



# BIG DATA

Un écosystème complexe en développement  
à l'aube de profondes mutations

## BIG DATA

### Un écosystème complexe en développement à l'aube de profondes mutations

Érigé au début de la décennie 2010 au rang des principaux piliers de la révolution numérique des entreprises françaises, le *big data* est aujourd'hui devenu une réalité dans leur pratique, notamment au sein de certaines fonctions.

Concept complexe et protéiforme longtemps resté au stade expérimental dans les très grandes sociétés, il a véritablement amorcé une phase de déploiement plus massif dans l'ensemble des entreprises françaises au milieu de la décennie 2010. Les avancées technologiques, l'implication forte des pouvoirs publics, et la croissance exponentielle des flux de données héritée entre autres de l'essor de l'Internet des objets, lui ont donné une réelle impulsion. La croissance des investissements consentis en services et en moyens humains et matériels en témoigne.

Partie intégrante de l'écosystème informationnel, le *big data* se place au cœur des grands enjeux économiques et stratégiques des entreprises. Il a progressivement conquis de nouveaux territoires, élargissant ses champs d'application ou se développant dans de nouveaux secteurs d'activité.

Complexe et onéreux, son déploiement sur le terrain soulève cependant des problématiques techniques et sécuritaires. Conjuguées aux nouveaux enjeux technologiques et législatifs, ces limites poussent à une nouvelle approche du *big data*, voué à amorcer aujourd'hui une phase de transformation en vue d'un développement généralisé.



# DANS CE DOSSIER

POINTS-CLÉS ET ENJEUX .....	4
LE BIG DATA : UN CONCEPT COMPLEXE, UNE APPROCHE STRATÉGIQUE.....	8
Un objet complexe polymorphe .....	8
Des enjeux stratégiques multiples pour les entreprises .....	9
Les outils et ressources humaines : deux prérequis au déploiement .....	10
Un secteur cloisonné .....	11
UN MARCHÉ ENTRÉ DANS UNE PHASE D'INDUSTRIALISATION .....	12
Le tournant du plan français <i>big data</i> .....	12
Des flux de données exponentiels hérités des avancées technologiques.....	13
Des évolutions technologiques en amont déterminantes .....	14
Les porteurs de projets investissent dans les outils et services <i>analytics</i> .....	15
Des entreprises en quête de recrutement de profils experts.....	17
Les infrastructures délocalisées de stockage se déploient sur le territoire .....	20
DES USAGES TRANSVERSAUX DANS DES DOMAINES VARIÉS .....	22
Le <i>big data</i> , enjeu clé pour l'analyse marketing.....	22
Systèmes d'information et bases de données : l'efficacité par le <i>big data</i> .....	25
Le renforcement algorithmique : le <i>big data</i> au service de l'intelligence artificielle.....	28
La gestion de flottes et de réseaux, guidée par la donnée.....	29
Générer des rendements supérieurs grâce aux données .....	32
Mieux détecter la fraude avec le <i>big data</i> .....	33
UN FUTUR PROMETTEUR GUIDÉ PAR DES SECTEURS PORTEURS .....	34
Un potentiel élevé dans l'assurance.....	34
De nombreuses perspectives dans la santé .....	35
Un gisement encore peu exploité dans les ressources humaines.....	37
La transformation du métier d'archiviste par le <i>big data</i> .....	38
Le véhicule autonome, candidat à un usage extensif de la donnée .....	39
Des opportunités variées pour les organisations publiques .....	40
LE BIG DATA À L'AUBE DE PROFONDES TRANSFORMATIONS .....	41
Vers une nouvelle architecture du <i>big data</i> : le <i>edge computing</i> .....	41
Une migration inexorable vers la <i>smart data</i> .....	43
Le <i>big data</i> et l'IA : une convergence inéluctable, gage d'efficacité .....	44
UN ESSOR DISPARATE QUI S'ACCOMPAGNE DE QUESTIONNEMENTS .....	46
LES FORCES EN PRÉSENCE.....	48
Liste des start-up françaises du <i>big data</i> .....	48
Liste des entreprises citées dans l'étude .....	51
SOURCES UTILISÉES .....	54
LEXIQUE.....	58

# POINTS-CLÉS ET ENJEUX

Ce qu'il faut retenir

## Le *big data* se déploie à marche accélérée, stimulé par les enjeux économiques

Le *big data*, souvent présenté par les spécialistes comme une révolution en rupture avec la business intelligence, **s'impose comme l'un des principaux piliers de la transformation numérique des entreprises**. Pour elles, cette dernière représente un accélérateur potentiel majeur de la croissance. L'avènement des réseaux sociaux, la généralisation des smartphones, l'essor du multi-équipement numérique dans les foyers, ou encore le boom des objets connectés sont à l'origine d'une **explosion des volumes de données extraites de la géolocalisation des appareils, des médias sociaux ou des capteurs placés dans les équipements**.

Pour les entreprises, cette **prolifération exponentielle des flux d'informations représente une opportunité de créer de la valeur**. Elle offre la possibilité de **tirer un avantage compétitif** de l'exploitation de ces derniers, à travers **une optimisation de leur offre commerciale** pour répondre aux besoins des clients, **la création de nouveaux biens ou services** ou encore **la gestion optimisée** de la production, de la logistique ou de l'organisation. Conscientes de son potentiel, un nombre croissant d'entreprises françaises ont pris le virage du *big data*, une tendance qui s'est amplifiée à partir du milieu de la décennie 2010-2020.

Avec la crise économique héritée de la pandémie de la Covid-19, **les enjeux liés au *big data* deviennent plus prégnants** pour les secteurs économiques ayant fortement subi ses répercussions, comme le tourisme, la restauration ou l'aéronautique. Le *big data* sonne comme une solution crédible pour ces derniers afin de préparer la reprise en **transformant la récession en opportunité d'évolution**. Pour d'autres secteurs plus avancés dans le déploiement, à l'image du e-commerce, la crise a été **l'occasion d'accélérer dans le *big data*, dans une optique d'amélioration des performances**.

Si les intérêts économiques et stratégiques du *big data* s'avèrent importants, **les défis à relever par les entreprises françaises le sont tout autant**. Pour une très grande majorité de petites et très petites entreprises, la mise en place d'une stratégie *data driven* apparaît comme un processus à la fois complexe et coûteux. Il présuppose des investissements importants en technologies et en ressources humaines. Il implique également des mutations organisationnelles et managériales. Pour ces acteurs, le développement de nouvelles offres de services comme le *cloud computing* et l'essor de l'offre de logiciels et applications spécialisés s'imposent comme des opportunités à saisir pour s'engager dans la voie du *big data*.



## Tirer parti de la grande hétérogénéité des besoins pour se positionner

Les acteurs du *big data* disposent de multiples opportunités : de par sa nature polymorphe, **le *big data* s'avère à même de s'adapter à de nombreux secteurs**. Des logiciels d'analyse marketing aux intelligences artificielles, en passant par les modèles prédictifs dans le médical et l'industrie, ou encore les outils d'assistance au personnel : **la diversité des solutions possibles permet aux entreprises du *big data* d'adopter un positionnement différenciant**. La grande variété de clients potentiels renforce ce phénomène, limitant la concurrence entre les nombreuses start-up s'étant lancées sur le marché. Face aux grands groupes du conseil en ingénierie et des services numériques, ces dernières se montrent capables de **répondre de façon plus personnalisée aux besoins des acteurs**, publics comme privés. Les partenariats avec des structures plus importantes leur confèrent les moyens de développer leur activité et d'accroître la conversion de l'économie au *big data*.

L'extension de la collecte et du traitement de quantités massives de données entraîne le développement **d'une multitude d'applications et d'usages inédits**. Encore en mouvement, les contours du *big data* ne cessent d'évoluer en fonction des innovations technologiques. **De nouvelles potentialités apparaissent, y compris dans des secteurs encore peu digitalisés** tels que l'agriculture ou le stationnement urbain.

Des collectivités territoriales aux grandes entreprises et aux organismes publics, **l'ensemble des acteurs s'intéressent aux fonctionnalités permises** par l'analyse des données. Cette accélération des projets laisse parfois la place à une certaine précipitation. Parfois, la protection des informations sensibles n'est pas assurée, ce qui aboutit à l'interdiction de certaines initiatives. Les entreprises positionnées dans les solutions du *big data* doivent **déterminer quelles données collecter et exploiter sans céder à une accumulation excessive**. Outre le risque de congestion, elles s'exposent à des infractions de la réglementation, fragilisant la visibilité sur leurs investissements.

Favorisé notamment par les politiques d'*open data*, **l'accès à la donnée se révèle à la fois croissant et perçu comme stratégique**, sans que l'utilisation finale de ces informations ne soit parfaitement définie. Dans cet environnement changeant, entre innovations techniques et solutions commerciales novatrices, développement de l'offre et découverte de nouveaux usages, **le dynamisme du *big data* nécessite un accompagnement du reste de l'économie**. Trop peu au fait de ces avancées, de nombreuses organisations peuvent retarder leur transition vers ces nouvelles solutions : **accentuer la diffusion des connaissances dans le domaine s'avère incontournable pour garantir la pérennité du marché**.

## Le *big data* contraint de se transformer pour se démocratiser davantage

Collecter et analyser en temps réel les données les plus pertinentes et les plus fiables pour en tirer des enseignements probants et établir des prédictions devient un véritable enjeu pour les entreprises de nombreux secteurs économiques. Cet enjeu s'avère particulièrement central pour certaines fonctions stratégiques telles que le marketing. Pour les entreprises françaises, l'adoption d'une stratégie *big data* est synonyme d'une **accumulation massive de données**, que les technologies des objets connectés et les réseaux sociaux entre autres ne font qu'amplifier. **Toutes les données récoltées ne peuvent cependant être analysées**, ces dernières se révélant pour partie redondantes, non fiables, complexes ou encore trop volumineuses. **Leur traitement devient dès lors un processus complexe et une source potentielle d'inefficacité** car celui-ci mobilise de façon non optimale les ressources de l'entreprise.

L'**intelligence artificielle**, technologie révolutionnaire interdépendante du *big data*, **représente une opportunité technologique pertinente et de première importance dans l'optique d'adopter une approche *smart data***, focalisée sur l'exploitation des seules données pertinentes **et de tendre vers une analyse de type prédictive**.

En parallèle, si le développement de nouvelles offres de services informatiques au cours des dix dernières années, à l'instar principalement du *cloud computing*, a été l'occasion pour de nombreuses petites et moyennes entreprises au budget informatique restreint de s'engager à moindre coût dans la voie de la collecte et de l'analyse de données massives, force est de constater que **les entreprises se heurtent à des contraintes**.

La migration du *big data* dans le *cloud*, reflet de l'externalisation croissante d'une partie du processus *big data* au sein d'infrastructures délocalisées, soulève **certaines problématiques techniques** telles que la saturation de la bande passante ou l'allongement du temps de latence. Elle fait parallèlement peser **la menace d'une perte de souveraineté de l'entreprise en raison du risque de déficit de confidentialité et de sécurité des données**. Ces deux derniers points représentent de véritables enjeux pour les entreprises, contraintes de se conformer au récent règlement européen RGPD relatif à la protection des données.

Face à ces difficultés, il apparaît opportun pour les entreprises d'**ajuster leur approche technologique du *big data* en choisissant des architectures informatiques davantage adaptées à leurs besoins**.

# POINTS-CLÉS ET ENJEUX

Ce qu'il faut retenir

## LES MOTEURS

- La digitalisation croissante des entreprises
- La massification des flux de données
  - Les évolutions technologiques (*cloud computing*, *edge computing*, logiciels, applications, etc.)
- Les retombées économiques stratégiques liées à l'analyse de données massives pour les entreprises
  - L'accroissement du nombre d'entreprises spécialisées (éditeurs de logiciels et d'applications, prestataires de services)

## BIG DATA

## LES FREINS

- Les difficultés de recrutement de profils experts de la data
- Le coût de déploiement du *big data* (infrastructures, logiciels, ressources humaines)
  - La complexité des projets
  - La mesure difficile du retour sur investissement

# LE BIG DATA : UN CONCEPT COMPLEXE, UNE APPROCHE STRATÉGIQUE

## Un objet complexe polymorphe

Terme apparu dans les années 1960, **le big data**, s'est historiquement déployé chez les géants du web et se veut le synonyme de mégadonnées, grosses données ou encore données massives. Selon les experts, il n'existe aucune définition précise ou universelle. En outre, cette dernière est variable selon les communautés qui s'y intéressent, qu'il s'agisse d'utilisateurs ou de fournisseurs de services et/ou de solutions. En 2001, le cabinet Gartner, premier à avoir conduit des travaux pour le définir, désignait le **big data** comme les ressources d'informations dont les caractéristiques, en termes de volume, de vitesse et de variété, imposent l'utilisation de technologies et de méthodes analytiques particulières pour générer de la valeur. Ces technologies et ces méthodes dépassent en général les capacités d'une seule et unique machine et nécessitent des traitements parallélisés. **Le big data peut ainsi être assimilé à un ensemble de technologies, de méthodes et de pratiques spécifiques pour stocker et analyser rapidement de très grandes masses de données pour en tirer des enseignements essentiels.**

Dans son approche, **le big data englobe trois grandes catégories de données** : les données structurées, qui ont un format fixe et sont souvent numériques ; les données non-structurées, c'est-à-dire des informations qui ne sont pas organisées et qui n'ont pas de format prédéterminé ; les données semi-structurées qui peuvent contenir les deux types de données précédentes. Quelle que soit leur qualification, ces données sont des "données numériques produites par l'utilisation des nouvelles technologies à des fins personnelles ou professionnelles. Cela recoupe les données d'entreprise (courriels, documents, bases de

données, historiques de processeurs métiers...) aussi bien que des données issues de capteurs, des contenus publiés sur le web (images, vidéos, sons, textes), des transactions de commerce électronique, des échanges sur les réseaux sociaux, des données transmises par les objets connectés (étiquettes électroniques, compteurs intelligents, smartphones...), des données géolocalisées, etc." liste le site internet de Futura Sciences.

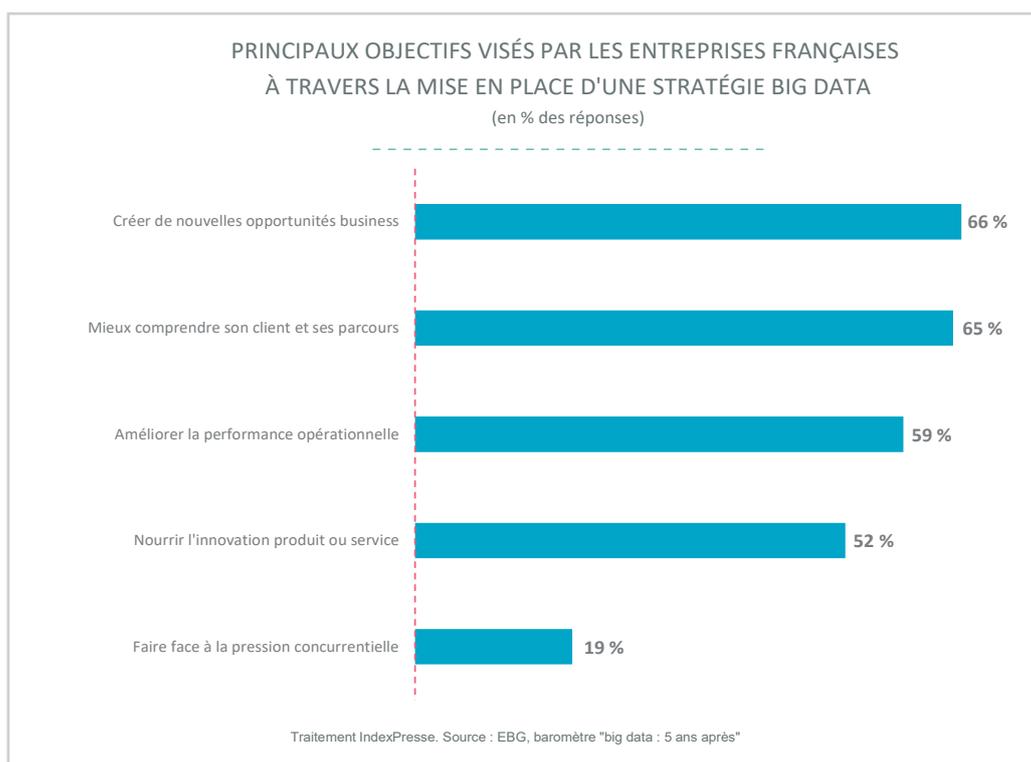
Dans son infrastructure, **le big data regroupe une famille d'outils de type informatique utilisés dans le cadre de la collecte, de l'analyse et de la visualisation de données qui répondent à une triple problématique dite règle des "3V", établie en 2001 par le cabinet Gartner.** Il s'agit d'un volume de données considérable à traiter, d'une grande variété d'informations et d'un certain niveau de vitesse à atteindre. Suite aux évolutions des technologies et à l'expérience des projets de déploiement du **big data** dans les entreprises, deux nouveaux V sont venus enrichir depuis la règle originelle, à savoir la véricité et la valeur.



# Des enjeux stratégiques multiples pour les entreprises

Les objectifs visés par les entreprises françaises engagées dans des projets de déploiement du *big data* sont multiples. Dans la première édition de son baromètre "Big data : 5 ans après" publié début 2019, l'EBG, principal *think-tank* français sur l'innovation digitale, en avait fait ressortir cinq principaux dans le cadre de son enquête menée entre mai et juillet 2018 auprès de 530 de ses adhérents de tous secteurs d'activité et de toutes tailles d'effectif. 38 % de ces derniers avaient, au moment de l'enquête, débuté leur projet il y a moins de 2 ans (en 2016). Un sur cinq l'avait amorcé il y avait plus de cinq ans (en 2013). D'après les résultats de ce sondage, **la création de nouvelles opportunités de business s'imposait en 2018 comme le principal objectif visé** avec les deux tiers des réponses. Le *big data* permet en effet d'enrichir ou de créer de nouveaux produits ou de nouveaux services sur la base des besoins identifiés grâce à lui. La quête de nouvelles opportunités business s'imposait juste devant **la meilleure**

**compréhension du client et de ses parcours** (65 % des réponses). Le *big data* offre la possibilité de mettre en place dans l'entreprise une stratégie de segmentation et de ciblage plus efficace. **L'amélioration de la performance opérationnelle apparaît également comme un objectif important** et arrive en troisième position avec 59 % des réponses. Le *big data* permet de réduire les coûts, d'optimiser les performances côté fournisseurs et de rendre plus efficient le processus décisionnel. Les données les plus utilisées par les entreprises engagées pour répondre à ces enjeux sont principalement issues de sources internes selon les résultats de l'enquête, que ces dernières soient générées par les clients et prospects de l'entreprise (86,5 %) ou issues de navigations loguées (46 %). Les données sociales, à la fois internes (*owned media*) et externes (*social listening*, par exemple) aux entreprises, arrivent en troisième position (38 %), devant les données *open data* et issues de l'internet des objets (IoT).

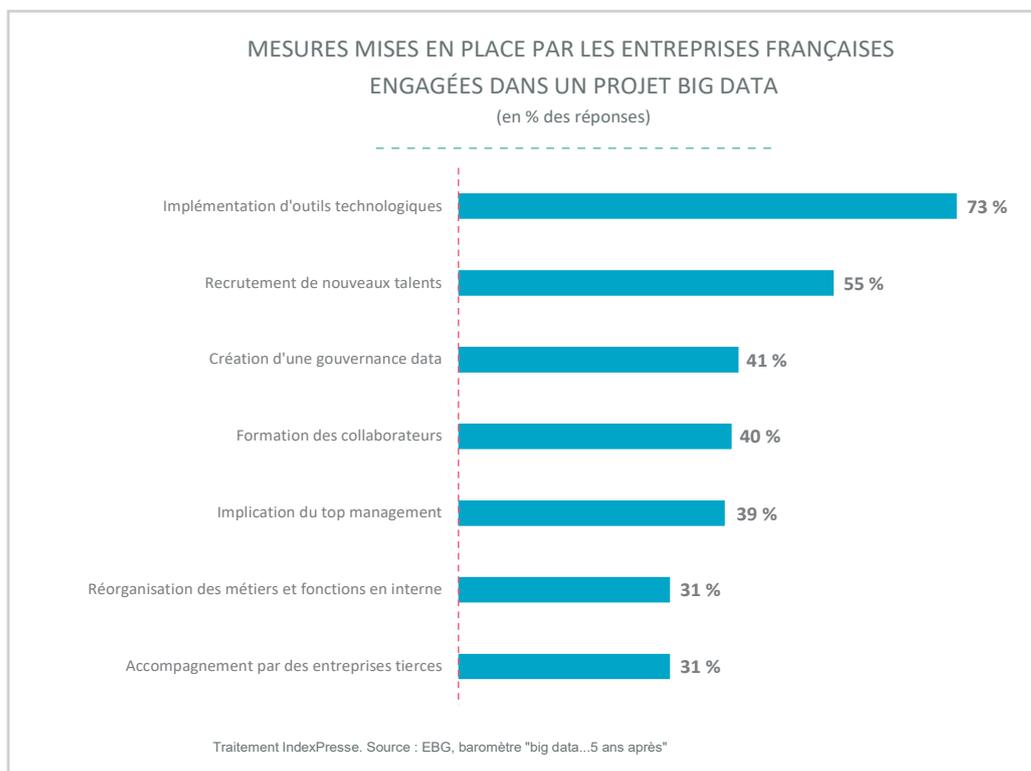


### Les outils et ressources humaines : deux prérequis au déploiement

Dans le cadre de la mise en place de leur projet de *big data*, **les entreprises françaises engagent de multiples actions**, les plus souvent conjointes. **L'implémentation d'outils technologiques** figure assez logiquement en tête des mesures. **La gestion des ressources humaines** occupe aussi une place prépondérante. Le recrutement d'experts était à mi-2018 le deuxième moyen le plus cité par les entreprises engagées selon les résultats du baromètre "*Big data : 5 ans après*" publié début 2019 par l'EBG pour le compte de Qlik, l'un des leaders du marché de la data et de l'analytics en France, et de Micropole, groupe international en conseil et technologies innovantes spécialisé dans la valorisation de la data et le digital. La formation des collaborateurs arrivait, quant à elle, en quatrième position avec 40 % des réponses. L'adaptation organisationnelle de l'entreprise, à travers notamment la création d'une gouver-

nance de la data (41 %) et la réorganisation des métiers et des fonctions en interne (31 %), sont également des leviers importants actionnés par les entreprises. Pour les porteurs de projets de *big data*, **la création d'une gouvernance ad hoc s'impose** notamment pour régir la sécurité des données, leur cadre d'utilisation, leur intégrité et leur fiabilité.

