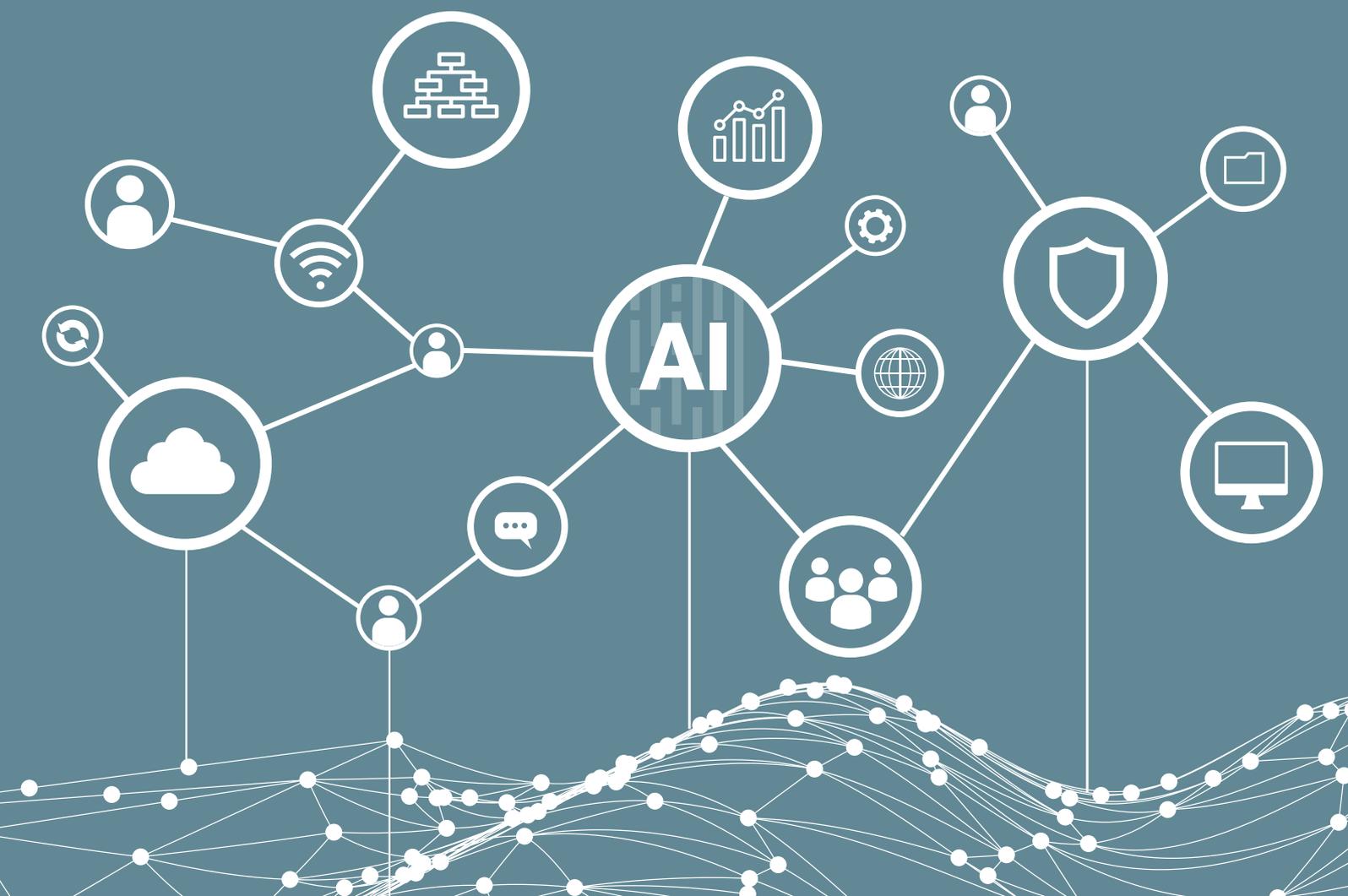


Mettre au point une organisation puissante de données et d'IA



Avant-propos

« Mettre au point une organisation puissante de données et d'IA » est un rapport MIT Technology Review Insights sponsorisé par Databricks. Pour réaliser ce rapport, MIT Technology Review Insights a mené une enquête internationale auprès de 351 Chief Data Officers, directeurs de l'analytics, DSI et autres décideurs technologiques. Les participants sont répartis de façon homogène entre l'Amérique du Nord, l'Europe et l'Asie-Pacifique. 14 secteurs d'activité sont représentés dans l'échantillon et toutes les personnes interrogées travaillent dans des organisations qui enregistrent un chiffre d'affaires annuel de plus d'un milliard de dollars. La recherche comprend également une série d'interviews avec des dirigeants en charge de la gestion des données, de l'analytics et de l'infrastructure associée au sein de leur organisation. Ce rapport a été rédigé par Denis McCauley, corrigé par Francesca Fanshawe et réalisé par Nicola Crepaldi. La recherche a été menée en toute indépendance sur le plan éditorial et les avis exprimés sont ceux de MIT Technology Review Insights.

Nous souhaitons remercier les personnes suivantes de nous avoir consacré du temps et apporté leur éclairage :

Patrick Baginski, Senior Director Data Science, McDonald's (États-Unis)

Bob Darin, Chief Data Officer, CVS Health, et Chief Analytics Officer, CVS Pharmacy (États-Unis)

Naveen Jayaraman, Vice President – Data, CRM & Analytics, L'Oréal (États-Unis)

Michel Lutz, Group Chief Data Officer, Total (France)

Mainak Mazumdar, Directeur des données et de la recherche, Nielsen (États-Unis)

Andy McQuarrie, Chief Technology Officer, Hivery (Australie)

Sol Rashidi, Chief Analytics Officer, The Estée Lauder Companies (États-Unis)

Ashwin Sinha, Chief Data and Analytics Officer, Macquarie Bank (Australie)

Don Vu, Chief Data Officer, Northwestern Mutual (États-Unis)

TABLE DES MATIÈRES

01 Résumé	4
02 Croissance et complexité	6
La vision de Databricks :	
La montée en puissance de l'effet Lakehouse	7
03 Harmonisation et réalisation d'une stratégie	9
Les très bons élèves en matière de données	11
Nielsen : data transformation pour une entreprise qui dépend fortement des données.....	13
04 Mise à l'échelle de l'analytics et apprentissage automatique (ML)	14
Changement de paradigme chez CVS Health	15
Obstacles à l'évolutivité	16
Protéger le retour sur investissement.....	17
Technologie, démocratie et culture.....	18
05 Visions de l'avenir	19
Les souhaits des directeurs des données pour une nouvelle architecture.....	19
06 Conclusion	21



Résumé

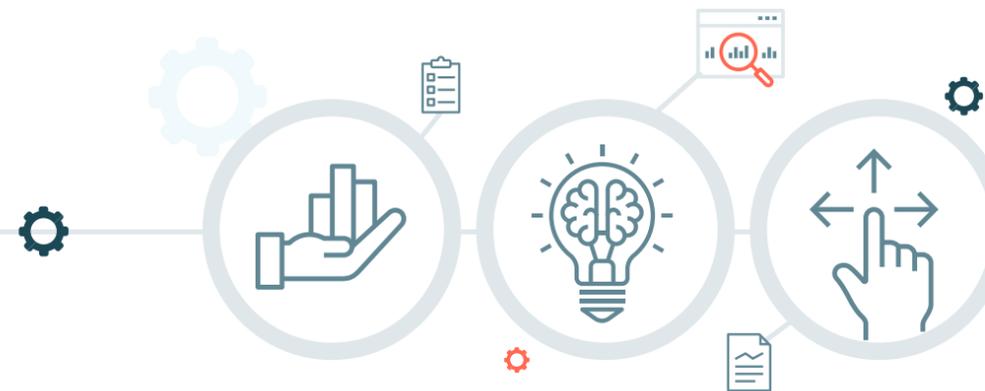
Les responsables de l'expérience client et les conseils d'administration reconnaissent que d'un point de vue stratégique, leur organisation doit absolument être capable de générer des insights exploitables, souvent en temps réel, à partir de données. Un constat largement appuyé par la fuite accélérée des consommateurs vers le digital au cours de l'année de crise qui vient de s'écouler. Afin d'être axées sur les données, les entreprises déploient des technologies cloud de plus en plus avancées, incluant des outils d'analytics et des fonctions d'apprentissage automatique (ML). Toutefois, ces outils n'apportent pas grand chose s'ils ne sont pas accompagnés de données à la fois nombreuses, de haute qualité et facilement accessibles.

C'est pourquoi une gestion des données efficace est l'un des piliers d'une organisation axée sur les données. Mais la gestion des données au sein d'une entreprise peut s'avérer très complexe. Tandis que de nouvelles technologies de données sont mises en œuvre, les systèmes hérités et silos de données deviennent un fardeau de plus en plus difficile à gérer, sauf s'ils peuvent être intégrés ou clôturés. Cette fragmentation de l'architecture est un véritable casse-tête pour de nombreux directeurs des données (CDO), pas seulement en raison des silos, mais aussi de la multitude d'outils on-premise et cloud utilisés par de nombreuses organisations. En plus de la mauvaise qualité des données, ces problèmes mis bout à bout privent les plateformes de données des organisations – ainsi que les modèles d'apprentissage automatique (ML) et d'analytics qu'elles prennent en charge – de la vitesse et de l'ampleur nécessaires à l'obtention des résultats commerciaux souhaités. Pour comprendre comment la gestion des données et les technologies sur lesquelles elle se fonde



évoquent face à de tels défis, MIT Technology Review Insights a interrogé 351 CDO (Chief Data Officers) et CAO (Chief Analytics Officers) – nous les désignons parfois par le terme « leaders des données » dans ce rapport – ainsi que des directeurs des systèmes d'information (DSI), directeurs techniques (CTO) et autres responsables technologiques. Nous avons également mené des entretiens plus poussés avec plusieurs autres décideurs technologiques. Voici les principales conclusions de cette étude :

- **Seulement 13 % des organisations parviennent à mettre en œuvre leur stratégie de données.** Ce petit groupe de « très bons élèves » a des résultats commerciaux quantifiables dans toute l'entreprise. Ces organisations réussissent car elles sont attentives aux piliers d'une bonne gestion et d'une bonne architecture des données, ce qui leur permet de « démocratiser » les données et de tirer profit de l'apprentissage automatique (ML). Ces piliers permettent de réduire la duplication des données, de simplifier l'accès à des données pertinentes, et de renforcer la capacité à traiter de grandes quantités de données à des vitesses élevées et à obtenir une meilleure qualité de données. Les très bons élèves ont massivement adopté le cloud. Pour 74 % d'entre eux, au moins la moitié de leurs services ou infrastructures de données fonctionne dans un environnement cloud.



