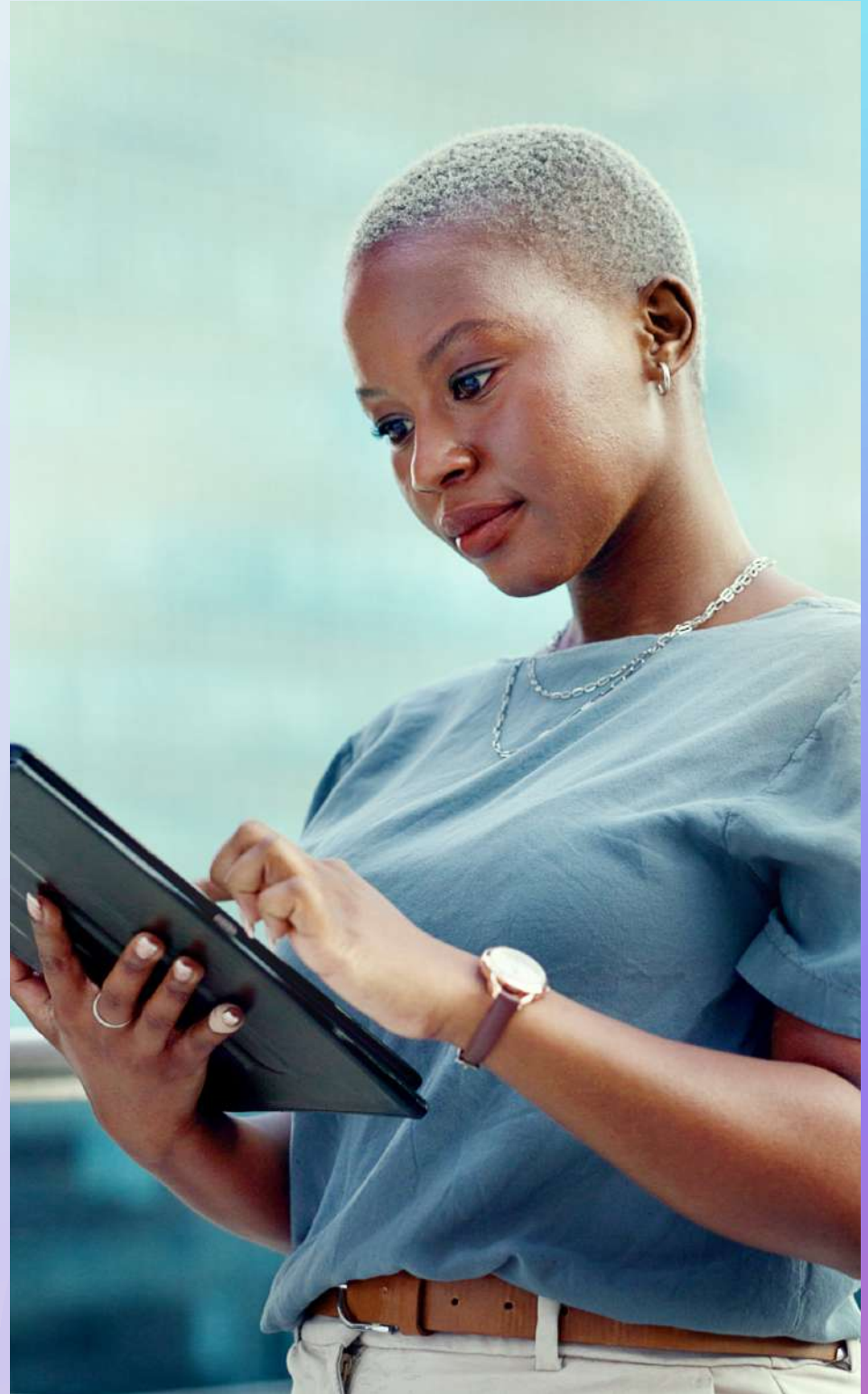




# Mise à l'échelle intelligente : tirer parti d'AWS pour une croissance rentable

Guide à l'intention des start-ups pour optimiser les coûts grâce aux données et à l'IA



# Table des matières

<b>Introduction</b>	<b>3</b>
ALTR	3
BioNTech	4

## Stratégie 1

<b>créer un modèle de données rentable dans le Cloud AWS</b>	<b>5</b>
Petal	6
Sayurbox	7
ClickHouse	8
Leadinfo	9
Woolworths	10
YOUGotaGift	11
BlocPower	12
Perplexity	13
Bazaarvoice	14

## Stratégie 2

<b>utiliser les informations et l'intelligence pour réduire les coûts</b>	<b>15</b>
Segment	16
Lenme	17

<b>Conclusion</b>	<b>18</b>
-------------------	-----------

## INTRODUCTION

# Réduire les coûts et accélérer l'innovation

Les données d'une start-up ont une valeur inestimable. Elles peuvent être utilisées pour stimuler l'innovation, découvrir de nouveaux rapports coûts-efficacité et gagner un avantage concurrentiel. Cependant, le stockage, le traitement et l'analyse de grandes quantités de données requièrent des ressources importantes, notamment en matériel, en logiciels et en personnel.

