

Choisir un service AWS d'analyse



Choisir un service AWS d'analyse: AWS Guide de décision

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

Guide de décision 1

Introduction 1

Comprenez 2

Tenez compte 6

Choix 15

Utiliser 21

Explorez 30

Historique du document 32

..... xxxiii

Choisir un service AWS d'analyse

Faire le premier pas

Objectif	Aidez à déterminer les services AWS d'analyse les mieux adaptés à votre organisation.
Dernière mise à jour	24 septembre 2025
Services couverts	<ul style="list-style-type: none">• Amazon Athena• AWS Clean Rooms• Amazon Data Firehose• Amazon DataZone• Amazon EMR• AWS Glue• Amazon Kinesis Data Streams• Service géré Amazon pour Apache Flink• Streaming géré par Amazon pour Apache Kafka• Amazon Managed Workflows for Apache Airflow• Amazon OpenSearch Service• Suite rapide• Amazon Redshift• Amazon S3• Amazon SageMaker

Introduction

Les données sont fondamentales pour les entreprises modernes. Les personnes et les applications doivent accéder aux données et les analyser en toute sécurité, qui proviennent de sources nouvelles et diverses. Le volume de données augmente également constamment, ce qui peut amener les entreprises à avoir du mal à capturer, stocker et analyser toutes les données nécessaires.

Pour relever ces défis, il faut créer une architecture de données moderne qui élimine tous vos silos de données pour les analyses et les informations, y compris les données tierces, et les rend accessibles à tous les membres de l'organisation, en un seul endroit, grâce à une gouvernance end-to-end. Il est également de plus en plus important de connecter vos systèmes d'analyse et d'apprentissage automatique (ML) pour permettre l'analyse prédictive.

Ce guide de décision vous aide à vous poser les bonnes questions pour construire votre architecture de données moderne basée sur les AWS services. Il explique comment décomposer vos silos de données (en connectant votre lac de données et vos entrepôts de données), vos silos de système (en connectant le ML et les analyses) et vos silos de personnel (en mettant les données entre les mains de tous les membres de votre organisation).

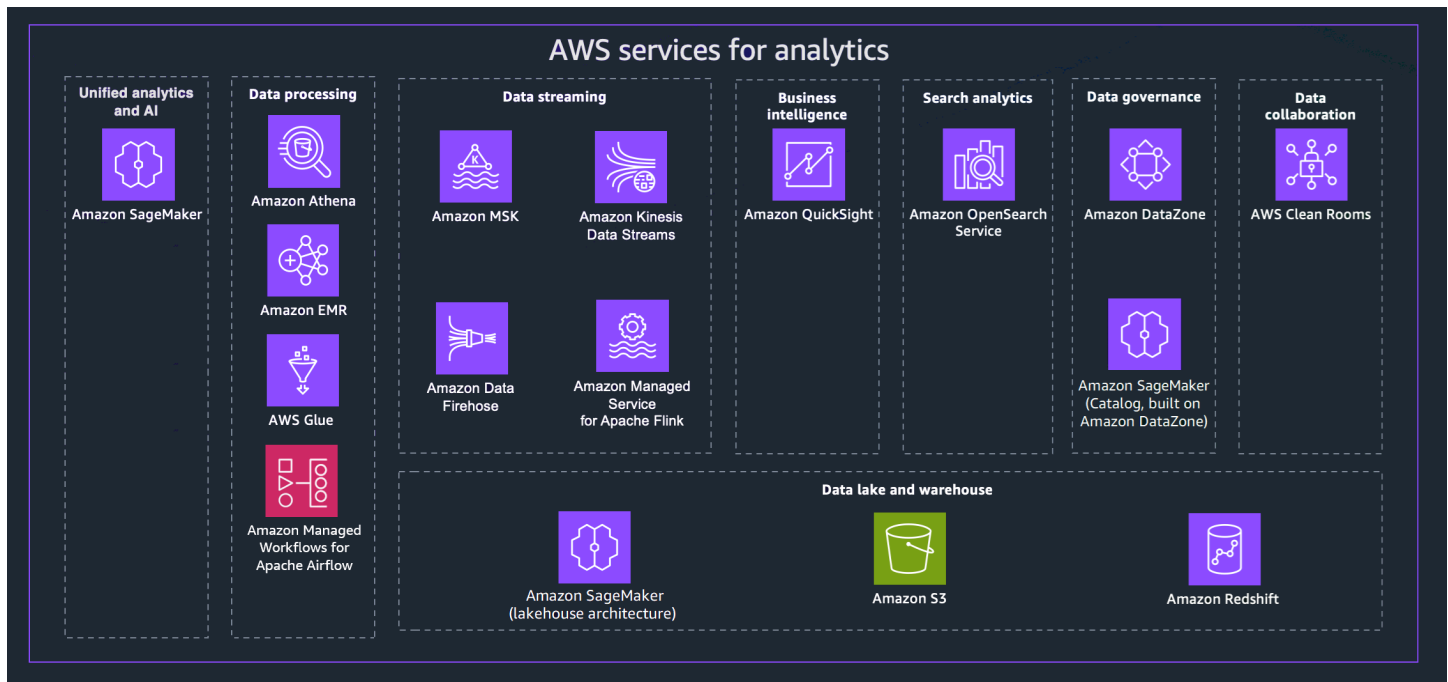
[Cet extrait de huit minutes est tiré d'une présentation d'une heure par Sirish Chandrasekaran et Rick Sears à re:Invent 2024. Il fournit un aperçu de la manière dont une entreprise fictive, Maxdome, utilise l'intelligence artificielle et l'analyse d' SageMaker Unified Studio, qui font partie de la prochaine génération d'Amazon, pour obtenir des informations sur les données SageMaker.](#)

Comprendre les services AWS d'analyse

Une stratégie de données moderne repose sur un ensemble de composants technologiques qui vous aident à gérer les données, à y accéder, à les analyser et à agir en fonction de celles-ci. Il vous offre également plusieurs options pour vous connecter aux sources de données. Une stratégie de données moderne doit permettre à vos équipes de :

- Utilisez vos outils ou techniques préférés
- Utilisez l'intelligence artificielle (IA) pour vous aider à trouver des réponses à des questions spécifiques concernant vos données
- Gérez qui a accès aux données grâce aux contrôles de sécurité et de gouvernance des données appropriés
- Éliminez les silos de données pour tirer le meilleur parti des lacs de données et des magasins de données spécialement conçus
- Stockez n'importe quelle quantité de données, à moindre coût et dans des formats de données ouverts et basés sur des normes
- Connectez vos lacs de données, entrepôts de données, bases de données opérationnelles, applications et sources de données fédérées au sein d'un ensemble cohérent

AWS propose une variété de services pour vous aider à mettre en place une stratégie de données moderne. Le schéma suivant décrit les AWS services d'analyse couverts par ce guide. Les onglets suivants fournissent des informations supplémentaires.



