissance tient au fait qu'ACCReD s'est ouvert aux entreprises sensibles (RATP, transporteurs, nucléaire, etc.) qui doivent s'assurer que leurs employés ne sont pas à risque. Or, beaucoup d'individus réapparaissent très fréquemment dans les fichiers à cause de la sensibilité de leur métier.

Dans le nucléaire, ACCReD permet de ne plus se contenter d'une étude de moralité à l'embauche mais de vérifier qu'il n'y a pas de dérives comportementales en interrogeant à la volée, des fichiers sources mis à jour dont le TAJ.

On retrouve le même mécanisme dans les applications de mobilité des agents. Auparavant, lors d'un contrôle routier, un gendarme devait faire autant de demandes qu'il y avait de fichiers indépendants à interroger : cartes grises, assurances, fichiers FOVeS des véhicules volés, titres d'identités, personnes recherchées, etc. Le temps de la saisie multiple ralentissait énormément le temps de traitement. Aujourd'hui, l'agent rentre une identité et/ou une immatriculation, il choisit les vérifications à faire. Talend et le SIR s'occupent de tout. La synthèse arrive généralement en 8 à 10 secondes.

C'est exactement le même principe qui permet aujourd'hui des contrôles automatisés aux frontières dans les aéroports français entre une et vingt secondes. L'application de la Police de l'Air et des Frontières consulte, là encore via le SIR et Talend, le fichier des personnes recherchées en temps réel.

« En l'espace de cinq ans, Talend est devenu un élément clé de notre Système d'Information ; il ne se passe plus une réunion d'élaboration d'un besoin sans que le SIR et Talend ne soient intégrés, tant la mécanique est fiable », se félicite le Commandant Tedesco.

TERADATA FRANCE

SÉVERINE VIGNERON

Marketing Manager

+33 (0)1 81 89 14 08

teradata.france@teradata.com

Antony Parc I, 2-6 place du General de Gaulle, 92160 ANTONY

www.teradata.com



teradata.



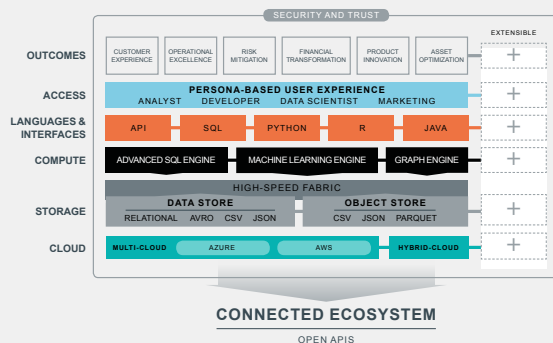
Teradata :

LEADER MONDIAL DE « PERVERSIVE DATA INTELLIGENCE »

Nous changeons la façon dont le monde utilise les données pour obtenir de meilleurs résultats. Teradata fournit des réponses intelligentes en temps réel, exploitant 100 % des données pertinentes, indépendamment de l'échelle ou du volume des requêtes. Et nous le faisons sur site, dans le Cloud et à n'importe quel endroit entre les deux. C'est ce qu'on appelle « Pervasive Data Intelligence ». Aujourd'hui, Teradata est la seule entreprise à disposer des produits, de l'expertise et des services de pointe pour proposer cette solution.

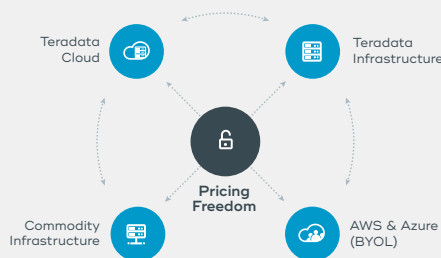
VANTAGE : Plateforme leader de l'analyse de données

Teradata Vantage est la plateforme leader alliant gestion de la donnée et de l'analytique dans un environnement Cloud hybride qui tire profit de 100 % de vos données pour tout type d'analyse, tout le temps, n'importe où.



Vantage permet une utilisation optimale du Cloud

L'analyse dans le cloud est un outil indispensable pour adresser les nouveaux challenges qui nécessite de l'analytique : augmentation des sources de données, réduction de l'empreinte des datacenters, pression pour prendre des décisions rapidement, pénuries des ressources et des talents.



Pour répondre à ces besoins grandissants, les entreprises se tournent vers l'analytique as-a-service dans le cloud pour moderniser les infrastructures, accélérer le time-to-value, réduire les risques financiers, se délasser des tâches fastidieuses afin de permettre de se concentrer sur les tâches à forte valeur ajoutée.