Les priorités des organisations en matière de données pour les deux prochaines années se situent dans trois domaines, tous couverts par l'adoption élargie de plateformes de cloud : **l'amélioration de la gestion des données, le renforcement de l'analytics de données et du ML, et le développement de tous les types de données d'entreprise**, y compris les données en continu (streaming) et non structurées.

- **La collaboration basée sur les nouvelles technologies crée une culture de travail axée sur les données.** Les CDO interrogés dans le cadre de l'étude accordent beaucoup d'importance à la démocratisation de l'analytics et des fonctions de ML. En les poussant jusqu'au bout grâce à des technologies de données perfectionnées, les utilisateurs finaux peuvent prendre des décisions business plus éclairées. C'est d'ailleurs la définition même d'une culture forte des données. Pour ce faire, il faut une architecture de données moderne. En résumé, comme le disait un CDO, une gestion des données est réussie quand les bons utilisateurs ont accès aux bonnes données afin de générer rapidement des insights apportant quelque chose sur le plan commercial.

- **L'impact business du ML est limité par les difficultés de gestion de son cycle de vie intégral.** La mise à l'échelle de scénarios d'utilisation du ML est un procédé terriblement complexe pour de nombreuses organisations. D'après 55 % des personnes interrogées, la première difficulté est l'absence d'un endroit centralisé dédié au stockage et à la découverte de modèles de ML. Si on y ajoute les transferts, sources d'erreurs, entre la data science et la production, et un manque de ressources en ML qualifiées – deux aspects cités par 39 % des sondés – il en résulte une collaboration difficile entre le ML, les données et les utilisateurs professionnels.

- **Les entreprises recherchent des plateformes de cloud native qui prennent en charge la gestion des données, l'analytics et l'apprentissage automatique (ML).** Les priorités des organisations en matière de données pour les deux prochaines années se situent dans trois domaines, tous couverts par l'adoption élargie de plateformes de cloud : l'amélioration de la gestion des données, le renforcement de l'analytics de données et du ML, et le développement de tous les types de données d'entreprise, y compris les données en continu (streaming) et non structurées. Quant aux « mauvais élèves » – les organisations qui ont du mal à mettre en œuvre une stratégie de données – l'amélioration de la gestion des données (citée par 59 % des répondants de ce groupe) éclipsé toutes les autres priorités. A contrario, la plupart des très bons élèves (53 %) souhaitent avant tout améliorer leurs scénarios d'utilisation de ML.

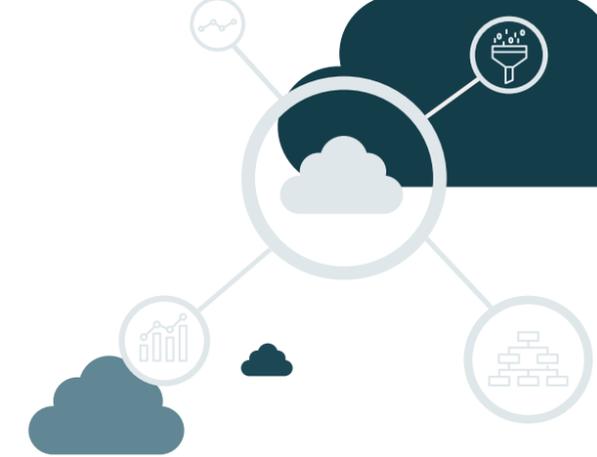
- **Les futures stratégies d'architecture des données exigent avant tout des normes ouvertes.** Si les personnes interrogées pouvaient mettre au point une nouvelle architecture de données pour leur entreprise, le principal avantage par rapport à l'architecture actuelle serait une meilleure inclusion des normes open source et des formats de données en accès libre (open data). Les dirigeants ont pris conscience de l'importance des normes open source pour accélérer l'innovation et permettre d'effectuer un choix dans la rentabilisation des meilleurs outils tiers. Sans surprise, une sécurité et une gouvernance renforcées se situent elles aussi tout en haut de la liste des exigences des répondants.

02 Croissance et complexité

Les organisations ne gèrent plus du tout leurs données de la même manière. Un changement qui a eu lieu à une vitesse à la fois stupéfiante et frustrante. Autrefois considérées par les dirigeants comme un sous-produit opérationnel, les données sont aujourd'hui à leurs yeux un moteur essentiel de valeur commerciale. Les volumes de données générées continuent d'augmenter rapidement pour les données structurées, semi-structurées et non structurées que les entreprises sont désormais en mesure de stocker pour les analyser.

Il n'y a pas si longtemps, les organisations faisaient appel à quelques géants de la technologie pour répondre à leurs besoins en outils et infrastructure de données. Aujourd'hui, les clients d'entreprise ont le choix entre des centaines de fournisseurs dans un large écosystème. Ces acteurs mettent régulièrement au point de nouveaux outils d'analytics de données – désormais souvent générés par l'apprentissage automatique – offrant une vitesse, une profondeur et une sophistication sans précédent. Les fournisseurs de services clouds en croissance perpétuelle fournissent aux organisations un vaste espace de stockage et un énorme pouvoir de traitement de leurs données, et ce en étant de plus en plus rentables. Dernier point, mais non des moindres, de nouveaux rôles et structures ont émergé à différents niveaux – comme le montre l'augmentation, entre autres, des Directeurs des données (Chief data officers / CDO) et Directeurs de l'analytics (Chief analytics officers / CAO) – afin d'orienter les capacités de l'organisation vers la création d'une nouvelle valeur business en accord avec les objectifs stratégiques.

« Avant, j'avais du mal à obtenir des données concernant de nombreux éléments de notre expérience client. Et en plus, c'était cher », déclare Bob Darin, Directeur des données de CVS Health (et Directeur de l'analytics de



Le cloud, autrefois considéré comme un environnement technologique optionnel, est devenu le pilier de la modernisation de la gestion des données : **63 % des répondants utilisent largement des services ou une infrastructure cloud dans leur architecture de données.**

CVS Pharmacy). « Maintenant, je peux obtenir des insights jusqu'ici impossibles à saisir sur nos clients, sur notre chaîne de valeur, ou encore sur la manière dont les gens travaillent. Nous disposons de tous les outils nécessaires à l'analyse de ces données dans les bonnes proportions, et le coût de ces outils est en baisse. Ce qui nous permet de générer des insights à grande échelle et de les intégrer à nos flux de travail patients et clients, dans le but de fournir une expérience plus personnalisée et pertinente. »

Le cloud, autrefois considéré comme un environnement technologique optionnel, est devenu le pilier de la modernisation de la gestion des données, pour une puissance informatique et un stockage de plus en plus importants à un coût de moins en moins élevé. 63 % des entreprises visées par l'étude utilisent largement des services ou une infrastructure de cloud dans leur architecture de données. Parmi elles, seulement un peu plus d'un tiers (34 %) utilisent plusieurs clouds.

Pourtant, la gestion des données est source de frustrations. Tandis que les entreprises cherchent à améliorer leurs plateformes de données, bon nombre d'entre elles restent aux prises avec des silos on-premise hérités résistant à toute intégration simple et engendrant des coûts élevés ou des problèmes dus à la duplication et à la mauvaise qualité des données. Le tout entraîne une

Point de vue des partenaires

Point de vue de Databricks : La montée en puissance de l'effet Lakehouse

Toutes les entreprises sont soumises à l'attrait de devenir « data-driven ». Elles accordent de plus en plus d'importance à l'IA pour concrétiser l'immense potentiel commercial que cette dernière peut offrir. Mais comme précisé dans ce rapport, seulement 13 % des organisations parviennent aujourd'hui à mettre en œuvre leur stratégie de données d'entreprise. Les décideurs des données et de l'analytics estiment que leur réussite est largement due à une bonne maîtrise des bases de la gestion des données. Alors pourquoi tant d'autres n'y parviennent pas ?

Tout part de l'architecture de données. D'après l'étude, les organisations doivent mettre au point quatre piles différentes de charges de travail en matière de données : business analytics, ingénierie des données, streaming et ML. Ces quatre piles

exigent des technologies très différentes et, malheureusement, elles ne fonctionnent pas toujours très bien ensemble.

L'écosystème technologique lié aux entrepôts et lacs de données vient compliquer l'architecture. Sa gestion finit par être onéreuse et par nécessiter des ressources considérables. Une complexité qui a des conséquences sur les équipes spécialisées dans les données. Les silos de données et ceux liés à l'organisation peuvent malencontreusement ralentir la communication, entraver l'innovation et créer des objectifs différents au sein des équipes. Il en résulte de multiples copies de données, une absence de modèle cohérent en termes de sécurité et de gouvernance, des systèmes fermés et des équipes de données moins productives.