Unified analytics and AI

La nouvelle génération d'[Amazon SageMaker](#) combine des fonctionnalités d'apprentissage AWS automatique (ML) et d'analyse largement adoptées pour offrir une expérience intégrée pour l'analyse et l'IA, fournissant un accès unifié à toutes vos données. [Amazon SageMaker Unified Studio](#) vous permet de collaborer et de créer plus rapidement à l'aide d'AWS outils courants pour le développement de modèles, le développement d'applications d'IA générative, le traitement des données et l'analyse SQL, le tout accéléré par Amazon Q Developer, notre assistant d'IA générative pour le développement de logiciels. Accédez à vos données à partir de lacs de données, d'entrepôts de données ou de sources tierces et fédérées, grâce à une gouvernance intégrée répondant aux exigences de sécurité de l'entreprise.

Data processing

- [Amazon Athena](#) vous aide à analyser les données non structurées, semi-structurées et structurées stockées dans Amazon S3. Par exemple, des formats de données CSV ou JSON, ou des formats en colonnes, tels qu'Apache Parquet et Apache ORC. Vous pouvez utiliser Athena pour exécuter des requêtes ad hoc en utilisant ANSI SQL, sans avoir besoin d'agréger ou de charger les données dans Athena. Athena s'intègre à Quick Suite et à [d'autres AWS services](#). AWS Glue Data Catalog Vous pouvez également analyser les données à grande

échelle avec [Trino](#), sans avoir à gérer l'infrastructure, et créer des analyses en temps réel à l'aide d'Apache Flink et d'Apache Spark.

- [Amazon EMR](#) est une plate-forme de cluster gérée qui simplifie l'exécution de frameworks de mégadonnées, tels qu'Apache Hadoop et Apache Spark, AWS pour traiter et analyser de grandes quantités de données. Grâce à ces infrastructures et des projets open source connexes, vous pouvez traiter des données à des fins d'analyse et pour des charges de travail business intelligence. Amazon EMR vous permet également de transformer et de déplacer de grandes quantités de données vers et depuis d'autres banques de AWS données et bases de données, comme Amazon S3.
- Avec [AWS Glue](#), vous pouvez découvrir et vous connecter à plus de 100 sources de données diverses et gérer vos données dans un catalogue de données centralisé. Vous pouvez créer, exécuter et surveiller visuellement des pipelines ETL pour charger des données dans vos lacs de données. Vous pouvez également rechercher et interroger immédiatement des données cataloguées à l'aide d'Athena, Amazon EMR et Amazon Redshift Spectrum.
- [Amazon Managed Workflows for Apache Airflow](#) (MWAA) est une implémentation entièrement gérée d'Apache Airflow qui facilite la création, la planification et le suivi des flux de travail de données dans le cloud. MWAA adapte automatiquement la capacité du flux de travail pour répondre à vos besoins et s'intègre aux services AWS de sécurité. Vous pouvez utiliser MWAA pour orchestrer les flux de travail dans l'ensemble de vos services d'analyse, notamment le traitement des données, les tâches ETL et les pipelines d'apprentissage automatique.

Data streaming

- Avec [Amazon Managed Streaming for Apache Kafka](#) (Amazon MSK), vous pouvez créer et exécuter des applications qui utilisent Apache Kafka pour traiter les données de streaming. Amazon MSK fournit les opérations de plan de contrôle, telles que les opérations de création, de mise à jour et de suppression de clusters. Il vous permet d'utiliser les opérations de plan de données Apache Kafka, telles que celles pour la production et la consommation de données.
- Avec [Amazon Kinesis Data](#) Streams, vous pouvez collecter et traiter de grands flux d'enregistrements de données en temps réel. Le type de données utilisé peut inclure des données de journaux d'infrastructure informatique, des journaux d'applications, des réseaux sociaux, des flux de données du marché et des données de flux de clics web.
- [Amazon Data Firehose](#) est un service entièrement géré qui fournit des données de streaming en temps réel vers des destinations telles qu'Amazon S3, Amazon Redshift, OpenSearch Amazon Service, Splunk et Apache Iceberg Tables. Vous pouvez également envoyer

des données à n'importe quel point de terminaison HTTP personnalisé ou à tout point de terminaison HTTP appartenant à des fournisseurs de services tiers pris en charge, notamment Datadog, Dynatrace, MongoDB, New LogicMonitor Relic, Coralogix et Elastic.

- Avec [Amazon Managed Service pour Apache Flink](#), vous pouvez utiliser Java, Scala, Python ou SQL pour traiter et analyser les données de streaming. Vous pouvez créer et exécuter du code sur des sources de streaming et des sources statiques pour effectuer des analyses de séries chronologiques, alimenter des tableaux de bord en temps réel et des métriques.

Business intelligence

[Quick Suite](#) donne aux décideurs la possibilité d'explorer et d'interpréter les informations dans un environnement visuel interactif. Dans un tableau de bord de données unique, Quick Suite peut inclure AWS des données, des données tierces, des mégadonnées, des données de feuille de calcul, des données SaaS, des données B2B, etc. Avec Quick Suite Q, vous pouvez utiliser le langage naturel pour poser des questions sur vos données et recevoir une réponse. Par exemple, « Quelles sont les catégories les plus vendues en Californie ? »

Search analytics

[Amazon OpenSearch Service](#) fournit toutes les ressources pour votre OpenSearch cluster et le lance. Il détecte et remplace également automatiquement les nœuds de OpenSearch service défectueux, réduisant ainsi les frais associés aux infrastructures autogérées. Vous pouvez utiliser OpenSearch Service direct query pour analyser les données dans Amazon S3 et d'autres AWS services.

Data governance

Avec [Amazon DataZone](#), vous pouvez gérer et gouverner l'accès aux données à l'aide de contrôles précis. Ces contrôles permettent de garantir un accès avec le bon niveau de privilèges et de contexte. Amazon DataZone simplifie votre architecture en intégrant des services de gestion des données, notamment Amazon Redshift, Athena, Quick Suite AWS Glue, des sources sur site et des sources tierces.

Data collaboration

[AWS Clean Rooms](#) est un espace de travail collaboratif sécurisé dans lequel vous pouvez analyser des ensembles de données collectifs sans donner accès aux données brutes. Vous pouvez collaborer avec d'autres entreprises en choisissant les partenaires avec lesquels vous souhaitez collaborer, en sélectionnant leurs ensembles de données et en configurant des contrôles renforçant la confidentialité pour ces partenaires. Lorsque vous exécutez des requêtes,

AWS Clean Rooms lisez les données à partir de leur emplacement d'origine et appliquez des règles d'analyse intégrées pour vous aider à garder le contrôle sur ces données.

