



## Papier blanc

### Mise à l'échelle des initiatives d'IA/ML : le rôle essentiel des données

Sponsorisé par : Snowflake Inc.

Ritu Jyoti  
Février 2022

## INTRODUCTION

---

L'intelligence artificielle (IA) est l'innovation technologique la plus perturbatrice de notre époque. À l'origine une discipline réservée aux cercles académiques, l'IA est désormais commercialement courante. Les entreprises adoptent l'IA/l'apprentissage automatique (ML) et exploitent une variété de types de données (structurées, non structurées et semi-structurées) dans tous les secteurs d'activité (LOB) et industries. Par exemple:

- Croissance des revenus :
    - Les équipes marketing et commerciales utilisent l'IA pour mieux cibler les clients potentiels, optimiser des campagnes de sensibilisation et hiérarchiser les prospects.
    - Les technologies d'intelligence artificielle permettent l'exploration des sentiments des médias sociaux, la sélection programmatique de propriétés publicitaires, mesurer l'efficacité des programmes de marketing, assurer la fidélité des clients et recommander des ventes intelligentes.
  - Efficacité économique/opérationnelle :
    - Les solutions de centre de contact basées sur l'IA accélèrent le délai de résolution et améliorent la expérience client (CX). Le traitement du langage naturel permet aux clients de parler facilement de ce dont ils ont besoin sans naviguer dans une arborescence téléphonique frustrante.
    - Les algorithmes d'apprentissage en profondeur (DL) accélèrent le diagnostic et le traitement des maladies graves et soutiennent la médecine de précision.
    - Les modèles IoT industriels peuvent désormais prédire quand une machine tombera en panne et recommander maintenance préventive, évitant ainsi tout temps d'arrêt potentiel.
  - Atténuation des risques :
    - Les institutions financières améliorent la souscription des prêts et réduisent les risques. L'IA peut également contribuer à réduire la criminalité financière grâce à une détection avancée des fraudes et à la détection d'activités anormales.
- L'IA joue aujourd'hui un rôle essentiel dans la cybersécurité. En améliorant la capacité des organisations à anticiper et déjouer les violations, protéger le nombre proliférant de surfaces de menaces avec des cadres de sécurité zéro confiance et rendre les mots de passe obsolètes, l'IA est essentielle pour sécuriser les périmètres de toute entreprise.

Les initiatives d'IA offrent plus que de simples économies de coûts. Selon l'enquête [IA Stratégies Voir l'enquête 2021](#) — un mondiale d'IDC auprès de 2 000 organisations, avec des décideurs et des influenceurs IT et LOB comme répondants - les perturbateurs de l'IA (organisations qui créent à plusieurs reprises une nouvelle valeur commerciale et un avantage concurrentiel durable grâce à l'IA) signalent une amélioration de 39% de l'expérience client et une amélioration de 33% l'efficacité des collaborateurs et l'accélération de l'innovation avec le déploiement des solutions d'IA en 2020.

Il s'agit d'une augmentation à deux chiffres de l'amélioration des résultats commerciaux par rapport à 2019. Il existe une corrélation directe entre la maturité de l'adoption de l'IA et des résultats commerciaux supérieurs. En tant que tel, les entreprises ont un besoin accru d'étendre stratégiquement leurs initiatives d'IA/ML.

Les données sont essentielles pour les initiatives d'IA/ML :

- L'IA/ML nécessite de vastes volumes de données pour former des modèles.
- Pour garantir des résultats impartiaux, il faut divers ensembles de données. ▪

Les modèles doivent être formés en permanence avec les informations les plus récentes pour maintenir les performances prédictives, en particulier dans les environnements commerciaux dynamiques.

En tirant parti de l'architecture de données consolidée, les perturbateurs de l'IA qui ont réussi ont pu exploiter la puissance de différents types de données et de l'écosystème associé pour stimuler l'innovation et la transformation.

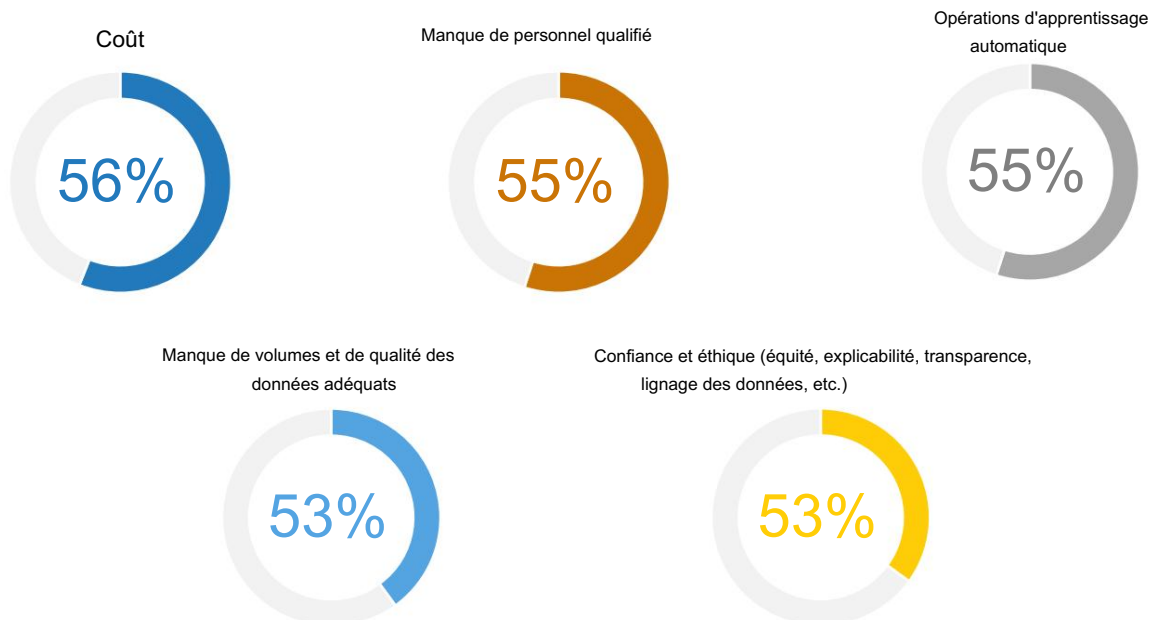
## SCALING AI : INHIBITEURS/DÉFIS/BESOINS

