

[Se lancer gratuitement](#)[Nous contacter](#)[Qu'est-ce que le cloud computing?](#) / [Hub des concepts liés au cloud computing](#) / [Analytique](#)

Qu'est-ce qu'Apache Iceberg ?

[Voir les offres gratuites pour les services d'analyse dans le cloud](#)

Explorez les offres d'analyses gratuites

Afficher les offres gratuites pour les services d'analytique dans le cloud



Découvrez les services d'analytique

Accélérez l'innovation grâce à la gamme de services d'analytique la plus complète



Parcourir les formations en analytique

Démarrez la formation sur l'analytique avec du contenu élaboré par des experts AWS



[Se lancer gratuitement](#)[Nous contacter](#)

Qu'est-ce qu'Apache Iceberg ?

Apache Iceberg est un format de table de données 100 % open source, distribué, piloté par la communauté, sous licence Apache 2.0, qui permet de simplifier le traitement des données sur de grands jeux de données stockés dans des lacs de données. Les ingénieurs de données utilisent Apache Iceberg parce qu'il est rapide, efficace et fiable à toutes les échelles et qu'il enregistre l'évolution des jeux de données au fil du temps. Apache Iceberg permet des intégrations faciles avec des frameworks de traitement de données populaires tels qu'Apache Spark, Apache Flink, Apache Hive, Presto, etc.

Qu'est-ce qu'un lac de données transactionnel ?

Un [lac de données](#) est un référentiel centralisé qui vous permet de stocker toutes vos données structurées et non structurées à n'importe quelle échelle. Une transaction de données est une série d'échanges de données effectués en une seule opération. Par exemple, lorsqu'un client retire de l'argent d'un compte bancaire, la banque procède à plusieurs échanges de données en même temps dans le cadre d'une seule transaction de données, notamment en vérifiant que le solde du compte est suffisant, en vérifiant l'identité et en débitant le retrait du compte. Un lac de données transactionnel est un type de lac de données qui non seulement stocke des données à grande échelle, mais prend également en charge les opérations transactionnelles, garantit l'exactitude et la cohérence des données et vous permet de suivre l'évolution des données et de leur structure au fil du temps. Ces propriétés sont regroupées sous le nom d'atomicité, de cohérence, d'isolation et de durabilité (ACID) :

- L'**atomicité** garantit que chaque transaction est un événement unique qui réussit ou échoue complètement : il n'existe pas de situation intermédiaire.
- La **cohérence** garantit que toutes les données écrites sont valides conformément aux règles définies du lac de données, garantissant ainsi l'exactitude et la fiabilité des données.
- L'**isolation** garantit que plusieurs transactions peuvent avoir lieu en même temps sans interférer les unes avec les autres, garantissant ainsi que chaque transaction s'exécute indépendamment.
- La **durabilité** signifie que les données ne sont ni perdues ni corrompues une fois qu'une transaction est soumise. Les données peuvent être récupérées en cas de défaillance du système, telle qu'une panne de courant.

[Se lancer gratuitement](#)[Nous contacter](#)

Parmi les principaux avantages de l'utilisation d'Apache Iceberg pour les lacs de données transactionnels, citons :