Data lake and data warehouse

- [La nouvelle génération d'Amazon SageMaker](#) est entièrement compatible avec Apache Iceberg, ce qui vous permet d'unifier les données entre les lacs de données Amazon Simple Storage Service (Amazon S3) et les entrepôts de données Amazon Redshift. Cela permet de créer des applications d'analyse, d'IA et d'apprentissage automatique (ML) sur une seule copie de données. Grâce aux intégrations sans ETL, vous pouvez diffuser des données provenant de sources opérationnelles en temps quasi réel, exécuter des requêtes fédérées sur plusieurs sources et accéder aux données à l'aide d'outils compatibles avec Apache Iceberg. Vous pouvez sécuriser vos données en définissant des autorisations précises qui sont appliquées à tous vos outils et moteurs d'analyse et de machine learning.
- [Amazon S3](#) peut stocker et protéger pratiquement n'importe quel type de données, que vous pouvez utiliser comme base de votre lac de données. Amazon S3 offre des fonctions de gestion qui vous permettent d'optimiser, d'organiser et de configurer l'accès à vos données aux fins de répondre aux exigences spécifiques de votre entreprise, de votre organisation et de votre conformité. Les tables Amazon S3 fournissent un stockage S3 optimisé pour les charges de travail d'analyse. À l'aide d'instructions SQL standard, vous pouvez interroger vos tables à l'aide de moteurs de requête compatibles avec Iceberg, tels qu'Athena, Amazon Redshift et Apache Spark.
- [Amazon Redshift](#) est un service d'entrepôt de données entièrement géré de plusieurs pétaoctets. Amazon Redshift peut être connecté à un data lakehouse d'Amazon SageMaker, ce qui vous permet d'utiliser ses puissantes fonctionnalités d'analyse SQL sur vos données unifiées dans les entrepôts de données Amazon Redshift et les lacs de données Amazon S3. Vous pouvez également utiliser Amazon Q dans Amazon Redshift, qui simplifie la création de code SQL grâce au langage naturel.

Tenez compte des critères applicables aux services AWS d'analyse

Il existe de nombreuses raisons de développer l'analyse des données sur AWS. Il se peut que vous deviez soutenir un nouveau projet ou un projet pilote comme première étape de votre migration vers le cloud. Il se peut également que vous migriez une charge de travail existante avec le moins de perturbations possible. Quel que soit votre objectif, les considérations suivantes peuvent vous être utiles pour faire votre choix.

Assess data sources and data types

Analysez les sources de données et les types de données disponibles pour acquérir une compréhension complète de la diversité, de la fréquence et de la qualité des données. Identifiez les défis potentiels liés au traitement et à l'analyse des données. Cette analyse est cruciale pour les raisons suivantes :

- Les sources de données sont diverses et proviennent de différents systèmes, applications, appareils et plateformes externes.
- Les sources de données ont une structure, un format et une fréquence de mise à jour des données uniques. L'analyse de ces sources permet d'identifier les méthodes et technologies de collecte de données appropriées.
- L'analyse des types de données, tels que les données structurées, semi-structurées et non structurées, détermine les approches appropriées en matière de traitement et de stockage des données.
- L'analyse des sources et des types de données facilite l'évaluation de la qualité des données et vous aide à anticiper les problèmes potentiels liés à la qualité des données (valeurs manquantes, incohérences ou inexactitudes).

Data processing requirements

Déterminez les exigences en matière de traitement des données relatives à la manière dont les données sont ingérées, transformées, nettoyées et préparées pour l'analyse. Les principales considérations sont les suivantes :

- Transformation des données : déterminez les transformations spécifiques nécessaires pour rendre les données brutes adaptées à l'analyse. Cela implique des tâches telles que l'agrégation, la normalisation, le filtrage et l'enrichissement des données.
- Nettoyage des données : évaluez la qualité des données et définissez des processus pour traiter les données manquantes, inexactes ou incohérentes. Mettez en œuvre des techniques de nettoyage des données pour garantir des données de haute qualité et des informations fiables.
- Fréquence de traitement : déterminez si un traitement en temps réel, en temps quasi réel ou par lots est requis en fonction des besoins analytiques. Le traitement en temps réel permet d'obtenir des informations immédiates, tandis que le traitement par lots peut être suffisant pour des analyses périodiques.

- **Évolutivité et débit** : évaluez les exigences d'évolutivité pour la gestion des volumes de données, la vitesse de traitement et le nombre de demandes de données simultanées. Assurez-vous que l'approche de traitement choisie peut s'adapter à la croissance future.
- **Latence** : considérez la latence acceptable pour le traitement des données et le temps nécessaire entre l'ingestion des données et les résultats de l'analyse. Cela est particulièrement important pour les analyses en temps réel ou sensibles au facteur temps.

Storage requirements

Déterminez les besoins de stockage en déterminant comment et où les données sont stockées tout au long du pipeline d'analyse. Les points importants à prendre en compte incluent :

- **Volume de données** : évaluez la quantité de données générées et collectées, et estimez la croissance future des données afin de prévoir une capacité de stockage suffisante.
- **Conservation des données** : définissez la durée pendant laquelle les données doivent être conservées à des fins d'analyse historique ou de conformité. Déterminez les politiques de conservation des données appropriées.
- **Modèles d'accès aux données** : découvrez comment les données seront consultées et interrogées afin de choisir la solution de stockage la plus adaptée. Tenez compte des opérations de lecture et d'écriture, de la fréquence d'accès aux données et de la localité des données.
- **Sécurité des données** : priorisez la sécurité des données en évaluant les options de chiffrement, les contrôles d'accès et les mécanismes de protection des données pour protéger les informations sensibles.
- **Optimisation des coûts** : optimisez les coûts de stockage en sélectionnant les solutions de stockage les plus économiques en fonction des modèles d'accès aux données et de leur utilisation.
- **Intégration aux services d'analyse** : garantisiez une intégration parfaite entre la solution de stockage choisie et les outils de traitement et d'analyse des données en cours de développement.

Types of data

Lorsque vous choisissez des services d'analyse pour la collecte et l'ingestion de données, prenez en compte différents types de données correspondant aux besoins et aux objectifs de votre

organisation. Les types de données courants que vous devrez peut-être prendre en compte sont les suivants :

- **Données transactionnelles** : incluent des informations sur les interactions ou transactions individuelles, telles que les achats des clients, les transactions financières, les commandes en ligne et les journaux d'activité des utilisateurs.
- **Données basées sur des fichiers** : font référence aux données structurées ou non structurées stockées dans des fichiers, tels que des fichiers journaux, des feuilles de calcul, des documents, des images, des fichiers audio et des fichiers vidéo. Les services d'analyse doivent prendre en charge l'ingestion de différents formats de fichiers.
- **Données sur les événements** : capture les événements ou incidents importants, tels que les actions des utilisateurs, les événements du système, les événements liés aux machines ou les événements commerciaux. Les événements peuvent inclure toutes les données qui arrivent à grande vitesse et qui sont capturées pour un traitement en continu ou en aval.

