Guide du développeur

AWS Lake Formation

AWS Lake Formation: Guide du développeur

Table of Contents

Qu'est-ce que c'est AWS Lake Formation ? 1
Caractéristiques de Lake Formation 2
Ingestion et gestion des données 2
Gestion de la sécurité 3
Partage des données 4
Comment ça marche 5
Flux de travail de gestion des autorisations de Lake Formation 5
Autorisations relatives aux métadonnées 7
Gestion de l'accès au stockage 10
Partage de données entre comptes dans Lake Formation 12
Composantes de la Lake Formation 13
Console Lake Formation 13
API et interface de ligne de commande de Lake Formation 13
Autres AWS services 13
Terminologie des Lake Formation 14
Lac de données 14
Accès aux données 14
Mode d'accès hybride 14
Plan 15
Flux de travail 15
Catalogue de données 15
Données sous-jacentes 16
Principal 16
Administrateur du lac de données 16
AWS intégrations de services avec Lake Formation 17
Ressources supplémentaires sur la Formation du Lac 19
Blogs 19
Discussions techniques et webinaires 19
Architecture moderne 19
Ressources de maillage de données 20
Guides des meilleures pratiques 20
Débuter avec Lake Formation 20
Premiers pas 21
Exécution des tâches AWS de configuration initiale 21

Inscrivez-vous pour un Compte AWS 21
Création d'un utilisateur doté d'un accès administratif 22
Octroi d’un accès par programmation 23
Configurez AWS Lake Formation 25
Configurer les ressources de Lake Formation à l'aide d' AWS CloudFormation un modèle 26
Création d'un administrateur de lac de données 27
Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride 32
Attribuer des autorisations aux utilisateurs de Lake Formation 34
Configurer un emplacement Amazon S3 pour votre lac de données 35
(Facultatif) Paramètres de filtrage des données externes 36
(Facultatif) Accordez l'accès à la clé de chiffrement du catalogue de données 37
(Facultatif) Créez un rôle IAM pour les flux de travail 37
Mise à niveau des autorisations de AWS Glue données pour le modèle Lake Formation 39
À propos de la mise à niveau vers le modèle d'autorisations de Lake Formation 40
Étape 1 : Répertorier les autorisations existantes 41
Étape 2 : configurer les autorisations de Lake Formation 44
Étape 3 : octroyer aux utilisateurs des autorisations IAM 44
Étape 4 : passer au modèle d'autorisations de Lake Formation 45
Étape 5 : Sécuriser les nouvelles ressources du catalogue de données 48
Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM 49
Étape 7 : Nettoyer les politiques IAM existantes 50
Configuration des points de terminaison Amazon VPC ()AWS PrivateLink 50
Considérations relatives aux points d'extrémité du VPC de Lake Formation 51
Création d'un point de terminaison VPC d'interface pour Lake Formation 51
Création d'une politique de point de terminaison VPC pour Lake Formation 52
Didacticiels 54
Création d'un lac de données à partir d'une AWS CloudTrail source 55
Public visé 57
Prérequis 57
Étape 1 : créer un utilisateur d'analyste de données 58
Étape 2 : ajouter des autorisations pour lire les AWS CloudTrail journaux au rôle de flux de travail 59
Étape 3 : créer un compartiment Amazon S3 pour le lac de données 59
Étape 4 : enregistrer un chemin Amazon S3 60
Étape 5 : accorder des autorisations de localisation des données 60
Étape 6 : Création d'une base de données dans le catalogue de données 61

Étape 7 : Accorder des autorisations de données 61
Étape 8 : Utiliser un plan pour créer un flux de travail 64
Étape 9 : Exécuter le flux de travail 65
Étape 10 : Autorisez SELECT sur les tables 66
Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena 67
Création d'un lac de données à partir d'une source JDBC 68
Public visé 68
Prérequis 69
Étape 1 : créer un utilisateur d'analyste de données 70
Étape 2 : créer une connexion dans AWS Glue 71
Étape 3 : créer un compartiment Amazon S3 pour le lac de données 72
Étape 4 : enregistrer un chemin Amazon S3 72
Étape 5 : accorder des autorisations de localisation des données 72
Étape 6 : Création d'une base de données dans le catalogue de données 73
Étape 7 : Accorder des autorisations de données 73
Étape 8 : Utiliser un plan pour créer un flux de travail 74
Étape 9 : Exécuter le flux de travail 76
Étape 10 : Autorisez SELECT sur les tables 77
Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena 77
Étape 12 : interroger les données du lac de données à l'aide d'Amazon Redshift Spectrum ... 78
Étape 13 : Accorder ou révoquer les autorisations de Lake Formation à l'aide d'Amazon Redshift Spectrum 82
Configuration des autorisations pour les formats de table ouverts dans Lake Formation 83
Public visé 84
Prérequis 84
Étape 1 : Approvisionnez vos ressources 86
Étape 2 : configurer les autorisations pour une table Iceberg 87
Étape 3 : configurer les autorisations pour une table Hudi 94
Étape 4 : configurer les autorisations pour une table Delta Lake 97
Étape 5 : Nettoyer les AWS ressources 99
Gestion d'un lac de données à l'aide d'un contrôle d'accès basé sur des balises 99
Public visé 101
Prérequis 102
Étape 1 : Approvisionnez vos ressources 103
Étape 2 : enregistrez l'emplacement de vos données, créez une ontologie LF-Tag et accordez des autorisations 104

Étape 3 : Création de bases de données Lake Formation 108
Étape 4 : Accorder les autorisations relatives aux tables 118
Étape 5 : exécuter une requête dans Amazon Athena pour vérifier les autorisations 120
Étape 6 : Nettoyer les AWS ressources 121
Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes 122
Public visé 122
Prérequis 123
Étape 1 : Approvisionnez vos ressources 124
Étape 2 : Requête sans filtres de données 125
Étape 3 : configurer les filtres de données et accorder des autorisations 127
Étape 4 : Requête à l'aide de filtres de données 129
Étape 5 : Nettoyer les AWS ressources 131
Partagez vos données en toute sécurité à l'aide de Lake Formation 131
Public visé 132
Configurer les paramètres de Lake Formation 134
Étape 1 : provisionnez vos ressources à l'aide AWS CloudFormation de modèles 136
Étape 2 : Conditions préalables au partage entre comptes de Lake Formation 139
Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises 142
Étape 4 : Implémentation de la méthode de ressource nommée 149
Étape 5 : Nettoyer les AWS ressources 153
Partage des ressources du catalogue de données avec des tiers à Comptes AWS l'aide d'un contrôle d'accès précis 153
Public visé 155
Prérequis 155
Étape 1 : Fournir un accès détaillé à un autre compte 156
Étape 2 : fournir un accès détaillé à un utilisateur du même compte 158
Permissions d'intégration à Lake Formation 160
Vue d'ensemble des autorisations relatives à Lake Formation 161
Méthodes de contrôle d'accès précis 163
Contrôle d'accès aux métadonnées 167
Contrôle d'accès aux données sous-jacent 171
Référence des personnalités de Lake Formation et des autorisations IAM 177
AWS Lake Formation personas 177
AWS politiques gérées pour Lake Formation 179
Autorisations suggérées par Persona 186

Modification des paramètres par défaut de votre lac de données 196
Permissions implicites de Lake Formation 200
Référence des autorisations de Lake Formation 201
Permissions de Lake Formation par type de ressource 202
Lake Formation accorde et AWS CLI révoque des commandes 204
Permissions de Lake Formation 209
Intégration d'IAM Identity Center 224
Prérequis 226
Connecter Lake Formation à l'IAM Identity Center 229
Mise à jour d'une intégration à IAM Identity Center 233
Suppression d'une connexion Lake Formation avec IAM Identity Center 235
Octroi d'autorisations aux utilisateurs et aux groupes 235
Ajouter un emplacement Amazon S3 à votre lac de données 239
Exigences relatives aux rôles utilisés pour enregistrer des sites 240
Enregistrement d'un emplacement Amazon S3 247
Enregistrement d'un emplacement Amazon S3 chiffré 251
Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte 256
Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes 259
Annulation de l'enregistrement d'un site Amazon S3 263
Mode d'accès hybride 264
Cas d'utilisation courants du mode d'accès hybride 266
Comment fonctionne le mode d'accès hybride 268
Configuration du mode d'accès hybride : scénarios courants 269
Supprimer les principes et les ressources du mode d'accès hybride 287
Affichage des principes et des ressources en mode d'accès hybride 288
Ressources supplémentaires 289
Création de tables et de bases de données du catalogue de données 290
Création d’une base de données 290
Création de tables 291
Utilisation des vues 312
Importation de données à l'aide de workflows 318
Plans et flux de travail 318
Création d'un flux de travail 320
Exécution d'un flux de travail 324
Gestion des autorisations relatives à Lake Formation 326
Octroi d'autorisations de localisation des données 326

Octroi d'autorisations de localisation des données (même compte) 327
Octroi d'autorisations de localisation des données (compte externe) 330
Octroi d'autorisations sur un emplacement de données partagé avec votre compte 333
Octroi et révocation des autorisations du catalogue de données 334
Autorisations IAM requises pour accorder les autorisations de Lake Formation 336
Octroi d'autorisations de data lake à l'aide de la méthode de ressource nommée 339
Contrôle d’accès basé sur les étiquettes 360
Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC 409
Exemple de scénario d'autorisations 416
Filtrage des données et sécurité au niveau des cellules 418
Vue d'ensemble du filtrage des données 418
Filtres de données 420
Support partiQL dans les expressions de filtre de ligne 424
Remarques et restrictions relatives au filtrage au niveau des colonnes 426
Autorisations requises pour interroger des tables avec filtrage au niveau des cellules 428
Gestion des filtres de données 429
Affichage des autorisations de base de données et de tables 444
Révocation des autorisations à l'aide de la console 449
Partage de données entre comptes 449
Prérequis 452
Mise à jour des paramètres de version de partage de données entre comptes 457
Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des responsables IAM 463
Octroi d'autorisations sur une base de données ou une table partagée avec votre compte .. 466
Octroi d'autorisations relatives aux liens vers 468
Accès aux données sous-jacentes d'une table partagée 471
Journalisation entre comptes CloudTrail 472
Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation 477
Afficher toutes les subventions entre comptes à l'aide de l'opération GetResourceShares API 481
Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci 482
Accepter une invitation à partager des AWS RAM ressources 484
Affichage des tables et des bases de données partagées du catalogue de données 486
Création de liens vers des ressources 488

Fonctionnement des liens vers des ressources 488
Création d'un lien de ressource vers une table partagée 491
Création d'un lien de ressource vers une base de données partagée 495
Gestion des liens de ressources dans les AWS Glue API 499
Accès aux tables dans toutes les régions 503
Flux de travail 504
Configuration de l'accès aux tables entre régions 508
Partage de données dans le Lake Formation 512
Gestion des autorisations pour les données dans un partage de données Amazon Redshift 513
Prérequis 514
Configuration des autorisations pour les partages de données Amazon Redshift 515
Interrogation de bases de données fédérées 519
Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes 520
Flux de travail 523
Prérequis 524
Connexion du catalogue de données à un métastore Hive externe 526
Ressources supplémentaires 530
Sécurité 531
Protection des données 531
Chiffrement au repos 533
Sécurité de l’infrastructure 533
Prévention du problème de l’adjoint confus entre services 534
Connexion aux événements de sécurité AWS Lake Formation 535
Intégration à Lake Formation 537
Utilisation de l'intégration de l'application Lake Formation 537
Comment fonctionne l'intégration de l'application Lake Formation 538
Rôles et responsabilités dans l'intégration de l'application Lake Formation 540
Lake Formationflux de travail pour les opérations d'API d'intégration d'applications 541
Enregistrement d'un moteur de requête tiers 543
Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications 544
Intégration des applications pour un accès complet aux tables 549
Collaboration avec d'autres AWS services 552
Amazon Athena 556
Support pour les formats de tables transactionnels 558

