



Le guide de référence par excellence pour construire une base de données à l'ère de l'IA générative

Trois attributs clés pour permettre à votre organisation de libérer toute la puissance de vos données.



Table des matières

Introduction	3
Devenir axé sur les données	5
Une approche complète	8
Une approche intégrée	13
Une approche gouvernée	18
Conclusion	24



INTRODUCTION

Les données sont l'étincelle qui entraîne l'innovation

Aujourd'hui, plus que jamais, les données sont au centre de chaque application, de chaque processus et de chaque décision opérationnelle. Elles représentent la genèse des inventions modernes. Dans le paysage complexe et en constante évolution d'aujourd'hui, la façon dont vous exploitez les données de votre organisation peut être la clé pour accélérer l'innovation et atteindre vos objectifs organisationnels. Les enjeux sont importants. Selon Forrester Research, les « entreprises [qui] disposent de capacités commerciales avancées axées sur la connaissance sont 8,5 fois plus susceptibles que... les entreprises en phase de démarrage de signaler que leur chiffre d'affaires annuel a augmenté de 20 % ou plus. »¹

L'avènement de l'IA générative met également l'accent sur l'importance des données. Lorsque vous voulez créer des applications d'IA générative selon vos besoins commerciaux, les données constituent le facteur de différenciation. Les données sont la clé pour remplacer les applications génériques par des applications d'IA générative qui créent une véritable valeur pour vos clients et votre entreprise.



Il n'est pas nécessaire de réinventer la roue

Devenir une organisation axée sur les données, c'est d'abord établir la bonne base de données. La bonne nouvelle, c'est qu'il existe déjà une base de données éprouvée et que les organisations en profitent déjà à l'aide d'Amazon Web Services (AWS). Par exemple, [AstraZeneca](#) intègre et met à l'échelle ses données et ses capacités d'intelligence artificielle dans l'ensemble de l'entreprise afin d'innover plus rapidement et améliorer les résultats des patients. Elle peut désormais exécuter plus de 51 milliards de tests statistiques en moins de 30 heures, ce qui facilite l'apport d'informations génomiques aux projets de découverte de médicaments. Le [groupe BMW](#) utilise les données pour optimiser sa chaîne d'approvisionnement et améliorer sa capacité de production. [LG AI Research](#), pour sa part, exploite ses données pour développer des applications d'IA générative afin de transformer les processus commerciaux, ce qui lui permet d'élargir l'accès à l'IA dans divers secteurs tels que la mode, la fabrication, la recherche, l'éducation et la finance.

Il est possible de bâtir la base de donnée qui saura transformer votre organisation. Lisez la suite pour découvrir les principes fondamentaux.

Principaux défis et considérations

Le volume de données générées et stockées est plus élevé que jamais

Les outils sur place et les magasins de données hérités ne peuvent pas répondre aux demandes actuelles. Les organisations ont besoin de magasins de données capables de s'adapter à des pétaoctets, voire même à des exaoctets de données. Ces magasins doivent également être en mesure de stocker ces données de manière rentable sans pour autant sacrifier le rendement.

Les données en silos réparties entre de nombreuses sources affectent la productivité et augmentent coûts.

Les entreprises doivent gérer divers types de données, notamment des fichiers journaux, des flux de clics, de données vocales et vidéo, qui sont généralement stockés en silos dans plusieurs magasins de données et au sein de plusieurs services. Il est donc difficile de tirer parti des données et d'en extraire des renseignements exploitables. Pour transformer l'infrastructure d'une source de complexité et de dépenses en un moteur de création de valeur, les entreprises doivent briser ces silos et unifier l'ensemble de leurs données.

L'état actuel de la prise de décision n'est pas viable

Gartner révèle que 65 % des décisions prises aujourd'hui sont plus complexes (car davantage de parties prenantes y participent ou elles impliquent plus de choix) qu'il y a cinq ans². Pour prendre de meilleures décisions et plus rapidement, les organisations doivent être en mesure d'effectuer des opérations d'analytique et d'apprentissage automatique de manière agile et rentable en faisant appel aux outils et aux performances nécessaires pour s'adapter à chaque cas d'utilisation. Les organisations ne peuvent plus perdre leur temps à redéployer et à reconfigurer constamment leur infrastructure pour faire évoluer les performances et la capacité.





Essayer de gérer la gouvernance des données est un travail à plein temps.

Les architectures de données traditionnelles ont besoin de procédures de gestion risquées et compliquées, car on accède aux données à partir de nombreux endroits. Accorder, suivre, contrôler et supprimer l'accès des employés tout en restant en conformité avec un nombre croissant de réglementations est un travail à plein temps. L'automatisation de ces tâches obligatoires de gouvernance des données permet aux équipes modernes de se concentrer à nouveau sur l'innovation.

L'adoption de l'analytique et de l'apprentissage automatique est toujours entravée par un manque de compétences et par l'inertie

De nombreuses entreprises ont du mal à progresser avec des outils d'analytique et d'apprentissage automatique à grande échelle. Gartner constate que les organisations qui investissent dans l'intelligence artificielle n'ont mis en production que 54 % de leurs projets pilotes de validation de concept d'intelligence artificielle³. Un manque persistant de compétences en matière de données et d'apprentissage automatique, ainsi que la quantité ou la qualité des données utilisées à des fins d'entraînement ne sont que quelques-uns des problèmes qui ralentissent les progrès dans ce domaine important. Cependant, il est de plus en plus nécessaire d'aider les utilisateurs professionnels à prendre des décisions axées sur les données.