Operational considerations

La responsabilité opérationnelle est partagée entre vous et AWS, la division des responsabilités variant selon les différents niveaux de modernisation. Vous avez la possibilité de gérer vous-même votre infrastructure d'analyse AWS ou de tirer parti des nombreux services d'analyse sans serveur pour alléger la charge de gestion de votre infrastructure.

Les options autogérées permettent aux utilisateurs de mieux contrôler l'infrastructure et les configurations, mais elles nécessitent davantage d'efforts opérationnels.

Les options sans serveur réduisent une grande partie de la charge opérationnelle en fournissant une évolutivité automatique, une haute disponibilité et des fonctionnalités de sécurité robustes, permettant aux utilisateurs de se concentrer davantage sur la création de solutions analytiques et l'obtention d'informations plutôt que sur la gestion de l'infrastructure et des tâches opérationnelles. Tenez compte des avantages des solutions d'analyse sans serveur :

- **Abstraction de l'infrastructure** : les services sans serveur font abstraction de la gestion de l'infrastructure, soulageant ainsi les utilisateurs des tâches de provisionnement, de dimensionnement et de maintenance. AWS gère ces aspects opérationnels, réduisant ainsi les frais de gestion.
- **Mise à l'échelle et performances automatiques** : les services sans serveur dimensionnent automatiquement les ressources en fonction des demandes de charge de travail, garantissant ainsi des performances optimales sans intervention manuelle.

- Haute disponibilité et reprise après sinistre : AWS fournissent une haute disponibilité pour les services sans serveur. AWS gère la redondance des données, la réplication et la reprise après sinistre afin d'améliorer la disponibilité et la fiabilité des données.
- Sécurité et conformité : AWS gère les mesures de sécurité, le chiffrement des données et la conformité pour les services sans serveur, conformément aux normes et aux meilleures pratiques du secteur.
- Surveillance et journalisation : AWS offre des fonctionnalités intégrées de surveillance, de journalisation et d'alerte pour les services sans serveur. Les utilisateurs peuvent accéder à des statistiques et à des journaux détaillés via Amazon CloudWatch.

Type of workload

Lors de la création d'un pipeline d'analyse moderne, il est essentiel de choisir les types de charge de travail à prendre en charge pour répondre efficacement aux différents besoins analytiques. Les principaux points de décision à prendre en compte pour chaque type de charge de travail sont les suivants :

Charge de travail par lots

- Volume et fréquence des données : le traitement par lots convient aux gros volumes de données soumis à des mises à jour périodiques.
- Latence des données : le traitement par lots peut entraîner un certain retard dans la fourniture d'informations par rapport au traitement en temps réel.

Analyse interactive

- Complexité des requêtes de données : l'analyse interactive nécessite des réponses à faible latence pour un feedback rapide.
- Visualisation des données : évaluez le besoin d'outils de visualisation de données interactifs pour permettre aux utilisateurs professionnels d'explorer les données visuellement.

Charges de travail en streaming

- Vitesse et volume des données : les charges de travail en streaming nécessitent un traitement en temps réel pour traiter les données à haute vitesse.

- Fenêtrage des données : définissez le fenêtrage des données et les agrégations temporelles pour le streaming des données afin d'en extraire des informations pertinentes.

Type of analysis needed

Définissez clairement les objectifs commerciaux et les informations que vous souhaitez tirer des analyses. Les différents types d'analyses répondent à des objectifs différents. Par exemple :

- L'analyse descriptive est idéale pour obtenir une vue d'ensemble historique
- L'analyse diagnostique permet de comprendre les raisons des événements passés
- L'analyse prédictive prédit les résultats futurs
- L'analyse prescriptive fournit des recommandations pour des actions optimales

Associez vos objectifs commerciaux aux types d'analyses pertinents. Voici quelques critères de décision clés qui vous aideront à choisir les bons types d'analyses :

- Disponibilité et qualité des données : les analyses descriptives et diagnostiques s'appuient sur des données historiques, tandis que les analyses prédictives et prescriptives nécessitent des données historiques suffisantes et des données de haute qualité pour créer des modèles précis.
- Volume et complexité des données : les analyses prédictives et prescriptives nécessitent d'importantes ressources informatiques et de traitement des données. Assurez-vous que votre infrastructure et vos outils peuvent gérer le volume et la complexité des données.
- Complexité des décisions : si les décisions impliquent plusieurs variables, contraintes et objectifs, l'analyse prescriptive peut être plus adaptée pour orienter les actions optimales.
- Tolérance au risque : les analyses prescriptives peuvent fournir des recommandations, mais elles s'accompagnent d'incertitudes. Assurez-vous que les décideurs comprennent les risques associés aux résultats d'analyse.

Evaluate scalability and performance

Évaluez les besoins en évolutivité et en performances de l'architecture. La conception doit gérer des volumes de données croissants, des demandes des utilisateurs et des charges de travail analytiques. Les principaux facteurs de décision à prendre en compte sont les suivants :

- Volume de données et croissance : évaluez le volume de données actuel et anticipez la croissance future.
- Vitesse des données et exigences en temps réel : déterminez si les données doivent être traitées et analysées en temps réel ou en temps quasi réel.
- Complexité du traitement des données : analysez la complexité de vos tâches de traitement et d'analyse des données. Pour les tâches de calcul intensives, des services tels qu'Amazon EMR fournissent un environnement évolutif et géré pour le traitement des mégadonnées.
- Simultanéité et charge utilisateur : tenez compte du nombre d'utilisateurs simultanés et du niveau de charge utilisateur sur le système.
- Capacités de mise à l'échelle automatique : considérez les services qui offrent des fonctionnalités d'auto-dimensionnement, permettant aux ressources d'augmenter ou de diminuer automatiquement en fonction de la demande. Cela garantit une utilisation efficace des ressources et une optimisation des coûts.
- Distribution géographique : envisagez des services dotés d'une réplication globale et d'un accès aux données à faible latence si votre architecture de données doit être distribuée sur plusieurs régions ou sites.
- Compromis coût-performance : trouvez un équilibre entre les besoins de performance et les considérations financières. Les services très performants peuvent avoir un coût plus élevé.
- Contrats de niveau de service (SLAs) : vérifiez les AWS services SLAs fournis pour vous assurer qu'ils répondent à vos attentes en matière d'évolutivité et de performance.