Ressources supplémentaires 561
Amazon Redshift Spectrum 561
Support pour les types de tables transactionnels 562
Ressources supplémentaires 563
AWS Glue 564
Support pour les types de tables transactionnels 565
Ressources supplémentaires 566
Amazon EMR 566
Support pour les formats de tables transactionnels 567
Ressources supplémentaires 568
Amazon QuickSight 569
Ressources supplémentaires 569
AWS CloudTrail lac 569
Appels d'API Logging AWS Lake Formation à l'aide de AWS CloudTrail 571
Informations sur la formation des lacs en CloudTrail 571
Comprendre les événements liés à la formation des lacs 572
Meilleures pratiques, considérations et limites en matière de formation des lacs 575
Meilleures pratiques et considérations relatives au partage de données entre comptes 575
Limites d'accès aux données entre régions 578
Considérations et limites relatives aux affichages du catalogue de données 578
Limites du filtrage des données 579
Considérations et limites relatives au mode d'accès hybride 581
Considérations et limites relatives au partage des données du magasin de métadonnées Hive 583
Limites du partage de données Amazon Redshift 584
Limites de l'intégration d'IAM Identity Center 586
Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation 587
Formats pris en charge et limites pour le compactage géré des données 590
Résolution des problèmes liés à la formation du Lake 592
Résolution de problème généraux 592
Erreur : Autorisations insuffisantes pour Lake Formation sur <Amazon S3 location> 592
Erreur : « Autorisations de clé de chiffrement insuffisantes pour l'API Glue » 593
Ma requête Amazon Athena ou celle d'Amazon Redshift qui utilise des manifestes échoue . 593
Erreur : « Autorisations de formation lacustres insuffisantes : création d'une balise requise dans le catalogue » 593

Erreur lors de la suppression d'administrateurs de lacs de données non valides 593
Résolution des problèmes d'accès entre comptes 593
J'ai accordé l'autorisation d'utiliser plusieurs comptes Lake Formation, mais le destinataire ne peut pas voir la ressource 594
Les utilisateurs principaux du compte destinataire peuvent voir la ressource du catalogue de données, mais ne peuvent pas accéder aux données sous-jacentes 595
Erreur : « L'association a échoué car l'appelant n'était pas autorisé » lors de l'acceptation d'une invitation de partage de AWS RAM ressources 595
Erreur : « Non autorisé à accorder des autorisations pour la ressource » 596
Erreur : « Accès refusé pour récupérer les informations de AWS l'organisation » 596
Erreur : « Organisation <organization-ID>introuvable » 596
Erreur : « Permissions de formation lacustres insuffisantes : combinaison illégale » 596
ConcurrentModificationException sur les demandes d'accord/de révocation adressées à des comptes externes 596
Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes 597
Résolution des problèmes liés aux plans et aux flux de travail 598
<role-ARN>Mon plan a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à exécuter : iam : PassRole on resource : » 598
<role-ARN>Mon flux de travail a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à effectuer : iam : PassRole on resource : » 599
Un crawler de mon flux de travail a échoué avec le message « La ressource n'existe pas ou le demandeur n'est pas autorisé à accéder aux autorisations demandées » 599
Un crawler de mon flux de travail a échoué avec « Une erreur s'est produite (AccessDeniedException) lors de l'appel de l' CreateTable opération... » 599
Problèmes connus pour AWS Lake Formation 599
Limitation du filtrage des métadonnées des tables 600
Problème lié au changement de nom d'une colonne exclue 601
Problème lié à la suppression de colonnes dans les tableaux CSV 601
Les partitions de table doivent être ajoutées sous un chemin commun 601
Problème lié à la création d'une base de données pendant la création du flux de travail 602
Problème lié à la suppression puis à la recréation d'un utilisateur 602
`GetTable`set `SearchTables` les API ne mettent pas à jour la valeur du `IsRegisteredWithLakeFormation` paramètre 602
Les opérations de l'API Data Catalog ne mettent pas à jour la valeur du `IsRegisteredWithLakeFormation` paramètre 603

Les opérations de Lake Formation ne prennent pas en charge AWS Glue le registre des schémas 603
Message d'erreur mis à jour 603
Lake Formation API 604
Autorisations 605
— operations — 605
— les types de données — 605
Paramètres du lac de données 606
— operations — 606
— les types de données — 606
Intégration à IAM Identity Center 606
— operations — 606
— les types de données — 606
Mode d'accès hybride 607
— operations — 607
— les types de données — 605
Vente d'informations d'identification 607
— operations — 607
— les types de données — 608
Identification 608
— operations — 608
— les types de données — 608
API de filtrage de données 609
— operations — 609
— les types de données — 609
Types de données courants 609
ErrorDetail 609
Modèles de chaîne 610
Régions prises en charge 611
Disponibilité générale 611
AWS GovCloud (US) 611
Optimisation des transactions et du stockage 611
Historique du document 614
AWS Glossaire 628
..... dcxxix

Qu'est-ce que c'est AWS Lake Formation ?

Bienvenue dans le guide du AWS Lake Formation développeur.

AWS Lake Formation vous permet de gérer, de sécuriser et de partager les données de manière centralisée à l'échelle mondiale à des fins d'analyse et d'apprentissage automatique. Avec Lake Formation, vous pouvez gérer un contrôle d'accès précis pour les données de vos lacs de données sur Amazon Simple Storage Service (Amazon S3) et ses métadonnées dans. AWS Glue Data Catalog

Lake Formation fournit son propre modèle d'autorisations qui complète le modèle d'autorisations IAM. Le modèle d'autorisations de Lake Formation permet un accès précis aux données stockées dans des lacs de données par le biais d'un simple mécanisme d'autorisation ou de révocation, un peu comme un système de gestion de base de données relationnelle (RDBMS). Les autorisations de Lake Formation sont appliquées à l'aide de contrôles granulaires au niveau des colonnes, des lignes et des cellules dans les services d' AWS analyse et d'apprentissage automatique, notamment Amazon Athena, Amazon Redshift Spectrum, Amazon QuickSight Amazon EMR et. AWS Glue

Le mode d'accès hybride Lake Formation pour vous AWS Glue Data Catalog permet de sécuriser et d'accéder aux données cataloguées en utilisant à la fois les autorisations Lake Formation et les politiques d'autorisations IAM pour Amazon S3 et AWS Glue les actions. Grâce au mode d'accès hybride, les administrateurs de données peuvent intégrer les autorisations de Lake Formation de manière sélective et progressive, en se concentrant sur un cas d'utilisation de lac de données à la fois.

Lake Formation vous permet également de partager des données en interne et en externe entre plusieurs AWS organisations ou directement avec les responsables d'IAM sur un autre compte Comptes AWS, offrant ainsi un accès détaillé aux AWS Glue Data Catalog métadonnées et aux données sous-jacentes.

Rubriques

- Caractéristiques de Lake Formation
- AWS Lake Formation : comment ça marche
- Composantes de la Lake Formation
- Terminologie des Lake Formation
- AWS intégrations de services avec Lake Formation

- Ressources supplémentaires sur la Formation du Lac
- Débuter avec Lake Formation

Caractéristiques de Lake Formation

Lake Formation vous aide à décloisonner les données et à combiner différents types de données structurées et non structurées dans un référentiel centralisé. Tout d'abord, identifiez les banques de données existantes dans Amazon S3 ou dans les bases de données relationnelles et NoSQL, puis déplacez les données vers votre lac de données. Ensuite, analysez, cataloguez et préparez les données à des fins d'analyse. Ensuite, offrez à vos utilisateurs un accès sécurisé en libre-service aux données grâce aux services d'analyse de leur choix.

Rubriques

- Ingestion et gestion des données
- Gestion de la sécurité
- Partage des données

Ingestion et gestion des données

Importer des données depuis des bases de données déjà présentes AWS

Une fois que vous avez indiqué où se trouvent vos bases de données existantes et que vous avez fourni vos identifiants d'accès, Lake Formation lit les données et leurs métadonnées (schéma) pour comprendre le contenu de la source de données. Il importe ensuite les données dans votre nouveau lac de données et enregistre les métadonnées dans un catalogue central. Avec Lake Formation, vous pouvez importer des données depuis des bases de données MySQL, PostgreSQL, SQL Server, MariaDB et Oracle exécutées sur Amazon RDS ou hébergées sur Amazon EC2. Le chargement de données en masse et incrémentiel est pris en charge.

Importer des données depuis d'autres sources externes

Vous pouvez utiliser Lake Formation pour déplacer des données depuis des bases de données locales en vous connectant à Java Database Connectivity (JDBC). Identifiez vos sources cibles et fournissez des informations d'accès dans la console. Lake Formation lit et charge vos données dans le lac de données. Pour importer des données à partir de bases de données autres que celles répertoriées ci-dessus, vous pouvez créer des tâches ETL personnalisées avec AWS Glue.

Cataloguez et étiquetez vos données

Vous pouvez utiliser des AWS Glue robots d'exploration pour lire vos données dans Amazon S3, extraire le schéma de base de données et de table et stocker ces données dans un outil de recherche AWS Glue Data Catalog. Utilisez ensuite Lake Formation Contrôle d'accès basé sur des balises Lake Formation (TBAC) pour gérer les autorisations sur les bases de données, les tables et les colonnes. Pour plus d'informations sur l'ajout de tables au catalogue de données, consultezCréation de tables et de bases de données du catalogue de données.

Gestion de la sécurité

Définissez et gérez les contrôles d'accès

Lake Formation fournit un endroit unique pour gérer les contrôles d'accès aux données de votre lac de données. Vous pouvez définir des politiques de sécurité qui limitent l'accès aux données au niveau de la base de données, de la table, de la colonne, de la ligne et de la cellule. Ces politiques s'appliquent aux utilisateurs et aux rôles IAM, ainsi qu'aux utilisateurs et aux groupes lors de la fédération via un fournisseur d'identité externe. Vous pouvez utiliser des contrôles précis pour accéder aux données sécurisées par Lake Formation dans Amazon Redshift Spectrum, Athena, ETL et Amazon EMR pour AWS Glue Apache Spark. Chaque fois que vous créez des identités IAM, veillez à suivre les meilleures pratiques IAM. Pour plus d'informations, consultez la section Bonnes pratiques en matière de sécurité dans le guide de l'utilisateur IAM.

Mode d'accès hybride

Le mode d'accès hybride Lake Formation offre la flexibilité nécessaire pour activer de manière sélective les autorisations Lake Formation pour les bases de données et les tables de votre AWS Glue Data Catalog. Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants. Pour plus d’informations, consultez Mode d'accès hybride.

Mettre en œuvre l'enregistrement des audits

Lake Formation fournit des journaux d'audit complets CloudTrail pour surveiller l'accès et démontrer la conformité aux politiques définies de manière centralisée. Vous pouvez auditer l'historique des accès aux données par le biais de services d'analyse et d'apprentissage automatique qui lisent les données de votre lac de données via Lake Formation. Cela vous permet de voir quels utilisateurs ou quels rôles ont tenté d'accéder à quelles données, avec quels services et à quel moment. Vous

pouvez accéder aux journaux d'audit de la même manière que vous accédez à tous les autres CloudTrail journaux à l'aide des CloudTrail API et de la console. Pour plus d'informations sur les CloudTrail journaux, consultezAppels d'API Logging AWS Lake Formation à l'aide de AWS CloudTrail.

Sécurité au niveau des lignes et des cellules

Lake Formation fournit des filtres de données qui vous permettent de restreindre l'accès à une combinaison de colonnes et de lignes. Utilisez la sécurité au niveau des lignes et des cellules pour protéger les données sensibles telles que les informations personnelles identifiables (PII). Pour plus d'informations sur la sécurité au niveau des lignes, consultez. Vue d'ensemble du filtrage des données

Contrôle d’accès basé sur les étiquettes

Utilisez le contrôle d'accès basé sur les balises Lake Formation pour gérer des centaines, voire des milliers d'autorisations de données en créant des étiquettes personnalisées appelées balises LF. Vous pouvez désormais définir des balises LF et les associer à des bases de données, à des tables ou à des colonnes. Partagez ensuite l'accès contrôlé entre les services d'analyse, d'apprentissage automatique (ML) et d'extraction, de transformation et de chargement (ETL) à des fins de consommation. Les balises LF permettent d'étendre facilement la gouvernance des données en remplaçant les définitions de politiques de milliers de ressources par quelques balises logiques. Lake Formation propose une recherche textuelle sur ces métadonnées, afin que vos utilisateurs puissent trouver rapidement les données à analyser.

