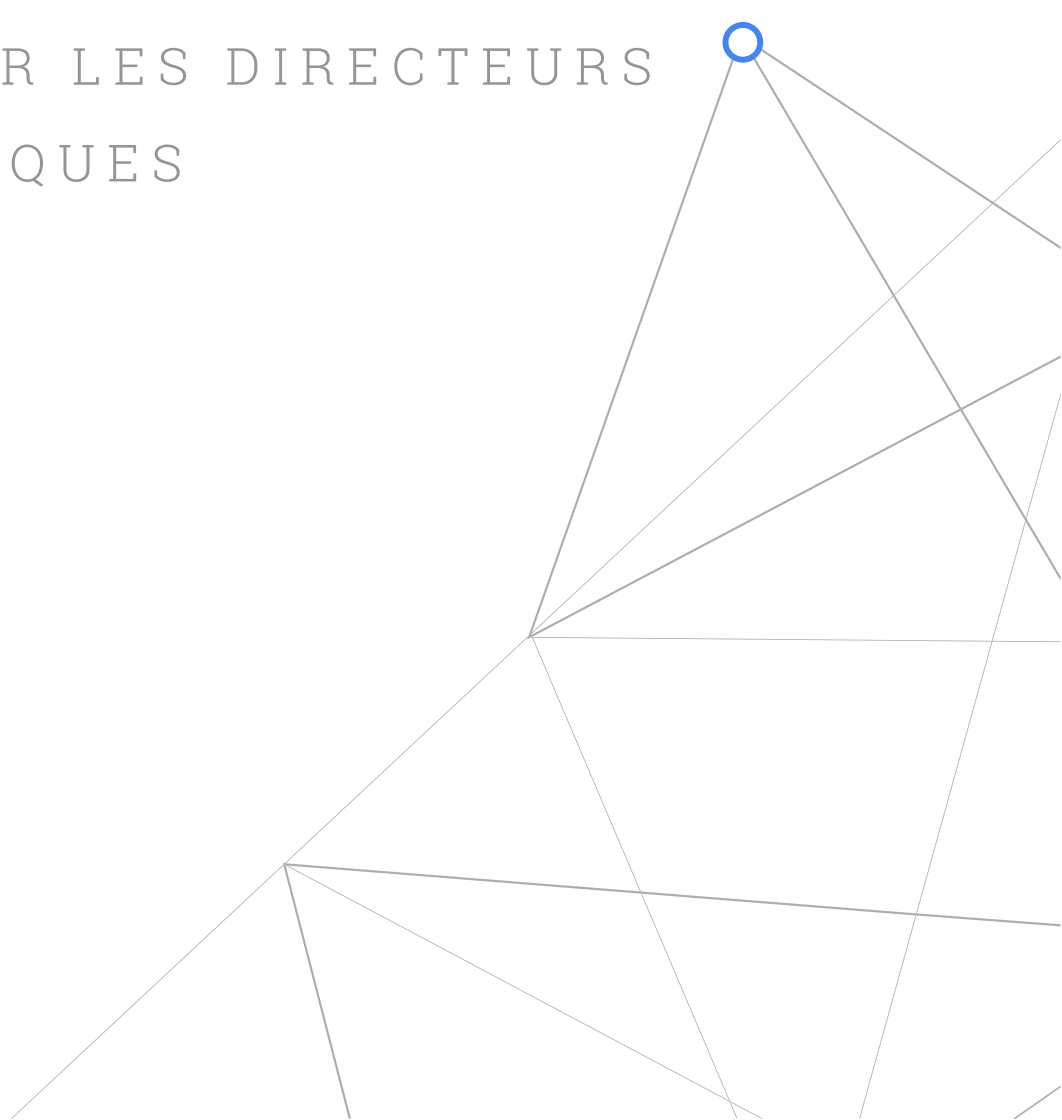


ANALYSE DES DONNÉES
ET MACHINE LEARNING :
GUIDE POUR LES DIRECTEURS
INFORMATIQUES



SOMMAIRE

Introduction 03

○ Un nouveau paysage de données 05

○ Stockage cloud et entreposage de données 09

○ Intégration des données en temps réel 16

○ Machine learning et intelligence artificielle 21

Conclusion 26

Travaux cités 27



INTRODUCTION

Dans les entreprises, prendre des décisions commerciales en se basant sur des données ne date pas d'hier. Auparavant, cela pouvait simplement signifier trouver un lien entre une campagne publicitaire dans la presse et une augmentation anecdotique des ventes. Les entreprises utilisaient tous les types de données à leur disposition, peu importe leur nature et le moment où elles les obtenaient.

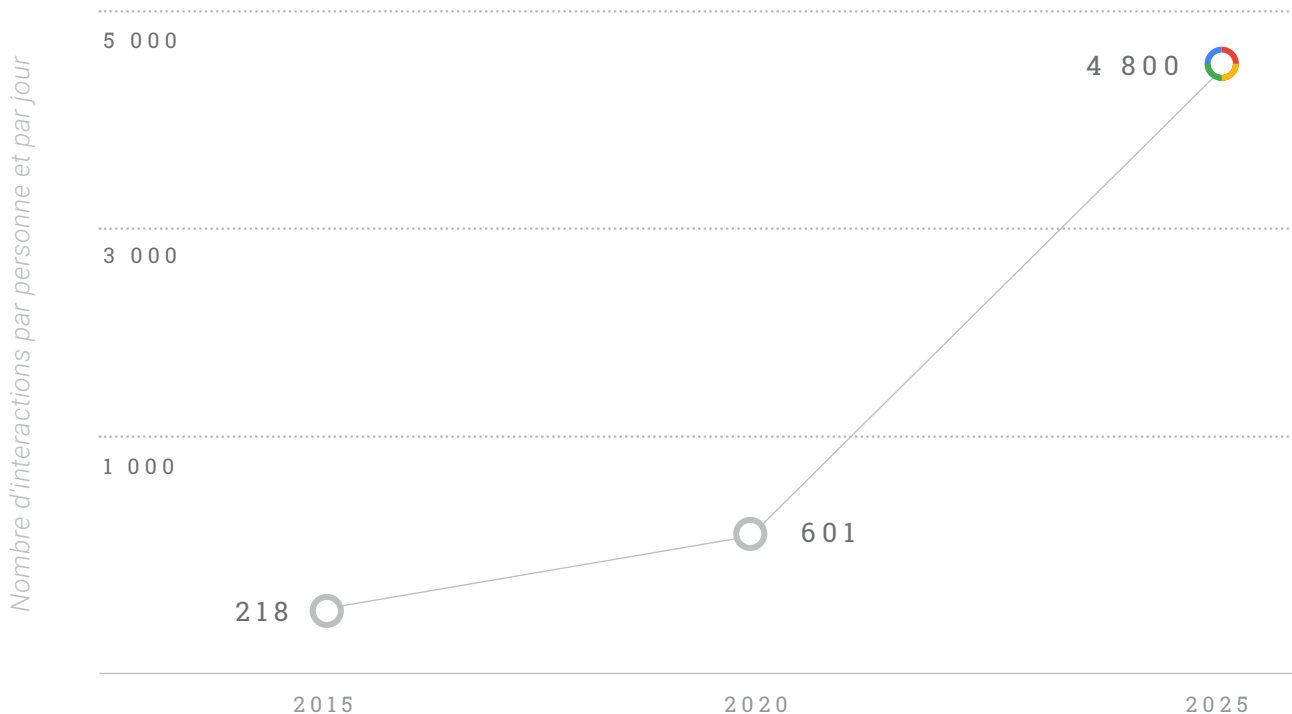
Aujourd'hui, les données sont omniprésentes. Elles sont générées depuis des appareils connectés à des vitesses vertigineuses, sous des formats multiples, par des milliards d'utilisateurs. Le Big Data est souvent qualifié d'opportunité, mais uniquement pour les entreprises capables de traiter le volume et la diversité des données. Pour les autres entreprises, le flot de données peut représenter un risque : les tendances potentielles sont inexploitées, les attentes des clients ne sont pas satisfaites, et les entreprises continuent de prendre des décisions non éclairées.

Deux facteurs permettent d'expliquer pourquoi ce phénomène est différent des précédentes évolutions. D'une part, on remarque une *augmentation exponentielle du volume et de la diversité des données générées* par des milliards d'utilisateurs et d'appareils. D'autre part, *la demande d'accès immédiat à des données et des informations de qualité est de plus en plus notable*. Chacun de ces facteurs engendre une nouvelle urgence dans la façon dont les entreprises gèrent les données. De plus, nombre des fonctionnalités du cloud sont arrivées à un équilibre coût-performances, permettant ainsi à toutes les entreprises d'avoir accès au machine learning (ML) et à l'intelligence artificielle (IA).

Bien que la valeur des données soit désormais largement reconnue, peu d'entreprises ont élaboré des stratégies modernes en matière de données¹. S'appuyant sur des recherches originales et sur les contributions de Google au cloud, ce guide a pour objectif d'aider les responsables informatiques et les dirigeants d'entreprise à mettre en place des stratégies cloud modernes pour la gestion des données. Chaque section présente des technologies permettant de transformer des paysages de données complexes et volumineux en informations pertinentes pour l'entreprise.



NOMBRE D'INTERACTIONS PAR PERSONNE CONNECTÉE ET PAR JOUR



D'ici 2025, une personne connectée interagira en moyenne 4 800 fois par jour avec des appareils connectés, soit l'équivalent d'une interaction toutes les 18 secondes².