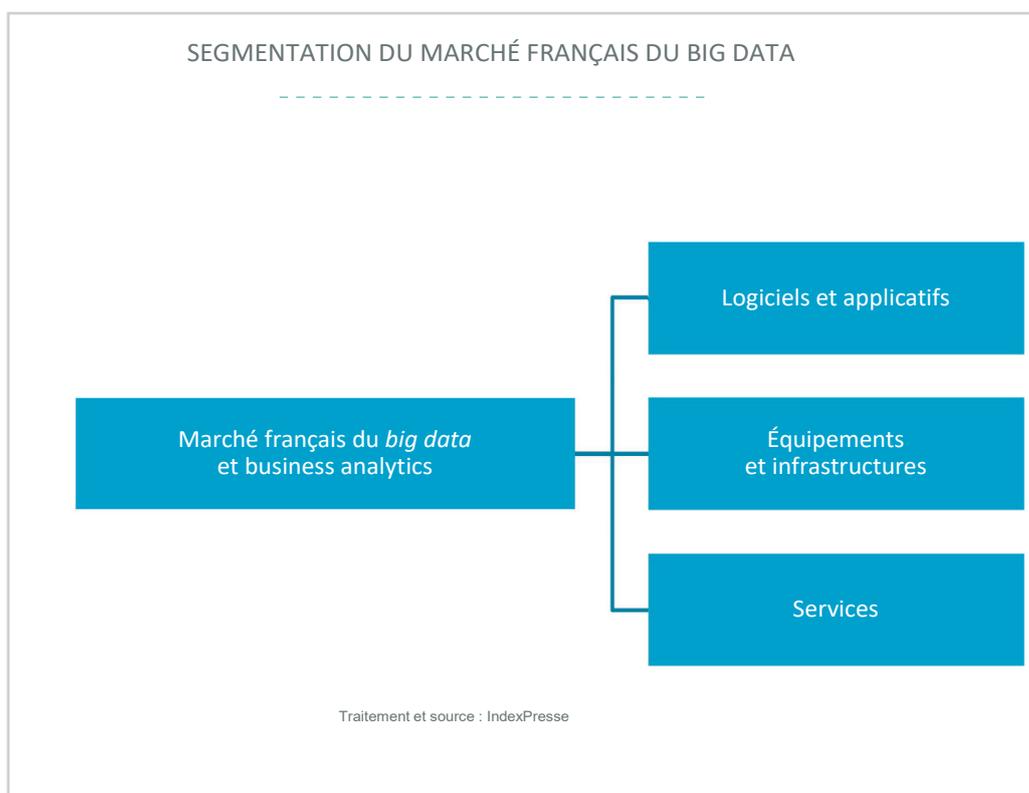
Pour la mise en place de leur projet, de nombreuses entreprises sont contraintes, pour des raisons de coût, de manque de compétences internes ou de complexité dans le déploiement, de **recourir à des prestataires extérieurs spécialisés**. Selon le baromètre de l'EBG, environ trois entreprises sur dix engagées en 2018 mentionnaient l'accompagnement par des entreprises tierces, telles que des sociétés de conseil en intégration, comme l'un des moyens mis en place.



## Un secteur cloisonné

Le marché français du *big data* peut être décomposé en trois grands segments distincts. Parmi eux, deux relèvent de la fourniture d'équipements. Un premier se compose des **équipements et infrastructures**, dans lequel sont intégrés les acteurs dont le cœur de métier repose sur la fourniture de solutions de stockage. Il s'agit principalement des fournisseurs de serveurs ainsi que les gestionnaires de centres de données (*data centers*). Le second grand segment, central, est représenté par les **logiciels et applications**. Selon la catégorisation retenue en 2015 par IDC France, il regroupe les applications analytiques (performance financière, CRM analytique, supply chain analytique, analytiques pour ressources hu-

maines, etc.), les plateformes de Business Intelligence (analyse prédictive ou de contenu, outils de requête et reporting, plateformes logicielles cognitives ou d'intelligence artificielle) ainsi que les plateformes d'intégration et de gestion des données analytiques (entrepôts de données relationnelles, magasins de données non relationnelles, outils d'intégration de données, etc.). Enfin un troisième segment concerne **les services spécialisés** tels que le conseil, la formation ou encore l'accompagnement dans la mise en place d'applications et la concrétisation des projets (analyse des besoins, développement et ou déploiement des outils, structuration, stockage et nettoyage de données, valorisation d'analyses, etc.).



# UN MARCHÉ ENTRÉ DANS UNE PHASE D'INDUSTRIALISATION

## Le tournant du plan français *big data*

Pour le marché français du *big data*, l'année 2014 a marqué un véritable tournant grâce à l'implication forte des pouvoirs publics. Dans le cadre du plan de la nouvelle France industrielle, 34 plans de reconquête industrielle avaient été actés cette année-là. Parmi eux figurait le plan "*big data*". Son ambition affichée était de positionner la France comme le pays de référence en Europe dans le domaine du *big data*, en associant à la fois des fournisseurs de technologies, de grands groupes utilisateurs et des instituts d'enseignement supérieur. La feuille de route adoptée, présentée début juillet 2014 au comité de pilotage, prévoyait plus spécifiquement la mise en place de sept grandes actions spécifiques de nature "à lever des verrous ou à développer le marché du *big data* par une politique volontariste de l'offre", indiquaient les rédacteurs du plan. Par conséquent, trois grandes catégories de mesures ont été préconisées dans le but de renforcer l'offre de tech-

nologies et/ou de services de *big data*, d'accélérer le développement de la demande et de favoriser la rencontre entre technologies et marchés verticaux. Selon la feuille de route initiale, les actions du plan représentaient un potentiel de création ou de consolidation de plus de 137 000 emplois sur la période 2015-2020, soit directement dans l'industrie informatique, soit dans les fonctions technologiques au sein des entreprises, soit encore dans les fonctions métiers des entreprises. Impulsé en 2014 à travers le lancement de projets pilotes labellisés, le développement du secteur du *big data* avait vocation à s'accélérer dès 2015 grâce à une phase de généralisation des projets. En 2015, une réorganisation des 34 plans industriels a été engagée par ministre de l'Économie, de l'Industrie et du Numérique. La trentaine de plans a été regroupée en 10 solutions. Trois de ces solutions visent spécifiquement le développement d'une industrie numérique française.

Plan <i>big data</i> - Feuille de route (juillet 2014)		
Développement de l'écosystème <i>big data</i>	Lancement d'initiatives sectorielles sur le <i>big data</i>	Adaptation du cadre réglementaire et législatif
<p><b>Objectifs :</b></p> <p>renforcer l'offre de technologies et/ou de services <i>big data</i></p> <p><b>Actions/mesures :</b></p> <ul style="list-style-type: none"><li>• mise en place de cursus de formation adaptés aux besoins (data scientists) ;</li><li>• soutien au financement et à l'accélération de start-up ;</li><li>• création d'un observatoire des usages du <i>big data</i> par la DGE</li></ul>	<p><b>Objectifs :</b></p> <p>accélérer le développement de la demande (usages du <i>big data</i>)</p> <p><b>Actions/mesures :</b></p> <ul style="list-style-type: none"><li>• soutien à des projets privés d'expérimentation sur de nouveaux usages de la donnée ;</li><li>• lancement de grands projets dans le secteur public</li></ul>	<p><b>Objectifs :</b></p> <p>lever les verrous réglementaires et législatifs</p> <p><b>Actions/mesures :</b></p> <ul style="list-style-type: none"><li>• soutien à des projets privés d'expérimentation sur de nouveaux usages de la donnée ;</li><li>• lancement de grands projets dans le secteur public</li></ul>

Traitement et source : IndexPresse



## Des flux de données exponentiels hérités des avancées technologiques

Depuis le milieu de la décennie 2010-2020, **les flux de données générées explosent en France**. Selon la troisième édition du rapport "Global Data Protection Index" publié en 2019 par Dell EMC et réalisé à partir d'un panel de décideurs IT d'entreprises de plus de 250 personnes, le volume de données générées en France a été multiplié par huit au cours de la période 2016-2018. Cette croissance s'est révélée supérieure à celle relevée à l'échelle mondiale (+569 % dans 18 pays). Cette explosion des flux est principalement à mettre en relation avec **l'avènement des réseaux sociaux et de l'Internet des Objets (IoT)**. Les marchés des solutions et des équipements connectés (IoT) ont connu **un essor considérable entre 2015 et 2020** en raison à la fois d'un accroissement du nombre d'acheteurs et d'une pratique grandissante du multi-équipement numérique au sein des foyers français.



Sur le territoire national, le chiffre d'affaires réalisé à travers les ventes BtoC d'objets connectés a notamment été multiplié par un peu plus de onze au cours de la période 2014-2019 d'après les chiffres communiqués par le cabinet d'études GfK, toutes catégories de produits confondues (domotique, montres, bracelets, drones, etc.). En 2020, elles auraient encore progressé de près de 25 %, franchissant ainsi la barre de deux milliards d'euros. Par ailleurs, la France comptait 57 millions de compteurs communicants déployés en 2020. **Les individus génèrent d'une quantité exponentielle de données** et de traces numériques à travers l'utilisation de leurs ordinateurs, smartphones, tablettes, caméras, capteurs, compteurs électriques ou encore des jouets connectés, dont les taux d'équipement augmentent pour chacune de ces catégories de solutions.

TAUX D'ÉQUIPEMENT DE LA POPULATION FRANÇAISE EN ÉQUIPEMENTS ET OBJETS CONNECTÉS EN 2019  
(en % de la population âgée de 12 ans et plus)

Produits	Taux d'équipement de la population française en 2019
Ordinateurs	76 %
Smartphones	77 %
Tablettes	42 %
Objets connectés	16 %
Enceintes connectées	10 %

Traitement IndexPresse. Source : ARCEP, baromètre du numérique

## Des évolutions technologiques en amont déterminantes

En France comme au plan mondial, l'apparition et le déploiement à plus grande échelle du *big data* dans les entreprises françaises à compter de 2015 n'auraient pu être possibles sans **les avancées, en amont, des technologies numériques et informatiques**. Ces dernières, qui se sont avant tout traduites par la création ou par l'essor d'offres de nouvelles solutions de stockage et de traitement ajustées ainsi que par l'émergence de nouveaux modes de calcul à haute performance, ont contribué à accélérer le déploiement des projets de *big data* dans les entreprises hexagonales. **Le décollage du *cloud computing* a été l'un des facteurs déterminants** de l'accélération des projets de *big data* depuis 2015. Progressivement, il tend à s'imposer dans les entreprises comme un véritable support des projets de *big data* **grâce à ses multiples atouts**. Il "permet aux sociétés de s'affranchir de tout ou partie de leurs investissements en infrastructures informatiques pour utiliser, à la place, des services en ligne : logiciels (software as a service, SaaS), stockage, puissance

de calcul, etc." résume Nadège Pradine, analyste à l'INSEE, dans un dossier consacré au *cloud computing* et au *big data* paru en novembre 2019 dans la collection INSEE Références. "La gestion externalisée des infrastructures informatiques est la caractéristique principale du *cloud computing* (informatique en nuage). Il consiste à fournir via internet des services informatiques aux entreprises, leur permettant ainsi de réaliser des économies et de transformer des investissements physiques (serveurs, réseau, etc.) en contrats de location de services en ligne", explique-t-elle. **L'exploitation du *big data* dans le cloud rend ainsi, notamment pour les entreprises à budget limité, le processus d'analyse de données moins coûteux** grâce à de multiples sources d'économies, en particulier au niveau des infrastructures informatiques : coûts liés à la maintenance et aux mises à niveau du système, consommation d'énergie, gestion des installations, etc. Par ailleurs, le *cloud* est une technologie flexible et évolutive et permet un partage des données en temps réel.

### Évolutions technologiques majeures à l'origine du déploiement du *big data*

#### Technologies de stockage

Passage de serveurs informatiques physiques internes à l'entreprise à des serveurs dit "*cloud*" (*cloud computing*), dotés d'une capacité de stockage bien supérieure.

En 2018, 20 % des sociétés de 10 personnes ou plus implantées en France (hors secteurs agricole, financier et d'assurance) et payant pour des services de *cloud computing* recourant à l'achat de puissance de calcul, 77% au stockage de fichiers.

#### Technologies de traitement ajustées

Explosion des outils d'analyse, d'extraction et de traitement des données de manière non structurée que cela soit du NoSQL (famille de systèmes de gestion de base de données qui s'écarte du paradigme classique des bases relationnelles) ou de nouveaux *frameworks* lié à l'écosystème de Hadoop (plateforme).

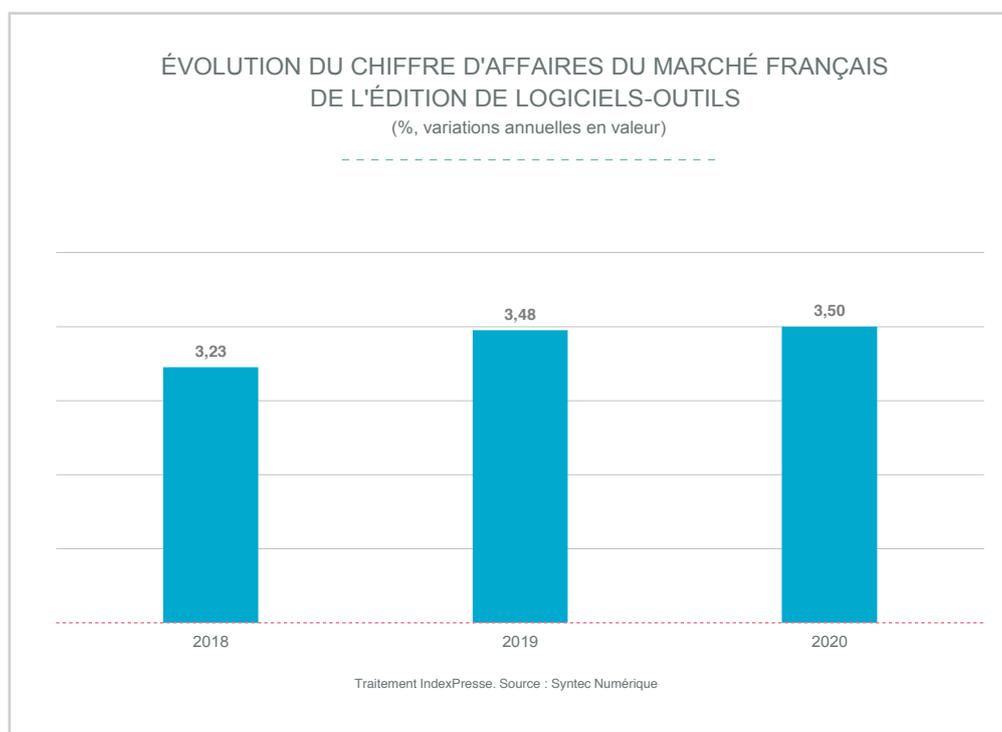
Traitement IndexPresse. Sources : presse spécialisée et INSEE, enquête TIC

## Les porteurs de projets investissent dans les outils et les services *analytics*

### Des ventes de logiciels outils à un niveau record en 2020

Segment clé du secteur du *big data*, le marché français des logiciels et applications connaît **une croissance solide depuis 2017**. Bien qu'en très fort ralentissement en 2020 compte tenu des répercussions de la crise sanitaire de la Covid-19, le chiffre d'affaires généré dans l'Hexagone par les éditeurs de logiciels outils (applications et outils analytiques, de gestion, de développement, d'intégration, d'orchestration et de mise en qualité de données) a atteint **le niveau record de 3,5 milliards d'euros** selon les données communiquées par Syntec Numérique, organisation professionnelle de l'écosystème numérique français qui rassemble quelque 700 éditeurs. **Depuis 2017**, l'indicateur a enregistré une progression de

8 % en valeur, soit l'équivalent de 270 millions d'euros supplémentaires. Ce dynamisme illustre "le reflet des investissements des donneurs d'ordres dans la *data* et dans la modernisation de leur patrimoine applicatif" concluait la fédération Syntec Numérique dans un communiqué de presse publié début décembre 2020 à l'occasion de la présentation des résultats annuels du secteur. Les entreprises souhaitant mettre en place une stratégie *big data* ont à leur disposition **une très grande diversité de logiciels et d'applications** spécifiques d'exploitation des données *big data* qui aident à collecter, à stocker, à analyser, à rapporter, et à présenter les données. À titre d'exemple, la société GetApp, figurant parmi les premières ressources de logiciels SaaS pour les entreprises, recense près de 90 outils *big data* sur sa plateforme de découverte de logiciels. Les outils les plus répandus et populaires sont



avant tout des plateformes dites *open source*, c'est-à-dire des logiciels informatiques dont le code source est librement accessible, modifiable et redistribuable. Ces plateformes ont indubitablement contribué à un déploiement plus massif du *big data* dans l'Hexagone. La plus connue et utilisée d'entre elles a longtemps été Hadoop, un *framework* gratuit écrit en Java et sous licence Apache. Le marché des plateformes gratuites Hadoop est dominé depuis plusieurs années par deux éditeurs majeurs ayant fusionné début 2019, Cloudera et Hortonworks. Le troisième grand acteur, MapR, a quant à lui été racheté à l'été 2019 par Hewlett-Packard Enterprise (HPE). "Des plateformes *big data* ont été déployées massivement dans les entreprises, toutes étant liées pour des raisons historiques à la technologie Hadoop. Mais alors que celles-ci constituaient à une époque le prérequis technologique nécessaire, ce n'est plus du tout vrai aujourd'hui", affirmait Yann Carpentier-Gregson, Practice Manager "Data Architectures" chez Umanis, dans une chronique publiée en septembre 2020 sur le site *decideo.fr*. Selon les experts, **les plateformes Hadoop souffrent d'inconvénients importants**. La complexité d'exploitation, le faible retour sur investissement, le manque d'agilité ou encore la faible adaptabilité au traitement de données en petites quantités figurent parmi les principaux d'entre eux selon l'expert.

"Pour faire face aux inconvénients constatés à l'usage des plateformes Hadoop, les fournisseurs de solutions *cloud* se sont mis à proposer des services spécialisés permettant d'assembler des plateformes dédiées répondant chacune à un usage particulier", rappelle Yann Carpentier-Gregson. "On peut citer par exemple : le stockage de volumes importants de données avec de bonnes performances d'entrées/sorties (Azure Datalake Store, AWS S3, Google Cloud Storage...), des capacités de calcul à mise en place rapide et ajustables (AWS EC2, K8S...), des outils de Business Intelligence (BI) dédiés au *big data* (Redshift, Azure Synapse, Snowflake, BigQuery...), des outils d'ingestion, transformation et modélisation haute vélocité (Databricks...), des plateformes Hadoop louées en mode *cloud* (HD Insight, EMR...)", explique-t-il.

### Le marché français des *analytics* décolle depuis 2015

Le marché français de l'*analytics* connaît une phase de forte croissance en France depuis le milieu de la décennie 2010-2020. Entre 2016 et 2020, **le chiffre d'affaires cumulé généré par les ventes de logiciels et les prestations de services informatiques associés dans ce domaine a triplé** dans l'Hexagone selon les données du syndicat professionnel Syntec Numérique. Une proportion importante de cette progression repose sur l'accroissement de l'offre (développement de nouvelles solutions, augmentation du nombre d'éditeurs de logiciels et applications spécialisées).

Malgré la conjoncture difficile marquée par la crise sanitaire de la Covid-19, le marché français a encore affiché **une augmentation de près de 3,5 % sur l'année 2020**. Cette progression, équivalente à 100 millions d'euros supplémentaires sur l'exercice, s'avère toutefois inférieure de 2,7 points à celle relevée pour l'ensemble du marché des SMACS (Social, Mobile, Analytique, Cloud et Sécurité), lequel regroupe en France quelque 3 250 sociétés actives au total à mi-mai 2021 selon la base de données *compuBase*. S'élevant à **2,5 milliards d'euros** cette dernière année selon les estimations de Syntec Numérique, le marché des *analytics* a atteint un niveau record depuis 2015 et représentait près de 6 % de l'ensemble du marché français des logiciels et des services informatiques.

208

Le nombre d'éditeurs de logiciels "*analytics*" en activité recensés en France à fin mars 2021

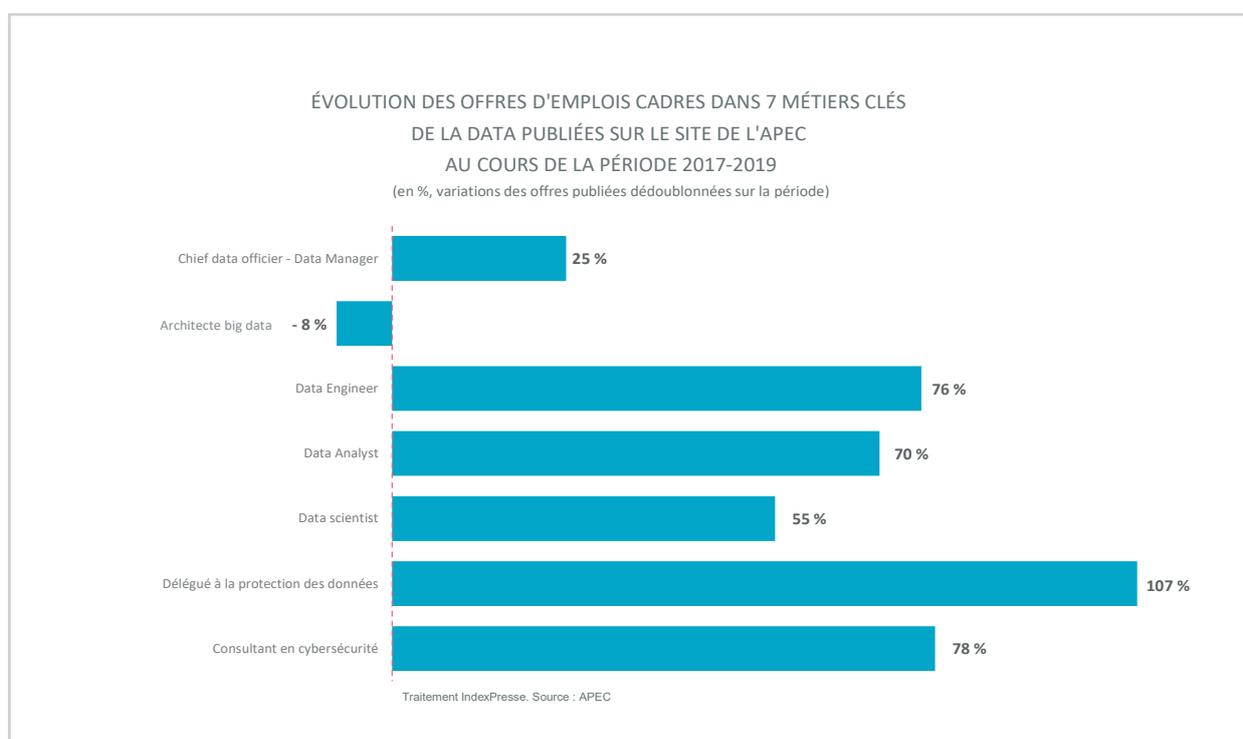
Source: *CumpuBase*

## Des entreprises en quête de recrutement de profils experts

### Des métiers de la data très recherchés

Face aux besoins croissants de main d'oeuvre qualifiée, **les recrutements de personnels spécialisés dans le domaine de la donnée augmentent fortement depuis 2015**. Selon une étude de l'APEC consacrée au sujet et publiée en octobre 2020, les offres d'emplois cadres parues sur le site *apec.fr* ont affiché **un bond de 60 % entre 2017 et 2019 pour les sept métiers clés de la data** analysés par l'association. En moyenne, près de 3 200 postes cadres ont été à pourvoir chaque année en France sur la période, avec un pic atteint en 2019 avant la crise sanitaire de la Covid-19, avec près de 3 800 offres recensées. **La plus forte croissance revient aux métiers dans le domaine de la protection et de la sécurisation des données (+ 83 %)**. Les métiers liés au traitement et à l'analyse de données (*data scientist*,

*data analyst*) ainsi qu'à la collecte et au stockage des données (architecte data, *data engineer*) complètent le podium, avec des hausses d'emplois à pourvoir de respectivement + 61 % pour la première catégorie et + 58 % pour la deuxième entre 2017 et 2019. En dehors des postes dans le domaine de la protection et de la sécurisation des données, **les offres pour les métiers de *data engineer* et de *data scientist* ont enregistré les plus fortes progressions** (respectivement + 76 % et + 70 %). Depuis quelques années, **le métier de *data curator* émerge** en parallèle, à la faveur des mutations du *big data*. Le rôle du *data curator* consiste à la fois à gérer les métadonnées, la protection, la gouvernance et la qualité des données. Cet opérateur est souvent responsable de présentations de données sous forme de tableaux de bord, de graphiques ou de diaporamas. Il interagit régulièrement avec les chercheurs, et planifie des ateliers éducatifs.