Data governance

La gouvernance des données est l'ensemble des processus, des politiques et des contrôles que vous devez mettre en œuvre pour garantir une gestion, une qualité, une sécurité et une conformité efficaces de vos actifs de données. Les principaux points de décision à prendre en compte sont les suivants :

- Politiques de conservation des données : définissez des politiques de conservation des données en fonction des exigences réglementaires et des besoins de l'entreprise et établissez des processus pour une élimination sécurisée des données lorsqu'elles ne sont plus nécessaires.
- Piste d'audit et journalisation : choisissez les mécanismes de journalisation et d'audit pour surveiller l'accès aux données et leur utilisation. Mettez en œuvre des pistes d'audit complètes

pour suivre les modifications des données, les tentatives d'accès et les activités des utilisateurs à des fins de surveillance de la conformité et de la sécurité.

- **Exigences de conformité** : comprenez les réglementations de conformité des données géographiques et sectorielles qui s'appliquent à votre organisation. Assurez-vous que l'architecture des données est conforme à ces réglementations et directives.
- **Classification des données** : Classez les données en fonction de leur sensibilité et définissez les contrôles de sécurité appropriés pour chaque classe de données.
- **Reprise après sinistre et continuité des activités** : planifiez la reprise après sinistre et la continuité des activités afin de garantir la disponibilité et la résilience des données en cas d'événements imprévus ou de défaillances du système.
- **Partage de données avec des tiers** : si vous partagez des données avec des entités tierces, mettez en œuvre des protocoles et des accords de partage de données sécurisés afin de protéger la confidentialité des données et d'empêcher toute utilisation abusive des données.

Security

La sécurité des données du pipeline d'analyse implique de protéger les données à chaque étape du pipeline afin de garantir leur confidentialité, leur intégrité et leur disponibilité. Les principaux points de décision à prendre en compte sont les suivants :

- **Contrôle d'accès et autorisation** : mettez en œuvre des protocoles d'authentification et d'autorisation robustes pour garantir que seuls les utilisateurs autorisés peuvent accéder à des ressources de données spécifiques.
- **Chiffrement des données** : choisissez les méthodes de chiffrement appropriées pour les données stockées dans les bases de données, les lacs de données et lors du transfert de données entre les différents composants de l'architecture.
- **Masquage et anonymisation des données** : pensez à la nécessité de masquer ou d'anonymiser les données pour protéger les données sensibles, telles que les informations personnelles ou les données commerciales sensibles, tout en permettant à certains processus analytiques de se poursuivre.
- **Intégration sécurisée des données** : établissez des pratiques d'intégration des données sécurisées pour garantir que les données circulent en toute sécurité entre les différents composants de l'architecture, en évitant les fuites de données ou les accès non autorisés lors du déplacement des données.

- Isolation du réseau : pensez aux services qui prennent en charge les [points de terminaison Amazon VPC](#) afin d'éviter d'exposer les ressources à l'Internet public.

Plan for integration and data flows

Définissez les points d'intégration et les flux de données entre les différents composants du pipeline d'analyse afin de garantir un flux de données fluide et une interopérabilité. Les principaux points de décision à prendre en compte sont les suivants :

- Intégration des sources de données : identifiez les sources de données à partir desquelles les données seront collectées, telles que les bases de données, les applications, les fichiers ou les sources externes APIs. Choisissez les méthodes d'ingestion de données (par lots, en temps réel, basées sur des événements) pour intégrer les données dans le pipeline de manière efficace et avec une latence minimale.
- Transformation des données : déterminez les transformations nécessaires pour préparer les données en vue de leur analyse. Choisissez les outils et les processus permettant de nettoyer, d'agréger, de normaliser ou d'enrichir les données au fur et à mesure de leur transfert dans le pipeline.
- Architecture de mouvement des données : choisissez l'architecture appropriée pour le mouvement des données entre les composants du pipeline. Envisagez le traitement par lots, le traitement par flux ou une combinaison des deux en fonction des exigences en temps réel et du volume de données.
- Réplication et synchronisation des données : choisissez les mécanismes de réplication et de synchronisation des données pour conserver les données up-to-date sur tous les composants. Envisagez des solutions de réplication en temps réel ou des synchronisations de données périodiques en fonction des exigences en matière de fraîcheur des données.
- Qualité et validation des données : mettez en œuvre des contrôles de qualité des données et des étapes de validation pour garantir l'intégrité des données lors de leur transfert dans le pipeline. Décidez des mesures à prendre lorsque les données ne sont pas validées, telles que les alertes ou la gestion des erreurs.
- Sécurité et chiffrement des données : déterminez comment les données seront sécurisées pendant le transit et au repos. Choisissez les méthodes de chiffrement pour protéger les données sensibles tout au long du pipeline, en tenant compte du niveau de sécurité requis en fonction de la sensibilité des données.
- Évolutivité et résilience : assurez-vous que la conception du flux de données permet une évolutivité horizontale et peut gérer des volumes de données et un trafic accru.

Architect for cost optimization

Le développement de votre pipeline d'analyse AWS offre diverses opportunités d'optimisation des coûts. Pour garantir la rentabilité, considérez les stratégies suivantes :

- **Dimensionnement et sélection des ressources** : dimensionnez correctement vos ressources en fonction des exigences de charge de travail réelles. Choisissez AWS des services et des types d'instances qui répondent aux besoins de performance des charges de travail tout en évitant le surprovisionnement.
- **Mise à l'échelle automatique** : implémentez l'auto-scaling pour les services soumis à des charges de travail variables. La mise à l'échelle automatique ajuste dynamiquement le nombre d'instances en fonction de la demande, réduisant ainsi les coûts pendant les périodes de faible trafic.
- **Instances ponctuelles** : utilisez les instances Amazon EC2 Spot pour les charges de travail non critiques et tolérantes aux pannes. Les instances Spot peuvent réduire considérablement les coûts par rapport aux instances à la demande.
- **Instances réservées** : envisagez d'acheter des instances AWS réservées pour réaliser des économies importantes par rapport à la tarification à la demande pour des charges de travail stables avec une utilisation prévisible.
- **Hierarchisation du stockage des données** : optimisez les coûts de stockage des données en utilisant différentes classes de stockage en fonction de la fréquence d'accès aux données.
- **Politiques relatives au cycle de vie des données** : établissez des politiques de cycle de vie des données pour déplacer ou supprimer automatiquement les données en fonction de leur ancienneté et de leurs habitudes d'utilisation. Cela permet de gérer les coûts de stockage et d'aligner le stockage des données sur leur valeur.

Choisissez les services AWS d'analyse

Maintenant que vous connaissez les critères permettant d'évaluer vos besoins en matière d'analyse, vous êtes prêt à choisir les services AWS d'analyse adaptés aux besoins de votre organisation. Le tableau suivant aligne les ensembles de services sur les capacités et les objectifs commerciaux communs.