Les données sont de plus en plus difficiles à sécuriser

Il fut un temps où les équipes des TI choisissaient entre l'accélération et la sécurisation de leurs architectures. À présent, elles doivent faire les deux. Parallèlement, 97 % des entreprises ont été confrontées à une augmentation des menaces de cybersécurité entre 2022 et 2023, selon le rapport State of Cybersecurity Resilience 2023 d'Accenture. Le taux d'atteintes à la sécurité réussies en provenance de réseaux externes demeure important, s'élevant à 61 %.⁴ Comment les organisations peuvent-elles optimiser la protection de la vie privée et la sécurité?

³ « [How to make better business decisions](#) », Gartner, 20 octobre 2021

⁴ « [État de la résilience en cybersécurité 2023 : comment la cybersécurité stimule la réinvention des entreprises pour renforcer leur résilience](#) », Accenture, 2023

Trois attributs clés pour aider votre organisation à exploiter pleinement les données

Les organisations fortement axées sur les données sont trois fois plus susceptibles d'améliorer de façon significative la prise de décision que celles qui s'appuient moins sur les données⁵.

La mise en œuvre d'une base de données qui facilite la gestion des données à chaque étape du parcours, depuis l'ingestion, le stockage et l'interrogation des données jusqu'à l'analyse, la visualisation et l'exécution des modèles d'apprentissage automatique, nécessite trois attributs fondamentaux. Quels que soient vos défis opérationnels, votre base de données doit être :

Complète : offrir les outils adéquats et proposer un rapport prix-performances optimal, pour tous les utilisateurs, tous les types de données et tous les cas d'utilisation.

Intégrée : brisez les silos pour relier toutes les données de votre organisation afin qu'elles puissent être utilisées efficacement

Gouvernée : outillez vos équipes pour leur permettre d'agir plus rapidement grâce à un accès aux données gouverné là et quand vos utilisateurs en ont besoin pour accélérer l'innovation

Le succès d'une organisation axée sur les données peut également nécessiter un changement d'état d'esprit plus large, dans lequel les objectifs et les décisions sont soutenus par une base de données qui tient compte des personnes, des processus, des outils et de l'éducation.



COMPLÈTE

Offrir les bons outils et proposer un prix optimal quel que soit l'utilisateur et le type de données

Les entreprises doivent élaborer des bases de données durables, capables de répondre à leurs besoins actuels et futurs. Pour exploiter efficacement les données, il ne suffit pas d'un lac de données, d'un entrepôt de données ou d'outils de veille stratégique. Il faut une base de données munie d'un ensemble complet d'outils qui tiennent compte de l'échelle et de la variété des données, ainsi que de nombreux objectifs pour lesquels vous souhaitez les utiliser.

Un fournisseur de services infonuagiques qui innove pour vous offrir en permanence tous les outils de données dont vous avez besoin, avec un rapport prix-performances adapté à votre cas d'utilisation, permet d'assurer la création d'une base de données qui évolue avec vous. AWS dispose de l'ensemble le plus large et le plus approfondi de capacités de données pour prendre en charge toutes les applications ou tous les cas d'usage liés aux données. Depuis les bases de données pour les applications jusqu'au stockage pour les lacs de données, à l'analytique, à l'apprentissage automatique et aux outils pour les utilisateurs finaux, AWS fournit les bonnes capacités dans chaque domaine, afin que vous n'ayez pas à faire de compromis sur les performances, le coût ou les résultats. AWS accélère continuellement son rythme d'innovation pour rester toujours à la hauteur de vos exigences en matière de données.

Mise à l'échelle des applications axées sur les données

Créez des applications sur une base de données moderne au meilleur prix et avec les meilleures performances pour votre cas d'utilisation à l'échelle grâce aux bases de données AWS. Plus de 100 000 organisations, par exemple, atteignent le rendement et la disponibilité des bases de données de qualité commerciale pour un dixième du coût grâce à [Amazon Aurora](#). Pour des cas d'utilisation tels que les graphiques, la diffusion en continu et les documents, AWS propose huit moteurs de base de données conçus de manière unique pour offrir des performances optimales à vos applications et transformer l'aspect économique de la possession d'une base de données.



Conseil en matière de base de données

Mettez en place une base de données rentable sans pour autant sacrifier le rendement. Permettez à votre organisation de maximiser ses capacités actuelles en optimisant les coûts :

Samsung a économisé 44 % sur ses coûts opérationnels mensuels et 22 % supplémentaires sur les frais de maintenance en procédant à une migration vers Amazon Aurora PostgreSQL.

Carrier a relié son réseau logistique de la chaîne du froid pour aider ses clients à optimiser les opérations de la chaîne du froid, à diminuer leur consommation d'énergie et à améliorer leurs résultats en réduisant les coûts, les retards, les pertes de marchandises et les dégradations en cours de transport.

United Airlines a créé un aéroport intelligent avec plus de 20 000 capteurs qui produisent des données permettant d'obtenir des informations en temps réel et qui optimisent les capacités des équipements au sol. Cela a permis d'économiser 120 millions de dollars, puisque l'achat de certains équipements n'était plus nécessaire.



AWS collabore avec plus de
**deux
millions
de clients**
*pour résoudre certains des
problèmes de données les
plus complexes au monde*



Alimenter les lacs de données et les entrepôts de données

Les organisations utilisent des lacs de données et des entrepôts de données pour unifier les jeux de données et appliquer l'analytique et l'apprentissage automatique afin d'aider à la prise de décision et de créer de nouvelles expériences client axées sur les données.