## UN MARCHÉ ENTRÉ DANS UNE PHASE D'INDUSTRIALISATION

Avec l'obsolescence des outils traditionnels de gestion de données, **de nouvelles compétences ont émergé pour répondre aux nouveaux besoins des directions des services informatiques**, dont les compétences internes s'avèrent souvent insuffisantes. Les profils d'experts en *data science* sont de plus en plus recherchés par les entreprises de tous secteurs qui cherchent à organiser, à analyser et à comprendre une très grande quantité d'informations dont elles disposent pour orienter leur stratégie *data driven*. **Certaines entreprises ont d'ailleurs été amenées à faire évoluer leur organisation vers un modèle dit *data centric***, avec la constitution d'équipes et/ou la mobilisation de compétences techniques dédiées, ou via l'acquisition de technologies et d'outils spécifiques.

**Les besoins futurs de recrutement de postes du domaine du *big data* sont encore appelés à croître.** "La croissance exponentielle des données et de la digitalisation de la société promet un bel avenir aux métiers liés à la *data*. L'arrivée de réglementations comme la RGPD en Europe par exemple, qui encadre l'utilisation des données et en renforce la protection, va obliger les entreprises à recruter de plus en plus de profils experts pour isoler ces données personnelles et les traiter correctement", explique Fany Pégard, rédactrice web pour *Futura Sciences*. Ce besoin identifié de recrutement de spécialistes de la donnée est également partagé par Stéphane Briffod, Directeur Avant-Vente chez Qlik, éditeur de logiciels de business intelligence et de data-visualisation, qui affirmait, dans un article paru en août 2020, que "le besoin en ingénieurs de la donnée est actuellement en train de croître". **La crise sanitaire de la Covid-19 pourrait même accélérer la tendance** compte tenu des nouveaux besoins des entreprises. "Dans un contexte de pandémie et de récession économique, les entreprises veulent naturellement se protéger du pire et multiplient les investissements pour optimiser leurs modèles de prédiction et la résilience de leurs activités face aux risques", argumentait Antoine Onnen, cofondateur et CTO de *Societeinfo.com*, spécialisée dans l'enrichissement de données BtoB dans un article publié début mars 2021 par *L'Usine Digitale*. Par ailleurs, "la crise invite les entreprises

à développer une connaissance plus fine de leur marché, de leurs clients, de l'écosystème. Le besoin de faire parler la donnée pour en soutirer de la valeur est critique en ces temps d'incertitudes mais aussi de nouvelles opportunités", expliquait de son côté Laurent Delattre, journaliste pour la revue spécialisée *IT for Business* dans un article publié en septembre 2020 consacré au marché de l'emploi IT et la rémunération des métiers de l'IT en 2021.

### OFFRES D'EMPLOIS DANS LES MÉTIERS DU BIG DATA PAR DOMAINE D'INTERVENTION (2017-2019)

#### 1 TRAITEMENT ET ANALYSE DE LA DONNÉE

(44 % DES OFFRES APEC  
SUR LA PÉRIODE 2017-2019 : + 61 %)

Chief data officer,  
Data Manager

#### 2 COLLECTE ET STOCKAGE DE LA DONNÉE

(30 % DES OFFRES APEC  
SUR LA PÉRIODE 2017-2019 : + 58 %)

Architecte Big Data,  
Data Engineer

#### 3 PROTECTION ET SÉCURITÉ DE LA DONNÉE

(19 % DES OFFRES APEC  
SUR LA PÉRIODE 2017-2019 : + 83 %)

Délégué à la protection des données,  
Consultant en cybersécurité

Traitement indexpresse. Source : APEC



### Des difficultés de recrutement persistentes

Si les besoins sont importants et augmentent ces dernières années, les entreprises éprouvent toutefois certaines des difficultés à recruter des profils experts de la *data*. “Aujourd’hui, **le facteur humain semble être le principal frein à la réussite et à la mise en place de projets *big data***. Les entreprises ne butent plus sur la technique mais sur l'acculturation, la collaboration entre profils métiers et experts techniques. 90 % des entreprises estiment que les difficultés rencontrées dans leurs projets *data* ne viennent pas de la technologie. Pour ces organisations, les difficultés dans les projets de données et de transformation digitale sont venues des personnes, des processus et de la culture”, résumait Manon Philippe, responsable des conférences du salon professionnel Big data Paris, dans une interview publiée en février 2020.

Les difficultés rencontrées par les entreprises sont multiples d’après les analystes. Ces dernières relèvent en premier lieu de **la complexité d’identifier le besoin**. “Il est en effet souvent difficile pour une organisation, surtout si elle n’a que très récemment mis en place une stratégie d’exploitation de ses données de pouvoir déterminer efficacement ses besoins en termes de ressources pour maîtriser le *big data*”. Par ailleurs, **la variété des métiers mobilisables peut constituer un frein**. “Face à cette variété et à la complexité, les entreprises doivent parfois mener une réflexion, voire demander une expertise extérieure en recrutement digital”. Enfin, “[...] il existe des raisons contextuelles, puisqu’il y a quelques années la *data* n’occupait pas une telle place au sein des entreprises et les primo-arrivants sur le marché de l’emploi, fraîchement diplômés, ne sont pas formés au traitement massif des données, qui nécessite de maîtriser de nouvelles solutions et de nouveaux langages, notamment dans le *cloud*”, expliquait Stéphane Briffod, Directeur Avant-Vente chez Qlik dans un article publié en août 2020 sur le site internet *decideo.fr*. **Les profils *big data* sont donc encore relativement rares dans les formations d’ingénieurs** en France. “Actuellement,

on compte environ 168 écoles d’ingénieurs et 150 000 étudiants pour l’ensemble du territoire. Parmi ces derniers, on estime que moins d’un tiers sont inscrits dans un cursus ‘informatique et sciences informatiques’ et visent des carrières de développeurs et/ou d’analystes”, affirmait le magazine *L’Usine Digitale* en mars 2021. “En plus d’être très sélective, la formation des futurs développeurs informatiques, *data analyst*, *data scientist*, etc., est très hétérogène. Par conséquent, **les entreprises font face à un pool réduit de candidats, avec des profils et compétences extrêmement variables**. Rares sont les talents qui réunissent à la fois des “*hard skills*” (bases informatiques et mathématiques suffisantes pour maîtriser le codage et le traitement lourd de données, bases en français et en anglais pour rédiger des contenus) et des “*soft skills*” (qui inclut l’expérience du terrain)”, expliquaient les analystes.

Par ailleurs, **les niveaux de salaires des métiers de la *data* sont jugés par les analystes comme peu attractifs** en France. À l’origine d’une fuite des compétences à l’étranger, ils représenteraient un autre facteur des difficultés de recrutement de profils experts de la part des entreprises. Face à cette problématique, **les grilles salariales tendent à être revalorisées** en France ces dernières années. Selon les résultats d’une étude sur les rémunérations conduites par le cabinet de recrutement PageGroup, filiale du groupe Michael Page France, entre 2019 et juin 2020 via son réseau de consultants, les salaires des professionnels des nouvelles technologies sont toutefois orientés à la hausse pour 2021. “Certains profils seront particulièrement sollicités en 2021. Parmi ceux qui seront les mieux rémunérés, on retrouve le *data analyst/data scientist* avec un salaire annuel de l’ordre de 35 000 à 45 000 euros en début de carrière, évoluant pour les profils les plus expérimentés (5-15 ans) à 60 000 € euros voire à 80 000 euros”, indiquait Véronique Arène, journaliste au groupe d’information et de services dédié aux professionnels de l’informatique en France *Le Monde Informatique* dans un article publié en septembre 2020.

## Les infrastructures délocalisées de stockage se déploient sur le territoire

La nette accélération des flux de données à traiter par les entreprises dans l'Hexagone depuis le milieu des années 2010 et le développement du *big data* dans le *cloud* ont engendré **un accroissement des besoins en infrastructures externalisées de stockage de données sur le territoire national**. Selon les données cartographiques publiées en mars 2021 par le magazine *Global Security Mag*, spécialisé en sécurité informatique, en France, **le parc en service de *data centers* neutres d'hébergement s'est étoffé d'une cinquantaine d'unités supplémentaires au cours de la période 2012-2020**, soit une hausse d'environ 7 unités par an. **À mi-mars 2021, 186 centres neutres de données étaient actifs** en France métropolitaine selon le dernier recensement effectué par le magazine, parmi lesquels environ 35 % d'entre eux étaient implantés dans la seule région Île-de-France. La France se positionnait au quatrième rang mondial sur le critère du nombre de *data centers* neutres actifs.

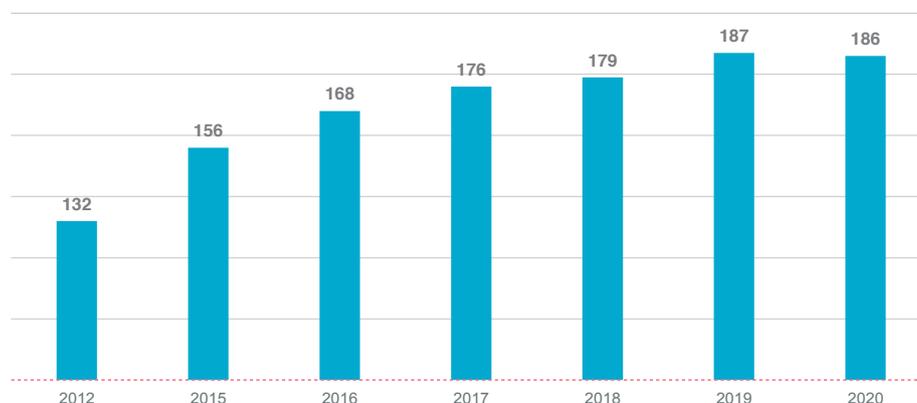
### FOCUS

#### LES DATA CENTERS

Un *data center* ou centre de données, est une infrastructure composée d'un réseau d'ordinateurs et d'espaces de stockage. Cette infrastructure peut être utilisée par les entreprises pour organiser, traiter, stocker et entreposer de grandes quantités de données. Un centre de données basique regroupe des serveurs, des sous-systèmes de stockage, des commutateurs de réseau, des routeurs, des firewalls, et des câbles et des racks physiques permettant d'organiser et d'interconnecter tout cet équipement informatique. Un *data center* non-neutre appartient à un opérateur de service. À l'opposé, les *data centers* neutres abritent de très nombreux fournisseurs de services et permettent donc d'obtenir au meilleur prix ces services.

Source : [lebigdata.fr](http://lebigdata.fr)

ÉVOLUTION DU PARC DE DATA CENTERS NEUTRES EN SERVICE EN FRANCE MÉTROPOLITAINE  
(nombre de data centers actifs)



Traitement IndexPresse. Source : Global Security Magazine

# 29

Le nombre  
de *data centers* neutres  
en construction  
en France métropolitaine  
à mi-mars 2021

Source : Global Security  
Magazine.

Tendance majeure observée dans le secteur, **la régionalisation des *data centers* s'est accélérée depuis 2015 en France**. De plus en plus de centres de stockage de données sont ainsi implantés en région, notamment à proximité ou dans les grandes métropoles de Lyon, Lille, Marseille, ou encore Bordeaux. "De plus en plus d'entreprises et d'organisations publiques externalisent leurs systèmes informatiques. Sachant que 80 % de près des 2,8 millions entreprises françaises se situent en région, les besoins en *data centers* ne peuvent que s'intensifier en dehors de l'Île-de-France", affirmait Jean-Pierre Tournemaine, Country Manager chez Vertiv France dans un article publié par *informatiquenews* en août 2020. "En région, de nombreuses infrastructures existent déjà. Des acteurs locaux se sont développés. Les façades méditerranéennes et atlantiques reçoivent les câbles sous-marins des continents africain, asiatique et américain. Elles offrent ainsi des emplacements idéaux pour déployer des infrastructures – des villes et région comme Marseille, Bordeaux ou la Bretagne sont ciblées", ajoutait-il.

**Le déploiement des *data centers* sur le territoire national est appelé à se poursuivre** dans l'avenir selon les experts. "L'explosion à venir du trafic, la transformation digitale des entreprises ou encore l'externalisation grandissante des infrastructures informatiques offrent par ailleurs de solides perspectives de croissance aux exploitants de *data centers* neutres", argumente Xerfi dans une étude sectorielle consacrée aux hébergeurs et aux gestionnaires de *data centers* publiée à l'été 2020. "L'arrivée de la 5G, l'accélération des développe-

ments de l'intelligence artificielle, l'élargissement de la gamme d'outils répondant aux nouveaux usages liés à la réalité virtuelle et la démocratisation de l'internet des objets, contribuent à porter le marché des *data centers* avec une croissance exponentielle du volume des données. Ces révolutions technologiques demandent toujours plus de disponibilité et de fiabilité en matière de stockage des données", résumait de son côté Michel Dernis, président d'Atrium Data, société d'audit et de conseil spécialisée dans le domaine des *data centers*, dans une interview parue en juin 2020 dans Forbes.

**La France dispose par ailleurs d'atouts qui en font un pays attractif pour l'implantation de *data centers***. "La France est un territoire favorable au développement des *data centers* car elle offre de bonnes performances en termes de cybersécurité et son réseau énergétique est l'un des plus fiables d'Europe, constituant ainsi un environnement idéal de ce point de vue" soulignait le cabinet d'études néerlandais en ingénierie de la construction Arcadis dans son étude comparative d'implantation des *data centers* dans le monde parue en mars 2021. En termes d'attractivité, **la France se classe neuvième sur une cinquantaine de marchés** intérieurs étudiés. "L'intérêt des investisseurs pourrait être décuplé par les initiatives gouvernementales récentes, comme l'allègement de la fiscalité énergétique en complément du développement du très haut débit sur l'ensemble du territoire", ajoute le cabinet. Pour les *data centers*, la croissance du *cloud* et du *serveless*, le développement du *edge computing* et du prédictif ainsi que l'arrivée de la 5G font partie des grands enjeux d'avenir. Toutefois, **la poursuite du déploiement du parc pourrait être en partie ralentie par l'essor des serveurs loués** selon certains analystes.

Si l'essor du *big data* observé depuis le milieu de la décennie 2010-2020 a avant tout été perceptible à travers une hausse des investissements des entreprises en moyens humains et matériels, son développement a parallèlement été visible à travers l'élargissement de ses usages et son déploiement dans de nouveaux secteurs.

# DES USAGES TRANSVERSAUX DANS DES DOMAINES VARIÉS

## Le *big data*, enjeu clé pour l'analyse marketing

### Détecter les tendances et les signaux faibles

En agglomérant de grandes quantités de données, le *big data* améliore la capacité d'analyse macroscopique des équipes marketing. Ces dernières parviennent à identifier des phénomènes de marché par le croisement d'informations provenant d'échantillons statistiques de grande envergure, l'exploitation de données inutilisées (*dark data*) ou encore à décrypter finement les attentes des consommateurs grâce à une veille massive et automatisée.

#### De nombreux acteurs se positionnent

Créée en 2000, Sidetrade se positionne sur l'aide au recouvrement des créances clients via l'intelligence artificielle et le *big data*. Dans une interview donnée à *BFM Business* en février 2021, le PDG de l'entreprise, Olivier Novasque, déclarait que l'entreprise "stock[ait] depuis 20 ans les comportements de paiement de plus de quatre millions d'entreprises". Il ajoutait que le volume d'informations collectés représentait "440 millions d'expériences de paiement sur les trois dernières années, [soit] mille milliards d'euros". L'entreprise peut ainsi détecter les méthodes et les moments les plus efficaces pour que son client engage des relances et récupère les sommes dues. Sidetrade a étendu le champ de son actions avec les rachats en 2016 de BrightTarget (Royaume-Uni) et d'IKO System (France), permettant de proposer une so-

lution intégrée d'augmentation du chiffre d'affaires et de génération de flux entrants de trésorerie. IKO System a été créée en 2010 et s'appuie sur le *big data* pour déterminer les prospects les plus pertinents pour le client.

De son côté, la société TINYCLUES propose une solution d'amélioration du ciblage des campagnes marketing. Créée en 2010, l'entreprise affirme détecter les signaux faibles permettant de s'adresser à des prospects n'ayant pas manifesté explicitement une intention d'achat. Ses algorithmes prédictifs seraient ainsi capables de fournir aux utilisateurs des publics cibles plus pertinents. David Bessi, CEO de TINYCLUES, déclarait en 2018 au site Frenchweb : "Nous captions les signaux implicites contenus dans les données afin de permettre aux marketeurs de créer, en quelques clics, des campagnes d'animation commerciale ciblées, sans mobiliser de data scientists". La société a levé 18 millions de dollars en 2018 pour se développer à l'international. Elle avait déjà récolté 5 millions d'euros en 2015.

Fondée en 2012, Contentsquare est quant à elle une société française spécialisée dans l'analyse des comportements des internautes. Les informations recueillies sont multiples, du parcours client aux mouvements de souris et au temps passé sur un texte ou une image. Les données agrégées permettent de détecter les éléments favorisant le passage à l'achat sur les sites de e-commerce, ou au contraire d'indiquer ceux à



modifier ou à améliorer pour accroître les ventes. L'entreprise a réalisé en mai 2021 **une levée de fonds de 500 millions de dollars**, notamment auprès du Japonais SoftBank. Elle avait déjà rassemblé 190 millions de dollars en 2019. La même année, elle avait fait l'acquisition du français PricingAssistant, spécialisé dans le suivi des prix sur Internet et l'e-merchandising.

La start-up PricingHub, fondée en 2018, se positionne également dans le domaine de **l'analyse des prix sur Internet**. L'objectif est de permettre une compréhension fine des attentes des consommateurs afin d'**optimiser la politique de prix, et ainsi de préserver davantage les marges** des clients de la société. Elle a levé deux millions d'euros fin 2019. Une autre start-up française, Pricemoov, se place en concurrent de PricingHub et a récolté trois millions d'euros en 2018.

Fondée en 2019, la société Appolo Plus a développé des algorithmes avec le CNRS. Ces derniers sont dédiés à l'analyse marketing, permettant de **réaliser des modèles prédictifs de la demande**.

La meilleure compréhension de celle-ci grâce aux données recueillies et traitées par la solution d'Appolo Plus conduit à **un positionnement de l'offre plus efficient**. La start-up affirme que ses clients connaissent en moyenne **une hausse de 15 % de leurs revenus** de par l'utilisation de son logiciel. Elle a mis en place en 2019 un partenariat avec la société de conseil Citwell afin d'accéder plus facilement à des clients ayant des problématiques liées à la chaîne logistique.

### Le big data se déploie aussi via des partenariats

Certaines entreprises ont noué des partenariats avec les grands acteurs du numérique pour mettre en place des projets *big data*. C'est notamment le cas dans la grande distribution.

Lancée en juin 2019 avec Microsoft, la **Data Factory d'Intermarché** constitue un élément phare de la transformation digitale de l'enseigne et de l'irruption du *big data* dans le secteur. Pour générer un chiffre d'affaires supérieur à surface de

magasin constante, des algorithmes collectent et analysent une variété d'informations : de la nature de la clientèle à la localisation du site en passant par son format. Le recoupage de ces données permet de **mieux assortir les produits proposés** dans les rayons à la zone de chalandise, accroissant ainsi les ventes. En accélérant la rotation des articles, la solution réduit également leur coût de stockage.

Dans une démarche similaire, **le groupe Carrefour a mis en place le Hub Digital en avril 2019 en partenariat avec Google**. L'enseigne en attend un gain de chiffre d'affaires significatif, comme l'a indiqué à LSA en février 2020 Amélie Oudéa-Castéra, directrice exécutive transformation digitale, e-commerce et *data* de Carrefour : "Les cinq principaux projets pourraient générer 250 à 280 millions d'euros de revenus incrémentaux sur l'année 2020". La supply chain représente aussi l'un des domaines sur lesquels se focalisent le Hub Digital. Ses algorithmes accumulent de nombreuses informations relatives au e-commerce alimentaire et les implémentent dans des modèles probabilistes. Ils sont parvenus à atteindre **des taux de précision supérieurs à 90 %, voire proches de 99 %**, ce qui équivaut à une marge d'erreur de seulement cinq commandes. Déterminer le niveau de la demande de façon aussi fine améliore les performances logistiques en **évitant les ruptures de stocks, les congestions et optimise la gestion du personnel** (besoins de main d'oeuvre et charge de travail).

5

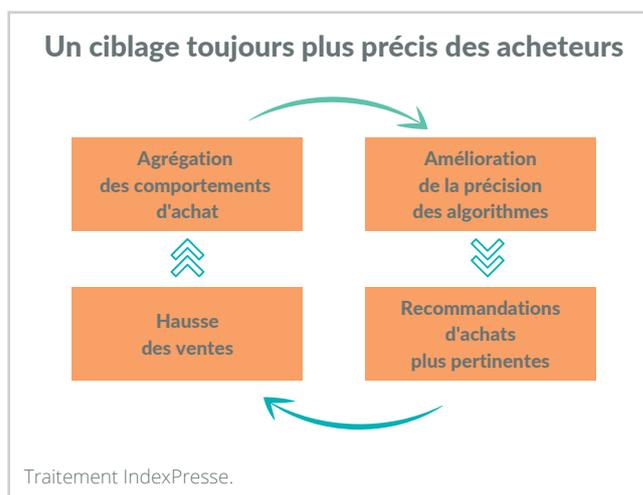
La marge d'erreur en nombre de commandes prévues des algorithmes du Hub Digital Carrefour-Google

## Perfectionner le ciblage personnalisé

La collecte d'un grand nombre de données sur les consommateurs permet d'**affiner les recommandations individuelles** ciblant ces derniers. Ce phénomène s'avère particulièrement marqué dans le e-commerce. En rassemblant les historiques d'achat et les produits consultés, les acteurs du secteur se montrent capables de **repérer automatiquement des comportements récurrents** (achats de plusieurs produits spécifiques en une seule commande par exemple). Ils peuvent ensuite proposer aux internautes **des offres plus personnalisées**. Croisant ces données avec d'autres informations telles que la localisation ou l'âge, les outils de *big data* renforcent d'autant plus la précision du ciblage.

