WENVISION

[sfeir]

DataOps

Revoir l'industrialisation et la gouvernance des données pour en tirer toute la valeur

Par Olivier Rafal

Sommaire

Didier Bove

DSI de Veolia

La nomination, il y a quelque temps, de Chief Digital Officers était l'expression d'une forme d'échec de la DSI, de ce qu'elle aurait dû faire. La situation s'est résorbée, mais on voit une résurgence de ce phénomène liée à la donnée, avec la nomination de Chief Data Officers, qui créent des applications, des produits, qui communiquent… Bref, qui recréent une entreprise à côté de l'entreprise.

Bien sûr, il y a un souci de gouvernance de données dans les entreprises, mais ce n'est pas en créant un poste qu'on le résoudra. Ni en achetant un outil. Classiquement, dans l'IT, 80 % des efforts passent dans le choix de l'outil, alors qu'il faudrait les faire porter sur les processus et l'organisation. Mais c'est évidemment beaucoup plus compliqué.

De notre côté, nous avons revu en profondeur la façon dont nous abordons la gouvernance et la valorisation de la donnée. Notre objectif est que les BU disposent de capacités opérationnelles locales tout en tirant parti de services créés en central (reporting, supervision…).

Nous sommes en effet une entreprise décentralisée, avec 12 zones géographiques et des activités qui peuvent s'avérer relativement différentes d'une zone à l'autre, en fonction de l'historique, des contraintes légales locales, etc. Cela n'aurait pas vraiment de sens de capter et stocker toutes les données. La période où on mesurait l'efficacité d'une DSI à la taille de son datalake est révolue !

L'enjeu aujourd'hui est de ne stocker que ce qui a du sens, ce qui nous permet de produire quelque chose d'utile pour les métiers sur le terrain ; nous sommes passés de "big is beautiful" à "useful is beautiful". Reboucler avec le terrain, conserver une certaine proximité, est fondamental. Il faut que nos applications produisent de la valeur pour la personne qui a émis la donnée au bout de la chaîne. Cela initie un cercle vertueux : la personne qui saisit la donnée fera beaucoup plus attention à sa qualité si elle sait qu'en retour, elle en obtiendra un bénéfice.

Pour y parvenir, il faut bien répartir les rôles : côté IT, nous assurons la partie technique, le nettoyage et le cycle de vie de la donnée, tandis que la partie fonctionnelle peut être gérée par les métiers.

Le passage au Cloud nous a d'ailleurs beaucoup aidé à concrétiser cette façon de fonctionner ; on peut très facilement démontrer aux métiers l'intérêt de certaines analyses, sans limitation, puisqu'on peut lancer des requêtes sur des centaines de millions de lignes sans se préoccuper de l'infrastructure sous-jacente, de la sauvegarde, etc.

Il faut également partager les mêmes définitions. Nous avons donc établi un référentiel d'objets métier fonctionnels (pompe, citerne, hauteur d'eau, etc.) mais aussi propres au service finances ou aux ressources humaines.

Enfin nous avons défini au niveau de l'IT groupe des axes stratégiques, porteurs de valeur, pour lesquels nous créons des "solutions digitales". Ces priorités découlent de la "raison d'être" définie en avril 2019 par Veolia, qui veut agir pour l'environnement. Elles orientent nos efforts mais aussi le discours : nous essayons de ne plus raisonner en termes de coûts IT mais de création de valeur. Il faut qu'au minimum deux pays demandent une solution pour qu'on initie le projet ; ils doivent s'entendre sur la valeur attendue de la solution et c'est cela qui va engager le travail de la donnée.

Une fois que tout cela est mis en place, que la donnée est propre, que les objets sont bien définis, on peut commencer à agréger d'autres types de données (venant des usines, notamment) et commencer à les exploiter, avec du machine learning. Cette partie "data science" est la cerise sur le gâteau. Elle arrive en dernier mais a le plus de saveur.

On peut réaliser des choses relativement simples, qui facilitent la vie des opérationnels sur le terrain. Nous avons par exemple conçu une application qui reconnaît les valeurs d'un index de consommation sur une photo (grâce à de l'IA, car les compteurs d'eau sont très peu normalisés). Nous avons aussi commencé à déployer un service d'aide à la saisie de formulaires, qui réduit le nombre de champs en fonction du contexte.

Nous travaillons aussi sur des algorithmes plus complexes, pour optimiser l'efficacité énergétique de nos stations d'épuration, par exemple. Lors-que nous parvenons à gagner quelques pourcents, tout le monde y gagne : Veolia, qui économise de l'énergie, le client, qui voit sa facture baisser, et l'environnement.

Dans une entreprise intelligente, la donnée circule à travers toute l'entreprise, l'intelligence artificielle est utilisée dans tous les départements, tout est connecté.

Jean-Paul Agon, Chairman & CEO L'Oréal

Source : https://www.bloomberg.com/news/videos/2021-01-14/leaders-with-lacqua-jean-paul-agon-l-oreal-chairman-ceo-video

Avant-propos

Didier Girard, co-CEO, VP Engineering de SFEIR
et Founder WENVISION, PhD Machine Learning

La période n'a jamais été plus propice pour tirer parti de la donnée.

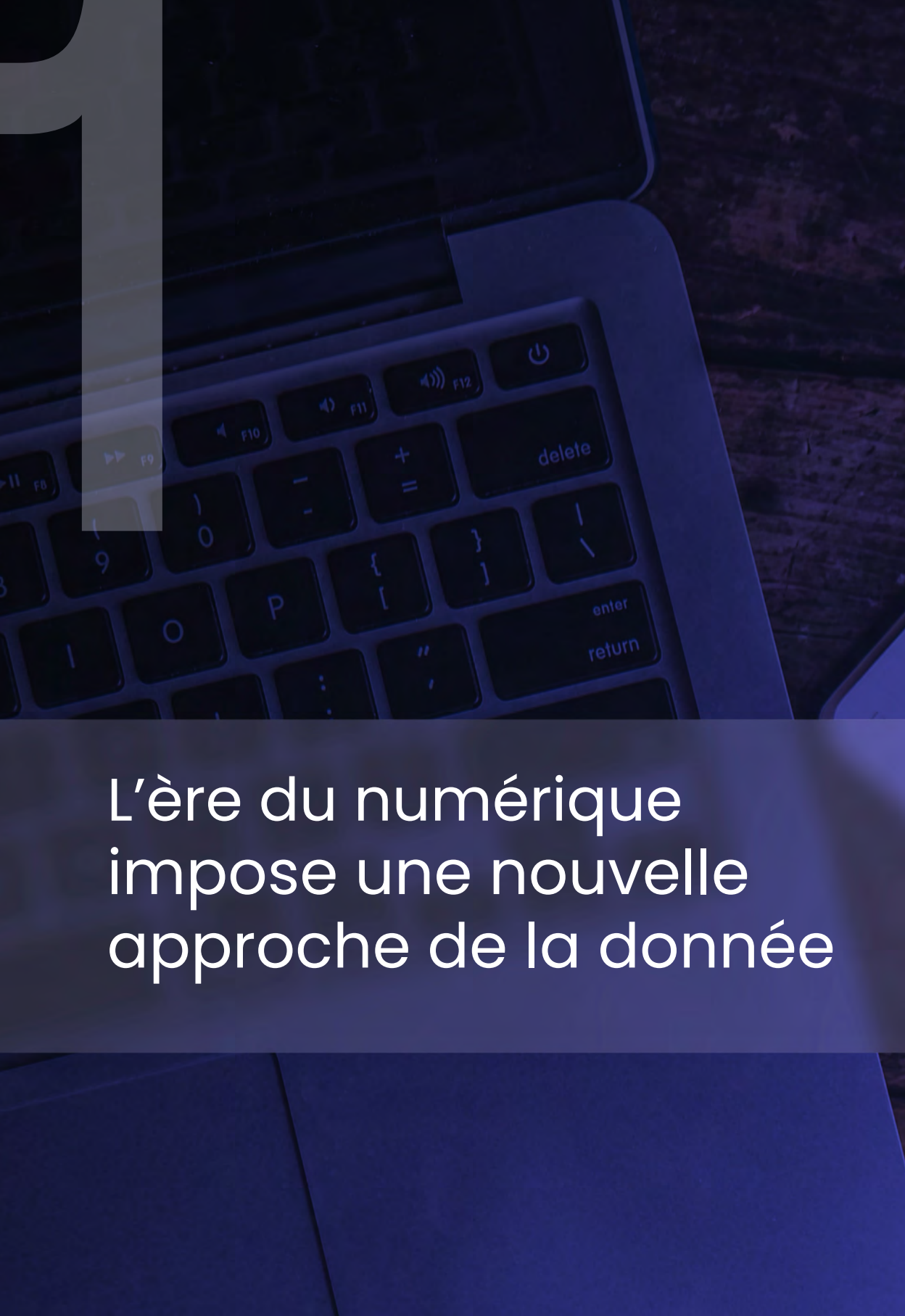
Elle existe en abondance. Plus on l'utilise, plus on la partage, plus elle se multiplie et crée de la valeur.
Le Cloud, ensuite, est la plateforme idéale pour valoriser la donnée. Il y a encore une dizaine d'années, on ne disposait pas forcément des moyens nécessaires pour analyser de grands volumes de données, en temps réel, ou bien cela s'avérait très complexe et coûteux à monter puis à maintenir.
Le Cloud simplifie énormément les problématiques de stockage, de traitement ou encore de partage de la donnée au sein d'un écosystème.
Il ne manque qu'une stratégie d'entreprise, une volonté d'avancer et de valoriser ses données, ainsi qu'un mode d'emploi - ou à tout le moins un accompagnement pour mettre en place les conditions du succès.

Car c'est l'ensemble de ces dimensions qui crée de la valeur, et pas seulement la technologie, la pertinence d'un algorithme ou le talent du data scientist. Valoriser la donnée, c'est un travail d'équipe, qui doit impliquer :

- **la direction générale de l'entreprise,** qui impulse la stratégie ;
- **un chef d'orchestre,** qui met cette stratégie en musique en s'appuyant sur les métiers et l'IT ;
- **l'IT,** dont on attend qu'elle mette en place une plateforme robuste et agile ;
- **les métiers,** qui détiennent le savoir sur la donnée et en sont les premiers bénéficiaires.

Tous les maillons sont essentiels pour réussir sa stratégie data. C'est cette approche collaborative que nous avons voulu mettre en avant ici, à travers ce livre blanc sur le DataOps.

Bonne lecture !



L’ère du numérique impose une nouvelle approche de la donnée


DAY OFF
BEAM 4-7 PM
DAY OFF
4-6 PM ELBM OPEN HOUSE
6PM KOG

La donnée en entreprise, c'est un peu comme l'attention centrée sur le client : tout le monde dit que c'est essentiel, mais au final, qui parvient vraiment à traduire ses intentions dans les faits ? Comment valider l'intuition que c'est dans la donnée que réside le levier qui permettra à l'entreprise d'optimiser ses processus, de gagner quelques pourcents sur chacune de ses transactions, de se lancer dans de nouveaux business models…

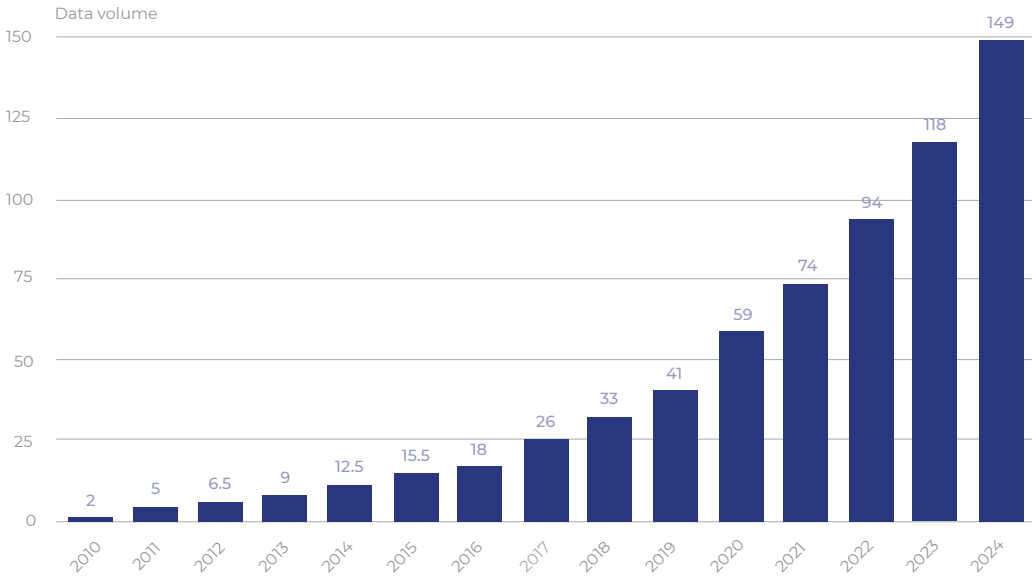
La grande époque du décisionnel a démontré son utilité. Mais ce n'est certainement pas en restant ancré dans les principes forgés au moment où il fallait mettre en place ce décisionnel que la donnée pourra révéler son plein potentiel. Il n'est pas question de remettre la BI et ses outils en cause ; il y aura toujours besoin de regarder dans le rétroviseur, d'analyser des chiffres de vente, de comparer des performances, etc.
De même, tout le monde n'a pas vocation à devenir "data analyst" ou statisticien. La majorité des acteurs d'une entreprise resteront des consommateurs de données ; il y aura donc aussi toujours besoin de fabriquer et d'envoyer des rapports.

Il faut cependant aussi savoir s'extraire de ce mode de pensée, qui a fini par structurer à la fois les processus, les métiers et les architectures technologiques autour de la donnée.
Une structure qui n'est plus adaptée aux challenges posés par l'économie numérique.

Repenser les usages de la donnée à l'aune du numérique

« La révolution numérique, ce n'est pas le smartphone ou l'objet connecté. C'est tout transformer en données pour tout transformer par les données. »
C'est par ces mots que Thierry Happe, cofondateur de Netexplo, a lancé les Assises de la Data Transformation, qui ont réuni de nombreux dirigeants, DSI et CDO de grandes entreprises, en janvier 2021. L'image peut paraître radicale, mais elle résume plutôt bien notre monde, où quasiment chacune de nos interactions produit de la donnée.