Bien que de nombreuses organisations comprennent l'importance de l'IA et son impact potentiel sur leur activité, elles ont souvent du mal à passer du pilote à la production. Selon les principaux défis d'IA pour les entreprises de 2021, de le solutions d'IA, citons, par ordre d'importance (voir Figure 1) :

- Coûts (c.-à-d., accélérateurs matériels et ressources de calcul) ▪ Manque de personnel qualifié (c.-à-d., talent) ▪ Manque d'outils et de technologies d'exploitation d'apprentissage automatique ▪ Manque de volume et de qualité de données adéquats ▪ Problèmes de confiance et de gouvernance

FIGURE 1

Aperçu des acheteurs de technologie : défis de mise en œuvre de l'IA/ML



n = 2 000

Source : IDC IA Sondage StrategiesView 2021, avril 2021

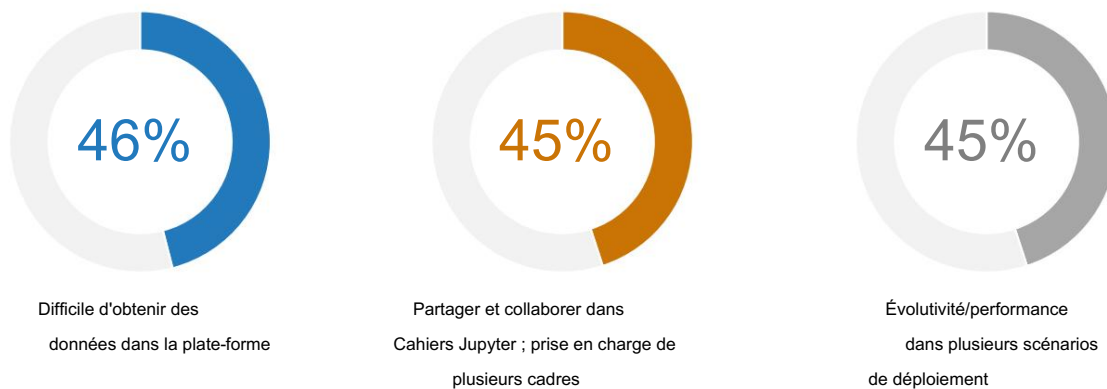
Non seulement les données sont au cœur de l'IA, mais elles constituent également un défi majeur. Plus de la moitié des organisations déclarent manquer des volumes et de la qualité des données nécessaires pour mettre en œuvre une solution d'IA. Mais ce n'est pas seulement une question d'approvisionnement.

Dans la phase de développement du modèle, l'un des plus grands défis pour les entreprises consiste à intégrer les données dans la plate-forme (voir Figure 2). Cela peut être difficile, surtout si les données ne sont pas facilement disponibles dans le bon format.

Fournir une source unique de vérité de données gouvernées de haute qualité est bénéfique non seulement pour les scientifiques des données, mais également pour les analystes et les autres équipes de données. Pour disposer d'une application d'IA pertinente, précise et évolutive, les entreprises doivent s'assurer que leurs données, en temps réel ou par lots, sont de haute qualité et faciles d'accès et de partage en toute sécurité au sein de l'organisation et avec le réseau d'entreprises de l'organisation. les partenaires.

FIGURE 2

### Défis de développement de modèles



n = 2 000

Source : IDC IA Sondage StrategiesView 2021, avril 2021

Pour exploiter toute la puissance des données avec l'IA/ML, les scientifiques des données et les ingénieurs en apprentissage automatique ont besoin des derniers frameworks logiciels et langages de programmation. Des frameworks généraux tels que TensorFlow, MXNet, Caffe, scikit-learn, Keras et PyTorch sont nécessaires, ainsi que des langages de programmation plus spécialisés tels que Python, Java et R.

Cependant, avoir la bonne technologie ne suffit pas. Les modèles d'apprentissage automatique ont besoin des données les plus pertinentes, qui ne se trouvent pas toujours à l'intérieur de l'organisation. Les données internes permettent uniquement aux entreprises de voir leurs propres opérations ou informations sur les clients. Cela ne donne pas une image complète. Les entreprises ont besoin d'accéder à un partage de données sécurisé.

Les données peuvent arriver en temps réel, il est donc important d'exploiter ces données également pour des prédictions en temps réel dans des cas d'utilisation tels que la détection de fraude ou les recommandations de produits.