Fondée en 2011, la société Netwave est spécialisée dans **la personnalisation des parcours clients**. Son outil adapte les produits, les publicités et les contenus aux visiteurs des sites marchands, permettant d'accroître les ventes réalisées. L'entreprise met en avant **le caractère omnicanal de sa solution**, utilisable tant par les *call centers* que les logiciels de campagnes e-mailing. Cette dernière réagirait en outre en temps réel aux actions des internautes, optimisant les données disponibles. Netwave a pour clients des acteurs du e-commerce comme Cdiscount, Rue du Commerce ou encore le spécialiste de l'aéronautique Safran.

L'entreprise Target2Sell a quant à elle été créée



en 2012. Elle se positionne également dans le ciblage personnalisé des clients et prospects. Sa solution permet d'**intégrer les différents outils et canaux de vente des utilisateurs** : magasins physiques, sites Internet, logiciel de CRM (relation-client), usages mobiles... Elle intervient tant au niveau des sites marchands que des campagnes marketing. La société insiste en outre sur les nombreux logiciels de partenaires compatibles avec la solution afin d'assurer son implémentation rapide. L'outil a été notamment adopté par Auchan, Sephora et Les Galeries Lafayette.

## Systemes d'information et bases de données : l'efficience par le *big data*

### L'essor du numérique dans la sécurité

Le *big data* accroît les capacités de recherche liées aux bases de données et permet de croiser des informations diversifiées. Dans ce cadre, le secteur de la sécurité développe des solutions permettant de collecter et d'exploiter des volumes importants de données, notamment les fichiers de police et des services de renseignement. Les acteurs opérant dans le domaine s'adressent principalement aux municipalités, la mise en place de ces systèmes s'effectuant au niveau local.

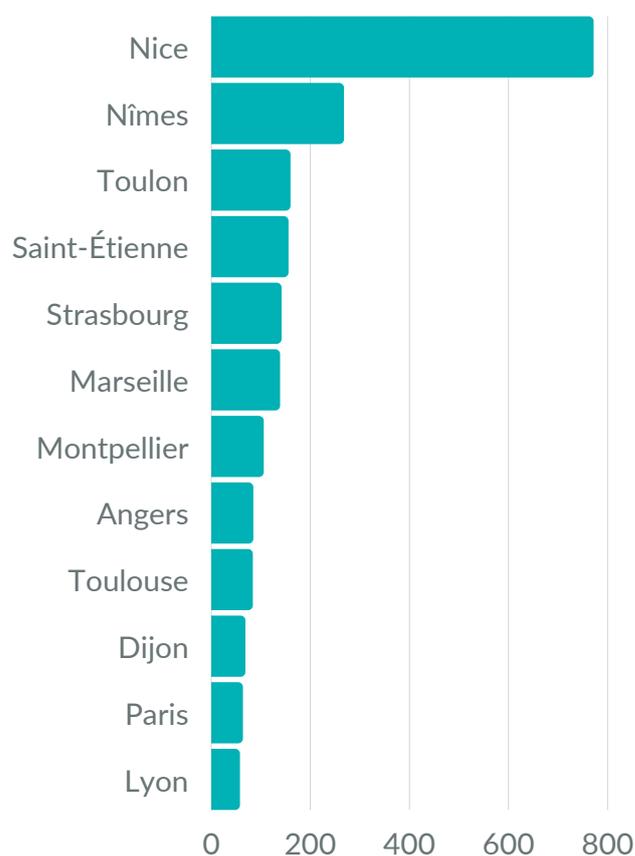
De nombreuses villes en France se sont équipées de matériel de vidéosurveillance, en particulier dans le Sud du pays. Si ce phénomène concerne surtout les plus grands centres urbains, certaines villes de taille moyenne s'avèrent bien dotées en équipements. Mulhouse a ainsi affiché un ratio de 240 caméras pour 100 000 habitants début 2020 selon *La Gazette des communes*, tandis que ce chiffre a atteint 210 pour Annecy et Perpignan. Parmi les agglomérations les plus peuplées, Nice fait figure d'exception : avec un taux de 771 caméras pour 100 000 habitants en 2019, elle se hisse loin devant les autres villes du classement. Ce volontarisme s'accompagne de programmes d'expérimentations de technologies liées à la sécurité comme la reconnaissance faciale, testée lors du carnaval local en 2019.

### La safe city, socle du *big data* dans le domaine sécuritaire

De telles initiatives s'inscrivent dans le concept de "ville sûre", ou *safe city*. Les municipalités multiplient les partenariats : la mairie de Nice s'est associée au spécialiste français de l'électronique de défense Thales en 2018 ; à Marseille, Engie Ineo (filiale de l'énergéticien français) a mis en place depuis 2017 le projet "d'observatoire *big data* de la tranquillité" ; la même année, la ville de

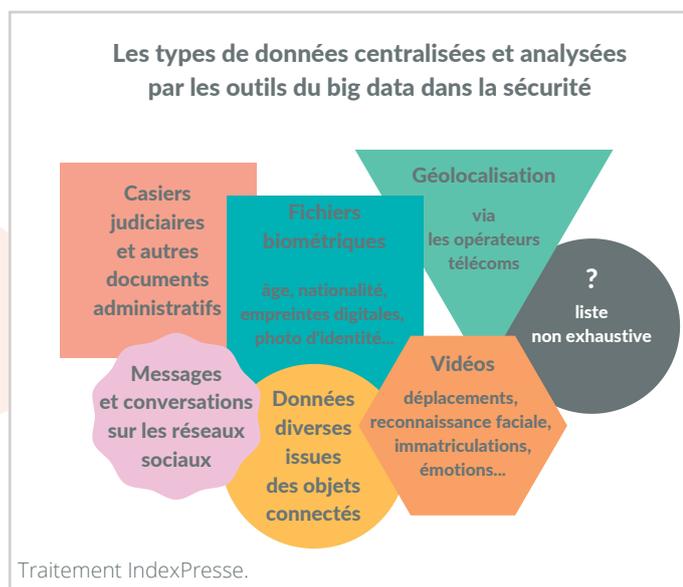
### Nombre de caméras de vidéosurveillance pour 100 000 habitants en 2019

(Les 12 villes françaises de plus de 150 000 habitants avec les ratios les plus élevés)



Traitement IndexPresse. Source : *La Gazette des communes*, Statista

Valenciennes a choisi le géant chinois des communications Huawei pour installer un système de surveillance et d'aide à la décision de police basé sur 360 caméras. Ces projets permettent d'inté-



grer différentes sources d'informations (fichiers des administrations, données de géolocalisation des opérateurs télécom, messages sur les réseaux sociaux...) afin d'identifier rapidement des individus, des comportements suspects, ou encore d'anticiper les menaces potentielles.

À Élancourt (Yvelines), la municipalité a lancé en 2005 un projet de "commissariat du futur", dont l'aboutissement s'avère prévu pour 2024. Mis en oeuvre en collaboration avec le spécialiste canadien des systèmes de gestion vidéo Genetec, le nouvel ensemble disposerait d'un budget de 25 millions d'euros. Cinq millions seraient destinés aux drones et autres technologies de surveillance d'après le journal *Le Parisien*. Le système présenterait diverses fonctionnalités comme la reconnaissance des plaques d'immatriculation des véhicules, la gestion des alarmes ou encore le contrôle d'accès.

Des start-up françaises se positionnent également dans le secteur. Fondée en 2010, Stereolabs s'est d'abord tournée vers le monde de l'audiovisuel (sport et cinéma) en développant une technologie d'analyse vidéo basée sur la 3D. Avec la pandémie de Covid-19 en 2020, l'entreprise s'est diversifié en permettant de détecter les distances trop faibles entre deux personnes dans un magasin.

Créée en 2018, Serenity se positionne pour sa part avant tout dans la cybersécurité. Elle a toutefois noué en 2019 un partenariat avec l'entreprise française Nomosense, spécialisée dans les logiciels d'exploitation des données liées aux objets connectés. Cette collaboration a permis de mettre au point des capteurs sonores pouvant être placés en ville. Ces derniers détectent les sons relatifs à différents types d'accidents et transmettent l'information aux autorités.

### En attente d'un cadre légal mieux défini pour sécuriser le marché

Les acteurs utilisant la *big data* dans leurs solutions de sécurité espèrent la mise en place d'une réglementation permettant le développement de l'activité. La multiplication des tests et des innovations a en effet conduit les pouvoirs publics à interdire ou stopper certaines initiatives, faute de débats suffisants et de lois pour les encadrer. La Commission nationale de l'informatique et des libertés (CNIL) a suspendu l'usage de la reconnaissance faciale dans des lycées du Sud de la France en octobre 2019, ainsi que des détecteurs de sons combinés aux caméras de la ville de Saint-Étienne à la même période. La protection des données personnelles et les risques d'atteintes à la vie privée demeurent les principales contraintes à la collecte massive d'informations dans un but sécuritaire, tant au niveau national qu'europpéen avec le RGPD.

La crise de Covid-19 de mars 2020 a accentué les questionnements, de nouveaux systèmes étant mis en place pour le contrôle sanitaire, comme les caméras thermiques installées en mai 2020 à Menecy, dans l'Essonne. Le Haut Conseil de la santé publique a toutefois émis un avis défavorable à ce type de mesures fin avril 2020. Les villes de Cannes et Nice ont respectivement fait appel aux start-up françaises Datakalab et Two-I pour équiper les systèmes de surveillance dans les transports en commun de logiciels permettant de détecter le port du masque ou le respect des gestes barrières. Pour l'instant, ces dispositifs ne sont pas autorisés à récolter des données biométriques, et analysent les flux de voyageurs grâce à leurs mouvements.

Fondée en 2016, Datakalab est spécialisée dans la reconnaissance faciale et l'analyse des émotions. Sa technologie peut s'appliquer à une grande variété de secteurs, étant initialement utilisée dans le domaine publicitaire. Capable également de mesurer la distance entre les personnes ou encore le temps d'attente dans une file, elle peut servir aux magasins et à tout autre structure accueillant du public.

Two-I a quant à elle été créée en 2017. Sa solution permet l'analyse intelligente des flux vidéo et répond pour l'instant surtout à des applications sécuritaires. "Nous proposons des solutions anti-intrusion capable de différencier une tentative hostile d'un simple animal de passage", déclarait ainsi en 2019 Guillaume Cazenave, le fondateur de l'entreprise.

**Ce développement de la safe city pourrait s'installer dans la durée** de par la nécessité de contrôler les accès aux lieux accueillant du public dans une logique de prévention sanitaire. L'adoption d'une législation moins contraignante pour les acteurs du secteur leur permettrait d'accroître l'extraction de données, d'améliorer leur traitement croisé par leurs algorithmes, et de renforcer la sécurité juridique des expérimentations.

### LES JO DE PARIS EN 2024, POSSIBLE ACCÉLÉRATEUR POUR LA SAFE CITY

L'afflux de visiteurs dans les enceintes sportives et de touristes transitant par les aéroports pourrait dynamiser le déploiement des solutions de sécurité basées sur le *big data*. Certains dispositifs de reconnaissance faciale ont déjà été testés en 2020, notamment dans l'aéroport de Lyon et le stade de football de Metz. Ces derniers recouvrent deux utilisations principales : l'authentification, pour faciliter les déplacements et éviter les usurpations d'identité • l'identification, qui permet de repérer des individus particuliers dans une foule, notamment en vue d'une interpellation et dans la lutte contre le terrorisme.

## Palantir : *big data* et renseignements

Fondée en 2004, cette société américaine se place comme **la spécialiste du big data dédié à la défense et aux services de renseignements**. Initialement financée par la CIA via son fonds In-Q-tel, elle a réalisé son introduction en Bourse en septembre 2020, atteignant **une valorisation de 21 milliards de dollars**. D'après *Challenges*, elle a investi plus de 1,5 milliard de dollars en R&D depuis sa création. Son logiciel Gotham collecte et analyse des informations provenant de centaines de bases de données différentes. Il est utilisé pour traquer des individus ou des activités jugées suspectes ou dangereuses. Outre les agences et administrations américaines (CIA, Pentagone, NSA, FBI...), Palantir a comme

client depuis 2016 les services secrets français, le contrat avec la DGSE ayant été renouvelé en 2019. **L'entreprise s'est également diversifiée vers le secteur privé**, sa solution Foundry permettant d'optimiser la chaîne de production des grands groupes tels qu'Airbus, BP ou Sanofi. Durant la pandémie de Covid-19 de 2020, **la société a proposé ses services aux systèmes de soins** pour améliorer leur réponse à la crise, et opère dans une douzaine de pays (dont les États-Unis et le Royaume-Uni). En 2020, **la moitié du chiffre d'affaires de Palantir serait généré par ses clients non gouvernementaux**, d'après le site SharesPost. Il s'est établi à 730 millions de dollars en 2019, pour 580 millions de pertes. Le fondateur de Palantir, Peter Thiel, investit également dans deux start-up alliant *big data* et sécurité : Clearview AI, capable de reconnaître une personne par analyse de milliers de photos postées sur Internet, et Anduril, spécialisée dans les drones et l'intelligence artificielle militaire.

## Le renforcement algorithmique: le *big data* au service de l'intelligence artificielle

Le développement des solutions d'intelligence artificielle repose fortement sur le *big data* pour **alimenter en données et entraîner les algorithmes**. Par le *machine learning*, les systèmes s'avèrent capables de **se perfectionner en autonomie**. Ce mécanisme nécessite le traitement de volumes considérables de données : visionnage de milliers de photos, enregistrement de grandes quantités de textes ou de contenus audio, collecte massive des comportements de passants ou d'utilisateurs d'un service... Cet apport d'informations permet notamment aux IA de **détecter des récurrences dans les éléments traités**, et ainsi de standardiser leur réponse en cas de similitudes entre deux situations.

Ce phénomène s'illustre dans le cas des **robots conversationnels**, qui regroupent les chatbots, les assistants vocaux, voire les robots physiques munis d'un dispositif de communication. La multiplication des échanges avec des humains leur permet de **rendre leur réponses à la fois plus précises et plus crédibles**. L'enjeu est double : à la pertinence du discours s'ajoute la fluidité de la discussion, nécessaire pour mettre à l'aise l'interlocuteur.

En France, **une variété de start-up se sont positionnées** dans le domaine, le site *MyFrenchStartup* en recensait 54 en 2017. Elles se déploient tant auprès des entreprises qu'en BtoC, une minorité se plaçant sur la fourniture de la technologie plutôt que sur les services associés.

L'entreprise Clac Des Doigts, créée en 2015, commercialise sous forme d'abonnement un assistant virtuel pouvant s'occuper des missions de conciergerie et de diverses commandes. La start-up permet aux organisations de dégager du temps pour leurs employés, qui peuvent déléguer les tâches les plus fastidieuses et chronophages au chatbot.

Fondée en 2017, la société AskMona a développé un chatbot dédié au monde de la culture. L'outil fournit des recommandations au sujet des événements culturels en fonction des préférences personnelles des utilisateurs. Ces dernières sont identifiées par le chatbot au fil de la discussion. La start-up s'adresse aussi aux institutions culturelles en leur proposant des chatbots personnalisés afin de mettre en valeur leurs expositions et d'assurer un service d'information auprès des visiteurs.

La société Dcode (marque ApiBots) a été créée en 2018. Elle développe pour les professionnels des chatbots destinés à l'amélioration et au suivi de la relation-client. Différents services connexes sont également proposés tels que la gestion des newsletters et le *community management*. La start-up insiste sur la facilité d'utilisation de son outil et a obtenu un succès particulier auprès des entreprises de l'immobilier.

Les chatbots s'inscrivent dans une logique plus large liée au **développement de l'interaction entre robots** (machine-to-machine). Les intelligences artificielles seront amenées à coopérer davantage entre elles, **échangeant des informations sans intervention humaine directe**. Un assistant virtuel pourra prendre un rendez-vous auprès d'un secrétaire numérique, des robots se coordonneront pour effectuer une tâche commune dans la défense, le service à la personne ou encore la construction... **Les potentialités s'avèrent multiples** et requièrent une toujours plus grande analyse de données pour **viabiliser l'autonomisation des systèmes basés sur l'IA**, qu'ils agissent seuls ou en groupe.

### ET AUSSI...

#### EXEMPLES DE FOURNISSEURS FRANÇAIS DE CHATBOTS

- DO YOU DREAM UP (2009)
- OVVY (2015)
- BOTDESIGN (2017)
- CLEVY (2017)
- AKV INNOVATIONS (2018)



# La gestion de flottes et de réseaux, guidée par la donnée

## La smart city, ou la ville améliorée par le numérique

Utilisé dans l'espace urbain, le **big data** constitue le socle de la ville intelligente (smart city). La diffusion des objets connectés et la multiplication des capteurs permettent la collecte et le traitement massifs de données dans le but d'**optimiser la gestion de l'agglomération**. Ce pilotage peut s'appliquer à des domaines variés, comme la consommation énergétique ou la circulation des véhicules.

### Le smart grid, une infrastructure augmentée par les données

Le **big data** contribue à une utilisation plus efficace de l'énergie ainsi qu'à la transition écologique. En outre, il accroît la capacité des infrastructures à maintenir un approvisionnement constant, sans coupure de courant. Pour y parvenir, il intervient dans les nouveaux systèmes énergétiques, rassemblés sous le concept de **smart grid**. Ces derniers agissent sur l'ensemble de la chaîne, de la production (centrales, parcs éoliens...) à la consommation (résidentiel, bureaux...) en passant par le réseau de transmission et de distribution. Le **smart grid** repose fortement sur les objets connectés (IoT), les capteurs placés sur les différents éléments du système fournissant les données nécessaires à sa gestion. D'autres technologies soutiennent son déploiement, notamment le développement du stockage de l'énergie, les gains de productivité des énergies renouvelables, ainsi que la digitalisation croissante de l'industrie comme des logements.

Un élément essentiel du **smart grid** réside dans la flexibilité qu'elle offre au réseau. Du côté des consommateurs d'électricité, les terminaux (comme les compteurs Linky en France) permettent de déterminer les besoins, qui peuvent ensuite être modulés si nécessaire pour faire face

## AVANTAGES DE LA SMART GRID

### RÉSILIENCE DE L'INFRASTRUCTURE

#### Un approvisionnement sécurisé



La collecte de données sur l'ensemble du système énergétique et la meilleure coordination entre les différentes sources de production permet de limiter les ruptures d'approvisionnement.

### OPTIMISATION DES FLUX

#### Un meilleur équilibrage entre production et consommation



La connaissance plus fine des variations de la consommation assure une plus grande adaptabilité au réseau. Les nouveaux systèmes permettent de moduler tant les besoins finaux que la quantité d'énergie injectée en cas de pic de demande.

### INTÉGRATION DES ÉNERGIES RENOUVELABLES

#### Faciliter la transition énergétique



Les capacités de stockage et la gestion décentralisée d'unités de production multiples limitent le caractère intermittent des énergies renouvelables, garantissant une stabilité accrue favorable à leur utilisation.

### IMPLICATION DU CONSOMMATEUR

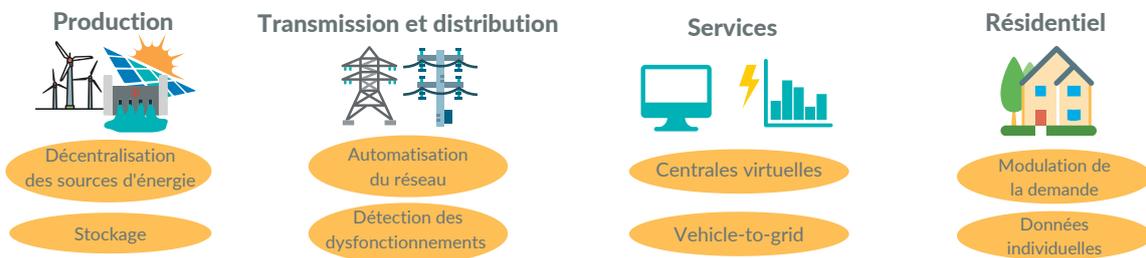
#### Vers un rôle plus actif dans la régulation du réseau



Les ménages peuvent mieux mesurer et piloter leur consommation d'énergie. Ils deviennent également des unités de production grâce aux panneaux solaires, et de stockage via leurs véhicules électriques.

Traitement IndexPresso.

### Interventions de la smart grid sur l'ensemble de la chaîne de l'électricité



Liste non exhaustive. Traitement IndexPresse.

aux pics de demande. La mise en place d'incitations (comme la variation du prix de l'électricité) peut encourager à décaler certaines consommations d'énergie tandis que d'autres peuvent être étalées dans le temps : **ralentir le rechargement** des véhicules électriques peut par exemple diminuer la pression sur le réseau.

**De nombreux acteurs se sont lancés sur le marché**, à l'instar de Smart Grid Energy. Fondée en 2011, la société est spécialiste de la gestion de capacités électriques décentralisées et de l'équilibrage du réseau en cas de risque de rupture. Elle a été acquise en 2016 par Vinci Energies. L'entreprise Voltalis, créée en 2006, place des boîtiers chez les particuliers permettant de réduire la consommation de certains appareils lorsque le réseau est en tension. Elle affirme exécuter l'opération sans pénaliser les ménages, leur faisant même réaliser des économies sur leur facture. Fondée en 2011, la société française EMBIX conçoit des projets de *smart grid* pour les villes et les assiste dans leur mise en place. Elle met également à leur disposition un outil logiciel de gestion des systèmes électriques intelligents, et s'occupe du pilotage opérationnel des solutions implémentées. L'entreprise a été impliquée dans le programme IssyGrid, mené entre 2012 et 2018 par plusieurs acteurs privés dont Bouygues, Total, Sopra Steria, Microsoft ou encore Schneider Electric. Les organisations publiques EDF et Enedis faisaient également partie du consortium. Réalisé en collaboration avec la ville d'Issy-les-Moulineaux (Hauts-de-Seine), le projet a permis de mesurer la consommation des quartiers concernés, soit d'environ 5 000 habitants et 10 000 employés de bureau selon les décomptes indiqués sur le site [issy.com](http://issy.com).

D'autres initiatives du même type ont été mises en place telles que Smile en Bretagne et Nice Smart Valley en région PACA.

### La data pour améliorer la mobilité urbaine

La **smart city** utilise également le **big data** pour fluidifier le trafic, facilitant les déplacements tout en réduisant la pollution générée par les congestions. Elle parie notamment sur la **multi-modalité** et l'**optimisation des trajets**. L'agrégation de données de différentes sources s'avère nécessaire : géolocalisation des automobilistes et des citoyens utilisant les flottes de véhicules partagés, flux de voyageurs dans les transports en commun, feux de signalisation ou encore places de parking disponibles. **L'espace urbain connecté recouvre plusieurs types de flux de données**, tels que le V2I (vehicle-to-infrastructure) et le V2V (vehicle-to-vehicle). Cette multiplicité implique des capteurs spécifiques dans les différents terminaux, des applications smartphones aux systèmes embarqués dans les voitures. L'intégration et le traitement de ces informations par les outils du **big data** permet d'**analyser la circulation en temps réel** afin de répondre tant aux besoins de mobilité qu'aux problématiques liées à la qualité de l'air en ville.

Division regroupant neuf entités du groupe français Bolloré, Blue Systems se spécialise dans les questions de gestion des flux de personnes et de marchandises. Elle a développé Smart Mobility manager, une solution centralisée pour superviser le trafic de voyageurs, le stationnement et les patrouilles de police.