La suite à la page suivante

infrastructure de données très complexe. Le cloud, malgré son impact révolutionnaire, peut également être vecteur de complexité si les organisations continuent de stocker leurs données chez de multiples fournisseurs afin de se protéger de tout verrouillage de leur part, de répondre à des besoins régionaux, ou bien encore de tirer profit de solutions best-of-breed optimisées. Les architectures de données ont par ailleurs évolué dans un laps de temps relativement court. Ainsi, les organisations risquent d'utiliser simultanément des bases de données on-premise, des entrepôts ou des lacs de données, ou bien d'autres architectures de données émergentes avec différents outils cloud en vue de la configuration, de la gouvernance ou d'autres fonctions.

« Les architectures sont devenues vraiment compliquées, mais seulement parce que nous avons tendance à les compliquer nous-mêmes », déclare Sol Rashidi, Directrice de l'analytics chez The Estée Lauder Companies. « Et ce, parce que nous oublions ce qui compte le plus. Trop souvent, nous utilisons les meilleures technologies et les

plateformes dernier cri en pensant qu'elles vont résoudre le problème. Mais si l'entreprise n'est pas prête à exploiter les outils, n'a pas la maturité pour extraire des insights et n'a fixé de processus ni de stratégie, elle ne fait que contribuer à une architecture spaghetti. »

Si les organisations sont incapables de gérer la complexité, il en résulte souvent un mélange d'occasions manquées (par exemple, si les scénarios d'utilisation du ML échouent à fournir des retours), de coûts plus élevés (notamment à cause de l'administration et de la prise en charge de multiples systèmes imbriqués), de difficultés à répondre aux exigences réglementaires croissantes sur les données et, enfin, d'exposition considérable à la concurrence.

Cependant, notre recherche montre bien que l'enthousiasme et l'optimisme l'emportent sur un éventuel sentiment de frustration parmi les décideurs interrogés quant à leur capacité, présente et future, à gérer efficacement les données pour leur entreprise.

Point de vue des partenaires

En parallèle de cela, le ML reste un objectif hors de portée. Avec l'émergence de l'architecture lakehouse, les organisations ne sont plus soumises aux limites et à la complexité des architectures héritées. En alliant la performance, la fiabilité et la gouvernance des entrepôts de données à l'extensibilité, les faibles coûts et la flexibilité de la charge de travail du lac de données, l'architecture lakehouse apporte une conception flexible et très performante pour différentes applications de données – comme le streaming en temps réel, le traitement par lots, l'analyse SQL, la science des données et le ML.

Chez Databricks, nous donnons vie à l'architecture lakehouse grâce à Databricks Lakehouse Platform. Delta Lake est le principal catalyseur derrière cette innovation. Delta Lake est au cœur de la plateforme. Il crée des lacs de données organisés qui apportent directement au lac de données existant la fiabilité, la performance et la gouvernance des entrepôts de données. Il permet aux organisations de mieux appréhender la gestion des données à l'échelle de l'entreprise.

La plateforme Lakehouse de Databricks excelle pour trois raisons :

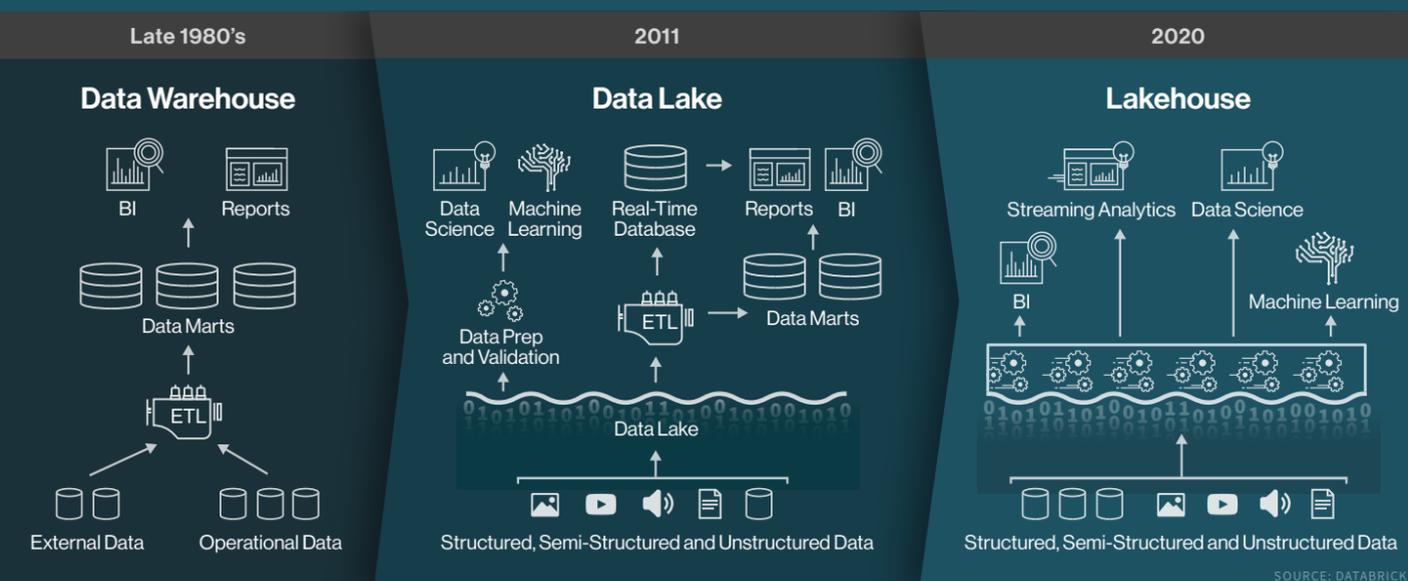
C'est simple : Les données ne doivent exister qu'une seule fois pour prendre en charge l'ensemble des charges de travail sur une plateforme commune.

C'est ouvert : À partir de normes ouvertes et open source, il est facile de travailler avec les outils existants et d'éviter les formats propriétaires.

C'est collaboratif : Les analystes, les ingénieurs de données et les data scientists peuvent travailler ensemble et donc plus efficacement.

Les économies, les rendements et les gains de productivité offerts par la plateforme Lakehouse de Databricks ont d'ores et déjà un impact global sur les entreprises, tous secteurs et toutes localisations géographiques confondus. Évitant les architectures trop complexes, Databricks fournit une fondation commune basée sur le cloud pour toutes les données et charges de travail, pour l'ensemble des fournisseurs de cloud. Les décideurs peuvent alors favoriser une culture axée sur les données créant avant tout de la valeur en supprimant, grâce à une maintenance préventive, le train-train quotidien lié à la planification et à toutes sa complexité.

De l'analytics de streaming vidéo à la valeur vie client, et de la prévention contre les maladies à l'exploration spatiale, les données sont essentielles. Comprendre les données, c'est ouvrir des portes.



SOURCE: DATABRICKS

03 Alignement et mise en œuvre stratégiques

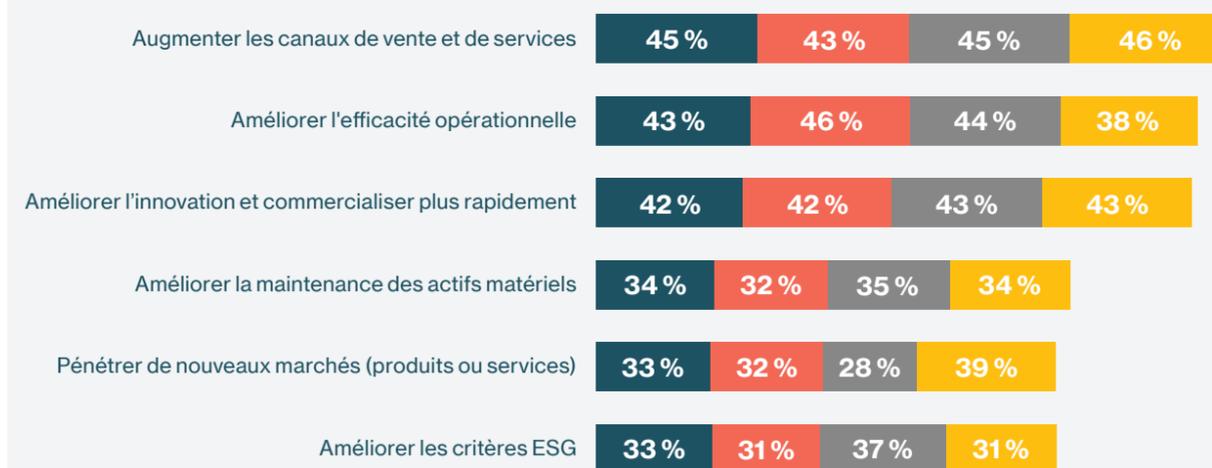


Dans un contexte de ralentissement économique jamais vu depuis près d'un siècle, on pourrait s'attendre à ce que les entreprises freinent leurs ambitions et se concentrent sur leur résultat net. Pourtant, bon nombre des entreprises représentées dans notre enquête semblent axées sur la croissance. Interrogés sur les objectifs commerciaux les plus importants qu'ils ont fixés pour leur stratégie de données d'entreprise au cours des deux prochaines années, les répondants sont plus nombreux à mettre l'accent sur la croissance du chiffre d'affaires, sous la forme d'un élargissement des canaux de vente et de service (cité par 45 %), que sur l'amélioration de l'efficacité (43 %). Suivie de près (42 %) par une meilleure innovation et une commercialisation plus rapide des produits nouveaux ou améliorés.

Pour les deux prochaines années, les principales initiatives en matière de données des entreprises interrogées renvoient clairement à une stratégie commerciale axée sur la croissance. Ces entreprises ont également conscience qu'il faut rapidement améliorer la gestion des données pour appuyer ces objectifs business. L'adoption généralisée de plateformes cloud natives viendra en renfort, facilitant la réalisation de ces objectifs ainsi que d'autres initiatives.