- La familiarité de SQL : le [langage SQL \(structured query language\)](#) est un langage de requête populaire fréquemment utilisé dans tous les types d'applications. Les analystes de données et les développeurs étudient et utilisent le langage SQL parce qu'il s'intègre bien à différents langages de programmation et qu'il est également assez facile à apprendre, car il utilise des mots-clés anglais courants dans ses instructions. Apache Iceberg permet à toute personne familiarisée avec le langage SQL (structured query language) de créer des lacs de données et d'effectuer la plupart des opérations sur les lacs de données sans avoir à apprendre un nouveau langage.
- Cohérence des données : Apache Iceberg assure la cohérence des données afin de garantir que tous les utilisateurs qui lisent et écrivent sur les données voient les mêmes données.
- Structure des données : Apache Iceberg permet de modifier facilement votre structure de données, également appelée évolution du schéma, ce qui signifie que les utilisateurs peuvent ajouter, renommer ou supprimer des colonnes d'une table de données sans perturber les données sous-jacentes.
- Gestion des versions des données : Apache Iceberg prend en charge la gestion des versions des données, ce qui permet aux utilisateurs de suivre les modifications apportées aux données au fil du temps. Cela active la fonctionnalité de voyage dans le temps, qui permet aux utilisateurs d'accéder aux versions historiques des données et de les interroger et d'analyser les modifications apportées aux données entre les mises à jour et les suppressions.
- Support multiplateforme : Apache Iceberg prend en charge différents systèmes de stockage et moteurs de requêtes, notamment Apache Spark, Apache Hive et Presto. Cela facilite l'utilisation d'Iceberg dans différents environnements de traitement de données.
- Traitement progressif : Iceberg prend en charge le traitement progressif, qui permet aux utilisateurs de traiter uniquement les données qui ont changé depuis la dernière exécution, également connu sous le nom de CDC (Change Data Capture, capture de changement des données). Cela peut contribuer à améliorer l'efficacité et les performances du traitement des données.

Quels sont les cas d'utilisation courants d'Apache Iceberg ?

[Se lancer gratuitement](#)[Nous contacter](#)

enregistrements. Cela est utile lorsque votre jeu de données nécessite des mises à jour fréquentes après l'arrivée des données, par exemple, des données de vente susceptibles de changer en raison d'événements ultérieurs tels que les retours de clients. Iceberg permet de mettre à jour des enregistrements individuels sans avoir à republier le jeu de données complet.

- Les tables de données dans des lacs de données qui présentent des modifications imprévisibles, telles que les tables SCD (Slowly Changing Dimension, dimension à évolution lente). Un exemple de SCD est une table d'enregistrement client qui inclut le nom, l'emplacement et les coordonnées qui peuvent changer au fil du temps à des intervalles inconnus.
- Lorsque les transactions avec le lac de données nécessitent une validité, une durabilité et une fiabilité garanties des données, les formats de table Apache Iceberg peuvent être déployés pour assurer les transactions ACID.
- Lorsqu'il est nécessaire de remonter le temps pour interroger des versions historiques de données afin d'effectuer une analyse des tendances, d'analyser les modifications apportées aux données sur une certaine période ou de restaurer ou de revenir à une version précédente afin de corriger des problèmes.

Qui utilise Apache Iceberg ?

Les ingénieurs, administrateurs, analystes et scientifiques des données font partie des personnes qui utilisent Apache Iceberg. Les ingénieurs et administrateurs de données peuvent utiliser Apache Iceberg pour concevoir et créer des systèmes de stockage de données évolutifs. Les analystes et scientifiques des données peuvent utiliser Apache Iceberg pour analyser efficacement de grands jeux de données.

Pourquoi choisir Apache Iceberg ?

Apache Iceberg offre un moyen rapide et efficace de traiter de grands jeux de données à grande échelle. Il offre les avantages suivants :

1. Open source : Apache Iceberg est un projet open source, ce qui signifie qu'il est gratuit et peut être personnalisé pour répondre à vos besoins spécifiques. Il rassemble également une communauté active de développeurs qui améliorent continuellement le projet et y ajoutent de nouvelles fonctionnalités.

[Se lancer gratuitement](#)[Nous contacter](#)

performances des requêtes, notamment le stockage en colonnes et des techniques de compression telles que le pushdown de prédicats et l'évolution des schémas.

4. Flexibilité : Apache Iceberg vous permet de modifier l'organisation de vos données afin qu'elles puissent évoluer au fil du temps sans que vous deviez réécrire vos requêtes ou reconstruire vos structures de données. Il prend également en charge plusieurs formats de données et sources de données, ce qui facilite son intégration aux systèmes existants.
5. Fiabilité : Apache Iceberg garantit la cohérence et la fiabilité des données grâce à sa prise en charge des transactions. Vous pouvez suivre l'évolution des données au fil du temps et revenir aux versions historiques pour corriger les problèmes.

Quels services AWS prennent en charge Iceberg ?