AWS aide depuis des années des centaines de milliers de clients à créer des lacs de données leur permettant de stocker de manière fiable et sécurisée des données structurées et non structurées grâce à des services tels que [Amazon Simple Storage Service \(Amazon S3\)](#), [AWS Glue](#) et [AWS Lake Formation](#). Des centaines de milliers de clients disposent de lacs de données alimentés par AWS. Pour un stockage de données permettant d'obtenir des résultats de requêtes ultra-rapides à partir de nombreuses sources de données différentes.

Optimisez le rendement de votre stockage et accélérez l'accès aux données

Pour vos données les plus fréquemment consultées, les classes de stockage telles que S3 Express One Zone peuvent multiplier par 10 les vitesses d'accès aux données et réduire les coûts de requête de 50 % par rapport à S3 Standard.

Fourniture d'analytique pour tous les cas d'usage

La véritable agilité aide les organisations à s'adapter rapidement à l'évolution des besoins opérationnels. Les services d'analyse AWS permettent aux équipes de votre entreprise d'ingérer, de combiner et d'exécuter des analyses historiques, en temps réel et prédictives, sur vos données. Cela comprend des services pour les requêtes SQL, l'analytique de journaux, la diffusion en continu et Apache Spark. Pour un stockage de données permettant d'obtenir très rapidement des résultats de requêtes sur de nombreuses sources de données différentes, [Amazon Redshift](#), un entrepôt de données à l'échelle du pétaoctet, offre un rapport prix-rendement jusqu'à six fois supérieur à celui d'autres entrepôts de données en nuage. Grâce à l'IA générative intégrée à Amazon Redshift, vous pouvez générer des requêtes SQL en langage naturel. De plus, la mise à l'échelle et les optimisations fondées sur l'IA d'Amazon Redshift Serverless vous aident à optimiser les coûts et le rendement en tirant des leçons de vos tendances passées.

Pour l'interrogation de mégadonnées, vous pouvez prendre en charge plus de cadres de mégadonnées que tout autre fournisseur à l'aide d' [Amazon EMR](#), qui vous permet d'obtenir des informations jusqu'à deux fois plus rapidement. Nos clients obtiennent des performances plus de trois fois supérieures grâce à Apache Spark lorsqu'ils utilisent nos exécutions pleinement prises en charge et optimisées par AWS pour EMR.

Innover plus rapidement grâce à des services qui rendent l'AA et l'IA plus accessibles

Les organisations utilisent l'AA pour ajouter de l'intelligence aux processus existants, automatiser les tâches manuelles fastidieuses et accélérer l'innovation grâce aux données. Aujourd'hui, grâce à l'IA générative, elles ont la possibilité de réinventer les expériences clients et les applications.

Avec AWS, vous avez accès à l'ensemble le plus complet de services d'intelligence artificielle et d'apprentissage automatique. [Amazon Bedrock](#) est le moyen le plus simple de créer et de mettre à l'échelle des applications d'IA générative avec des modèles de fondations (FM) pour créer de nouveaux contenus et de nouvelles idées, y compris des conversations, des histoires et des images. Grâce à Bedrock, vous pouvez utiliser vos propres données pour personnaliser facilement et en toute sécurité les FM d'AI21 Labs, d'Anthropic et de Stability AI ainsi que les modèles [Amazon Titan](#) par l'entremise d'une API. AWS propose également une large gamme de services qui vous permettent d'ajouter des capacités d'IA telles que la reconnaissance d'image, les prévisions et la recherche intelligente aux applications par un simple appel d'API. Pour créer vos propres modèles d'AA, [Amazon SageMaker](#) fournit les outils nécessaires à la création, à l'entraînement et au déploiement de modèles d'AA à grande échelle.

Mise à disposition des informations sur les données dans l'ensemble de l'organisation

Ce n'est plus l'apanage de quelques personnes compétentes en matière de données que de pouvoir extraire des informations précieuses et pertinentes à partir des données afin d'éclairer la prise de décision. Les solutions de veille stratégique alimentées par l'apprentissage automatique, comme [Amazon QuickSight](#), facilitent la connectivité aux sources de données. Les analystes commerciaux peuvent utiliser ces données pour présenter des nouvelles tendances et des prédictions sur des visualisations et des tableaux de bord de veille stratégique interactifs.

Amazon Q, notre assistant alimenté par l'IA générative, vous aide à créer des tableaux de bord et des histoires visuelles convaincantes dans [QuickSight Q](#) à partir des données de votre tableau de bord, le tout en langage naturel. Les utilisateurs professionnels peuvent facilement tirer parti de perspectives pertinentes en libre-service. Même s'ils posent des questions vagues en langage naturel, ils recevront des réponses complètes et adaptées au contexte qui expliquent les données de manière exhaustive à l'aide d'éléments visuels et narratifs.

Amélioration de la maîtrise des données

En sachant comment utiliser les données de manière efficace, vos employés peuvent aider votre organisation à atteindre ses objectifs en matière de données. Investissez dans la formation et l'amélioration des compétences de votre personnel en matière de données, d'analytique et d'apprentissage automatique avec la [formation et la certification AWS](#).