La première priorité, mentionnée par 48 % des répondants, est une meilleure gestion des données grâce à l'amélioration de leur qualité et de leur traitement. (Un chiffre qui passe à 74 % parmi les personnes interrogées travaillant dans des entreprises du secteur énergétique - pétrole et gaz - et à 67 % dans celles qui produisent des

Graphique 1 : Principaux objectifs commerciaux des entreprises dans le cadre de leur stratégie de données pour les deux prochaines années (réponses les plus fréquentes ; pourcentage des répondants)



Enquête MIT Technology Review Insights, 2021

■ Total ■ Amérique du Nord ■ Europe ■ Asie-Pacifique

Parvenir à une meilleure gestion des données en améliorant leur qualité et leur traitement permet aux objectifs axés sur la croissance, comme ceux pilotés par le ML, d'avancer à un rythme soutenu.

biens de consommation.) Ces démarches permettent aux objectifs axés sur la croissance, tels que ceux pilotés par le ML, d'avancer à un rythme soutenu. Pour Hivery, une société spécialisée dans les technologies pour points de vente et basée en Australie, dont les produits sont générés par l'Intelligence Artificielle (IA), la qualité des données clients est encore plus importante que sa capacité à intégrer de gros volumes. En réalité, explique Andy McQuarrie, CTO d'Hivery, plus nos données clients sont propres, moins il y a de problèmes d'intégration.

Si elles sont mises en œuvre, les autres priorités en matière de données des entreprises sondées – augmentation de l'adoption des plateformes de cloud (43 %), amélioration de l'analytics de données (43 %) et généralisation de l'application du ML (42 %) – apporteront aux équipes de données plus de capacité, de pouvoir et d'évolutivité pour, entre autres, exploiter rapidement les nouvelles opportunités de ventes et de service mais aussi pour soutenir le développement de produits liés aux données. Par ailleurs, ces autres priorités soutiennent aussi parfaitement l'objectif d'amélioration de l'efficacité opérationnelle. Autre priorité (citée par 38 %

des répondants) : l'utilisation généralisée des données en continu, non structurées et d'autres types de données.

Aujourd'hui, l'adéquation de cette stratégie de données vis-à-vis des objectifs commerciaux en général semble aller de soi. Mais cela n'a pas toujours été le cas. Selon Don Vu, CDO de la société américaine de services financiers Northwestern Mutual, l'adéquation des données vis-à-vis de la stratégie commerciale s'est nettement renforcée dans de nombreuses entreprises au fur et à mesure que les CDO ont exercé leur influence et que les responsabilités en matière de données ont été regroupées au sein de structures organisationnelles rationalisées. Dans l'entreprise, ajoute Don Vu, « les collaborateurs connaissaient l'importance de cette adéquation, mais celle-ci s'est cristallisée lorsque nos équipes ont réfléchi sérieusement à la manière de mettre en œuvre les différentes initiatives dans le cadre de la stratégie commerciale. Les salariés ont donc fait le rapprochement entre la stratégie commerciale et des notions comme la confiance envers les sources fondamentales de vérité, ou encore la démocratisation de l'utilisation des données. »

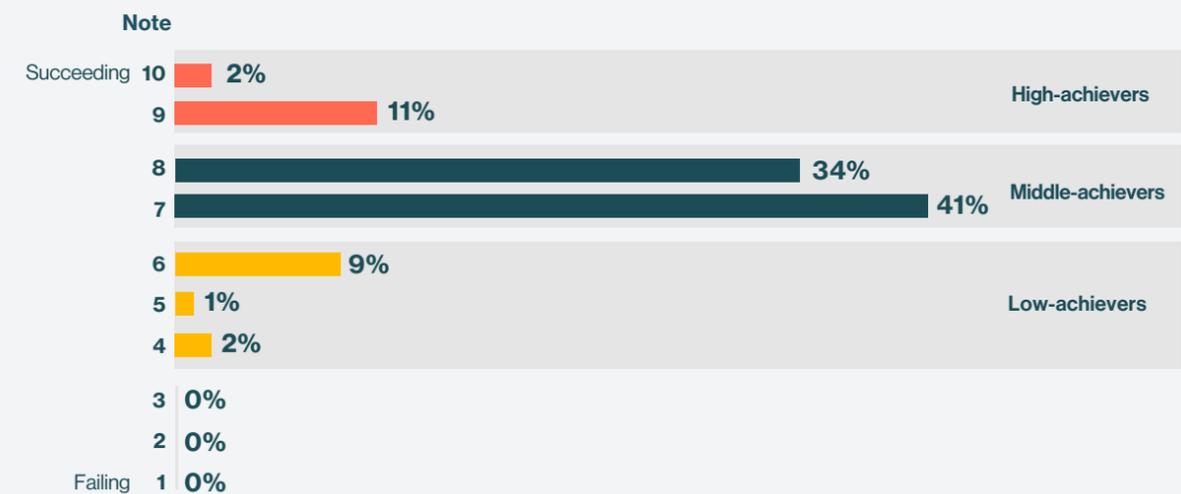
Les très bons élèves en matière de données

Très peu de grandes entreprises excellent dans la gestion des données. C'est ce que montre l'étude. Seulement 13 % des répondants estiment que leur organisation enregistre une très bonne performance dans la mise en œuvre d'une stratégie en matière de données, et lui attribuent les meilleures notes (entre 9 et 10, sur une échelle de 1 à 10). Ces « très bons élèves » en matière de

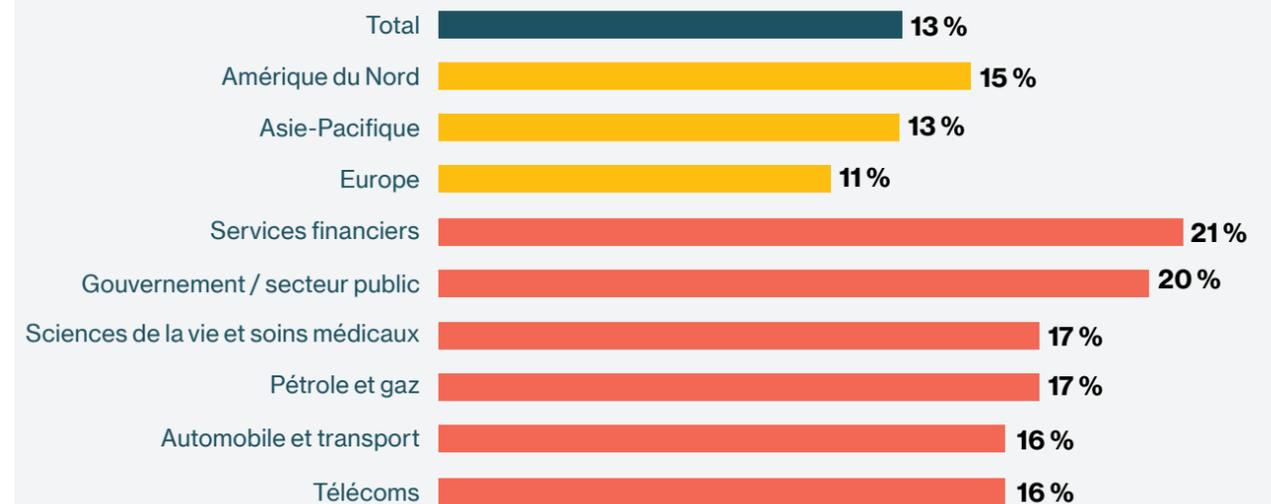
données obtiennent des résultats commerciaux mesurables au sein de plusieurs départements de l'entreprise, déclarent leurs dirigeants. À l'opposé, on trouve un groupe faisant à peu près la même taille : les « mauvais élèves » (12 % de l'échantillon). Leur performance en matière de données atteint tout au plus la note de 6 sur cette même échelle.

Ces deux groupes ont des caractéristiques et intentions

Graphique 3a : Dans quelle mesure les organisations parviennent à réaliser la stratégie de données de l'entreprise (autoévaluation sur une échelle de 1 à 10, 10=réussite totale)

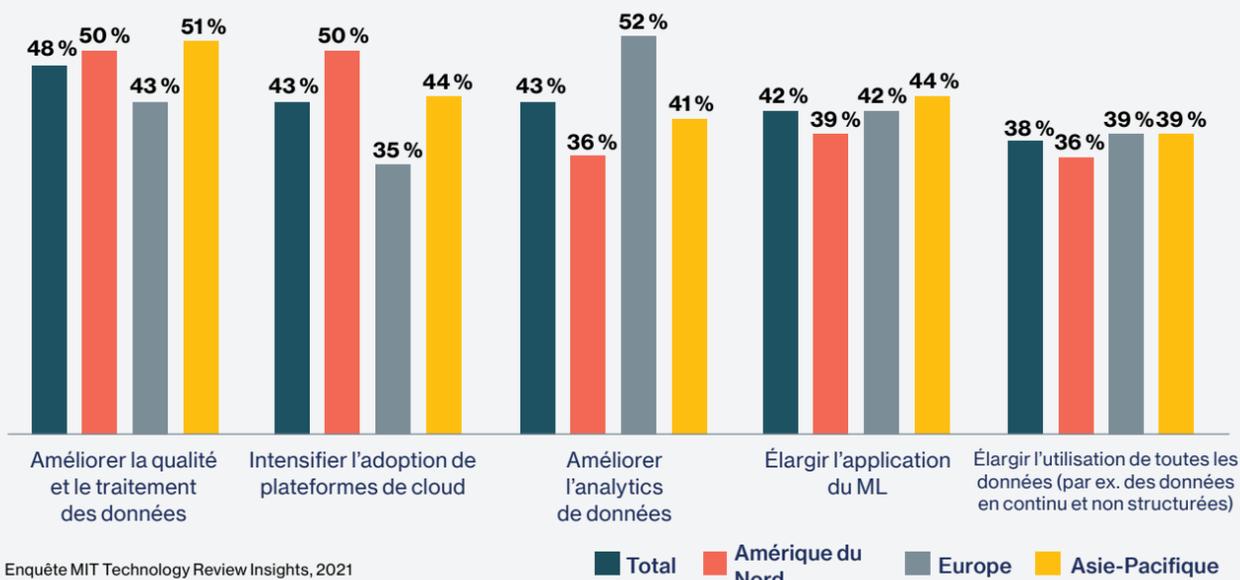


Graphique 3b : Très bon élèves : répondants qui attribuent à leur organisation une note de 9 ou 10 concernant la réalisation de la stratégie de données de l'entreprise, avec des résultats commerciaux quantifiables sur plusieurs unités opérationnelles (total, régions et secteurs sélectionnés)