À mesure que les utilisateurs et les cas d'utilisation prolifèrent, les applications basées sur l'apprentissage automatique doivent être capables de gérer la charge supplémentaire. Si l'application ne parvient pas à évoluer, les goulots d'étranglement des performances peuvent diminuer la valeur de l'utilisation de l'IA/ML. Par exemple, si un client ne reçoit pas les recommandations de produits ou de services qu'il souhaite en temps opportun, il peut être moins susceptible de revenir.

Bien que le développement d'un système évolutif puisse être difficile, il est essentiel de le faire pour gérer la demande croissante des entreprises. Ne pas faire évoluer le système peut entraîner une perte d'activité et des opportunités de revenus manquées. Par exemple, les retards peuvent entraîner des paniers abandonnés ou l'incapacité de faire des recommandations en temps opportun. Lors de la mise à l'échelle de votre système, les organisations doivent être préparées à des problèmes techniques potentiels tels que l'optimisation de l'infrastructure (performances de traitement et élasticité), l'interopérabilité (comme les langages de programmation pris en charge et les cadres ML) et l'intégration des opérations d'apprentissage automatique avec les outils et pratiques DevOps existants.

En s'appuyant sur une infrastructure élastique et intelligente qui ne nécessite pratiquement aucune gestion, les organisations peuvent gérer plus efficacement de gros volumes de données et traiter les données sans goulots d'étranglement, quel que soit le nombre d'utilisateurs, de manière rentable et rapide. Il y a plusieurs avantages inhérents à cette approche, notamment :

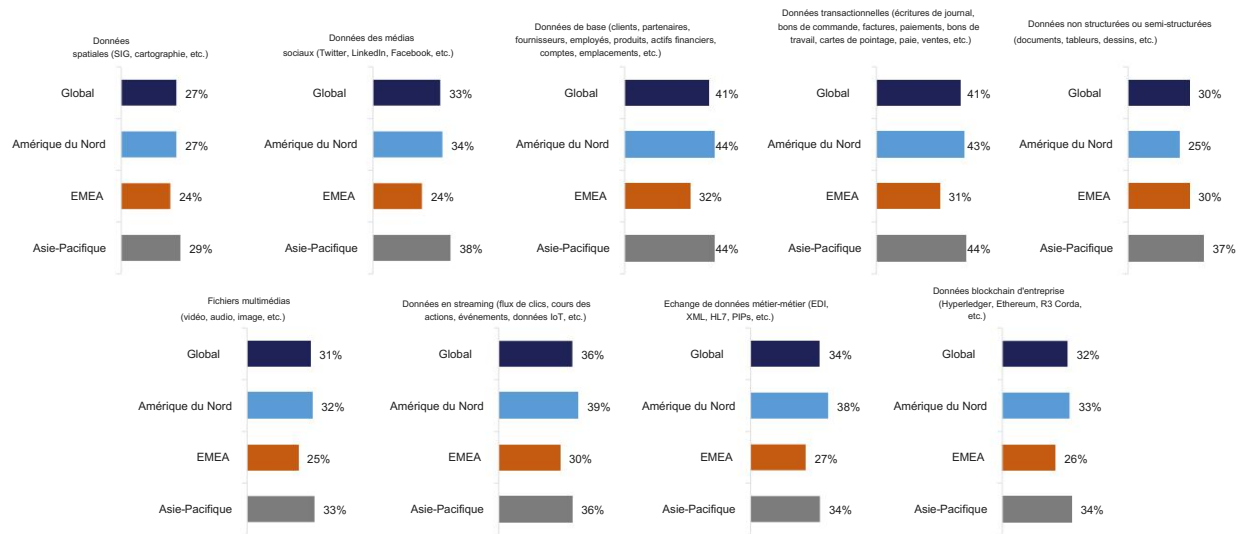
- Productivité plus élevée. Les organisations peuvent être plus agiles et créatives en disposant d'un pipeline qui permet des exécutions rapides de chaque étape (préparation des données, expérimentation, formation de modèle et déploiement).
- Collaboration. Il est avantageux que la formation et les résultats du modèle formé puissent être exploités par toutes les personnes impliquées dans le cycle de vie de l'apprentissage automatique.
- Optimisation du rapport coût/valeur. Cela ne fait jamais de mal d'optimiser les coûts et la valeur. La mise à l'échelle aide à utiliser de manière optimale les ressources disponibles et fait un compromis entre le coût marginal et la précision. L'architecture rationalisée qui prend en charge plusieurs personnes élimine le besoin de systèmes redondants. ▪ Délai de valorisation accéléré. Le pipeline doit être aussi automatisé que possible afin que les professionnels de la science des données puissent se concentrer sur des tâches plus complexes (par exemple, générer de la valeur à partir des données plutôt que de créer des intégrations ou de gérer l'infrastructure, créant ainsi un chemin plus rapide vers les déploiements de production).

Une IA efficace nécessite la diversité des données. De même, l'impact transformateur complet de l'IA peut être réalisé en utilisant un large éventail de types de données. L'ajout de couches de données peut améliorer la précision des modèles et l'impact éventuel des applications. Par exemple, les données démographiques de base d'un consommateur fournissent un aperçu approximatif de cette personne. Si vous ajoutez plus de contexte comme l'état civil, l'éducation, l'emploi, le revenu et les préférences comme la musique et les choix alimentaires, une image plus complète commence à se former. Avec des informations supplémentaires sur les achats récents, l'emplacement actuel et d'autres événements de la vie, le portrait prend vraiment vie.