Les créateurs d'entreprises dont les ressources sont limitées recherchent constamment l'équilibre subtil qui permet d'optimiser les coûts liés à la mise en œuvre de leurs données sans entraver l'innovation et la croissance. À tous les stades, les start-ups sont confrontées au défi que représentent les volumes croissants de données et les coûts associés à leur gestion. Dans le cadre d'une étude menée par Deloitte en 2022, 45 % des directeurs technologiques ont déclaré que la collecte et la protection de volumes toujours croissants de données étaient leur principale préoccupation.

Aujourd'hui, l'émergence de l'intelligence artificielle (IA) générative ajoute un autre niveau de complexité et de coût dont les créateurs d'entreprises doivent tenir compte. Elle offre des capacités allant des chatbots à la génération de code logiciel en passant par la recherche et le développement, et à mesure qu'elle évolue, son potentiel semble illimité. Alors que *Forbes* prévoit que les coûts des centres de données vont « **augmenter pour atteindre plus de 76 milliards d'USD d'ici 2028** », en raison de l'IA générative, les start-ups qui peuvent gérer ces coûts de manière rentable seront bien placées pour réaliser les promesses de celle-ci et réussir dans un environnement de plus en plus concurrentiel.

Bien qu'il n'existe pas d'approche unique pour optimiser les coûts des données, les clients d'Amazon Web Services (AWS) ont réduit leurs coûts tout en tirant le meilleur parti de leurs données grâce à ces deux stratégies :

- 1 Créer un modèle de données rentable dans le cloud
- 2 Utiliser des informations et les prévisions tirées de leurs données pour trouver des possibilités d'optimisation des coûts dans tous les secteurs de leur entreprise



À l'aide de technologies sans serveur et d'AWS, **ALTR** a créé un service extrêmement évolutif sans avoir à mettre à l'échelle les coûts.

« Notre activité a doublé au cours de la dernière année et nos dépenses de production AWS n'ont augmenté que de 2,4 % », a déclaré Chris Struttman, cofondateur et directeur de la technologie chez ALTR.



# Optimisation des coûts : par où commencer ?

Les investisseurs attendent beaucoup des dépenses technologiques pour assurer un retour sur investissement important. Les créateurs d'entreprises reconnaissent la valeur de la création d'un modèle de données solide à même de répondre à leurs besoins actuels et futurs. Construire un modèle qui exploite la valeur des données n'est pas un processus linéaire. Pour exploiter efficacement les données, il ne suffit pas d'un lac de données, d'un entrepôt de données, d'un outil de veille stratégique ou d'un modèle d'IA générative. De la collecte initiale à la prise de décision en passant par l'analyse, il est essentiel de prendre le temps de créer un modèle de données de bout en bout dans le cloud pour s'assurer que les données sont gérées efficacement tout au long de leur cycle de vie. Un modèle de données solide peut non seulement réduire les coûts globaux de gestion des données et de l'infrastructure associée, mais aussi vous aider à utiliser les données pour trouver des solutions efficaces afin de réduire davantage les coûts.

La première étape consiste à évaluer votre infrastructure de données actuelle et à déterminer les domaines dans lesquels il est possible de réduire les coûts. Cette étape peut notamment impliquer le regroupement des sources de données, l'optimisation du stockage des données et la priorisation des solutions qui offrent le meilleur rapport performances-prix. Cela peut également impliquer l'adoption de logiciels open source ou de solutions basées sur le cloud pour réduire les coûts d'infrastructure sur site ou mettre en œuvre des politiques de gouvernance des données davantage axées sur l'efficacité.

Une fois le modèle de données mis en place, vous pouvez en tirer des informations qui vous permettront d'examiner les pratiques actuelles et d'identifier les domaines dans lesquels vous pouvez améliorer les processus et réaliser des économies dans l'ensemble de votre entreprise.

En adoptant une approche globale et stratégique de la gestion des données, vous pouvez créer un modèle de données rentable qui offre une valeur mesurable et soutient la croissance de votre start-up. Lisez ce guide pour en savoir plus.

## BIONTECH

**BioNTech** a utilisé AWS pour accélérer le traitement des données pour les flux de travail protéomiques de 500 fois tout en réduisant considérablement le coût des instances de calcul. « Sur AWS, nos scientifiques génèrent et partagent un nombre exponentiel de données dans le but de trouver des thérapies efficaces, ciblées et personnalisées pour les patients », a expliqué Michael McCarthy, architecte de solutions chez BioNTech.

« C'est vraiment l'imagination qui vous limite, et je n'ai encore jamais trouvé de solution que je ne puisse créer dans AWS. » En augmentant la vitesse et la capacité de mise à l'échelle à davantage de flux de travail avec AWS, BioNTech a considérablement amélioré l'efficacité tout en réduisant les coûts. « Nous pourrions refaire tout le travail des sept dernières années en 60 heures pour une fraction du prix », a indiqué Akhil Chaudhary, ingénieur de données chez BioNTech.