Accès entre comptes

Les fonctionnalités de gestion des autorisations de Lake Formation simplifient la sécurisation et la gestion des lacs de données distribués sur plusieurs AWS comptes grâce à une approche centralisée, fournissant un contrôle d'accès précis au catalogue de données et aux sites Amazon S3. Pour plus d’informations, consultez Partage de données entre comptes dans Lake Formation.

Partage des données

La fonctionnalité de partage de données vous permet de configurer des autorisations sur des ensembles de données stockés dans différentes sources de données telles qu'Amazon Redshift sans migrer les données ou les métadonnées vers Amazon S3 ou. AWS Glue Data Catalog Vous pouvez utiliser les méthodes suivantes pour partager des données dans Lake Formation :

Pour plus d'informations, voir Partage de données dans Lake Formation.

- Intégration de Lake Formation au partage de données Amazon Redshift : utilisez Lake Formation pour gérer de manière centralisée les autorisations d'accès aux bases de données, aux tables, aux colonnes et aux lignes des partages de données Amazon Redshift et pour restreindre l'accès des utilisateurs aux objets d'un partage de données.
- Connexion AWS Glue Data Catalog à des métastores externes : connectez-vous AWS Glue Data Catalog à des métastores externes pour gérer les autorisations d'accès aux ensembles de données dans Amazon S3 à l'aide de Lake Formation. Aucune migration de métadonnées vers le AWS Glue Data Catalog n'est nécessaire.

  Pour plus d’informations, consultez Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes.
- Intégrer Lake Formation à AWS Data Exchange — Lake Formation prend en charge l'octroi de licences d'accès à vos données via AWS Data Exchange. Si vous souhaitez obtenir une licence pour vos données de Lake Formation, consultez AWS Data Exchange le guide de l'AWS Data Exchange utilisateur.

AWS Lake Formation : comment ça marche

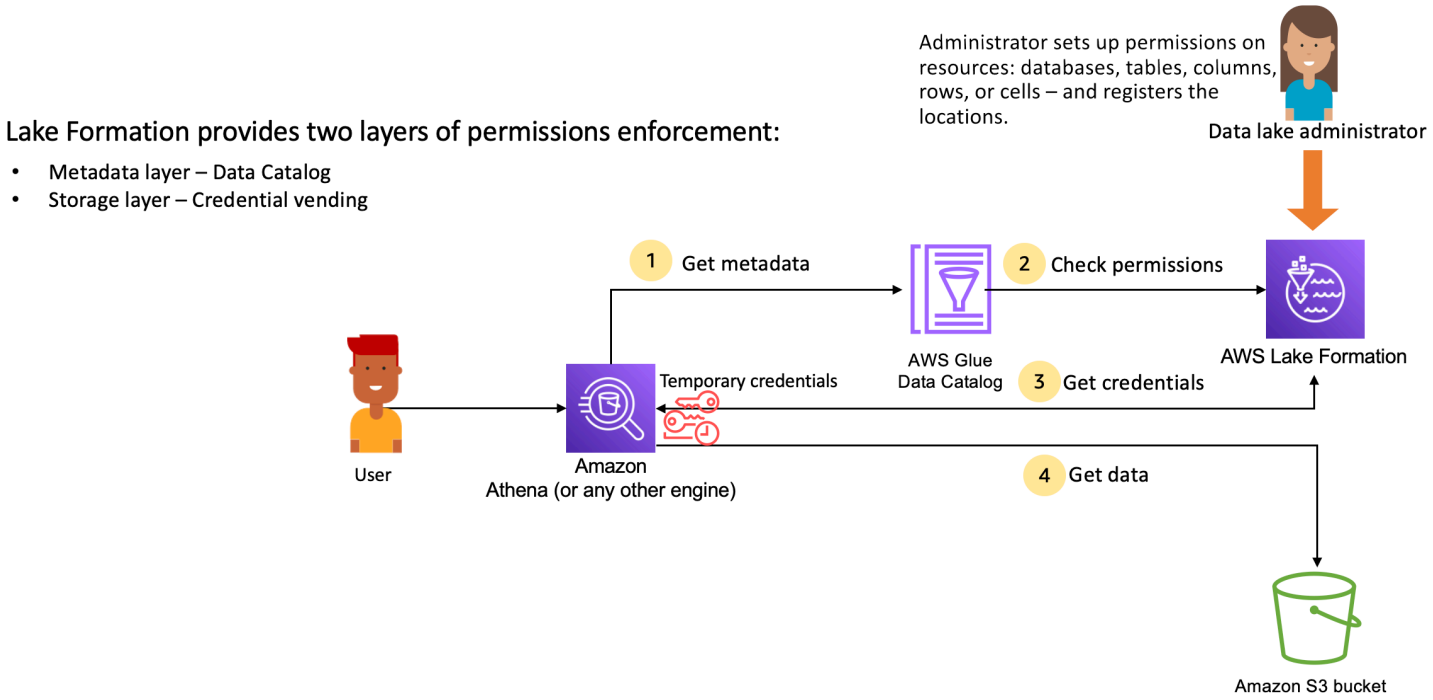
AWS Lake Formation fournit un modèle d'autorisations du système de gestion de base de données relationnelle (RDBMS) pour accorder ou révoquer l'accès aux ressources du catalogue de données telles que les bases de données, les tables et les colonnes contenant des données sous-jacentes dans Amazon S3. Les autorisations Lake Formation, faciles à gérer, remplacent les politiques complexes relatives aux compartiments Amazon S3 et les politiques IAM correspondantes.

Dans Lake Formation, vous pouvez implémenter des autorisations à deux niveaux :

- Application des autorisations au niveau des métadonnées sur les ressources du catalogue de données, telles que les bases de données et les tables
- Gestion des autorisations d'accès au stockage sur les données sous-jacentes stockées dans Amazon S3 pour le compte de moteurs intégrés

Flux de travail de gestion des autorisations de Lake Formation

Lake Formation s'intègre aux moteurs d'analyse pour interroger les magasins de données Amazon S3 et les objets de métadonnées enregistrés auprès de Lake Formation. Le schéma suivant illustre le fonctionnement de la gestion des autorisations dans Lake Formation.



Étapes de haut niveau de la gestion des autorisations de Lake Formation

Avant que Lake Formation ne puisse fournir des contrôles d'accès aux données de votre lac de données, un administrateur de lac de données ou un utilisateur disposant d'autorisations administratives définit des politiques utilisateur individuelles pour les tables du catalogue de données afin d'autoriser ou de refuser l'accès aux tables du catalogue de données à l'aide des autorisations de Lake Formation.

Ensuite, l'administrateur du lac de données ou un utilisateur délégué par l'administrateur accorde des autorisations Lake Formation aux utilisateurs sur les bases de données et les tables du catalogue de données, et enregistre l'emplacement de la table sur Amazon S3 auprès de Lake Formation.

1. Obtenir des métadonnées — Un principal (utilisateur) envoie une requête ou un script ETL à un moteur d'analyse intégré tel qu'Amazon Athena, AWS Glue Amazon EMR ou Amazon Redshift Spectrum. Le moteur d'analyse intégré identifie la table demandée et envoie une demande de métadonnées au catalogue de données.
2. Vérifier les autorisations — Le catalogue de données vérifie les autorisations de l'utilisateur auprès de Lake Formation, et si l'utilisateur est autorisé à accéder à la table, renvoie les métadonnées qu'il est autorisé à voir au moteur.
3. Obtenir des informations d'identification — Le catalogue de données indique au moteur si la table est gérée par Lake Formation ou non. Si les données sous-jacentes sont enregistrées auprès de

Lake Formation, le moteur d'analyse demande à Lake Formation de fournir un accès temporaire aux données.

4. Obtenir des données — Si l'utilisateur est autorisé à accéder à la table, Lake Formation fournit un accès temporaire au moteur d'analyse intégré. À l'aide de l'accès temporaire, le moteur d'analyse extrait les données d'Amazon S3 et effectue le filtrage nécessaire, tel que le filtrage par colonne, ligne ou cellule. Lorsque le moteur a terminé d'exécuter la tâche, il renvoie les résultats à l'utilisateur. Ce processus s'appelle la vente d'informations d'identification.

   Si la table n'est pas gérée par Lake Formation, le deuxième appel du moteur d'analyse est directement envoyé à Amazon S3. La politique de compartiment Amazon S3 et la politique utilisateur IAM concernées sont évaluées pour l'accès aux données.

   Chaque fois que vous utilisez des politiques IAM, veillez à respecter les bonnes pratiques IAM. Pour plus d'informations, consultez la rubrique Bonnes pratiques IAM du Guide de l'utilisateur IAM.

Rubriques

- Autorisations relatives aux métadonnées
- Gestion de l'accès au stockage
- Partage de données entre comptes dans Lake Formation

Autorisations relatives aux métadonnées

Lake Formation fournit l'autorisation et le contrôle d'accès au catalogue de données. Lorsqu'un rôle IAM appelle l'API du catalogue de données depuis n'importe quel système, le catalogue de données vérifie les autorisations de données de l'utilisateur et renvoie uniquement les métadonnées auxquelles l'utilisateur est autorisé à accéder. Par exemple, si un rôle IAM n'a accès qu'à une seule table dans une base de données et qu'un service ou un utilisateur assumant le rôle effectue l'`GetTables`opération, la réponse ne contiendra qu'une seule table, quel que soit le nombre de tables de la base de données.

Paramètres par défaut - autorisations **`IAMAllowedPrincipal`** de groupe

AWS Lake Formation, par défaut, attribue des autorisations à toutes les bases de données et tables à un groupe virtuel nomméIAMAllowedPrincipal. Ce groupe est unique et visible uniquement au sein de Lake Formation. Le `IAMAllowedPrincipal` groupe inclut tous les principaux IAM qui ont accès aux ressources du catalogue de données via les politiques principales IAM et AWS Glue les

politiques de ressources. Si cette autorisation existe sur une base de données ou une table, tous les principaux auront accès à la base de données ou à la table.

Si vous souhaitez fournir des autorisations plus détaillées sur une base de données ou une table, supprimez `IAMAllowedPrincipal` l'autorisation et Lake Formation appliquera toutes les autres politiques associées à cette base de données ou table. Par exemple, s'il existe une politique qui permet à l'utilisateur A d'accéder à la base de données A avec `DESCRIBE` des autorisations, et `IAMAllowedPrincipal` qu'elle existe avec toutes les autorisations, l'utilisateur A continuera à effectuer toutes les autres actions jusqu'à ce que l'`IAMAllowedPrincipal`autorisation soit révoquée.

En outre, par défaut, le `IAMAllowedPrincipal` groupe dispose d'autorisations sur toutes les nouvelles bases de données et tables lors de leur création. Il existe deux configurations qui contrôlent ce comportement. Le premier est au niveau du compte et de la région, ce qui permet cela pour les bases de données nouvellement créées, et le second au niveau de la base de données. Pour modifier le paramètre par défaut, voirModifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride.

Octroi d’autorisations

Les administrateurs des lacs de données peuvent accorder des autorisations de catalogue de données aux principaux afin que ceux-ci puissent créer et gérer des bases de données et des tables, et accéder aux données sous-jacentes.