Développez un processus décisionnel axé sur les données dans l'ensemble de votre organisation

Permettez aux utilisateurs professionnels de :

- répondre à différents besoins analytiques à partir de la même source d'authenticité grâce à la veille stratégique unifiée sur [Amazon QuickSight](#);
- extraire des perspectives clés plus rapidement à l'aide d'[Amazon Q](#), un assistant alimenté par l'IA générative;
- construire des tableaux de bord et créer des récits visuels convaincants en utilisant le langage naturel avec Amazon Q dans QuickSight;
- générer des prédictions d'AA précises sans expérience préalable grâce à [Amazon SageMaker Canvas](#), et ce, sans avoir à coder;
- simplifier l'accès gouverné aux données pour les utilisateurs professionnels avec [Amazon DataZone](#);
- accéder à plus de 150 cours de développement professionnel liés aux données, à l'analytique et à l'AA par l'entremise de la [formation et la certification AWS](#).



ADP prend 312 mille milliards de décisions par mois grâce aux processus analytiques



ADP aide plus de 900 000 entreprises à gérer 70 millions d'employés grâce à son processus de gestion du personnel et des salaires. Cette gestion génère une quantité massive de données. En fait, ADP traite plus de 2,5 pétaoctets de données avec plus de 25 milliards de points de données individuels représentés. ADP utilise Amazon Redshift et [Amazon Neptune](#) pour gérer certains aspects de son traitement global des données. Ces services de données aident les entreprises à mesurer, à comparer, à prédire et à appliquer des informations sur leur main-d'œuvre. Grâce aux services AWS, ADP permet également aux organisations de créer des tableaux de bord sur l'équité salariale, ce qui aide plus des deux tiers des entreprises à afficher une amélioration de l'équité salariale.

« L'heure est venue d'utiliser les données pour aider les gens à comprendre les actions que nous pouvons entreprendre pour créer un environnement de travail plus inclusif, diversifié et équitable tout en bâtissant l'avenir que nous voulons tous créer. »

Jack Berkowitz,
directeur des données, ADP

[Lire l'intégralité de l'article >>](#)



Le Groupe BMW démocratise l'utilisation des données à grande échelle



Le Groupe BMW a anonymisé les données provenant des capteurs des véhicules et d'autres sources à l'échelle de l'entreprise, et ce, pour les rendre facilement accessibles aux équipes internes qui créent des applications internes et destinées aux clients avec AWS. L'entreprise a choisi un lac de données centralisé basé sur AWS pour son agilité et sa capacité à traiter quotidiennement des téraoctets de données de télémétrie provenant de millions de véhicules. La constitution d'un catalogue de données lisible par l'homme et l'affichage clair des ressources de données se sont avérés essentiels. Cela a stimulé la productivité des analystes de données, des scientifiques des données et des ingénieurs.

« Nous ne faisons que commencer notre parcours avec AWS, et nous sommes impatients d'aider notre entreprise à accomplir sa stratégie axée sur l'innovation dans l'avenir. »

Kai Demtröder,
vice-président de la transformation des données, de l'intelligence artificielle, des plateformes de données et DevOps, Groupe BMW

[Lire l'intégralité de l'article >>](#)



INTÉGRÉ

Brisez les silos pour relier toutes les données de votre organisation afin qu'elles puissent être utilisées efficacement

Les occasions de transformer votre entreprise à l'aide des données se retrouvent sur l'ensemble de la chaîne de valeur. Toutefois, pour opérer une telle transformation, vous devez avoir une vue d'ensemble de votre client et de votre entreprise. Avec des données réparties entre plusieurs départements, services, bases de données sur site et applications tierces, vous devez être en mesure d'intégrer facilement les données des différents silos pour obtenir les meilleures informations. Les entreprises peuvent adopter différentes approches en matière d'unification des données (maillage de données, lake house, tissu de données, etc.), mais en général, elles utilisent un lac de données comme élément de base. Les lacs de données vous permettent de recueillir, de stocker, d'organiser et de traiter les données précieuses de vos silos de données et de les mettre à la disposition des outils d'IA, de visualisation et d'analytique d'une façon contrôlée.

Adopter un avenir zéro-ETL

De nombreuses organisations possèdent plusieurs lacs de données en plus des entrepôts de données, des bases de données des outils d'analytique, des outils d'apprentissage automatique et des applications SaaS. L'intégration des données à travers les silos nécessite des pipelines d'extraction, de transformation et de chargement (ETL) complexes, qui peuvent prendre des heures (voire des jours). Ce n'est tout simplement pas assez rapide pour une prise de décision moderne. Les organisations doivent adopter des technologies qui automatisent ou éliminent l'ETL lorsque cela est possible.

AWS investit actuellement dans un avenir zéro-ETL pour que vous puissiez vous connecter rapidement et facilement à toutes vos données et les exploiter, où qu'elles se trouvent. Cela inclut :

- Permettre l'analyse en temps quasi réel de vos données transactionnelles avec des intégrations zéro-ETL entre Amazon Redshift et plusieurs bases de données AWS, telles qu'Amazon Aurora MySQL, Amazon Aurora PostgreSQL, Amazon RDS pour MySQL et [Amazon DynamoDB](#)





- Rapprocher l'AA de la source de données sans mouvement de données ni exportation de l'AA grâce à l'intégration de SageMaker dans Amazon Redshift, Aurora, Athena et Neptune.
- Utilisation de l'IA générative en vue de faciliter et d'accélérer l'intégration et le traitement des données à l'aide d'Amazon Glue
- Fournir une intégration zéro-ETL entre Amazon OpenSearch Service et Amazon DynamoDB afin que vous puissiez exécuter des requêtes de recherche en temps quasi réel. L'intégration zéro-ETL entre Amazon OpenSearch Service et S3 vous permet d'exécuter facilement des requêtes et des visualisations complexes et de relever des anomalies sans déplacer les données.
- Vous pouvez ainsi exécuter des requêtes sur des données stockées dans des bases de données opérationnelles, dans des entrepôts de données et dans des lacs de données pour fournir des informations sur plusieurs sources de données sans déplacer les données grâce à [Amazon Athena](#) et à Amazon Redshift.