# Créez un modèle de données rentable dans le Cloud AWS

## Modernisez votre infrastructure de données

Le passage au Cloud AWS peut réduire les coûts d'infrastructure de données en supprimant le besoin de personnel de support dédié et la nécessité de gérer le matériel et les logiciels sur site. Les services Cloud AWS permettent un accès immédiat à des ressources informatiques de pointe telles que le calcul et le stockage, sans coût initial ni frais d'exploitation.

Et avec la facturation à la seconde, vous ne payez que les ressources que vous consommez. Les services de données gérés dans le cloud éliminent les contraintes opérationnelles, vous laissant ainsi libre de vous concentrer sur votre cœur de métier au lieu de devoir investir du temps, de l'argent et des ressources dans la gestion de l'infrastructure informatique.



## Choisir la base de données adaptée à vos applications

Les bases de données constituent la base des applications, mais leur gestion peut nécessiter des ressources importantes. Les services de base de données entièrement gérés d'AWS déchargent votre équipe des tâches chronophages comme la mise en service, l'application de correctifs et les sauvegardes. Ils assurent également une surveillance continue, un stockage avec mécanisme d'autorégénération et une mise à l'échelle automatisée, permettant ainsi aux développeurs de se concentrer sur la création de la prochaine application magnifique.

AWS offre des bases de données conçues spécifiquement pour votre cas d'utilisation à grande échelle, que ce soit pour des applications Web à fort trafic, la mise en cache et les applications géospatiales, la gestion de contenu, les applications industrielles à grande échelle, la détection de fraudes, les applications Internet des objets (IoT), les systèmes d'enregistrement ou encore les applications d'IA générative, le tout au meilleur rapport qualité-prix et avec des performances optimales.

**Amazon Aurora**, par exemple, offre les performances et la disponibilité d'une base de données relationnelle de qualité commerciale pour un coût dix fois moins élevé, ce qui constitue un avantage économique évident. Selon une **étude IDC** commandée par AWS, une entreprise ayant un chiffre d'affaires annuel moyen de 1 milliard de dollars pourrait réaliser une valeur commerciale annuelle moyenne de 9 millions de dollars en :

- accroissant la productivité informatique et l'efficacité du personnel informatique, des développeurs et des équipes en charge des bases de données ;
- améliorant les délais de mise sur le marché, la fidélisation des clients et la croissance technologique ;
- réduisant les risques liés aux temps d'arrêt imprévus ;
- renforçant les performances des bases de données.

De plus, avec une réduction de 32 % des dépenses liées aux bases de données, l'entreprise peut s'attendre à un retour sur investissement en 13 mois, ainsi qu'à un retour sur investissement de 439 % sur trois ans.

## Petal

La start-up de cartes de crédit **Petal** est une pionnière dans l'élargissement de l'accès au crédit à la consommation. L'approche transformatrice de l'entreprise permet aux consommateurs d'utiliser leurs antécédents bancaires pour bénéficier des cartes de crédit Petal, au lieu de se fier uniquement aux antécédents de crédit pour démontrer leur solvabilité. L'infrastructure de Petal, axée sur les données, fonctionne sur les services AWS, de l'infrastructure backend au paysage des applications front-office. Lorsque les clients demandent une carte de crédit, ils sont dirigés vers l'application Web pour remplir les informations personnelles requises. Petal héberge l'interface utilisateur de ces premières pages Web à l'aide d'**Amazon Simple Storage Service** (Amazon S3), un service de stockage d'objets offrant une capacité de mise à l'échelle, une disponibilité des données, une sécurité et des performances de pointe. « Nous voulons que la première interaction soit très résiliente », a déclaré John Wang, vice-président de l'ingénierie chez Petal. « En utilisant Amazon S3, nous pouvons maintenir une haute disponibilité de notre page de dépôt de candidature pour nos millions de candidats. »



# sayurbox

Fondée en 2016, **Sayurbox** s'est donnée comme mission d'améliorer la chaîne d'approvisionnement agricole d'Indonésie. En 2019, l'entreprise a migré vers AWS à partir d'un autre fournisseur de cloud afin de réduire les frais généraux et de tirer parti d'offres gérées telles qu'**Amazon DocumentDB** et **Amazon Elasticsearch Service**. Les clients peuvent désormais commander sur le site Web et sur l'application de Sayurbox ou même sur WhatsApp. Grâce à des services gérés évolutifs sur AWS, Sayurbox maintient des coûts bas, ce qui permet aux clients d'économiser environ 30 % par rapport aux achats effectués dans les supermarchés et aux agriculteurs de gagner jusqu'à 20 % de plus sur leurs récoltes. « Avec AWS, il est plus facile d'augmenter nos ressources et de suivre les dépenses au fur et à mesure », a indiqué Nilesch Kumar, vice-président de l'ingénierie chez Sayurbox.

L'intérêt croissant pour les applications d'IA générative place le choix des bases de données au premier plan. Les applications d'IA générative ont besoin de bases de données capables de stocker, d'indexer, d'extraire et de rechercher des intégrations vectorielles, qui sont des représentations numériques de données non structurées telles que du texte, des images, du son ou d'autres éléments.