On pense bien sûr en premier lieu aux réseaux sociaux (Twitter compte 6000 nouveaux tweets par seconde !) ou aux plateformes de contenu. En 2019, Youtube a dépassé les 500 heures de vidéo ajoutées chaque minute. On pense aussi bien sûr aux sites Web, aux smartphones, aux autres applications en ligne… Mais c'est oublier que toute l'activité économique s'est numérisée, dans tous les secteurs, tous les métiers. Dans nos activités professionnelles aussi bien que nos activités personnelles, ce sont toutes nos actions qui produisent des données et déclenchent des événements enregistrés dans différents systèmes : créer ou consulter un document, enregistrer un nouveau client, réaliser ou écouter un podcast, produire ou visionner une série, vendre ou acheter en ligne ou en magasin, retirer de l'argent à un distributeur, prendre rendez-vous chez un médecin, établir un diagnostic, emprunter les transports en commun, se promener dans la rue…



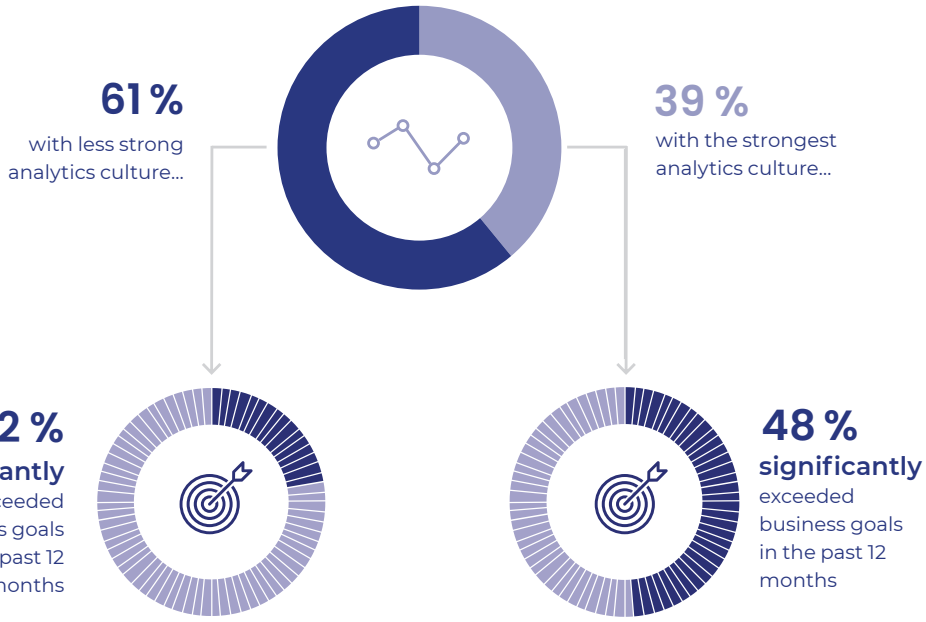
Volume de données créées, capturées, copiées et consommées dans le monde, en zettaoctets (millions de pétaoctets), entre 2010 et 2024[1].

Du point de vue de l'entreprise, toutes ces données représentent un actif extraordinaire. Là aussi, tous les métiers sont concernés, peuvent s'emparer du sujet et s'appuyer sur la donnée, qu'il s'agisse d'améliorer la connaissance de ses clients, d'analyser ses performances, d'optimiser un processus, de réaliser des prévisions plus justes… La plupart du temps, quand la donnée n'est pas utilisée dans l'entreprise, ce n'est pas parce qu'elle manque ; c'est parce que les collaborateurs n'ont pas conscience de ce que cela peut leur apporter. Dans certains cas, c'est parce que la DSI n'aura pas mis en place les moyens appropriés pour s'appuyer sur cette donnée. Mais soyons honnêtes : si l'IT pêche de ce point de vue, c'est généralement par manque d'instructions claires et de support de la part de la direction générale.

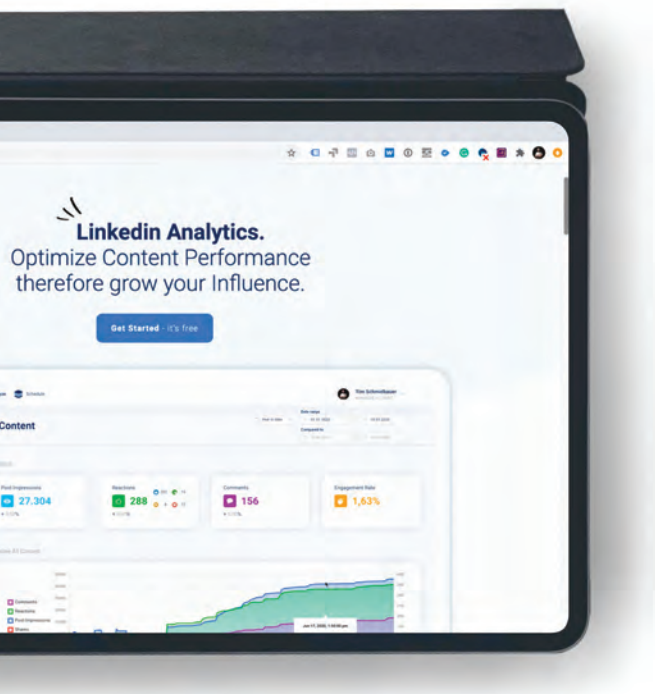
Devenir une entreprise data-driven doit se décréter au plus haut niveau… et être suivi dans les faits par un accompagnement et un support forts. Depuis les dirigeants de l'entreprise jusqu'aux opérationnels métiers sur le terrain, tous doivent prendre conscience de la nécessité de valoriser la donnée.

Aux Assises de la Data Transformation, c'est Claire Pedini, responsable à la fois des ressources humaines de Saint-Gobain et de la transformation numérique du groupe, qui insistait sur cette priorité : *« Soyons obsédés par la donnée, on ne l'est pas assez, tous les process de l'entreprise doivent être revus sous l'angle de la data. »* Pour cela, pas de secret, tous les collaborateurs doivent en être persuadés, dit Claire Pedini : *« Avant, il fallait que tout le monde parle anglais. Aujourd'hui, il faut être data-fluent, que tout le monde dans l'entreprise sache parler data. »*

(1) Source : Statista (https://www.statista.com/statistics/871513/worldwide-data-created/)



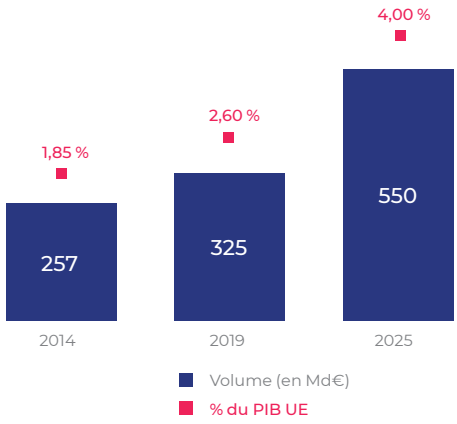
Étude Deloitte “Becoming an Insight-Driven Organization”, 2019[1].



Dans une étude de 2019 auprès de plus de 1000 personnes au sein de grandes entreprises (au-delà de 500 employés), Deloitte souligne les bénéfices d'avoir créé une culture de l'analyse de données ; les entreprises s'appuyant sur la donnée sont plus de deux fois plus nombreuses que les autres à dépasser leurs objectifs (cf. ci-contre). Cette nécessité de s'appuyer sur la donnée est évidemment renforcée par la pandémie, qui a accéléré la bascule vers les activités en ligne et fragilisé nombre d'entreprises. Ainsi, pour McKinsey, la donnée est un levier clé pour sortir de la crise.
Dans un rapport publié en août 2020 (The recovery will be digital), le cabinet recommande d'analyser *« plusieurs sources de données, sur une base hebdomadaire (ou plus fréquente) afin d'évaluer les besoins changeants de ses clients et ses partenaires - ainsi que ses propres performances »*.

La donnée est aussi mise en avant par l'Union européenne comme un levier de compétitivité. Les économistes de l'Union européenne estiment que la création de valeur basée sur l'analyse et la vente de données dans l'Europe des 27 a représenté 325 milliards d'euros en 2019 (400 milliards avec le Royaume-Uni), soit 2,6 % du PIB global. Surtout, la croissance de cette économie de la donnée est bien plus forte que la croissance moyenne du PIB : en pourcentage du PIB, elle représentait seulement 1,85 % en 2014 et pourrait atteindre 4 % en 2025 (prévisions établies avant la crise du Covid, laquelle a certes ralenti l'activité économique mais encore accéléré la transformation numérique).

Un scénario optimiste, dit "de forte croissance", table même sur un taux composite de croissance annuelle de 11,5 % de l'économie de la donnée entre 2020 et 2025 - si les États libèrent les données, les valorisent, et que les entreprises réalisent les investissements nécessaires, dans les compétences, les organisations et les technologies, notamment l'intelligence artificielle. La politique d'Open Data en France, par exemple, contribue activement à la croissance de cette économie. Elle peut même être non marchande mais créer de la valeur, comme on l'a vu avec Covid Tracker et Vite Ma Dose.

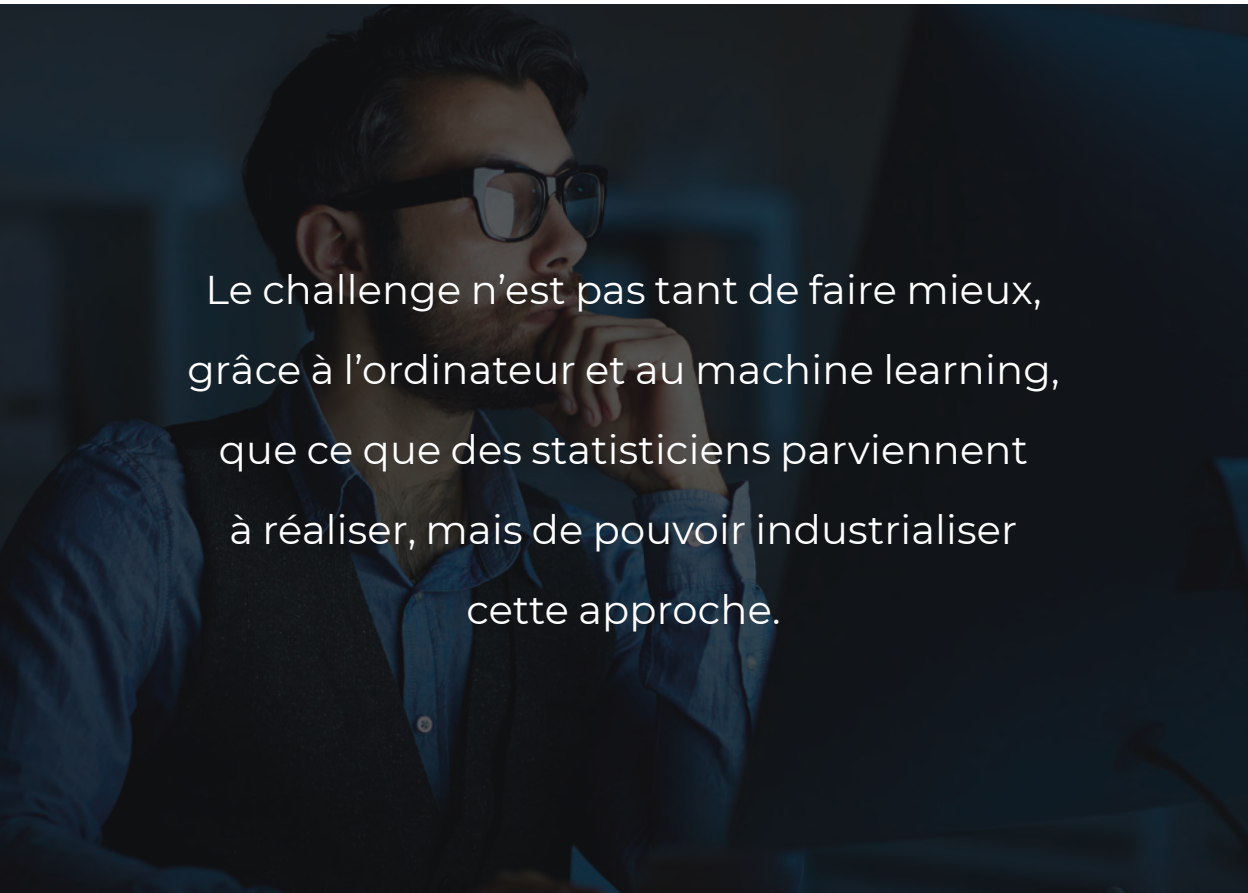


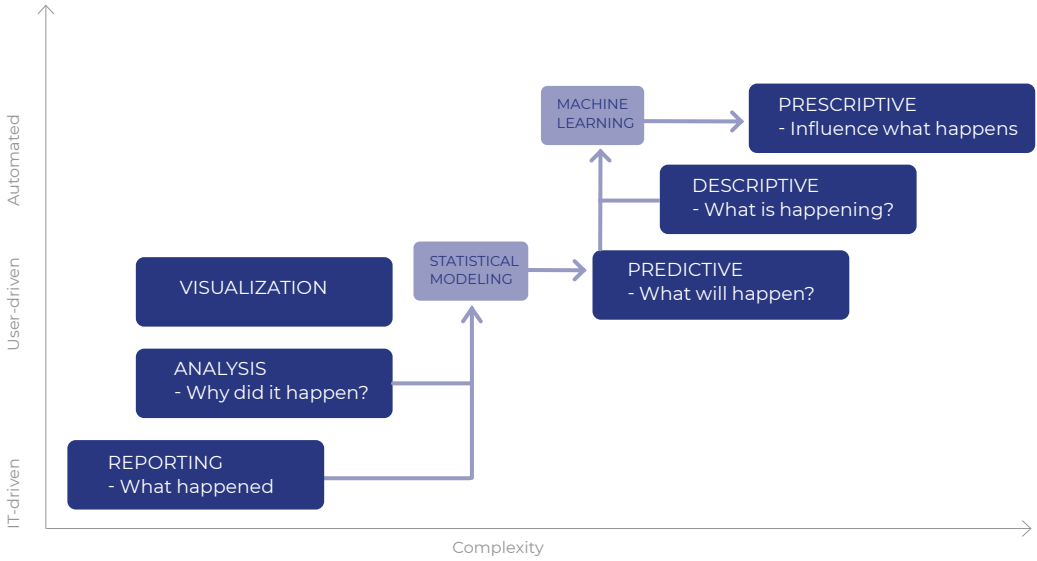
Estimations de l'économie de la donnée au sein de l'UE27[2].