L'aperçu de l'acheteur de technologie illustré à la figure 3 montre que même si les organisations utilisent un large éventail de types de données, l'utilisation de données non structurées est encore largement inexploitée. De plus, les données continuent d'être cloisonnées (voir la figure 4), ce qui les rend difficiles à accéder et à gérer de manière appropriée.

FIGURE 3

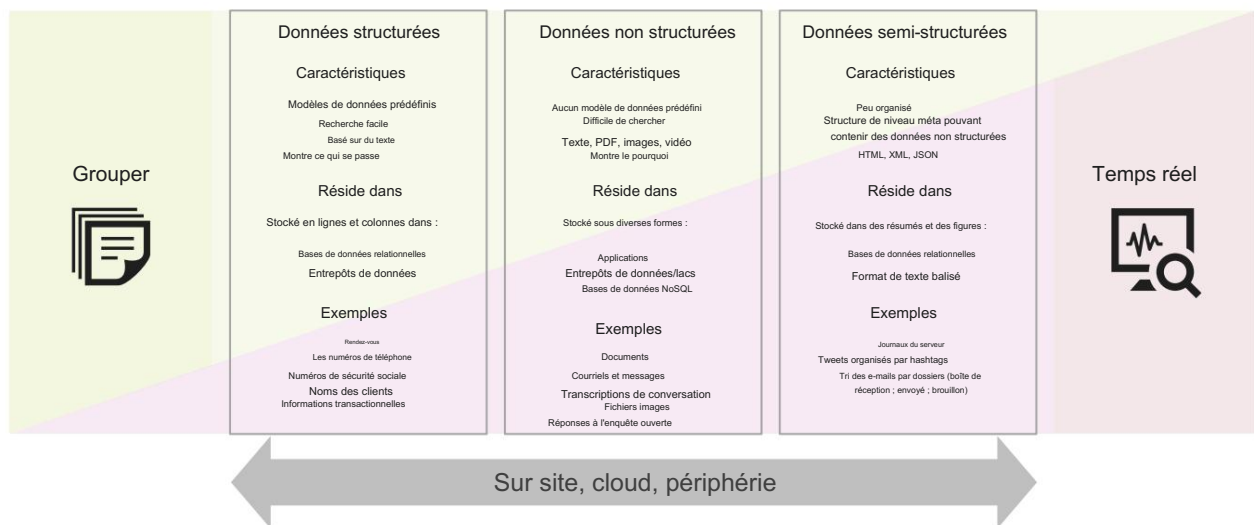
Types de données utilisés pour les solutions IA/ML par région



Source : IDC IA Sondage StrategiesView 2021, avril 2021

FIGURE 4

Données cloisonnées en fonction de la structure des données



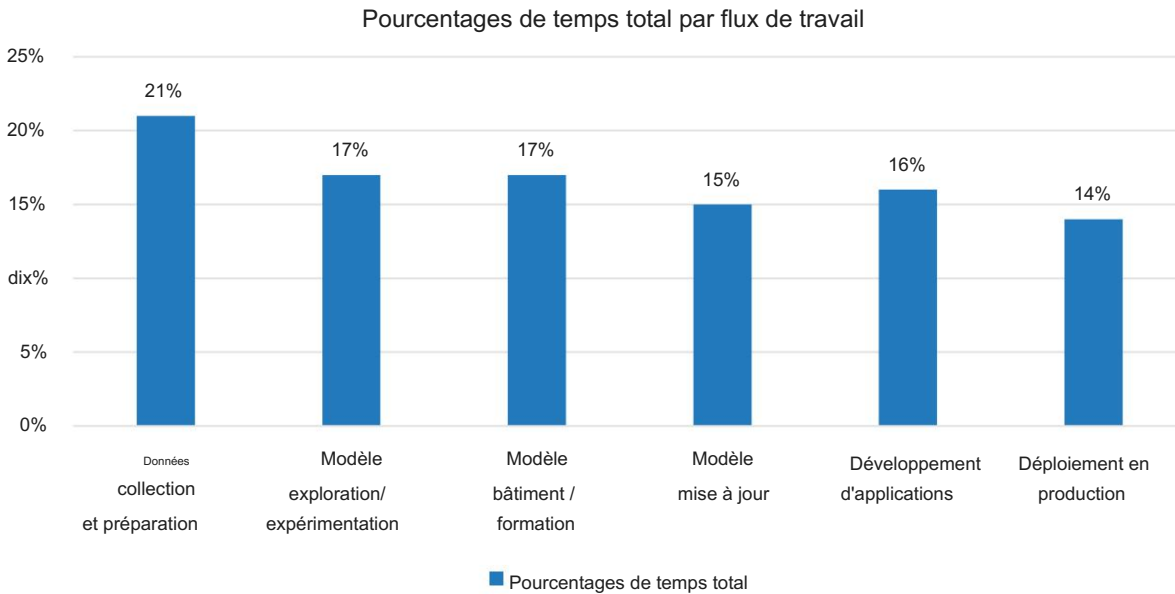
Source : IDC, 2022

Malheureusement, en raison de ces défis, les organisations consacrent plus de temps à des tâches qui ne relèvent pas de la science des données. Par exemple, IDC a constaté que les organisations consacrent le plus grand pourcentage (21%) de leur collecte/préparation des données (voir Figure 5).