Les start-ups privilégient l'utilisation de bases de données qu'elles maîtrisent déjà pour stocker des intégrations vectorielles, car elles ont déjà fait leurs preuves en production et répondent aux exigences de stabilité, de disponibilité, de stockage, de calcul et de rapport performances-prix de leur activité. Et lorsque vos vecteurs et vos données commerciales sont stockés au même endroit, vos applications s'exécutent plus rapidement, car il n'est pas nécessaire de synchroniser ou de déplacer les données. AWS offre des capacités vectorielles dans ses magasins de données les plus populaires, notamment Amazon Aurora, **Amazon Relational Database Service (Amazon RDS)**, **Amazon OpenSearch Service**, **Amazon OpenSearch sans serveur**, **Amazon Neptune** et **Amazon DocumentDB (compatible avec MongoDB)**.



## Explorer les lacs de données

De nos jours, les start-ups recueillent des données structurées et non structurées provenant de diverses sources (transactions client en ligne, capteurs sur les appareils IoT, commentaires des clients, et ainsi de suite), et ce référentiel en constante expansion s'étend à de nombreux services et systèmes sur site. Les start-ups peuvent regrouper ces données disparates dans un lac de données afin de les utiliser plus efficacement à des fins d'analytique avancée et d'IA.

AWS aide des centaines de milliers de clients à établir des bases solides et rentables pour les lacs de données grâce à des services tels qu'Amazon S3, AWS Glue et **AWS Lake Formation** depuis des années.

Amazon S3 offre plusieurs façons d'optimiser les coûts de stockage des volumes croissants de données. **Amazon S3 Intelligent-Tiering** automatise la gestion du cycle de vie des données, en migrant les données vers le niveau le plus rentable lorsque les modèles d'accès changent. Par rapport à Amazon S3 Standard, S3 Intelligent-Tiering a permis aux clients d'économiser 2 milliards d'USD en coûts de stockage depuis son lancement en 2018. L'utilisation de classes de stockage telles qu'**Amazon S3 Express One Zone** pour vos données les plus fréquemment consultées permet de multiplier par dix la vitesse d'accès aux données et de réduire les coûts de requête de 50 %.

Amazon S3 Intelligent-Tiering a permis aux clients d'économiser

**2 MILLIARDS** USD  
d'  
**en coûts de stockage** par rapport à Amazon S3  
depuis 2018.

## ClickHouse

**ClickHouse** est la base de données open source la plus rapide et la plus économe en ressources pour les applications et l'analytique en temps réel. ClickHouse Cloud est une base de données en tant que service cloud natif qui s'exécute sur AWS.

« ClickHouse Cloud est utilisé pour l'analytique en temps réel, de l'observabilité aux données de trading sur le marché. Cela exige les meilleures performances de stockage à grande échelle. C'est pourquoi nous utilisons la classe de stockage Amazon S3 Express One Zone pour le cache chaud de nos données dans les services alimentant les applications analytiques en temps réel les plus exigeantes. S3 Express One Zone offre un temps de latence des requêtes constant de quelques millisecondes sur des centaines de milliers de transactions par seconde, ce qui permet à ClickHouse Cloud de traiter les données encore plus rapidement qu'auparavant. Avec S3 Express One Zone, nous offrons des performances exceptionnelles à nos utilisateurs en accélérant les requêtes à partir de n'importe quelle source de données et en générant des rapports analytiques à l'aide de requêtes SQL en temps réel. En ajoutant S3 Express One Zone à notre architecture, nous avons amélioré les performances des requêtes jusqu'à 283 % et réduit notre coût total de possession global de 65 %. »

Tanya Bragin, vice-présidente des produits, ClickHouse



## Extraire rapidement et à moindre coût des informations basées sur les données

Extrayez rapidement des informations de vos données grâce à des outils spécialement conçus et optimisés pour offrir les meilleures performances et capacités de mise à l'échelle au meilleur coût afin de répondre à vos besoins. Les services analytiques d'AWS tels qu'**Amazon Redshift** vous permettent d'ingérer vos données, de les combiner et d'exécuter en toute sécurité l'analytique historique, en temps réel et prédictive sur ces données en quelques clics seulement. De plus, les recherches renvoient les résultats de requête ultra-rapidement. Grâce à des performances optimisées pour les charges de travail réelles des clients, Amazon Redshift offre un rapport performances-prix jusqu'à sept fois supérieur pour les charges de travail de requêtes à haute simultanéité et faible latence, et un rapport performances-prix jusqu'à six à sept fois supérieur à celui de tout autre entrepôt de données. En fonction de la charge de travail, les nouvelles capacités de mise à l'échelle et les optimisations pilotées par l'IA pour **Amazon Redshift sans serveur** permettent au service d'apprendre de vos modèles et de s'adapter de manière proactive sur plusieurs aspects, notamment les utilisateurs simultanés, la variabilité des données et la complexité des requêtes. Tout cela en tenant compte de vos objectifs en matière de rapport performances-prix afin de vous permettre d'optimiser les coûts et les performances.

Pour des applications de Big Data plus rapides et des analyses de données à l'échelle du pétaoctet à moins de la moitié du coût des solutions sur site, **Amazon EMR**, un service de big data géré compatible avec Apache Hadoop et Apache Spark, offre diverses options tarifaires pour vous aider à optimiser les coûts d'analyse en fonction de vos modèles d'utilisation uniques.