(1) Source : https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html
(2) Source : UE & IDC (https://datalandscape.eu/study-reports/final-study-report-european-data-market-monitoring-tool-key-facts-figures-first-policy)

McKinsey recommande aussi de s'appuyer sur la donnée pour investir dans l'intelligence artificielle afin de reconstruire des modèles adaptés à une ère numérique, post-Covid : *« De la même manière que de nombreuses entreprises ont dû reconstruire les modèles financiers et de risque qui ont échoué lors de l'effondrement financier de 2008, les modèles devront également être remplacés en raison des changements économiques et structurels massifs causés par la pandémie. Par exemple, les modèles qui utilisent des séries chronologiques, des prix du pétrole ou des données sur le chômage devront être entièrement reconstruits.*

Les données doivent également être réévaluées. À mesure que les entreprises construisent ces modèles, les équipes d'analystes devront probablement rassembler de nouveaux ensembles de données et utiliser des techniques de modélisation améliorées pour prévoir la demande et gérer les actifs avec succès. »



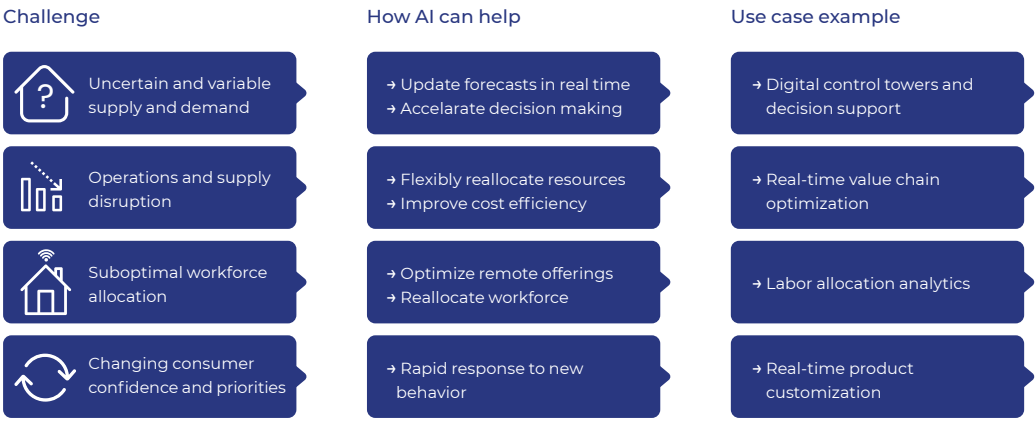


Le paysage de l'analytique selon le cabinet 451 Research

Le challenge n'est pas tant de faire mieux, grâce à l'ordinateur et au machine learning, que ce que des statisticiens parviennent à réaliser, mais de pouvoir industrialiser cette approche, de façon à traiter davantage de données, plus souvent, s'attaquer à des problèmes plus complexes… Jusqu'à, in fine, automatiser un certain nombre de traitements.

Cette progression des techniques analytiques depuis la BI façon rétroviseur jusqu'à l'automatisation et l'intelligence artificielle est résumé par le cabinet d'analyse 451 Research dans le graphe ci-dessus.

De la modélisation "avancée" à l'automatisation de processus opérationnels, les techniques d'intelligence artificielle peuvent aider l'entreprise à s'adapter aux changements, à la croissance du nombre de sujets à traiter et à l'accélération des interactions. Dans le retail, par exemple, les acteurs doivent tout à la fois assurer un renouvellement plus fréquent de leurs gammes, optimiser la mise en rayon, être capable d'échanger avec les clients et de vendre sur tout un éventail de canaux (magasin, site Web, applications mobiles, drive, etc.), multiplier les promotions pertinentes, individualiser les communications par courriel… Impossible de faire l'ensemble de ces tâches, de façon correcte, à un rythme élevé, sans s'appuyer sur des outils se nourrissant de données et de modèles analytiques.



Exemples de recours à l'IA pour résoudre des challenges actuels selon le BCG.[1]

De la même manière, l'introduction d'outils d'analyse et d'applications automatisant certains traitements dans le secteur public permet aux agents concernés de se libérer du temps pour traiter de façon plus approfondie certains dossiers, recevoir le public, etc. En filigrane, on entend bien aussi les risques de ne pas se préoccuper réellement de cet actif. Des collaborateurs débordés, démotivés car ils n'ont pas le temps de faire leur travail correctement. Des processus opérationnels qui coûtent de l'argent, sans qu'on sache très bien comment les améliorer. Des clients qui restent une masse anonyme, alors que les concurrents leur proposent une expérience personnalisée…

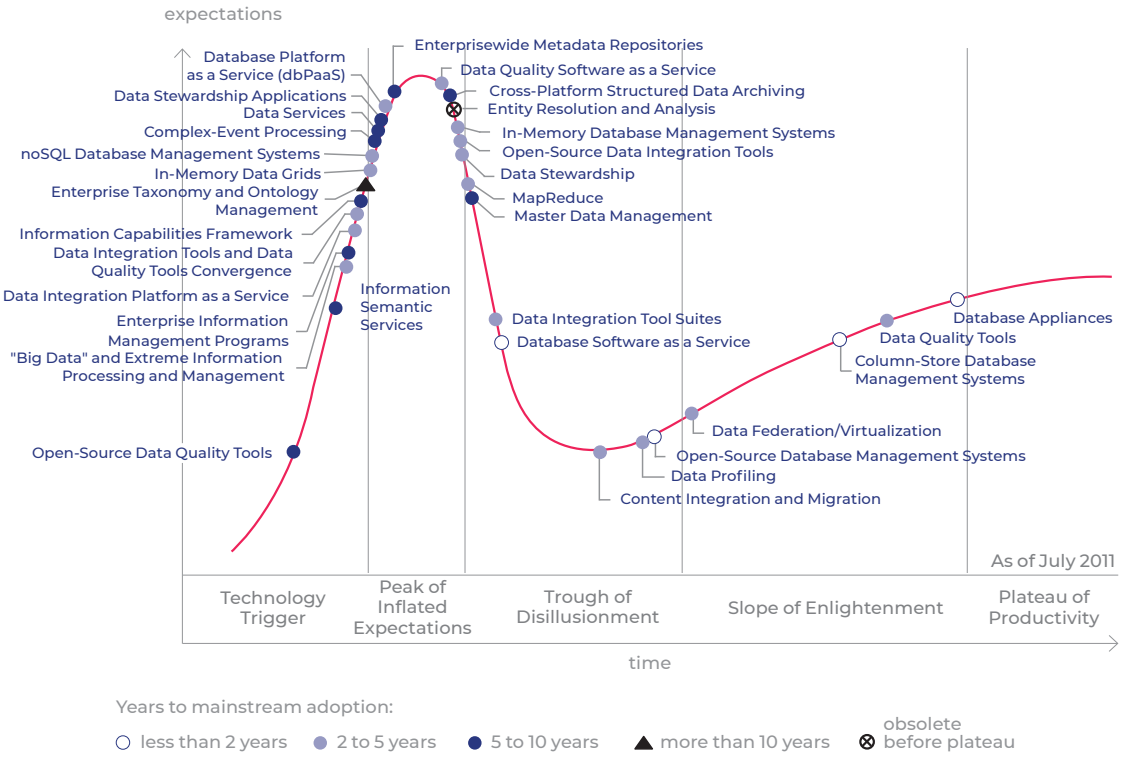
On a longtemps glosé sur le vilain terme d'ubérisation. Toujours est-il que dans la réalité, c'est bien la donnée et une façon intelligente de l'utiliser qui ont permis à des acteurs comme Uber, Netflix, Google ou Amazon de s'imposer. Au détriment d'acteurs pourtant bien implantés, qui ont tardé à saisir l'importance des changements induits par le numérique.

La donnée est au cœur de cette transformation numérique ; c'est même ce qui a conduit nombre d'entreprises dans les années 2010 à investir massivement dans le Big Data.

Apprendre des années "Big Data"

Quelle entreprise n'a pas construit son cluster Hadoop au début des années 2010 ? Le Big Data était partout, il montait fièrement à l'assaut du pic des attentes exagérées de Gartner (cf. ci-contre).
Chacun rivalisait pour construire son lac de données, beaucoup plus versatile et moins cher que les entrepôts de données, construits pour un type de données et un usage bien précis. C'était aussi la frénésie pour former et recruter des "data scientists", métier qualifié *"le plus sexy du 21e siècle"* par la Harvard Business Review en octobre 2012.

(1) Source : https://www.bcg.com/fr-fr/publications/2020/business-applications-artificial-intelligence-post-covid



Positionnement des technologies “big data” dans l’édition 2011 du Hype Cycle for Data Management, Gartner

Du point de vue du marché, cette effervescence a créé une véritable dynamique. Mais pour les entreprises, le retour sur investissement n’a pas été évident.

Quelques années après les débuts fracassants du Big Data, les cabinets d’analystes multiplient les constats et prévisions alarmistes :

→ **87 %** des projets de data science ne vont jamais en production (VentureBeat, 2019) ;

→ **77 %** des entreprises disent que l’adoption des projets Big Data et IA représente un gros challenge (NewVantage, 2019) ;

→ **60 %** des projets Big Data échouent à aller en production (Gartner, 2016) ;

→ **85 %** des projets Big Data échouent à aller en production (Gartner, 2017).

Il y avait pourtant une véritable justification à cet engouement pour le Big Data. Les entrepôts de données traditionnels étaient, et sont encore, d’onéreuses mécaniques de précision, ingérant une donnée structurée nettoyée, qualifiée, destinée à alimenter des tableaux de bord et outils de BI, construits par l’IT selon les spécifications des métiers. L’analyse de la donnée telle qu’elle se pratiquait jusqu’alors consistait donc essentiellement à regarder dans un rétroviseur des éléments définis deux ans auparavant.

Ce schéma convient parfaitement pour analyser par exemple des chiffres de vente, par produit, par zone géographique, etc. En revanche, tout changement de structure doit passer par un long processus IT. Impossible donc de s'adapter rapidement aux évolutions de l'entreprise, à la montée en puissance de l'économie numérique avec ses concurrents extrêmement agiles, à la multiplication de la donnée… De même, impossible avec les mêmes outils de stocker et d'analyser de la donnée non structurée (texte, vidéo…) pour proposer des services innovants aux collaborateurs ou aux clients.

C'est en réponse à ces limitations que le Big Data apparaît : il s'agit de répondre aux fameux 3 V de l'époque :

- **vélocité :** pour que le business puisse obtenir plus rapidement des réponses ;
- **volume :** pouvoir stocker davantage de données sans se préoccuper du coût ou de la structure ;
- **variété :** pouvoir traiter des données non structurées.

De nombreux clusters de machines Linux se montent, s'appuyant sur les technologies open source Hadoop et MapReduce, complétées à un rythme effréné par la Fondation Apache. C'est une décennie très riche qui s'écoule alors, jusqu'à ce que le journal LeBigData lui-même pose la question : "La fin du Big Data ?"

"L'Apache Software Foundation a déclaré dernièrement l'abandon de 13 projets Apache centrés sur le Big Data comme Apex, Falcon et Sentry. Une page du Big Data se tourne avec ce déclin, relatif, de Apache Hadoop, figure proéminente du secteur jusque-là."

Extrait de l'article "La fin du Big Data ? Apache abandonne une dizaine de projets Hadoop", 15/04/21, LeBigData, https://www.lebigdata.fr/abandon-projets-apache-hadoop

En réalité, ce n'est pas véritablement le Big Data qui prend fin aujourd'hui : il y a toujours besoin, et même plus que jamais, de pouvoir réagir rapidement ou de pouvoir analyser de grands volumes de données, notamment pour créer et entraîner des modèles prédictifs et autres algorithmes de machine learning. Ce qui disparaît - et c'est tant mieux - est cette croyance en une technologie qui d'un coup rendrait les entreprises data-driven. Tous les projets n'ont pas échoué, heureusement. Mais l'analyse de cette décennie est sans conteste en demi-teinte.

On peut distinguer au moins six grandes raisons ayant abouti à ce résultat :

→ **Une très forte complexité.** Les technologies estampillées Big Data étaient complexes à mettre en œuvre et à maintenir. On a beaucoup parlé de la pénurie de data scientists, mais avant cela, il fallait des architectes et des ingénieurs de haut niveau pour concevoir ces architectures de données et les mettre à jour en fonction des incessantes nouveautés proposées par la Fondation Apache et les acteurs du marché. Et même si on parle de serveurs Linux, il fallait des moyens conséquents pour monter et faire tourner une telle infrastructure en entreprise.

→ **Un focus sur la quantité plus que la qualité.** Beaucoup de projets sont tombés dans l'excès inverse de la BI traditionnelle, où toute donnée devait être fiable. La capacité de pouvoir collecter des tonnes de données brutes a conduit à vouloir réaliser des analyses statistiques sur ces grandes masses. La plupart du temps, les résultats ont validé l'adage "garbage in, garbage out" : avec des données de mauvaise qualité en entrée, difficile de produire des résultats de qualité.

→ **L'oubli de la notion de valeur.** De très nombreux projets se sont montés pour stocker de la donnée, toute la donnée, sans véritablement se préoccuper de la finalité. S'il n'y a pas de réflexion en amont sur les cas d'usage, il devient impossible de mesurer la valeur des projets Big Data pour le business, sans même parler de calculer un ROI.

→ **Une gouvernance axée sur la sécurité et la conformité.** La gouvernance des données ne consiste pas seulement à s'assurer que les rôles ont bien été distribués et que chacun n'accède qu'aux données auxquelles il a le droit d'accéder. Au-delà des règles, il s'agit de réorganiser complètement la façon dont les équipes IT, les équipes data et les métiers peuvent travailler ensemble en bonne intelligence pour produire de la valeur.

→ **L'absence d'une stratégie globale.** Une stratégie data-driven ne découle pas subitement d'un projet informatique ; c'est l'infrastructure de données qui doit être conçue en fonction de la stratégie de l'entreprise et des impératifs business. Le rôle crucial de la donnée doit être clair pour tout le monde, c'est ce qui déterminera la gouvernance, l'organisation, les processus et l'infrastructure technologique.

→ **Une approche trop centralisatrice.** Vouloir tout mettre au même endroit, qu'il s'agisse des données ou des ressources humaines, peut s'avérer payant dans certains cas, mais cela ne devrait pas être une règle. Abolir les silos n'est pas forcément synonyme de tout regrouper ; la donnée peut être mieux exploitée par des technologies et des équipes différentes, le tout étant de savoir construire des ponts et d'avoir une vision globale.

Cette décennie Big Data aura au moins mis en évidence les nouveaux besoins des entreprises et des utilisateurs. Car si les projets n'ont pas forcément délivré tout leur potentiel, le besoin demeure. Il se fait même plus pressant, avec le transfert croissant de nos activités en ligne et le passage d'un monde centré sur la fabrication et la vente de produits à un monde qui se focalise davantage sur le service et la satisfaction client.
Nous avons tous observé en tant que consommateurs l'importance de se voir proposer des courriers individualisés, des offres personnalisées ou encore des services d'abonnement pour des objets dont l'usage nous intéresse plus que la possession. Cette "servicisation" de l'industrie, qu'on retrouve aussi sous le nom d'économie de la souscription, nécessite une attention toute particulière portée aux clients.

Dans les entreprises orientées clients, les métiers doivent être particulièrement réactifs. Ils doivent pouvoir s'adapter aux évolutions technologiques, aux changements d'habitudes, aux modes, à l'apparition de nouveaux concurrents... Dans les télécoms ou la banque, par exemple, les acteurs en place ont dû lancer des structures de zéro, en mode start-up, pour garantir cette agilité, et tabler sur l'analyse de données pour concevoir et ajuster les meilleures offres possibles.