Autorisations au niveau de la base de données et de la table

Lorsque vous accordez des autorisations au sein de Lake Formation, le concédant doit spécifier le principal auquel les autorisations doivent être accordées, les ressources pour lesquelles les autorisations doivent être accordées et les actions que le bénéficiaire doit avoir accès pour effectuer. Pour la plupart des ressources de Lake Formation, la liste principale et les ressources auxquelles accorder des autorisations sont similaires, mais les actions qu'un bénéficiaire peut effectuer varient en fonction du type de ressource. Par exemple, `SELECT` des autorisations sont disponibles pour les tables pour lire les tables, mais `SELECT` pas pour les bases de données. L'`CREATE_TABLE`autorisation est autorisée sur les bases de données, mais pas sur les tables.

Vous pouvez accorder AWS Lake Formation des autorisations de deux manières :

- Méthode de ressource nommée : vous permet de choisir les noms de base de données et de tables tout en accordant des autorisations aux utilisateurs.

- Contrôle d'accès basé sur les balises LF (LF-TBAC) : les utilisateurs créent des balises LF, les associent aux ressources du catalogue de données, accordent des autorisations sur les balises LF, associent des autorisations à des utilisateurs individuels et rédigent des politiques d'`Describe`autorisation LF à l'aide de balises LF destinées à différents utilisateurs. Ces politiques basées sur des balises LF s'appliquent à toutes les ressources du catalogue de données associées à ces valeurs de balises LF.

  > **Note**
  >
  > Les balises LF sont propres à Lake Formation. Ils ne sont visibles que dans Lake Formation et ne doivent pas être confondus avec les balises de AWS ressources.

  Le LF-TBAC est une fonctionnalité qui permet aux utilisateurs de regrouper les ressources dans des catégories définies par l'utilisateur de balises LF et d'appliquer des autorisations à ces groupes de ressources. C'est donc le meilleur moyen d'étendre les autorisations à un grand nombre de ressources du catalogue de données.

  Pour plus d’informations, consultez Contrôle d'accès basé sur des balises Lake Formation.

Lorsque vous accordez des autorisations à un directeur, Lake Formation évalue les autorisations comme une union de toutes les politiques relatives à cet utilisateur. Par exemple, si vous avez deux politiques sur une table pour un principal où l'une accorde des autorisations aux colonnes col1, col2 et col3 par le biais d'une méthode de ressource nommée, et l'autre politique accorde des autorisations à la même table et au même principal à col5, et col6 via des balises LF, les autorisations effectives seront une union des autorisations qui seraient col1, col2, col3, col5 et col6. Cela inclut également les filtres de données et les lignes.

Autorisations de localisation des données

Les autorisations de localisation des données permettent aux utilisateurs non administrateurs de créer des bases de données et des tables sur des sites Amazon S3 spécifiques. Si un utilisateur tente de créer une base de données ou une table dans un emplacement qu'il n'est pas autorisé à créer, la tâche de création échoue. Cela permet d'empêcher les utilisateurs de créer des tables à des emplacements arbitraires dans le lac de données et de contrôler les endroits où ces utilisateurs peuvent lire et écrire des données. Il existe une autorisation implicite lors de la création de tables dans l'emplacement Amazon S3 au sein de la base de données dans laquelle elles sont créées. Pour plus d’informations, consultez Octroi d'autorisations de localisation des données.

Créer des autorisations de table et de base de données

Par défaut, les utilisateurs non administrateurs ne sont pas autorisés à créer des bases de données ou des tables au sein d'une base de données. La création de bases de données est contrôlée au niveau du compte à l'aide des paramètres de Lake Formation afin que seuls les principaux autorisés puissent créer des bases de données. Pour plus d'informations, consultez Création d'une base de données. Pour créer une table, le principal a besoin d'une `CREATE_TABLE` autorisation sur la base de données dans laquelle la table est créée. Pour plus d'informations, consultez Création de tables.

Autorisations implicites et explicites

Lake Formation fournit des autorisations implicites en fonction du personnage et des actions qu'il effectue. Par exemple, les administrateurs des lacs de données obtiennent automatiquement `DESCRIBE` des autorisations pour toutes les ressources du catalogue de données, des autorisations de localisation des données pour tous les emplacements, des autorisations pour créer des bases de données et des tables dans tous les emplacements, ainsi que `Grant` `Revoke` des autorisations sur n'importe quelle ressource. Les créateurs de bases de données obtiennent automatiquement toutes les autorisations de base de données sur les bases de données qu'ils créent, et les créateurs de tables obtiennent toutes les autorisations sur les tables qu'ils créent. Pour plus d'informations, consultez Permissions implicites de Lake Formation.

Autorisations pouvant être accordées

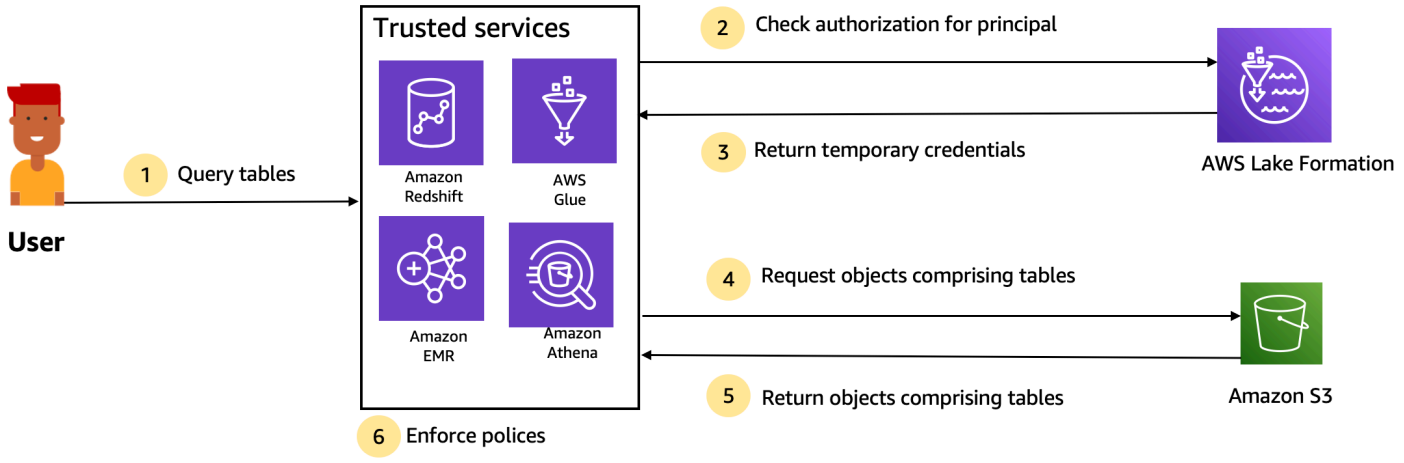
Les administrateurs de data lake ont la possibilité de déléguer la gestion des autorisations à des utilisateurs non administratifs en fournissant des autorisations pouvant être accordées. Lorsqu'un principal reçoit des autorisations pouvant être accordées sur une ressource et un ensemble d'autorisations, ce principal peut accorder des autorisations à d'autres principaux sur cette ressource.

Gestion de l'accès au stockage

Lake Formation utilise la fonctionnalité de distribution automatique d'informations d'identification pour fournir un accès temporaire aux données Amazon S3. La vente d'informations d'identification, ou de jetons, est un modèle courant qui fournit des informations d'identification temporaires aux utilisateurs, aux services ou à une autre entité dans le but d'accorder un accès à court terme à une ressource.

Lake Formation s'appuie sur ce modèle pour fournir un accès à court terme à des services AWS d'analyse tels qu'Athena afin d'accéder aux données pour le compte du principal appelant. Lorsqu'ils accordent des autorisations, les utilisateurs n'ont pas besoin de mettre à jour leurs politiques de compartiment Amazon S3 ou leurs politiques IAM, et ils n'ont pas besoin d'un accès direct à Amazon S3.

Le schéma suivant montre comment Lake Formation fournit un accès temporaire aux sites enregistrés :



Trusted services enforce AWS Lake Formation policies (distributed enforcement with fail close).

1. Un principal (utilisateur) saisit une requête ou une demande de données pour une table par le biais d'un service intégré fiable tel qu'Athena, Amazon EMR, Redshift Spectrum ou. AWS Glue
2. Le service intégré vérifie l'autorisation de Lake Formation pour le tableau et les colonnes demandées et prend une décision d'autorisation. Si l'utilisateur n'est pas autorisé, Lake Formation refuse l'accès aux données et la requête échoue.
3. Une fois l'autorisation réussie et l'autorisation de stockage activée pour la table et l'utilisateur, le service intégré récupère les informations d'identification temporaires de Lake Formation pour accéder aux données.
4. Le service intégré utilise les informations d'identification temporaires de Lake Formation pour demander des objets à Amazon S3.
5. Amazon S3 fournit les objets Amazon S3 au service intégré. Les objets Amazon S3 contiennent toutes les données de la table.
6. Le service intégré assure l'application nécessaire des politiques relatives à la Formation des Lacs, telles que le filtrage au niveau des colonnes, au niveau des lignes et/ou au niveau des cellules. Le service intégré traite les requêtes et renvoie les résultats à l'utilisateur.

Activer l'application des autorisations au niveau du stockage pour les tables du catalogue de données

Par défaut, l'application au niveau du stockage n'est pas activée pour les tables du catalogue de données. Pour activer l'application au niveau du stockage, vous devez enregistrer l'emplacement Amazon S3 de vos données sources auprès de Lake Formation et fournir un rôle IAM. Les

autorisations au niveau du stockage seront activées pour toutes les tables ayant le même chemin d'emplacement de table ou le même préfixe que l'emplacement Amazon S3.

Lorsqu'un service intégré demande l'accès à l'emplacement des données pour le compte d'un utilisateur, le service Lake Formation assume ce rôle et renvoie les informations d'identification au service demandé avec des autorisations limitées sur la ressource afin que l'accès aux données puisse être effectué. Le rôle IAM enregistré doit disposer de tous les accès requis à l'emplacement Amazon S3, y compris AWS KMS les clés.

Pour plus d’informations, consultez Enregistrement d'un emplacement Amazon S3.

AWS Services pris en charge

AWS des services analytiques tels qu'Athena, Redshift Spectrum, Amazon AWS Glue EMR Amazon QuickSight et Amazon SageMaker s'intègrent à AWS Lake Formation à l'aide des opérations de l'API de vente automatique d'informations d'identification de Lake Formation. Pour consulter la liste complète des AWS services intégrés à Lake Formation, ainsi que le niveau de granularité et les formats de table qu'ils prennent en charge, voirCollaboration avec d'autres AWS services.

Partage de données entre comptes dans Lake Formation

Avec Lake Formation, vous pouvez partager les ressources du catalogue de données (bases de données et tables) au sein d'un AWS compte et entre les comptes dans une configuration simple à l'aide de la méthode des ressources nommées ou des balises LF. Vous pouvez partager une base de données complète ou sélectionner des tables d'une base de données avec tous les principaux IAM (rôles et utilisateurs IAM) d'un compte, vers d'autres AWS comptes au niveau du compte ou directement avec les principaux IAM d'un autre compte.

Vous pouvez également partager les tables du catalogue de données avec des filtres de données afin de restreindre l'accès aux détails au niveau des lignes et des cellules. Lake Formation utilise AWS Resource Access Manager (AWS RAM) pour faciliter l'octroi d'autorisations entre comptes. Lorsqu'une ressource est partagée entre deux comptes, AWS RAM envoie des invitations au compte du destinataire. Lorsqu'un utilisateur accepte une invitation de AWS RAM partage, AWS RAM fournit les autorisations nécessaires à Lake Formation pour que les ressources du catalogue de données soient disponibles et pour activer l'application des niveaux de stockage. Pour plus d’informations, consultez Partage de données entre comptes dans Lake Formation.

Lorsque l'administrateur du lac de données du compte destinataire accepte le AWS RAM partage, les ressources partagées sont disponibles dans le compte du destinataire. L'administrateur du lac de données accorde d'autres autorisations Lake Formation sur la ressource partagée aux principaux

IAM supplémentaires du compte destinataire, s'il dispose d'`GRANTABLE`autorisations sur la ressource partagée.