Lorsque qu'un processus d'ETL est nécessaire, notamment si vous devez combiner plusieurs ensembles de données ou ajouter de la valeur grâce à des transformations, AWS continue d'investir dans AWS Glue, un service d'ETL et d'intégration de données évolutif et sans serveur qui facilite la découverte, la préparation, le déplacement et l'intégration de données provenant de sources multiples à des fins d'analyse et d'apprentissage automatique.

Analyse de l'ensemble de vos données et celles de tierces parties

Pour éliminer les silos de données, il ne suffit pas d'établir des liens avec certaines de vos sources de données. Vous devez être en mesure de vous connecter de manière transparente à toutes vos sources de données, qu'elles se trouvent dans AWS ou dans des applications tierces externes, sur site ou même dans un autre environnement infonuagique. AWS investit actuellement vers un avenir dans lequel vous pourrez intégrer automatiquement des centaines de sources de données, quel que soit l'emplacement de celles-ci.

De plus en plus souvent, les organisations exploitent également les données de tierces parties pour enrichir leurs informations en les associant à leurs propres données. [AWS Data Exchange](#) permet aux clients d'AWS d'accéder à des données de tierces parties à partir de fichiers, de tableaux et d'API provenant de plus de 300 fournisseurs de données et de plus de 3 500 produits de données, le tout à partir d'un seul endroit. Les données de tierces parties fournies par des partenaires et par des clients sont également utilisées, ce qui accroît le besoin de politiques de gouvernance exhaustives pour protéger les données. Les salles blanches (environnements protégés où plusieurs parties peuvent analyser des données combinées sans jamais exposer les jeux de données brutes) sont considérées comme une solution. [AWS Clean Rooms](#) aide les entreprises et leurs partenaires commerciaux à analyser et à collaborer en toute sécurité sur leurs jeux de données, sans partager ni révéler les données sous-jacentes.



À propos de KINTO Technologies Corporation

KINTO Technologies Corporation est un acteur majeur du secteur des plateformes de mobilité et est l'entreprise technologique responsable du développement du service KINTO en tant que société de services financiers de Toyota.

Défi

KINTO a été confronté à des défis de rendement persistants qui ont eu des répercussions sur la charge de travail de la production en raison d'une solution personnalisée pour les changements en continu à partir des bases de données principales.

« Grâce à l'intégration zéro-ETL d'Aurora MySQL avec Amazon Redshift, nous disposons en tout temps d'un accès en temps quasi réel aux données dans Amazon Redshift, ce qui élimine les heures passées par les développeurs à gérer manuellement les pipelines de données pour les opérations ETL ou à gérer les effets sur le rendement de nos charges de travail. Il en résulte un fardeau opérationnel allégé. »

Hitoshi Kageyama,
vice-président principal de KINTO Technologies Corporation

La solution AWS

KINTO fait appel à l'intégration zéro-ETL d'Amazon Aurora avec Amazon Redshift pour obtenir un pipeline de données plus résilient. KINTO peut désormais appliquer les fonctions analytiques avancées d'Amazon Redshift à ses données opérationnelles en temps quasi réel.

Résultat

Grâce à l'intégration zéro-ETL d'Aurora MySQL avec Amazon Redshift, KINTO dispose de données en temps quasi réel, ce qui élimine les heures passées par les développeurs à gérer manuellement les pipelines de données pour les opérations ETL. De plus, le nombre d'heures passées à gérer les effets d'un faible rendement sur l'exécution des charges de travail a été considérablement réduit, ce qui a allégé la charge opérationnelle.

Produits AWS utilisés

Amazon Aurora MySQL

Amazon Redshift



Comment la livraison de données permet à Goldman Sachs de travailler plus intelligemment

Goldman Sachs

Goldman Sachs a vu la demande de données sur les marchés financiers augmenter de façon exponentielle, doublant généralement d'année en année. Pour améliorer l'efficacité globale de la consommation de données tierces, l'entreprise a migré ses flux de données de marché existants vers AWS Data Exchange, un service géré qui permet de centraliser, de rationaliser et d'automatiser les processus de données dans la mesure du possible. L'amélioration de l'efficacité permet aux équipes de se concentrer davantage sur l'analyse à valeur ajoutée des données et de passer moins de temps à manipuler les données pour les rendre prêtes à l'emploi. Une fourniture de données sans friction est essentielle au succès continu de l'organisation.

« AWS Data Exchange est un élément clé de la stratégie financière infonuagique de Goldman Sachs, car ce service fluidifie l'approvisionnement en données financières auprès de fournisseurs tiers nouveaux et existants. Il nous permet également de nous concentrer sur la fourniture de nos services de base et d'analytique de données différenciée pour mieux servir nos clients. »

Marco Argenti,
codirigeants principal de l'information, Goldman Sachs

[Lire l'intégralité de l'article >>](#)



ENGIE accélère sa transition vers le zéro carbone



[ENGIE](#), une entreprise mondiale de services publics en transition vers une économie zéro carbone, a créé son lac de données Common Data Hub (plateforme de données communes) sur AWS. ENGIE a collaboré avec les [services professionnels AWS](#) pendant la conception et la mise en œuvre de la solution et a formé une équipe de service interne pour superviser la plateforme. ENGIE a éliminé ses silos de données à travers plus d'un millier de projets avec le Common Data Hub, ce qui a permis à chaque service de bénéficier d'un accès équitable à un cadre de données commun.