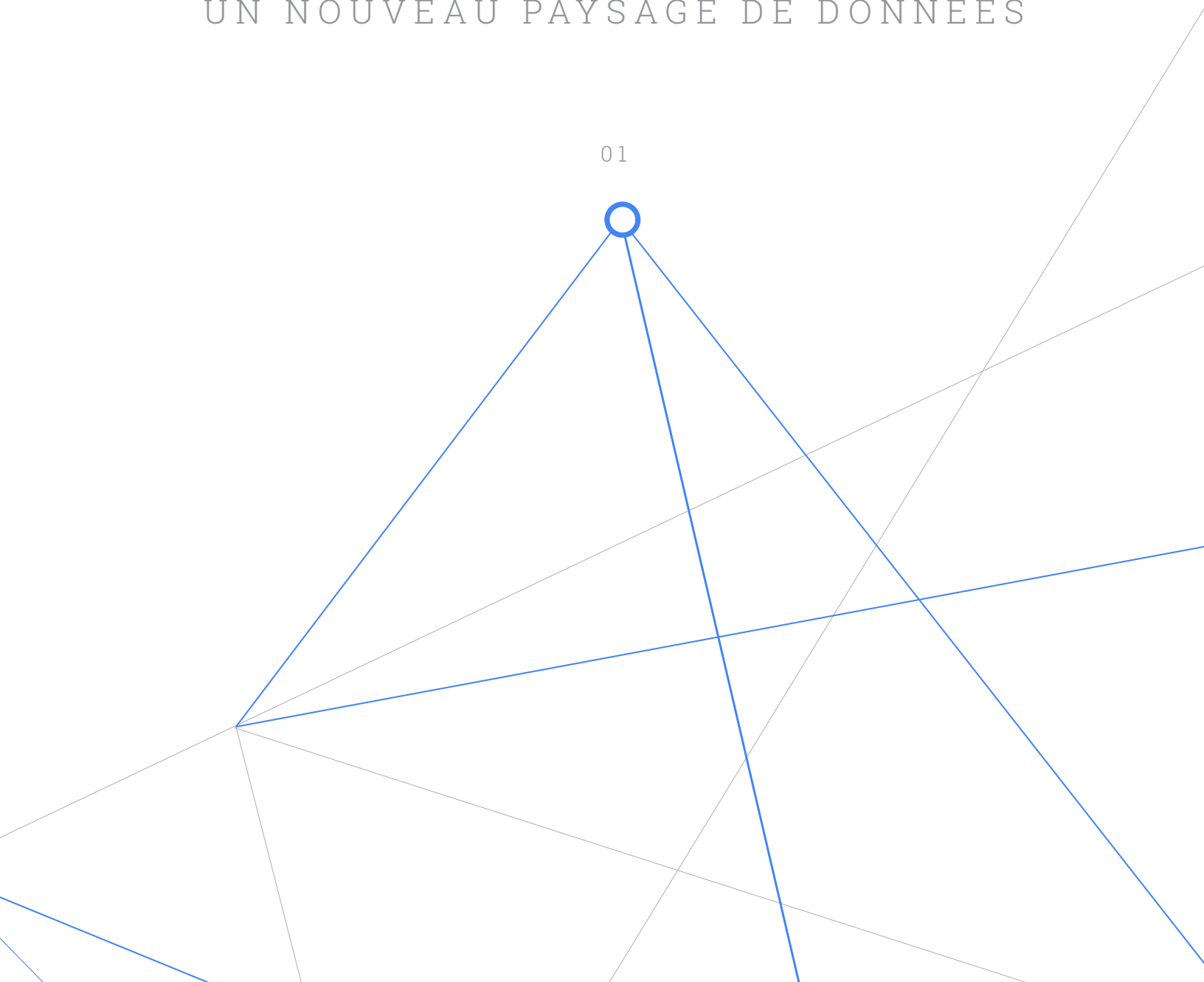
NOS ORIGINES

Le *Guide de l'analyse de données et du machine learning* de Google Cloud met à profit le résultat de vingt années de travail mené par Google pour résoudre les problèmes de données les plus complexes que rencontrent les entreprises. Au fil du temps, nous avons participé à de nouvelles recherches qui ont permis de façonner le paysage du Big Data. Nous avons contribué d'une part à deux articles de recherche à la fin [2003](#) et en [2004](#) qui, ensemble, ont donné naissance au mouvement Hadoop ; et d'autre part, à [l'article Dremel](#), qui forme le socle sur lequel repose la fonctionnalité d'entrepôt de données cloud dont il est question dans ce guide.

Nous avons conçu, créé et déployé [Spanner](#), le premier système compatible avec les transactions réparties et cohérentes en externe qui permet de distribuer des données à l'échelle mondiale. En 2017, nous l'avons [rendu accessible](#) à l'ensemble de nos clients³. Plus récemment, [Google Brain](#) a contribué à alimenter l'intérêt des entreprises pour l'intelligence artificielle, menant ainsi au lancement de notre [projet TensorFlow](#) en Open Source⁴. Grâce à ce guide, nous espérons partager notre expérience avec des responsables stratégiques qui cherchent à tirer parti des promesses du machine learning et de l'IA.

UN NOUVEAU PAYSAGE DE DONNÉES

01





UN NOUVEAU PAYSAGE DE DONNÉES

01

La gestion des données serait plus facile si leur augmentation était limitée à quelques sources ou si elles étaient uniformes. Toute la difficulté réside dans la diversité des sources et des formats. Prenons par exemple le volume croissant de données non structurées : les e-mails, les journaux système, les pages Web, les enregistrements des clients, les documents, les diapositives, les chats et l'explosion du nombre d'images et de vidéos rich media en HD. D'immenses volumes d'informations sont instantanément disponibles depuis n'importe quel appareil connecté à Internet, et de nouvelles attentes apparaissent quant à la disponibilité et l'immédiateté des données.

Les applications grand public (la recherche, la messagerie, l'e-commerce, les réseaux sociaux et les vidéos en ligne par exemple) ont été les premières à faire face à ce problème. Il a fallu créer de nouveaux systèmes pour traiter le trafic à l'échelle du Web dans son intégralité, tout en affichant instantanément les informations. Ces avancées sont désormais disponibles (et prennent une place de plus en plus grande) pour toutes les entreprises, aidant les fabricants à gérer plus efficacement leur chaîne d'approvisionnement ou les médecins à améliorer la précision des diagnostics médicaux.

Les équipes informatiques se retrouvent au milieu de ces évolutions. Elles s'échinent à fournir une *vue de l'organisation en temps réel* tout en gérant un *paysage de données plus vaste et plus complexe*. Lors du lancement d'un logiciel, il est nécessaire de réduire la complexité pour augmenter ses chances de réussite.

Dans ce guide, nous expliquerons comment toutes les entreprises peuvent bénéficier de services cloud gérés pour relever les défis de la gestion de données. Ce guide illustre le processus par lequel les données brutes d'entreprise sont envoyées sur l'espace de stockage cloud. À mesure que se posent les questions stratégiques, les outils cloud sont capables de préparer et d'organiser les données brutes à la demande. Une fois préparées, les données sont ensuite intégrées à un entrepôt de données cloud, où elles

sont immédiatement exploitables pour analyse. Cette réserve de données sert de "base principale" à partir de laquelle les entreprises peuvent recueillir, préparer et analyser tous types de données, quelle que soit leur source. La nature entièrement gérée des services cloud permet de rationaliser l'intégralité du processus (y compris l'analyse en temps réel), sans que les équipes informatiques n'aient besoin de connaître l'infrastructure sous-jacente. Sur cette base, le guide conclut en illustrant la manière dont les entreprises peuvent utiliser ce cycle de collecte et de préparation des données pour mettre en place le machine learning et l'IA.

UNE STRUCTURE SANS SERVEUR : UNE MEILLEURE PRODUCTIVITÉ DES TECHNOLOGIES DE L'INFORMATION

Les architectures sans serveur modernes sont le résultat des efforts déployés pour *réduire les responsabilités* que portent les développeurs et les équipes informatiques en matière de gestion. L'informatique sans serveur a essentiellement pour but d'éliminer les tâches standardisées (la gestion des clusters de serveurs, le partitionnement des bases de données, l'équilibrage de charge, la planification de la capacité et la surveillance de la disponibilité) afin que les équipes informatiques puissent se concentrer sur les questions clés de l'entreprise. Grâce à l'informatique sans serveur, la distinction est nette entre les *tâches informatiques standardisées* (le travail de maintenance courant commun à toutes les entreprises) et le *travail différencié* qui élève le service informatique au rang de fournisseur direct de plus-value pour l'entreprise.

RÉSUMÉ DU CHAPITRE 1

1 Les entreprises font face à trois nouveaux défis :

- Le *volume* de données créées
- La *diversité* des formats et des sources de données
- La *vitesse* à laquelle les consommateurs et les collaborateurs internes attendent les informations

2 Le **cloud computing** aide les entreprises à relever ces défis grâce à *une gestion des données rapide à grande échelle*, sans qu'elles n'aient besoin de se soucier de l'infrastructure.

3 **Plus précisément, les entreprises sont désormais en mesure de moderniser leurs stratégies de données** en se concentrant sur le *stockage cloud* et sur les *entrepôts de données* qui constituent la pierre angulaire nécessaire au fondement du machine learning et de l'IA.

TÉMOIGNAGE CLIENT

FIS

FIS analyse les événements et les perturbateurs du marché à l'aide de services cloud entièrement gérés.