Catégories	Pour quoi est-il optimisé ?	Services
Analyses unifiées et intelligence artificielle	<p>Analyse et développement de l'IA</p> <p>Optimisé pour utiliser un environnement de développement unique, Amazon SageMaker Unified Studio, afin d'accéder aux données, aux analyses et aux fonctionnalités d'intelligence artificielle.</p>	Amazon SageMaker
Traitement des données	<p>Analyses interactives</p> <p>Optimisé pour effectuer une analyse et une exploration de données en temps réel, ce qui permet aux utilisateurs d'interroger et de visualiser les données de manière interactive.</p>	Amazon Athena
	<p>Traitement des mégadonnées</p> <p>Optimisé pour le traitement, le déplacement et la transformation de grandes quantités de données.</p>	Amazon EMR
	<p>Data catalog (Catalogue de données)</p> <p>Optimisé pour fournir des informations détaillées sur les données disponibles, leur structure, leurs caractéristiques et leurs relations.</p>	AWS Glue

Catégories	Pour quoi est-il optimisé ?	Services
	<p>Orchestration du flux de travail</p> <p>Optimisé pour créer, planifier et surveiller des flux de données à l'aide d'Apache Airflow afin de coordonner les processus d'analyse et les tâches ETL.</p>	<u>Amazon MWAA</u>
Streaming de données	<p>Traitement des données de streaming par Apache Kafka</p> <p>Optimisé pour utiliser les opérations du plan de données Apache Kafka et exécuter des versions open source d'Apache Kafka.</p>	<u>Amazon MSK</u>
	<p>Traitement en temps réel</p> <p>Optimisé pour la saisie et l'agrégation rapides et continues des données, notamment les données des journaux de l'infrastructure informatique, les journaux des applications, les réseaux sociaux, les flux de données de marché et les données de navigation sur le Web.</p>	<u>Amazon Kinesis Data Streams</u>

Catégories	Pour quoi est-il optimisé ?	Services
	<p>Livraison de données en streaming en temps réel</p> <p>Optimisé pour fournir des données de streaming en temps réel vers des destinations telles qu'Amazon S3, Amazon Redshift, OpenSearch Service, Splunk, Apache Iceberg Tables et tout point de terminaison HTTP personnalisé ou appartenant à des fournisseurs de services tiers pris en charge.</p>	Amazon Data Firehose
	<p>Création d'applications Apache Flink</p> <p>Optimisé pour utiliser Java, Scala, Python ou SQL pour traiter et analyser les données de streaming.</p>	Service géré Amazon pour Apache Flink
Intelligence d'affaires	<p>Tableaux de bord et visualisations</p> <p>Optimisé pour représenter visuellement des ensembles de données complexes et fournir des requêtes en langage naturel pour vos données.</p>	Suite rapide

Catégories	Pour quoi est-il optimisé ?	Services
Analyses de recherche	<p>OpenSearch Clusters gérés</p> <p>Optimisé pour l'analyse des journaux, la surveillance des applications en temps réel et l'analyse du flux de clics.</p>	Amazon OpenSearch Service
Gouvernance des données	<p>Gestion de l'accès aux données</p> <p>Optimisé pour configurer la gestion, la disponibilité, l'utilisabilité, l'intégrité et la sécurité appropriées des données tout au long de leur cycle de vie.</p>	Amazon DataZone
Collaboration sur les données	<p>Salles blanches de données sécurisées</p> <p>Optimisé pour collaborer avec d'autres entreprises sans partager de données sous-jacentes brutes.</p>	AWS Clean Rooms

Catégories	Pour quoi est-il optimisé ?	Services
Lac de données et entrepôt	<p>Accès unifié aux lacs de données et aux entrepôts de données</p> <p>Construit sur une architecture Lakehouse pour optimiser l'unification de l'accès aux données entre les lacs de données Amazon S3, les entrepôts de données Amazon Redshift, les bases de données opérationnelles et les sources de données tierces et fédérées.</p>	Amazon SageMaker
	<p>Stockage d'objets pour les lacs de données</p> <p>Optimisé pour fournir une base de lac de données avec une évolutivité pratiquement illimitée et une durabilité élevée.</p>	Amazon S3
	<p>Entreposage de données</p> <p>Optimisé pour le stockage, l'organisation et la récupération centralisés de gros volumes de données structurées et parfois semi-structurées provenant de diverses sources au sein d'une organisation.</p>	Amazon Redshift

Utiliser les services AWS d'analyse

Vous devez désormais avoir une idée précise de vos objectifs commerciaux, ainsi que du volume et de la rapidité des données que vous allez ingérer et analyser pour commencer à créer vos pipelines de données.

Pour découvrir comment utiliser et en savoir plus sur chacun des services disponibles, nous avons fourni un parcours permettant d'explorer le fonctionnement de chacun des services. Les sections suivantes fournissent des liens vers une documentation détaillée, des didacticiels pratiques et des ressources pour vous aider à passer d'une utilisation de base à des plongées approfondies plus avancées.

Amazon Athena

- Commencer à utiliser Amazon Athena

Découvrez comment utiliser Amazon Athena pour interroger des données et créer une table à partir d'exemples de données stockés dans Amazon S3, interroger la table et vérifier les résultats de la requête.

[Commencez avec le didacticiel](#)

- Commencez avec Apache Spark sur Athena

Utilisez l'expérience de bloc-notes simplifiée de la console Athena pour développer des applications Apache Spark à l'aide de Python ou d'Athena Notebook. APIs

[Commencez avec le didacticiel](#)

- Cataloguez et gérez les requêtes fédérées Athena avec l'architecture Amazon Lakehouse SageMaker

Découvrez comment vous connecter, gérer et exécuter des requêtes fédérées sur les données stockées dans Amazon Redshift, DynamoDB et Snowflake via le data lakehouse d'Amazon. SageMaker

[Lisez le blogue](#)

- Analyse des données dans Amazon S3 à l'aide d'Athena

Découvrez comment utiliser Athena sur les journaux des Elastic Load Balancers, générés sous forme de fichiers texte dans un format prédéfini. Nous vous montrons comment créer une table,

partitionner les données dans un format utilisé par Athena, les convertir au format Parquet et comparer les performances des requêtes.

[Lire le billet de blog](#)

AWS Clean Rooms

- Configuration AWS Clean Rooms

Découvrez comment procéder à la configuration AWS Clean Rooms dans votre AWS compte.

[Lisez le guide](#)

- Accédez à des informations sur les données dans des ensembles de données multipartites à l'aide de la résolution d' AWS entités AWS Clean Rooms sans partager les données sous-jacentes

Découvrez comment utiliser la préparation et le rapprochement pour améliorer la mise en correspondance des données avec les collaborateurs.