Filiale du groupe de transport Transdev, la société Cityway, créée en 1999, se présente comme pionnière dans la "*Mobility as a Service*" (MaaS). La société a mis au point des solutions logicielles pour favoriser la réduction des embouteillages et de la pollution. Elle a également fondé en 2019 Le Hangar, un incubateur dédié aux start-up de la mobilité comme Visitmoov, axée sur l'optimisation des voyages touristiques.

# La maintenance prédictive optimise la disponibilité des capacités opérationnelles grâce aux algorithmes

## Disposer de sa flotte de véhicules en permanence

Fournies par des capteurs embarqués, les données des véhicules connectés peuvent servir aux logiciels de *big data* pour **prévoir les futures réparations**.

**Mobilisée pour des opérations extérieures de longue durée** (en particulier sur le continent africain), **l'armée française a concentré ses efforts** sur les équipements dédiés à ces interventions. La Ministre des Armées, Florence Parly, a ainsi déclaré en 2018 que "si le taux de disponibilité de nos équipements terrestres est excellent en opération, plus de 90 %, il est plus mitigé ici, en France". **Peu de véhicules de l'armée de Terre atteignent le taux de 70 % de disponibilité** début 2020, seuil nécessaire pour remplir ses engagements sur le sol national, d'après le site *opex360.com*. Les véhicules blindés de combat d'infanterie (VBCI) affichaient ainsi un taux de 64 % à cette période.

Depuis 2016, la réforme du Maintien en condition opérationnelle terrestre (MCO-T) puis **l'utilisation de la solution HUMS (Health and Usage Monitoring System)** de l'entreprise française de défense Arqus a permis d'équiper les blindés de catégories Jaguar, Griffon et Serval **de capteurs servant à la maintenance prédictive**. Le volume d'équipements en immobilisation technique a diminué de 3 % entre 2018 et 2019, d'après *Terre Info Magazine*. L'objectif serait à présent d'atteindre un taux de disponibilité de 80 % de ces véhicules. "L'immobilisation d'un matériel terrestre doit devenir l'exception" grâce à ces nouvelles technologies, selon *Terre Info Magazine*.

## Garantir une production constante dans l'industrie

Aussi appelée *smart factory* ou industrie 4.0, **l'intégration de systèmes cyber-physiques dans les**

**usines** constitue une innovation majeure pour augmenter la productivité industrielle. Dans ce cadre, **la maintenance prédictive représente un axe majeur pour assurer un flux de production continu**. "Une minute d'arrêt machine non programmée peut engendrer des coûts pouvant aller jusqu'à plusieurs dizaines de milliers d'euros", selon Peter Livaudais de l'entreprise britannique de maintenance prédictive Senseye, dans la revue *Maintenance & Entreprises* de novembre 2019.

Afin d'assurer un fonctionnement constant de ses usines et d'éviter des interventions complexes, **le groupe métallurgiste français Eramet** a mis en place en 2020 un projet de maintenance prédictive. Initialement lancé dans son site de production de ferronickel de Nouvelle-Calédonie, il a été étendu à d'autres lieux en France et à l'étranger, notamment au Gabon et en Norvège. Chief Digital Officer du groupe, Ludovic Donati estimait ainsi fin 2020 que "les premiers résultats sont visibles sur la productivité de nos mines et de nos usines". Le projet a été conduit **en partenariat avec la start-up française Moxino**, spécialiste de la maintenance prédictive industrielle. Fondée en 2013, elle a développé un outil basé sur les objets connectés pour optimiser la disponibilité des équipements industriels. Eramet souhaite aller plus loin en détectant les risques d'incidents sur ses différents sites grâce au croisement de diverses données, dont celles provenant de drones exploités par la société française Delair.

De son côté, la SNCF a lancé début 2021 un projet basé sur la maintenance prédictive avec Engie Solutions et le cabinet spécialiste du numérique Wavestone. Baptisé "Smart Station", il vise à réduire la consommation énergétique des gares en signalant plus tôt les dysfonctionnements de certains dispositifs. Les équipements de la gare seront également entretenus de façon anticipée pour garantir une meilleure fluidité du trafic des voyageurs. L'ensemble du programme devrait être mis en place pour les Jeux olympiques de 2024 à Paris.

## Générer des rendements supérieurs grâce aux données

Dans certains secteurs, le **big data** permet **directement d'accroître la production**. C'est par exemple le cas dans l'agriculture avec le concept de *smart farming*. En multipliant la collecte et les sources de données, les agriculteurs peuvent analyser de manière précise l'état de leurs récoltes, et ainsi en optimiser le rendement. De nouvelles solutions logicielles croisent les différentes informations dans une logique d'**aide à la décision des exploitants**. La diminution de la consommation d'eau pourrait atteindre 15 % à 20 % selon une évaluation du site *Terre-net.fr* en 2017, tandis que l'utilisation des engrais serait réduite de 30 % à 40 % d'après *bigdata.fr* (2018).

La start-up Live NBalance, fondée en 2019 par l'entreprise américaine John Deere et Airbus Defence and Space, suit l'évolution de la consommation d'azote des parcelles en s'appuyant sur des capteurs au niveau des machines agricoles et sur des images satellitaires. La solution permet de détecter des quantités absorbées anormales et de mesurer la productivité de ces apports en termes de croissance des récoltes.

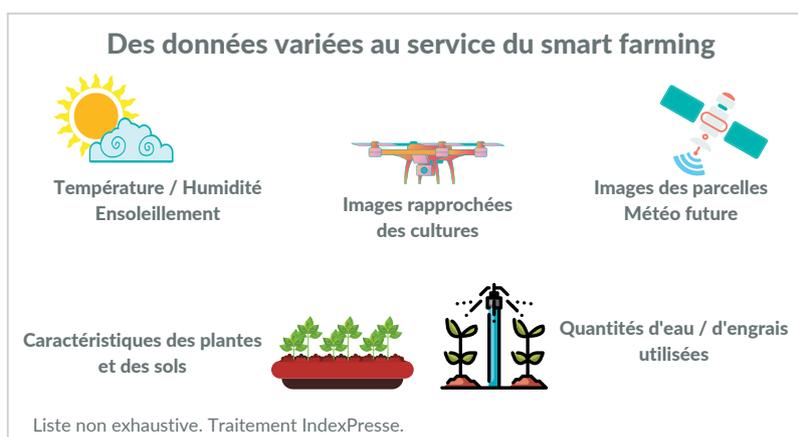
Créée en 2014 et soutenue par l'Inra, l'Institut du végétal et l'Allemand Bosch, la société Hiphen a mis au point en 2019 le système Field Sensor, basé notamment sur des capteurs placés au niveau du sol. Les données collectées sur les plantes et la météo sont croisées avec celles recueillies par drone ou par satellite afin d'évaluer le

développement des cultures dans le temps. Des algorithmes permettent de détecter les maladies ou d'éventuelles fragilités (feuilles nécrosées).

Fondée en 2014, la start-up Weenat a développé un dispositif de capteurs relié à un logiciel commercialisé en SaaS. Il permet aux agriculteurs d'effectuer et de centraliser de nombreuses mesures telles que l'humidité de l'air, les températures ou encore la pluviométrie. Ces dernières peuvent ensuite être analysées par l'outil pour **assurer une meilleure gestion des parcelles**. Fin 2019, la jeune société a levé 2 millions d'euros pour soutenir son développement à l'international, et enregistrait jusqu'alors un taux de réabonnement de 98 %.

L'entreprise Carré, créée en 1938, est spécialiste du matériel agricole. Elle a développé en 2019 un véhicule électrique autonome de désherbage des sols, récoltant différentes données durant cet activité. Celui-ci transfère sur tablette ou smartphone la température au sol, la luminosité, l'hygrométrie ou encore des informations relatives aux plantes cultivées.

La **maîtrise des données** générées sur les parcelles **constitue un enjeu majeur pour le monde agricole**. En France, le domaine se structure avec la création en 2017 de l'Institut Convergence Agriculture Numérique, ou #DigitAg.



## Mieux détecter la fraude avec le *big data*

### Lutter contre la contrefaçon

L'accumulation de données et leur traitement de plus en plus pointu et minutieux permettent de **repérer les manœuvres frauduleuses avec une plus grande acuité**. Le secteur du luxe s'avère particulièrement confronté à de tels agissements, la contrefaçon constituant **un phénomène affectant tant le chiffre d'affaires que l'image des marques**. En France, les enseignes du vêtements, de la chaussure et des accessoires auraient subi des pertes de 3,6 milliards d'euros en 2019 du fait de ces fraudes, d'après un rapport de l'Office de l'Union Européenne pour la propriété intellectuelle (EUIPO). **Des solutions basées sur le *big data* ont été développées par des start-up** pour répondre aux besoins des grandes maisons de luxe :

- depuis 2011, l'entreprise Webdrone dispose **d'outils de recherche automatisés** sur Internet pour détecter les ventes de produits contrefaits. Une investigation est ensuite menée par des enquêteurs humains pour déterminer l'identité des trafiquants ;
- fondée en 2012, la société américaine Entrupy propose **un dispositif de visualisation de précision**, doté d'une intelligence artificielle. Entraîné grâce aux algorithmes à **analyser les formes et les matériaux d'un vêtement** ou d'un sac à main par exemple, il peut identifier les produits fabriqués avec des matières premières non conformes (notamment de qualité inférieure) ;
- créée en 2013, la start-up française Data&Data, spécialiste du traitement massif de données, **passé en revue les sites de vente en ligne, les places de marchés virtuelles ou encore les réseaux sociaux** pour traquer les contrefaçons et autres activités frauduleuses pouvant attenter à l'entreprise cliente.

### La finance contre les activités illicites

Dans le secteur financier, **la surveillance des flux financiers suspects** constitue un élément incontournable dans la lutte contre les activités criminelles. Les banques s'avèrent tenues de contrôler les virements effectués par leur intermédiaire, et de transmettre les dossiers sujets à caution aux autorités (Tracfin). La société de conseil et d'audit Mazars propose ainsi des solutions contre le blanchiment en s'appuyant sur l'expertise de Panatio, entreprise allemande spécialisée dans l'analyse *big data*.

### Les assurances face aux fraudeurs

Impactés par les déclarations trompeuses et les accidents intentionnels, les acteurs de l'assurance voient dans le *big data* une technologie contribuant à **mieux évaluer l'authenticité des sinistres**. Fondée en 2013, l'entreprise française Shift Technology a développé un algorithme croisant de nombreuses données (polices d'assurance, déclarations, éléments extérieurs...) pour identifier les fraudes, en particulier pour les sinistres liés aux voitures et aux habitations. La start-up avait levé 1,4 million d'euros fin 2014, et commercialise son logiciel sous le modèle SaaS.

Fin 2019, l'Agence française de lutte contre la fraude (Alfa) a mis en place avec la Fédération française de l'assurance (FFA) **un système décryptant les millions de déclarations relatives aux accidents de voiture**. Ce dernier peut permettre d'établir des récurrences dans les tentatives visant à tromper les assurances, et ainsi les repérer de manière automatisée.

**3,6 milliards d'euros**

Les pertes causées par la contrefaçon au secteur textile et maroquinerie en France en 2019

Source: EUIPO, 2019.

# UN FUTUR PROMETTEUR GUIDÉ PAR DES SECTEURS PORTEURS

## Un potentiel élevé dans l'assurance

### Personnaliser davantage les contrats

Dans le domaine assurantiel, le **big data** peut **comporter des avantages en termes de ciblage** des clients. Une utilisation plus poussée de telles solutions permettrait **d'évaluer plus précisément les risques au niveau individuel**, et ainsi de pouvoir moduler les primes associées.

La classification actuelle des risques en grandes catégories homogènes (âge, profession, état de santé...) alloue des probabilités à chacune d'elles. Intégrant cette typologie, les individus s'avèrent associés à ces statistiques en fonction de leurs paramètres personnels. **L'attribution de la prime de risque** pour chaque client **se révèle donc indirecte** : c'est la précision de la catégorisation qui la définit et non des critères calculés individuellement. **Cette relative standardisation constitue le socle de la mutualisation**, les personnes d'une même catégorie finançant celles connaissant un sinistre.

Outre l'amélioration des modèles probabilistes pour chaque type de profils, le **big data** pourrait permettre de dépasser le cadre actuel pour aller vers **une individualisation plus importante des contrats d'assurance**. L'analyse des données de consommation, de santé, liées à l'exercice physique ou encore à la conduite automobile grâce aux objets connectés représenterait pour le secteur un potentiel marketing considérable. **Des changements interviendraient dans la nature même des contrats** : à un prix fixé lors de la souscription se substituerait **une variabilité du coût** liée aux nouvelles informations recueillies.

### L'usage du **big data** se développe dans le secteur

Les acteurs adoptent progressivement cette nouvelle technologie. Anthony Jabre, responsable maîtrise indemnisation à **Malakoff Humanis**, a expliqué début 2021 "[qu']en prévoyance, en fonction des caractéristiques du sinistre, **un algorithme permet d'aider les experts** à décider de la suite à donner au dossier, allant jusqu'à déclencher un contrôle."

Fondée en 2012, la start-up française Apidata est spécialisée dans l'assurance de personnes, intervenant principalement dans la santé et la prévoyance. Elle a développé à destination des assureurs un outil de gestion leur permettant d'avoir **une meilleure vision sur leurs données, les rendant plus fiables et davantage exploitables**. Xavier Garcia, CEO d'Apidata, affirmait à *Forbes* fin 2018 que "l'ennemi de la **data**, c'est la main humaine. À chaque fois qu'une information est saisie manuellement la qualité de la donnée est altérable, car le potentiel d'erreur de saisie est important". Il ajoutait que sa solution permettait de "traiter la **data** par des robots [...] en excluant les tâches manuelles, qui sont réservées à l'analyse."

L'entreprise française Saagie a quant à elle été créée en 2013. Elle est spécialisée dans la mise en place de projets **big data**, **facilitant l'extraction et l'exploitation des données de ses clients**. Lors d'une levée de fonds de 4,2 millions d'euros en 2016, le groupe mutualiste Matmut a fait son entrée au capital de la start-up, considérant que "le **big data** est un enjeu central pour l'avenir, susceptible de modifier notre proposition de valeur et le business model de notre activité". Saagie a levé 25 millions d'euros en juin 2020, notamment auprès de compagnies d'assurance telles que Crédit Mutuel Innovation et AG2R La Mondiale.



### Des difficultés à lever

La personnalisation plus poussée des contrats permise par le **big data** nécessite toutefois un **accès renforcé aux données** des individus, tant au niveau de leur consommation, de leur mobilité qu'au sujet de leur santé. Or, les différentes législations telles que le RGPD en Europe restreignent ce type d'utilisations. **Le poids croissant de l'économie numérique tend à renforcer la mise en place de régulations**, avec par exemple le Digital Market Act et le Digital Services Act dans le cadre de l'Union européenne.

Dans le cas où les assureurs parviendraient à faire lever les obstacles réglementaires les plus importants, **il n'est pas certain qu'ils puissent facilement accéder aux données personnelles** de

leurs clients. La valeur supplémentaire prise par ces dernières pourraient **inciter les acteurs les collectant à davantage les conserver** pour en tirer un profit supérieur. Le réseau social Facebook avait déjà freiné les compagnies d'assurance britanniques en 2016 dans leur volonté d'acquérir des données de ses utilisateurs.

**L'individualisation croissante des contrats** associée à un calcul du risque de plus en plus précis pourrait également conduire à la **remise en cause de la mutualisation**. Générateur de profits pour les assureurs, ce changement amènerait à augmenter les frais payés par les personnes les plus à risque, ce qui pourrait entraîner **des contestations sur le plan social et politique**.

## De nombreuses perspectives dans la santé

### Un usage en essor

Le domaine médical peut **considérablement bénéficier du big data**. Accroître son usage permettrait de réaliser des analyses sur un grand nombre de personnes, **facilitant la recherche** et le développement de nouveaux traitements. **L'aide au diagnostic et la médecine prédictive** (notamment pour le dépistage) s'avèreraient également enrichies par cette technologie, qui profitera en outre du **développement de la télémédecine**.

Encore relativement peu répandu dans le secteur, **le big data nécessite un plus grand accès aux données** de santé. Ces dernières peuvent être individuelles, collectées par exemple via les objets connectés. **Agrégées et anonymisées, elles présentent moins de risques** en termes de respect de la vie privée. C'est cette seconde voie qui a été prise en France, avec **l'ouverture progressive des données de santé (open data) depuis 2013**, la création du dossier médical partagé en 2015, et celle du Système National des Données de Santé (SNDS) en 2016.

Le sujet du **big data** dans le domaine médical s'avère également pris au sérieux en Wallonie (Belgique), où **un institut de recherche dédié**

(Inha, Institute of Analytics for Health) **a ouvert ses portes en 2018**. Collectant les informations auprès des hôpitaux et des professionnels de la médecine générale, il se montre accessible aux laboratoires et aux entreprises de l'industrie pharmaceutiques. Ces organisations peuvent notamment **rechercher plus facilement certains profils de patients** spécifiques à une étude particulière. Toutefois, **les données sont conservées au sein de l'Institut** : dans un souci de confidentialité, ce dernier transmet uniquement les résultats des demandes et non les chiffres bruts.

### Un défi tant technique que méthodologique

**Une difficulté** concernant l'exploitation des données de santé **consiste en leur structuration**. Interviewé en octobre 2019 dans la revue *Pharmaceutiques*, Xose Fernandez, chief data officer à l'Institut Curie, a déclaré que les informations "produites par l'homme (...) représentent 80 % des données que l'on possède, et (...) ne sont pas structurées". Sous forme de textes rédigées à la main, elles peuvent se révéler **complexes à lire par les outils du big data**. Issues de la vie quo-

tidienne comme d'études cliniques, les données doivent être analysées avec minutie afin d'**éviter les erreurs d'interprétations**. Comme l'a indiqué dans la même revue Meriem Sefta, responsable des partenariats de l'entreprise française Owkin, spécialisée dans l'IA médicale : "Les données de vie réelle présentent l'avantage d'être riches en volume mais sont potentiellement intrinsèquement porteuses de biais de risque méthodologique, contrairement à des données d'essais cliniques par exemple". **Les informations médicales collectées proviennent en outre de multiples sources**, telles que les hôpitaux, la médecine de ville, les entreprises, les administrations ou les patients eux-mêmes. **L'intégration de ces données dans des systèmes uniques s'avère difficile**. D'une part, la centralisation, nécessaire à l'analyse, demande une standardisation des éléments recueillis. D'autre part, leur diversité doit apparaître afin de qualifier le contexte dans lequel ils ont été collectés : le degré de confiance dans les données, leur sens et les conclusions des professionnels de santé pourront différer selon la nature des sources utilisées.

### Un écosystème en voie de structuration

Malgré ces freins au développement du *big data* dans la santé, **les innovations se multiplient**. **Souvent portés par des partenariats** entre des start-up et des groupes pharmaceutiques ou des centres de recherche publics, les projets mettent la donnée au service d'algorithmes prédictifs :

- fondée en 2013, la start-up française Therapixel a développé **une solution d'imagerie pour la détection du cancer du sein**. Basée sur l'analyse de centaines de milliers de mammographies, l'outil a permis de diviser par cinq le nombre de faux positifs dans les services de l'Hôpital Cochin, à Paris, selon Olivier Clatz, le dirigeant de l'entreprise ;
- mi-2018, l'Institut du cerveau et de la moëlle épinière (ICM) et l'Institut national de recherche en informatique et en automatique (Inria) ont noué un partenariat pour **développer des algorithmes au sujet de la maladie de Parkinson**. Ces derniers permettent de créer **des modélisations neuronales retraçant l'évolution de la pathologie** chez les patients étudiés, contribuant à sa compréhension et à la personnalisation de la prise en charge ;
- créée en 2016, la start-up Owkin travaille depuis 2018 avec l'Institut Curie pour exploiter ses bases de données afin d'**améliorer la détection et le traitement des cancers**. La société mène également le projet Substra, regroupant de nombreux hôpitaux et centres de recherche français. Lancé en 2018, il vise à mutualiser les efforts dans le développement d'algorithmes pour le secteur médical ;
- fondée en 2015, la société OpenHealth Company rassemble les données de ventes de médicaments des pharmacies françaises. Elle reçoit également des informations relatives aux prescriptions, fournies par les médecins généralistes. Ces données permettent un suivi de la consommation, ce qui peut alerter sur le développement d'une maladie et faciliter la veille concurrentielle des laboratoires. L'entreprise commercialise aussi une solution pour analyser les comportements des patients. Elle a levé 5 millions d'euros à deux reprises, en 2017 et en 2019 ;
- la start-up française Cardiologs, créée en 2014, a développé **une solution facilitant la lecture des électrocardiogrammes**. Elle permet notamment d'accélérer la visualisation des images suggérant le risque d'accident vasculaire cérébral (AVC). Elle était l'une des lauréates du Concours Mondial d'Innovation 2030 mis en place par l'État en mars 2014 ;
- la société française Ad Scientam, fondée en 2013, a multiplié les partenariats pour créer des **dispositifs médicaux sur smartphone** : en 2019 avec Johnson & Johnson sur **la réaction aux anti-dépresseurs**, et avec l'Américain Vertex Pharmaceuticals concernant **la mucoviscidose** ;
- **Median Technologies**, société française créée en 2002, est spécialisée dans la détection des tumeurs grâce à l'amélioration et à la standardisation de la lecture des images médicales par des algorithmes.

# Un gisement encore peu exploité dans les ressources humaines

L'utilisation du *big data* reste peu développée au sein des directions des ressources humaines des entreprises comme des administrations. Elle pourrait pourtant **améliorer de nombreux aspects relatifs à la gestion du personnel** : recrutement plus pertinent, meilleur ciblage des formations proposées aux collaborateurs, répartition mieux planifiée de la charge de travail... Le *big data* peut aussi constituer **un outil d'évaluation du climat social** (absentéisme, risque de démissions, motivations des collaborateurs) et **renforcer l'efficacité des solutions de recherche d'emploi**, privées comme publiques.

Peu de projets ont vu le jour dans le domaine, cependant **quelques entreprises se sont lancées dans cette voie**. La compagnie d'assurance française Axa a mis au point en juin 2019 **un algorithme servant à accroître la précision des suggestions de formations** pour ses salariés. Combinant la catégorie de la fonction exercée, les notes attribuées par les employés aux différents contenus, ainsi que les centres d'intérêts déclarés par chacun, **la solution a fait augmenter la fréquentation du site de 50 %** par rapport à 2018. **Le nombre de formations suivies a doublé**, tandis que celles dispensées uniquement en ligne ont connu une hausse de 75 %. Responsable de la formation et du développement chez l'assureur, Damien Bourgeois a expliqué fin 2019 dans la revue *Entreprise & Carrières* qu'il est "possible de croiser les données. Si par exemple des gens investissent du temps de consultation sur certains types de formations ou sur un sujet particulier, pas forcément lié à leur métier, il est possible d'en déduire une appétence. Nous ne le faisons pas pour l'instant mais **il est envisageable de détecter des centres d'intérêt complémentaires** par rapport à ceux qui sont aujourd'hui collectés."

Fondée en 1994, la société Horoquartz se présente comme un spécialiste de la gestion des plannings et du personnel. Elle a développé entre 2019 et 2020 **une solution permettant d'automatiser la validation des absences** en calculant la charge de travail supplémentaire à répar-

tir sur les autres collaborateurs. L'entreprise vise également à **renforcer la précision des systèmes de planification des tâches entre les équipes**, en intégrant des éléments prédictifs dans leurs modèles. En déterminant la probabilité d'événements tels que les embouteillages ou les intempéries, la disponibilité réelle des employés pourrait être mieux anticipée.