« Nous étions convaincus qu'AWS était une bonne solution pour plusieurs raisons, notamment pour le modèle de coûts et plus particulièrement pour le stockage de données. »

Gregory Wolowiec,
responsable de l'équipe technique chez ENGIE Data Programs

[Lire l'intégralité de l'article >>](#)



GOUVERNÉE

Libérez vos équipes pour agir plus rapidement grâce à une stratégie gouvernée d'accès aux données

Au-delà de l'approche complète et intégrée, il est tout aussi important de veiller à ce que tous les utilisateurs de vos données, qu'il s'agisse de personnes, d'applications, de moteurs ou de modèles d'AA ou d'IA, puissent accéder aux données où et quand ils en ont besoin, avec le niveau de contrôle adéquat. Avec la bonne stratégie de gouvernance des données en place, vous pouvez agir plus rapidement et assurer un accès aux données dont les utilisateurs ont besoin, quand ils en ont besoin.

À mesure qu'un nombre croissant de données migre vers le nuage et que les nouveaux modèles d'IA et d'AA consomment des quantités considérables de données, il appartient à l'entreprise de faire évoluer ses modèles de gouvernance des données au même rythme. Les chefs de file des TI et les dirigeants d'entreprise ont besoin de politiques actualisées pour protéger les données alors qu'elles se déplacent entre différents référentiels, mais aussi pour s'adapter à l'évolution des réglementations en matière de confidentialité et de sécurité des données concernant l'endroit où elles peuvent être stockées.

Simplification des autorisations d'accès aux données

La mise en œuvre d'une stratégie de gouvernance des données réussie présente un ensemble unique de défis. Fournir aux consommateurs internes ou externes leurs données, tout en assurant un niveau d'accès adéquat à des ensembles de données particuliers, s'avère une tâche complexe et chronophage. Les équipes ont souvent recours à des techniques lourdes, telles que des scripts manuels ou un examen de grappes de données individuelles, pour déterminer quels consommateurs ont accès à quelles données.

Le travail manuel peut également entraîner des problèmes coûteux en termes de qualité des données au sein des équipes et des services. Sans outils de gouvernance centralisés, les données sont enfermées dans des silos, ce qui signifie que vous ne serez pas en mesure d'accéder à toutes les données dont vous pourriez avoir besoin et de les analyser pour résoudre des problèmes ou identifier de vastes domaines d'opportunités.

« La clé d'une bonne gouvernance consiste à définir l'accès pour se libérer de cette méthode. Autrement dit, il faut créer des processus d'exception et non se limiter à une approche selon laquelle vous devez vous adresser à une personne de l'organisation centrale chaque fois que vous avez besoin de données⁶. »

Rahul Pathak,
vice-président des moteurs de bases
de données relationnelles chez AWS

Portefeuille de gouvernance AWS

- Cataloguer, découvrir, partager et gouverner les données pour permettre à tous les membres de votre organisation d'interagir avec ces dernières à l'aide d'Amazon DataZone.
- Gouverner et vérifier l'ensemble du cycle de développement de l'IA et de l'AA à l'aide des fonctionnalités d'Amazon SageMaker.
- Cataloguer les données dans l'ensemble des lacs de données, des entrepôts de données et des bases de données à l'aide d'AWS Glue
- Gouverner et vérifier les interactions avec les données dans votre lac de données sur Amazon S3 ou gouverner le partage des données dans Amazon Redshift à l'aide d'AWS Lake Formation.



Développement d'une stratégie de gouvernance des données

[Un rapport d'informations d'AWS et de MIT](#) montre que la gouvernance des données est la principale priorité des directeurs des données (DD), plus de 65 % d'entre eux indiquant que « l'établissement d'une gouvernance des données claire et efficace » est leur principale responsabilité. Les DD consacrent également beaucoup de temps à la gouvernance : plus de 63 % d'entre eux déclarent que les initiatives de gouvernance des données constituent un domaine d'action prioritaire⁷.

Sans une approche de gouvernance des données soutenant l'innovation, les organisations auront du mal à devenir axées sur les données et, en fin de compte, à rester compétitives. Après tout, plus les travailleurs passent de temps à se débattre avec les données, moins ils s'en servent pour innover. Pour réussir, une stratégie de gouvernance des données doit s'harmoniser aux initiatives commerciales financées. En effet, une stratégie de gouvernance des données autonome sans intégration aux initiatives commerciales financées risque d'échouer.

AWS investit tout au long du parcours lié aux données pour permettre une gouvernance des données de bout en bout et à moindre effort. Grâce à la gouvernance des données sur AWS, les entreprises contrôlent l'emplacement de leurs données, les personnes qui y ont accès et ce qu'elles peuvent en faire à chaque étape du flux de données. AWS propose un portefeuille de services qui permettent aux organisations de comprendre, de conserver et de protéger les données.