[Lire le billet de blog](#)

- Comment la confidentialité différentielle permet de débloquent des informations sans révéler de données au niveau individuel

Découvrez comment la confidentialité AWS Clean Rooms différentielle simplifie l'application de la confidentialité différentielle et contribue à protéger la vie privée de vos utilisateurs.

[Lisez le blogue](#)

Amazon Data Firehose

- Tutoriel : Création d'un stream Firehose depuis la console

Découvrez comment utiliser le AWS Management Console ou un AWS SDK pour créer un flux Firehose vers la destination de votre choix.

[Lisez le guide](#)

- Envoyer des données vers un flux Firehose

Découvrez comment utiliser différentes sources de données pour envoyer des données vers votre flux Firehose.

[Lisez le guide](#)

- Transformer les données sources dans Firehose

Découvrez comment invoquer votre fonction Lambda pour transformer les données sources entrantes et transmettre les données transformées aux destinations.

[Lisez le guide](#)

Amazon DataZone

- Commencer à utiliser Amazon DataZone

Apprenez à créer le domaine DataZone racine Amazon, à obtenir l'URL du portail de données, à parcourir les DataZone flux de travail Amazon de base pour les producteurs et les consommateurs de données.

[Commencez avec le didacticiel](#)

- Annonce de la disponibilité générale du lignage des données dans la prochaine génération d'Amazon SageMaker et d'Amazon DataZone

Découvrez comment Amazon DataZone utilise la capture automatique du lignage pour se concentrer sur la collecte et le mappage automatiques des informations de lignage provenant d'Amazon Redshift AWS Glue et d'Amazon Redshift.

[Lisez le blogue](#)

Amazon EMR

- Commencer à utiliser Amazon EMR

Découvrez comment lancer un exemple de cluster à l'aide de Spark et comment exécuter un PySpark script simple stocké dans un compartiment Amazon S3.

[Commencez avec le didacticiel](#)

- Commencer à utiliser Amazon EMR sur Amazon EKS

Nous vous montrons comment commencer à utiliser Amazon EMR sur Amazon EKS en déployant une application Spark sur un cluster virtuel.

[Explorez le guide](#)

- Commencez avec EMR Serverless

Découvrez comment Amazon EMR Serverless fournit un environnement d'exécution sans serveur qui simplifie le fonctionnement des applications d'analyse utilisant les derniers frameworks open source.

[Commencez avec le didacticiel](#)

AWS Glue

- Commencer avec AWS Glue DataBrew

Découvrez comment créer votre premier DataBrew projet. Vous chargez un exemple de jeu de données, vous exécutez des transformations sur cet ensemble de données, vous créez une recette pour capturer ces transformations et vous exécutez une tâche pour écrire les données transformées sur Amazon S3.

[Commencez avec le didacticiel](#)

- Transformez les données avec AWS Glue DataBrew

Découvrez AWS Glue DataBrew un outil visuel de préparation des données qui permet aux analystes de données et aux data scientists de nettoyer et de normaliser facilement les données afin de les préparer à des fins d'analyse et d'apprentissage automatique. Apprenez à créer un processus ETL à l'aide de AWS Glue DataBrew.

[Commencez avec le laboratoire](#)

- AWS Glue DataBrew journée d'immersion

Découvrez comment nettoyer et normaliser les données AWS Glue DataBrew à des fins d'analyse et d'apprentissage automatique.

[Commencez par l'atelier](#)

- Commencer à utiliser le AWS Glue Data Catalog

Découvrez comment créer le premier AWS Glue Data Catalog, qui utilise un compartiment Amazon S3 comme source de données.

[Commencez avec le didacticiel](#)

- Catalogue de données et robots d'exploration dans AWS Glue

Découvrez comment utiliser les informations du catalogue de données pour créer et surveiller vos tâches ETL.

[Explorez le guide](#)

Amazon Kinesis Data Streams

- Tutoriels de mise en route pour Amazon Kinesis Data Streams

Découvrez comment traiter et analyser les données boursières en temps réel.

[Commencez avec les didacticiels](#)

- Modèles architecturaux pour l'analyse en temps réel à l'aide d'Amazon Kinesis Data Streams, partie 1

Découvrez les modèles architecturaux courants de deux cas d'utilisation : l'analyse de données de séries chronologiques et les microservices pilotés par les événements.

[Lisez le blogue](#)

- Modèles architecturaux pour l'analyse en temps réel à l'aide d'Amazon Kinesis Data Streams, partie 2

Découvrez les applications d'intelligence artificielle avec Kinesis Data Streams dans trois scénarios : intelligence économique générative en temps réel, systèmes de recommandation en temps réel et diffusion et inférence de données liées à l'Internet des objets.

[Lisez le blogue](#)

Amazon Managed Service for Apache Flink

- Qu'est-ce qu'Amazon Managed Service pour Apache Flink ?

Découvrez les concepts fondamentaux d'Amazon Managed Service pour Apache Flink.

[Explorez le guide](#)

- Atelier Amazon Managed Service pour Apache Flink

Dans cet atelier, vous apprendrez comment déployer, exploiter et dimensionner une application Flink avec Amazon Managed Service pour Apache Flink.

[Participez à l'atelier virtuel](#)

Amazon MSK

- Commencer à utiliser Amazon MSK

Découvrez comment créer un cluster Amazon MSK, produire et consommer des données, et surveiller l'état de santé de votre cluster à l'aide de métriques.

[Commencez avec le guide](#)

- Atelier Amazon MSK

Approfondissez vos connaissances grâce à cet atelier pratique sur Amazon MSK.

[Commencez par l'atelier](#)

Amazon MWAA

- Commencer à utiliser Amazon MWAA

Découvrez comment créer votre premier environnement MWAA, télécharger un DAG sur Amazon S3 et exécuter votre premier flux de travail.

[Commencez avec le didacticiel](#)

- Création de pipelines de données avec Amazon MWAA

Découvrez comment créer end-to-end des pipelines de données qui orchestrent d'autres services d'AWS analyse tels que Glue, EMR et Redshift. Ce billet de blog explore une approche rationalisée et axée sur la configuration pour orchestrer les tâches dbt Core à l'aide de MWAA et Cosmos, les tâches exécutant des transformations sur Amazon Redshift.

[Lire le billet de blog](#)

- Atelier Amazon MWAA

Découvrez des exercices pratiques pour apprendre à déployer, configurer et utiliser Amazon MWAA pour l'orchestration du flux de travail des données.

[Commencez par l'atelier](#)

- Bonnes pratiques pour Amazon MWAA

Découvrez les modèles architecturaux et les meilleures pratiques pour utiliser Amazon MWAA dans vos flux de travail d'analyse.

[Lisez le guide](#)

OpenSearch Service

- Commencer à utiliser le OpenSearch service

Découvrez comment utiliser Amazon OpenSearch Service pour créer et configurer un domaine de test.