Le groupe français de services administratifs ADP commercialise en France depuis janvier 2020 **une solution de comparaison de rémunérations** baptisée Compensation Benchmark. Basée alors sur un million de fiches de paie, elle fournit aux entreprises **la possibilité de connaître avec précision le niveau de salaire de leurs employés par rapport aux standards du marché**. L'outil offre divers critères de segmentation, du métier à la zone géographique en passant par la taille de la société, afin d'affiner la recherche. Cette solution permet notamment **d'accroître la rétention des compétences** en évitant des départs dus à des rémunérations inférieures à celles proposées par la concurrence.

Malgré ces avancées dans l'usage du *big data* pour les ressources humaines, **des freins à son développement demeurent**. **L'une des difficultés réside dans le partage des données**, même agrégées et anonymisées. Les entreprises s'avèrent réticentes à les diffuser, un phénomène s'expliquant par leur organisation interne selon Thierry Bobineau, directeur marketing chez Horoquartz : "Dans une entreprise, personne ne vous reprochera de ne pas avoir communiqué de données à l'extérieur." Par ailleurs, **les directions RH ne se révèlent pas toujours suffisamment formées aux enjeux du *big data***, tandis que l'intégration de spécialistes des données dans ces départements ne constitue pas forcément une réponse pertinente pour entraîner les équipes dans le projet. **Les craintes concernant le remplacement de certaines fonctions RH par les algorithmes** retardent également l'implémentation de telles solutions.

# La transformation du métier d'archiviste par le *big data*

## Vers un rôle plus centré sur la gestion des données

Peu investi par les solutions d'analyse massive de données, **l'archivage pourrait constituer un domaine émergent** pour le secteur. L'utilisation du *big data* présente des avantages pour les archivistes, en particulier concernant **la mise en valeur des grandes quantités de documents conservés**. Les capacités de traitement macroscopique permises par ces outils faciliteraient par exemple les travaux des historiens, et rendraient possible des approches de recherche novatrices.

Dans ce cadre, **le rôle de l'archiviste évolue** : à la connaissance de la base de données pour trouver le document demandé s'ajoute la gestion de son organisation afin de pouvoir générer des analyses globales. Cité par la revue *Archimag* en octobre 2020 pour son mémoire de fin d'études consacré au *big data* dans l'archivage, Alexandre Vieira décrypte les changements à l'oeuvre : “[*le big data*] **nécessite que les archives commencent à créer des index et des tables** (...) pour permettre ces analyses (...) et non plus seulement pour trouver les documents unitaires”. Il ajoute que **l'archiviste “doit organiser, structurer” la donnée** afin qu'elle “puisse apporter une valeur significative (...) pour la recherche, mais aussi pour toute organisation, privée ou publique”.

## Des projets se sont mis en place

La réflexion sur le sujet a démarré en 2010 avec **la création d'un groupe de travail au sein de l'Association des archivistes français (AAF)**. Ce dernier a été poursuivi par **une mission d'information au Sénat** en 2014, réalisée par une délégation de l'AAF. Depuis, quelques organisations, surtout publiques, ont mis en oeuvre des plans d'archivage basés sur le *big data*.

Développé entre 2014 et 2017, **le projet européen eArchiving Building Block** (initialement E-Ark) a rassemblé des institutions et des entre-

prises de plusieurs pays afin de fournir **une solution d'archivage de long terme pour des données de toute nature** et issues des sources les plus variées. Il s'adresse aux divers producteurs d'informations, aux services d'archives et aux développeurs de logiciels et de services souhaitant garantir dans le temps la sécurité et l'intégrité des données.

En France, la Société Nationale des Chemins de Fer Français (SNCF) a lancé en 2017 deux portails permettant d'accéder à ses bases de données : *Data.sncf* et *Openarchives.sncf*. Le premier agrègeait 221 ensembles de données diverses fin 2020. Ces dernières concernent la fréquentation des gares ou encore les objets trouvés. La seconde base s'avère dédiée aux documents iconographiques tels que les films, les photographies et les dessins techniques pour la construction des infrastructures.

Fin 2018, **les Archives nationales ont organisé un hackathon** (événement de développement numérique) : les participants devaient proposer des pistes d'amélioration et créer des solutions plus abouties pour **mettre en valeur les grandes bases de données de l'institution**. Celles-ci regroupent des documents variés dont l'ancienneté s'étale sur quatorze siècles. L'événement s'est structuré autour de thématiques comme **l'analyse des archives notariales de Paris ou rendre plus accessible l'historique de la Légion d'Honneur**. Interrogée en mars 2019 par *Archimag* quand elle était directrice des Archives nationales, Françoise Banat-Berger a affirmé que le hackaton était “un très beau succès en termes de participants, de qualité des projets présentés et (...) s'inscrit dans une démarche de transformation des Archives nationales”.

Ces différents projets ont amené les archivistes à **travailler avec d'autres acteurs**, comme les développeurs informatiques et les juristes. Ils peuvent ainsi **mettre en avant leur expertise** des documents archivés au sein de cette collaboration.

# Le véhicule autonome, candidat à un usage extensif de la donnée

Réalisée surtout par des grands groupes comme Google et Tesla, ou Valéo en France, **la recherche sur les voitures autonomes n'en est encore qu'à ses débuts**. Il faudra ainsi "des décennies avant qu'[elles] ne sillonnent les routes", a déclaré fin 2020 John Krafcik, responsable de la filiale de Google Waymo, qui développe cette innovation. Malgré la complexité des technologies à maîtriser, **ces véhicules pourraient à terme trouver leur place dans de nombreux secteurs**, dans les villes, sur les sites industriels ou encore au niveau des plateformes logistiques. Marc Le Flohic, président du spécialiste français du laser Lumibird, reste optimiste sur **l'essor futur de ce type de véhicules** : "On sait que la demande en lidars [technologie de détection et de calcul de la distance basée sur le laser, utilisée notamment dans les systèmes embarqués d'aide à la conduite] pour la voiture autonome explosera, on ne sait pas quand".

**Le développement de tests et d'innovations** dans le domaine progresse toutefois en France.

Fondé en 1923, l'équipementier automobile français Valéo place **l'autonomie comme l'un de ses axes de développement** de la mobilité du futur. Il a mis au point en 2015 un prototype de voiture autonome baptisé Cruise4U, suivi par un second en 2018, le modèle Drive4U. Ces derniers ont été testés en conditions réelles, notamment sur le périphérique parisien. Début 2020, le groupe a révélé le eDeliver4U, un véhicule utilitaire autonome développé en partenariat avec le Chinois Meituan Dianping, spécialiste du e-commerce de services.

En France, la société Navya, fondée en 2014 et spécialisée dans la mobilité autonome, propose **des navettes sans chauffeur et des véhicules pour les aéroports et l'industrie**. Elle a par exemple mis au point un tracteur pour déplacer

les bagages, opérant sur la zone aéroportuaire de Toulouse-Blagnac depuis décembre 2019. D'autres sites se révèlent concernés par ses solutions, tels que les usines de fabrication automobile, les mines et les entrepôts. L'entreprise a également testé en novembre 2020 **la livraison de colis sur un campus universitaire** de Corée du Sud, en partenariat avec Korea Post. Utilisant les véhicules de Navya pour le transport de personnes, le groupe français Bertolami a annoncé à la même période la réalisation d'essais en 2021 pour **la livraison en zone rurale**. Citant en février 2021 Benjamin Baudet, directeur de l'entreprise, la revue *L'Officiel des transporteurs* indique qu'un nouveau test d'une solution autonome pourrait advenir durant l'année, et mêlerait "transport de passagers et de marchandises". La société a déjà effectué une expérimentation concernant une navette sans conducteur à Val Thorens en 2019.

Le Ministère des Transports a lancé **une série de tests mi-2019 afin d'accélérer le déploiement des véhicules autonomes en France**. Les zones rurales s'avéraient particulièrement ciblées, la ministre Élisabeth Borne ayant déclaré à cette période : "Je suis convaincue que ça peut être une solution dans des territoires où aujourd'hui, on a pas d'alternative à la voiture". Elle avait également ajouté "[qu]on a besoin d'engranger de l'expérience, c'est ce qu'on va faire avec ces seize expérimentations". Divers domaines d'applications sont concernés comme le transport de personnes, le fret ou encore la livraison de marchandises en ville. Le budget alloué au programme s'élève à 200 millions d'euros.

Si les véhicules autonomes demeurent jusqu'ici théoriques, **les tâches semi-automatisées devraient s'accroître dans le futur**, fournissant un **potentiel de marché élevé pour le big data**.

# Des opportunités variées pour les organisations publiques

Le *big data* présente un potentiel élevé pour les organismes gouvernementaux et les collectivités territoriales. **Certains usages s'avèrent déjà relativement répandus**, comme la prévention des catastrophes naturelles ou la préservation de l'environnement. Le croisement de nombreuses données peut permettre de **modéliser les dégâts éventuels causés par une tempête ou une inondation** par exemple. La multiplication de capteurs plus performants représente un atout pour mieux suivre l'évolution des populations animales et des milieux. L'accélération des changements climatiques devrait encourager l'utilisation dans ces domaines.

D'autres usages du *big data* se révèlent moins investis et devraient s'accroître à l'avenir. Il s'agit notamment des prévisions concernant la conjoncture économique et des analyses démographiques et financières liées aux régimes de retraite.

Les questions macroéconomiques constituent des thématiques cruciales pour le gouvernement : les variations du PIB (produit intérieur brut), de l'inflation ou du commerce extérieur doivent être calculées le plus précisément possible pour garantir la prospérité du pays. Prévoir l'évolution de la situation économique recouvre deux avantages principaux. Le premier consiste à **mieux adapter les politiques publiques en fonction de l'impact estimé d'une récession** ou au contraire d'une croissance très dynamique. Le second s'incarne dans **l'anticipation des effets d'une mesure gouvernementale sur l'économie** du pays. **Les réformes gagnent ainsi en efficacité** tandis que **leurs éventuelles conséquences négatives peuvent être circonscrites** par des politiques compensatoires. En agrégeant des données recueillies sur les réseaux sociaux, les médias, auprès des opérateurs mobiles ou encore liées aux transactions financières de toute nature, **les outils du *big data* peuvent considérablement affiner les modèles prédictifs actuels.**

L'analyse massive de données s'avère également être **un atout au niveau des problématiques démographiques**. La Caisse Nationale d'Assurance Vieillesse (CNAV) gère ainsi des dizaines de millions de comptes d'actifs et de retraités, ce qui lui fournit des quantités considérables d'informations. En multipliant les échanges avec d'autres services administratifs, les données peuvent être rassemblées pour une exploitation plus approfondie. L'usage du *big data* **favorise la lutte contre le non-recours et contre la fraude**, ou permet de **détecter les risques chez les bénéficiaires vieillissants**. En effectuant des simulations toujours plus fines des parcours de vie, les solutions de traitement de la donnée prodiguent **des estimations de long terme précises sur les besoins en financement du système de retraite**. Son pilotage devient plus aisé, le montant des dépenses se révélant mieux anticipé tandis que ces dernières peuvent être fléchées de façon plus efficace.

Primordiales à l'échelle d'une entreprise ou d'une municipalité, **les données se montrent d'autant plus importantes au niveau d'un pays entier**. La gestion du territoire et les activités sociales génèrent et nécessitent de grandes quantités d'informations, parfois **qualifiées d'infrastructures de par leur caractère à la fois structurel et stratégique**. Tout comme leurs homologues dans les transports ou l'énergie, les infrastructures digitales s'avèrent incontournables pour le développement d'un pays : **leur maîtrise et leur utilisation à bon escient constituent des déterminants majeurs pour sa compétitivité.**

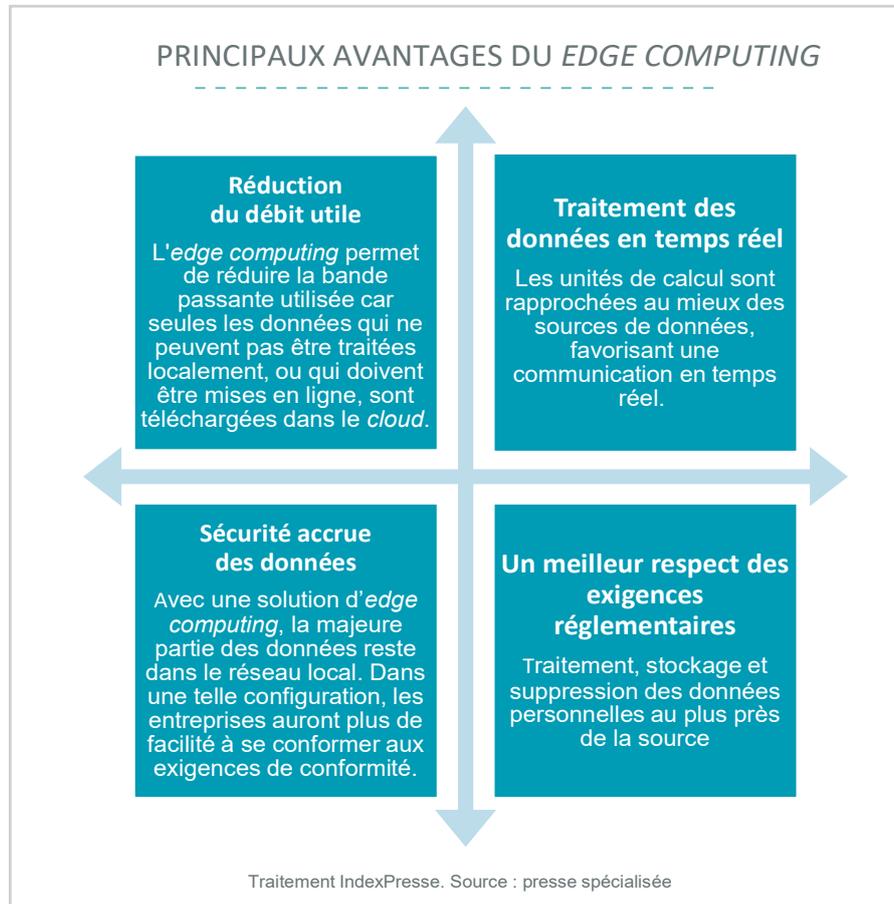
Concept multidimensionnel et transversal, le *big data* s'est diffusé au sein de nombreux secteurs depuis 2015 et **s'ancre durablement dans le paysage économique** français. Propulsé par l'innovation technologique, il se décline en **une variété d'usages et d'approches** tant sur le plan commercial que sur le volet technique. Toutefois, **à mesure de son déploiement se pose de nouvelles problématiques**, lesquelles le contraignent à des transformations inéluctables.

# LE BIG DATA À L'AUBE DE PROFONDES MUTATIONS

## Vers une nouvelle architecture du *big data* : le *edge computing*

Le *edge computing*, anglicisme signifiant "traitement des données à la périphérie du réseau", connaît un **fort développement depuis 2017 en France**. "Les solutions d'analyse de données et d'intelligence artificielle se déplacent de plus en plus vers le *edge computing*", affirmait le spécialiste Laurent Delattre dans un article paru dans le magazine *IT for Business* en février 2021. Selon la définition du cabinet spécialisé IDC, le *edge computing* peut être considéré comme un réseau mail-

lé de micro *data centers* qui traitent ou stockent les données critiques localement. Les données sont ensuite transmises vers un *data center* central ou un stockage *cloud*. **Il s'impose véritablement comme le chaînon manquant entre l'objet connecté et le *cloud computing***. "De plus en plus de technologies d'analyse de données s'animent en dehors des centres de données et du *cloud*, se rapprochant des actifs physiques. Cela réduit ou élimine la latence des solutions centrées sur les



données et permet une valeur plus temps réel", explique de son côté Gartner, entreprise américaine spécialisée dans les technologies de l'information. Forme d'architecture informatique distribuée ouverte faisant office d'alternative au *cloud computing*, le *edge computing* "répond à un besoin croissant d'analyser les données directement là où elles sont produites afin d'agir en temps réel". "Il apparaît pour les entreprises de plus en plus pertinent pour gagner en flexibilité en vue d'en dégager un avantage concurrentiel", explique Eric Delattre. Grâce à cette technologie, souvent qualifiée d'"à la frontière du réseau", **les données ne sont plus transmises à un *data center* distant mais sont traitées directement par le périphérique qui les génère** (objet connecté, smartphone...) ou par l'intermédiaire d'un ordinateur ou d'un serveur local.

Pour le spécialiste Eric Delattre, **plusieurs facteurs sont à l'origine de cette tendance**. Le premier et principal est **la croissance exponentielle des applications mobiles et de l'internet des objets (IoT)** depuis le milieu de la décennie 2010-2020. Cette dernière "continue à alimenter le besoin en plateformes sécurisées et évolutives à haute disponibilité et performances, à faible latence, pour traiter toutes les données générées et consommées en périphérie" explique-t-il. **Parallèlement, le *cloud* présente des limites pour les utilisateurs** : "Le *cloud* est limité par son modèle de traitement centralisé, qui introduit des contraintes de coût, de bande passante, de performance, de sécurité et de réglementation", synthétise Eric Delattre. Enfin, les **avancées technologiques en matière de capacités informatiques de calcul et de stockage** contribuent à un déploiement plus massif des architectures *edge computing*. D'autres analyses expliquent parallèlement l'essor du *edge computing* par la maturité croissante des plateformes du *edge*, le développement croissant de technologies telles que la connectivité 5G et Wi-Fi 6, et l'augmentation des cas d'utilisation pour l'analyse et le *machine learning*. Le déploiement de centres de données *edge computing* peut prendre de multiples formes. Il peut principalement prendre appui soit sur des appareils locaux remplissant une fonction spécifique soit sur des centres de données locaux

offrant des capacités de traitement et de stockage importantes, soit encore sur des centres de données régionaux desservant un grand nombre d'utilisateurs locaux.

Le *edge computing* est, selon les experts, **amené à se développer de manière exponentielle durant les cinq prochaines années**, compte tenu des avantages qu'il offre en complémentarité avec le *cloud* traditionnel. Selon les prédictions établies par le cabinet spécialisé dans les marchés des technologies de l'information IDC, plus de la moitié des nouvelles infrastructures IT seront déployées en périphérie au plan mondial à l'horizon 2023. Selon ce même cabinet, d'ici 2022, environ 74 % de toutes les données devront être analysées et traitées directement sur les appareils pour réduire la bande passante mondiale. Par ailleurs, d'après le cabinet de conseil français Aster Fab, la proportion de données traitées à l'échelle mondiale par *edge computing* pourrait passer de 9 % en 2020 à 75 % en 2025. Enfin, selon Forrester, les investissements *edge* au premier trimestre 2020 ont été plus élevés que sur l'ensemble de l'année 2019. "Au cours des trois prochaines années, les acheteurs déplaceront bon nombre de leurs stratégies de *cloud* vers le *edge*", prédit le cabinet d'étude spécialisé dans les technologies.

**Le secteur du *edge computing* est encore peu investi par les entreprises spécialisées en France**. La plateforme d'origine américaine Tracxn, spécialisée dans le recensement de start-up, répertoriait au total 268 start-up spécialisées dans l'*edge computing* à travers le monde à mi-décembre 2020, dont une extrême minorité sont de nationalité française.

9

Le nombre  
de start-up françaises  
spécialisées dans le *edge  
computing* recensées  
en France en 2020

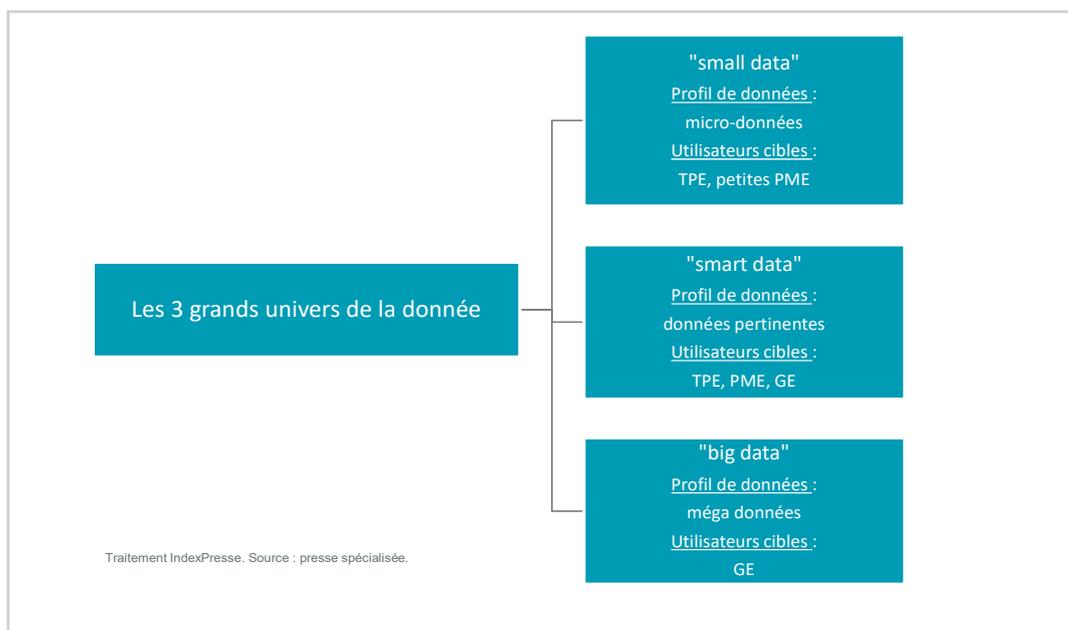
Source: Tracxn

## Une migration inexorable vers la *smart data*

Depuis 2015, le **déploiement intensif du *big data* dans les entreprises a entraîné une explosion du volume de données disponibles et à traiter, avec comme limite principale des données bien souvent en partie inexploitable et/ou peu pertinentes**. Selon le site internet Culture Data, développé par la société d'expertise en base de données et data intelligence Cartégie, ce caractère inexploitable s'explique par deux raisons majeures. "D'une part, parce que toutes ne sont pas utiles pour des actions marketing. De l'autre, parce que le flux est tel qu'il mélange des informations correctes et factuelles (les commandes, les informations saisies dans le cadre de la commande) avec des informations dont l'analyse est plus aléatoire (données de navigation, données saisies avant un abandon de panier, navigation sans commande...)". Par ailleurs, les données peuvent ne plus être à jour (car une donnée correcte d'il y a deux ans peut être fausse aujourd'hui) ou être mal renseignées. En définitive, **une proportion, jugée faible, de données est sous-exploitée par les en-**

**treprises**. Face à cette problématique majeure inhérente à la massification des flux d'informations, **une nouvelle approche de la donnée émerge** à travers le monde : la ***smart data***, aussi appelée "analyse en streaming". Évolution souvent présentée comme inéluctable par les spécialistes, la *smart data* se focalise ainsi sur les deux derniers V de la règle des 5V du *big data* : véracité et valeur. Il ouvre la possibilité de **tendre vers une analyse prédictive**, autrement dit d'anticipation des événements et résultats grâce à des algorithmes. Les enjeux du recours à cette approche résident dans l'optimisation des ressources techniques et l'amélioration du ROI (ou retour sur investissement). La *smart data* a vocation à **se focaliser sur l'analyse en temps réel des seules données dites "intelligentes", c'est-à-dire celles qui sont réellement utiles à la prise de décision**. L'approche *smart data* doit ainsi faire le tri entre des données brutes et difficilement utilisables et des données pertinentes à forte valeur ajoutée.