Enquête MIT Technology Review Insights, 2021

Graphique 2 : Principales initiatives dans le cadre de la stratégie de données à l'échelle de l'entreprise pour les deux prochaines années (réponses les plus fréquentes ; pourcentage de sondés)



Enquête MIT Technology Review Insights, 2021

Graphique 4 : Les principaux facteurs de réussite permettant aux organisations faisant partie des « très bons élèves » de mettre œuvre leurs initiatives dans le cadre de leur stratégie de données (réponses les plus fréquentes ; pourcentage de répondants)



Enquête MIT Technology Review Insights, 2021

bien différentes. Par exemple, le cloud est nettement plus représenté dans l'architecture de données des très bons élèves : pour 74 % d'entre eux, au moins la moitié de leurs services ou infrastructures de données fonctionne dans un environnement cloud, contre 60 % pour les mauvais élèves. Quant aux priorités en matière de données, la plupart des mauvais élèves (59 %) souhaitent améliorer la gestion des données (qualité et traitement) au cours des deux prochaines années, tandis que l'initiative la plus citée par les très bons élèves (53 %) est l'application généralisée du ML.

Loin de tenir les bases pour acquises, les très bons élèves pensent que leur réussite tient à l'attention portée aux fondements d'une bonne gestion des données. Ces fondements incluent la réduction de la duplication des données, un accès simplifié aux données pour les utilisateurs finaux de l'entreprise – véritable marqueur de la « démocratisation » des données – et le traitement de gros volumes de données à des vitesses élevées. La

duplication des données dans les grandes organisations a lieu à plusieurs niveaux, notamment dans les entrepôts de données, les systèmes opérationnels, les rapports, les tableaux de bord et les outils bureautiques. Or, cela a un coût et des conséquences sur la gestion des risques et sur la fiabilité, précise Ashwin Sinha, Directeur des données et de l'analytics chez Macquarie Bank. Par ailleurs, la duplication des données a un impact sur la capacité à élargir et utiliser efficacement l'apprentissage automatique dans l'ensemble de l'organisation.

Lorsqu'on leur demande quels sont les freins à leur progression, la plupart des mauvais élèves mentionnent l'évolutivité limitée de leur plateforme de gestion des données. Parmi les autres obstacles fréquemment cités, on retrouve la lenteur de traitement des gros volumes de données et les difficultés à promouvoir la collaboration. Comme nous allons le voir, les difficultés à obtenir l'évolutivité, la vitesse et la collaboration touchent les organisations pour l'ensemble des opérations de données.

Graphique 5 : Les principales difficultés qui empêchent les organisations faisant partie des « mauvais élèves » de réaliser leurs initiatives dans le cadre de leur stratégie de données (réponses les plus fréquentes ; pourcentage de répondants)



Enquête MIT Technology Review Insights, 2021

Nielsen : transformation des données pour une entreprise qui dépend fortement des données



Nielsen connaît l'importance d'une bonne gestion des données. Elle est une des rares organisations existant depuis plus d'un siècle qui, depuis sa création, a placé les données au cœur de son modèle économique. Les panels de Nielsen indiquent aux sociétés qui produisent des biens de consommation quels sont les produits que les clients achètent et comment les comportements évoluent. Les panels conseillent également ces sociétés sur le lieu et l'heure à privilégier pour diffuser leurs spots publicitaires. Pour sa deuxième année en tant que CDO, et sa cinquième en tant que directeur de la recherche, Mainak Mazumdar a dirigé une véritable transformation de la gestion et de l'infrastructure des données de Nielsen. « Il y a quelques années seulement », se souvient-il, « nous avions des problèmes de fragmentation – beaucoup de données en silos et un accès réservé à des experts dans ce domaine –, d'absence de métadonnées et de manque de gouvernance, le tout alors que les volumes de données explosaient. »

Aujourd'hui, Mainak Mazumdar dresse un tout autre tableau : « Aujourd'hui, nous sommes capables de passer rapidement de 20 à 30 spécialistes sur une plateforme à plus de 300. Nous sommes sur une plateforme de cloud avec un lac de données au sein duquel les données sont organisées, étiquetées, définies, consolidées et où elles possèdent des métastores. Nous avons

mis au point notre propre moteur d'analytics. En fait, mon équipe, affectée directement à la production, fait la même chose que les ingénieurs logiciel par le passé. » D'après lui, ces changements ont réduit de moitié la durée de cycle de son équipe. « Aujourd'hui, la vitesse de nos modèles est environ 50 fois plus élevée. Nous mettons une minute, voire moins, à réaliser une tâche qui auparavant nous en prenait 20. Et en même temps, nous intégrons et traitons d'immenses volumes de données facilement accessibles pour la data science. C'est un changement considérable. »

Ainsi, Nielsen a tiré profit de ce potentiel pour favoriser la croissance et a lancé un nouveau produit de notations sur les 200 marchés locaux américains de la société. Grâce au traitement de gros volumes de données issues de boîtiers de décodeurs TV, le « modèle d'apprentissage profond à degré élevé de reconnaissance » permet à Nielsen de dire aux clients non seulement ce qu'un téléspectateur est susceptible de regarder à un moment donné, mais aussi quel membre du foyer regarde, chose jusqu'ici impossible à dire. « Sans les changements que nous avons opérés dans la gestion des données, jamais nous n'aurions pu lancer ce produit et intégrer de tels volumes », précise Mainak Mazumdar. « Et c'est grâce à cette intégration de volumes de données encore plus importants que les modèles – et le produit – ne cessent de s'améliorer. »

Mettre à l'échelle l'analytics et l'apprentissage automatique



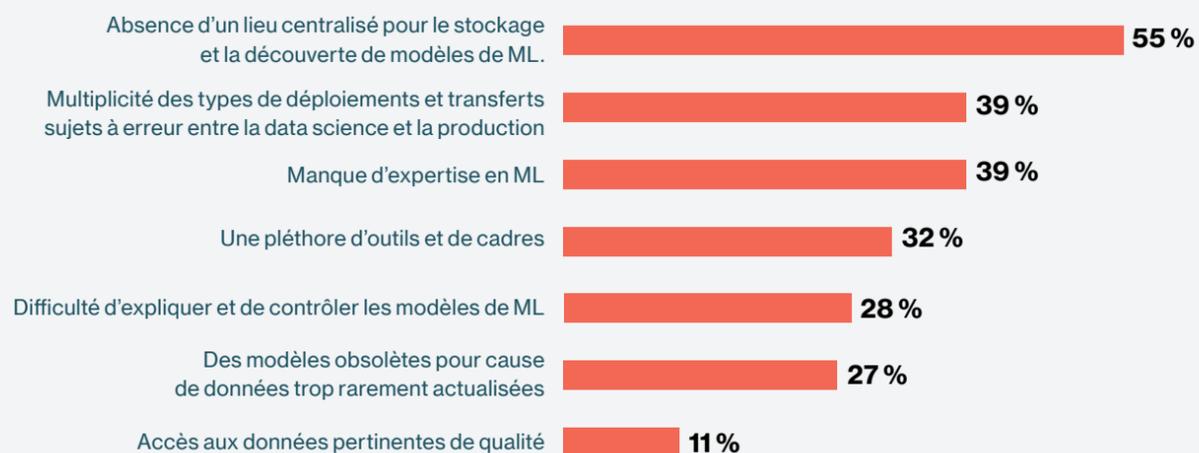
Les chefs d'entreprise savent que la capacité de leur entreprise à rester en phase avec la demande et à l'anticiper, à gérer la pression de la concurrence, à innover efficacement et à fonctionner de manière efficiente repose sur leur maîtrise de l'analytics et du ML. Les organisations de la quasi-totalité des secteurs travaillent à la mise au point de scénarios d'utilisation d'analytics et de ML qui leur permettront d'obtenir de meilleurs résultats commerciaux. Pour la plupart des grandes entreprises, il est devenu indispensable de posséder une large gamme de cas d'usage en production et à l'échelle. Les CDO et leurs équipes sont de plus en plus jugés sur leur capacité à fournir de tels scénarios.

C'est un problème pour de nombreuses organisations. Elles ont notamment du mal à atteindre la mise à l'échelle

requis pour générer un impact notable. D'après Sol Rashidi, d'Estée Lauder, cela s'explique en partie par une trop grande ambition : « Trop souvent, les entreprises grillent les étapes en matière de ML et veulent tout de suite aller à 100 à l'heure sans avoir pris le temps de maîtriser les bases. » D'autres CDO, à l'instar de Don Vu, chez Northwestern Mutual, estiment qu'il faut avant tout sélectionner les bons cas d'usage à déployer en production. Sans l'apport d'utilisateurs professionnels, ajoute-t-il, on augmente le risque de sélectionner des scénarios ne correspondant pas clairement à un objectif commercial.

Un dialogue solide entre les concepteurs de modèles de ML et les utilisateurs professionnels permet d'empêcher l'effet « algorithme lustré », où l'équipe de data science privilégie les scénarios d'utilisation de ML les plus attrayants alors même que les utilisateurs ne leur

Graphique 6 : Les principales difficultés que rencontrent les entreprises pour la mise à l'échelle des cas d'usage (réponses les plus fréquentes ; pourcentage de sondés)



Enquête MIT Technology Review Insights, 2021

Changement de paradigme chez CVS Health



Les pharmacies ont toujours joué un rôle essentiel dans nos sociétés, mais ce rôle n'a sans doute jamais été aussi important qu'au cours de l'année passée. Des chaînes de pharmacie comme CVS, la plus grande des États-Unis en termes de chiffre d'affaires, utilisent des fonctions de données plus perfectionnées que jamais afin que leurs clients soient bien à jour dans leurs traitements et dans leur utilisation d'autres services de santé, de peur que des affections sous-jacentes ne viennent aggraver leur état de santé. En tant que CDO chez CVS Health et directeur de l'analytics de son activité de pharmacie de détail, Bob Darin dirige bon nombre de démarches allant dans ce sens.