Cependant, les principaux ne peuvent pas interroger les ressources partagées à l'aide d'Athena ou de Redshift Spectrum sans lien de ressource. Un lien de ressource est une entité du catalogue de données similaire à un concept Linux-Symlink.

L'administrateur du lac de données du compte destinataire crée un lien de ressource sur la ressource partagée. L'administrateur accorde des `Describe` autorisations sur le lien de ressource avec les autorisations requises sur la ressource partagée d'origine à d'autres utilisateurs. Un utilisateur du compte destinataire peut ensuite utiliser le lien de ressource pour interroger la ressource partagée à l'aide d'Athena et Redshift Spectrum. Pour plus d'informations sur les liens vers des ressources, consultezCréation de liens vers des ressources.

Composantes de la Lake Formation

AWS Lake Formation repose sur l'interaction de plusieurs composants pour créer et gérer votre lac de données.

Console Lake Formation

Vous utilisez la console Lake Formation pour définir et gérer votre lac de données et pour accorder et révoquer les autorisations de Lake Formation. Vous pouvez utiliser des plans sur la console pour découvrir, nettoyer, transformer et ingérer des données. Vous pouvez également activer ou désactiver l'accès à la console pour les utilisateurs individuels de Lake Formation.

API et interface de ligne de commande de Lake Formation

Lake Formation fournit des opérations d'API via plusieurs SDK spécifiques au langage et le AWS Command Line Interface ().AWS CLI L'API Lake Formation fonctionne conjointement avec l'AWS GlueAPI. L'API Lake Formation se concentre principalement sur la gestion des autorisations de Lake Formation, tandis que l'AWS GlueAPI fournit une API de catalogue de données et une infrastructure gérée pour définir, planifier et exécuter des opérations ETL sur vos données.

Pour plus d'informations sur l'AWS GlueAPI, consultez le guide du AWS Glue développeur. Pour plus d'informations sur l'utilisation du AWS CLI, consultez la référence des AWS CLI commandes.

Autres AWS services

Lake Formation utilise les services suivants :

- AWS Gluepour orchestrer les tâches et les robots d'exploration pour transformer les données à l'aide des AWS Glue transformations.
- L'IAM va accorder des politiques d'autorisation aux responsables de Lake Formation. Le modèle d'autorisation Lake Formation complète le modèle d'autorisation IAM pour sécuriser votre lac de données.

Terminologie des Lake Formation

Voici quelques termes importants que vous rencontrerez dans ce guide.

Lac de données

Le lac de données correspond à vos données persistantes stockées dans Amazon S3 et gérées par Lake Formation à l'aide d'un catalogue de données. Un lac de données stocke généralement les éléments suivants :

- Données structurées et non structurées
- Données brutes et données transformées

Pour qu'un chemin Amazon S3 se trouve dans un lac de données, il doit être enregistré auprès de Lake Formation.

Accès aux données

Lake Formation fournit un accès sécurisé et granulaire aux données grâce à un nouveau modèle d'accord/de révocation des autorisations qui renforce les politiques (IAM). AWS Identity and Access Management

Les analystes et les data scientists peuvent utiliser le portefeuille complet de services d' AWS analyse et d'apprentissage automatique, tels qu'Amazon Athena, pour accéder aux données. Les politiques de sécurité configurées de Lake Formation permettent de garantir que les utilisateurs ne peuvent accéder qu'aux données auxquelles ils sont autorisés à accéder.

Mode d'accès hybride

Le mode d'accès hybride vous permet de sécuriser et d'accéder aux données cataloguées à l'aide des autorisations Lake Formation et des autorisations IAM et Amazon S3. Le mode d'accès hybride

permet aux administrateurs de données d'intégrer les autorisations de Lake Formation de manière sélective et progressive, en se concentrant sur un cas d'utilisation de lac de données à la fois.

Plan

Un plan est un modèle de gestion des données qui vous permet d'ingérer facilement des données dans un lac de données. Lake Formation fournit plusieurs plans, chacun correspondant à un type de source prédéfini, tel qu'une base de données relationnelle ou AWS CloudTrail des journaux. À partir d'un plan, vous pouvez créer un flux de travail. Les flux de travail se composent de AWS Glue robots d'exploration, de tâches et de déclencheurs générés pour orchestrer le chargement et la mise à jour des données. Les plans utilisent la source de données, la cible de données et le calendrier comme entrées pour configurer le flux de travail.

Flux de travail

Un flux de travail est un conteneur pour un ensemble de AWS Glue tâches, de robots d'exploration et de déclencheurs connexes. Vous créez le flux de travail dans Lake Formation, qui s'exécute dans le AWS Glue service. Lake Formation peut suivre l'état d'un flux de travail en tant qu'entité unique.

Lorsque vous définissez un flux de travail, vous sélectionnez le plan sur lequel il est basé. Vous pouvez ensuite exécuter des flux de travail à la demande ou selon un calendrier.

Les flux de travail que vous créez dans Lake Formation sont visibles dans la AWS Glue console sous la forme d'un graphe acyclique dirigé (DAG). À l'aide du DAG, vous pouvez suivre la progression du flux de travail et résoudre les problèmes.

Catalogue de données

Le catalogue de données est votre magasin de métadonnées permanent. Il s'agit d'un service géré qui vous permet de stocker, d'annoter et de partager des métadonnées dans le AWS cloud de la même manière que vous le feriez dans un métastore Apache Hive. Il fournit un référentiel uniforme dans lequel des systèmes disparates peuvent stocker et trouver des métadonnées pour suivre les données dans des silos de données, puis utiliser ces métadonnées pour interroger et transformer les données. Lake Formation utilise le catalogue de AWS Glue données pour stocker les métadonnées relatives aux lacs de données, aux sources de données, aux transformations et aux cibles.

Les métadonnées relatives aux sources de données et aux cibles se présentent sous la forme de bases de données et de tables. Les tables stockent des informations de schéma, des informations de

localisation, etc. Les bases de données sont des ensembles de tables. Lake Formation fournit une hiérarchie d'autorisations pour contrôler l'accès aux bases de données et aux tables du catalogue de données.

Chaque AWS compte possède un catalogue de données par AWS région.

Données sous-jacentes

Les données sous-jacentes font référence aux données sources ou aux données contenues dans les lacs de données vers lesquels pointent les tables du catalogue de données.

Principal

Un principal est un utilisateur ou un rôle AWS Identity and Access Management (IAM) ou un utilisateur Active Directory.

Administrateur du lac de données

Un administrateur de lac de données est un mandant qui peut accorder à n'importe quel principal (y compris lui-même) n'importe quelle autorisation sur n'importe quelle ressource du catalogue de données ou sur l'emplacement des données. Désignez un administrateur de lac de données comme premier utilisateur du catalogue de données. Cet utilisateur peut ensuite accorder des autorisations plus détaillées sur les ressources à d'autres principaux.

> **Note**
>
> Les utilisateurs administratifs IAM, c'est-à-dire les utilisateurs dotés de la politique `AdministratorAccess` AWS gérée, ne sont pas automatiquement des administrateurs de lacs de données. Par exemple, ils ne peuvent pas accorder d'autorisations Lake Formation sur les objets du catalogue à moins d'en avoir obtenu l'autorisation. Ils peuvent toutefois utiliser la console ou l'API Lake Formation pour se désigner comme administrateurs de lacs de données.

Pour plus d'informations sur les fonctionnalités d'un administrateur de lac de données, consultezPermissions implicites de Lake Formation. Pour plus d'informations sur la désignation d'un utilisateur en tant qu'administrateur de data lake, consultez. Création d'un administrateur de lac de données

AWS intégrations de services avec Lake Formation

Vous pouvez utiliser Lake Formation pour gérer les autorisations d'accès aux bases de données, aux tables et aux colonnes sur les données stockées dans Amazon S3. Une fois vos données enregistrées auprès de Lake Formation, vous pouvez utiliser des services AWS d'analyse tels AWS Glue qu'Amazon Athena, Amazon Redshift Spectrum, Amazon EMR pour interroger les données. Les AWS services suivants s'intègrent aux autorisations de Lake Formation AWS Lake Formation et les respectent.

AWS Service	Détails de l'intégration
AWS Glue	Thème de référence : Utilisation AWS Lake Formation avec AWS Glue AWS Glueet Lake Formation partagent le même catalogue de données. Pour les opérations de console (telles que l'affichage d'une liste de tables) et toutes les opérations d'API, les AWS Glue utilisate urs ne peuvent accéder qu'aux bases de données et aux tables sur lesquelles ils disposent des autorisations Lake Formation.
Amazon Athena	Thème de référence : Utilisation AWS Lake Formation avec Amazon Athena Utilisez Lake Formation pour autoriser ou refuser les autorisations de lecture des données dans Amazon S3. Lorsque Amazon Athena les utilisateurs sélectionnent le AWS Glue catalogue dans l'éditeur de requêtes, ils ne peuvent interroger que les bases de données, les tables et les colonnes sur lesquelles ils sont autorisés à Lake Formation. Les requêtes utilisant des manifestes ne sont pas prises en charge. Actuellement, Lake Formation ne prend pas en charge la gestion des autorisations sur les opérations d'écriture telles que `VACUUMMERGE`, `UPDATE` et `OPTIMIZE` sur les tables dans Open Table Formats. Outre les principaux utilisateurs qui s'authentifient auprès d'Athena via AWS Identity and Access Management (IAM), Lake Formation prend en charge les utilisateurs d'Athena qui se connectent via le

AWS Service	Détails de l'intégration
	pilote JDBC ou ODBC et s'authentifient via SAML. Les fournisse urs SAML pris en charge incluent Okta et Microsoft Active Directory Federation Service (AD FS).
Amazon Redshift Spectrum	Thème de référence : Utilisation AWS Lake Formation avec Amazon Redshift Spectrum Lorsque les utilisateurs d'Amazon Redshift créent un schéma externe sur une base de données dans le AWS Glue Data Catalog, ils peuvent uniquement interroger les tables et les colonnes de ce schéma pour lesquelles ils disposent des autorisations de Lake Formation.
Édition Amazon QuickSight Enterprise	Référence : Utilisation AWS Lake Formation avec Amazon QuickSigh t Lorsqu'un utilisateur QuickSight d'Amazon Enterprise Edition interroge un ensemble de données dans un emplacement Amazon S3, il doit disposer de l'`SELECT`autorisation Lake Formation sur les données.
Amazon EMR	Référence : Utilisation AWS Lake Formation avec Amazon EMR Vous pouvez intégrer les autorisations de Lake Formation lorsque vous créez un cluster Amazon EMR doté d'un rôle d'exécution. Un rôle d'exécution est un rôle IAM que vous associez à des tâches ou à des requêtes Amazon EMR, puis Amazon EMR utilise ce rôle pour accéder aux ressources. AWS

Lake Formation travaille également avec AWS Key Management Service(AWS KMS) pour vous permettre de configurer plus facilement ces services intégrés afin de chiffrer et de déchiffrer les données sur les sites Amazon Simple Storage Service (Amazon S3).