FIS a élaboré un outil de reconstitution du marché permettant de déterminer les potentielles origines des événements qui perturbent le marché boursier, comme le "Flash Crash" de 2010. Non seulement le système de FIS est en mesure de stocker des milliards de transactions, mais il permet aux responsables de la conformité et de l'évaluation des risques de mener des opérations de surveillance, d'interroger le système à la demande et de procéder à une reconstitution du marché.

Selon FIS, le système est capable de traiter et d'associer, en une journée, jusqu'à 15 téraoctets de données en quatre heures et de les stocker pendant six ans, comme la loi le prescrit. "Cela fait un total de 30 pétaoctets de données" commente Neil Palmer, directeur des services technologiques Advanced Technology chez FIS. "Peu de structures atteignent cette envergure, et certainement pas dans le domaine des services financiers. C'est un projet colossal."

L'équipe de Neil Palmer avait besoin d'une plate-forme dotée d'une forte puissance de traitement, mais elle souhaitait aussi éliminer les coûts et la maintenance associés à la création et au fonctionnement d'un système sur site. Neil Palmer ajoute "Google Cloud Platform offre un avantage incomparable, celui de l'évolutivité. Dans ce scénario, une solution informatique traditionnelle reposant sur du matériel fixe signifiait investir des millions de dollars dans des équipements qui sont souvent inutilisés les journées boursières."

ENTREPRISE

FIS

SECTEUR D'ACTIVITÉ

Services financiers

À PROPOS

Leader international dans le domaine des technologies de services financiers, FIS est une entreprise spécialisée dans les services bancaires de détail et institutionnels, les paiements, la gestion d'actifs et de fortune, la gestion des risques et la conformité, le conseil et les solutions d'externalisation.

STOCKAGE CLOUD ET
ENTREPOSAGE DE DONNÉES

02





STOCKAGE CLOUD ET ENTREPOSAGE DE DONNÉES

Pour se moderniser, les entreprises peuvent commencer par centraliser les données brutes issues des principaux processus de travail dans un espace de stockage cloud. Ce faisant, elles se positionnent pour profiter des fonctions d'analyse qu'offre le cloud.

02

Les données cloisonnées qui sont disséminées dans l'entreprise ne font que ralentir les équipes informatiques et commerciales, tandis que de nouveaux cloisonnements sont créés tous les jours (pour des raisons organisationnelles ou techniques, ou les deux)⁵. Le magazine *Harvard Business Review* a [publié un article](#) sur la nécessité de disposer d'une source de données unique et fiable ainsi que de filtres distincts permettant aux différentes équipes de consulter les données⁶.

Grâce au stockage cloud et à l'entreposage de données, les entreprises peuvent atteindre ces deux objectifs : entretenir à la fois un dépôt central unique et permettre aux différents responsables commerciaux d'analyser les données en fonction de leurs besoins spécifiques (avec plus de rapidité et de flexibilité qu'auparavant). Ensemble, ces fonctionnalités participent à la création d'une vue panoramique de l'entreprise au-delà des cloisonnements.

Recueillir des données brutes pour les analyses futures

IDC estime que sur la totalité des fichiers, moins de 1 % est analysé⁷. Les 99 % restants (en fonction des besoins commerciaux et du moment où ils se présentent) contiennent des éléments d'informations utiles pour prendre des décisions. Les entreprises sont dans l'incapacité de prévoir les questions stratégiques qui vont se poser. C'est pourquoi elles ont besoin de trouver des méthodes simples et flexibles pour stocker de gros volumes de données à moindre coût, a fortiori pour les fichiers non structurés qui représentent la majorité des données générées⁸.

Le cloud représente un moyen économique de stocker d'énormes volumes de fichiers (moins d'un centime d'euro par gigaoctet lors de l'écriture)⁹. Les données nécessaires sont "gardées au chaud" (elles sont disponibles à l'échelle mondiale pour diffuser les applications ou effectuer des analyses), tandis que les données dont la valeur est encore inexploitée sont archivées (leur stockage est moins cher). Les espaces de stockage en ligne les plus puissants sont à même de récupérer des données, même celles qui sont archivées, avec une très faible latence.



IDC ESTIME QUE **MOINS DE 1 %** DU
TOTAL DES DONNÉES EST ANALYSÉ⁷.

< 1%

Au-delà des économies réalisées, le stockage cloud sert de base pour effectuer des analyses puissantes. Les entreprises peuvent recueillir des fichiers structurés ou non structurés très facilement dans leur format d'origine. Étant donné que le stockage est délibérément distinct du traitement et de l'analyse, les équipes peuvent différer la structuration des données nécessaire à l'analyse jusqu'à ce que les questions stratégiques se posent. Avoir la possibilité de restructurer facilement des données d'un même ensemble pour répondre à de nouvelles questions à la demande est donc encore plus important. Le stockage cloud est réellement différent, car il permet de recueillir et de reconvertir les données de façon extrêmement efficace. Pour orienter une entreprise de façon à bénéficier des fonctions d'analyse, les équipes doivent veiller à ce que les données brutes de leurs processus de travail soient recueillies et centralisées.

Grâce à cette flexibilité, les entreprises utilisent de plus en plus le cloud comme dépôt pour leurs données non structurées : environ la moitié des entreprises des États-Unis, d'Europe et d'Asie-Pacifique prévoient une hausse d'au moins 5 % du stockage de leurs données non structurées dans le cloud pour l'année à venir. Nombre d'entre elles annoncent une augmentation supérieure à 10 %¹⁰.

INTERNET DES OBJETS

D'après une enquête menée auprès de plus de 500 responsables informatiques par la revue *MIT Sloan Management Review* à la demande de Google Cloud, le cloud est en pleine expansion : d'ici 2019, la majorité (65 %) des applications, des données et/ou des infrastructures seront basées sur le cloud.

L'Internet des objets (IdO) joue un rôle central dans cette transition vers le cloud. En effet, 91 % des personnes interrogées déclarent s'intéresser à l'IdO : 59 % déploient actuellement des solutions permettant de transférer des données provenant d'appareils connectés à l'IdO vers le cloud, et 32 % prévoient de le faire. Parmi les principaux avantages du déploiement des données IdO dans le cloud, les personnes interrogées ont cité la faculté d'intégration à de nouveaux outils et plates-formes (à 33 %), un déploiement et une itération plus rapides (à 31 %), une plus grande flexibilité dans les processus de travail et dans le choix des fournisseurs (à 29 %) et une sécurité renforcée (à 28 %).

Pour tirer le meilleur parti de l'utilisation des données IdO, les entreprises doivent être capables de les comprendre en contexte. Avec un entrepôt de données cloud compatible avec des entrées par lot et par flux associé à une puissante plate-forme d'analyse, vos données IdO peuvent vous fournir des tendances en temps réel.

Gérer les données au-delà des cloisonnements

Recueillir tous types de données à moindre coût est désormais possible. Les entreprises peuvent alors concentrer leurs efforts pour dresser un état des lieux rigoureux de leurs principaux processus de travail. Tandis que le stockage cloud centralise les données brutes dans leur format d'origine, un entrepôt de données cloud permet aux entreprises de rassembler des données cloisonnées en vue de les analyser, comme le font les entrepôts de données classiques. Avec le cloud, pour un investissement en capital minime, les entreprises peuvent gérer de gros volumes de données, bénéficier d'une évolutivité quasi indéfinie et ne payer que pour leur consommation réelle. Les services cloud gérés vont encore plus loin : les équipes informatiques n'ont plus à se soucier de l'infrastructure sous-jacente. Les entreprises n'ont plus qu'à définir les questions stratégiques auxquelles elles souhaitent répondre ainsi que les données dont elles ont besoin pour trouver ces réponses.

Exemples :

- Quels sont les objectifs commerciaux principaux que doivent servir mes données ? Comprendre les interactions des utilisateurs avec mes systèmes, identifier des tendances, augmenter les ventes, fidéliser les clients ?
- D'où proviennent les données les plus importantes (transactions, journaux des serveurs, services cloud, appareils/IdO, réseaux sociaux) ? Sont-elles déjà importées dans l'espace de stockage cloud ?
- À quelle vitesse mon système doit-il intégrer les nouvelles données dans les rapports et les visualisations ?
- Le fait de prendre des décisions en s'appuyant sur les données est-il monnaie courante au sein de mon entreprise (pas uniquement pour les spécialistes informatiques et les analystes données) ? Qui doit avoir accès à la plate-forme d'analyse ?

Une fois que les objectifs commerciaux sont définis, les entreprises doivent identifier les sources des données d'entrée parmi les cloisonnements afin de les importer dans un entrepôt de données cloud pour les analyser. Voici une liste des sources d'entrée classiques :

Stockage cloud

Les données issues du stockage cloud peuvent être importées dans un entrepôt de données pour être analysées¹². À ce stade, il peut être utile de formaliser un schéma qui s'articule autour des questions stratégiques posées afin de structurer les données brutes avant l'analyse.

Bases de données analytiques et transactionnelles

Les données stockées dans des bases de données analytiques et transactionnelles peuvent être chargées par lots, ou importées en flux continu ligne par ligne, directement dans un entrepôt de données cloud.

Données stockées dans les services cloud

Les données stockées par les fournisseurs de SaaS courants peuvent être importées dans un entrepôt de données cloud, bien souvent de façon automatique.

Diffusion de données en flux continu

Les données des applications Web, mobiles et IdO peuvent contourner le stockage cloud et être importées directement en flux continu dans un entrepôt de données cloud (reportez-vous au [chapitre 3 : Intégration des données en temps réel](#)).