Cela fait longtemps que l'entreprise utilise des systèmes de données afin d'encourager ses clients à rester à jour dans leurs traitements. Une pratique qui se traduit, par exemple, par la prise de contact avec le patient via des appels téléphoniques ou des SMS, par des instructions au comptoir de la pharmacie, ou encore par des recommandations visant à orienter le patient vers son praticien au sujet de suivis spécifiques ou d'exams des traitements. Selon Bob Darin, ces initiatives, pour être plus personnalisées, sont devenues indissociables de la data science au cours de ces dernières années. « Aujourd'hui, nous savons quels rappels, programmes, modes de communication et messages sont les plus efficaces pour aider les gens à se souvenir de prendre leurs traitements,

de planifier leurs collectes de médicaments et de comprendre les possibilités d'économies. » Ils sont tous alimentés par des modèles d'analytics, explique-t-il, alors qu'auparavant, c'était un processus uniforme.

« Le tout résulte d'un véritable changement de paradigme pour nous », affirme Bob Darin. « Non seulement dans notre manière d'utiliser l'analytics, mais aussi dans celle dont nous gérons nos architecture et plateforme de données. » Il ajoute que la société a dû investir dans l'entreposage de données, le reporting opérationnel et l'analytics. Toutefois, ses équipes sont actuellement dans une démarche consistant à « faire passer un plus grand nombre de fonctions commerciales vers un environnement de données hybride et multidimensionnel, qui en plus de prendre en charge l'entreposage de données et des scénarios d'utilisation d'analytics descriptive, est capable d'utiliser un apprentissage automatique plus perfectionné, de développer un algorithme et d'optimiser des techniques dans les bonnes proportions. »

Ce type d'environnement, précise Bob Darin « doit prendre en charge des applications de data science plus perfectionnées ainsi qu'une analytique commerciale quotidienne, comme les descriptifs et requêtes ad hoc – le type d'informations que les gens utilisent pour prendre chaque jour leur décisions commerciales. Il nous faut des architectures et des plateformes de données qui prennent en charge ces deux éléments. »

Les trois quarts des personnes travaillant dans le secteur de la production électrique déclarent que l'absence d'un endroit centralisé pour stocker et découvrir des modèles de ML est un obstacle majeur à la mise à l'échelle. Même chose pour plus de deux tiers des répondants travaillant dans les secteurs gouvernementaux et des biens de consommation.

attribuent qu'une faible valeur. « Si votre cas d'usage d'apprentissage automatique ne correspond pas très étroitement à un scénario d'utilisation business, alors je me demande ce que vous faites avec », se demande Andy McQuarrie d'Hivery.

Obstacles à la mise à l'échelle

Les entreprises doivent faire face à la complexité de la gestion du cycle de vie du ML de bout en bout, comme le montre la liste des difficultés rencontrées par les organisations interrogées concernant la mise à l'échelle des cas d'usage (Graphique 6). La difficulté la plus courante, citée par 55 % des répondants, est l'absence d'un lieu centralisé pour le stockage et la découverte de modèles de ML. Les trois quarts des personnes travaillant dans la production d'électricité et autres infrastructures

publiques liées à l'énergie déclarent qu'il s'agit d'un obstacle majeur à la mise à l'échelle. Même chose pour plus de deux tiers des sondés travaillant dans les agences gouvernementales et dans les sociétés produisant des biens de consommation. Pour 39 % des répondants, une collaboration inadaptée entre la data science et la production, laquelle se traduit par de multiples déploiements et des transferts sujets à erreur, est un autre inconvénient majeur vis-à-vis de la mise à l'échelle. Un chiffre qui passe à 50 % parmi les personnes interrogées qui travaillent dans les sciences de la vie et les organisations de soins médicaux. Et de nombreuses organisations sont confrontées à de multiples outils et cadres, mais aussi à une actualisation trop rare des données de modèles (points cités respectivement par 32 % et 27 %).



« Les multiples déploiements et transferts sujets à l'erreur entre la science des données et la production représentent un immense problème. Il y a souvent une différence importante entre le résultat de la science des données et le résultat obtenu après l'avoir mise en pratique. »

Naveen Jayaraman, Vice-président – données, CRM et analytics, L'Oréal

Les CDO que nous avons interrogés confirment le problème des multiples déploiements et des transferts sujets à erreur entre la data science et la production. « C'est un très gros défi », déclare Naveen Jayaraman, Vice-président – données, CRM et analytics chez L'Oréal. « Il y a souvent une différence importante entre le résultat théorique de la data science et le résultat obtenu après l'avoir mise en pratique. »

prises avec ce manque ; 59 % ont déclaré qu'il s'agit d'un défi majeur pour eux.

Protéger le retour sur investissement

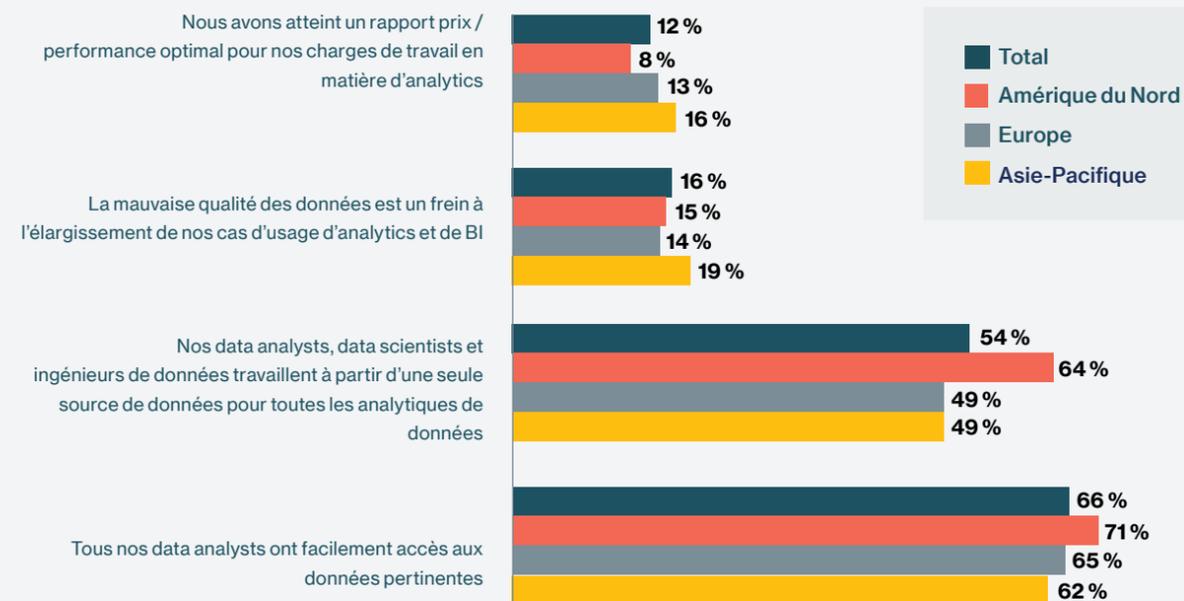
Les difficultés à mettre en place des cas d'usage d'analytics et de ML à l'échelle souhaitée ont des conséquences sur le retour sur investissement (ROI). Par exemple, seulement 12 % des personnes interrogées ont déclaré avoir atteint un rapport prix / performance optimal pour leurs charges de travail d'analytics. Les problèmes de duplication des données cités plus haut sont susceptibles de réduire la performance, tout comme le coût d'affectation des ressources limitées pour se concentrer sur le nettoyage des données de mauvaise qualité. Et, comme le précise Naveen Jayaraman, un élargissement à l'échelle de l'entreprise implique souvent un investissement dans l'infrastructure du cloud et dans la gestion des données. « Cela n'est pas évident », ajoute-t-il. « À moins qu'un cas d'usage ne présente des gains importants en termes de ventes ou d'efficacité, vous aurez toujours un gros problème de ROI. »

Selon Bob Darin, CVS Health est parvenue à réduire les problèmes de transfert : « Auparavant, des data scientists mettaient au point des modèles et en tiraient des conclusions. Ils transmettaient alors ces dernières à une autre équipe pour la mise en œuvre. Cette démarche en cascade ne fonctionnait pas et les modèles avaient du mal à gagner en évolutivité. » Il ajoute que CVS Health a réussi à réduire ces difficultés en intégrant la data science et la production en une même équipe qui travaille en partenariat avec le service informatique. « C'est tellement mieux que d'avoir une équipe pour développer les conclusions et une autre pour la mise à l'échelle. »

Pour les organisations qui cherchent à mettre en place des pratiques de ML efficaces, le manque de compétences reste un problème systémique. 39 % des sondés estiment que le manque d'expertise en ML est un obstacle majeur à la mise à l'échelle des cas d'usage, et les industries manufacturières interrogées (64 %) le voient comme un problème spécifique. Cependant, les très bons élèves semblent avoir surmonté ce manque, puisque seulement 27 % le considèrent comme un défi du ML. À l'inverse, les mauvais élèves sont encore largement aux

D'un autre côté, comme le souligne Patrick Baginski, Senior director data science chez McDonald's, le géant du fast-food, on obtient bien plus qu'un simple avantage de ROI si on atteint ou dépasse les objectifs de performance. « C'est le moment où jamais de valoriser le ML. Plus vite serez capable de montrer la valeur du ML et de la data science, plus vite vous obtiendrez l'adhésion des utilisateurs et gagnerez la confiance de la direction dans ce qu'ils peuvent tous deux apporter à l'organisation. » Par ailleurs, ML et data science contribuent largement au développement d'une culture des données.