Les personnes impliquées et interrogées comprennent les scientifiques des données, les architectes de données, les ingénieurs de données, les ingénieurs en apprentissage automatique, les développeurs d'applications et le personnel d'exploitation.

FIGURE 5

Temps passé à chaque étape du cycle de vie de l'application AI/ML



n = 1 366

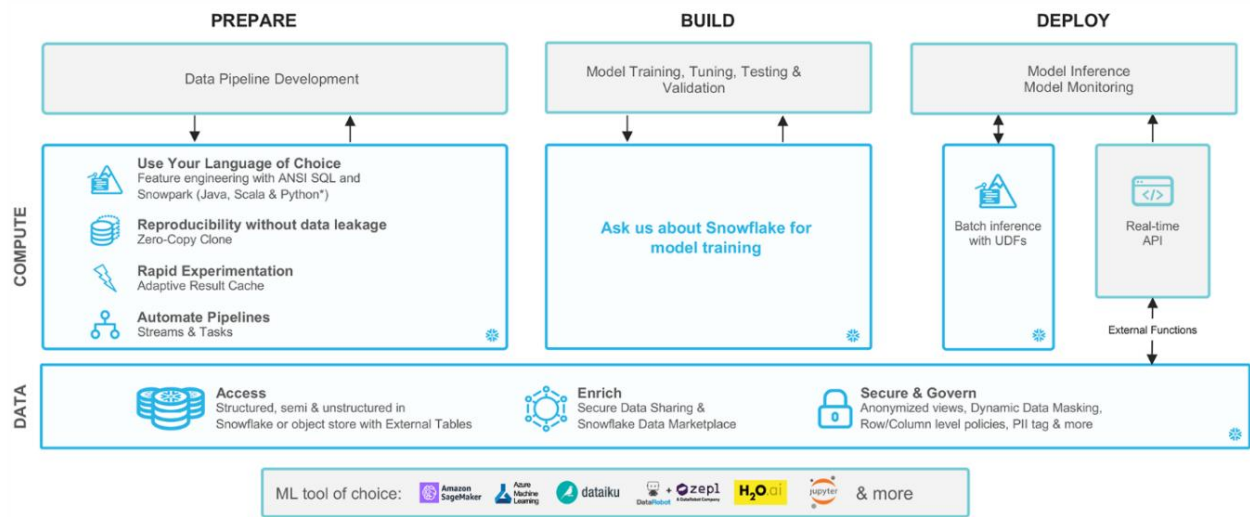
Source : IDC IA Sondage StrategiesView 2021, avril 2021

## CONSIDÉRER SNOWFLAKE POUR LA SCIENCE DES DONNÉES

Le Data Cloud de Snowflake permet aux organisations de consolider plusieurs types et structures de données provenant de nombreuses sources en une seule source de vérité. Cette consolidation permet à toutes les personnes impliquées dans le cycle de vie de l'IA/ML - de la préparation des données à la création de modèles en passant par le déploiement d'applications (voir Figure 6) - de partager des données et de collaborer efficacement pour obtenir rapidement des informations précieuses.

FIGURE 6

## Flocon de neige pour IA/ML



\* Snowpark pour Python est en avant-première.

Source : Flocon de neige, 2022

Les sections qui suivent expliquent comment les organisations et les personnes impliquées dans le cycle de vie de l'IA/ML peuvent bénéficier de l'utilisation de Snowflake.

### Accès facile à toutes les données pertinentes

Snowflake aide les data scientists à découvrir et à accéder à des données structurées, semi-structurées et non structurées (en préversion) pour les workflows de data science, avec une prise en charge native de JSON, AVRO, XML, ORC et Parquet. Même si vos données sont stockées dans un magasin d'objets tel qu'Amazon S3, Apache Glacier (en préversion), Azure Blob Storage et Google Cloud Storage, avec Snowflake, vous pouvez facilement interroger ces données à l'aide de tables externes. L'utilisation d'un ensemble d'outils pour tous les types de données raccourcit le cycle de découverte et de préparation des données. Une fois les données d'entraînement d'un modèle préparées, à l'aide de la fonction Zero-Copy Cloning de Snowflake, les scientifiques des données peuvent prendre un instantané des données pour obtenir une vue ponctuelle de leurs données de production pour la reproductibilité de l'entraînement du modèle sans créer de copies redondantes.

"Les ensembles de données fiables et organisés de Snowflake Data Marketplace rendent les choses tellement plus faciles, et nous sommes ravis de tirer parti de ces ensembles de données pour améliorer les performances de nos modèles d'apprentissage automatique."  
— Mash Syed, scientifique principal des données, Chipotle

Les clients peuvent également utiliser Snowflake pour stocker et matérialiser des fonctionnalités à l'aide de flux et de tâches, ou leurs outils ELT existants tels qu'Airflow et dbt pour gérer les transformations. Snowflake s'associe à des entreprises comme AtScale, Rasgo et Tecton, permettant aux utilisateurs d'adopter des plates-formes prêtes à l'emploi qui améliorent la gestion, la découverte et le traitement des fonctionnalités à grande échelle grâce à des intégrations approfondies avec Snowflake.