[Commencez avec le didacticiel](#)

- Visualisation des appels d'assistance client avec le OpenSearch service et OpenSearch les tableaux de bord

Découvrez un aperçu complet de la situation suivante : une entreprise reçoit un certain nombre d'appels de support client et souhaite les analyser. Quel est l'objet de chaque appel ? Combien étaient positifs ? Combien étaient négatifs ? Comment les gestionnaires peuvent-ils chercher ou revoir les transcriptions de ces appels ?

[Commencez avec le didacticiel](#)

- Atelier de mise en route avec Amazon OpenSearch Serverless

Découvrez comment configurer un nouveau domaine Amazon OpenSearch Serverless dans la AWS console. Explorez les différents types de requêtes de recherche disponibles, concevez des visualisations accrocheuses et découvrez comment sécuriser votre domaine et vos documents en fonction des privilèges utilisateur attribués.

[Commencez par l'atelier](#)

- Base de données vectorielle optimisée en termes de coûts : introduction aux techniques OpenSearch de quantification d'Amazon Service

Découvrez comment OpenSearch Service prend en charge les techniques scalaires et de quantification des produits afin d'optimiser l'utilisation de la mémoire et de réduire les coûts d'exploitation.

[Lire le billet de blog](#)

Quick Suite

- Commencer à utiliser l'analyse de données Quick Suite

Découvrez comment créer votre première analyse. Utilisez des échantillons de données pour créer une analyse simple ou plus avancée. Vous pouvez également vous connecter à vos propres données pour créer une analyse.

[Explorez le guide](#)

- Visualisation avec Quick Suite

Découvrez l'aspect technique de la business intelligence (BI) et de la visualisation des données avec AWS. Découvrez comment intégrer des tableaux de bord dans des applications et des sites Web, et comment gérer en toute sécurité les accès et les autorisations.

[Commencez avec le cours](#)

- Ateliers Quick Suite

Prenez une longueur d'avance dans votre aventure avec Quick Suite grâce aux ateliers

[Commencez par les ateliers](#)

Amazon Redshift

- Commencer à utiliser Amazon Redshift Serverless

Découvrez le flux de base d'Amazon Redshift Serverless pour créer des ressources sans serveur, vous connecter à Amazon Redshift Serverless, charger des échantillons de données, puis exécuter des requêtes sur les données.

[Explorez le guide](#)

- Atelier de plongée approfondie sur Amazon Redshift

Découvrez une série d'exercices qui aident les utilisateurs à commencer à utiliser la plateforme Amazon Redshift.

[Commencez par l'atelier](#)

Amazon S3

- Commencer à utiliser Amazon S3

Découvrez comment créer votre premier DataBrew projet. Vous chargez un exemple de jeu de données, vous exécutez des transformations sur cet ensemble de données, vous créez une recette pour capturer ces transformations et vous exécutez une tâche pour écrire les données transformées sur Amazon S3.

[Commencez avec le guide](#)

Amazon SageMaker

- Commencer avec SageMaker

Apprenez à créer un projet, à ajouter des membres et à utiliser le JupyterLab bloc-notes d'exemple pour commencer à le créer.

[Lisez le guide](#)

- Présentation de la nouvelle génération d'Amazon SageMaker : le centre de toutes vos données, analyses et intelligence artificielle

Découvrez comment démarrer avec le traitement des données, le développement de modèles et le développement d'applications d'IA générative.

[Lisez le blogue](#)

- Qu'est-ce que SageMaker Unified Studio ?

Découvrez les fonctionnalités d' Amazon SageMaker Unified Studio et comment y accéder lorsque vous utilisez Amazon SageMaker.

[Lisez le guide](#)

- Débuter avec l'architecture Lakehouse d'Amazon SageMaker

Apprenez à créer un projet et à parcourir, télécharger et interroger des données pour vos cas d'utilisation professionnels sur Amazon SageMaker.

[Lisez le guide](#)

- Connexions de données dans l'architecture Lakehouse d'Amazon SageMaker

Découvrez comment l'architecture Lakehouse fournit une approche unifiée pour gérer les connexions de données entre les AWS services et les applications d'entreprise.

[Lisez le guide](#)

- Cataloguez et gérez les requêtes fédérées Athena avec l'architecture Lakehouse SageMaker

Découvrez comment vous connecter, gérer et exécuter des requêtes fédérées sur les données stockées dans Amazon Redshift, DynamoDB et Snowflake pour vos projets Amazon SageMaker

[Lisez le blogue](#)

Découvrez les moyens d'utiliser les services AWS d'analyse

Editable architecture diagrams

Schémas d'architecture de référence

Explorez les diagrammes d'architecture pour vous aider à développer, à faire évoluer et à tester vos solutions d'analyse AWS.

[Explorez les architectures de référence analytiques](#)

Ready-to-use code

Solution en vedette	AWS Des solutions
Analyses évolutives à l'aide d'Apache Druid sur AWS	Explorez les solutions déployables préconfigurées

Déployez du code AWS intégré pour vous aider à configurer, exploiter et gérer Apache Druid sur un AWS environnement d'hébergement rentable, hautement disponible, résilient et tolérant aux pannes.

[Découvrez cette solution](#)

et leurs guides de mise en œuvre, conçus par. AWS

[Découvrez toutes les solutions AWS de sécurité, d'identité et de gouvernance](#)

Documentation

Livres blancs sur les analyses

Consultez les livres blancs pour obtenir des informations supplémentaires et des meilleures pratiques sur le choix, la mise en œuvre et l'utilisation des services d'analyse les mieux adaptés à votre organisation.

[Découvrez les livres blancs sur les analyses](#)

AWS Blogue sur les mégadonnées

Consultez les articles de blog qui abordent des cas d'utilisation spécifiques du Big Data.

[Découvrez le blog sur le AWS Big Data](#)

Historique du document

Le tableau suivant décrit les modifications importantes apportées à ce guide de décision. Pour recevoir des notifications concernant les mises à jour de ce guide, vous pouvez vous abonner à un flux RSS.

Modification	Description	Date
Mises à jour de re:Invent	Liens mis à jour dans le guide de décision et ajout d'Amazon Managed Workflows pour Apache Airflow.	24 septembre 2025
Mises à jour de re:Invent	Références mises à jour à Amazon SageMaker, Amazon SageMaker Unified Studio (qui n'est plus disponible en version préliminaire) et Amazon SageMaker Lakehouse tout au long du guide de décision.	9 septembre 2025
Mises à jour de re:Invent	Ajout d' SageMaker AI Unified Studio et AWS Clean Rooms. Document entièrement mis à jour avec les nouvelles fonctionnalités et capacités de l'IA.	20 février 2025
Publication initiale	Guide publié pour la première fois.	17 novembre 2023

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.