Ressources supplémentaires sur la Formation du Lac

Rubriques

- Blogs
- Discussions techniques et webinaires
- Architecture moderne
- Ressources de maillage de données
- Guides des meilleures pratiques

Blogs

- AWS Lake Formation Bilan de l'année 2022
- Architecture de données moderne multirégionale hautement résiliente
- Partage entre comptes à l'aide de balises LF pour diriger les principaux IAM
- Tableau de bord de l'inventaire des autorisations de Lake Formation
- Maillage de données piloté par événements

Discussions techniques et webinaires

- re:Invent 2020 — Lacs de données : créez, sécurisez et partagez facilement avec AWS Lake Formation
- re:Invent 2022 — Création et exploitation d'un lac de données sur Amazon S3
- AWS Summit SF 2022 — Comprendre et mettre en place une architecture de données moderne
- AWS Summit ATL 2022 — Des lacs de données modernes avec AWS Lake Formation Amazon Redshift et AWS Glue
- AWS Summit ANZ 2022 — Lacs de données, maisons lacustres et maillage de données : quoi, pourquoi et comment ?
- AWS Discussions techniques en ligne — Simplifier les autorisations et la gouvernance dans votre lac de données

Architecture moderne

- Modèles d'architecture moderne

Ressources de maillage de données

- Créez une architecture de données moderne et un modèle de maillage de données à grande échelle à l'aide d'un contrôle d'accès AWS Lake Formation basé sur des balises
- Comment JPMorgan Chase a créé une architecture de maillage de données pour générer une valeur significative afin d'améliorer sa plateforme de données d'entreprise
- Créez un maillage de données sur AWS

Guides des meilleures pratiques

- AWS Lake Formation guides de bonnes pratiques

Débuter avec Lake Formation

La lecture de ces sections est indispensable:

- AWS Lake Formation : comment ça marche— Découvrez la terminologie essentielle et la façon dont les différents composants interagissent.
- Débuter avec Lake Formation— Obtenez des informations sur les prérequis et effectuez les tâches de configuration importantes.
- Didacticiels— Suivez les step-by-step tutoriels pour apprendre à utiliser Lake Formation.
- Sécurité dans AWS Lake Formation— Découvrez comment vous pouvez contribuer à sécuriser l'accès aux données de Lake Formation.

Débuter avec Lake Formation

Si vous ne vous êtes pas inscrit AWS ou si vous avez besoin d'aide pour démarrer, assurez-vous d'effectuer les tâches suivantes.

Rubriques

- Exécution des tâches AWS de configuration initiale
- Configurez AWS Lake Formation
- Mise à niveau AWS Glue des autorisations de données vers le AWS Lake Formation modèle
- AWS Lake Formation et points de terminaison VPC d'interface ()AWS PrivateLink

Exécution des tâches AWS de configuration initiale

Pour utiliser AWS Lake Formation, vous devez au préalable effectuer les tâches suivantes :

Rubriques

- Inscrivez-vous pour un Compte AWS
- Création d'un utilisateur doté d'un accès administratif
- Octroi d’un accès par programmation

Inscrivez-vous pour un Compte AWS

Si vous n'en avez pas Compte AWS, procédez comme suit pour en créer un.

Pour vous inscrire à un Compte AWS

1. Ouvrez https://portal.aws.amazon.com/billing/signup.
2. Suivez les instructions en ligne.

   Dans le cadre de la procédure d‘inscription, vous recevrez un appel téléphonique et vous saisirez un code de vérification en utilisant le clavier numérique du téléphone.

   Lorsque vous vous inscrivez à un Compte AWS, un Utilisateur racine d'un compte AWSest créé. Par défaut, seul l‘utilisateur racine a accès à l‘ensemble des Services AWS et des ressources de ce compte. Pour des raisons de sécurité, attribuez un accès administratif à un utilisateur et

utilisez uniquement l'utilisateur root pour effectuer les tâches nécessitant un accès utilisateur root.

AWS vous envoie un e-mail de confirmation une fois le processus d'inscription terminé. Vous pouvez afficher l'activité en cours de votre compte et gérer votre compte à tout moment en accédant à https://aws.amazon.com/ et en choisissant Mon compte.

Création d'un utilisateur doté d'un accès administratif

Une fois que vous vous êtes inscrit à un utilisateur administratif Compte AWS, que vous Utilisateur racine d'un compte AWS l'avez sécurisé AWS IAM Identity Center, que vous l'avez activé et que vous en avez créé un, afin de ne pas utiliser l'utilisateur root pour les tâches quotidiennes.

Sécurisez votre Utilisateur racine d'un compte AWS

1. Connectez-vous en AWS Management Consoletant que propriétaire du compte en choisissant Utilisateur root et en saisissant votre adresse Compte AWS e-mail. Sur la page suivante, saisissez votre mot de passe.

   Pour obtenir de l'aide pour vous connecter en utilisant l'utilisateur racine, consultez Connexion en tant qu'utilisateur racine dans le Guide de l'utilisateur Connexion à AWS .
2. Activez l'authentification multifactorielle (MFA) pour votre utilisateur racine.

   Pour obtenir des instructions, voir Activer un périphérique MFA virtuel pour votre utilisateur Compte AWS root (console) dans le guide de l'utilisateur IAM.

Création d'un utilisateur doté d'un accès administratif

1. Activez IAM Identity Center.

   Pour obtenir des instructions, consultez Activation d' AWS IAM Identity Center dans le Guide de l'utilisateur AWS IAM Identity Center .
2. Dans IAM Identity Center, accordez un accès administratif à un utilisateur.

   Pour un didacticiel sur l'utilisation du Répertoire IAM Identity Center comme source d'identité, voir Configurer l'accès utilisateur par défaut Répertoire IAM Identity Center dans le Guide de AWS IAM Identity Center l'utilisateur.

Connectez-vous en tant qu'utilisateur disposant d'un accès administratif

- Pour vous connecter avec votre utilisateur IAM Identity Center, utilisez l‘URL de connexion qui a été envoyée à votre adresse e-mail lorsque vous avez créé l‘utilisateur IAM Identity Center.

  Pour obtenir de l'aide pour vous connecter en utilisant un utilisateur d'IAM Identity Center, consultez la section Connexion au portail AWS d'accès dans le guide de l'Connexion à AWS utilisateur.

Attribuer l'accès à des utilisateurs supplémentaires

1. Dans IAM Identity Center, créez un ensemble d'autorisations conforme aux meilleures pratiques en matière d'application des autorisations du moindre privilège.

   Pour obtenir des instructions, voir Création d'un ensemble d'autorisations dans le guide de AWS IAM Identity Center l'utilisateur.
2. Affectez des utilisateurs à un groupe, puis attribuez un accès d'authentification unique au groupe.

   Pour obtenir des instructions, voir Ajouter des groupes dans le guide de AWS IAM Identity Center l'utilisateur.

Octroi d’un accès par programmation

Les utilisateurs ont besoin d'un accès programmatique s'ils souhaitent interagir avec AWS l'extérieur du AWS Management Console. La manière d'accorder un accès programmatique dépend du type d'utilisateur qui y accède AWS.

Pour accorder aux utilisateurs un accès programmatique, choisissez l’une des options suivantes.

Quel utilisateur a besoin d’un accès programmatique ?	Pour	Par
Identité de la main-d’œuvre (Utilisateurs gérés dans IAM Identity Center)	Utilisez des informations d'identification temporaires pour signer les demandes programmatiques adressées	Suivez les instructions de l’interface que vous souhaitez utiliser.

Quel utilisateur a besoin d'un accès programmatique ?	Pour	Par
	aux AWS CLI AWS SDK ou AWS aux API.	• Pour le AWS CLI, voir Configuration du AWS CLI à utiliser AWS IAM Identity Center dans le guide de AWS Command Line Interface l'utilisateur. • Pour les AWS SDK, les outils et les AWS API, consultez la section Authentification IAM Identity Center dans le Guide de référence AWS des SDK et des outils.
IAM	Utilisez des informations d'identification temporaires pour signer les demandes programmatiques adressées aux AWS CLI AWS SDK ou AWS aux API.	Suivez les instructions de la section Utilisation d'informations d'identification temporaires avec AWS les ressources du Guide de l'utilisateur IAM.

Quel utilisateur a besoin d'un accès programmatique ?	Pour	Par
IAM	(Non recommandé) Utilisez des informations d'identification à long terme pour signer les AWS CLI demandes programmatiques adressées aux AWS SDK ou AWS aux API.	Suivez les instructions de l'interface que vous souhaitez utiliser. • Pour le AWS CLI, voir Authentification à l'aide des informations d'identification utilisateur IAM dans le Guide de l'AWS Command Line Interface utilisateur. • Pour les AWS SDK et les outils, voir Authentifier à l'aide d'informations d'identification à long terme dans le Guide de AWS référence des SDK et des outils. • Pour les AWS API, consultez la section Gestion des clés d'accès pour les utilisateurs IAM dans le guide de l'utilisateur IAM.

Configurez AWS Lake Formation

Les sections suivantes fournissent des informations sur la configuration de Lake Formation pour la première fois. Toutes les rubriques de cette section ne sont pas obligatoires pour commencer à utiliser Lake Formation. Vous pouvez utiliser les instructions pour configurer le modèle d'autorisations de Lake Formation afin de gérer vos AWS Glue Data Catalog objets et emplacements de données existants dans Amazon Simple Storage Service (Amazon S3).

1. Création d'un administrateur de lac de données
2. Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride
3. the section called “Configurer un emplacement Amazon S3 pour votre lac de données”

4. the section called “Attribuer des autorisations aux utilisateurs de Lake Formation”
5. the section called “Intégration d'IAM Identity Center”
6. the section called “(Facultatif) Paramètres de filtrage des données externes”
7. the section called “(Facultatif) Accordez l'accès à la clé de chiffrement du catalogue de données”
8. (Facultatif) Créez un rôle IAM pour les flux de travail

Cette section explique comment configurer les ressources de Lake Formation de deux manières différentes :

- Utilisation d'un AWS CloudFormation modèle
- Utilisation de la console Lake Formation

Pour configurer Lake Formation à l'aide de AWS la console, rendez-vous surCréation d'un administrateur de lac de données.

Configurer les ressources de Lake Formation à l'aide d' AWS CloudFormation un modèle

> **Note**
>
> La AWS CloudFormation pile exécute les étapes 1 à 6 ci-dessus, à l'exception des étapes 2 et 5. Effectuez Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride et the section called “Intégration d'IAM Identity Center” manuellement depuis la console Lake Formation.

1. Connectez-vous à la AWS CloudFormation console à l'adresse https://console.aws.amazon.com/cloudformation en tant qu'administrateur IAM dans la région USA Est (Virginie du Nord).
2. Choisissez Launch Stack.
3. Choisissez Suivant sur l'écran Créer une pile.
4. Entrez un nom de pile.
5. Pour DatalakeAdminNameet DatalakeAdminPassword, entrez votre nom d'utilisateur et votre mot de passe pour l'utilisateur administrateur de Data Lake.

6. Pour DatalakeUser1Name et DatalakeUser1Password, entrez votre nom d'utilisateur et votre mot de passe pour l'utilisateur Data Lake Analyst.
7. Pour DataLakeBucketName, entrez le nouveau nom du bucket qui sera créé.
8. Choisissez Suivant.
9. Sur la page suivante, choisissez Next.
10. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
11. Choisissez Créer.

    La création de la pile peut prendre jusqu'à deux minutes.

Nettoyage des ressources

Si vous souhaitez nettoyer les ressources de la AWS CloudFormation pile :

1. Désenregistrez le compartiment Amazon S3 créé par votre stack et enregistré en tant qu'emplacement de lac de données.
2. Supprimez la AWS CloudFormation pile. Cela supprimera toutes les ressources créées par la pile.

Création d'un administrateur de lac de données

Les administrateurs de lacs de données sont initialement les seuls utilisateurs ou rôles AWS Identity and Access Management (IAM) qui peuvent accorder à Lake Formation des autorisations sur les emplacements des données et les ressources du catalogue de données à n'importe quel principal (y compris à lui-même). Pour plus d'informations sur les fonctionnalités d'administration des lacs de données, consultezPermissions implicites de Lake Formation. Par défaut, Lake Formation vous permet de créer jusqu'à 30 administrateurs de lacs de données.

Vous pouvez créer un administrateur de lac de données à l'aide de la `PutDataLakeSettings` console Lake Formation ou de l'API Lake Formation.

Les autorisations suivantes sont requises pour créer un administrateur de lac de données. L'`Administrator`utilisateur dispose implicitement de ces autorisations.

- `lakeformation:PutDataLakeSettings`
- `lakeformation:GetDataLakeSettings`

Si vous accordez cette `AWSLakeFormationDataAdmin` politique à un utilisateur, celui-ci ne sera pas en mesure de créer d'autres utilisateurs administrateurs de Lake Formation.