Pour fournir aux scientifiques des données un accès à toutes les données pertinentes, même au-delà de leur organisation, le Snowflake Data Cloud simplifie le partage de données entre partenaires, fournisseurs, vendeurs et clients en utilisant le partage de données sécurisé et l'accès aux données tierces via le Snowflake Data Marketplace. Cela offre un accès à des ensembles de données uniques qui peuvent aider à augmenter la précision des modèles sans pipelines de données complexes.

Le partage de données sécurisé dans Snowflake ne nécessite pas de transfert de données via FTP ou la configuration d'API pour lier les applications. Il simplifie l'intégration ETL et synchronise automatiquement les données "en direct" entre les fournisseurs de données et les consommateurs de données. Étant donné que les données source sont partagées plutôt que copiées, les clients n'ont pas besoin de stockage cloud supplémentaire. La place de marché et l'échange de données de Snowflake permettent aux data scientists de collaborer facilement sur des modèles en partageant des données brutes et traitées.

Pour rendre les résultats du modèle facilement accessibles aux utilisateurs qui peuvent agir sur ces résultats de modèle, Snowflake permet aux utilisateurs de les consommer via des tableaux de bord, des rapports et des outils d'analyse commerciale en tirant parti des connexions avec des partenaires de l'écosystème tels que Looker, Sigma, Tableau et ThoughtSpot. Avec Snowflake, vos données peuvent être stockées dans n'importe quel cloud et dans n'importe quelle région pour répondre au mieux à vos besoins avec une expérience de données cohérente pour la collaboration et la migration en cas de besoin.

Dans l'ensemble, en offrant une plate-forme unique, Snowflake élimine le besoin d'exécuter des systèmes séparés chaque fois qu'il y a un changement d'outils, de bibliothèques ou de langues. De plus, les résultats des modèles des scientifiques des données sont introduits dans Snowflake afin qu'ils soient disponibles pour les applications de données et pour les utilisateurs non techniques afin de générer de la valeur commerciale.

## Flexibilité du langage et du cadre

En fournissant aux développeurs une plate-forme unique prenant en charge le langage de leur choix et des solutions open source et commerciales populaires, Snowflake permet aux développeurs de passer plus de temps à générer des informations commerciales exploitables.

Snowpark est un nouveau framework de développement pour Snowflake. Il permet aux ingénieurs de données, aux data scientists et aux développeurs de données de coder à leur manière avec le langage de leur choix, y compris Python, Scala et Java, et d'exécuter des pipelines de données et des flux de travail ML plus rapidement et de manière plus sécurisée sur une seule plate-forme. Les développeurs veulent de la flexibilité lorsqu'ils travaillent avec des données, des environnements évolutifs de manière élastique qui ne nécessitent pratiquement aucun travail d'administration et de maintenance, et un accès immédiat aux données dont ils ont besoin.

Avec Snowpark, les développeurs peuvent libérer l'échelle et les performances du moteur de Snowflake et tirer parti des contrôles de gouvernance et de sécurité natifs intégrés à la plate-forme de Snowflake.

Grâce au vaste écosystème de partenaires de Snowflake, les clients peuvent profiter de connexions directes aux outils et langages de science des données existants et émergents tels que Python, R, Java et Scala ; bibliothèques open source telles que PyTorch, XGBoost, scikit-learn et TensorFlow ; des cahiers comme Jupyter et Zeppelin ; et des plates-formes telles qu'Amazon SageMaker, Dataiku, DataRobot et H2O.ai. Selon Snowflake, les utilisateurs d'Amazon SageMaker peuvent soit tirer parti des intégrations prédéfinies vers Amazon SageMaker Data Wrangler ou Amazon SageMaker Autopilot, soit utiliser le connecteur Snowflake pour Python pour remplir directement les Pandas DataFrames dans leurs instances de bloc-notes. Cette connexion à haut débit permet d'accélérer le développement de modèles, ainsi que d'optimiser la préparation des données et l'ingénierie des fonctionnalités. Dataiku, DataRobot et H2O.ai ont une intégration Snowflake intégrée où ses utilisateurs peuvent rapidement connecter leur compte à Snowflake et pousser le traitement à travers plusieurs étapes du flux de travail vers le moteur de performance élastique de Snowflake.

Grâce au partenariat et aux intégrations de produits Snowflake et Anaconda, les utilisateurs de Snowflake peuvent désormais accéder de manière transparente à l'un des écosystèmes les plus populaires de bibliothèques open source Python sans le



besoin d'installations manuelles et de gestion des dépendances de packages. L'intégration peut stimuler la productivité des développeurs Python.

En résumé, avec Snowpark pour Python (en préversion), les équipes data peuvent :

- Accélérer leur rythme d'innovation en utilisant la syntaxe familière de Python et l'écosystème florissant de bibliothèques open source pour explorer et traiter les données là où elles se trouvent.
- Optimisez le temps de développement en supprimant le temps passé à gérer des environnements Python défectueux grâce à un gestionnaire de dépendances de packages Python intégré. • Opérez avec une confiance et une sécurité améliorées en éliminant les copies non contrôlées des données avec tout le code exécuté dans un bac à sable hautement sécurisé directement à l'intérieur de Snowflake.

Pour mettre toutes ces fonctionnalités à la disposition de leur riche écosystème de partenaires, le programme accéléré Snowpark met en avant les partenaires avec des intégrations Snowpark qui étendent le moteur de Snowflake à leurs clients.