Retrouvez tous nos retours d'expérience clients sur notre site : www.teradata.com

teradata.

RETOUR D'EXPERIENCE

LARRY H. MILLER
SPORTS & ENTERTAINMENT
A CHOISI TERADATA :



Que vous soyez un fan passionné ou un observateur passif, les équipes sportives ont besoin d'un avantage concurrentiel et d'un différenciateur pour vaincre la concurrence. L'utilisation de la donnée est omniprésente dans les nouvelles générations d'analyse sportives.

Les enjeux métiers sont :

- Détenir des biens telles que la franchise NBA et des espaces de divertissement.
- Nécessité d'un **nouvel avantage concurrentiel** pour battre la concurrence.
- Le Management veut des **indicateurs plus précis** sur les caractéristiques des athlètes performants.
- Nécessité de nouvelle technologie de données et d'**analyse** qui intègre de **nouveaux types de données**, comme la santé des joueurs, les habitudes alimentaires et les habitudes de sommeil.
- **Augmenter les ventes** en augmentant la satisfaction client.

La solution :

- **Migration** de Microsoft SQL vers Teradata Intellicloud **en moins de 6 mois**, incluant la roadmap initiale de construction d'analytiques avancés et des tableaux de bord opérationnels pour chaque département métier.
- Teradata Data Lab, Teradata QueryGrid™.
- Utilisation Teradata Consulting et d'applications tiers comme Tableau, Attunity and WhereScape.
- Sélection de **Teradata Vantage** avec un **modèle à la consommation**.

Les bénéfices :

- Permettre au Management d'améliorer ses statistiques pour des **analyses accrues sur les performances** des joueurs à partir de nouvelles sources de données tel que la santé des athlètes.
- **Perfectionner** la capacité de gérer la **satisfaction client** à travers les données de contact incluant les données mobiles et les informations de stationnement dans les stades.
- **Améliorer l'expérience client** grâce à des balises permettant d'orienter les offres de concessions mobiles vers les concessions les plus proches des stands, optimisation du stationnement, renforcement de la sécurité.

STOP BUYING “ANALYTICS”



IT'S TIME TO INVEST IN ANSWERS.

Analytics. It's a \$200 billion industry, fueled by the speed, scale, and competition of the rising digital economy. But, with all of the investment being made by businesses looking to gain an edge in this world, there's one question no one seems to be asking: When do we stop buying into partial solutions that overpromise and underdeliver? The answer to that question is “now.”

At Teradata, we transform how businesses work and people live through the power of data.

We exist to deliver real-time, intelligent answers, leveraging 100% of your relevant

data, regardless of scale or volume of query. And we do it on-prem, in the cloud, and anywhere in between. We call this Pervasive Data Intelligence. And it comes to life today.

Introducing Vantage, the platform for Pervasive Data Intelligence.

Only Vantage can leverage 100% of your relevant data, all of the time. So you can analyze anything, deploy anywhere, and deliver analytics on what matters most to your business. It's the answer to the frustration of today's approach to analytics.

Get the answer at [Teradata.com](https://www.teradata.com).

teradata.



TIBCO Software, le partenaire de l'entreprise data-driven

TIBCO aide les organisations à transformer leurs données en actions, à améliorer l'expérience client, à accroître l'excellence opérationnelle et à innover avec la création de nouveaux services.

Les technologies de TIBCO interconnectent collaborateurs, systèmes, équipements et API de l'écosystème de l'entreprise. Elles capturent les données en temps réel et augmentent l'intelligence numérique de chaque organisation en optimisant et accélérant prises de décisions et actions.

Connected Intelligence

CONNECT

any application, device, or data source.



API-led Integration



Event-driven Applications



Data Fabric



Information Management



Analytics



Data Science & Streaming

UNIFY

all data for better access, trust, and control.

AUGMENT

Capture and process information to make faster decision

Transformez les données en actions

CONNECT

Interconnectez collaborateurs, systèmes, processus et objets dans le cloud et on-premise avec une solution d'intégration hybride basée sur les API, les conteneurs et les microservices.

UNIFY

Donnez et pilotez l'accès à toutes vos sources de données et assurez la fiabilité de ces données avec le data management. Cette offre intègre le MDM, la gestion des métadonnées et des données de référence ainsi que la data virtualisation.