Elle se montre **particulièrement pertinente lors-**



qu'elle est appliquée au domaine du marketing, notamment au sein des entreprises qui adoptent une stratégie *data driven*. Les données qualifiées donnent des informations précieuses pour la segmentation et le ciblage mais aussi pour améliorer le parcours client et le taux de fidélisation. "La *smart data* permet de prédire les comportements des consommateurs, de personnaliser les messages et de réhabiliter de l'efficacité dans les stratégies marketing et commerciales pour offrir une approche plus ROIste aux retailers", explique l'entreprise d'ingénierie informatique ARCA Computing sur son blog. Selon un sondage réalisé entre mai et juillet 2018 par l'EBG, principal *think-tank* français sur l'innovation digitale, en partenariat avec Micropole et Qlik auprès d'un panel de 530 entreprises adhérentes à l'EBG, la meilleure

connaissance du client et de ses parcours se hisse d'ailleurs en 2018 au deuxième rang des objectifs du *big data* du point de vue des utilisateurs, avec 65 % des répondants.

L'approche *smart data*, qui nécessite de mobiliser des logiciels et outils informatiques spécifiques, se veut particulièrement **opportune dans le domaine du marketing, où la personnalisation s'impose comme un véritable facteur clé** de succès. Grâce à la *smart data*, les données qualifiées donnent des informations précieuses pour la segmentation et le ciblage mais aussi pour améliorer le parcours client et le taux de fidélisation. L'efficacité des campagnes marketing s'en trouve accrue.

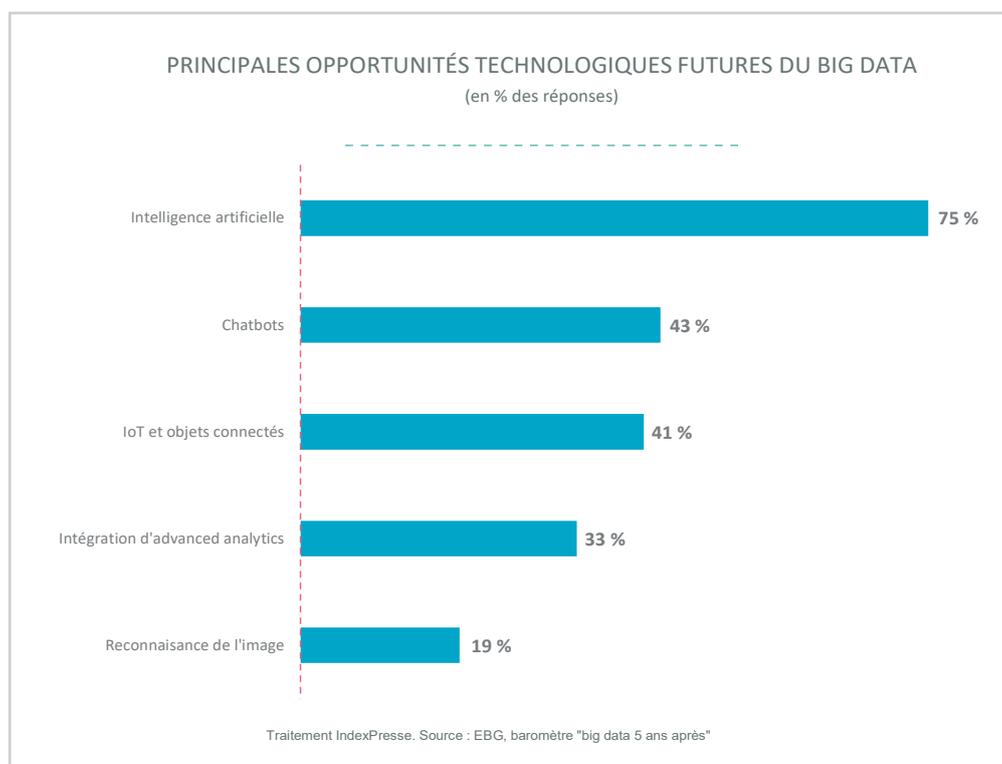
## Le *big data* et l'IA : une convergence inéluctable, gage d'efficacité

Selon les analystes, **une convergence du *big data* et de l'intelligence artificielle au cours de la prochaine décennie paraît inévitable**, ces deux technologies étant présentées comme complémentaires. Dans un sondage réalisé en 2018, l'EBG avait mis en évidence que **l'intelligence artificielle s'imposait de très loin comme la principale opportunité technologique** selon les professionnels de l'économie du numérique interrogés. Trois répondants sur quatre à l'enquête voyaient dans l'intelligence artificielle une opportunité technologique pour concevoir de nouveaux services innovants, très loin devant les chatbots (robots logiciels pouvant dialoguer avec un individu ou un consommateur), l'IoT et les objets connectés. Selon un consensus des spécialistes, une telle convergence serait **une révolution technologique d'une ampleur supérieure à celle d'internet, voire, pour certains, des technologies de communication mobiles**. L'intelligence artificielle (IA) offre la possibilité aux ordinateurs de simuler l'intelligence humaine grâce notamment au *machine learning*. Elle permet aux machines

d'apprendre sans avoir été au préalable programmées spécifiquement à cet effet. Les algorithmes développés dans ce cadre donnent aux ordinateurs la capacité d'acquiescer de nouvelles informations, de raisonner, et de s'auto-corriger par eux-mêmes. Mais, dans ces objectifs, la machine doit intégrer une grande quantité de données car, comme toute loi de statistiques et des probabilités, plus l'intelligence artificielle a de données à traiter, plus elle a de chance d'en tirer une tendance générale. Le *big data* représente dès lors l'essence du *machine learning*. De son côté, le *big data* a pour finalité de collecter, de stocker et d'analyser, via de multiples techniques, de très grands volumes de données. Or, les données brutes en grande quantité sont difficiles à analyser par un algorithme standard et les outils analytiques traditionnels ne sont pas suffisamment performants pour exploiter pleinement la valeur du *big data*. L'IA, qui ouvre la voie à des algorithmes sophistiqués, permet alors une analyse intelligente des informations issues du *big data*.

Cette convergence de deux technologies dans une logique de déploiement de l'analyse prédictive apporte des bénéfices : un gain de temps pour les équipes opérationnelles en charge d'analyser ces données, une communication facilitée entre les différentes entités productrices et consommatrices de cette donnée, et une prise de décision davantage éclairée grâce à la garantie d'analyse d'une donnée qualitative. Une étude réalisée en juin et juillet 2020 par l'agence Adelanto pour le compte d'Axians France et de HPE auprès de 144 responsables informatiques de PME et de grands comptes, a mis en évidence que l'amélioration de la capacité décisionnelle représentait le principal avantage perçu de l'IA selon les interrogés, avec 67 % des réponses, juste devant la capacité d'innovation. Selon cette étude, les principaux projets déployés ou en cours de réflexion portent principalement sur les modèles prédictifs (72 % des réponses). Si la convergence entre IA et *big data* dans les années à venir constituera une tendance profonde selon les analystes, force est

de constater que **des défis majeurs sont à relever**. Tout d'abord, "il n'existe aucune règle stricte qui régit la gestion de l'intelligence artificielle", explique le journaliste Yves Frandmontagne dans un article publié en février 2017 dans la revue *IT Social*. Defait, il existe **des craintes en matière de sécurité**. "Les problèmes de sécurité liés à ce concept sont à aborder au moment opportun afin de réduire les risques d'impacts qui deviennent plus importants", explique le journaliste. De plus, il n'existerait **aucune disposition légale** régularisant la vente et l'achat de données ou de logiciels d'intelligence artificielle. Dans sa mise en place, les entreprises se heurtent parallèlement à **des difficultés de déploiement**. D'après les résultats de l'étude réalisée par l'agence Adelanto à l'été 2020, le manque de compétences internes ressort comme le principal frein, totalisant 63 % des réponses. Suivent l'insuffisance des ressources informatiques (49 %), puis le budget élevé du prototype (près de 46 %).



# UN ESSOR DISPARATE QUI S'ACCOMPAGNE DE QUESTIONNEMENTS

## Un déploiement qui s'accélère...

Depuis 2015, les concrétisations de projets de déploiement de *big data* se sont accélérées en France. À la faveur de la conjonction de facteurs à la fois technologiques, politiques et stratégiques, le *big data* s'est progressivement développé à plus grande échelle. Longtemps resté au stade expérimental dans les grandes entreprises, il s'étend aujourd'hui aux petites sociétés.

Les ambitions gouvernementales affichées en 2014 de faire de la France un exemple en Europe dans le domaine du *big data* ont trouvé écho auprès des entreprises françaises. Selon les dernières données de l'enquête sur les technologies de l'information et de la communication réalisée par Eurostat, office statistique de l'Union européenne, la proportion d'entreprises françaises de 10 salariés et plus des secteurs marchands qui recouraient en 2017 à l'analyse de données massives ressortait en croissance de 5 points par rapport à 2015, hors secteurs agricole, de la finance et de l'assurance. Elle atteignait par ailleurs un niveau à la fois élevé et supérieur à la moyenne de l'ensemble de l'Union européenne à 28 pays. Selon l'enquête, 16 % des entreprises françaises de 10 salariés et plus y avaient recours, soit un ratio de quatre points plus élevé que la moyenne de l'Union européenne.

Bien que positive, cette tendance masque toutefois de fortes disparités selon les secteurs et les tailles d'entreprises considérées. Le *big data* apparaît avant tout à un stade particulièrement avancé au sein des grandes et très grandes structures. D'après les dernières données disponibles issues de l'enquête TIC 2019 d'Eurostat, alors que la proportion de grandes entreprises (250 salariés ou plus) recourant à l'analyse de données massives atteignait 37 %, celle des PME (entre 10 et 250 salariés) ressortait dans le même temps

inférieure à 10 %. Cette dichotomie s'explique notamment par le différentiel des moyens financiers mobilisables pour concrétiser des projets de déploiement de *big data*, dont le coût à supporter se révèle très élevé.

## ...mais soulève des questionnements

Concept multidimensionnel et transversal, le *big data* se diffuse dans de nombreux secteurs et s'ancre durablement dans le paysage économique, en France comme à l'international. Propulsé par l'innovation technologique, il se décline en une variété d'usages et d'approches tant sur le plan commercial que sur le volet technique. La massification du traitement de l'information conduit à de nouvelles manières d'aborder la donnée. Marchandise de valeur à acquérir et échanger, matière première brute à exploiter ou infrastructure essentielle à un écosystème, sa définition s'avère multiple et changeante au gré des évolutions quant à son utilisation. Dans ce domaine, la France part avec des atouts : elle dispose de grands groupes (comme Atos et Thales), d'un grand nombre de start-up, ainsi que de politiques publiques favorables au développement du *big data*. Ces dernières passent notamment par un soutien financier au secteur et la mise à disposition des données par l'*open data*.

Alors que le déploiement des outils du *big data* s'accélère et que ses limites ne cessent de s'étendre, des dilemmes se posent au sujet du développement de cette technologie. Le premier d'entre eux concerne la complexité et l'efficacité des dispositifs, tant la masse de données pouvant être collectées se révèle considérable. L'exploitation des *dark data*, ces données disponibles mais ignorées jusqu'alors, accroît encore le phénomène. Déjà soumises à des flux d'infor-

## UN ESSOR DISPARATE QUI S'ACCOMPAGNE DE QUESTIONNEMENTS

mations constants, les organisations font face à de tels volumes qu'il leur est **nécessaire de bien identifier les gisements les plus pertinents** pour leur activité. En réponse à ces problématiques, **le big data bifurque déjà vers des technologies et pratiques émergentes** comme le *edge computing* et la *smart data*.

Souvent présentée comme une solution au défi climatique, l'analyse massive de données pourrait **optimiser l'utilisation des ressources énergétiques**. La meilleure gestion du système électrique (*smart grid*), des flux de véhicules et de passagers dans les transports en commun (*smart city*) ou encore l'amélioration des performances du tissu industriel (*smart factory*) seraient de nature à **limiter l'impact environnemental des activités humaines**. La réalité de ce scénario demeure toutefois incertaine, tant **le big data se révèle ambigu concernant les questions écologiques**. La connectivité accrue au sein des villes, des foyers, des entreprises et des administrations nécessite **des capacités de captation et de stockage des données considérables**.

Énergivores et requérant l'extraction polluante de métaux spécifiques, elles peuvent **provoquer une augmentation des usages** en rendant toujours plus fluide la production ou les trajets.

Enfin, une autre problématique concerne **la régulation du secteur**. Frein à l'innovation pour les uns, élément de structuration pour d'autres, elle s'impose dans **une logique de protection des données personnelles**. Question particulièrement sensible, la confidentialité de ces informations peut constituer **un obstacle au développement de certains algorithmes**. Dans un univers concurrentiel où l'accès toujours plus grand à la donnée renforce les solutions du *big data* et de l'IA, **l'équilibre entre le respect de la vie privée et la compétitivité des entreprises européennes** va se poser avec d'autant plus d'intensité. Face à des espaces économiques plus permissifs sur l'utilisation des données comme la Chine et les États-Unis, **la capacité de l'Europe à protéger ses acteurs et ses données** s'avèrera déterminant pour le futur du secteur.

# LES FORCES EN PRÉSENCE

## Liste des start-up françaises du *big data*

Nom de l'entreprise	Activité	Effectifs salariés	Année de création
AD SCIENTAM	Applications et tests connectés pour le domaine médical	10 à 19	2013
AKHEROS	Logiciels de cybersécurité pour le M2M	1 à 2	2012
AMIRAL TECHNOLOGIES	Outils de maintenance prédictive	nd*	2018
APIDATA	Logiciel de gestion des flux de données	3 à 5	2012
APOLLO PLUS	Modèles prédictifs pour le marketing	nd*	2019
BATVOICE TECHNOLOGIES	Solution d'analyse de la voix	nd*	2015
BEEBRYTE	Solution dédiée au <i>smart grid</i>	nd*	2015
BRAINCUBE	Solution d'IoT pour l'industrie	nd*	2007
CALLDESK	Robot conversationnel pour les <i>call centers</i>	10 à 19	2016
COSMO COMPANY	Outils d'aide à la décision pour l'industrie	50 à 99	2010
DATAGALAXY	Outil de <i>big data</i> collaboratif	3 à 5	2015
DATAKALAB	Analyse des flux de personnes par l'image	3 à 5	2016
DATASCIENTEST	Formations en <i>data science</i>	nd*	2017
DATASWATI	Solution d'IA pour l'industrie	6 à 9	2016
DATAVALORIS	<i>Deep learning</i> appliqué au marketing	6 à 9	2015
DCBRAIN	Modèles prédictifs pour l'industrie	20 à 49	2014
DEEPOMATIC	Logiciel pour améliorer la visualisation du travail des ouvriers	20 à 49	2014

\* : non disponible  
 Traitement IndexPresse. Sources : Kompass, societe.com, et sites web des entreprises concernées.

Nom de l'entreprise	Activité	Effectifs salariés	Année de création
DHATIM	Assistant IA pour augmenter la productivité	20 à 49	2008
DREAMQUARK	Solution d'IA dédiée aux assurances	20 à 49	2014
EARTHCUBE	Logiciels de <i>big data</i> , surtout pour les données géospatiales	80	2016
ERNERGISME	Solution dédiée à la <i>smart grid</i>	20 à 49	2004
EXOTEC SOLUTIONS	Système de robots pour la logistique	50	2015
FACTEN	Logiciel axé sur l'analyse des brevets et des articles scientifiques	1 à 2	2012
FIELDBOX.AI	Créateur et opérateur d'IA pour l'industrie	20 à 49	2011
FORTIA (20S)	Logiciel de création d'applications et d'IA	nd*	2012
GOLEM.AI	Solution IA d'assistance aux employés	10 à 19	2016
GOSHABA	Logiciel RH d'automatisation de la sélection des candidatures	6 à 9	2014
INCALIA	Solutions sur-mesure de <i>machine learning</i> pour les start-up	3 à 5	2016
INQOM	Logiciel de comptabilité basé sur l'IA	20 à 49	2016
JENJI	Gestion des données financières par l'IA	6 à 9	2013
KAYRROS	Outil d'aide à la décision dans des secteurs variés, notamment pour le risque climatique	50 à 99	2016
KILI TECHNOLOGY	Solution d'annotation pour créer des données d'apprentissage	nd*	2018
KLS GROUP	Logiciels de gestion de la chaîne logistique	4	2008
LA FORGE	Outil de design par l'IA	nd*	
LA JAVANESS	Solution d'IA pour accroître la productivité	20 à 49	2015
LEVIA.AI	Robot conversationnel	nd*	
LILI.AI	Assistant virtuel	3 à 5	2016
LOKAD	Solution d'optimisation de la chaîne logistique	20 à 49	2008
METEO*SWIFT	Traitement de données pour les systèmes électriques	1 à 2	2015
MINDSAY	Robot conversationnel et automatisation des procédures	nd*	2016
MYDATAMODELS	Solution de <i>smart data</i>	6 à 9	2018
MYRIAD	Automatisation des procédures	nd*	
NAM.R	Solution de géolocalisation	20 à 49	2017
NEHOOV	Outil d'analyse et de prévision sans recours à un <i>data scientist</i>	1 à 2	2016

\* : non disponible  
 Traitement IndexPresse. Sources : Kompass, societe.com, et sites web des entreprises concernées.

Nom de l'entreprise	Activité	Effectifs salariés	Année de création
OCTOPEEK	Conseil en <i>big data</i> et marketing	20 à 49	2010
OPENIO	Infrastructure de stockage modulaire et open source optimisée pour le <i>big data</i>	20 à 49	2015
OWKIN	Solution <i>big data</i> pour le domaine médical	102	2016
PACKETAI	Solution autonome de prédiction des incidents IT	nd*	2018
PREVISION.AI	Solution de création de modèles prédictifs à partir de bases de données	3 à 5	2016
PRICINGHUB	Solution d'analyse et de fixation des prix, notamment dans le e-commerce	nd*	
PROPHESSEE	Système de vision neuromorphique pour les machines	50 à 99	2014
QEMOTION	Outil d'analyse des émotions pour améliorer le parcours client	6 à 9	2015
QUAVITRA (TIMPROVE)	Solution RH de mesure de l'expérience collaborateur et de son impact sur l'entreprise	nd*	2019
QUIVIDI	Mesure et optimisation de la communication pour le digital "out of home"	3 à 5	2006
SIDETRADE	Solution <i>big data</i> pour le marketing et les prévisions de ventes	100 à 199	2000
SYNAPLUS	Exploitation et valorisation des données avec le logiciel CosMo	3 à 5	2014
TARGET2SELL	Solution IA pour améliorer l'expérience client	20 à 49	2012
THERAPIXEL	Solution d'imagerie basée sur le <i>big data</i> pour la détection du cancer du sein	1 à 2	2013
TILKAL	Solution de traçabilité de la chaîne logistique	10 à 19	2017
TILKEE	Application <i>cloud</i> pour accroître les ventes	20 à 49	2012
TOUCAN TOCO	Solution de reporting et data visualisation	50 à 99	2014
TRUSTPAIR	Solution anti-fraude aux virements	3 à 5	2017
TWO-I	Analyse des flux de personnes par l'image	6 à 9	2017
VEKIA	Solution d'optimisation de la chaîne logistique	20 à 49	2007
WAKEO	Outil de suivi en temps réel de la chaîne logistique	3 à 5	2016
XLEARN	Outil de prédiction des besoins en RH	nd*	2017
ZEENEA	Outil de <i>big data</i> collaboratif	35	2017

\* : non disponible  
 Traitement IndexPresse. Sources : Kompass, societe.com, et sites web des entreprises concernées.