## Performances à travers les étapes du flux de travail ML

Snowflake peut gérer simultanément de grandes quantités de données et d'utilisateurs. Son infrastructure de calcul multicluster intelligente s'adapte automatiquement pour répondre aux demandes d'ingénierie de fonctionnalités sans aucun goulot d'étranglement ni limitation de la simultanéité des utilisateurs. Pour automatiser et mettre à l'échelle les pipelines d'ingénierie de fonctionnalités, les utilisateurs peuvent tirer parti des flux et des tâches pour que leurs données soient prêtes pour l'inférence de modèle.

Pour l'inférence en bloc, Snowflake simplifie le chemin vers la production avec la possibilité de déployer des modèles dans Snowflake en tant que fonctions définies par l'utilisateur (UDF). Des partenaires tels que Dataiku, DataRobot et H2O.ai créent une expérience plus intégrée pour que les utilisateurs disposent d'un flux de travail guidé pour déployer des modèles formés dans Snowflake sans effort. Pour l'inférence en temps réel, les utilisateurs peuvent déployer des modèles dans une couche externe (par exemple, Docker) et demander facilement des prédictions directement depuis Snowflake en utilisant des fonctions externes pour communiquer avec le point de terminaison de l'API du modèle.

## Sécurité et gouvernance de niveau entreprise

Snowflake permet aux organisations d'appliquer des contrôles de gouvernance et une sécurité cohérents au niveau de l'entreprise dans tous les flux de travail IA/ML, limitant ainsi les biais de l'IA. Le Data Cloud de Snowflake est construit sur une base de sécurité multicouche qui comprend le chiffrement, le contrôle d'accès, la surveillance du réseau et les mesures de sécurité physiques, permettant la robustesse contradictoire nécessaire à une solution AI/ML. En plus des certifications technologiques standard de l'industrie telles que ISO/IEC 27001 et SOC 1/SOC 2 Type 2, Snowflake est conforme aux réglementations gouvernementales et industrielles importantes telles que les certifications PCI, DSS, HIPAA/Health Information Trust Alliance et FedRAMP. Toute cette conformité est essentielle pour les déploiements d'IA/ML dans les cas d'utilisation industriels. Les fonctionnalités évolutives de gouvernance et de sécurité des données de Snowflake permettent aux organisations de gérer facilement leurs initiatives de confiance en matière d'apprentissage automatique. Grâce à des fonctionnalités de sécurité telles que les vues anonymisées, le masquage dynamique des données et les politiques au niveau des lignes/colonnes, les organisations peuvent s'assurer que les scientifiques des données ne peuvent pas utiliser des informations sensibles susceptibles d'entraîner des biais dans les modèles.

## RECOMMANDATIONS

---

Aujourd'hui, les entreprises sont confrontées à un ensemble complexe de défis commerciaux, notamment un rythme d'activité croissant, un volume croissant de données commerciales, la nécessité de réfléchir à des stratégies de données partagées pour tirer une véritable valeur des données, une portée croissante du commerce mondial et une multitude de risques pour les clients, les employés et les fournisseurs. Le volume de clients et de fournisseurs, ainsi que la complexité réglementaire et les entreprises multisectorielles, signifient que la complexité est courante dans les entreprises mondiales.

Les entreprises rationalisent, modernisent et transforment leurs portefeuilles d'applications d'entreprise.

L'apprentissage automatique, le traitement du langage naturel, les interfaces utilisateur d'assistance et les analyses avancées associées à des ensembles de données organisés font progresser les applications traditionnelles pour qu'elles deviennent intelligentes.

Ces applications intelligentes permettent d'obtenir plus d'informations sur les employés en automatisant les transactions qui étaient auparavant bloquées et en apportant plus de données dans l'équation afin que les organisations puissent prendre de meilleures décisions immédiatement. Les organisations ont besoin d'une stratégie de données pour l'IA, qui variera considérablement en fonction de la taille, de la nature et de la complexité de leur entreprise et de leur stratégie d'IA. Pour accélérer l'innovation et le délai de valorisation et bénéficier d'un avantage concurrentiel durable, il est conseillé aux acheteurs de technologies de :

- Construire un vivier de talents d'experts techniques et du domaine de l'industrie tels que des ingénieurs de données, scientifiques et ingénieurs en apprentissage automatique.
- Obtenez l'adhésion et la confiance des employés pour la stratégie de données avec inclusivité et transparence. • Créer un flux de travail pour intégrer des sources de données tierces et/ou nouvelles dans l'organisation, y compris les tests, l'achat et l'intégration transparente avec les ensembles de données et les processus internes existants.
- Assurez-vous que le processus est interfonctionnel entre l'informatique, les achats, le juridique, la conformité et la sécurité. • Sélectionnez une plate-forme de données sécurisée et gouvernée prenant en charge tous les types de données  
Flux de travail du cycle de vie AI/ML.
- Assurez la flexibilité de la programmation grâce à la prise en charge de plusieurs langages de programmation tels que Python, Java et Scala, ainsi qu'à des workflows d'apprentissage automatique de premier plan tels que TensorFlow, PyTorch et scikit-learn.
  
- Adoptez une grille de données intelligente qui aide à :
  - Automatisez et appliquez des politiques universelles de données et d'utilisation dans les écosystèmes multcloud. • Automatisez la façon dont les données sont découvertes, cataloguées et enrichies pour les utilisateurs. • Automatisez la façon d'accéder, de mettre à jour et d'unifier les données réparties sur des données distribuées et des paysages cloud sans avoir à effectuer de déplacement ou de réplication de données.