AUGMENT

Améliorez vos performances opérationnelles et commerciales grâce à la compréhension et la maîtrise des données. Pour ce faire, TIBCO associe data visualisation, data science, machine learning, analyse temps réel et analyse prédictive dans son offre TIBCO Spotfire®.

Avec TIBCO, vous capturez, unifiez, visualisez, comprenez et agissez sur les données :

- en partageant des connaissances basées sur l'IA avec les collaborateurs qui en ont besoin,
- en analysant les données en temps réel,
- en automatisant les décisions pour relier intelligence et action,
- en brisant les barrières pour une analyse de qualité s'appuyant sur les métadonnées, le MDM et la virtualisation des données.

Mercedes-AMG Petronas Motorsport

Data and insights that fuel championships

"To win you need to understand what's important, and what's not. We need to make sure we're adapting to the situation."

—James Vowles, Chief Strategist,
Mercedes-AMG Petronas Motorsport



Découvrez TIBCO Spotfire X
Langage naturel, machine learning
et temps réel pour une nouvelle
expérience utilisateur

Official Team Partner

TIBCO®



AMG
PETRONAS
MOTORSPORT

www.tibco.com

TIGERGRAPH

DEBORAH IVES

EMEA Head of Marketing

+44 7450895688

deborah.ives@tigergraph.com

3 Twin Dolphin Drive, Ste 225 CA 94605 USA

www.tigergraph.com



Graph is Now Accessible to Everyone

TigerGraph Cloud solves the accessibility to graph database and analytics issues with an easy-to-use cloud-based graph database-as-a-service. TigerGraph Cloud is built for people who'd rather be building innovative applications to deliver new insights than managing databases. TigerGraph Cloud enables everyone to benefit from graph-based analysis. With TigerGraph Cloud anyone can get started in minutes, build a proof-of-concept model in hours and deploy a solution to production in days.



Price/Performance

TigerGraph's parallel processing reduces the memory needed to scale to billions of entities (vertices) and relationships (edges)



SaaS

No hardware to buy and manage, easy integration and scaling, pay as you go



Fully Managed by the Experts

Managed by the same engineers who developed TigerGraph, with automatic backup, automatic updates, security hardening, 24/7 maintenance and support.



Flexible Pricing and Sizing Options

A range of instances to fit a variety of needs - choose from one hour to yearly

Graph is Easier than Ever

TigerGraph Cloud Starter Kits contain pre-built schemas, with sample dataset & queries and graph algorithms. These not only cover the most popular use cases such as fraud detection, explainable AI, machine learning, recommendation, knowledge graph, hub or influencer detection, referral networks, community detection, social network and supply chain analysis, but also cover specific industries such as healthcare, pharmaceutical, financial services, internet, eCommerce and telecom.

TigerGraph GraphStudio is an intuitive solution development kit (SDK) with an integrated data science workbench purpose-built for connected data. TigerGraph GraphStudio makes it simple to traverse connected business entities, find connections among payments, claims, orders, customers and citizens, discover meaningful behavior patterns such as fraud rings and propensity to buy a specific product, and search the dataset for matching entities and patterns, all within an easy-to-use, intuitive graphical user interface.

TRY IT NOW FOR FREE AT TIGERGRAPH.COM/CLOUD



TigerGraph

The Only Scalable Graph Database for the Enterprise

Based on the industry's first Native Parallel Graph technology, TigerGraph unleashes the power of interconnected data offering deeper insights and better outcomes in real time, no matter how large or complex the dataset.

**VISIT TIGERGRAPH
AT BIG DATA PARIS BOOTH A10**





Nous créons la plateforme analytique nommée 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 difficultés généralement associées aux systèmes analytiques et les remplace par une interface réactive et conviviale 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*".

TIMI est composé de 4 outils:

- **TIMI Modeler**: Création automatisée de modèles prédictifs de haute précision (Auto-ML). Tout en self-service.
- **Anatella**: L'outil de data management (ETL) le plus rapide du marché. Tout en self-service. Inclut des fonctionnalités spécifiques très utiles pour l'advanced analytics (tels que le graph mining, text mining, etc.) et la Business Intelligence (par ex.: connecteurs natifs à Tableau, Qlick, Kibella).
- **Stardust**: Outil de segmentation: segmentation multivariée en 3D de toute votre base de clients.
- **Kibella**: création illimitée de Dashboards en self-service.