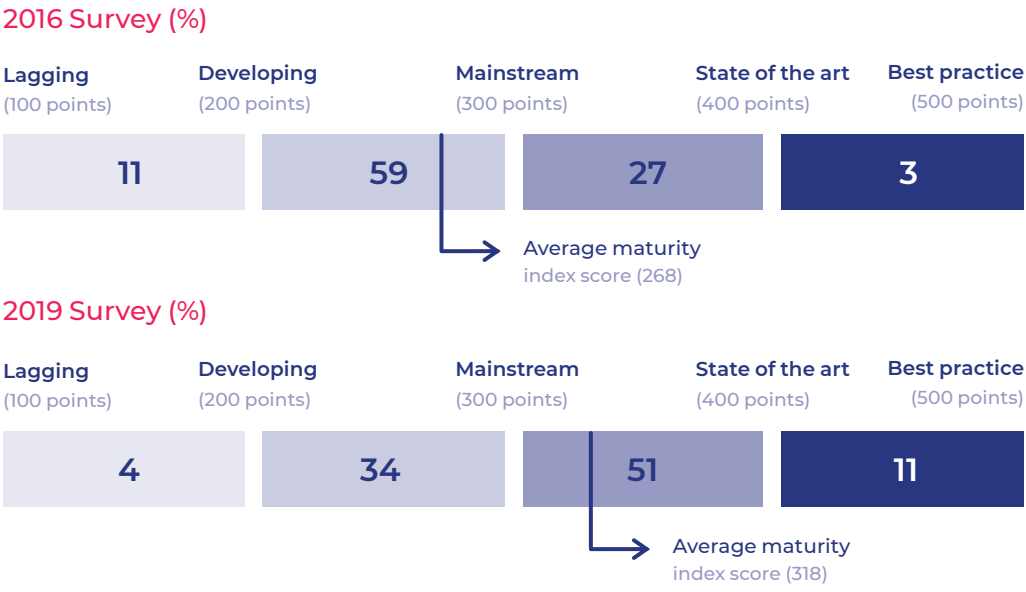
Pour répondre efficacement à ces enjeux, les métiers doivent être le plus autonomes possible avec les données. Là aussi, les années Big Data ont vu l'émergence d'un phénomène souvent poussé à l'extrême, qui a de même montré ses limites : la démocratisation de l'accès aux données.

Ce qui est à la base une bonne idée s'est souvent traduit dans les faits par une BI self-service anarchique, avec des jeux de données poussés aux utilisateurs et ouverts à toutes les interprétations. Le BCG fait le constat d'une maturité assez faible en 2016, qui aurait bien progressé en 2019 avec un score passé de 268 à 318 sur 500 (cf. ci-dessous).

De façon beaucoup plus empirique, ce que nous constatons sur le terrain, c'est que cette maturité reste relativement faible, ce qui est cause de beaucoup d'incompréhensions et de frustrations. Les équipes IT sont accusées par les utilisateurs de ne pas aller assez vite, de ne pas proposer de solutions amenant de la valeur. Les utilisateurs sont accusés de manquer de maturité, de précision dans leurs demandes, mais aussi de compétences pour manipuler et comprendre la donnée.

Tandis que les directions générales se demandent pourquoi leur entreprise n'est toujours pas data-driven.

Résoudre ces sujets de tension ne se fera pas du jour au lendemain. Il est donc urgent de s'attaquer à ce problème, pour aborder sereinement la décennie qui commence et les prochaines années. De bons choix en matière d'architecture et de gouvernance au sens large constitueront déjà un bon début pour industrialiser l'approche de la donnée.



Enquête BCG sur la maturité des entreprises en matière de gouvernance de la donnée[1].

(1) Source : https://www.bcg.com/publications/2019/rough-road-to-data-maturity

Laurent Ostiz

Global Chief Data Officer, Adeo

Il y a ceux qui vendent du rêve, qui voient la data comme une baguette magique. Ceux qui tentent de faire peur, en ne parlant que de biais algorithmiques, de vie privée et d'enfermement. Ou encore ceux qui viennent vous conseiller les 10 algorithmes que toute entreprise devrait mettre en place pour être performante. Je préfère une approche un peu plus scientifique, et considérer la donnée comme un outil pour résoudre des problèmes. Cela évite de tomber dans le travers de faire de la data pour faire de la data. Si on formule ce qu'on essaie de faire sous la forme d'un problème à résoudre, alors on saura mesurer si on réussit à le faire.

Ma deuxième conviction, c'est que ce problème doit venir des métiers. On peut influencer, inspirer, mais surtout pas pousser une solution ex nihilo. Il faut aller voir les métiers, comprendre leurs problèmes, faire ressortir les sujets qui les préoccupent le plus : des informations qui manquent, des tâches qui pourraient être optimisées, voire automatisées… Si le sujet n'est pas une priorité du métier, s'ils ne se cassent pas les dents dessus depuis un certain temps, il n'y aura pas d'utilisation dans la durée.

Mais le plus grand défi n'est pas là. Une fois que vous avez créé votre 'produit data', vous n'avez fait que la moitié du chemin : il faut penser à l'expérience utilisateur. Faire en sorte qu'il serve vraiment. S'il s'agit de mettre en place de l'automatisation, il n'y a pas de souci, la démarche ira au bout. Mais si la data doit aider les métiers à prendre des décisions, alors il faut penser à la façon dont on la présente et dont elle sera utilisée. Même si votre algorithme est plus malin que les règles de gestion traditionnelles, cela ne sert à rien de présenter l'information dans un rapport ou un dashboard à côté si l'utilisateur n'a pas l'habitude de manipuler ce type de données. Le défi, et on l'a appris en chemin, c'est de rendre cela familier, de l'intégrer dans un processus métier adapté. C'est cela, le sens du MLOps : le machine learning devient une fonctionnalité d'un produit digital.

Enfin, il y a un dernier élément qu'on oublie souvent : s'assurer que la donnée présentée sera utile. Il faut mesurer ce que les gens font de cette information, se nourrir de ça et reboucler. Pour faire progresser l'algorithme, bien sûr, mais aussi pour ajuster l'UX, le moment ou encore la façon dont on fournit cette information.

Il n'y a pas de réponse absolue à la question "comment réussir sa stratégie data". La réalité d'une entreprise, c'est qu'on ne peut pas tout faire ; la stratégie data doit s'inscrire dans la stratégie de l'entreprise. De notre côté, nous ne considérons nos produits data comme des succès que lorsque l'usage de la donnée s'inscrit dans la durée, que les équipes métier l'utilisent de façon régulière et familière.

Cela paraît très simple, dit comme ça, mais parfois les évidences mettent du temps à remonter à la surface. Nous l'avons appris au fur et à mesure. Parce qu'au début, on se concentre sur la partie de l'iceberg qui n'est pas la plus importante. Et c'est tout à fait normal, il y a de nombreux challenges techniques et organisationnels à gérer lorsqu'on démarre : les datalakes, la plateforme MLOps, les outils, la qualité de la donnée, la gouvernance… Si bien qu'on peut se perdre en route.

Notre organisation évolue également. Nous sommes partis d'un modèle très centralisé, mieux adapté pour créer la plateforme et bâtir nos outils et nos process. Aujourd'hui, nous rapprochons les équipes data des équipes métier. C'est plus complexe à gérer, mais je suis convaincu que cela apporte de la valeur : les équipes créent de la connivence, il n'y a même plus besoin de faire remonter les sujets.

En revanche, la plateforme reste commune, de même que l'outillage : pour des raisons de partage, d'économies d'échelle, de gestion centralisée du cycle de vie de la donnée, d'expertise… Toutes les enseignes partagent l'essentiel de leurs données sur cette plateforme, peuvent venir se brancher dessus et utiliser le moteur de recherche pour trouver les données dont elles ont besoin. Et vu le prix du stockage des données froides dans le Cloud, on peut se permettre de conserver l'historique de toutes ces données, de façon à être plus pertinents lorsque les métiers ont des besoins spécifiques.

2

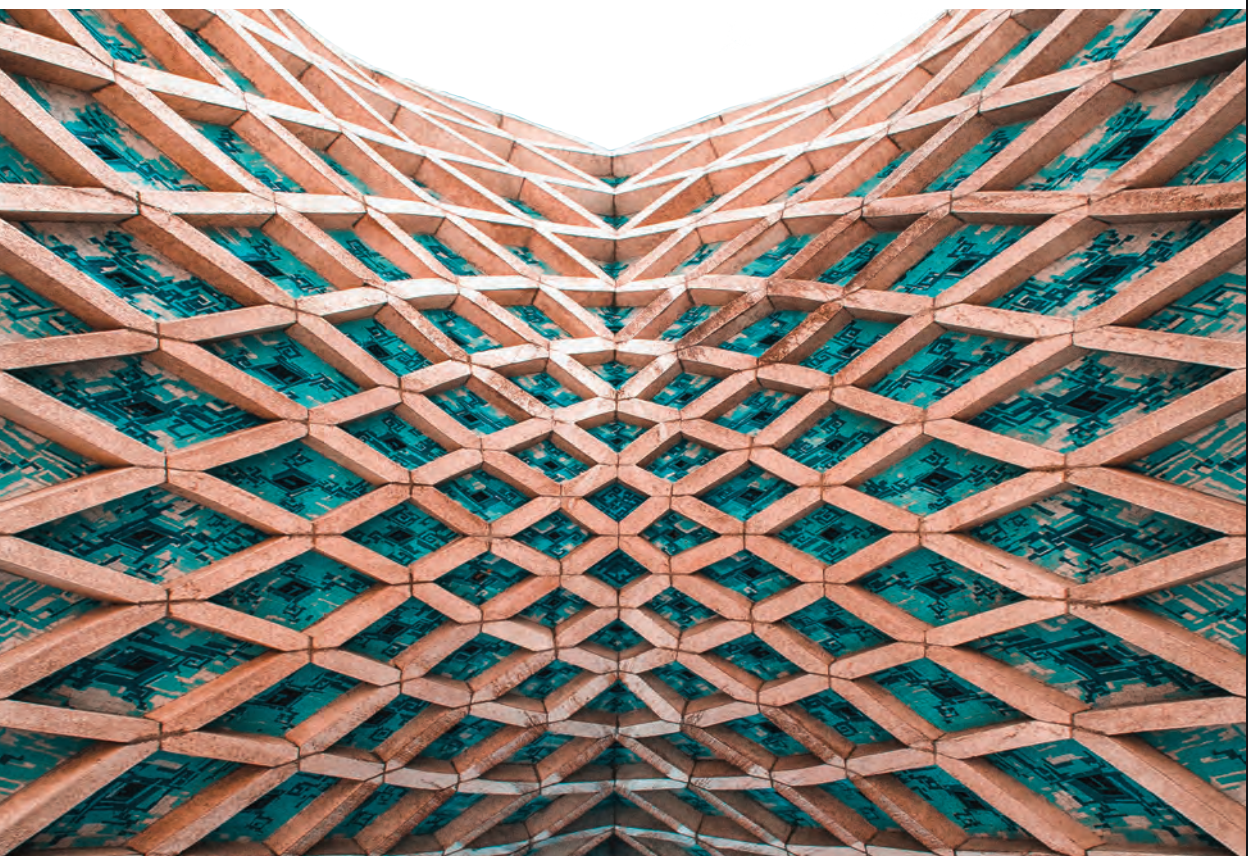
L’industrialisation réussie passe par une refonte complète de la gouvernance

Fini les tâtonnements, place à l'approche industrialisée de la donnée : le DataOps. Si on tire bien les leçons des années Big Data, qui ont coïncidé avec les débuts de la transformation numérique de la société, c'est bien de cela que les entreprises ont besoin : une infrastructure solide mais agile, qui offre aux métiers la possibilité de s'appuyer sur la donnée pour prendre rapidement des décisions pertinentes, voire automatiser un certain nombre de processus.

Gartner définit le DataOps ainsi :
« DataOps est une pratique collaborative de gestion des données qui vise à améliorer la communication, l'intégration et l'automatisation des flux de données entre les gestionnaires et les consommateurs de données au sein d'une organisation.

L'objectif de DataOps est de fournir de la valeur plus rapidement en créant une livraison prévisible et une gestion du changement des données, des modèles de données et des artefacts connexes. DataOps utilise la technologie pour automatiser la conception, le déploiement et la gestion de la livraison des données avec des niveaux de gouvernance appropriés, et il utilise les métadonnées pour améliorer la convivialité et la valeur des données dans un environnement dynamique. »

Il s'agit donc de la combinaison d'un environnement technique (que nous aborderons à la fin de ce chapitre) et d'une gouvernance modernisée, prenant en compte la stratégie, les rôles, les processus, ou encore les métadonnées.



Les principes du DataOps

1. Satisfaire continuellement le client en lui fournissant rapidement et en permanence des informations analytiques précieuses.

2. Valoriser une analytique avec des données pertinentes fournies par des systèmes robustes.

3. Embrasser le changement, encourager les clients internes à faire évoluer leurs besoins pour optimiser l'avantage concurrentiel.

4. Diversifier les rôles, les compétences et les opinions pour augmenter l'innovation et la productivité : il s'agit d'un sport d'équipe.

5. Prévoir des interactions quotidiennes entre les clients, les équipes analytiques et DevOps tout au long du projet.

6. S'auto-organiser afin de faire émerger les meilleurs algorithmes, architectures, exigences et conceptions.

7. Réduire l'héroïsme : face au rythme et à l'ampleur des besoins en analytique, les équipes et processus doivent être durables et répétables.

8. Affiner les performances opérationnelles en prenant régulièrement en compte les retours des clients, ses propres retours, ainsi que les statistiques opérationnelles.

9. Garder à l'esprit que l'analytique, c'est du code : les outils pour intégrer, modéliser ou visualiser les données produisent du code, qui décrit le travail effectué sur la donnée.

10. Orchestrer de bout en bout les données, les outils, le code, les environnements et le travail des équipes d'analyse s'avère un facteur clé de succès.

11. Rendre les résultats reproductibles en prenant soin de tout versionner : les données, les configurations matérielles et logicielles, le code et la configuration de chaque outil.

12. Minimiser le coût en fournissant aux équipes analytiques des environnements techniques faciles à créer, isolés, sûrs et jetables qui reflètent leur environnement de production.

13. Privilégier la simplicité, ou l'art de maximiser la quantité de travail qu'on n'a pas à effectuer, pour contribuer à l'agilité globale.

14. Considérer l'analytique comme du 'lean manufacturing' : il faut se concentrer sur les processus de façon à atteindre une efficacité continue dans la fabrication de la connaissance analytique.

15. Mettre l'accent sur la qualité : les pipelines analytiques doivent pouvoir détecter automatiquement les anomalies et les problèmes de sécurité.

16. Surveiller la qualité et les performances en permanence pour détecter les variations inattendues et générer des statistiques opérationnelles.

17. Réutiliseraumaximumpouraméliorer l'efficacité de la production d'informations analytiques.

18. Réduire les cycles, de la conception à la réutilisation après réarrangement, en passant par le développement et la diffusion en tant que processus de production répétable.

Le Manifeste du DataOps (traduction : SFEIR).