## Automatiser les processus coûteux et chronophages

La gestion de l'infrastructure de données et la valorisation des données demandent beaucoup de temps et de personnel. L'adoption de technologies qui aident à automatiser les processus de gestion des données répétitifs et chronophages peut permettre d'économiser des ressources qui pourront être affectées à des projets plus innovants, contribuant ainsi à accélérer la création de valeur.



**Leadinfo** est une entreprise à grande échelle basée aux Pays-Bas qui aide les clients B2B (des organisations sportives aux agences de marketing numérique en passant par les entreprises de l'industrie lourde dans le domaine de l'acier et de l'emballage) à cibler les clients potentiels en analysant les informations sur les visiteurs de leur site Web. L'entreprise a vu son activité croître rapidement après que la pandémie de COVID-19 a mis un terme à de nombreux appels de vente en face à face. En migrant ses données clients de plusieurs bases de données sur AWS vers Amazon Aurora, Leadinfo a réduit ses coûts informatiques mensuels de 30 % tout en améliorant l'efficacité de ses développeurs de 50 %. Cela lui a permis de déployer rapidement de nouvelles fonctionnalités pour ses clients et de jeter les bases d'une croissance continue de l'entreprise.



## Obtenir plus d'informations avec moins d'efforts grâce à l'approche zéro ETL

Les informations les plus pertinentes basées sur les données proviennent de la connexion entre toutes vos sources de données, qu'il s'agisse de départements, de services, d'outils sur site ou d'applications tierces. Traditionnellement, la connexion des données implique des pipelines d'extraction, de transformation et de chargement (ETL) complexes, qui prennent souvent des heures ou des jours, ce qui représente un temps précieux susceptible de bloquer ou de retarder les décisions ou les actions. Afin d'alléger les coûts et les ressources importants liés à l'ETL, AWS propose plusieurs intégrations zéro ETL, ce qui vous permet de vous connecter rapidement et facilement à toutes vos données et de les exploiter.

Cela inclut des **intégrations zéro ETL** entre Amazon Redshift et Amazon Aurora PostgreSQL, **Amazon Relational Database Service (Amazon RDS) for MySQL** et **Amazon DynamoDB** afin de vous permettre de tirer plus facilement parti de l'analytique en temps quasi réel. En pratique, cela signifie que quelques secondes après l'écriture des données dans une base de données, vous pouvez utiliser Amazon Redshift pour effectuer des analyses et du machine learning (ML) sur des pétaoctets de données. Outre Amazon Redshift, AWS a également étendu sa prise en charge de l'approche zéro ETL à Amazon OpenSearch Service, utilisé par des dizaines de milliers de clients pour la recherche, la surveillance et l'analyse en temps réel des données commerciales et opérationnelles, et des intégrations zéro ETL avec DynamoDB et Amazon S3.

Dans de nombreux cas, AWS peut contribuer à éliminer complètement l'ETL, mais les start-ups peuvent tout de même avoir besoin de transformations telles que le nettoyage, la déduplication et la possibilité de combiner des jeux de données à des fins d'analyse et de ML. Dans de tels cas, AWS Glue permet une transformation rapide et évolutive des données. AWS Glue exploite l'IA générative pour vous permettre de créer plus facilement des tâches d'intégration de données, en réduisant le temps et les ressources nécessaires. Cela englobe la fourniture de suggestions de code et de corrections de syntaxe en temps réel, voire même la création de tâches de pipeline de données en langage naturel.



Woolworths, pionnier de la vente au détail, continue de mettre l'accent sur l'innovation, la valeur et la durabilité. L'entreprise utilise l'intégration zéro ETL d'Amazon Aurora avec Amazon Redshift pour obtenir des informations sur les promotions et les événements urgents. L'approche zéro ETL signifie que les analyses et les campagnes dont le développement prenait auparavant deux mois peuvent être réalisées en une journée. Le résultat ? Gains de temps, réduction de la charge technique, diminution des points de défaillance associés à la gestion des pipelines, et économies de coûts importantes.

Pour en savoir plus, consultez [Intégrations de données avec AWS](#).





« Amazon Ads utilise Amazon Q Developer au quotidien pour être plus productif.

Les développeurs Python de l'équipe ont découvert que l'utilisation d'Amazon Q Developer a rendu l'écriture en Java presque aussi fluide que l'écriture en Python. Ils trouvent cela particulièrement utile dans deux cas d'utilisation : l'écriture de tests unitaires, où cela permet de réduire de moitié le temps nécessaire, et lorsqu'ils doivent résoudre des problèmes ésotériques liés à CDK, où cela leur évite d'avoir à consulter la documentation complète. Amazon Ads est capable de coder 25 à 35 % plus rapidement avec Amazon Q Developer. »