TIMI résout les 4 principaux obstacles rencontrés lors de projets de type (Advanced) Analytique/AI/Machine Learning/Big data:

1. **Coût d'Infrastructure Prohibitifs**: Construire (et utiliser) de bons modèles prédictifs et de bons KPI's implique d'exécuter des transformations de données complexes sur des volumétries relativement importantes (plusieurs dizaines de TB). Presque tous nos clients éprouvaient de larges difficultés sur ce point. C'est pourquoi nous avons créé Anatella: un ETL « Big Data » aux performances incroyables.
2. **Complexité très élevée**: La maintenance de ces transformations de données complexes et la création de bons modèles prédictifs (ML & AI) ne sont plus réservées « à une petite élite intellectuelle ». Avec TIMi, tout est en self-service, à la souris. La complexité d'utilisation est similaire à MS-Excel.

3. **Temps d'implémentation déraisonnable**: Grâce à sa rapidité et sa convivialité, TIMi vous donne la possibilité d'investiguer facilement vos données pour y trouver la perle rare.

4. **Capacité de stockage et de traitement des données limitées**: Les algorithmes de compression originaux inclus dans TIMi apportent une solution adéquate à tout problème de stockage.

Dès 2007, notre solution de Auto-ML, TIMi Modeler, révolutionnait la façon dont les entreprises analysent leurs données. Mais, en 2007, nous n'avions pas encore Anatella (notre ETL) et cela restait une vraie «épine dans le pied». En effet, un "data scientist" consacre plus de 80% de son temps à utiliser un ETL (par ex.: pour la recherche, l'unification et le nettoyage des données) (Source: IDG). Ce fait est aussi parfois appelé le "*80/20 data science dilemma*" (google it!). Par conséquent, une solution rapide et conviviale pour la transformation de données à grande vitesse est un élément important pour le confort, l'efficacité et la productivité de tout "data scientist".

La supériorité technologique de TIMi en matière d'outil de Auto-ML n'est plus à démontrer.

Pour démontrer la supériorité technologique de TIMi en matière de ETL, nous le comparerons à Spark. Spark est l'outil le plus rapide dans l'écosystème Hadoop. Pour la comparaison, nous allons mesurer le temps requis pour exécuter les 22 transformations de données qui répliquent exactement les 22 requêtes SQL incluses dans le célèbre benchmark open source et universitaire: le TPC-H. Cette charge de travail est intéressante car elle est universellement reconnue comme étant représentative des transformations de données couramment utilisées dans la vie quotidienne de tous les "data scientists".

Les résultats de ce benchmark (càd les temps d'exécution des 22 requêtes du TPC-H) sont dans la table ci-dessus, à droite. Comme Spark est une solution de calcul distribuée, nous reportons deux temps d'exécution: (1) **Colonne A**: le temps d'exécution sur une seule machine et (2) **Colonne B**: le temps d'exécution de Spark sur un cluster composé d'une infinité de machines (temps dit « incompressible » suivant la terminologie de la « Loi de Amdahl »).

La colonne C contient les mêmes informations que la colonne B (càd le « temps incompressible », mais elle est exprimée en pourcentage. Chaque fois que cette colonne C dépasse 50%, les cellules sont sur fond **rouge**. Veuillez noter que, presque tout le temps, la colonne C est rouge: elle dépasse 50% presque tout le temps. Une réduction de 50% du temps d'exécution (càd un temps incompressible de 50%) correspond à un gain en vitesse de 2 dans le meilleur des cas (càd sur un cluster de taille infinie). Ce gain en vitesse (qui est limité à 2

TPC-H Query (100GB data-base)	Spark (2019/2)			Anatella v1.84	
	running on 1CPU (seconds)	Incompressible runtime (when running on infinite number of CPU's)		Run time (seconds)	Speed-up vs Spark infinite CPU
		A	B		
Q1	184	37	20 %	27.1	1.4
Q2	956	649	68 %	5.7	113.5
Q3	929	572	62 %	34.5	16.6
Q4	830	230	28 %	33.7	6.8
Q5	2275	674	30 %	43.7	15.4
Q6	102	21	20 %	6.2	3.3
Q7	1113	762	68 %	45.6	16.7
Q8	1621	1009	62 %	49.3	20.5
Q9	2059	1227	60 %	200	6.1
Q10	1035	699	68 %	38.9	17.9
Q11	441	304	69 %	4.2	71.8
Q12	454	263	58 %	47.7	5.5
Q13	377	143	38 %	105	1.4
Q14	373	275	74 %	3.2	85.5
Q15	error	error	error	9.7	
Q16	839	587	70 %	31.4	18.7
Q17	1255	665	53 %	26.7	24.9
Q18	1135	717	63 %	36.9	19.4
Q19	972	119	12 %	44.1	2.7
Q20	972	643	66 %	21.7	29.6
Q21	3815	2235	59 %	127	17.5
Q22	206	112	55 %	44.5	2.5

pour Spark) est inattendu. Cela rend le système Spark dans son ensemble presque inutilisable, car la promesse majeure de Spark (évolutivité horizontale: offrir une vitesse supérieure sur une infrastructure plus grande) n'est pas réalisée.