Gouvernance des données

La croissance exponentielle du volume global des données n'est pas le seul problème auquel sont confrontées les entreprises. Selon Forrester, dans leurs efforts de veille stratégique, les entreprises font face à plusieurs grands défis, notamment la rapide évolution des exigences en matière d'analyse et de création de rapports ainsi que la discordance entre les services commerciaux et informatiques¹³. De plus, avec le manque de spécialistes en science des données (**reportez-vous à la section "Et si tout le monde devenait analyste ?"**), les entreprises doivent suivre de nouvelles approches pour développer une expertise analytique.

Dès lors qu'il bénéficie d'un accès basé sur un rôle, un employé ou un développeur d'applications peut interroger les données stockées dans un entrepôt de données cloud, générer des rapports ou consulter des visualisations. Les entrepôts de données cloud permettent de gérer différents accès de façon personnalisée et ciblée. Les contrôles d'accès personnalisés ajoutés à la fonctionnalité d'audit complet participent à démocratiser la science des données, tout en préservant des dispositifs de sécurité. En effet, plus de la moitié des entreprises des États-Unis, d'Europe et d'Asie-Pacifique déclarent avoir mis en place des outils de veille stratégique en libre-service dans l'entreprise, être en train de le faire ou même être sur le point d'élargir leur application¹⁴.

ET SI TOUT LE MONDE DEVENAIT ANALYSTE ?

La responsabilité de tirer des conclusions statistiquement rigoureuses à partir de données était autrefois l'apanage des analystes professionnels. Toutefois, d'ici 2018 selon [McKinsey](#), "les États-Unis pourraient manquer de 140 000 à 190 000 personnes disposant de compétences analytiques approfondies, et de 1,5 million de responsables et analystes capables d'exploiter les analyses du Big Data pour prendre des décisions efficaces"¹⁵.

Et comme la concurrence s'intensifie, la plupart des entreprises diversifient leurs stratégies pour trouver des personnes compétentes. Le terme "*analyste citoyen*" (voir la définition de "citizen data scientist" dans cet [article d'InformationWeek](#)) désigne une personne qui tire parti de l'analyse des données, mais dont les missions principales ne concernent ni les statistiques, ni les analyses. Son travail vient judicieusement compléter celui des analystes internes, en particulier pour les entreprises qui souhaitent développer une culture de la science des données¹⁶.

Pour réussir, ces personnes doivent :

- avoir accès aux données ;
- être curieuses ;
- maîtriser SQL ;
- être spécialisées dans un domaine ;
- savoir collaborer.

RÉSUMÉ DU CHAPITRE 2

- 1 Le **stockage cloud** permet aux entreprises de recueillir tous types de données, structurées ou non, dans leur format d'origine. La centralisation des données dans un espace de stockage cloud permet de mettre en place les bases pour l'analyse, les détails seront traités lorsque l'entreprise devra interroger les données pour répondre à une question stratégique concrète.
- 2 Grâce aux **entrepôts de données cloud**, les entreprises peuvent rassembler des données cloisonnées disparates afin de les analyser, peu importe leur provenance (espace de stockage cloud, bases de données analytiques ou transactionnelles sur site ou dans le cloud, ou autres services cloud). Les entreprises peuvent interroger les données, générer des rapports et créer des visualisations sans avoir à gérer l'infrastructure sous-jacente.
- 3 L'**accès personnalisé** basé sur un rôle permet de démocratiser l'analyse dans l'entreprise. Un entrepôt de données cloud peut s'étendre à l'échelle de l'entreprise, mais peut également s'adapter en fonction de la structure de l'organisation.

TÉMOIGNAGE CLIENT

COLORADO CENTER FOR PERSONALIZED MEDICINE

Le Colorado Center for Personalized Medicine (CCPM) mène des recherches innovantes grâce à l'analyse de l'ADN de patients, afin de prédire les risques de maladies et d'élaborer des traitements ciblés basés sur les gènes d'un individu. Le CCPM s'appuie sur *Health Data Compass*, un entrepôt de données professionnel dédié. L'entrepôt *Health Data Compass* intègre les données génomiques des patients du CCPM et les dossiers médicaux électroniques issus d'UCHealth, de l'hôpital pour enfants du Colorado et de la faculté de médecine du Colorado. D'autres informations externes sont prises en comptes, telles que les déclarations de sinistres, les dossiers médicaux publics et les données environnementales.

Auparavant, *Health Data Compass* reposait sur un système classique sur site pour stocker et analyser les données. Cependant, cette approche s'est révélée coûteuse (frais de maintenance élevés) et non adaptée aux besoins du centre en matière d'analyse (sans parler de future croissance). Suite à un projet pilote complet de six mois, *Health Data Compass* a migré vers GCP et Tableau, qui peuvent à eux deux gérer des ensembles de données massifs et effectuer des analyses de données puissantes, tout en réduisant les coûts et en permettant une évolutivité facile au fur et à mesure de la croissance du CCPM. La décision du CCPM a été influencée par la capacité de GCP, et notamment de l'entrepôt de données BigQuery de Google Cloud, à respecter la conformité avec la loi HIPAA.

"Nous prenons très au sérieux notre responsabilité de protéger les données des patients. Par rapport aux systèmes sur site, Google Cloud Platform offre des avantages certains quant à la sécurité des données. Ce système nous permet également d'être en conformité avec la loi HIPAA", déclare Michael Ames, directeur associé de *Health Data Compass* et directeur de l'architecture d'entreprise pour CCPM¹⁷.

ENTREPRISE

*Colorado Center for
Personalized Medicine*

SECTEUR D'ACTIVITÉ

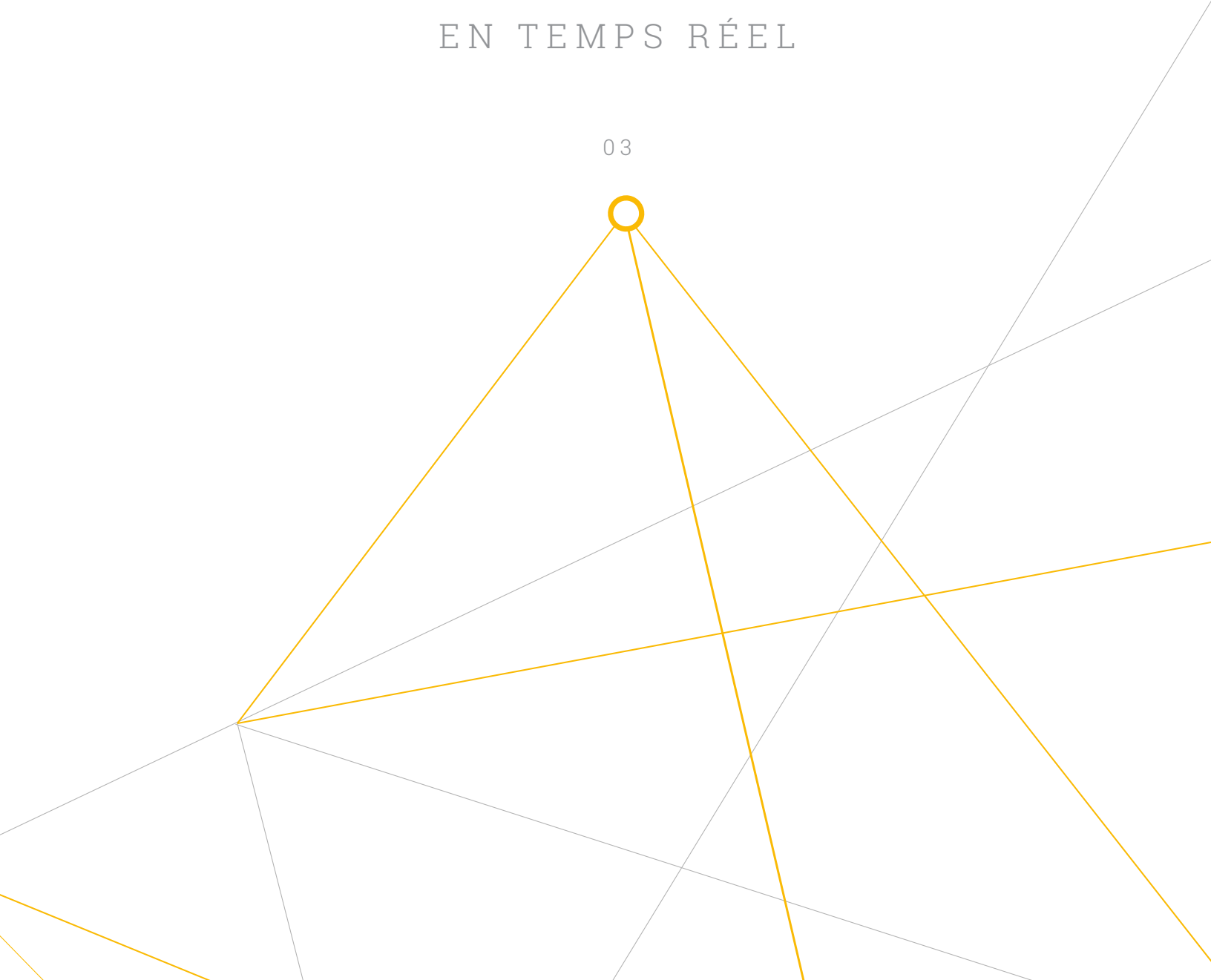
Santé

À PROPOS

Le centre Colorado Center for Personalized Medicine (CCPM) est le fruit d'un partenariat entre l'université du Colorado à Denver, les établissements UCHealth, l'hôpital pour enfants du Colorado et la faculté de médecine du Colorado. Il se situe à Denver, dans l'état du Colorado.