Aneesh Shukla, ingénieur en développement logiciel, Amazon Ads



**YOU GotaGift** fournit des solutions de cartes-cadeaux numériques et des devises de marque aux clients du Moyen-Orient. En 2017, lorsque l'entreprise a rencontré des problèmes de mise à l'échelle et de performance avec son fournisseur de cloud local, elle a migré vers AWS. Elle a maintenant conteneurisé son application, commencé à utiliser des microservices et déplacé sa base de données vers **Amazon Aurora sans serveur**. L'entreprise a également migré de son propre outil d'analyse et de veille stratégique vers **Amazon QuickSight**, qui fournit la veille stratégique et aide les entreprises à prendre de meilleures décisions fondées sur les données. Ashin K N, directeur de la technologie, explique que QuickSight accélère le développement et augmente la productivité. « Avant d'utiliser QuickSight, nous devions créer des tableaux de bord uniques pour chaque produit », a déclaré Ashin K N. « Désormais, ces tableaux de bord sont créés automatiquement. »

La plateforme de YOU GotaGift est plus fiable et les opérations sont rationalisées à l'aide d'AWS. « Nous avons cinq personnes qui gèrent l'ensemble de notre infrastructure cloud », a indiqué le directeur de la technologie. « Si nous n'utilisions pas AWS et cette gamme variée de services, nous aurions besoin d'une quinzaine d'ingénieurs pour créer et maintenir les services en interne. L'utilisation d'AWS nous a aidés à réduire les délais de développement et les coûts. »



## Choisir des services de données qui intègrent l'intelligence

L'utilisation de services intégrant une technologie intelligente peut également aider à éliminer les charges lourdes et à améliorer l'efficacité.

Par exemple, **Amazon Q Developer** est un outil de productivité basé sur l'IA pour l'IDE et la ligne de commande, qui aide les développeurs avec des suggestions de code allant des extraits aux fonctions complètes en temps réel.

**Amazon Q**, un nouveau type d'assistant optimisé par l'IA générative, peut être adapté à votre entreprise en utilisant vos propres données pour prendre en charge pratiquement tous les domaines de votre activité. Les utilisateurs peuvent poser des questions à Amazon Q en langage naturel afin d'obtenir des informations exploitables et utiliser celles-ci pour gérer les données et simplifier les tâches répétitives et courantes liées aux données. Par exemple, cet outil peut vous aider à créer des tableaux de bord et des récits visuels captivants à partir des données de vos tableaux de bord dans QuickSight en langage naturel.

Amazon Q dans Amazon Redshift peut également vous aider à générer des requêtes SQL en langage naturel, accélérant ainsi votre capacité à interroger des données. Grâce à l'IA générative intégrée à **Amazon DataZone**, vous pouvez profiter de l'automatisation des descriptions détaillées de votre catalogue de données, ce qui facilite grandement la recherche et l'utilisation des données pour les membres de votre start-up.



## Opter pour l'architecture sans serveur et consacrer plus de temps à l'innovation

L'architecture sans serveur, qui consiste à créer sans gérer de serveurs, vous permet de créer et d'exécuter des applications et des analyses sans nécessiter d'investissements initiaux importants dans l'infrastructure ni de coûts opérationnels continus. De plus, il s'agit d'un moyen plus rentable, flexible et efficace de gérer vos données. AWS offre des options sans serveur pour plusieurs services de données afin de vous aider à réduire le temps et les efforts nécessaires pour gérer l'infrastructure.

Amazon Aurora sans serveur permet aux clients d'économiser jusqu'à 90 % par rapport au coût de la capacité de mise en service pour les pics de charge. Tous les services d'analytique AWS, y compris Amazon Redshift et Amazon EMR, sont disponibles sans serveur, ce qui facilite l'analyse des données à n'importe quelle échelle sans avoir à configurer, mettre à l'échelle ou gérer l'infrastructure sous-jacente pour les services de données. Amazon Redshift sans serveur utilise des capacités de mise à l'échelle et des optimisations pilotées par l'IA pour apprendre de vos modèles et s'adapter de manière proactive sur plusieurs aspects, notamment les utilisateurs simultanés, la variabilité des données et la complexité des requêtes. Tout cela en tenant compte de vos objectifs en matière de rapport performances-prix afin de vous permettre d'optimiser les coûts et les performances. Cela permet de réduire les coûts d'infrastructure, d'accroître la capacité de mise à l'échelle, de réduire les frais généraux opérationnels et d'accélérer l'accès aux informations.

AWS offre désormais le plus grand nombre d'options sans serveur pour l'analytique des données dans le cloud.



Afin d'améliorer l'expérience utilisateur de BlocMaps, sa solution SaaS qui fournit aux municipalités et aux services publics des informations sur la décarbonisation des bâtiments, le leader de la technologie climatique **BlocPower** a migré ses données vers une combinaison de solutions de stockage de données basées sur le cloud, dont Amazon Redshift, un entrepôt de données dans le cloud rapide, simple et largement utilisé. BlocPower stocke les données qu'elle recueille à partir de 100 millions de profils de bâtiment dans Amazon Simple Storage Service (Amazon S3), qui offre un stockage d'objets conçu pour récupérer n'importe quelle quantité de données à partir de n'importe quel emplacement. « Notre application a été beaucoup plus performante, et notre facturation a bénéficié d'Amazon Redshift sans serveur », a déclaré Sean Davis, architecte de données chez BlocPower. Plus précisément, la start-up a pu traiter et interroger ses données en quelques minutes, soit 10 fois plus rapidement qu'avec son architecture précédente. Non seulement BlocPower a augmenté ses possibilités de revenus, mais la start-up a également optimisé ses coûts de calcul. Après avoir adopté Amazon Redshift sans serveur, BlocPower ne paie plus pour les clusters « inactifs ». « Le modèle sans serveur a été parfait pour nous », a indiqué M. Davis. « Nous payons moins cher pour nos processus et nous obtenons plus de ressources de calcul lorsque nous en avons besoin. »