A l'opposé, TIMi utilise une stratégie de calcul distribué infiniment évolutive mais, 99% du temps, il n'est même pas nécessaire d'utiliser du calcul distribué tant TIMi/Anatella est déjà rapide sur un CPU. Par exemple, La requête Q2 du TPC-H est plus de 100 fois plus rapide sur une seule machine équipée de TIMi que sur un cluster Spark de taille infinie (113.5 fois plus rapide: voir la cellule **bleue** dans la table ci-dessus).

C'est ainsi que, grâce aux performances exceptionnelles de TIMi, la société Snype re-calcule chaque jour les revenus nets de ***tous*** les télécoms du Vietnam à partir des **données brutes des CDR** (100 millions abonnés, 120e6 SIMs, 8e12 de records par jour, 200 To/jour) sur une seule machine (112 cœurs, 512 GB, 2 PB de stockage). Seul TIMi est capable d'une telle prouesse.

Je vous invite à vous rendre sur <https://timi.eu> pour y télécharger la version communautaire et gratuite de TIMi! Bon Amusement à redécouvrir vos données!

LA SOLUTION

LA PLUS

RAPIDE DU MARCHÉ

1 SERVEUR TIMI
PLUS RAPIDE QUE
276 SERVEURS
SPARK

MANIPULEZ AISÉMENT
DES MONTAGNES
DE DONNÉES

A l'aide d'un simple PC de bureau, TIMi analyse et manipule des bases de données contenant des milliards d'enregistrements et des milliers de variables. En utilisant seulement 100GB d'espace disque, les algorithmes de compression originaux de TIMi permettent de stocker et manipuler un volume de données équivalent à 5 TB en BDD.

TIMI effectue des transformations complexes sur plusieurs terabytes de données en quelques minutes.


INTEGRATED DATAMINING

INTERFACE CONVIVIALE
EN SELF-SERVICE

Apprentissage court et aisé. Créez des transformations de données complexes et construisez des modèles de qualité en quelques click de souris. Aucun doctorat en "Data Science" n'est nécessaire.

ILS NOUS FONT CONFIANCE

Téléchargez gratuitement la Community Edition sur <https://timi.eu>





Trifacta accélère la préparation des données pour les déclarations réglementaires des plus grandes institutions financières

Anticipant la multiplication par deux des coûts liés à la conformité d'ici 2022, un nombre croissant d'institutions financières remplace les processus de reporting à base tableurs et optent pour Trifacta, pour davantage de rapidité et d'exactitude

Trifacta, leader mondial de la préparation de données, a annoncé ce jour une augmentation de l'adoption de ses produits Wrangler par les entreprises du secteur des services financiers, qui y voient le moyen de gérer des initiatives de conformité et de risques de plus en plus exigeantes comme CCAR, BCBS 239 et AML. L'envolée de Trifacta dans le domaine des déclarations et du reporting réglementaires fait écho à la présence croissante de l'entreprise dans le secteur des services financiers, pour divers usages opérationnels et analytiques.

Les exigences réglementaires évoluent constamment, et deviennent de plus en plus granulaires et nuancées. En effet, les institutions financières sont tenues de suivre et de traiter en moyenne 200 nouveaux amendements réglementaires internationaux par jour selon le service de veille réglementaire de Thomson Reuters. La conformité aux différentes obligations de chaque juridiction pèse plus lourd en ressources sur les entreprises de services financiers.

De plus, les institutions financières gèrent une quantité et une diversité énormes de données de transactions, ce qui peut engendrer des problèmes de qualité majeurs à régler avant même de pouvoir commencer à s'atteler à déchiffrer la masse toujours croissante d'exigences réglementaires. La préparation des données à elle seule peut représenter jusqu'à 80 % d'un projet d'analytique — ce qui signifie que le manque d'efficacité des processus et les problèmes de qualité et d'exactitude des données peuvent faire effet boule de neige sur les coûts opérationnels.