INTÉGRATION DES DONNÉES EN TEMPS RÉEL

03



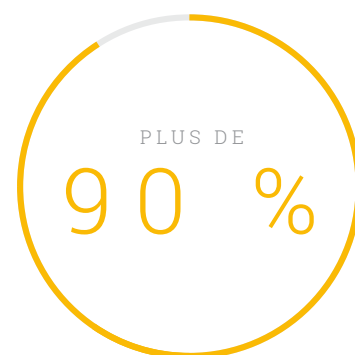
INTÉGRATION DES DONNÉES EN TEMPS RÉEL

03

Les analystes déclarent passer 50 à 80 % de leur temps à préparer les données afin de les analyser (nettoyage, organisation et interprétation des données)¹⁸. Le provisionnement de ressources et l'adaptation de la capacité des clusters de serveurs pour des charges de travail imprévisibles représentent un véritable fardeau pour les équipes chargées de la préparation des données sur site¹⁹.

Des services gérés pour moins de préparation

Grâce aux services cloud entièrement gérés, les équipes informatiques sont débarrassées des tâches relatives à l'infrastructure qui vont de pair avec la préparation des données à grande échelle et leur intégration. Imaginons un thermostat intelligent qui cherche à connaître les préférences de chaque équipe au sein d'un bureau et à s'y adapter. Lorsque le thermostat fonctionne, le cloud ingère les données d'utilisation brutes, telles que les paramètres de température et les niveaux de consommation énergétique tout au long de la journée. Tandis que les données affluent, un pipeline de traitement peut être appelé à la demande pour préparer les données brutes : il veille à ce que les données correspondent à une plage valable, il convertit la température et la consommation énergétique en unités exploitables et il formate les données temporelles. Le pipeline structure ces données de façon cohérente, puis il transfère les résultats transformés dans un entrepôt de données cloud. Les requêtes, les visualisations et les rapports sont disponibles instantanément.



des entreprises

**DÉCLARENT ÊTRE
INTÉRESSÉES PAR LE
DÉPLOIEMENT D'OUTILS
DE PRÉPARATION DES
DONNÉES EN LIBRE-
SERVICE POUR RÉPONDRE
AUX INITIATIVES
DU BIG DATA²⁰.**

Avec les services cloud entièrement gérés, les ressources de l'infrastructure nécessaires pour traiter le flux de travail sont automatiquement affectées avant d'être démobolisées juste après. Les entreprises ne paient que pour les ressources qu'elles utilisent et évitent ainsi de perdre du temps à faire des prévisions.

Vers l'analyse des données en temps réel

Les systèmes classiques visaient à analyser les données hors connexion "par lots". Avec la demande d'informations en temps réel, il est nécessaire d'adopter une nouvelle approche. Les systèmes d'analyse en flux continu basés sur le cloud sont conçus pour gérer en temps réel les flux de données provenant d'applications Web, de smartphones ou de millions de capteurs IoT. Des centaines de milliers de capteurs peuvent être installés sur les équipements industriels pour rendre compte de leur état brut et transmettre ces données dans le cloud à des fins de traitement et de surveillance. Des flux visuels peuvent être analysés en temps réel pour des applications de détection d'anomalies ou de reconnaissance de visages/d'objets par exemple. Des services cloud largement déployés, testés et éprouvés sont utilisés dans de tels cas, et l'analyse de flux de données peut être mise en place en quelques jours seulement.

Avec l'analyse de flux de données en temps réel, les données sont directement importées dans les pipelines de traitement. Les données transformées peuvent ensuite être intégrées dans un entrepôt de données cloud où elles attendent d'être interrogées, visualisées ou présentées sous forme de rapports en quelques secondes. Dans cette configuration, le pipeline de traitement joue le rôle d'intergiciel qui peut être appelé ou démobolisé à la demande, capable de rassembler les flux de données en temps réel à l'aide d'ensembles de données extraits de l'espace de stockage. Les données peuvent être structurées de façon flexible pour répondre aux questions stratégiques d'une entreprise lorsqu'elles se posent.

Ainsi, les entreprises disposent de deux méthodes complémentaires (l'analyse par lots et par flux) pour rassembler, préparer et intégrer les données de n'importe quelle source vers n'importe quelle cible. Avec les services cloud gérés, ces deux méthodes peuvent être utilisées facilement.

TIRER LE MEILLEUR PARTI DE VOS INVESTISSEMENTS EN BIG DATA

De nombreuses entreprises innovantes utilisent déjà le Big Data, souvent basé sur des outils Open Source, tels qu'[Apache Hadoop](#) et [Apache Spark](#). Pour ces entreprises, il est possible de protéger leurs investissements en matière d'employés qualifiés et d'outils tout en profitant au maximum des avantages de productivité qu'offre le cloud.

L'utilisation des outils Big Data Open Source est largement répandue. Elle est même en expansion. Nombreuses sont les entreprises dans le monde qui stockent de gros volumes de données non structurées sur des systèmes de fichiers cloud publics (y compris Hadoop). Plus d'un tiers des personnes interrogées aux États-Unis et en Europe (et plus de la moitié en Asie-Pacifique) déclarent avoir mis en place Hadoop, être en train de le faire, ou même être sur le point d'élargir son application (y compris HBASE, Accumulo, MapR, Cloudera et Hortonworks). De même, environ la moitié des personnes interrogées aux États-Unis et en Europe (jusqu'à 60 % en Asie-Pacifique) déclarent avoir mis en place des plates-formes de données en mémoire, être en train de le faire, ou même être sur le point d'élargir son application (y compris Apache Spark, SAP Hana, Kognitio, Terracotta et Gigaspaces).

Pour ces entreprises, le cloud offre deux options clés :

- Elles peuvent continuer à gérer des projets Big Data à l'aide d'outils Open Source qu'elles maîtrisent bien, et à la fois migrer vers des machines virtuelles dans le cloud. Elles bénéficieront de tous les avantages du cloud : élimination des investissements CapEx onéreux, passage à un modèle de facturation OpEx où seules les données stockées et traitées sont facturées, et évolutivité simple. Sachez que dans ce modèle de progression, les développeurs et les équipes informatiques sont toujours chargés de la gestion de leur stockage et de leurs pipelines de traitement des données. Toutefois, il s'agit de la stratégie la plus simple pour exploiter au mieux les compétences, les outils et les relations avec les fournisseurs existants.
- Le cloud propose des versions entièrement gérées pour les principaux outils Open Source de Big Data. Par exemple, exécuter [Apache Hadoop](#), [Apache Spark](#), [Apache Pig](#) et [Apache Hive](#) dans le cloud permet de supprimer les tâches courantes de gestion des données, telles que le déploiement, la journalisation et la surveillance²¹. Cette option convient parfaitement aux équipes qui cherchent à exploiter au mieux le système sur site et dans le cloud.

Non seulement chacune de ces options permet aux entreprises de protéger leurs investissements préexistants en déployant le Big Data, mais elles utilisent les atouts du cloud pour contrôler leurs coûts et gagner en flexibilité.

RÉSUMÉ DU CHAPITRE 3

- 1 Grâce aux **pipelines de traitement des données cloud**, les entreprises peuvent extraire, transformer, préparer et intégrer les données de n'importe quelle source vers n'importe quelle cible (sur site ou dans le cloud).
- 2 Grâce aux **approches sans serveur** appliquées à la préparation des données, l'infrastructure sous-jacente est intégralement gérée ; les ressources sont affectées automatiquement en fonction des besoins de chaque pipeline de traitement des données.
- 3 Les **analyses de flux cloud** permettent aux données provenant du Web, des applications mobiles ou de l'IdO d'être diffusées en temps réel dans les pipelines de traitement. À partir de là, les données peuvent être préparées et intégrées à un entrepôt de données pour offrir un aperçu de l'entreprise en temps réel.

ÉTUDE DE CAS

CITIBANK UK

Dans cette démonstration de faisabilité, l'objectif de l'équipe était de montrer avec quelle facilité Citibank pouvait utiliser [Google BigQuery](#) et [Google Cloud Pub/Sub](#) pour analyser et ingérer des données tick par tick quasiment en temps réel et des données historiques fournies par Thomson Reuters, soit l'équivalent d'environ 1 000 instruments financiers. Ce travail a été réalisé en collaboration avec Sean Micklethwaite, responsable du développement chez Citibank, et Sebastian Fuchs, spécialiste des solutions chez Thomson Reuters.

Sean Micklethwaite explique : "Nous souhaitons disposer d'une API que nous pourrions interroger à la demande, au sujet de données historiques, sans avoir à gérer notre propre entrepôt de données, ni assumer les coûts et les frais généraux associés. Nous avons également besoin de mises à jour en temps réel sur les données boursières ne dépassant pas le niveau de latence d'une personne. Avec Google Cloud, nous avons accès à toutes les données dont nous avons besoin via une plate-forme unique. BigQuery se charge de nos besoins en données tick par tick historiques et est assez puissant pour traiter les ticks bruts à haute fréquence sur des plages étendues. Cloud Pub/Sub répond à nos exigences de données en temps réel, et nous recevons toutes les données dans un format parfaitement cohérent."