**Perplexity** développe actuellement l'un des premiers moteurs de réponses conversationnelles au monde utilisant la puissance de l'IA générative pour aider les utilisateurs à trouver des connaissances pertinentes. Face au défi que représente l'optimisation de ses modèles en termes de précision et d'exactitude, Perplexity avait besoin d'une solution robuste capable de répondre à ses besoins en matière de calcul. En utilisant **Amazon SageMaker HyperPod**, Perplexity est capable de transférer des données entre différents GPU beaucoup plus rapidement, ce qui a réduit le temps d'entraînement des modèles de ML jusqu'à 40 %.

« Les bibliothèques parallèles de données et de modèles intégrées d'Amazon SageMaker HyperPod nous ont permis d'optimiser le temps d'entraînement sur les GPU et de doubler le débit d'entraînement », a déclaré Aravind Srinivas, PDG et cofondateur. « En conséquence, nos expériences d'entraînement peuvent désormais s'exécuter deux fois plus rapidement, ce qui signifie que nos développeurs peuvent itérer plus rapidement, accélérant ainsi le développement de nouvelles expériences d'IA générative pour nos clients. »

## Innover avec l'intelligence artificielle, l'IA générative et le machine learning de manière rentable

La plupart des start-ups considèrent l'IA, et désormais maintenant l'IA générative, comme des éléments essentiels de leur stratégie de données. Cependant, cela peut s'avérer coûteux, en particulier si l'entreprise ne dispose pas du financement et des ressources nécessaires pour développer et mettre à l'échelle efficacement des projets d'IA et d'IA générative. AWS fournit l'ensemble le plus complet de capacités d'analytique et d'IA pour aider les entreprises à déduire des informations de leurs données opérationnelles, afin de rationaliser les opérations et de réduire les coûts dans l'ensemble de l'entreprise.

Amazon Bedrock est un service entièrement géré qui propose un choix de modèles de fondation (FM) performants issus de grandes sociétés d'IA telles que AI21 Labs, Anthropic, Cohere, Meta, Mistral AI, Stability AI et Amazon via une API unique, ainsi qu'un large éventail de fonctionnalités nécessaires pour créer des applications d'IA générative en alliant sécurité, confidentialité et IA responsable. Amazon Bedrock vous aide à évaluer des modèles pour vos cas d'utilisation, à personnaliser les modèles avec vos propres données et à accélérer le développement de l'IA générative. Avec AWS, que vous créiez vos propres FM ou que vous personnalisiez des FM existants ou des applications d'IA générative, vous bénéficiez également de l'infrastructure et des fonctionnalités les plus performantes, évolutives et sécurisées.

**Amazon SageMaker** fournit tous les outils nécessaires pour créer, entraîner et déployer facilement des modèles de ML, y compris des FM, à grande échelle. Cet outil est essentiel pour normaliser le développement du ML. Une fois que les pratiques de ML sont normalisées, elles peuvent être facilement reproduites et appliquées à différents cas d'utilisation ou services, ce qui améliore la cohérence et renforce la collaboration entre les spécialistes des données et les autres parties prenantes.

Outre les outils permettant de créer des applications de ML et d'IA générative, pour aider les clients à améliorer leurs performances tout en réduisant les coûts, AWS a développé un portefeuille de puces personnalisées, dont **AWS Trainium**, qui réduit le coût d'entraînement des modèles de ML jusqu'à 50 % par rapport à des instances GPU similaires, et **AWS Inferentia**, qui offre des performances élevées au coût le plus bas pour les applications d'inférence du ML.



bazaarvoice

**Bazaarvoice**, un fournisseur leader d'évaluations de produits et de solutions de contenu générées par les utilisateurs, a migré ses charges de travail de ML vers Amazon SageMaker. SageMaker a permis à l'entreprise d'accélérer le déploiement des modèles de ML, de réduire les coûts et de donner à ses ingénieurs la possibilité de proposer plus rapidement de nouvelles fonctionnalités aux clients, tout en réduisant les coûts de 82 %. Et elle réinvestit ces économies pour améliorer encore son service.

## Fournir des solutions de données en libre-service aux utilisateurs professionnels

En permettant aux utilisateurs professionnels de prendre eux-mêmes des décisions fondées sur les données sans faire appel à des équipes d'ingénierie ou de science des données, vous pouvez étendre de manière rentable les décisions fondées sur les données et l'innovation à l'ensemble de votre entreprise.

Amazon QuickSight est notre outil de veille stratégique (BI) sans serveur optimisé par le ML qui permet aux analystes métier de créer, de publier et d'intégrer facilement des visualisations de données interactives et des tableaux de bord pour obtenir des informations à partir des données.