Pour créer un administrateur de lac de données (console)

1. Si l'utilisateur qui doit être administrateur du lac de données n'existe pas encore, utilisez la console IAM pour le créer. Sinon, choisissez un utilisateur existant qui sera l'administrateur du lac de données.

   > **Note**
   >
   > Nous vous recommandons de ne pas sélectionner d'utilisateur administratif IAM (utilisateur doté de la politique `AdministratorAccess` AWS gérée) comme administrateur du lac de données.

   Associez les politiques AWS gérées suivantes à l'utilisateur :

Politiques	Obligatoire ?	Remarques
`AWSLakeFormationDataAdmin`	Obligatoire	Autorisations d'administrateur de base du lac de données. Cette politique AWS gérée contient un refus explicite du fonctionnement de l'API Lake Formation, `PutDataLakeSetting` qui empêche les utilisateurs de créer de nouveaux administrateurs de lacs de données.

Politiques	Obligatoire ?	Remarques
`AWSGlueConsoleFullAccess` , `CloudWatchLogsReadOnlyAccess`	Facultatif	Joignez ces politiques si l'administrateur du lac de données doit résoudre les problèmes liés aux flux de travail créés à partir des plans de Lake Formation. Ces politiques permettent à l'administrateur du lac de données de consulter les informations de dépannage dans la AWS Glue console et dans la Amazon CloudWatch Logs console. Pour plus d'informations sur les flux de travail, consultezthe section called “Importation de données à l'aide de workflows”.
`AWSLakeFormationCrossAccountManager`	Facultatif	Joignez cette politique pour permettre à l'administrateur du lac de données d'accorder et de révoquer des autorisations entre comptes sur les ressources du catalogue de données. Pour plus d’informations, consultez Partage de données entre comptes dans Lake Formation.
`AmazonAthenaFullAccess`	Facultatif	Joignez cette politique si l'administrateur du lac de données doit exécuter des requêtes dans Amazon Athena.

2. Joignez la politique intégrée suivante, qui accorde à l'administrateur du lac de données l'autorisation de créer le rôle lié au service Lake Formation. Le nom suggéré pour la politique est`LakeFormationSLR`.

   Le rôle lié à un service permet à l'administrateur du lac de données d'enregistrer plus facilement les sites Amazon S3 auprès de Lake Formation. Pour plus d'informations sur le rôle lié au service Lake Formation, consultez. the section called “Utilisation des rôles liés à un service”

> **Important**
>
> Dans toutes les politiques suivantes, remplacez-le <account-id>par un numéro de AWS compte valide.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:AWSServiceName": "lakeformation.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PutRolePolicy"
            ],
            "Resource": "arn:aws:iam::<account-id>:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess"
        }
    ]
}
```

3. (Facultatif) Attachez la politique `PassRole` intégrée suivante à l'utilisateur. Cette politique permet à l'administrateur du lac de données de créer et d'exécuter des flux de travail. L'`iam:PassRole`autorisation permet au flux de travail d'assumer le rôle `LakeFormationWorkflowRole` de créer des robots d'exploration et des tâches, et d'associer le rôle aux robots et aux tâches créés. Le nom suggéré pour la politique est`UserPassRole`.

> **Important**
>
> Remplacez <account-id>par un numéro de AWS compte valide.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole"
            ]
        }
    ]
}
```

4. (Facultatif) Joignez cette politique en ligne supplémentaire si votre compte accordera ou recevra des autorisations entre comptes Lake Formation. Cette politique permet à l'administrateur du lac de données de consulter et d'accepter AWS Resource Access Manager (AWS RAM) les invitations de partage de ressources. En outre, pour les administrateurs de lacs de données du compte de AWS Organizations gestion, la politique inclut une autorisation permettant d'autoriser les subventions entre comptes aux organisations. Pour plus d'informations, consultez Partage de données entre comptes dans Lake Formation.

   Le nom suggéré pour la politique est`RAMAccess`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ram:AcceptResourceShareInvitation",
                "ram:RejectResourceShareInvitation",
                "ec2:DescribeAvailabilityZones",
                "ram:EnableSharingWithAwsOrganization"
            ],
            "Resource": "*"
        }
    ]
```

```
}
```

5. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/ et connectez-vous en tant qu'utilisateur administrateur que vous avez créé Création d'un utilisateur doté d'un accès administratif ou en tant qu'utilisateur doté d'une politique AWS gérée par `AdministratorAccess` l'utilisateur.
6. Si une fenêtre Welcome to Lake Formation apparaît, choisissez l'utilisateur IAM que vous avez créé ou sélectionné à l'étape 1, puis choisissez Get started.
7. Si la fenêtre Welcome to Lake Formation ne s'affiche pas, effectuez les étapes suivantes pour configurer un administrateur de Lake Formation.

   a. Dans le volet de navigation, sous Administrateurs, sélectionnez Rôles et tâches administratifs. Dans la section Administrateurs du lac de données de la page de console, choisissez Ajouter.

   b. Dans la boîte de dialogue Ajouter des administrateurs, sous Type d'accès, choisissez Data lake administrator.

   c. Pour les utilisateurs et les rôles IAM, choisissez l'utilisateur IAM que vous avez créé ou sélectionné à l'étape 1, puis cliquez sur Enregistrer.

Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride

Lake Formation démarre avec les paramètres « Utiliser uniquement le contrôle d'accès IAM » activés pour garantir la compatibilité avec le AWS Glue Data Catalog comportement existant. Ces paramètres vous permettent de gérer l'accès à vos données dans le lac de données et à ses métadonnées par le biais de politiques IAM et de politiques de compartiment Amazon S3.

Pour faciliter la transition des autorisations de lac de données d'un modèle IAM et Amazon S3 vers les autorisations de Lake Formation, nous vous recommandons d'utiliser le mode d'accès hybride pour Data Catalog. Avec le mode d'accès hybride, vous disposez d'un chemin incrémentiel qui vous permet d'activer les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les autres utilisateurs ou charges de travail existants.

Pour plus d’informations, consultez Mode d'accès hybride.

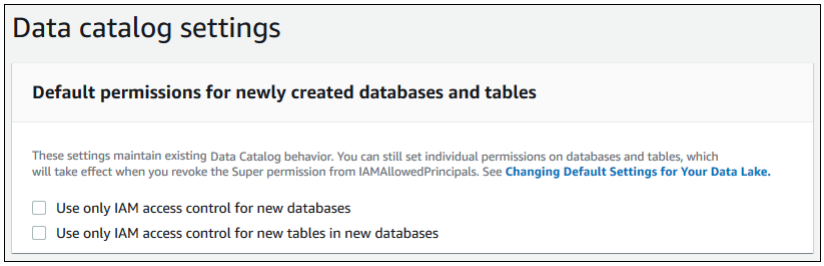
Désactivez les paramètres par défaut pour déplacer tous les utilisateurs existants d'une table vers Lake Formation en une seule étape.

> **Important**
>
> Si vous avez des AWS Glue Data Catalog bases de données et des tables existantes, ne suivez pas les instructions de cette section. Au lieu de cela, suivez les instructions de the section called “Mise à niveau des autorisations de AWS Glue données pour le modèle Lake Formation”.

> **Warning**
>
> Si vous avez mis en place une automatisation qui crée des bases de données et des tables dans le catalogue de données, les étapes suivantes peuvent entraîner l'échec des tâches d'automatisation et d'extraction, de transformation et de chargement (ETL) en aval. Procédez uniquement après avoir modifié vos processus existants ou accordé des autorisations explicites de Lake Formation aux principaux responsables requis. Pour plus d'informations sur les autorisations de Lake Formation, consultezthe section called “Référence des autorisations de Lake Formation”.

Pour modifier les paramètres par défaut du catalogue de données

1. Continuez dans la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/. Assurez-vous d'être connecté en tant qu'utilisateur administrateur que vous avez créé Création d'un utilisateur doté d'un accès administratif ou en tant qu'utilisateur doté de la politique `AdministratorAccess` AWS gérée.
2. Modifiez les paramètres du catalogue de données :
   a. Dans le volet de navigation, sous Administration, sélectionnez Paramètres du catalogue de données.
   b. Décochez les deux cases et choisissez Enregistrer.



3. Révoquez `IAMAllowedPrincipals` l'autorisation accordée aux créateurs de bases de données.

   a. Dans le volet de navigation, sous Administration, sélectionnez Administrative roles and tasks.

   b. Sur la page de console des rôles et tâches administratifs, dans la section Créateurs de base de données, sélectionnez le `IAMAllowedPrincipals` groupe, puis choisissez Révoquer.

      La boîte de dialogue Révoquer les autorisations apparaît, indiquant que `IAMAllowedPrincipals` vous disposez de l'autorisation Créer une base de données.

   c. Choisissez Révoquer.

Attribuer des autorisations aux utilisateurs de Lake Formation

Créez un utilisateur qui aura accès au lac de données dans AWS Lake Formation. Cet utilisateur dispose des autorisations de moindre privilège pour interroger le lac de données.

Pour plus d'informations sur la création d'utilisateurs ou de groupes, consultez la section Identités IAM dans le guide de l'utilisateur IAM.

Pour autoriser un utilisateur non administrateur à accéder aux données de Lake Formation

1. Ouvrez la console IAM à https://console.aws.amazon.com/iam et connectez-vous en tant qu'utilisateur administrateur que vous avez créé Création d'un utilisateur doté d'un accès administratif ou en tant qu'utilisateur avec la politique `AdministratorAccess` AWS gérée.
2. Choisissez Utilisateurs ou Groupes d'utilisateurs.
3. Dans la liste, sélectionnez le nom de l'utilisateur ou du groupe auquel intégrer une politique.

   Choisissez Autorisations.
4. Choisissez Ajouter des autorisations, puis choisissez Joindre directement les politiques. Entrez `Athena` dans le champ de texte Politiques de filtrage. Dans la liste des résultats, cochez la case correspondant à`AmazonAthenaFullAccess`.
5. Cliquez sur le bouton Créer une politique. Sur la page Créer une politique, choisissez l'onglet JSON. Copiez et collez le code suivant dans l'éditeur de politiques.