Sebastian Fuchs ajoute : "Nous avons commencé à utiliser BigQuery sans avoir à fournir de gros efforts de planification de la capacité en amont. La capacité augmente en fonction des besoins, qu'il s'agisse de provisionnement de contenu ou du nombre de requêtes des utilisateurs."

ENTREPRISE

Citibank UK

SECTEUR D'ACTIVITÉ

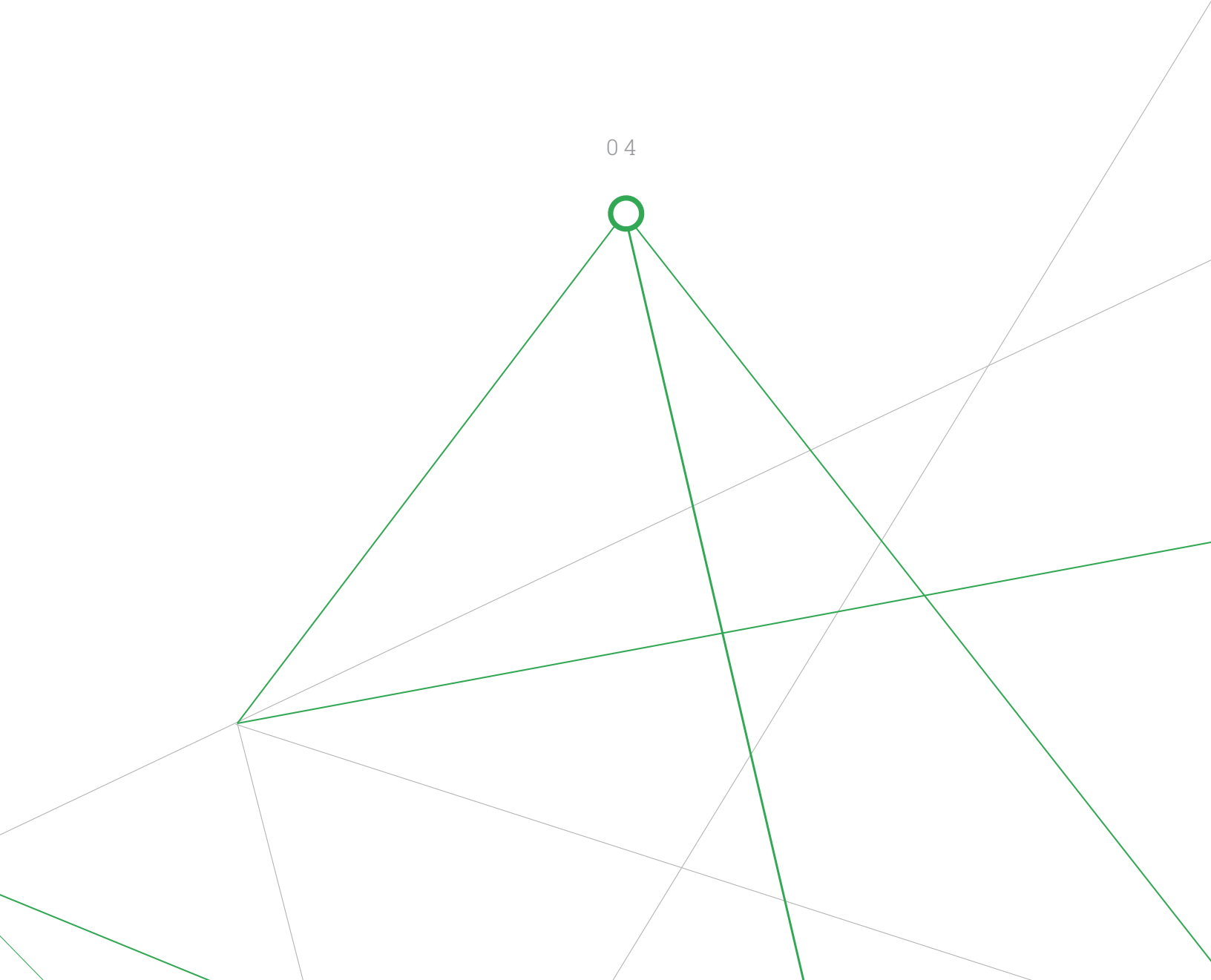
Services financiers

À PROPOS

Dans le cadre d'une démonstration de faisabilité, Google Cloud s'est associé à l'agence Thomson Reuters pour montrer aux services bancaires et boursiers de Citibank comment, en combinant les principales technologies de traitement des données de Google et les informations sur les marchés financiers de Thomson Reuters, ils pouvaient profiter de certains avantages.

MACHINE LEARNING ET IA

04



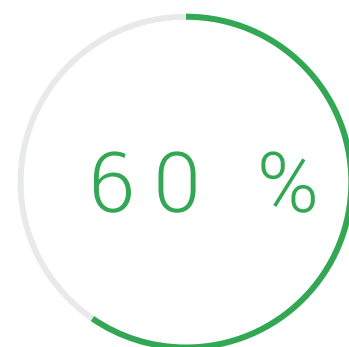
MACHINE LEARNING ET IA

Les récentes avancées en machine learning (ML) et en intelligence artificielle (IA) font régulièrement les titres des journaux. [Les ordinateurs ont battu les plus grands champions du monde](#) de Go, un jeu de plateau où le nombre de combinaisons est supérieur au nombre d'atomes dans l'univers²². Les ordinateurs [maîtrisent les jeux vidéo populaires](#) et, plus important encore, ils ont appris à [reconnaître les chats](#)²³. Plus récemment, des travaux sur l'IA ont permis de [réaliser d'importantes économies sur les coûts énergétiques](#), décrivant le machine learning comme un "système universel permettant de comprendre les dynamiques holomorphes"²⁴. Progressivement, ce système trouve des domaines d'application (et affiche des résultats prometteurs) dans de multiples secteurs d'activité.

Le concept de l'IA est simple : pour un logiciel, il s'agit de sa capacité à progresser sans programmation explicite. Plutôt que de mobiliser les développeurs à l'écriture manuelle d'un nouveau code, l'IA s'appuie sur des algorithmes capables de devenir "plus intelligents" en ingérant toujours plus de données du monde réel. La centralisation du stockage et la préparation des données dans le cloud (thèmes qui ont été traités respectivement aux chapitres 2 et 3) créent les bases idéales pour former et améliorer les modèles d'IA.

L'IA est une opportunité qui va bien au-delà de la simple automatisation des tâches qui étaient auparavant manuelles. Dans le cadre de la vente en ligne par exemple, les algorithmes de machine learning peuvent ingérer et analyser d'énormes volumes de données client pendant que les prospects parcourent la boutique en ligne ou l'application mobile du marchand. Plus le modèle ingère de données, plus il est capable de prévoir avec précision le moment que choisit un client donné pour réaliser un achat spécifique, ainsi que les motivations qui le poussent à passer à l'action. À terme, cet apprentissage devient intuitif et permet au marchand de présenter le bon produit auprès de la bonne personne au bon moment. Ce niveau de personnalisation, qui caractérisait autrefois les marchands des petites villes qui connaissaient les prénoms et les anniversaires des enfants de leurs clientes, est désormais possible à grande échelle.