## Liste des acteurs cités dans l'étude

Société	Nature de l'entreprise	Pays d'origine
Ad Scientam	Applications et tests connectés pour le médical	France
Adelanto	Agence de web marketing	France
ADP	Groupe spécialiste des services administratifs	France
Airbus	Groupe d'aéronautique et de défense	France
Akheros	Logiciels de cybersécurité pour le M2M	France
Alibaba	Groupe du numérique spécialiste du e-commerce	Chine
Allianz	Compagnie d'assurance	Allemagne
Amazon	Groupe du numérique spécialiste du e-commerce	États-Unis
Amiral Technologies	Outils de maintenance prédictive	France
Anduril	Big data et IA pour la défense	États-Unis
APEC	Association pour l'emploi des cadres	France
Apidata	Logiciel de gestion des flux de données	France
Apollo Plus	Modèles prédictifs pour le marketing	France
Arca Computing	Agence de conseil et développement web	France
Arcadis	Entreprise d'ingénierie, de conseil et de gestion de projets	Pays-Bas
Arquus	Entreprise de défense spécialiste des équipements terrestres	France
Atos	Groupe spécialiste des services numériques	France
Atrium Data	Entreprise d'audit et de conseil spécialisée dans le domaine des <i>data centers</i>	France
Audi	Constructeur automobile	Allemagne
Aviva	Compagnie d'assurance	Royaume-Uni
Axa	Compagnie d'assurance	France
Axians France	Spécialiste des solutions ICT (information et communication)	France
Baidu	Groupe du numérique spécialisé dans les moteurs de recherche	Chine
Batvoice Technologies	Solution d'analyse de la voix	France
BeeBryte	Solution dédiée à la <i>smart grid</i>	France
Berkshire Hathaway	Société d'investissement	États-Unis
BMW	Constructeur automobile	Allemagne
BotSociety	Fournisseur de robots conversationnels (chatbots)	États-Unis
Braincube	Solution d'IoT pour l'industrie	France
British Petroleum	Compagnie pétrolière	Royaume-Uni
CallDesk	Robot conversationnel pour les <i>call centers</i>	France
Cardiologs	Solution <i>big data</i> pour la cardiologie	France
Carrefour	Enseigne de la grande distribution	France
Cartégie	Entreprise spécialisée en base de données et data intelligence	France
Clearview AI	Solution d'identification par analyse de photos en ligne	États-Unis
Cloudera	Editeur de logiciels <i>big data</i>	États-Unis
Cognigy	Fournisseur de robots conversationnels (chatbots)	Allemagne
CoSMo Company	Outils d'aide à la décision pour l'industrie	France
Data&Data	Solution de <i>big data</i> anti-contrefaçon	France
DataGalaxy	Outil de <i>big data</i> collaboratif	France
Datakalab	Analyse des flux de passants par l'image	France
Datascientest	Formations en <i>data science</i>	France
Dataswati	Solution d'IA pour l'industrie	France
Datavaloris	Deep learning appliqué au marketing	France

Société	Nature de l'entreprise	Pays d'origine
DCBrain	Modèles prédictifs pour l'industrie	France
Deepomatic	Logiciel pour améliorer la visualisation du travail des ouvriers	France
Dell EMC	Editeur de logiciels et de systèmes de stockage	États-Unis
Dhatim	Assistant IA pour augmenter la productivité	France
Dreamquark	Solution d'IA dédiée aux assurances	France
Earthcube	Logiciels de <i>big data</i> , surtout pour les données géospatiales	France
EBG	Fédération du marketing, du digital et de l'innovation	France
Energisme	Solution dédiée à la <i>smart grid</i>	France
Engie Ineo	Systèmes électroniques dans l'énergie et la sécurité	France
Entrupy	Solution de <i>big data</i> anti-contrefaçon	États-Unis
Eurostat	Institut statistique de l'Union européenne	
Exotec Solutions	Système de robots pour la logistique	France
Facebook	Réseau social	États-Unis
Facten	Logiciel axé sur l'analyse des brevets et des articles scientifiques	France
Fieldbox.ai	Créateur et opérateur d'IA pour l'industrie	France
Finferries	Opérateur de ferrys	Finlande
Fortia (2OS)	Logiciel de création d'applications et d'IA	France
Frost et Sullivan	Cabinet d'études et conseil	États-Unis
Gartner	Cabinet d'études et conseil en technologies avancées	États-Unis
Genetec	Spécialiste des solutions de sécurité	Canada
GfK	Cabinet d'études de marché et d'audit marketing	Allemagne
Golem.ai	Solution IA d'assistance aux employés	France
Google	Groupe du numérique spécialisé dans les moteurs de recherche	États-Unis
Goshaba	Logiciel RH d'automatisation de la sélection des candidatures	France
Groupe Bertolami	Spécialiste du transport de passagers	France
HBO	Chaîne de télévision	États-Unis
Hewlett Packard Enterprise (France)	serveurs, réseau, stockage, logiciels et services aux entreprises	États-Unis
Horoquartz	Spécialiste de la gestion des plannings et du personnel	France
Hortonworks	Editeur de logiciels informatiques	États-Unis
Huawei	Groupe spécialisé dans les systèmes de communication	Chine
Humanis	Groupe mutualiste de protection sociale	France
IDC France	Cabinet d'études et conseil en nouvelles technologies de l'information	États-Unis
Incalia	Solutions sur-mesure de <i>machine learning</i> pour les start-up	France
Inqom	Logiciel de comptabilité basé sur l'IA	France
In-Q-tel	Fonds d'investissement de la CIA	États-Unis
Insee	Organisme public des statistiques économiques	France
Intermarché	Enseigne de la grande distribution	France
Janssen	Laboratoire pharmaceutique	Belgique
Jenji	Gestion des données financières par l'IA	France
Johnson & Johnson	Laboratoire pharmaceutique	États-Unis
JP Morgan	Banque	États-Unis
Kayros	Outil d'aide à la décision, notamment pour le risque climatique	France
Kili Technology	Solution d'annotation pour créer des données d'apprentissage	France
KLS Group	Logiciels de gestion de la chaîne logistique	France
La Forge	Outil de design par l'IA	France
La Javaness	Solution d'IA pour accroître la productivité	France
Levia.ai	Robot conversationnel	France
Lili.ai	Assistant virtuel	France
Lokad	Solution d'optimisation de la chaîne logistique	France
Lumibird	Spécialiste du laser pour le médical et la défense	France
Meteo*Swift	Traitement de données pour les systèmes électriques	France
Michael Page France	Entreprise spécialiste du recrutement de cadres confirmés	France
Micropole	Entreprise de conseil spécialisée dans la transformation digitale	France
Microsoft	Groupe du numérique spécialisé dans les systèmes d'exploitation	États-Unis
Mindsay	Robot conversationnel et automatisation des procédures	France
Multiasistance	Gestion des sinistres pour les assurances	France
MyDataModels	Solution de <i>smart data</i>	France
Myriad	Automatisation des procédures	France
Nam.r	Solution de géolocalisation	France
Navya	Constructeur de véhicules autonomes	France

Société	Nature de l'entreprise	Pays d'origine
Nehoov	Outil d'analyse et de prévision sans recours à un data scientist	France
Next Big Sound	Spécialiste du <i>big data</i> pour l'industrie musicale	États-Unis
Octopeek	Conseil en <i>big data</i> et marketing	France
OpenIO	Infrastructure de stockage modulaire et open source optimisée pour le Big data	France
Oracle	Gestionnaire de bases de données	États-Unis
Owkin	Solution <i>big data</i> pour le médical	France
PacketAI	Solution autonome de prédiction des incidents IT	France
PageGroup	Groupe spécialisé dans le recrutement de profils qualifiés	Royaume-Uni
Palantir	<i>Big data</i> surtout pour la défense, le renseignement et l'industrie	États-Unis
Pandorobot	Fournisseur de robots conversationnels (chatbots)	États-Unis
Persistent Systems	Entreprise de services numériques	Inde
Prévision.io	Solution de création de modèles prédictifs à partir de bases de données	France
PricingHub	Solution d'analyse et de fixation des prix, notamment dans le e-commerce	France
Prophesee	Système de vision neuromorphique pour les machines	France
Qemotion	Outil d'analyse des émotions pour améliorer le parcours client	France
Qlik	Editeur de logiciels de business intelligence et de data-visualisation	États-Unis
Qlink	Solution de <i>big data</i> pour les paiements	Afrique du Sud
Quavitra (Timprove)	Solution RH de mesure de l'expérience collaborateur	France
Quividi	Mesure et optimisation de la communication pour le digital "out of home"	France
Qwant	Développeur d'un moteur de recherche	France
Rolls-Royce	Constructeurs de voitures et de moteurs	Royaume-Uni
Sanofi	Laboratoire pharmaceutique	France
Savana	Spécialiste du <i>big data</i> dans le médical	Espagne
Sidetrade	Solution <i>big data</i> pour le marketing et les prévisions de ventes	France
Snatchbot	Fournisseur de robots conversationnels (chatbots)	Israël
SNCF	Entreprise publique du ferroviaire	France
Societeinfo.com	Entreprise spécialisée dans l'enrichissement de données	France
Sophia Genetics	Solution <i>big data</i> pour la génétique	Suisse
Steam	Plateforme de jeux vidéo	États-Unis
Synaplus	Exploitation et valorisation des données avec le logiciel CosMo	France
Syntec Numérique	Syndicat des acteurs du numérique	France
Taikang Online	Compagnie d'assurance digitale	Chine
Target2Sell	Solution IA pour améliorer l'expérience client	France
Tars	Fournisseur de robots conversationnels (chatbots)	États-Unis
Tencent	Groupe du numérique aux activités diversifiées	Chine
Tesla	Constructeur de voitures électriques	États-Unis
Thales	Groupe spécialiste de l'électronique de défense	France
The Warranty Group	Compagnie d'assurance	États-Unis
Therapixel	Solution <i>big data</i> pour la détection du cancer du sein	France
Tilkal	Solution de traçabilité de la chaîne logistique	France
Tilkee	Application cloud pour accroître les ventes	France
Toucan Toco	Solution de reporting et data visualisation	France
Trustpair	Solution anti-fraude aux virements	France
Twitter	Réseau social	États-Unis
Two-I	Analyse des flux de passants par l'image	France
Umanis	Prestataires de services numériques (informatique décisionnelle, big data)	France
Valéo	Équipementier automobile	France
Vekia	Solution d'optimisation de la chaîne logistique	France
Vertex Pharmaceuticals	Laboratoire pharmaceutique	États-Unis
Vertiv France	Fournisseur de solutions, équipements et infrastructures numériques	États-Unis
Wakeo	Outil de suivi en temps réel de la chaîne logistique	France
Webdrone	Solution de <i>big data</i> anti-contrefaçon	France
Xerfi	Cabinets d'études économiques et sectorielles	France
xLearn	Outil de prédiction des besoins en RH	France
Zeenea	Outil de <i>big data</i> collaboratif	France
Zynga	Développeur de jeux mobiles	États-Unis

Traitement IndexPresse.

# SOURCES UTILISÉES

- Amandi Lorena, "Webdrone ou l'intelligence artificielle pour lutter contre la contrefaçon", *carnetsduluxe.com*, juillet 2018
- APEC, "Les métiers de la data (étude APEC)", *apex.fr*, octobre 2020
- Arène Véronique, "Hausse des salaires pour les pros des data, du code et de la sécurité", *lemondeinformatique.fr*, septembre 2020
- Arène Véronique, "Syntec Numérique table sur 1% de croissance dans l'IT en 2021", *lemondeinformatique.fr*, décembre 2020
- Azouyan Philippe, "Quelle est la prochaine étape pour le Edge Computing ?", *Global Security Mag*, février 2021
- Bally Pascal, "Agriculture et Big data : la donnée est dans le pré", *inneance.fr*, avril 2019
- Bionne Delphine, "Big data...5 ans après (baromètre EBG)", *micropole.com*, janvier 2019
- Blaizot Coline, "Rolls-Royce et Finferries lancent le premier ferry autonome au monde", *objetconnecte.net*, décembre 2018
- Bodet Guillaume, "Data, comment les entreprises françaises peuvent-elles rattraper leur retard face aux géants américains et asiatiques ?", *itforbusiness.fr*, décembre 2020
- Boivigny Camille, "Transformer le big data en or", *Pharmaceutiques*, octobre 2019, p.16-19
- Brenugat Valérie, "SPIE Industrie et tertiaire : du capteur à la Big data", *Trends*, novembre-décembre 2019, p.20-24
- Breuil Pascale, "Les données au coeur de la stratégie numérique de la branche retraite", *Revue française des affaires sociales*, octobre 2017, p.150-158
- Briffod Stéphane, "2020 : l'année pivot pour recruter des ingénieurs de la donnée", *decideo.fr*, août 2020
- Carpentier-Gregson Yann, "Quelles technologies pour une stratégie Big data réussie en 2020", *informatiquenews.fr*, septembre 2020
- Chantrel Flavien, "Décryptage : quelles évolutions pour le Big data en 2020 ?", *blogdumoderateur.com*, février 2020
- Charpentier Arthur, "Big data, GAFA et assurance", *Réalités industrielles*, février 2020, p.53-57
- Charpentier Arthur, "Quel avenir pour les probabilités prédictives en assurance ?", *Réalités industrielles*, février 2020, p.74-77
- Chenevoy Clotilde, "Data : pourquoi les distributeurs s'allient à des géants de la tech", *LSA*, février 2020, p.24-27
- Chokogoue Juvenal, "Le Big data 5 ans après : phénomène de mode ou réalité ?", *lemagit.fr*, octobre 2018
- Cochelin Pascal, "Volume, vitesse et variété : la production en trois V", *terre-net.fr*, février 2017
- Colmont Christine, "IA et Pharma. Les accords se multiplient", *Pharmaceutiques*, octobre 2019
- De Caevel Christophe, "La Wallonie avance vers la médecine de demain", *Trends*, juillet 2018
- Débes Florian, "Palantir, la licorne aux deux visages", *Les Échos Week-end*, avril 2020, p.18-21
- Delattre Laurent, "L'IA émerge enfin dans les projets des entreprises françaises", *itforbusiness.fr*, novembre 2020
- Delattre Laurent, "Marché de l'emploi IT et rémunération des métiers de l'IT en 2021", *itforbusiness.fr*, septembre 2020
- Delattre Laurent, "Vers la fin du Big data : 10 tendances Data & Analytics en 2021", *itforbusiness.fr*, février 2021

# SOURCES UTILISÉES

- El Ghouzzi Julie, "25 start-up de l'Intelligence Artificielle au prisme du luxe", *centreduluxe.com*, juillet 2018
- Elie Mathilde, "La "safe city" sous surveillance", *Entreprise & Carrières*, février 2020, p.30-37
- Ernst & Young, "Formations et compétences sur l'Intelligence Artificielle en France (étude OPIIEC)", *opiiec.fr*, octobre 2019
- Faquet Romain, "Numérisation des entreprises françaises", *tresor.economie.gouv.fr*, novembre 2020
- Fernandez Rodriguez Laura, "La crise sanitaire, nouveau laboratoire de la safe city", *La Gazette des communes*, mai 2020, p.12-13
- Fouquet Caroline, "Bilan 2020 et perspectives 2021 du secteur numérique : le secteur ne ralentit pas autant que prévu", *syntec-numerique.fr*, décembre 2020
- Garcia-Montero Célia, "Marché de l'IoT en France : tous les chiffres", *lejournaldunet.fr*, avril 2021
- Garreau Marion, "La reconnaissance faciale progresse en France", *La Gazette des communes*, novembre 2020, p.24-35
- Gaube Emeline, "Le jeu vidéo, secteur en pointe dans le Big data", *larevuedesmedias.ina.fr*, octobre 2015
- Germain Sabine, "Big data : le champ des possibles", *L'Argus de l'assurance*, décembre 2019, p.27
- Gros Maryse, "Secteur IT : Syntec Numérique prévoit 4,2% de croissance en 2019", *lemondeinformatique.fr*, juin 2019
- Hamon Grégoire, "Navya nourrit des ambitions dans la logistique", *L'Officiel des transporteurs*, février 2021
- Hoang Mathilde, "Les données publiques : qu'en est-il de leur ouverture et de leur utilisation ?", *Cahiers français*, janvier-février 2021, p.36-45
- Ingrand Cédric, "Google dévoile Meena, un "chatbot" presque humain", *lci.fr*, janvier 2020
- INSEE, "Les TIC et le commerce électronique dans les entreprises en 2019 (enquête INSEE)", *insee.fr*, avril 2020
- Jeanne François, "La convergence agriculture et numérique a (enfin) son institut", *IT for Business*, juillet 2017, p.14
- Krim Mourad, "Centres de données : la France dans le top 10 mondial des pays les plus attractifs", *itsocial.fr*, mars 2021
- Krim Mourad, "L'adoption de l'IA par les entreprises françaises freinée à cause du déficit de compétences", *itsocial.fr*, mars 2021
- L. Bastien, "Big data 2019 : toutes les tendances et prédictions", *lebigdata.fr*, décembre 2018
- L. Bastien, "Big data 2020 : tendances et prédictions pour l'année à venir", *lebigdata.fr*, décembre 2019
- L. Bastien, "Intelligence artificielle et big data : une convergence révolutionnaire", *lebigdata.fr*, novembre 2018
- L. Bastien, "Agriculture : comment le Big data révolutionne l'industrie agricole ?", *lebigdata.fr*, mars 2018
- L. Bastien, "Nvidia : comment le fabricant de GPU est devenu un géant du Big data", *lebigdata.fr*, mars 2021
- Lagneau Laurent, "Pour l'armée de Terre, l'immobilisation d'un matériel terrestre doit "devenir l'exception"", *opex360.com*, février 2020
- Lamigeon Vincent, "Palantir arme ses datas", *Challenges*, octobre 2020, p.60-61
- Lemke Coralie, "Comment l'intelligence artificielle aide à lutter contre la maladie de Parkinson", *usinenouvelle.com*, avril 2018

# SOURCES UTILISÉES

- Lemke Coralie, "L'IA, nouvel outil de précision des médecins", *Maintenance & Entreprise*, novembre 2018, p.35-42, 44-45
- Livaudais Peter, "Maintenance prédictive : quand la Big data se met au service d'une industrie plus efficiente", *Maintenance & Entreprise*, novembre-décembre 2019, p.18-19
- Mansard Laurent, "Élancourt : le commissariat du futur pourrait voir le jour pour les J. O.", *leparisien.fr*, juillet 2019
- Monier Nicolas, "Ces machines qui enquêtent sous surveillance", *Le nouvel Économiste*, novembre 2019, p.24, 27
- Onnen Antoine, "Entreprise française cherche talent du Big data", *usine-digitale.fr*, mars 2021
- Onnen Antoine, "Wanted : Entreprise française cherche talent du Big data", *usine-digitale.fr*, mars 2021
- Passily Augustine, "Explosion des nouvelles technologies : pour qui ? pour quoi ?", *L'Information Agricole*, juin 2017, p.12-23
- Pechon Bastien, "Moisson de données", *Réalités industrielles*, juillet 2018, p.58-61
- Pégard Fanny, "Quels sont les métiers d'avenir dans la data ?", *futura-sciences.com*, novembre 2020
- Pradine Nadège, "Cloud computing et big data : la dématérialisation au service des sociétés européennes", *insee.fr*, novembre 2019
- Richard Philippe, "Big data : 6 tendances en 2020", *itsocial.fr*, janvier 2020
- Richaud Nicolas, "Fondé par un Français, App Annie est devenu une référence mondiale dans le mobile", *lesechos.fr*, mars 2019
- Robert Augustin, "Smart City et mobilité : les enjeux d'une ville moderne", *widoobiz.com*, janvier 2021
- Rosencrance Linda, "San Francisco multiplie les projets Smart City dans les transports et le stationnement", *lemagit.fr*, novembre 2017
- S. Elina, "Plus de violations de données malgré des dépenses accrues en cybersécurité", *lebigdata.fr*, mars 2021
- Sequeira Martins Gilmar, "Le big data doit encore convaincre les RH", *L'Usine Nouvelle*, octobre 2019, p.12-15
- Simon Jean-Paul, "Jeu vidéo : à qui profite le big data ?", *larevuedesmedias.ina.fr*, décembre 2015
- Simon Jean-Paul, "Le Big data : un enjeu pour les industries créatives", *larevuedesmedias.ina.fr*, janvier 2015
- Tassi Philippe, "Big data et statistiques", *Économie et statistiques*, avril 2019, p.5-193
- Texier Bruno, "Les archivistes face au management des données", *Archimag*, octobre 2020, p.20-21
- Tournemaine Jean-Pierre, "Les infrastructures de datacenter se déploient en région", *informatiquenews.fr*, août 2020
- Tournemaine Jean-Pierre, "Les infrastructures de datacenter se déploient en région.", *informatiquenews.fr*, août 2020
- Tréguer Félix, "La "ville sûre" ou la gouvernance par les algorithmes", *Le Monde Diplomatique*, juin 2019, p.22-23
- "Contrefaçon : un manque à gagner de 6,8 milliards d'euros pour notre économie", *leparisien.fr*, juin 2019
- "Data Centers en France : les enjeux pour 2020", *lesechos.fr*, décembre 2019
- "La donnée, pétrole du futur pour les uns, lumière du soleil pour les autres", *Le nouvel Économiste*, mars 2020, p.2, 4, 6-8
- "La nouvelle France industrielle : feuille de route big data", *economie.gouv.fr*, juillet 2014
- "La Tech, là où on ne l'attend pas", *Les Échos Week-end*, octobre 2020, p.82-83

# SOURCES UTILISÉES

"Les 7 métiers porteurs du Big data en 2020-2021", *data-transitionnumerique.com*, septembre 2020

"Les data centers sont un socle en matière de développement régional", *forbes.fr*, juin 2020

"Lumibird. De la voiture au médical, une histoire de laser", *Investir, Le Journal des Finances*, janvier 2021, p.13

"Métiers du Big data : tout savoir sur les métiers de la donnée", *lebigdata.fr*

"Zoom sur les Smart Cities, ces villes qui connectent leurs infrastructures aux voitures", *frandroid.com*, octobre 2020

# LEXIQUE

## Algorithme

Séquence informatique d'instructions et d'opérations sur des données pour aboutir à un résultat.

## Analytics

Décryptage de données structurées ou non à l'aide d'algorithmes afin d'en retirer des informations et des conclusions opérationnelles.

## Cloud computing

Infrastructure dans laquelle la puissance de calcul et le stockage sont gérés par des serveurs distants. Il s'agit d'une externalisation de ses outils numériques par le client.

## Data mining

Concept à la base du *big data*. Extraction massive de données et incorporation de ces dernières dans des outils de traitement dédiés.

## Dark data

Données disponibles facilement mais ignorées et non exploitées jusqu'ici, par absence de connaissance sur leur potentiel ou de moyens techniques suffisants.

## Data-driven

Littéralement, "guidé par la donnée." Se dit d'une stratégie, d'une approche, d'une solution où la donnée s'avère centrale et oriente les décisions.

## Edge computing

Architecture informatique dans laquelle les données sont traitées de façon décentralisée au niveau des terminaux où elles sont générées (ou par un serveur local). Elles sont ensuite transmises à un serveur de stockage central.

## IoT (Internet of Things)

Phénomène basé sur la multiplication des biens intégrant des fonctionnalités numériques. Potentiellement connectés entre eux et à Internet, ils collectent de nombreuses données et peuvent alimenter les outils de *big data*.

## Machine learning

Processus de renforcement et d'auto-apprentissage des algorithmes par traitement d'un grand nombre de données. Il se place à la base de l'amélioration des performances des intelligences artificielles.

## Machine-to-machine (M2M)

Se dit de technologies servant à assurer l'échange d'informations et la coordination directement entre robots ou interfaces numériques, sans passer par une intervention humaine.

## Smart city

Ensemble de solutions fondées sur la donnée et la connectivité liées au milieu urbain. Il recouvre des thématiques comme la mobilité, l'énergie ou encore la sécurité.

## Think tank

Organisme de réflexion spécialisé sur des sujets économiques, sociétaux, technologiques... Ses membres produisent des notes et des études pour promouvoir des idées et des propositions dans le débat public et auprès des décideurs.

## Vehicle-to-grid

Méthode de gestion du système énergétique dans lequel les véhicules électriques servent d'unités de stockage de l'électricité. Ils en injectent dans le réseau depuis leurs batteries en cas de pic de consommation.



# La collection IndexPresse *Business Etude*

Comment accéder à des données fiables, pertinentes et surtout synthétisées, alors que l'information n'a jamais été aussi accessible en apparence ?

Voilà une question à laquelle sont confrontés quotidiennement les décideurs dans les entreprises lorsqu'il s'agit de prendre les bonnes décisions.

C'est pourquoi nous avons créé la collection **IndexPresse Business Etude**, des études sectorielles complètes, réalisées à partir des plus grands titres de la presse

économique et professionnelle. En s'appuyant sur des informations fiables et de qualité, les études d'IndexPresse offrent des synthèses analytiques et éclairées sur les secteurs d'activité émergents ou en mutation.

Vous aurez ainsi toutes les clés en main pour accompagner votre réflexion stratégique, en vous appuyant sur l'examen des enjeux de votre marché, afin d'anticiper ses évolutions et valider, ou modifier, votre positionnement dans le jeu concurrentiel.

## **IndexPresse** *Business Etude*

Date de parution - avril 2021.



**Bertrand PERRI**

bertrand.perri@indexpresse.fr

Auteur

Étude rédigée en collaboration avec **Renaud Hammamy**

Avec un chiffre d'affaires des logiciels et services "analytics" qui a triplé au cours de la période 2016-2020, le marché du *big data* est en plein essor en France depuis le milieu de la décennie 2010-2020. Le *big data* trouve des usages nouveaux dans des secteurs où il n'était jusque-là peu ou pas déployé. Son potentiel de développement est encore conséquent. Face à ces perspectives et aux nouveaux besoins des entreprises, de nombreuses start-up se positionnent dans le secteur pour proposer une offre spécialisée.

Comment les entreprises françaises abordent-elles le déploiement du *big data* ? Quels leviers actionnent-elles pour le mettre en place ? Quelles difficultés rencontrent-elles ? Dans quels secteurs le *big data* est-il à un stade avancé ? Sur quels marchés offre-t-il des perspectives de développement ? À quelles transformations est-il voué dans l'avenir ?

Cette étude apporte des éléments de réponse et de réflexion pour comprendre les enjeux et les perspectives du *big data*, décrypter ses usages et ses applications potentielles, et identifier les orientations à venir pour un déploiement plus massif en phase avec les défis actuels.

Photo de couverture : ©ici se trouve le copyright de la photo de couverture