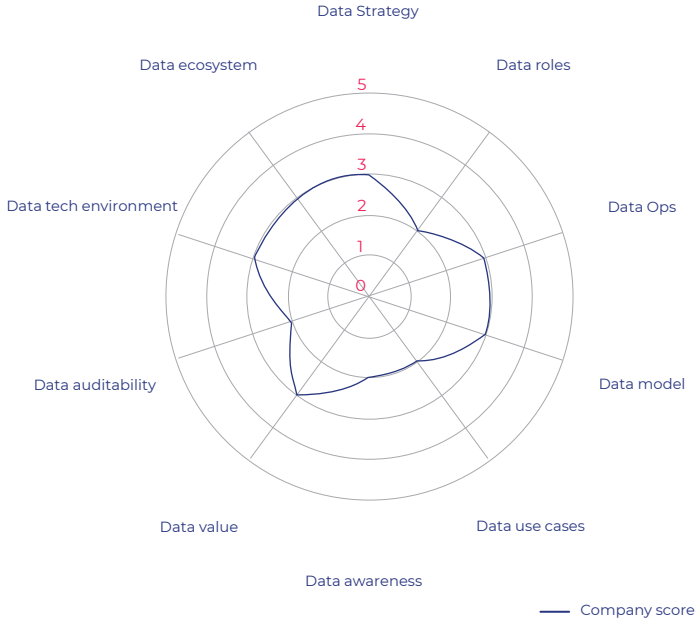
Source : https://www.dataopsmanifesto.org/

Instaurer une nouvelle collaboration IT et métiers

« Pendant un moment, on nous demandait de créer des datalakes, juste pour collecter des données, observe Florent Legras, Head of Data chez SFEIR. *Je pense qu'il est temps de remettre le business au centre, que les plateformes de données soient guidées par la valeur apportée. »*

Les sujets autour de la donnée ne sont ni complètement IT ni complètement métiers. S'ils sont guidés uniquement par l'IT, ils échoueront, faute de sponsor, de direction claire et d'utilité. S'ils restent entre les mains des métiers, chacun fera ce qu'il souhaite de son côté - créer une usine à gaz personnelle avec un tableur ou un outil de BI, recourir à un fournisseur externe d'analyse, etc. - avec dans tous les cas un accroissement du shadow IT (toute cette part de l'informatique d'entreprise qui échappe à toute supervision), une perte du contrôle des données ou encore une perte de savoir-faire et de connaissance métier. Dans ce cas, comment valider que les analyses sont pertinentes ? Que les données sont cohérentes au sein de l'entreprise ? Et comment faire lorsqu'un besoin nouveau émerge ?

Une collaboration entre IT et métiers est absolument nécessaire pour progresser, délivrer la valeur ajoutée attendue et ne pas tomber dans ces travers. Cela peut sembler une évidence, mais dans les faits, cette collaboration est loin d'être évidente. Nous le constatons régulièrement dans les audits de maturité de la gouvernance de données que nous réalisons (cf. un exemple de score ci-dessous).



Matrice de maturité SFEIR de la gouvernance de données avec un exemple fictif de score

Ces problèmes de collaboration entre IT et métiers impactent plusieurs axes d'analyse :

→ La stratégie :
la volonté de s'appuyer sur la donnée, de la considérer comme un actif stratégique, doit venir de la direction générale de l'entreprise et être déclinée à tous les niveaux, pour que chaque projet soit aligné avec les objectifs de l'entreprise. Sans ce sponsoring explicite, difficile de faire collaborer des gens qui ont leurs propres priorités et urgences, et à plus forte raison de faire en sorte qu'ils se sentent responsables des attributions et rôles qu'on peut leur confier.

→ Les rôles :
il ne suffit pas de nommer des data owners, des data stewards ni même un Chief Data Officer pour que, magiquement, tout s'organise. En revanche, ne pas désigner des responsables ou omettre de clarifier leurs rôles et responsabilités sont des risques majeurs. Comment sensibiliser quelqu'un sur la notion de qualité d'une donnée, par exemple, s'il ne s'en sent pas responsable ? Ces différentes attributions doivent donc être claires pour tous, à commencer par les responsables eux-mêmes, dont les périmètres et missions doivent être précis. Au besoin, il peut être nécessaire de créer de nouveaux postes (cf. pages suivantes).

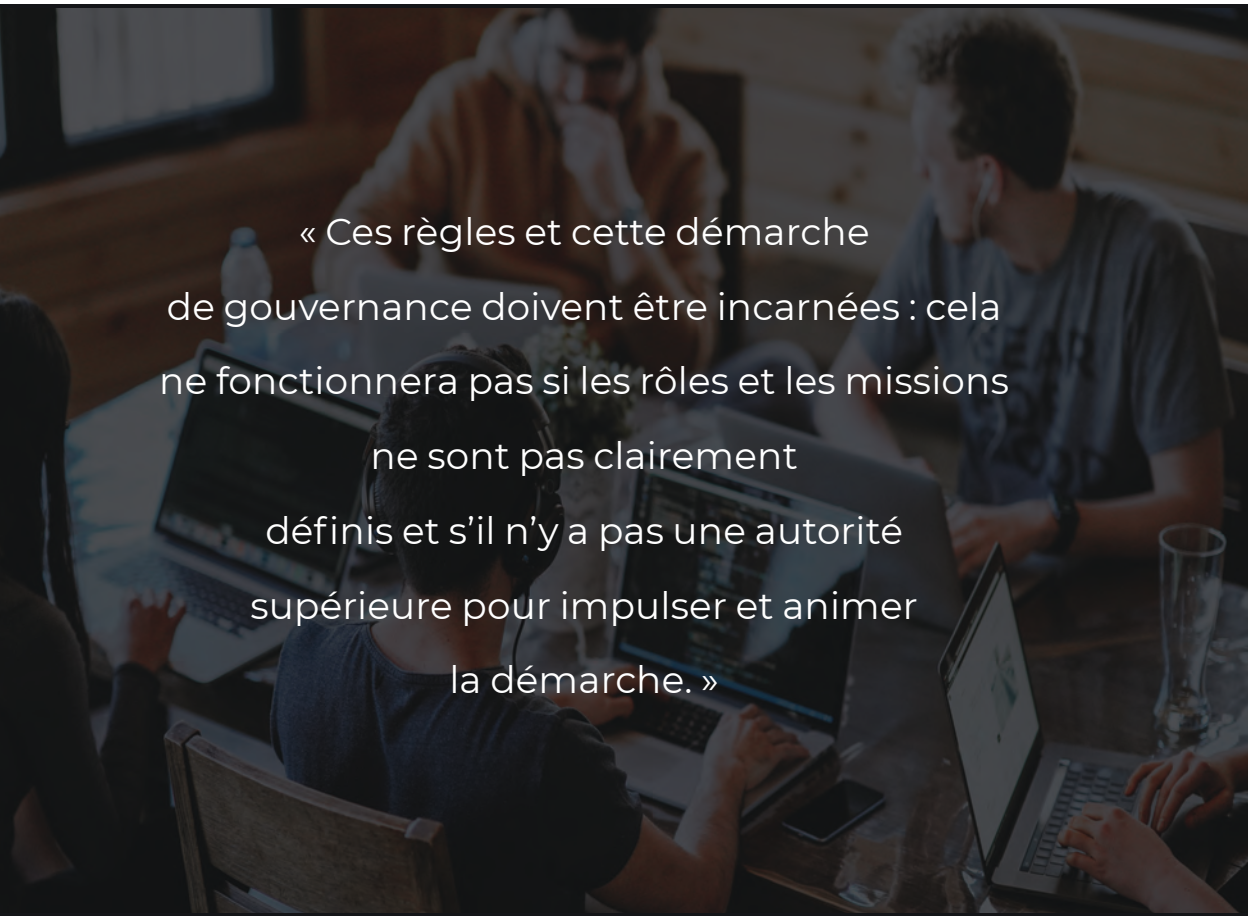
→ Les processus (le DataOps) :
la façon dont IT et métiers travaillent ensemble dépend de l'organisation et donc des rôles, mais aussi des outils mis à disposition par l'IT pour faciliter cette collaboration, faire en sorte qu'un dialogue s'instaure dès les phases initiales de cadrage et que les métiers soient les plus autonomes possibles tout en respectant les contraintes légales et techniques. Cela évite de devoir reconsidérer un projet à mi-chemin parce qu'il n'aurait pas respecté une phase de validation ou serait parti sur des technologies exotiques. Un processus clair évite également les incompréhensions, sur les délais d'implémentation notamment.

Une plateforme de données va nécessiter côté IT un data architect, pour la concevoir en fonction des besoins de l'entreprise, des data engineers, pour la réaliser et la faire évoluer, des ingénieurs machine learning, pour industrialiser les modèles, mais aussi côté métier des data analysts et des data scientists pour tirer parti des données.

Cependant, des règles de gouvernance sont nécessaires pour faire vivre cette plateforme au quotidien et s'assurer que l'entreprise en tire de la valeur, tout en opérant en toute conformité avec les réglementations en vigueur.

Ces règles et cette démarche de gouvernance doivent être incarnées : cela ne fonctionnera pas si les rôles et les missions ne sont pas clairement définis et s'il n'y a pas une autorité supérieure pour impulser et animer la démarche. Cette autorité doit découler de la direction générale et s'incarner dans un chief data officer.

Ce rôle peut évidemment porter un autre nom. Dans le schéma ci-dessous, nous l'appelons responsable de la transformation data-driven : une façon d'insister sur la nécessité de changer les habitudes et les processus en s'appuyant sur la donnée.



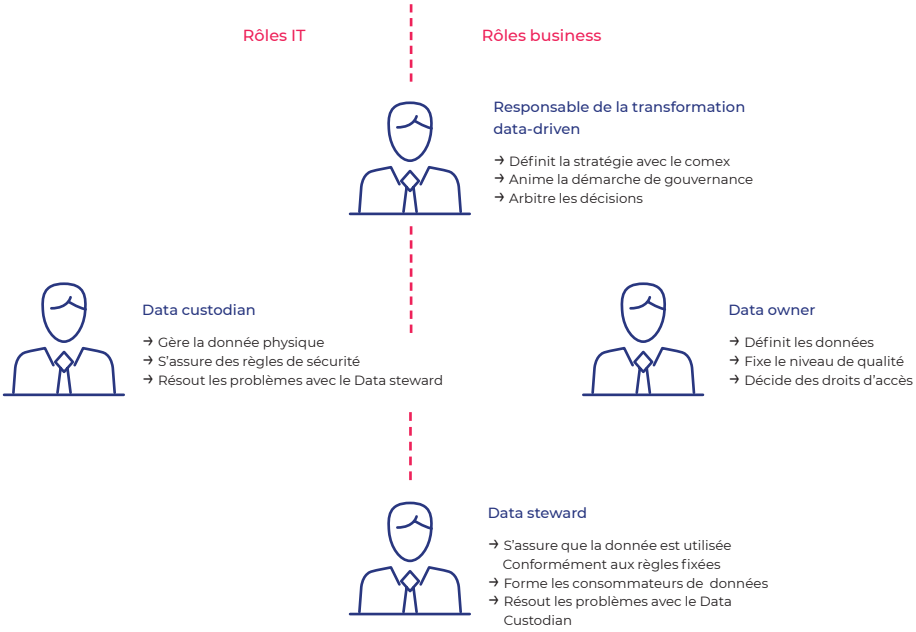
Les quatre rôles indispensables :

→ Responsable de la transformation data-driven : issue ou rattachée à la direction générale, cette personne orchestre la mise en œuvre concrète de l'orientation stratégique de l'entreprise. Cela implique conduite du changement, arbitrage et vision panoptique de la gouvernance des données au sein de l'entreprise.

→ Data owners : responsables d'un domaine métier et donc d'un type de données, ils sont à même d'offrir des définitions des données (qu'est-ce qu'un chiffre d'affaires, de quoi il se compose, comment il se calcule…) ainsi que la qualité attendue, le niveau de sécurité, les droits d'accès et la réglementation qui s'y rattache.

→ Data stewards : à cheval entre IT et métiers, les stewards constituent une équipe data qui facilite l'usage au quotidien des données au sein de l'entreprise, veille à ce que leur implémentation par l'IT corresponde aux usages attendus et à ce que leur usage par les métiers soit conforme aux règles édictées par les data owners.

→ Data custodians : intégrés au sein de l'équipe IT, ils s'assurent du bon fonctionnement de la data platform, gèrent le cycle de vie technique des données et des technologies, leur sécurité, leur sauvegarde…



Les nouveaux rôles de la gouvernance de données

La nécessité d'une collaboration saine entre IT et métiers se retrouve dans bien d'autres axes d'analyse de notre matrice de maturité de la gouvernance de données - à l'instar de la nécessité d'assurer la qualité, la confidentialité ou la conformité des données - mais les éléments listés sont vraiment déterminants. S'ils ne font pas l'objet d'un travail spécifique, le projet de valorisation de la donnée ne pourra donner de résultats satisfaisants, quelle que soit la qualité intrinsèque de la plateforme de données mise en œuvre.

« La construction de cette plateforme doit répondre avant tout aux usages que l'entreprise va en avoir, prévient Aurélien Allienne, Head of Data chez SFEIR. *La finalité pèse sur les premiers choix. Dans un premier temps, il faut faire abstraction des concepts techniques pour parler uniquement des besoins, des données qu'on veut y mettre… tout en attirant l'attention sur le fait que certaines données ne sont pas prises en compte alors qu'elles pourraient s'avérer intéressantes. »*

Cette approche par les cas d'usage est un excellent moyen d'entamer la conversation entre IT et métiers. À condition de ne pas rester dans les généralités, mais bien d'entamer un dialogue constructif où chacun challenge l'autre. Par exemple, si un métier dit qu'il veut disposer d'une vue à 360° de son client, qu'est-ce que cela signifie exactement ? Quels sont les objectifs recherchés ? Inversement, que propose l'IT pour simplifier la création de produits (offres ou services) basés sur la donnée, par des équipes "data fluent" sans avoir à passer à chaque fois par un nouveau cycle de projet IT ?

Un autre élément essentiel du dialogue entre toutes les parties prenantes est de disposer d'un vocabulaire commun, formalisé dans un dictionnaire auquel tout le monde peut se référer. Notamment des personnes issues d'autres départements business qui pourraient avoir besoin de données qui ne les concernent pas a priori. Décloisonner ces données s'avère fortement créateur de valeur dans différents cas d'usage.

Il s'agit donc de s'accorder sur les grands domaines sémantiques (les clients, les produits, les services, etc.) ainsi que sur les définitions individuelles. Ce sera le travail des data owners, sous l'égide du chief data officer (ou responsable de la transformation data-driven). Ce travail doit en effet être coordonné à l'échelle de l'entreprise, pour que la donnée puisse sortir des silos et contribuer à nourrir l'ensemble des projets dans tous les départements de l'entreprise.

Utiliser la donnée sur la donnée

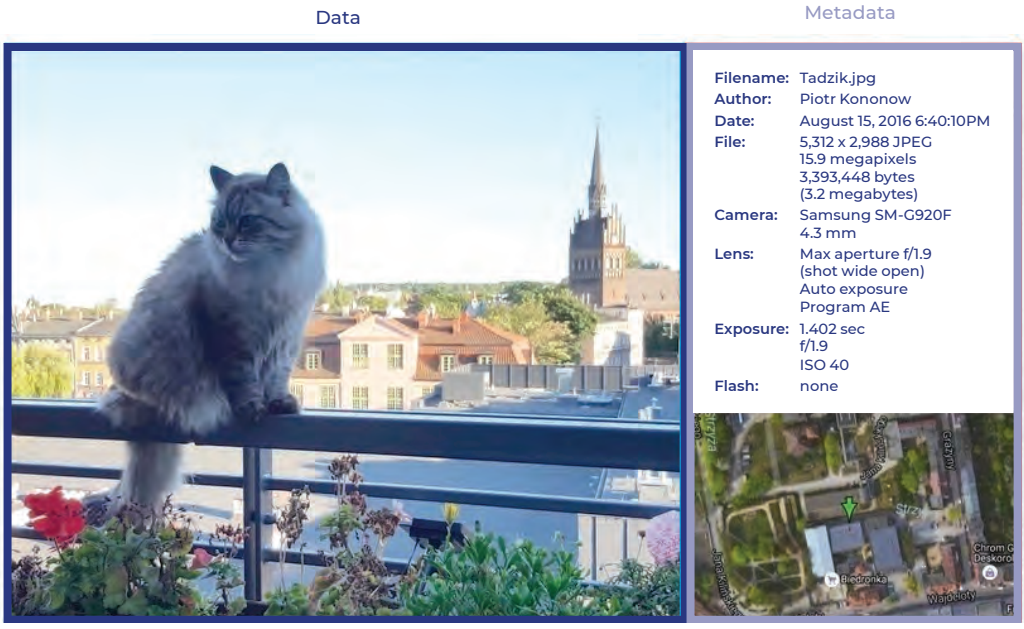
Le concept de métadonnée, ou donnée sur la donnée, est souvent négligé dans le processus de création d'une 'data platform'.