Avec **Amazon Q dans QuickSight**, les utilisateurs peuvent interroger leurs données en langage naturel sans avoir à écrire une seule ligne de code.

La tarification à la session vous permet de ne payer que lorsque vos utilisateurs accèdent aux tableaux de bord ou aux rapports, ce qui est efficace et réduit les coûts pour les déploiements avec de nombreux utilisateurs.

**Amazon SageMaker Canvas**, une interface visuelle de type pointer-cliquer, permet aux analystes métier de générer des prédictions ML précises sans avoir d'expérience préalable en ML, comme la prévision de la perte de clientèle à partir de données sur la consommation de produits et l'historique des achats, ou la prévision de la maintenance imprévue à l'aide de données historiques sur les événements et l'exploitation.



# Utiliser les informations et l'intelligence pour réduire les coûts

## Comment les clients utilisent les informations pour réduire les coûts

AWS fournit l'ensemble le plus complet de capacités d'analytique et d'IA pour aider les start-ups à tirer des informations de leurs données opérationnelles afin de générer des réductions de coûts dans l'ensemble de l'entreprise, tout en atteignant leurs objectifs commerciaux uniques.



Traite désormais

# 450 MILLIARDS

d'événements par mois à des fins de personnalisation  
sans créer son propre pipeline de ML

## Unifier les profils des clients pour accroître l'efficacité du marketing

Segment, par Twilio, est une entreprise d'infrastructure de données clients qui utilise AWS pour aider ses clients à collecter et à unifier des données sur leurs utilisateurs et à créer des recommandations personnalisées à partir de ces données avec Amazon Personalize. L'entreprise traite 450 milliards d'événements par mois en utilisant des milliers d'instances Amazon Elastic Compute Cloud (Amazon EC2) et exécute plus de 16 000 conteneurs Docker sur Amazon Elastic Container Service (Amazon ECS). Selon Calvin French-Owen, directeur de la technologie et cofondateur de Segment, de nombreux clients de l'entreprise ont besoin de personnaliser leurs produits à l'aide du ML, mais ne disposent pas de suffisamment de données d'entraînement pour élaborer les modèles de prédiction nécessaires. « Il s'agit d'une belle synergie qui leur permet de lancer très facilement Segment dès le premier jour et de commencer à collecter toutes leurs données, » a indiqué M. French-Owen. « Grâce à Amazon Personalize, ils peuvent ensuite utiliser ces données pour formuler des recommandations sans avoir à mettre en place leur propre pipeline de machine learning. »



## Utiliser l'analytique et l'intelligence artificielle pour trouver des opportunités d'économies

**Lenme**, un service par abonnement, a révolutionné le secteur du prêt en tirant parti d'AWS pour automatiser une plateforme qui résout désormais les défis de longue date liés à l'acquisition, à la vérification et à l'évaluation des emprunteurs. Lenme a relevé ce défi en utilisant les services AWS pour vérifier et qualifier les emprunteurs en seulement trois clics grâce aux fonctionnalités d'IA de l'**API de vérification de l'identité d'Amazon Rekognition**, qui aide Lenme à vérifier les clients avec une grande précision et en quelques secondes. **Amazon Rekognition** est un service d'IA entièrement géré qui offre des capacités de vision par ordinateur pré-entraînées et personnalisables pour extraire des informations à partir d'images et de vidéos. La nouvelle technologie de Lenme permet à l'entreprise de proposer des produits peu coûteux tout en s'imposant comme un leader de confiance dans le secteur des prêts financiers.

# Lenme

« Notre plateforme est désormais plus rapide et plus efficace, nous permettant de vérifier et d'authentifier les clients en trois clics et quelques secondes. Cela nous permet de fournir à nos prêteurs davantage de données et de réduire les risques de prêt jusqu'à 80 %. »

Mark Maurice, PDG de Lenme





## CONCLUSION

# Optimiser l'innovation axée sur les données tout en minimisant les coûts

Avec le bon modèle de données, vous pouvez comprendre votre start-up, vous adapter aux besoins des clients, rationaliser les processus, prendre de meilleures décisions et innover plus rapidement. Dans le paysage commercial actuel, complexe et en constante évolution, l'optimisation des coûts de vos données est essentielle pour rester compétitif.

En choisissant un fournisseur de cloud qui innove en permanence pour vous apporter tous les outils de données dont vous aurez besoin, avec un rapport performances-prix adapté à votre cas d'utilisation, vous serez en mesure de créer un modèle de données qui évoluera avec votre start-up. Des bases de données d'applications à l'IA générative personnalisée basée sur vos données en passant par les lacs de données, l'analytique et le ML, AWS offre tout ce dont vous avez besoin pour créer un modèle de données cloud efficace de bout en bout, avec des options de tarification et de service qui vous permettent d'optimiser les coûts à chaque étape du processus. Exploitez le potentiel de vos données et préparez l'avenir avec les services de données AWS.

**Découvrez comment AWS permet aux start-ups de maximiser la valeur commerciale grâce à un modèle de données de bout en bout ›**

**Découvrez comment AWS aide les start-ups à tirer parti de la valeur commerciale de l'IA générative ›**