04

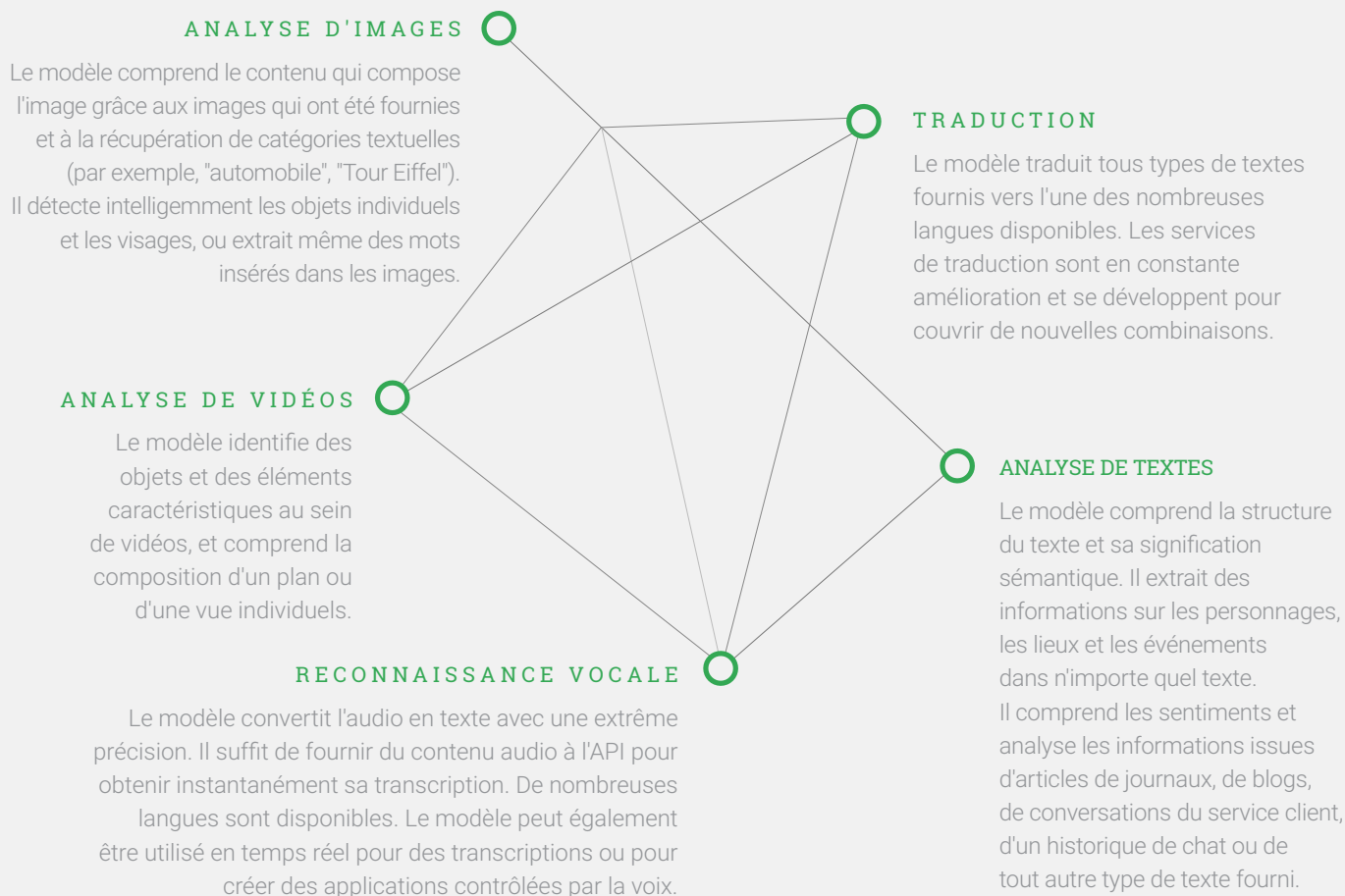


des personnes interrogées
SONT CONVAINCUES QUE
LA RÉUSSITE DE LEUR
ENTREPRISE DÉPEND
DES RÉSULTATS DE
LA MISE EN PLACE DU
MACHINE LEARNING²⁵.

De nombreuses petites entreprises, souvent tournées vers la technologie, récoltent déjà les fruits du ML. Pour d'autres entreprises plus anciennes, c'est une chance unique d'exploiter la richesse des données de leur historique²⁶. Les résultats du ML dépendent de la quantité de données brutes disponibles pour former les modèles d'apprentissage (**reportez-vous à la section "Modèles pré-éduqués : premiers pas vers l'IA"**). Les entreprises établies peuvent exploiter leurs données first party (tous les éléments dont elles sont propriétaires, des journaux système informatiques aux transactions financières en passant par les transcriptions d'appels du service client) afin de former et d'optimiser ces modèles, et ainsi d'obtenir des statistiques qui leur sont propres.

MODÈLES PRÉ-ÉDUQUÉS : PREMIERS PAS VERS L'IA

Le moyen le plus simple pour se lancer en IA consiste à utiliser les modèles de machine learning pré-éduqués, disponibles instantanément via le cloud. Aucune connaissance préalable en ML n'est nécessaire pour les utiliser. Ceux qui utilisent des applications grand public connaissent peut-être déjà ces capacités. En effet, dans leurs prédictions, certains modèles ont atteint des niveaux de précision qui surpassent l'homme.



Ces services sont généraux (ils ne sont pas attachés à des applications grand public) et ils peuvent être facilement intégrés à tous types d'applications, via des appels d'API. Les développeurs n'ont pas besoin de connaître les détails sous-jacents. Comme elles n'ont pas besoin de développer ces services en interne, les entreprises peuvent instantanément tirer profit de ces fonctionnalités innovantes en tant que service.

Les entreprises et les industriels dont la place n'est plus à faire sur le marché accumulent souvent plusieurs décennies de données first party : des transactions financières, des journaux système, des données brutes issues de la fabrication ou de la vente, des données d'e-commerce recueillies depuis des années ou les résultats de campagnes marketing. Une fois affinées et utilisées pour effectuer l'apprentissage des modèles de machine learning personnalisés, ces données deviennent une source de puissance prédictive. Plutôt que de reconverter des services caducs, ces établissements peuvent utiliser les données first party afin d'optimiser leur processus de travail pour leurs clients. Ils disposent d'une source puissante de différenciation.

Les cas d'utilisation s'étendent à de nombreux secteurs et révèlent de très prometteuses applications pour l'IA. Par exemple, la détection des fraudes pour les services financiers et la maintenance préventive pour les secteurs industriels témoignent de son potentiel. Le système est capable de faire apparaître les anomalies dans un océan de transactions et de journaux désordonnés, répondant ainsi à un besoin commun à de nombreux secteurs. Le machine learning rend également la classification plus facile, autre caractéristique bien utile. Cette capacité est illustrée par les suggestions de diagnostic et de traitement dans le domaine de la santé ou par les jugements sur la solvabilité.

Un cercle vertueux : collecte, préparation, apprentissage et prédiction

Les capacités que nous avons décrites aux chapitres 2 et 3 sont la pierre angulaire de l'entraînement des modèles de machine learning via les données first party. Lorsque les données brutes sont centralisées au préalable dans un espace de stockage cloud et un entrepôt de données cloud, les pipelines de données sans serveur peuvent extraire ces données en continu et les préparer pour entraîner les modèles de ML sur mesure. Étant donné que ces modèles de ML peuvent aussi être hébergés dans le cloud, ils sont à disposition immédiate des applications pour effectuer des prédictions. Cette boucle forme un cercle vertueux, où les modèles de ML hébergés dans le cloud ne cessent de progresser grâce aux nouvelles données d'apprentissage ; en retour, ces données permettent aux modèles d'être pertinents et constamment entraînés.

MESURER LES RÉSULTATS

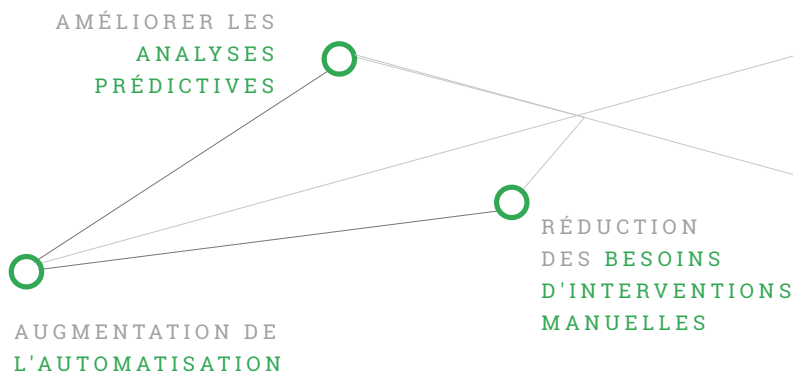
En partenariat avec l'agence de recherche M-Brain, Google Cloud a conduit une enquête auprès de 20 responsables informatiques et dirigeants d'entreprise qui ont mis en place des projets de machine learning pour connaître les bénéfices qu'ils en tirent. Voici les principaux avantages cités :

- Gain de temps
- Économies
- Gestion des risques améliorée
- Analyses de meilleure qualité
- Augmentation du chiffre d'affaires

Les autres avantages mentionnés sont les suivants : automatisation, amélioration des services et meilleure planification des stocks.²⁷

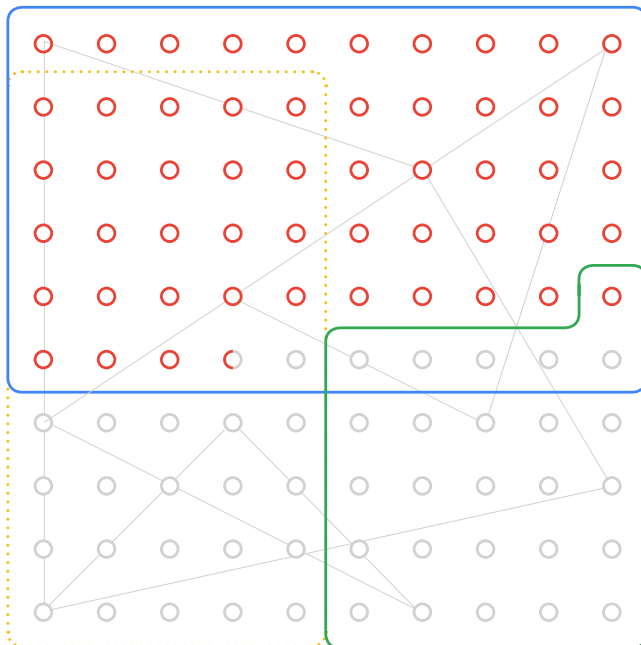
PRINCIPAUX BESOINS DES ENTREPRISES

des secteurs suivants²⁸ :



MACHINE LEARNING : LE NOUVEAU TERRAIN D'EXPÉRIMENTATION POUR SURPASSER LA CONCURRENCE

L'ère du machine learning est arrivée et bat même déjà son plein au sein des petites entreprises tournées vers la technologie, selon une nouvelle enquête menée auprès de dirigeants d'entreprise et d'experts en technologie par le *MIT Technology Review Custom*. Voici les principaux résultats²⁹ :



60 % des personnes interrogées ont déjà **mis en œuvre des stratégies de machine learning**.

PLUS DE 50 % des personnes ayant déjà mis en œuvre des stratégies de machine learning enregistrent un retour sur investissement.

45 % ont **amélioré la qualité de leurs analyses de données et de leurs informations**.

26 % indiquent bénéficier d'un **meilleur avantage concurrentiel**.