Par exemple, avec Amazon DataZone, les administrateurs et les directeurs des données peuvent gérer et régir l'accès aux données, tandis que les ingénieurs de données, les scientifiques des données, les gestionnaires de produits, les analystes et les autres utilisateurs professionnels peuvent découvrir et utiliser ces données ainsi que collaborer avec ces dernières afin d'obtenir des données utiles pour l'entreprise. Grâce à la nouvelle capacité d'IA générative d'Amazon DataZone (recommandations d'IA pour les descriptions), vous pouvez générer automatiquement des descriptions et des contextes professionnels pour vos sources de données, notamment en mettant en évidence les colonnes essentielles, en suggérant des cas d'utilisation analytique pertinents et même en soulignant les risques potentiels tels que les données d'identification personnelle (PII). Pour vos modèles d'IA et d'AA, Amazon SageMaker peut vous aider à gouverner et à vérifier le cycle de développement de l'AA de bout en bout. En outre, les ingénieurs et développeurs de données peuvent cataloguer les données dans l'ensemble des lacs de données, des entrepôts de données et des bases de données à l'aide d'AWS Glue, qui comprend des règles de qualité des données visant à vérifier l'actualisation, l'exactitude et l'intégrité des données. Pour conclure, grâce à AWS Lake Formation, vous pouvez gouverner et vérifier les actions entreprises avec les données de votre lac de données sur Amazon S3 et le partage des données sur Amazon Redshift.

⁷ « [Perspectives des DD 2024 : données et IA générative](#) » Rapport de perspectives d'AWS et de MIT, 2024

Utiliser l'IA et l'AA de manière responsable

L'IA et l'AA font partie des technologies les plus transformatrices que nous rencontrerons au cours de notre génération pour résoudre les problèmes commerciaux et sociétaux, améliorer l'expérience des clients et stimuler l'innovation. Avec l'utilisation généralisée et l'ampleur croissante de l'IA, il est essentiel que la technologie soit créée de manière responsable.

Alors que la plupart des entreprises ont commencé leur périple vers une IA responsable, la majorité d'entre elles (95 %) ont des difficultés à mettre en opération tous les éléments clés d'une IA responsable, notamment l'équité, l'explicabilité, la robustesse, la sécurité et la protection, la transparence et la gouvernance. Veillez à ce que vos modèles d'IA et d'AA soient créés de manière responsable, avec des pratiques en matière de données qui atténuent les préjugés et protègent la confidentialité des données grâce à vos ressources AWS.

- Tirez parti de capacités précises telles qu'[Amazon SageMaker Clarify](#) pour relever et corriger les distorsions potentielles dans les modèles d'apprentissage automatique des entreprises.
- Favorisez la transparence au moyen d'outils tels que les Cartes de service AWS AI et les Cartes de modèles Amazon SageMaker.
- Former vos équipes à l'équité et aux préjugés avec un cours de la [Machine Learning University](#)
- Donnez aux scientifiques des données des contrôles de gouvernance afin qu'ils puissent obtenir une visibilité de bout en bout sur les modèles d'AA, y compris l'entraînement, l'historique des versions et les performances des modèles, le tout grâce à SageMaker.
- Développez des applications d'IA générative de manière responsable avec les modèles de fondations Amazon Titan, capables de détecter et de supprimer les contenus préjudiciables dans les données, de rejeter les contenus inappropriés dans les entrées de l'utilisateur et de filtrer les sorties de modèles qui contiennent du contenu inapproprié, comme des discours haineux, des injures ou de la violence. Titan Image Generator dispose de protections intégrées contre les biais et inclut des filigranes invisibles sur toutes les images afin de les identifier comme étant générées par l'IA.
- Utilisez Guardrails pour Amazon Bedrock afin de mettre en œuvre des mesures de protection pour vos cas d'utilisation et vos politiques d'IA responsable.

Conseil en matière de base de données

D'ici 2025, Gartner estime que 80 % des organisations cherchant à développer leur activité numérique échoueront parce qu'elles n'adoptent pas une approche moderne de la gouvernance des données et de l'analyse.⁸ Pour réussir, vous devez veiller à ce que vos données soient sur la bonne voie, qu'elles soient accessibles à ceux qui en ont besoin et qu'une utilisation adéquate en est faite. Téléchargez le guide [Cours de maître sur la gouvernance des données](#) pour obtenir des conseils étape par étape sur la gouvernance moderne des données, qu'il s'agisse de l'étape de la planification, de l'organisation, de la création ou du déploiement.

En simplifiant la gouvernance des données, OneFootball a vu l'utilisation de sa plateforme d'analytique augmenter de 40 %

1/ ONEFOOTBALL

OneFootball s'est rapidement développée pour devenir l'une des plateformes de médias numériques les plus populaires au monde pour les amateurs de soccer. Pour exploiter davantage les données au profit de l'entreprise et des 70 millions de fans du « beau jeu », OneFootball a créé une solution plus souple sur AWS en quelques jours seulement. Depuis l'intégration des données à partir de ses bases de données dorsales inefficaces dans son lac de données en nuage, OneFootball a radicalement simplifié l'ingestion des données et a éliminé complètement les applications ETL traditionnelles. Un beau jeu, en effet.

Grâce à AWS Lake Formation, OneFootball a simplifié la gestion et la gouvernance des données de sa sécurité à l'échelle, en améliorant l'accessibilité des données dans toute l'entreprise. En fin de compte, cela a permis à l'organisation de constater une augmentation du nombre d'utilisateurs d'analyses actifs par semaine et d'autres indicateurs de performance clés, ce qui a contribué à une hausse de 40 % de l'utilisation de la plateforme d'analyses.