Trifacta permet aux institutions financières de se défaire de leurs outils habituels de type tableurs dans le cadre de leurs déclarations réglementaires afin d'améliorer la rapidité, la collaboration et la précision des équipes chargées de la conformité. Plus important encore, la capacité de Trifacta à assurer la traçabilité des données offre une meilleure visibilité tant en interne qu'en externe sur la façon dont les données ont été transformées, une fonctionnalité critique pour la conformité. L'équipe des risques de marché du Crédit Agricole a quant à elle pu simplifier radicalement son processus d'émission de rapports à l'aide de Trifacta.

« Trifacta a permis aux experts métiers de transformer les données pour répondre à différentes normes de conformité, » témoigne Pierre Castellani, Responsable de l'informatique des risques de marché au Crédit Agricole. « Grâce à Trifacta, nos experts métiers bénéficient d'une compréhension immédiate des données, et peuvent les préparer eux-mêmes en utilisant les suggestions intelligentes, ce qui a accéléré la préparation des rapports exigés par la réglementation. Avant Trifacta, une équipe dédiée codait manuellement les rapports de données. Maintenant que ce procédé est fortement réduit, elle peut consacrer plus de temps à l'analyse. »

« Avant Trifacta, une équipe dédiée codait manuellement les rapports de données. Maintenant que ce procédé est fortement réduit, elle peut consacrer plus de temps à l'analyse. »

Pierre Castellani, Responsable de l'informatique des risques de marché au Crédit Agricole

En plus du soutien à la conformité, Trifacta observe une utilisation accrue de Wrangler au sein des institutions financières pour les initiatives de vues 360 degrés clients, de gestion des risques, d'apprentissage automatique et d'intelligence artificielle (IA). La course à l'innovation redéfinit le domaine des services financiers, à mesure que de plus en plus d'institutions adoptent des procédés axés sur les données pour la gestion de la relation client, ainsi que l'apprentissage automatique et l'IA pour optimiser leurs opérations et l'expérience client. Des données épurées sont le point de départ de toutes ces initiatives, faisant de la plateforme de préparation de données de Trifacta un outil essentiel à leur réussite pour un nombre croissant d'institutions.

Parmi les autres institutions financières qui utilisent Trifacta figurent Deutsche Börse, Commerzbank, Nordea Bank, la bourse du Luxembourg et la Royal Bank of Scotland.

Pour en savoir plus sur Trifacta en tant que solution au service de la conformité financière et des rapports réglementaires, visionnez notre webinaire avec Nordea.

Accelerate data cleaning & preparation with Trifacta for cloud data lakes & warehouses

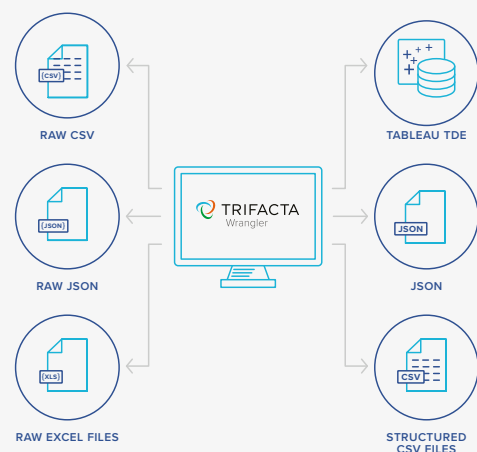
Try Trifacta
Wrangler
14-Day Trial
FREE!



For success of your analytics, ML & data onboarding initiatives

Benefits of Trifacta Wrangler

- Profile, clean and deliver quality data faster
- Wrangle data from virtually any source
- Full platform functionality (no limitations)
- Cloud-based and no coding experience needed



Visit Stand n°A21 or
trifacta.com/start-wrangling





GRUPE LASTMINUTE.COM : BIG DATA ANALYTICS, AU CŒUR DE LA MISSION D'ENTREPRISE

La société

Fondé en 2006, le groupe lastminute.com est un leader européen de la vente en ligne de voyages et de loisirs. Les sites Web et les applications mobiles de la société sont disponibles dans 40 pays.

Ces dernières années, le groupe a défini une mission corporate très précise : devenir l'entreprise de voyage la plus pertinente et inspirante, engagée à enrichir la vie des voyageurs. Pour atteindre cet objectif l'entreprise préconise l'utilisation d'applications Big Data innovantes.