## CONCLUSION

---

De nombreuses entreprises adoptent l'IA au cours de leur transformation numérique, non seulement parce qu'elles le peuvent, mais parce qu'elles le doivent. L'IA est la technologie qui aide les entreprises à être agiles, innovantes et évolutives.

Les entreprises qui réussiront deviendront des organisations "IA first" capables de synthétiser les informations (c'est-à-dire d'utiliser l'IA pour convertir les données en informations puis en connaissances), d'apprendre (c'est-à-dire d'utiliser l'IA pour comprendre les relations entre les connaissances et d'appliquer l'apprentissage aux problèmes de l'entreprise) et de fournir des informations à grande échelle (c'est-à-dire utiliser l'IA pour prendre en charge les décisions et l'automatisation). L'IA devient omniprésente dans tous les domaines fonctionnels d'une entreprise. IDC prévoit que le marché global des logiciels d'IA atteindra 596 milliards de dollars de revenus d'ici 2025, avec un TCAC de 17,7 %.

Les données sont au cœur des initiatives d'IA. Les organisations doivent renforcer leur stratégie de données pour l'IA et adopter une plate-forme de données sécurisée, gouvernée, collaborative et évolutive qui aide les professionnels de la science des données à se concentrer sur la science des données et à faire évoluer les initiatives d'IA de manière transparente.

## DÉFINITIONS

---

- L'intelligence artificielle (IA) est définie comme des systèmes informatiques qui apprennent, raisonnent, s'adaptent et s'auto correct.
- L'apprentissage automatique (ML) est un sous-ensemble de techniques d'IA qui permet aux systèmes informatiques d'apprendre sans programmation par un humain.
- L'apprentissage en profondeur (DL) est un sous-ensemble de techniques de ML qui rend les réseaux de neurones multicouches informatiques réalisables.
- La science des données est un domaine interdisciplinaire qui utilise des méthodes, des processus, des algorithmes et des systèmes scientifiques pour extraire des connaissances et des idées à partir de données bruyantes, structurées et non structurées et appliquer des connaissances et des idées exploitables à partir de données dans un large éventail de domaines d'application. La science des données est liée à l'exploration de données, à l'apprentissage automatique et au big data. ▪ Une grille ou structure de données intelligente est une approche architecturale qui se tisse et automatise le cycle de vie des données et de l'IA sur plusieurs clouds et sources de données.

[APPRENDRE ENCORE PLUS](#)

---

## Recherche connexe

- Accélérez l'innovation et l'avantage concurrentiel durable avec IA (IDC #US48354321, novembre 2021) <sup>un</sup> Stratégie de données solide pour
- Perspective d'analyse de marché : logiciel mondial d'intelligence artificielle, 2021 (IDC #US48243221, septembre 2021)
- Gérer les risques commerciaux liés à l'IA/ML et prospérer grâce à la confiance<sup>IA</sup> (IDC #US48235521, septembre 2021)
- Prévisions mondiales des logiciels d'intelligence artificielle, 2021-2025 (IDC #US48125621, août 2021)
- Magasins de fonctionnalités : essentiels pour <sup>Mise à l'échelle</sup> ML Initiatives et accélération à la fois du haut et du bas Impact de ligne (IDC #US47223320, janvier 2021)
- MLOps : le nouvel avantage concurrentiel de votre entreprise (IDC #US46643620, juillet 2020)

MESSAGE DU COMMANDITAIRE

Pour en savoir plus sur l'approche et les offres de Snowflake, sélectionnez l'une des options :

Visitez <https://www.snowflake.com/>

Visitez [l'étude de cas Chipotle](#)

Visitez [l'étude de cas de Kount](#)

## À propos d'IDC

International Data Corporation (IDC) est le premier fournisseur mondial d'informations commerciales, de services de conseil et d'événements pour les marchés des technologies de l'information, des télécommunications et des technologies grand public. IDC aide les professionnels de l'informatique, les dirigeants d'entreprise et la communauté des investisseurs à prendre des décisions fondées sur des faits concernant les achats de technologie et la stratégie commerciale. Plus de 1 100 analystes d'IDC fournissent une expertise mondiale, régionale et locale sur les opportunités et les tendances technologiques et industrielles dans plus de 110 pays à travers le monde. Depuis 50 ans, IDC fournit des informations stratégiques pour aider nos clients à atteindre leurs principaux objectifs commerciaux. IDC est une filiale d'IDG, le leader mondial des médias technologiques, de la recherche et de l'événementiel.

## Siège mondial

140, rue Kendrick  
Bâtiment B  
Needham, MA 02494  
cerf  
508.872.8200  
Twitter : @IDC  
blogs.idc.com  
www.idc.com

---

### Copyright

Publication externe d'informations et de données d'IDC — Toute information d'IDC destinée à être utilisée dans des publicités, des communiqués de presse ou des supports promotionnels nécessite l'approbation écrite préalable du vice-président ou du directeur national d'IDC concerné. Une ébauche du document proposé doit accompagner toute demande de ce type. IDC se réserve le droit de refuser l'approbation d'une utilisation externe pour quelque raison que ce soit.

Copyright 2022 IDC. La reproduction sans permission écrite est interdite.