Apache Iceberg prend en charge des frameworks de traitement de données populaires tels qu'Apache Spark, Apache Flink, Apache Hive et Presto. Les services AWS tels qu'[Amazon Athena](#), [Amazon EMR](#) et [AWS Glue](#) incluent la prise en charge native des frameworks de lacs de données transactionnels, notamment Apache Iceberg. Apache Iceberg, associé aux services AWS pris en charge, permet de créer un lac de données transactionnel, souvent basé sur le stockage dans [S3](#).

- **Amazon Athena** est un service d'analyse sans serveur et interactif basé sur des frameworks open source, prenant en charge les formats de table et de fichier ouverts. Athena permet d'analyser plusieurs téraoctets de données là où elles se trouvent, de manière simple et flexible. Athena fournit un support natif des requêtes de lecture, de voyage dans le temps, d'écriture et DDL pour les tables Apache Iceberg qui utilisent le format Apache Parquet pour les données et le catalogue AWS Glue pour leur métastore.
- **Amazon EMR** est une solution de big data pour le traitement des données à l'échelle du pétaoctet, l'analyse interactive et le machine learning à l'aide de frameworks open source, tels qu'Apache Spark, Hadoop, Presto et Hive. À partir d'Amazon EMR 6.5.0, vous pouvez utiliser Apache Spark 3 sur des clusters Amazon EMR au format de table Iceberg. Les frameworks EMR tels que Spark, Trino, Flink et Hive prennent en charge Apache Iceberg.
- **AWS Glue** est un service d'intégration des données sans serveur qui facilite la découverte, la préparation, le déplacement et l'intégration des données depuis des sources multiples pour l'analyse, le machine learning (ML) et le développement des applications. AWS Glue 3.0 et versions ultérieures prennent en charge le framework Apache Iceberg pour les lacs de données. Vous pouvez utiliser AWS Glue pour effectuer des opérations de lecture et d'écriture sur des tables Iceberg dans Amazon S3, ou travailler avec des tables Iceberg à l'aide du Catalogue de

[Se lancer gratuitement](#)[Nous contacter](#)

Étapes suivantes sur AWS



Consultez d'autres ressources liées aux produits

[Consulter les services d'analytique »](#)

Créer un compte gratuit

Obtenez un accès instantané à l'offre gratuite AWS.

[S'inscrire »](#)

Commencez à créer sur la console

Démarrez la création dans la console de gestion AWS.

[Se connecter »](#)[Connectez-vous à la console](#)

Se renseigner sur AWS

[Qu'est-ce qu'AWS ?](#)[Qu'est-ce que le cloud computing ?](#)

Ressources pour AWS

[Mise en route](#)[Formation et certification](#)[Bibliothèque de solutions AWS](#)[Centre d'architecture](#)

[Se lancer gratuitement](#)[Nous contacter](#)[Sécurité dans le Cloud AWS](#)[Nouveautés](#)[Blogs](#)[Communiqués de presse](#)

Développeurs sur AWS Aide

[Centre pour développeurs](#)[Kits SDK et outils](#)[.NET sur AWS](#)[Python sur AWS](#)[Java sur AWS](#)[PHP sur AWS](#)[JavaScript sur AWS](#)[Contactez-nous](#)[Obtenez l'aide d'experts](#)[Soumettre un ticket de support](#)[AWS re:Post](#)[Centre de connaissances](#)[Présentation d'AWS Support](#)[Domaine juridique](#)[Emplois AWS](#)[Créer un compte AWS](#)

Amazon est un employeur qui souscrit aux principes d'équité en matière d'emploi : *minorités, femmes, handicaps, seniors, identité de genre, orientation sexuelle, âge.*

Langue

[عربي |](#)[Bahasa Indonesia |](#)[Deutsch |](#)[English |](#)[Español |](#)[Français |](#)[Italiano |](#)[Português |](#)[Tiếng Việt |](#)[Türkçe |](#)



Se lancer gratuitement

Nous contacter

[中文 \(繁體\)](#)

[Confidentialité](#)

|

[Conditions d'utilisation du site](#)

|

[Préférences des cookies](#)

|

© 2023, Amazon Web Services, Inc. ou ses sociétés apparentées. Tous droits réservés.