Le choix d'une solution logicielle

Pour mettre en œuvre des projets Big Data, l'entreprise avait besoin de transformer son architecture technique.

Le projet, appelé Data M.A.G. (Marketing Analytics in Google), a abouti à la mise en place d'un datalake interne, implémenté sur Google Cloud comme infrastructure de calcul. La création d'un référentiel unique pour toutes les données brutes a été choisi afin de corréler les différents types de données provenant des divers processus métiers de l'entreprise.

Le cloud IaaS (Infrastructure-as-a Service) offrait des avantages de coût évidents pour lastminute.com, réduisant l'investissement dans le stockage et permettant de minimiser le transfert de données des systèmes en ligne internes et externes.

Vertica, base de données en colonnes à très haute performance analytique, a été choisie comme moteur analytique pour le Big Data. Vertica réduit les temps de réponse des requêtes, améliorant ainsi la prise de décision et la capacité à fournir des analyses en temps réel et en continu. Vertica se distingue également par sa capacité à s'intégrer facilement à l'écosystème décisionnel existant de lastminute.com ainsi que par son évolutivité horizontale efficace, permettant une augmentation des volumes de données et des charges de travail analytiques au sein du groupe.

Après avoir testé d'autres solutions en mode SaaS, lastminute.com a décidé de déployer Vertica comme plate-forme analytique globale du groupe pour ses fonctionnalités techniques mais aussi parce qu'elle offrait le modèle de tarification le plus efficace pour le plan d'expansion.

Le groupe lastminute.com a progressivement consolidé toutes les données structurées existantes sur Vertica et complète

maintenant l'environnement analytique pour stocker des données non structurées liées à l'activité des clients en ligne.

Les données anonymisées recueillies servent à étudier le comportement des clients sur tous les canaux de lastminute.com et à analyser toutes les étapes du parcours client, depuis les recherches préliminaires en ligne jusqu'à l'acte d'achat.

Cette solution d'analyse intégrée permet à l'entreprise de corréler les données issues des processus décisionnels traditionnels avec les données issues des campagnes marketing, du CRM analytique et des analyses de machine learning. Les Data Scientists peuvent ainsi améliorer l'analyse de la relation client avec des modèles prédictifs avancés.

Les développements futurs

L'un des objectifs de développement est d'optimiser le budget marketing en ligne. À l'aide d'algorithmes d'attribution, l'entreprise peut identifier les événements particuliers pour lesquels l'investissement dans une campagne marketing génère un meilleur rendement et, par conséquent, allouer plus efficacement les ressources disponibles.

En résumé

L'implémentation de Vertica a déjà prouvé son impact positif sur le ROI des campagnes marketing, et donc sur le chiffre d'affaires de l'entreprise. De plus, les algorithmes d'attribution permettront à l'entreprise d'optimiser les coûts marketing globaux, ce qui entraînera une augmentation des profits.

En conclusion, le groupe lastminute.com a mis en place le socle technologique pour la poursuite des projets Big Data et pour rester à la pointe de l'innovation. Conformément à cet objectif, elle a modifié sa propre structure organisationnelle en créant une équipe Data Science qui soutiendra de plus en plus l'ensemble du groupe.

Tester Vertica: vertica.com/try

Nous contacter: vertica.com

VERTICA

ADVANCED ANALYTICS PLATFORM

Performance
Scalability
Reliability

in an open world



vertica.com

YSANCE

JEAN-MARIE UZEL

VP Data Services

+33 (0)6 37 77 33 05

jean-marie.uzel@ysance.com

24 rue du sentier 75002 Paris FRANCE

www.ysance.com

Analytique

API

Applications

Consultants/SSII
Intégrateurs

Dataviz

Infrastructure

Institutionnels

ysance



ysance



Seuls au monde face à vos problèmes Data ?

Rencontrez nos experts ils seront plus bavards que Wilson.

Découvrez nos offres



RCU



Omnichannel
Analytics



e-RFM

Ysance, 24 rue du Sentier 75002 Paris - +33 (0)6 37 77 33 05 - www.ysance.com



Nous sommes en 2020, quelles grandes tendances observez-vous en Data dans le secteur du Retail ?

Cela fait près de 10 ans que nous collaborons avec des Retailers qui ont mis l'omnicanal et le client au centre de leur stratégie. Nous les appelons les innovateurs et ils ont largement intégré des KPIs clients au cœur de leur Comex.