« Désormais, nous pouvons passer plus de temps à comprendre les problèmes opérationnels plutôt qu'à gérer différents types d'extractions de bases de données. Voir de plus en plus de personnes utiliser l'analytique au quotidien au sein de l'entreprise est une grande réussite. Une intégration fiable de toutes les sources de données pertinentes était une condition préalable. »

Stephan Durry,
responsable des données et des informations, OneFootball

[Lire la suite >](#)



Pinterest donne la priorité aux clients grâce à la gouvernance des données

Pinterest

Afin que ses données croissantes ne dépassent pas les contrôles existants, Pinterest a mis sur pied un système de contrôle d'accès précis (FGAC) évolutif et automatisé à l'aide d'Amazon S3. Ce système contrôle l'accès aux données et se fonde sur de multiples critères offrant des options telles que le contrôle d'accès basé sur les rôles et la sécurité pour les jeux de données à l'échelle du pétaoctet. L'entreprise a également permis aux créateurs et aux entreprises présentes sur la plateforme de se définir comme membres d'un groupe sous-représenté, tout en garantissant que les données sensibles ne seraient pas utilisées à d'autres fins (p. ex. la publicité).

[Lire l'intégralité de l'article >>](#)

« Les répercussions des efforts de gouvernance de Pinterest sur les clients incluent l'utilisation de données d'identification d'une « manière très contrôlée » afin de soutenir les entreprises appartenant à des personnes noires à l'occasion du Juneteenth. Les créateurs peuvent également ajouter des badges à leurs profils (ce qui permet au contenu d'apparaître dans des espaces thématiques sur Pinterest) pour montrer que les entreprises appartiennent à une personne qui s'identifie à un groupe sous-représenté. »

David Chaiken,
architecte en chef, Pinterest



Vers une sécurité plus stratégique

AWS donne la priorité à la sécurité depuis le premier jour, avec une infrastructure protégée en permanence, hautement performante, résiliente et efficace pour vos charges de travail et vos applications. Les experts en sécurité de classe mondiale qui surveillent l'infrastructure AWS élaborent et maintiennent également notre large sélection de services de sécurité innovants, qui peuvent vous aider à simplifier la complexité de vos propres exigences en matière de sécurité et d'exigences.

Les services et les solutions de sécurité AWS peuvent offrir un éventail d'avantages importants :

- **Accéder plus rapidement aux informations** : fournissez le bon niveau d'accès à vos ressources à tout moment tout en maintenant la confiance dans la protection de vos données. Les [services de sécurité AWS](#) sont conçus dans un souci de performances. Leur objectif est de vous offrir une protection maximale et une gouvernance des données qui ne vous ralentit pas.
- **Réduire les temps d'arrêt** : une sécurité infonuagique plus stricte et plus moderne permet à votre entreprise de rester active (ainsi, vous n'avez pas besoin d'arrêter d'analyser des données pour exécuter un processus de sécurité distinct). Cette sécurité peut être intégrée à chaque étape du processus.
- **Respecter votre budget** : AWS garantit une sécurité économique et s'adapte à l'évolution de vos risques et de vos exigences en matière de sécurité, protégeant ainsi les investissements de votre organisation et son engagement en faveur des initiatives liées aux données.
- **Garder le cap** : qu'il s'agisse d'infrastructure ou de services, AWS est une solution sûre, la sécurité étant prise en compte à chaque étape du processus. Vous pouvez ainsi consacrer davantage de temps à transformer des données en de meilleures décisions qui stimulent les résultats opérationnels et moins de temps à vous soucier de la sécurité et de la gouvernance.



Une histoire de fiabilité et de sécurité inégalées

- **Amazon S3** : stockez et récupérez n'importe quelle quantité de données avec la meilleure sécurité possible
- **AWS Lake Formation** : créez un lac de données sécurisé en quelques jours avec un contrôle d'accès précis
- **Régions multi-AZ** : assurez des basculements fluides si une zone de disponibilité (AZ) est perturbée



CONCLUSION

La prochaine vague d'innovation viendra des données et de l'IA

Des dirigeants se sont engagés dans une quête visant à faire des données un atout stratégique pour alimenter l'innovation grâce aux données et à l'IA. L'innovation prend de l'ampleur partout. Qu'il s'agisse d'AA, d'IA ou d'IA générative, le succès dépend de données pertinentes et de haute qualité, c'est pourquoi les dirigeants doivent faire preuve de ténacité afin de mettre sur pied une base de données solide.

Créer la base de données qui convient le mieux à votre organisation est possible, quels que soient sa taille, son emplacement ou ses besoins commerciaux. AWS fournit l'ensemble de services le plus complet sur le marché pour toutes les charges de travail, tous les types de données et tous les résultats souhaités.

Découvrez pourquoi AWS est le meilleur endroit pour exploiter la valeur de nos données et transformer des informations en temps réel en innovations significatives. Découvrez également comment nous pouvons aider vos équipes en matière d'infrastructure, d'outils et de soutien à la mise en œuvre via le premier réseau mondial de services professionnels et de partenaires. En matière de données et d'IA, les clients d'AWS sont plus expérimentés.

Découvrez la façon de réinventer votre organisation pour qu'elle devienne axée sur les données ›

Découvrez les principaux cas d'usage des données afin de maximiser la valeur opérationnelle

Comment pouvez-vous tirer parti de vos données pour améliorer l'expérience client, optimiser et réinventer les chaînes d'approvisionnement, améliorer la prise de décision, créer des applications modernes et plus encore? Explorez les principales possibilités d'amélioration pour votre organisation dans le livre numérique « Huit domaines essentiels de solutions basées sur les données pour les dirigeants : Optimiser la valeur opérationnelle grâce aux données ».