Construire une plateforme fonctionnelle, s'assurer qu'elle tienne la charge, qu'elle réponde aux besoins, tout cela s'avère déjà relativement complexe. Mais cette plateforme doit pouvoir vivre dans le temps, chaque utilisateur doit pouvoir se l'approprier facilement, qu'il s'agisse d'un ingénieur ou d'un métier. Pour cela, documenter la plateforme ne suffit pas : la donnée elle-même doit être documentée. C'est le rôle des métadonnées.

Les objets numériques naissent avec leur lot de métadonnées : une photo, une vidéo, un livre embarquent automatiquement un certain nombre d'informations qui vont faciliter leur indexation et leur recherche et permettre de les catégoriser selon un certain nombre d'attributs (langue, dimension, durée, localisation, etc.). Ils peuvent également comporter des métadonnées pour indiquer les droits de propriété intellectuelle s'y rapportant ou les droits d'accès.

> **Je pense qu'il est temps de remettre le business au centre, que les plateformes de données soient guidées par la valeur apportée.**

Florent Legras
Head of Data chez SFEIR



Exemple de données sur la donnée[1]

Si les données se rapportent à des clients, produits, usines, ou encore à du chiffre d'affaires, il faudra s'assurer de créer et gérer les métadonnées permettant de les manipuler sur le long terme. C'est l'un des piliers assurant une gouvernance stable. Les métadonnées peuvent apporter plusieurs types d'informations, qu'on peut regrouper dans deux dimensions :

→ Des informations sémantiques, qui aident à comprendre le sens des données et garantiront la cohérence des projets analytiques. Par exemple, une personne dans la colonne client a-t-elle déjà nécessairement fait un acte d'achat ? Un item de la colonne produit désigne-t-il un article et toutes ses déclinaisons (coloris, matériaux…) ou bien juste un SKU (stock keeping unit : unité en stock) ?

Idéalement, ces métadonnées seront directement disponibles pour les métiers au travers d'un dictionnaire.

→ Des informations techniques, qui permettent de connaître le cycle de vie de la donnée, son "lignage" et ses dépendances (où est-elle née, par quel pipeline est-elle passée, quel traitement a-t-elle subi, quelle est sa date de péremption…), son niveau de qualité (niveau de complétude d'une colonne, indice de confiance lié à la source…) ou encore les droits associés.

(1) Source : https://dataedo.com/kb/data-glossary/what-is-metadata

Ces métadonnées sont indispensables pour assurer le bon fonctionnement de la plateforme de données. Elles doivent donc être prévues lors de la création des tables et traitements, et générées dès la création de la donnée. Cela nécessite un chantier avec un volet humain important, pour aboutir à un consensus sur les informations et leur sens.
« Il faut que chaque équipe projet se sente responsable de la qualité de la donnée qu'elle produit, prévient Amel Hafsa, Cloud data architect chez SFEIR. Il y a donc toute une conduite du changement à mettre en place. L'idée derrière, c'est qu'on pourra chercher dans un catalogue non seulement toutes les tables et colonnes contenant une donnée précise, mais aussi tous les indicateurs de qualité associés. »

Bien sûr, cet aspect de la gouvernance comporte aussi un volet technique, pour outiller la démarche : il serait illusoire de vouloir maintenir cet ensemble de métadonnées à la main. Cela prendrait trop de temps pour un résultat pas forcément fiable et surtout un suivi s'effritant dans le temps, alors même que les volumes de données et les besoins ne feront qu'augmenter ; se priver d'automatisation finit très vite par décourager les équipes et fragiliser toute la démarche.

> **La construction de la plateforme de données doit répondre avant tout aux usages business que l'entreprise va en avoir.**

Aurélien Allienne
Head of Data chez SFEIR

Concept de Data Pipeline selon IDC

Il ne faut pas non plus négliger les données sur la plateforme de données elle-même et ses différents événements. Expert des sujets data, François Nguyen consacre sur son blog plusieurs articles sur les notions de DataOps et de "Data as a platform", ou plateformes de données, où il insiste sur la nécessité d'industrialiser son approche. Il écrit ainsi : *« La donnée, ce n'est pas un projet ou une liste de projets. En fin de compte, ce que vous construisez, c'est une plateforme ou du moins un cadre, car de plus en plus de données vont arriver. Vous pouvez bien sûr les gérer comme un projet unique, mais vous commencerez à créer des silos, même s'il s'agit de la même équipe. La notion de plateforme de données signifie que vous construisez une destination où il sera facile d'ingérer, de transformer, de surveiller et d'exposer les données avec rapidité et à l'échelle »*[1]

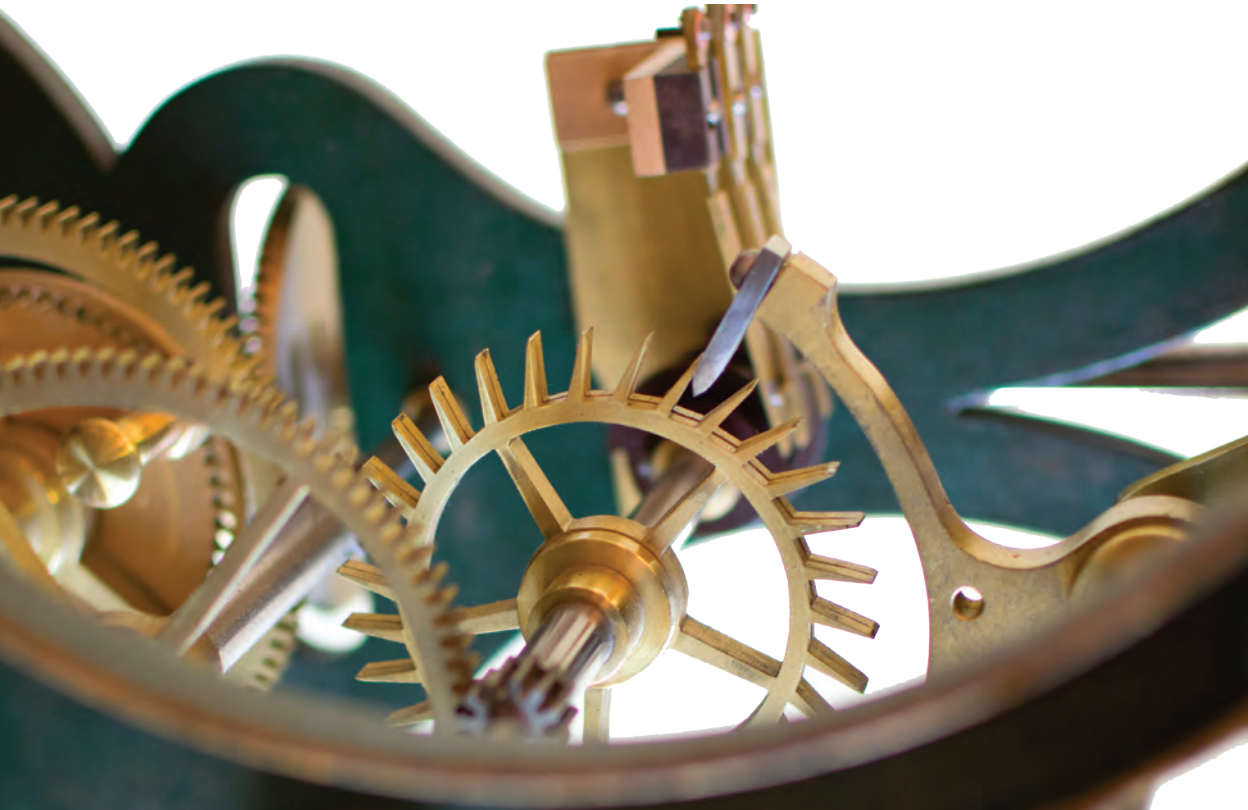
Le seul moyen de faire cela correctement, explique François Nguyen, est de s'appuyer sur la donnée. D'abord en considérant que l'ensemble du pipeline de données repose sur du code et un ensemble de paramètres : il peut donc facilement être reproduit et personnalisé. Ce pipeline représente l'ensemble du processus de traitement de la donnée, depuis son identification jusqu'à la phase de création de valeur pour l'entreprise, ainsi que le cabinet IDC le représente (cf. ci-dessus).

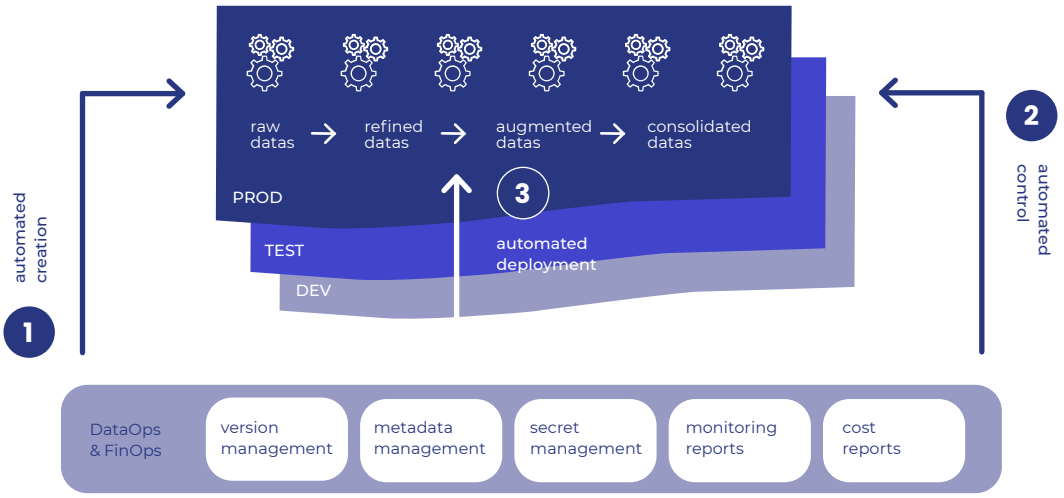
(1) Traduction SFEIR - Source : https://francois-nguyen.blog/2021/01/02/the-devops-part-of-dataops/.

Il s'agit ensuite de monitorer tous les événements, donc de capturer de la donnée sur chaque action du pipeline. Cela permettra en premier lieu une gestion plus efficace des incidents. Un des grands problèmes aujourd'hui en matière d'analytique, c'est l'indisponibilité de certaines données à cause de transferts ayant échoué. Or, on ne s'en aperçoit pas forcément tout de suite, et relancer un transfert peut s'avérer complexe, si la donnée en question provient d'une cascade de systèmes et que plusieurs jobs doivent être relancés.

Le monitoring permet de savoir ce qui se passe, à l'endroit et au moment où cela se passe. Dans le domaine bancaire, on enregistre ces événements de façon à pouvoir rejouer des séquences en cas d'audit. Bien évidemment, le monitoring de ces événements vise aussi à pouvoir prendre les mesures appropriées, de façon manuelle ou automatisée.

L'automatisation des actions sur événements s'inscrit dans une démarche EDA (Event-driven architecture, architecture orientée événements) qui réduit les interdépendances et facilite l'orchestration des processus.
« On avait tendance à définir toutes les étapes dans l'orchestration, explique François Nguyen. *Aujourd'hui, tous les départs de flux ne sont pas orchestrés dans un seul pipeline, mais de façon événementielle. Cela fiabilise l'ensem ble : on ne déclenche une étape que si la précédente est finie. Cela offre aussi plus de flexibilité pour développer de nouveaux objets, qui peuvent se brancher à tel ou tel endroit. Il suffit de définir les événements déclencheurs dans la base de configuration. »*





Démarche DataOps & FinOps (source : SFEIR)

La donnée sur les processus de traitement peut donc servir à automatiser la création des pipelines, leur contrôle ainsi que leur déploiement, depuis la plateforme de test jusqu'à la plateforme de production (cf. ci-dessus). Cette métadonnée peut également être exploitée avec profit, dans le cadre d'une démarche FinOps, pour optimiser la création de valeur en maîtrisant ses coûts. Une requête peut coûter plus ou moins cher selon la puissance et le temps de traitement qu'elle requiert. De même, recourir à certaines technologies spécifiques ou bien stocker de grands volumes de données peut s'avérer coûteux au-delà d'un certain seuil et sur le long terme. Avec une démarche FinOps, la décision de lancer cette requête ou de stocker ou non ces données se prend en toute connaissance de cause.

Tout cela est très variable en fonction des technologies utilisées, des licences impliquées, etc. Mais dans tous les cas, cela demande d'être rigoureux à la fois dans les choix techniques et dans les décisions qui relèvent davantage du business. Les principes du FinOps ne se bornent pas à analyser les coûts. Il s'agit avant tout de faire collaborer métiers et IT, que chaque partie comprenne les contraintes de l'autre.
La DSI doit être capable de dire combien coûte telle ou telle demande des métiers, de conseiller des alternatives, afin que les métiers puissent hiérarchiser leurs besoins, trier et affiner leurs demandes. Peut-être accepteront-ils de bon gré de lancer une analyse sur les 6 derniers mois plutôt que sur les 3 dernières années. Au moins, la décision sera prise sur la base de données fiables et pertinentes.

Philippe Girolami

VP of Engineering, Data engineering & Machine Learning, Dailymotion

Il n'y aura jamais assez d'ingénieurs dans une entreprise pour tout faire, il faut donc outiller les personnes qui se servent de la data et bien articuler les rôles et responsabilités entre les uns et les autres.

Sur la partie Analytics, par exemple, jusqu'à quel point les ingénieurs ont-ils besoin de comprendre le métier ? S'il y a un défaut de compréhension et qu'ils tentent de faire de la modélisation, le résultat ne sera pas bon. Inversement, les analystes peuvent être tentés de réaliser les requêtes de transformation eux-mêmes et, possiblement, aboutir à des dégradations de performances, des hausses de coûts ou, pire, des contresens dus à la complexité de l'outil de transformation - alors que tout cela aurait pu être évité avec une démarche d'ingénierie.