Nous assistons depuis quelques mois à une nouvelle vague d'entreprises qui s'équipent, certains sous la pression du marché et des chiffres, c'est à dire se transforment en profondeur autour du client. Tous les métiers (des achats à la logistique) parlent maintenant du Client. Au delà des plateformes SaaS Marketing qu'Ysance a conçu dès 2015 pour les accompagner, nous avons développé des offres pré-packagées pour qu'ils puissent redessiner leur SI interne autour du client, ce sont nos offres de RCU, d'Analytics Clients et de nouvelle segmentation RFM.

Depuis le temps que le digital existe, vous sous-entendez que les entreprises ne connaissent pas bien leurs clients ?

Il est bon de se rappeler que la majorité des entreprises du secteur de la distribution se sont d'abord construites autour de l'offre produit et des emplacement physiques puis des différents canaux de communication qui sont apparus avec les innovations technologiques. La notion même de client a finalement été sous-investie. Ce n'est plus acceptable en 2020.

Pouvez-vous préciser comment s'intègre un RCU dans un système d'information marketing et client et quels sont ses principaux bénéfices ?

La plateforme RCU va en premier lieu recevoir les données clients de l'ensemble des systèmes qui en possèdent (crm, marketing automation, support clients, autres...), puis il va falloir nettoyer, consolider, dédoublonner puis fusionner la donnée avant de l'enrichir au sein de ce qu'on appelle un Golden Record, à savoir la fiche client unique. On en profite pour y adjoindre un ensemble d'indicateurs pré-calculés (appétences, segmentation/scoring, CLTV, ...). On va ensuite ré-alimenter tous les systèmes qui consomment ces données clients avec une information unique, propre et ainsi agrégée.

En terme de bénéfices constatés, ils sont assez rapides: Si vous envoyez par exemple des catalogues papiers, grâce au dédoublonnage, vous allez économiser sur le volume d'envoi. Bénéficiez de ciblage marketing centralisés et plus précis. Ré-alimentez au final en temps réel via votre RCU, les autres systèmes avec les informations de vos consommateurs.

Dans les comités de direction, on parle encore de la promotion de la semaine prochaine plutôt que des clients à qui on va s'adresser. Vous confirmez ?

Tout à fait. Et c'est une grande contradiction qui fait partie de notre quotidien. On lit partout depuis plusieurs années que les organisations veulent devenir "customer-centric", mais posez-vous sincèrement la question des tableaux de bord et autres KPIs clients qui sont mis à disposition des comités de direction et des équipes internes, à quel point la donnée client influence, permet d'arbitrer des décisions. Demandez à l'ensemble de vos collaborateurs s'ils savent combien de clients différents ont acheté le mois dernier, combien de nouveaux clients vous avez gagné, perdu ?

Le poids de l'historique reste fort, on cherche avant tout à vendre des produits, plus qu'à s'adresser à des clients, car il faut assurer le quotidien, tout en se transformant.

Le chemin reste long, il est celui du changement.

Qu'appellez-vous eRFM ?

La RFM opérait principalement les transactions effectives. La question que nous adressons aujourd'hui est celle de la E-influence (mesure de l'impact du digital sur les ventes). Il est aujourd'hui indispensable d'intégrer le comportement digital dans toute segmentation client. Pour ce faire, il faut réconcilier les données clients (comportement digital - qu'il soit email ou navigation - et transactions) sur tous les canaux, et s'appuyer sur ces nouvelles segmentations pour décider des meilleures communication avec lui.

Pour illustrer l'intérêt par un exemple très simple: Un client qui n'a pas acheté depuis 12 mois mais qui pourtant ouvre ses emails et continue à visiter le site, n'est pas réellement un client perdu.



LEAD WITH DATA™

Pilotez par la donnée.

La plateforme de gestion et d'analyse des données de bout en bout de Qlik est conçue pour transformer toute votre entreprise à l'ère du digital.

qlik.com





PEP-IT

Rejoignez PEP-IT :
la mine d'or pour les
pros de la DATA !



PEP-IT

la plateforme qui réunit
prestataires qualifiés & entreprises
pour leur projet DATA.

www.pep-it.io





Google Cloud
SUMMIT

Google Cloud Summit est de retour à Paris.

Jeudi 5 novembre 2020
Paris Expo - Porte de Versailles

Rejoignez-nous à l'occasion de cet événement gratuit qui rassemble la communauté cloud pour une journée exceptionnelle.