Figure 7 : Répondants entièrement ou plutôt d'accord avec les déclarations sur l'analytics et l'informatique décisionnelle (BI) au sein de leur organisation (pourcentage de sondés)



Enquête MIT Technology Review Insights, 2021

Technologie, démocratie et culture

Malgré l'ampleur de la transformation qu'a subi la gestion des données au niveau des systèmes, du leadership et des perceptions de sa valeur commerciale, la plupart des organisations souffrent d'une inégalité importante entre les équipes de données et les utilisateurs finaux, à savoir les collaborateurs du front ou du back-office qui ont besoin des conclusions fournies par les données afin de prendre des décisions au quotidien. De nombreux CDO tentent de combler cette inégalité en intégrant directement des data scientists aux unités opérationnelles afin qu'ils puissent interagir régulièrement avec les utilisateurs. Autre méthode : mettre l'analytics à la disposition des utilisateurs afin qu'ils puissent tirer eux-mêmes les conclusions dont ils ont besoin. Selon Patrick Baginski de McDonald's, il faut également pousser les données « jusqu'au bout », là où se trouve l'utilisateur.

Un objectif que partagent tous les CDO et les directeurs de l'analytics que nous avons interrogés. « Tandis que les outils progressent et que les collaborateurs maîtrisent de plus en plus l'analytique perfectionnée et la data science, nous devons permettre aux utilisateurs de procéder eux-mêmes à des analytics et non les laisser consommer les analytiques produites par quelqu'un d'autre », déclare Bob Darin

C'est la mise en pratique de la démocratisation des données. Par ailleurs, certains CDO estiment que c'est non seulement un moyen de développer la maîtrise des données, mais aussi d'encourager la croissance de la culture des données au sein d'une organisation. L'infrastructure et les outils de données y participent directement. « Nous sommes une équipe de données centrale avec pour objectif la création d'outils qui fédèrent une grande partie de la responsabilité et du pouvoir pour les autres », déclare Don Vu. « Nous essayons donc de pousser

les capacités de notre plateforme de données jusqu'au bout. Nous souhaitons instaurer de la gravité parmi les équipes d'application et utilisateurs du point de vue des fonctionnalités. Nous voulons que les gens aient envie d'être sur cette plateforme parce qu'elle prend en compte la qualité des données, la confidentialité et l'accès. »

D'après Ashwin Sinha de Macquarie Bank, en simplifiant l'accès des utilisateurs finaux aux données et à l'analytics, on contribue directement à l'amélioration de la culture des données de l'organisation. « Cela ne se joue pas uniquement au niveau des analystes », précise-t-il. « Les principaux dirigeants de l'organisation avec un esprit analytique utilisent des outils de visualisation des données avec des interfaces simples afin d'analyser les données et d'obtenir des insights. » Ashwin Sinha ajoute que les outils de visualisation ont permis de faciliter l'accès, au même titre que l'utilisation de l'infrastructure cloud par l'équipe a permis d'améliorer la fiabilité et la performance. « Au cours de ces trois dernières années, nos charges de travail – réglementaires, d'apprentissage automatique et analytiques – ont été déplacées vers une plateforme de données cloud. La plateforme de données emploie divers services de cloud, paquets open source et outils de gestion des données afin de garantir la standardisation, l'intégration et l'approvisionnement des données dans l'ensemble de l'organisation de manière adaptée à l'objectif, avec un lignage et une traçabilité complètes des données. »

Toujours d'après Ashwin Sinha, l'alliance d'une solide gouvernance des données, de la facilité d'accès et de la simplification permet d'améliorer la confiance des utilisateurs envers les données et l'analytics. Il est impossible de développer une culture des données sans confiance envers une plateforme de données fiable.

« Définissez les données les plus importantes, donnez-leur la priorité, établissez la discipline nécessaire à leur protection et à leur gouvernance, puis démocratisez les données afin de permettre à vos spécialistes et utilisateurs finaux d'extraire les insights dont ils ont besoin pour innover. »

Sol Rashidi, Directrice de l'analytics, The Estée Lauder Companies

05 Visions de l'avenir

La moitié des dirigeants interrogés (et environ deux tiers de ceux qui travaillent dans des entreprises technologiques ou manufacturières) déclarent être actuellement en train d'évaluer ou de mettre en place une nouvelle plateforme de données afin de répondre à leurs difficultés en matière de données. 9 % aimeraient faire de même, mais sont confrontés à des obstacles au changement de leur architecture. Là encore, on observe une nette différence entre les très bons et les mauvais élèves. La plupart des très bon élèves (56 %) sont satisfaits de leur architecture actuelle, même si près d'un tiers d'entre eux sont à la recherche de nouvelles plateformes ou testent des fournisseurs. À l'inverse, seulement 24 % des mauvais élèves sont satisfaits de leur architecture existante, et 59 % cherchent activement à la changer. Les entreprises technologiques et manufacturières sont bien plus susceptibles que celles des autres secteurs d'entreprendre une modernisation de leur plateforme

Ce que veulent les CDO comme nouvelle architecture

Et si les leaders des données et de la technologie pouvaient mettre au point une nouvelle architecture de données pour leur activité ? Nous leur avons demandé quels en seraient les principaux avantages par rapport à leur architecture existante. Leur souhait numéro un, cité par 50 % des répondants (et par 60 % ou plus parmi les sondés travaillant dans le secteur des technologies, des

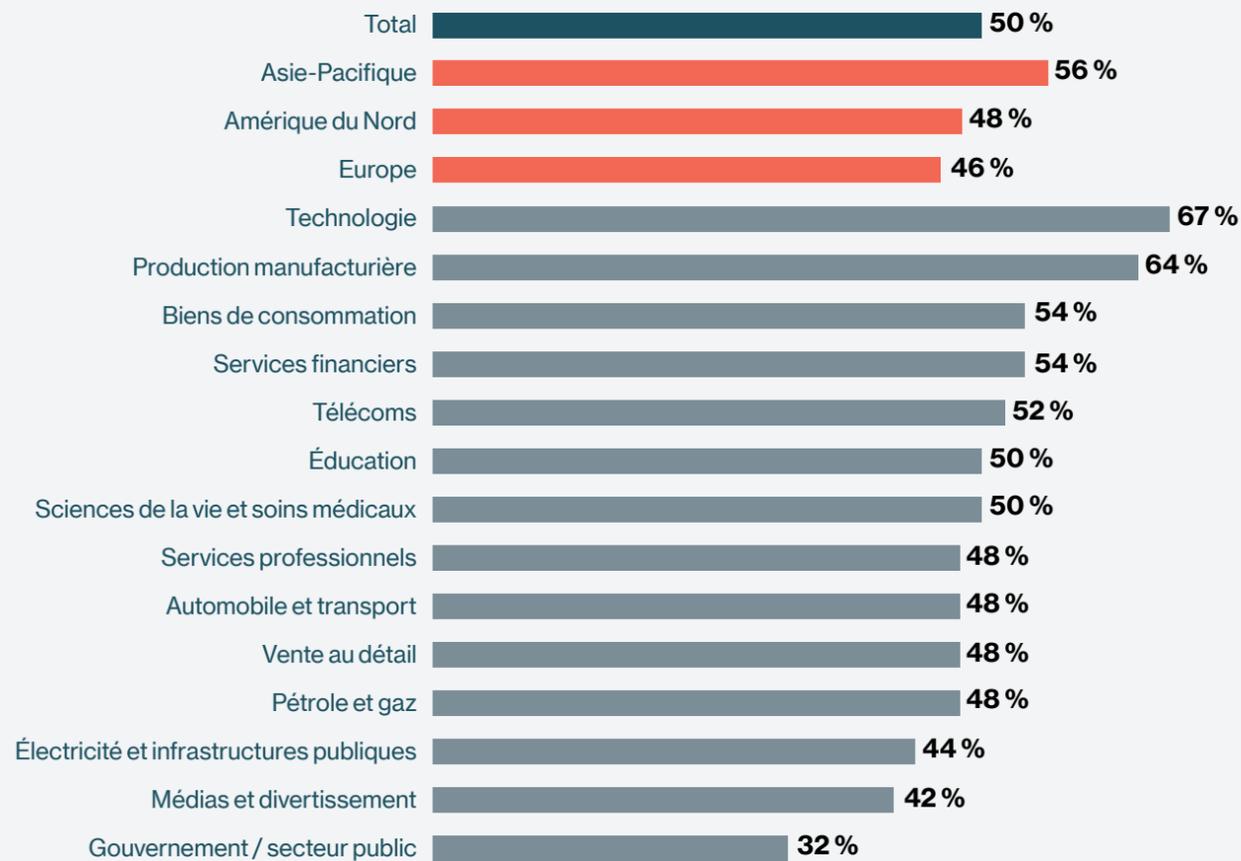


infrastructures publiques et les administrations centrales) sont des normes open source et des formats de données en accès libre. Pour Andy McQuarrie d'Hivery, ce n'est pas étonnant : « Les normes ouvertes facilitent la progression sur la courbe de maturité. Vous pouvez consommer des services grâce à un service administré à court terme. Lorsque votre activité grandit et que vous devez ajouter un composant, l'open source vous permet de le faire sans modifier l'ensemble de l'architecture. Plus besoin de changer les technologies comme vous deviez le faire auparavant. »

Ensuite, les CDO souhaitent non pas de nouvelles fonctions, mais des fonctions renforcées dans des domaines où les leaders de la technologie sont en recherche permanente d'amélioration : une sécurité renforcée, une gouvernance plus solide et un meilleur rapport prix / performance pour l'infrastructure, les opérations, la maintenance et d'autres éléments de l'architecture. Les sondés mettent aussi en avant le fait qu'une nouvelle architecture puisse prendre en charge tous les cas d'usage d'analytics, qu'ils soient basés sur le ML, la data science ou la business intelligence.