L'organisation autour de la donnée est forcément propre à chaque entreprise. Chez Dailymotion, la partie engineering est centralisée. Il s'agit de construire des jeux de données capables de monter en charge, en respectant un certain nombre de bonnes pratiques pour que la qualité soit au rendez-vous. Je ne dirais pas que nous sommes parvenus à mettre en place une gouvernance idéale, parfaitement structurée, mais nous y travaillons encore et toujours. Aujourd'hui, nous avançons vers une organisation où les jeux de données centraux sont sous l'entière responsabilité des ingénieurs data et où les analystes ont pleine autonomie pour créer des jeux de données secondaires servant leurs besoins grâce à une plateforme fournie par les ingénieurs. Concrètement, dans ce schéma, les ingénieurs data travaillent en amont sur la plateforme, au centre sur la création et la maintenance des jeux de données centraux et en aval pour valider les requêtes des analystes pour créer les jeux de données secondaires.

Sur la partie Machine Learning, le même défi de clarification des rôles se pose entre ingénieurs machine learning et ingénieurs data. Chez Dailymotion, le travail d'un ingénieur de machine learning va au-delà de produire un modèle Python dans un notebook : il n'a fini son travail que quand son modèle est en production. Nous avons commencé par toujours associer des ingénieurs data aux ingénieurs machine learning afin de traiter pour eux les problématiques de montée en charge, de CI/CD, d'extraction des données, etc. Nous allons à présent vers un modèle similaire à la partie analytics : les ingénieurs data fournissent aux ingénieurs machine learning les composants nécessaires à leur autonomie (chaînes de traitement normalisées, composants standard réutilisables entre chaînes de traitement, cluster de GPU "as-a-service", template de CI, orchestrateur propre au ML, etc.).

Un autre défi, propre au Machine Learning, est de savoir définir le “besoin”. C’est quelque chose sur lequel nous avons progressé : au début, nous avons travaillé sur des algorithmes sans suffisamment clarifier les contraintes issues du produit. Renforcer la collaboration avec le produit et définir les attendus en termes de spécifications ont été clefs pour arriver à monter en production tous nos projets ces 4 dernières années : quel est le KPI de succès et quelles contraintes s’imposent à la solution, quoi qu’il advienne.

À la fois pour les Analytics et pour le Machine Learning, l’utilisation du Cloud est un élément crucial pour la réussite de nos projets car il apporte beaucoup de souplesse en éliminant des obstacles en début de projet : bons de commande, installation, capacity planning au cordeau. C’est d’autant plus important pour des projets Machine Learning où le niveau d’incertitude sur la faisabilité est élevé en début de projet.

L’élasticité du Cloud est aussi fondamentale pour stocker et analyser nos données. Nous disposons de plusieurs pétaoctets de données et malgré cela, nous n’avons jamais de problème, les performances sont au rendez-vous.

Le Cloud nous apporte la tranquillité d’esprit : en passant moins de temps sur l’ingénierie, on peut se concentrer davantage sur notre mission, qui est de répondre aux besoins du métier.

Mais le Cloud fonctionne tellement bien que cela en devient presque un inconvénient : quand on compte en pétaoctets et en centaines de chaînes de traitement, cela finit par alourdir sensiblement la facture. Il faut donc se poser la question de l’utilité de certaines données ou de certains traitements. Nous avons adopté une démarche FinOps : toutes nos ressources cloud ont été labellisées (buckets, datasets, pipelines, compute…) pour pouvoir les identifier dans la facture et savoir à quels produits elles sont rattachées. Cela nous permet d’offrir aux analystes comme aux ingénieurs machine learning une plateforme de données dont on maîtrise à la fois la qualité et les coûts.

3

Cloud et IA, le duo gagnant pour accélérer la transition

« Aujourd'hui, il faut avoir d'excellentes raisons, liées principalement aux aspects réglementaires, pour construire une infrastructure de données on premises plutôt que dans le Cloud, juge Vincent Costanza, data engineer chez SFEIR. *Cela coûte tellement cher de maintenir une infrastructure on premises. Et au-delà du coût, c'est l'aspect managé qui facilite grandement le provisionnement et la gestion des ressources. Les outils sont plus simples à appréhender et utiliser, et les fournisseurs de Cloud sont toujours en avance de phase sur l'évolution des technologies. »*

Ce jugement fait aujourd'hui consensus dans l'industrie. Henri d'Agrain, délégué général du Cigref, le Club informatique des grandes entreprises françaises, disait ainsi en avril dernier sur BFM Business : *« Ce mouvement du Cloud est inéluctable. Dans 10 ans, l'ensemble des services, des systèmes d'information des entreprises, des administrations publiques, seront dans le Cloud. »*

L'industrialisation de la démarche de valorisation de la donnée est l'un des tout premiers moteurs de ce passage au Cloud. Pionniers de cette démarche, Amazon et Google ont montré la voie. Leur exemple montre qu'on peut bâtir des infrastructures de données offrant une élasticité inouïe mais aussi la possibilité de s'aventurer sur le terrain du machine learning de façon industrielle.

Gagner en agilité, en rapidité et en élasticité

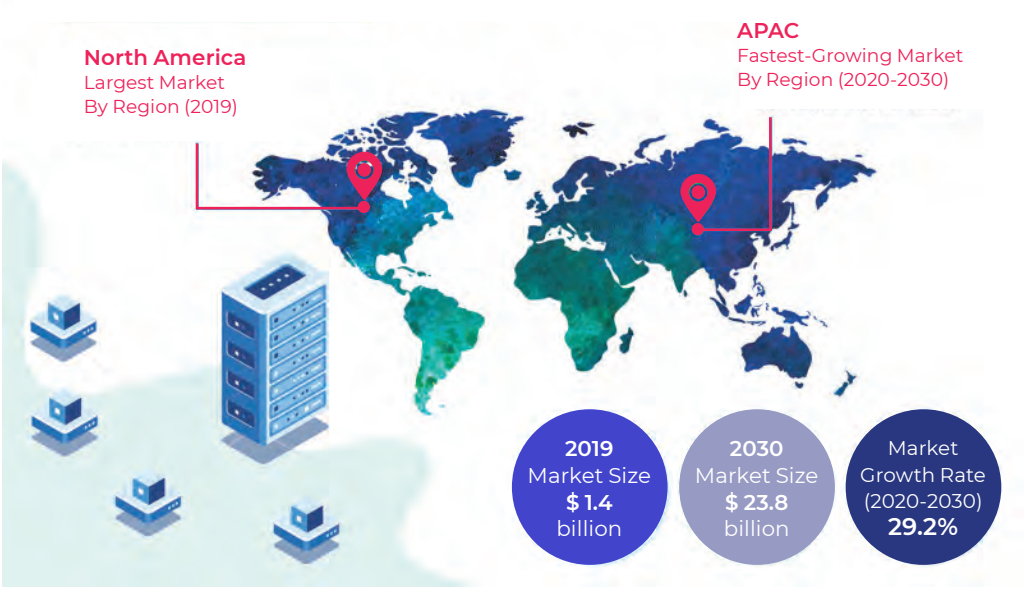
Les plateformes de données Cloud bénéficient d'abord des gains génériques du modèle Cloud : élasticité, montée en charge, disponibilité, évolution, sécurité… Data engineer et Cloud architect chez SFEIR, Pascal Castéran est par exemple intervenu dans une entreprise réalisant du séquençage d'ADN : *« Les clusters de machines on premises étaient chers à l'achat et rapidement obsolètes. De plus, leur capacité fixe de traitement n'était pas non plus optimale au regard de l'activité de l'entreprise ; avec des périodes d'utilisation intensive où des jobs pouvaient être bloqués en attente et, à l'opposé, des périodes d'inactivité Nous avons ainsi adapté leurs pipelines de traitement de données pour tirer parti d'une infrastructure Cloud permettant la mise à l'échelle automatique en fonction de la charge de travail. Cela a permis de réduire de façon conséquente les temps de traitement tout en maîtrisant la facturation associée. »*

Pour une entreprise de petite taille ou de taille moyenne, le Cloud est l'assurance de pouvoir disposer d'une infrastructure de qualité, digne des plus grands, à un tarif abordable car lié à l'usage. Dans une plus grande entreprise, cela redonne de l'agilité. En poste chez un grand industriel, Amel Hafsa constate ainsi les bienfaits de la migration : *« Là où nous devions attendre 3 ou 4 mois pour instancier un accès à l'infrastructure, il est devenu possible de commencer à travailler dans la journée, alors même que les process administratifs restent relativement contraignants. Ensuite, il est possible de créer une nouvelle source en quelques minutes, les traitements sont beaucoup plus rapides, on peut monter en charge… Cela facilite la vie des data scientists et des data analysts ! »*

Cela facilite également la vie des ingénieurs devant installer, configurer et maintenir les technologies. La technologie open source Kafka, par exemple, qui sert à recevoir, ordonner et délivrer des messages, dans le cadre d'une architecture orientée événements, nécessite plusieurs mois d'implémentation dans une architecture on premises ainsi qu'une équipe pour la maintenir. Des offres managées, dans le Cloud, accélèrent la mise en œuvre et exonèrent l'entreprise des efforts de maintenance et d'évolution. Il en va de même pour de nombreuses technologies de bases de données, dont les hyperscalers offrent des versions managées - quand ces technologies ne sont tout simplement pas exclusives au Cloud.

La sécurité des données s'en sort renforcée également, grâce aux mécanismes de sécurité inclus ou proposés en option avec les offres de Cloud - qu'il faudra donc activer et paramétrer correctement ! Les accès aux données, en particulier, sont beaucoup plus surveillés dans une infrastructure Cloud où le mode Zero Trust est la norme (ou devrait l'être !) : on part du principe que chaque accès vient potentiellement de l'extérieur et on ne fait confiance à personne à l'intérieur du réseau ; chaque tentative d'accès doit donc disposer des autorisations ad hoc. Cela facilite aussi le partage sécurisé des données avec des partenaires ou des clients.

Dans ces conditions, il n'est pas étonnant de voir les études converger sur une croissance extrêmement forte des plateformes de données dans le Cloud. L'étude de 2020 de Prescient Strategic Intelligence sur le marché du "data-warehouse as a service" conclut ainsi à un taux moyen de croissance annuelle de presque 30 % entre 2019 et 2030 (cf. ci-dessous).



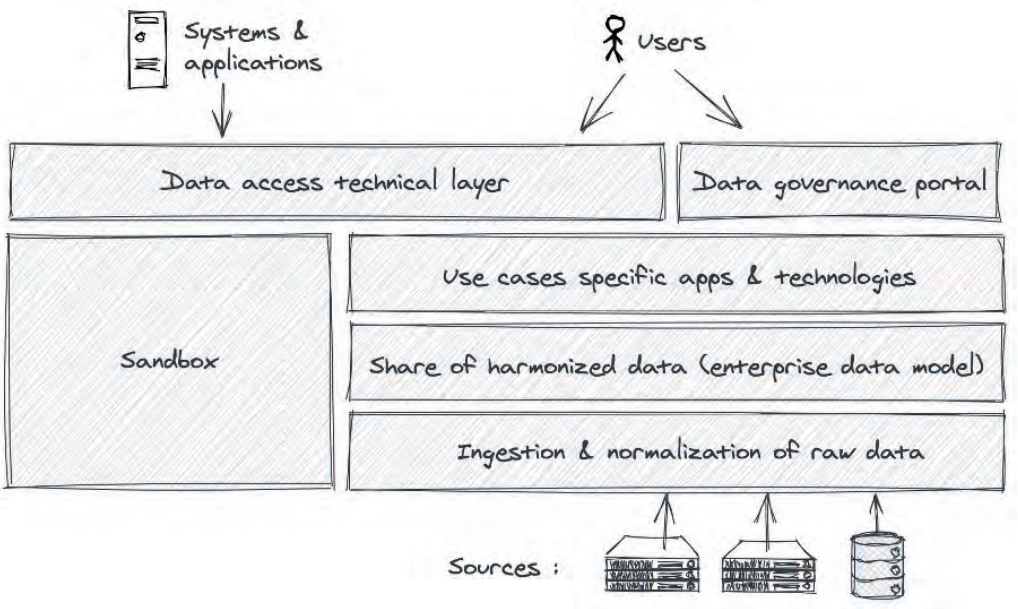
Marché mondial du 'datawarehouse as a service' selon Prescient Strategic Intelligence[1]

(1) Source : https://www.psmarketresearch.com/market-analysis/data-warehouse-as-a-service-market

Plutôt qu'on premises vs Cloud, le choix qui se pose à l'architecte est de déterminer s'il préfère opter pour une certaine dépendance vis-à-vis de son fournisseur de Cloud ou s'il est préférable de maîtriser complètement la technologie utilisée. *« Il faut évaluer les coûts de migration et la complexité liée à la réversibilité des données »*, prévient Pierre-Gilles Mialon, Senior Cloud Consultant chez SFEIR. Choisir une technologie unique ou un service managé accroît en effet l'adhérence à un fournisseur et donc aux risques associés à ce "lock-in".

C'est un fait dont il faut être conscient mais qui ne doit pas inquiéter outre mesure : le "lock-in" a toujours existé en informatique et la réversibilité des données pose sensiblement moins de problèmes lorsqu'il faut migrer d'un Cloud à un autre que lorsqu'il faut changer d'ERP. En l'occurrence, le passage d'une infrastructure de données on premises à une infrastructure dans le Cloud permet de débloquer des situations et d'ouvrir tout un champ des possibles au sein des entreprises qui le pratiquent. Et si l'entreprise préfère éviter le plus possible les solutions managées des hyperscalers, *« de très nombreuses offres open source existent,* souligne Pierre-Gilles Mialon, *qui permettent de réaliser sa propre implémentation ».*

L'autre question à se poser concerne le multi-Cloud. *« Tous les fournisseurs de Cloud ont leurs différenciants, on peut avoir envie d'utiliser le meilleur de chacun,* note Aurélien Allienne, Head of Data chez SFEIR. *En revanche, si des échanges sont prévus, il faut prévoir des coûts liés au réseau. Et si on brasse des téraoctets, cela peut vite chiffrer. »* Il faut aussi compter avec la latence entre les zones géographiques : elle peut s'avérer incompatible avec le service qu'on souhaite proposer.