Téléchargez l'intégralité de l'enquête en cliquant ici. [📄](#)

RÉSUMÉ DU CHAPITRE 4

- 1 Le concept de l'intelligence artificielle et du machine learning sous-jacent est simple : il s'agit de la capacité d'un logiciel à progresser **sans programmation explicite**.
- 2 L'intelligence artificielle s'appuie sur d'importants volumes de données d'apprentissage pour offrir aux entreprises établies un **avantage unique**, celui d'exploiter la richesse des données qu'elles ont accumulées tout au long de leur existence.
- 3 Le stockage cloud, l'entreposage de données, l'intégration de données et l'analyse **constituent le fondement naturel** de l'intelligence artificielle et du machine learning, car ils mettent les données à disposition pour l'apprentissage et l'optimisation en temps réel, nourrissant ainsi un cercle vertueux, en perpétuelle progression.



CONCLUSION

Dans un univers où les données sont nombreuses et les réponses immédiates, savoir extraire la valeur des données (qu'importe leur source, leur taille et les exigences de rapidité) donne aux entreprises un avantage concurrentiel important.

La première étape consiste à repenser intégralement la stratégie de données. Les outils cloud d'aujourd'hui permettent aux entreprises de gérer d'énormes volumes de données (de tous types) plus efficacement et plus économiquement que jamais auparavant. Les entreprises qui adoptent une approche moderne lorsqu'il s'agit de recueillir, stocker, préparer et analyser leurs données disposent de solides bases pour tirer profit du machine learning et de l'intelligence artificielle. En définitive, ces nouvelles fonctionnalités permettront de renforcer les liens entre les entreprises et leurs clients en offrant aux entreprises des outils de prédiction pour chaque interaction.

DÉCOUVREZ LES AVANTAGES QUE PEUT APPORTER [GOOGLE CLOUD](#) À VOTRE ENTREPRISE.

*Espace de stockage
et bases de données*

Solutions pour le Big Data

Machine learning et IA

TRAVAUX CITÉS

1. 88 % des cadres supérieurs interrogés par Ernst & Young sont d'accord sur le fait que les données doivent être au cœur de la prise de décision, seuls 31 % ont considérablement restructuré leurs opérations pour y intégrer le Big Data et seulement 23 % ont mis en place des stratégies de données à l'échelle de l'entreprise. Ernst & Young, *Becoming an Analytics-Driven Organization* (Transformer l'entreprise grâce à l'analyse de données) (2015) ([lien](#)).
2. David Reinsel et al., *Data Age 2025: The Evolution of Data to Life-Critical* (Les données en 2025 : évolution vers une nécessité cruciale) (IDC, 2017) ([lien](#)).
3. Cade Metz, "Exclusive: Inside Google Spanner, the Largest Single Database on Earth", (Exclusif : au cœur de Google Spanner, la plus grande base de données unique au monde) *Wired* (26 novembre 2012) ([lien](#)).
Cade Metz, "Spanner, the Google Database that Measured Time, Is Now Open to Everyone", (Spanner, la base de données de Google qui défie le temps, est désormais accessible à tous) *Wired* (14 février 2017) ([lien](#)).
4. Robert McMillan, "Inside the Artificial Brain that's Remaking the Google Empire", (Au cœur du cerveau artificiel qui réinvente l'empire Google) *Wired* (16 juillet 2014) ([lien](#)). TensorFlow ([lien](#)).
5. Forrester, *Forrester's Global Business Technographics Data and Analytics Survey* (Enquête sur les données Technographics et les stratégies d'analyse des entreprises internationales, Forrester) (2016) ([lien](#)).
6. Leandro DalleMule et Thomas H. Davenport, "What's Your Data Strategy?" (Quelle est votre stratégie de données ?) *Harvard Business Review* (mai 2017) ([lien](#)).
7. John Gantz et David Reinsel, *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East* (Le monde numérique de 2020 : Big Data, ombres numériques supérieures et croissance soutenue en Extrême-Orient) (IDC, 2012) ([lien](#)).
8. Tracie Kambies et al., *Tech Trends 2017: Dark Analytics: Illuminating Opportunities Hidden within Unstructured Data* (Tech Trends 2017 – Dark Analytics : À la découverte de la valeur cachée des données) (Deloitte University Press, 2017) ([lien](#)).
9. *Tarifs de Google Cloud Storage*, Google Cloud Platform ([lien](#)).
10. Forrester, *Forrester's Global Business Technographics Data and Analytics Survey* (Enquête sur les données Technographics et les stratégies d'analyse des entreprises internationales, Forrester) (2016) ([lien](#)).
11. "Three Ways Marketing Organizations Can Make Data More Actionable", (Trois méthodes pour rendre les données d'entreprise exploitables) *Harvard Business Review* (9 août 2016) ([lien](#)).
12. Les entrepôts de données cloud modernes sont compatibles avec l'importation automatique de nombreux formats partiellement structurés (ils sont même capables de les interroger ponctuellement). En ce qui concerne les données non structurées qui doivent d'abord être transformées (l'ETL par exemple), reportez-vous au **chapitre 3 : Préparation des données**.
13. Forrester, *Forrester's Global Business Technographics Data and Analytics Survey* (Enquête sur les données Technographics et les stratégies d'analyse des entreprises internationales, Forrester) (2016) ([lien](#)).
14. Forrester, *Forrester's Global Business Technographics Data and Analytics Survey* (Enquête sur les données Technographics et les stratégies d'analyse des entreprises internationales, Forrester) (2016) ([lien](#)).
15. James Manyika et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity* (Big Data : la prochaine frontière pour l'innovation, la compétitivité et la productivité) (McKinsey Global Institute, 2011) ([lien](#)).
16. Lisa Morgan, "Citizen Data Scientists: 7 Ways to Harness Talent", (Analystes citoyens : sept façons d'exploiter au mieux les compétences) *InformationWeek* (24 juillet 2015) ([lien](#)).

17. *Colorado Center for Personalized Medicine: Improving Healthcare by Integrating Patient Records and Genetic Data Using Google Cloud Platform and Tableau* (Colorado Center for Personalized Medicine : améliorer les soins en intégrant les dossiers médicaux et les données génétiques à l'aide de Google Cloud Platform et de Tableau) (Google Cloud Platform, 2017) ([lien](#)).
18. Steve Lohr, "For Big-Data Scientists, Janitor Work Is Key Hurdle to Insights", (Pour les analystes Big Data, la préparation freine l'obtention d'informations) *New York Times* (17 août 2014) ([lien](#)).
19. Forrester, *Forrester's Global Business Technographics Data and Analytics Survey* (Enquête sur les données Technographics et les stratégies d'analyse des entreprises internationales, Forrester) (2016) ([lien](#)).
20. Forrester, *Forrester's Global Business Technographics Data and Analytics Survey* (Enquête sur les données Technographics et les stratégies d'analyse des entreprises internationales, Forrester) (2016) ([lien](#)).
21. *Apache Hadoop*, The Apache Software Foundation ([lien](#)).
Apache Spark, The Apache Software Foundation ([lien](#)).
Apache Pig, The Apache Software Foundation ([lien](#)).
Apache Hive, The Apache Software Foundation ([lien](#)).
22. Paul Mozur, "Google's A.I. Program Rattles Chinese Go-Master As It Wins Match", (Défaite du champion chinois de Go contre le programme d'intelligence artificielle développé par Google) *New York Times* (25 mai 2017) ([lien](#)).
23. Nicola Twilley, "Artificial Intelligence Goes to the Arcade", (L'intelligence artificielle se mesure aux jeux d'arcade) *The New Yorker* (25 février 2015) ([lien](#)).
 John Markoff, "How Many Computers to Identify A Cat? 16,000", (Combien d'ordinateurs pour identifier un chat ? Réponse : 16 000) *The New Yorker* (25 juin 2012) ([lien](#)).
24. James Vincent, "Google Uses DeepMind AI to Cut Data Center Energy Bills", (Google utilise l'intelligence artificielle DeepMind pour réduire les factures énergétiques des centres de données) *The Verge* (21 juillet 2016) ([lien](#)).
25. *Harvard Business Review Analytic Services Global Data and Analytics Survey* (Enquête mondiale sur l'analyse et les données, Harvard Business Review Analytic Services), sponsorisée par Google (2017).
26. Une enquête menée par *MIT Technology Review* montre que les petites entreprises sont en voie d'adoption du machine learning et en tirent profit : sur un panel de 375 sondés, 60 % d'entre eux (parmi lesquels environ deux tiers comptent moins de 1 000 employés) sont issus des secteurs de la technologie, des affaires et des services financiers. *MIT Technology Review Custom et Google Cloud, Machine Learning: The New Proving Ground for Competitive Advantage* (Machine learning : le nouveau terrain d'expérimentation pour surpasser la concurrence) (2017) ([lien](#)).
27. Anna Rader, *Machine Learning Initiatives Across Industries: Practical Lessons from IT Executives* (Initiatives autour du machine learning dans différents secteurs : leçons pratiques de cadres informatiques) (M-Brain, sponsorisé par Google, 2017) ([lien](#)).
28. Anna Rader et Irida Jano, *Machine Learning Market Research: How Leading Industries Are Adopting AI* (Étude de marché sur le machine learning : comment les grands industriels se tournent vers l'intelligence artificielle) (M-Brain 2017) ([lien](#)).
29. *MIT Technology Review Custom et Google Cloud, Machine Learning: The New Proving Ground for Competitive Advantage* (Machine learning : le nouveau terrain d'expérimentation pour surpasser la concurrence) (2017) ([lien](#)).



© 2017 Google Inc.
1600 Amphitheatre Parkway, Mountain View, CA 94043, États-Unis