Ce que veulent en premier les leaders de la technologie, ce sont des normes open source et des formats de données en accès libres, puis une sécurité et une gouvernance renforcées, ainsi qu'un meilleur rapport prix / performance des éléments de l'architecture.

Graphique 8 : Répondants actuellement en phase d'évaluation ou de mise en œuvre d'une nouvelle plateforme de données permettant à leurs solutions ou stratégie de données à l'échelle de l'entreprise de répondre à leurs problèmes actuels
(pourcentage de sondés)



Enquête MIT Technology Review Insights, 2021

« L'un de nos principaux objectifs est l'intégration native aux systèmes clients. Comment faciliter le transfert des données de nos clients vers nos systèmes ? »

– Andy McQuarrie, Chief Technology Officer, Hivery

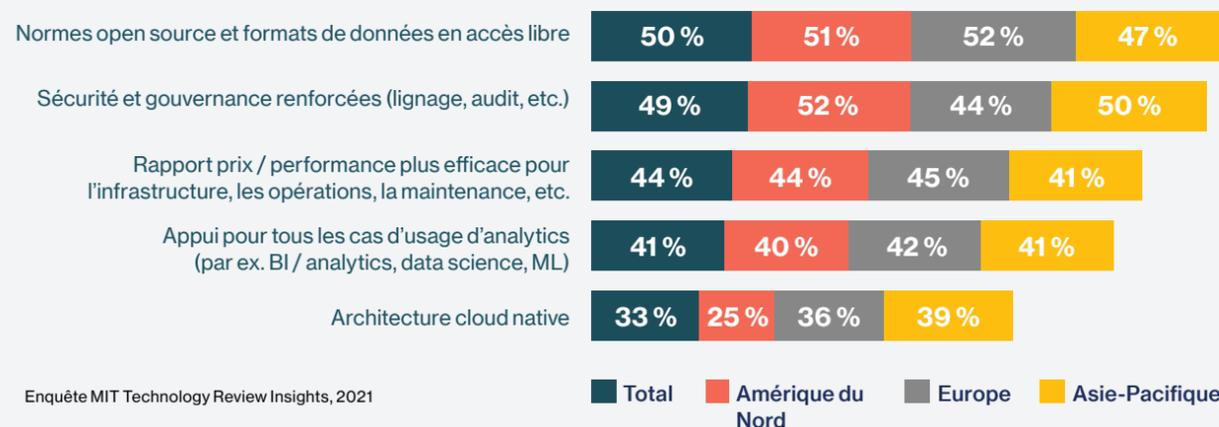
Les CDO que nous avons interrogés ont d'autres attentes et projets liés à l'évolution de leur infrastructure et gestion des données. Michel Lutz, Directeur des données groupe chez TotalEnergies, a pour objectif d'étendre l'automatisation basée sur le ML de sa gestion des métadonnées. Cela a déjà été fait pour les données géoscientifiques, mais il prévoit de l'étendre à tous les domaines de données du groupe au cours de l'année prochaine. Par ailleurs, Michel Lutz surveille de près le potentiel de l'architecture maillée de données. « Ce sera probablement la prochaine étape pour notre architecture de données, car elle permet une plus grande mise à l'échelle organisationnelle et plus de spécialisation des domaines », explique-t-il.

Naveen Jayaraman de L'Oréal partage ce point de vue quant aux avantages des produits de données axés sur les domaines pour des entreprises comme la sienne qui cherchent à distribuer des fonctions de gestion des données à grande échelle. Comme le précise Naveen Jayaraman, les approches architecturales qui emploient un maillage de données ont un véritable potentiel pour ses objectifs, à partir du moment où certains éléments, comme un stockage polyglote, peuvent être centralisés.

Mainak Mazumdar de Nielsen va bientôt mettre en œuvre « IA pour IA » – principalement la sélection automatisée de modèles de ML. « Nous souhaitons intégrer une autre couche d'intelligence qui prendra les meilleures décisions concernant les modèles à utiliser. »

À plus long terme, Ashwin Sinha réfléchit aux possibilités offertes par une « architecture souple et en temps réel », suffisamment souple pour s'adapter aux futures innovations de l'ingénierie des données. Cela permettrait, explique-t-il, de mettre à l'épreuve du futur une architecture de conception récente, ne serait-ce qu'à moyen terme. « Nous aimerions pouvoir adopter une approche architecturale suffisamment flexible pour ne pas être obsolète deux ou trois ans plus tard. »

Graphique 9 : Les principaux avantages de la nouvelle architecture idéale des répondants par rapport à leur architecture existante (réponses les plus fréquentes ; pourcentage de sondés)



Enquête MIT Technology Review Insights, 2021

■ Total ■ Amérique du Nord ■ Europe ■ Asie-Pacifique

06 Conclusion

L'un des fils conducteurs de notre étude, en particulier de nos discussions avec les CDO, a été l'identification d'un lien direct qui part de l'infrastructure des données, se poursuit vers la data science et va jusqu'aux attitudes des collaborateurs vis-à-vis des données, à tous les niveaux de l'organisation, pour enfin se fixer aux résultats commerciaux. C'est la culture des données. La recherche apporte des enseignements aux CDO qui souhaitent utiliser leurs actifs pour encourager la culture des données au sein de leur entreprise.

- **Simplicité et flexibilité.** De nombreux CDO veulent simplifier les architectures trop complexes, mais ils doivent également veiller à ce que les interfaces pour les utilisateurs finaux d'analytics soient faciles à utiliser et, dans la mesure du possible, ludiques. Par ailleurs, l'architecture doit rester suffisamment flexible pour répondre aux futurs besoins commerciaux sans avoir recours à de grandes migrations ou refontes pour l'ajout de nouvelles technologies. Plus il y aura d'utilisateurs capables d'interagir avec des données mises à jour en permanence à l'aide de leurs outils favoris et de tirer à la demande des conclusions à partir de ces données, plus le taux d'engagement augmentera.

- **Une bonne gouvernance.** Il n'y a rien de plus dommageable pour l'expérience d'analytics de la part des collaborateurs que des données erronées. Les CDO connaissent l'importance d'une gouvernance solide des données, mais certains ignorent la portée des répercussions au sein de l'organisation en cas d'échecs de gouvernance. La mise en place et le maintien d'une seule source de vérité est essentielle pour tout le monde.

- **Expliquer et promouvoir.** Certains CDO chargent une partie de leurs équipes d'éduquer les unités opérationnelles et d'autres collaborateurs à la data science et de les former à l'utilisation de l'analytics et d'autres outils. Par ailleurs, de nombreux CDO pensent que la promotion à grande échelle de la valeur et des résultats positifs obtenus grâce à l'utilisation des données et de l'analytics permet de motiver leurs collaborateurs à utiliser ces dernières. La promotion de l'exactitude des données grâce à une gouvernance solide permettra aussi de gagner leur confiance.

- **Créer des liens avec d'autres cadres supérieurs.** Les CDO et Directeurs de l'analytics savent à quel point ils se doivent de renforcer leurs liens avec d'autres cadres supérieurs, mais toute déconnexion des DSI et Directeurs technologiques vis-à-vis de la gestion des données peut devenir source de méfiance si elle n'est pas réglée rapidement. La stratégie en matière de données doit prendre en compte l'infrastructure technologique existante et veiller à toujours rester axée sur la réalisation des priorités de l'entreprise.

À propos de MIT Technology Review Insights

MIT Technology Review Insights est le service de publication sur mesure de *MIT Technology Review*, le magazine sur les technologies le plus ancien du monde, soutenu par le principal organisme de technologie au monde – lequel crée des événements et études sur les principaux défis commerciaux et technologiques de notre époque. Insights mène des recherches et analyses qualitatives et quantitatives aux États-Unis et en dehors. Il publie aussi des contenus très divers, comme des articles, rapports, infographies, vidéos et podcasts. Et grâce à son MIT Technology Review Global Panel en pleine expansion, Insights bénéficie d'un réseau hors pair de dirigeants, d'innovateurs et de leaders visionnaires du monde entier auprès desquels il peut mener des enquêtes et réaliser des interviews.

De la part du sponsor

Databricks est l'entreprise alliant IA et données. Basée sur une architecture lakehouse moderne, Databricks combine le meilleur des entrepôts et des lacs de données pour vous offrir une plateforme simple, ouverte et collaborative pour toutes les charges de travail liées aux données. Plus de 5 000 organisations dans le monde entier, dont Shell, Comcast, CVS Health, HSBC, T-Mobile et Regeneron, font confiance à Databricks pour mener à bien l'ingénierie des données à grande échelle, la data science exploratoire, l'apprentissage automatique à cycle de vie complet et l'analytics commercial. Forte de sa présence dans le monde entier et de ses centaines de partenaires, dont Microsoft, Amazon, Google et Tableau, Databricks a pour mission d'aider les équipes en charge des données à répondre aux problèmes les plus complexes.



Illustrations

Davooda : 5, 9, 13, 14 ; fad82 : 13 ; Friday Studio : 2, 3, 6, 15, 17, 22 ; Nadienko : 18 ; ninamalin : 18, papipo : 15 ; Prostock Studio : Couvertures et quatrièmes de couverture, 2, 3, 4, 5, 13, 19 ; Rashes Ashor : 19 ; Starline : Couvertures et quatrièmes de couverture, 2. Toutes les illustrations sont fournies par Shutterstock et rassemblées par Scott Shultz Design.

Même si nous mettons tout en œuvre pour vérifier leur exactitude et leur fiabilité, MIT Technology Review Insights ne saurait être tenu responsable des informations, opinions ou conclusions exposées dans le présent rapport.

© Copyright MIT Technology Review Insights, 2021. Tous droits réservés.



MIT Technology Review Insights

 www.technologyreview.com

 @techreview @mit_insights

 insights@technologyreview.com