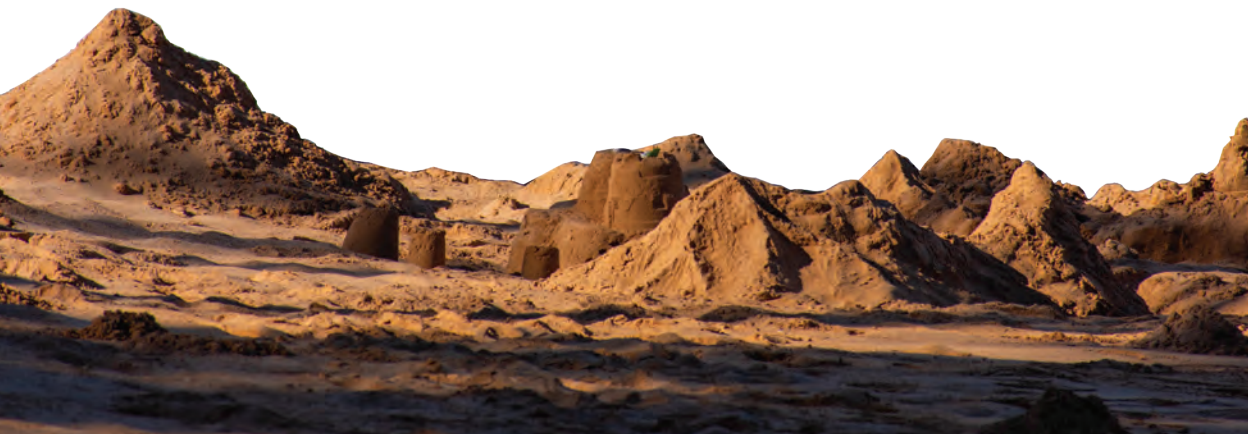
Exemple de Cloud Data Platform reproductible (source : SFEIR)

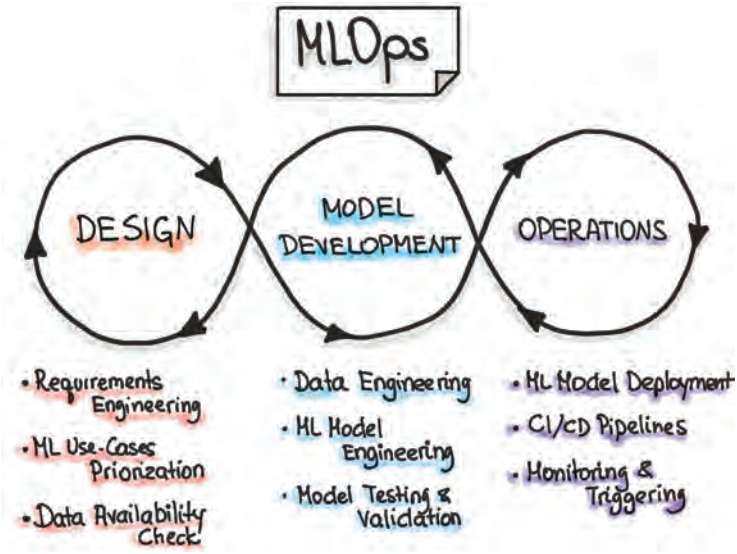
En revanche, normaliser une architecture et les outils d'une plateforme de données (cf. la page précédente) de façon à pouvoir la déployer à l'identique dans chaque plaque géographique voire dans chaque pays est parfaitement possible (sauf cas très particulier, comme en Russie ou en Chine). Et cela résout un casse-tête récurrent des grands groupes qui souhaitent laisser une certaine latitude à leurs filiales pour organiser au mieux leurs données en fonction de leurs problématiques et cas d'usage, tout en maintenant une certaine cohérence. De façon à introduire de l'industrialisation dans la mise en œuvre et l'exploitation des plateformes de données, tout en donnant de l'autonomie.

C'est un changement majeur dans l'approche que les entreprises peuvent avoir de la donnée et des plateformes analytiques. Elles ne sont plus limitées, finalement, que par leur imagination. Une fois que les briques fondamentales sont mises en place, les équipes data et métier peuvent se concentrer sur ce qui crée de la valeur : les cas d'usage.
Des « bacs à sable » peuvent être rapidement et simplement mis en place pour tester des idées, des concepts, des modèles, etc. et être détruits le soir et réinitialisés le lendemain. Cette flexibilité ainsi que la disponibilité d'énormes jeux de données et de processeurs spécialisés s'avèrent particulièrement utiles pour mettre au point et tester des algorithmes de machine learning.

Profiter du MLOps clé en main

Le cycle du DataOps s'inspire du cycle DevOps pour automatiser l'implémentation des pipelines de données et permettre aux métiers et spécialistes de l'analyse de données d'être les plus autonomes possibles - en fiabilisant et en industrialisant toutes les phases de conception et de mise en œuvre. L'approche MLOps cherche à mettre en place le même type de cycle pour la partie machine learning.





Les principes du MLOps (source : INNOQ)

Très souvent, trop souvent, même, création des modèles, développement et mise en production sont totalement décorrélés. *« On se retrouve souvent dans une situation où il manque le fil conducteur pour aller du point de départ au point d'arrivée,* témoigne Jonathan Chauvin, Data Engineer chez SFEIR. *On nous amène des données, on nous dit ce qu'on souhaite en concluant quasiment par fais-moi de la magie. »* La partie création de modèles, en mode expérimentation, parvient à des résultats qui semblent intéressants, voire éblouissants. Mais la mise en œuvre à l'échelle de ces modèles, leur industrialisation, et surtout la maintenance de dizaines de modèles en production laissent tellement à désirer que cela aboutit à des effets déceptifs : les promesses liées à l'intelligence artificielle sont fortes, mais les gains sur le terrain grâce au machine learning sont beaucoup moins impressionnants.

De toute évidence, cela n'aide pas à promouvoir les efforts de tous autour de la donnée. Alors que ce sont justement les bénéfices concrets que les métiers en tirent qui devraient les inciter à faire davantage attention à la qualité de la donnée qu'ils font entrer dans le système. C'est donc toute la chaîne qu'il faut revoir, l'intégralité du cycle MLOps : conception, développement, tests, déploiement et monitoring

Sur son blog, François Nguyen établit un parallèle cruel entre la beauté du machine learning tel qu'il est présenté sur des diapositives et la vraie vie (cf. tableau de la page suivante). Le mythe du data scientist, alternativement perçu comme un mouton à 5 pattes ou un super-héros, s'effondre. Il ne s'agit pas de dénigrer sa valeur, ou son rôle, mais le data scientist ne peut pas tout à lui tout seul. Il doit s'habituer à travailler en équipe et n'être que le maillon d'une chaîne qui doit aboutir à la mise en production d'un modèle pertinent pour les métiers ; cela signifie que ce modèle doit pouvoir s'appliquer aux données réelles et être constamment mis à jour pour conserver sa pertinence.

Le ML dans les slides	Le ML dans la vraie vie	L'apport du MLOps
Le Data Scientist est un super-héros…	Le Data Scientist n'est pas omnipotent. Ce n'est pas un ingénieur logiciel, pour commencer.	MLOps organise un travail d'équipe autour de la donnée et du machine learning.
…son essai fonctionne…	C'est une boîte noire.	MLOps introduit de la validation, de l'intégration continue, du versioning.
…sur son ordinateur personnel…	Il faut retravailler pour passer en production.	MLOps implique une chaîne de déploiement continue.
…avec une précision de 99%…	La pertinence laisse à désirer.	Des points de comparaison cherchent à établir une symétrie entre développement et production.
…sur une donnée de qualité…	Les données, au quotidien, présentent en fait une qualité… inégale.	MLOps introduit une validation sur les données réelles.
…le jour de sa mise en place.	Dès le lendemain et au fil du temps, les résultats se dégradent.	Un monitoring permet d'identifier les dégradations, le modèle est continûment entraîné, les métadonnées associées donnent de l'information sur le cycle de vie.

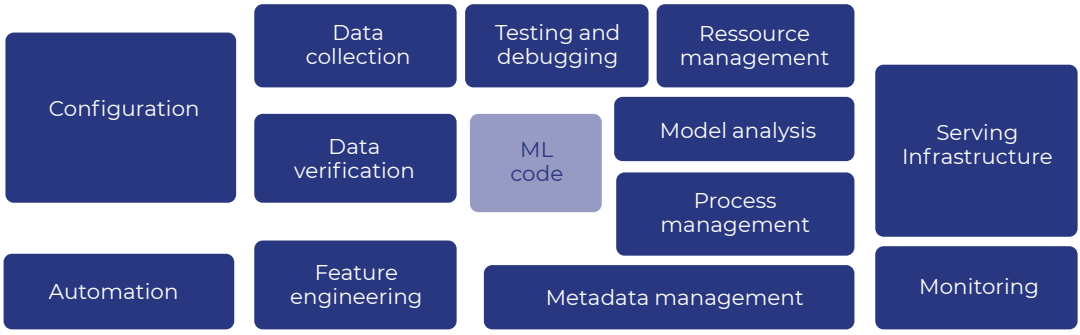
Comparaison entre le ML imaginé et dans la vie réelle, adaptée de François Nguyen[1]

Ne pas faire cet effort ne fera qu'alourdir la dette technique sans créer de valeur. C'est donc un changement culturel très fort mais indispensable qui est demandé aux data scientists, qui doivent passer d'une forme d'art de la donnée à une forme d'artisanat encadré, reproductible. Il s'agit de réduire les risques, insiste dans son introduction l'ouvrage d'O'Reilly sur le sujet (Introducing MLOps, How to scale machine learning in the enterprise, paru en novembre 2020) :

- Le risque que le modèle soit indisponible pendant une certaine période de temps ;
- Le risque que le modèle retourne une prédiction erronée selon le jeu de données analysé ;
- Le risque que la pertinence ou l'équité du modèle se dégrade avec le temps ;
- Le risque que les compétences nécessaires à la maintenance du modèle ne soient plus là.

(1) Source : https://francois-nguyen.blog/2021/01/10/ai-ml-powerpoint-vs-real-life/

Ce changement qui s'opère dans le monde du machine learning peut s'appuyer sur les plateformes technologiques qui ont énormément gagné en maturité. C'est le pari qui a été fait chez SFEIR, par exemple : aider les data scientists des entreprises à industrialiser leur approche en tirant parti de l'environnement technologique. Il faut arriver à prendre conscience que le code ML ne représente qu'une infirme partie du système (cf. schéma ci-dessous), explique Florent Legras, Head of Data : *« La plupart des data scientists n'ont pas de connaissances sur le fonctionnement d'un système d'information. De plus en plus d'entreprises se retrouvent donc avec des modèles inutilisables, car ne prenant pas en compte le processus d'intégration d'un algorithme. Ou bien elles ont des algorithmes en production, mais qui n'ont pas été révisés depuis un an ou deux - et dont on ne sait pas sur quelles données ils ont été entraînés. Il faut industrialiser grâce au MLOps, que les modèles soient documentés, testés, reproductibles, modifiables… »*



Le code machine learning ne représente qu'une petite fraction de la chaîne MLOps[1]

(1) Source : Hidden Technical Debt in Machine Learning Systems, https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf)

Qu'il s'agisse d'outils indépendants ou des plateformes de machine learning proposées par les hyperscalers, il est désormais possible d'industrialiser la démarche, les outils eux-mêmes imposant une refonte de la gouvernance des projets ML, dès lors qu'on suit les recommandations d'implémentation et d'usage.
Les plateformes Cloud vont même plus loin en proposant du "ML as a service" : un ensemble d'outils et de modèles pré-entraînés pour répondre à un certain nombre d'enjeux génériques. *« Dans ces domaines, les data scientists peuvent obtenir de meilleurs résultats que les algorithmes fournis par les modèles pré-entraînés,* observe Florent Legras, *mais cela leur demanderait plusieurs mois de travail, alors que là, le modèle peut être livré en une journée. Le gain de performance vaut-il la peine d'y consacrer autant de temps ? D'autant que les data scientists sont souvent assez spécialisés, et leur contribution sera plus utile sur des cas d'usage spécifiques. »*

Dans un contexte où les data scientists sont une ressource rare, les algorithmes sur étagère et les plateformes Cloud de MLOps 'as a service' représentent donc un levier non négligeable pour démarrer - et faire progresser sa politique de valorisation de la donnée dans de bonnes conditions.

Remerciements

L'auteur, Didier Girard, SFEIR et WENVISION remercient chaleureusement Didier Bove, DSI deVeolia, Laurent Ostiz, CDO d'Adeo, Philippe Girolami, VP Data engineering & MLde Dailymotion, ainsi que François Nguyen, Data & Analytics CIO et auteur de YetAnother Blog on Data, qui nous ont fait l'honneur et l'amitié de partager leur expérience de la mise en œuvre des plateformes de données et des principes du DataOps et du MLOps.

Nous remercions également tous les collaborateurs de SFEIR qui travaillent au quotidien dans l'univers de la Data et ont pu ainsi apporter leur expertise, ainsi que leurs anecdotes, qui ont nourri la réflexion ayant précédé la réalisation de ce document : Florent Legras, Aurélien Allienne, Amel Hafsa, Séléna Coquil, Jérôme Danger, Sébastien Friess, Alaeddine Bouattour, Pascal Castéran, Pierre-Gilles Mialon, Stéphane Barat, Vincent Costanza, Jonathan Chauvin…

Merci enfin à vous, lecteurs, qui ne manquerez pas de nous faire part de vos avis, expériences et recommandations, et avec qui nous écrirons la suite de livre blanc.

À propos de l'auteur

Olivier RAFAL

Olivier Rafal est Consulting Director - Strategy chez WENVISION. Il réalise des missions de conseil auprès de DSI et dirigeants d'entreprise sur l'optimisation conjointe des stratégies numériques et business.

Il réalise des missions d'audit et de conseil auprès des DSI et dirigeants d'entreprise sur l'optimisation conjointe des stratégies numériques et business, ainsi que sur la stratégie data-driven et la gouvernance de la donnée.

Olivier Rafal était auparavant VP de teknowlogy Group. Il a été journaliste dans le domaine IT pendant 15 ans, notamment en tant que rédacteur en chef de lemondeinformatique.fr, puis analyste et consultant pendant 10 ans, avant de rejoindre SFEIR en 2020.

À propos de SFEIR

SFEIR est une société de conseil en stratégie digitale et en développement d'innovation technologique Nous accompagnons les entreprises dans leur transformation numérique en mettant à leur disposition un accompagnement et des ressources spécialisées dans le développement d'applications de pointe. Nous aidons nos clients à concevoir avec succès la dernière génération de solutions numériques et d'applications métier.

Présents dans 7 agences en France et au Luxembourg, nous nous appuyons sur l'expertise de nos 700 expert.es spécialisé.e.s dans le développement d'applications de dernière génération et sur nos partenaires technologiques pour capitaliser sur les meilleures solutions afin de dynamiser la stratégie de nos clients.

Notre offre est fondée sur 3 métiers :
→ **CONSEIL** | Créez une stratégie gagnante basée sur l'innovation grâce à notre offre de conseil et d'accompagnement technique et technologique.

→ **RÉALISATION** | Concevez, créez, et scalez avec nos meilleurs experts pour le développement optimal de vos projets.

→ **FORMATION SFEIR Institute** | Développez vos compétences front, back, cloud, SRE, data et machine learning.

→ [sfeir]

SFEIR - 48 rue Jacques Dulud - 92200 Neuilly sur Seine

[sfeir]

contact@sfeir.com
www.sfeir.com

WENVISION

contact@wenvision.com
www.wenvision.com

Paris

48, rue Jacques Dulud
92200 Neuilly-sur-Seine
+33 1 41 38 52 00

Lille

74, rue des Arts
59800 Lille
+33 3 66 72 61 32

Strasbourg

Crystal Park
1, avenue de l'Europe
67300 Schiltigheim
+33 3 88 47 04 38

Luxembourg

5 Place de la Gare
1616 Luxembourg
+352 26 54 471

Bordeaux

Centre les Grands Hommes
Place des Grands Hommes
33300 Bordeaux

Nantes

Halle 6 Est
40 rue de la Tour d'Auvergne
44200 Nantes
+33 2 55 59 07 00

Belgique

Avenue des Arts 6
1210 Bruxelles
+32(0)2 899 83 70