```
{
    "Version": "2012-10-17",
    "Statement": [
```

```
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
            ],
            "Resource": "*"
        }
    ]
}
```

6. Cliquez sur le bouton Suivant en bas jusqu'à ce que la page de révision de la politique s'affiche. Entrez un nom pour la politique, par exemple,`DatalakeUserBasic`. Choisissez Créer une politique, puis fermez l'onglet Politiques ou la fenêtre du navigateur.

Configurer un emplacement Amazon S3 pour votre lac de données

Pour utiliser Lake Formation afin de gérer et de sécuriser les données de votre lac de données, vous devez d'abord enregistrer un emplacement Amazon S3. Lorsque vous enregistrez un emplacement, ce chemin Amazon S3 et tous les dossiers situés sous ce chemin sont enregistrés, ce qui permet à Lake Formation de faire appliquer les autorisations de niveau de stockage. Lorsque l'utilisateur demande des données à un moteur intégré tel qu'Amazon Athena, Lake Formation fournit l'accès aux données au lieu d'utiliser les autorisations de l'utilisateur.

Lorsque vous enregistrez un emplacement, vous spécifiez un rôle IAM qui accorde des autorisations de lecture/écriture sur cet emplacement. Lake Formation assume ce rôle lorsqu'il fournit des informations d'identification temporaires aux AWS services intégrés qui demandent l'accès aux données de l'emplacement Amazon S3 enregistré. Vous pouvez spécifier le rôle lié au service (SLR) de Lake Formation ou créer votre propre rôle.

Utilisez un rôle personnalisé dans les situations suivantes :

- Vous prévoyez de publier des statistiques dans Amazon CloudWatch Logs. Le rôle défini par l'utilisateur doit inclure une politique d'ajout de journaux dans CloudWatch les journaux et de publication de métriques en plus des autorisations SLR. Pour un exemple de politique intégrée qui accorde les CloudWatch autorisations nécessaires, voirExigences relatives aux rôles utilisés pour enregistrer des sites.
- L'emplacement Amazon S3 existe dans un autre compte. Pour plus de détails, consultez the section called “Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte”.
- L'emplacement Amazon S3 contient des données chiffrées avec un Clé gérée par AWS. Pour plus d'informations, consultez Enregistrement d'un emplacement Amazon S3 chiffré et Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes.
- Vous prévoyez d'accéder à l'emplacement Amazon S3 à l'aide d'Amazon EMR. Pour plus d'informations sur les exigences relatives aux rôles, consultez la section Rôles IAM pour Lake Formation dans le guide de gestion Amazon EMR.

Le rôle que vous choisissez doit disposer des autorisations nécessaires, comme décrit dansExigences relatives aux rôles utilisés pour enregistrer des sites. Pour obtenir des instructions sur la façon d'enregistrer un emplacement Amazon S3, consultezAjouter un emplacement Amazon S3 à votre lac de données.

(Facultatif) Paramètres de filtrage des données externes

Si vous avez l'intention d'analyser et de traiter les données de votre lac de données à l'aide de moteurs de requêtes tiers, vous devez autoriser les moteurs externes à accéder aux données gérées par Lake Formation. Si vous ne vous inscrivez pas, les moteurs externes ne pourront pas accéder aux données des sites Amazon S3 enregistrés auprès de Lake Formation.

Lake Formation prend en charge les autorisations au niveau des colonnes afin de restreindre l'accès à des colonnes spécifiques d'un tableau. Les services d'analyse intégrés tels qu' Amazon Athena Amazon Redshift Spectrum et Amazon EMR extraient les métadonnées de table non filtrées à partir du. AWS Glue Data Catalog Le filtrage proprement dit des colonnes dans les réponses aux requêtes relève de la responsabilité du service intégré. Il est de la responsabilité des administrateurs tiers de gérer correctement les autorisations afin d'éviter tout accès non autorisé aux données.

Pour autoriser les moteurs tiers à accéder aux données et à les filtrer (console)

1. Continuez dans la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/. Assurez-vous que vous êtes connecté en tant que principal disposant de

l'autorisation IAM pour le fonctionnement de l'`PutDataLakeSettings`API Lake Formation. L'utilisateur administrateur IAM que vous avez créé Inscrivez-vous pour un Compte AWS possède cette autorisation.

2. Dans le volet de navigation, sous Administration, sélectionnez Paramètres d'intégration des applications.
3. Sur la page des paramètres d'intégration des applications, procédez comme suit :

   a. Cochez la case Autoriser les moteurs externes à filtrer les données dans les sites Amazon S3 enregistrés auprès de Lake Formation.

   b. Entrez les valeurs des balises de session définies pour les moteurs tiers.

   c. Pour les identifiants de AWS compte, entrez les identifiants de compte à partir desquels les moteurs tiers sont autorisés à accéder aux emplacements enregistrés auprès de Lake Formation. Appuyez sur Entrée après chaque identifiant de compte.

   d. Choisissez Enregistrer.

Pour autoriser les moteurs externes à accéder aux données sans validation des balises de session, voir Intégration des applications pour un accès complet aux tables

(Facultatif) Accordez l'accès à la clé de chiffrement du catalogue de données

Si elle AWS Glue Data Catalog est cryptée, accordez des autorisations AWS Identity and Access Management (IAM) sur la AWS KMS clé à tous les principaux qui doivent accorder des autorisations à Lake Formation sur les bases de données et les tables du catalogue de données.

Pour plus d'informations, consultez le Guide du développeur AWS Key Management Service .

(Facultatif) Créez un rôle IAM pour les flux de travail

Avec AWS Lake Formation, vous pouvez importer vos données à l'aide de flux de travail exécutés par des AWS Glue robots d'exploration. Un flux de travail définit la source de données et le calendrier d'importation des données dans votre lac de données. Vous pouvez facilement définir des flux de travail à l'aide des plans ou des modèles fournis par Lake Formation.

Lorsque vous créez un flux de travail, vous devez lui attribuer un rôle AWS Identity and Access Management (IAM) qui accorde à Lake Formation les autorisations nécessaires pour ingérer les données.

La procédure suivante suppose une bonne connaissance de l'IAM.

Pour créer un rôle IAM pour les flux de travail

1. Ouvrez la console IAM à https://console.aws.amazon.com/iam et connectez-vous en tant qu'utilisateur administrateur que vous avez créé Création d'un utilisateur doté d'un accès administratif ou en tant qu'utilisateur avec la politique `AdministratorAccess` AWS gérée.
2. Dans le volet de navigation, choisissez Rôles, puis Créer un rôle.
3. Sur la page Create role, choisissez AWS service, puis Glue. Choisissez Suivant.
4. Sur la page Ajouter des autorisations, recherchez la stratégie AWSGlueServiceRolegérée et cochez la case à côté du nom de la stratégie dans la liste. Complétez ensuite l'assistant de création de rôle en nommant le rôle`LFWorkflowRole`. Pour terminer, choisissez Create role.
5. De retour sur la page Rôles, recherchez `LFflowRole` et choisissez le nom du rôle.
6. Sur la page Récapitulatif des rôles, sous l'onglet Autorisations, choisissez Créer une politique intégrée. Sur l'écran Créer une politique, accédez à l'onglet JSON et ajoutez la politique en ligne suivante. Le nom suggéré pour la politique est`LakeFormationWorkflow`.

> **Important**
>
> Dans la politique suivante, remplacez <account-id>par un Compte AWS numéro valide.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                 "lakeformation:GetDataAccess",
                 "lakeformation:GrantPermissions"
             ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": ["iam:PassRole"],
            "Resource": [
                "arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole"
            ]
```

```
        }
    ]
}
```

Voici une brève description des autorisations définies dans cette politique :

- `lakeformation:GetDataAccess`permet aux tâches créées par le flux de travail d'écrire sur l'emplacement cible.
- `lakeformation:GrantPermissions`permet au flux de travail d'accorder l'`SELECT`autorisation sur les tables cibles.
- `iam:PassRole`permet au service d'assumer le rôle de `LakeFormationWorkflowRole` créer des robots d'exploration et des tâches (instances de flux de travail), et d'associer le rôle aux robots et aux tâches créés.

7. Vérifiez que deux politiques sont associées au rôle`LakeFormationWorkflowRole`.
8. Si vous ingérez des données situées en dehors de l'emplacement du lac de données, ajoutez une politique intégrée autorisant la lecture des données sources.

Mise à niveau AWS Glue des autorisations de données vers le AWS Lake Formation modèle

AWS Lake Formation les autorisations permettent un contrôle d'accès précis pour les données de votre lac de données. Vous pouvez utiliser le modèle d'autorisations de Lake Formation pour gérer vos AWS Glue Data Catalog objets et emplacements de données existants dans Amazon Simple Storage Service (Amazon S3).

Le modèle d'autorisations de Lake Formation utilise des autorisations grossières AWS Identity and Access Management (IAM) pour l'accès aux services d'API. Il restreint les données auxquelles vos utilisateurs et ces services peuvent accéder via la fonctionnalité Lake Formation. À titre de comparaison, le AWS Glue modèle accorde l'accès aux données via des autorisations IAM de contrôle d'accès détaillées. Pour effectuer le changement, suivez les étapes décrites dans ce guide.

Pour plus d’informations, consultez Vue d'ensemble des autorisations relatives à Lake Formation .

Rubriques

- À propos de la mise à niveau vers le modèle d'autorisations de Lake Formation
- Étape 1 : Répertorier les autorisations existantes des utilisateurs et des rôles

- Étape 2 : configurer des autorisations équivalentes pour Lake Formation
- Étape 3 : Donnez aux utilisateurs l'autorisation IAM d'utiliser Lake Formation
- Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations de Lake Formation
- Étape 5 : Sécuriser les nouvelles ressources du catalogue de données
- Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM pour l'accès futur au data lake
- Étape 7 : Nettoyer les politiques IAM existantes

À propos de la mise à niveau vers le modèle d'autorisations de Lake Formation

Pour maintenir la rétrocompatibilité avecAWS Glue, AWS Lake Formation accorde par défaut l'`Super`autorisation au `IAMAllowedPrincipals` groupe sur toutes les ressources du catalogue de AWS Glue données existantes, et accorde l'`Super`autorisation sur les nouvelles ressources du catalogue de données si les paramètres de contrôle d'accès Utiliser uniquement IAM sont activés. Ainsi, l'accès aux ressources du catalogue de données et aux emplacements Amazon S3 est contrôlé uniquement par des politiques AWS Identity and Access Management (IAM). Le `IAMAllowedPrincipals` groupe inclut tous les utilisateurs et rôles IAM autorisés à accéder aux objets de votre catalogue de données par vos politiques IAM. L'`Super`autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée.

Vous pouvez commencer à utiliser Lake Formation pour gérer l'accès à vos données en enregistrant les emplacements des ressources du catalogue de données existantes dans Lake Formation ou en utilisant le mode d'accès hybride. Lorsque vous enregistrez l'emplacement Amazon S3 en mode d'accès hybride, vous pouvez activer les autorisations de Lake Formation en optant pour les principes pour les bases de données et les tables situées sous cet emplacement.

Pour faciliter la transition des autorisations de lac de données d'un modèle IAM et Amazon S3 vers les autorisations de Lake Formation, nous vous recommandons d'utiliser le mode d'accès hybride pour Data Catalog. Avec le mode d'accès hybride, vous disposez d'un chemin incrémentiel qui vous permet d'activer les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les autres utilisateurs ou charges de travail existants.

Pour plus d’informations, consultez Mode d'accès hybride.

Désactivez les paramètres par défaut du catalogue de données pour déplacer tous les utilisateurs existants d'une table vers Lake Formation en une seule étape.

Pour commencer à utiliser les autorisations de Lake Formation avec vos bases de AWS Glue données et tables de catalogue de données existantes, vous devez effectuer les opérations suivantes :

1. Déterminez les autorisations IAM existantes de vos utilisateurs pour chaque base de données et chaque table.
2. Répliquez ces autorisations dans Lake Formation.
3. Pour chaque site Amazon S3 contenant des données :
   a. Révoquez l'`Super`autorisation accordée au `IAMAllowedPrincipals` groupe sur chaque ressource du catalogue de données qui fait référence à cet emplacement.
   b. Enregistrez l'emplacement auprès de Lake Formation.
4. Nettoyez les politiques précises de contrôle d'accès IAM existantes.

> **Important**
>
> Pour ajouter de nouveaux utilisateurs pendant le processus de transition de votre catalogue de données, vous devez configurer des AWS Glue autorisations granulaires dans IAM comme auparavant. Vous devez également reproduire ces autorisations dans Lake Formation, comme décrit dans cette section. Si les nouveaux utilisateurs disposent des politiques IAM grossières décrites dans ce guide, ils peuvent répertorier toutes les bases de données ou tables auxquelles l'autorisation est accordée. `Super` `IAMAllowedPrincipals` Ils peuvent également consulter les métadonnées de ces ressources.

Suivez les étapes décrites dans cette section pour passer au modèle d'autorisations Lake Formation. Commencez parthe section called “Étape 1 : Répertorier les autorisations existantes”.

Étape 1 : Répertorier les autorisations existantes des utilisateurs et des rôles

Pour commencer à utiliser AWS Lake Formation les autorisations avec vos AWS Glue bases de données et tables existantes, vous devez d'abord déterminer les autorisations existantes de vos utilisateurs.

> **Important**
>
> Avant de commencer, assurez-vous d'avoir terminé les tâches dansPremiers pas.

Rubriques

- Utilisation de l'opération API
- À l'aide du AWS Management Console
- En utilisant AWS CloudTrail

Utilisation de l'opération API

Utilisez l'opération d'ListPoliciesGrantingServiceAccessAPI AWS Identity and Access Management (IAM) pour déterminer les politiques IAM associées à chaque principal (utilisateur ou rôle). À partir des politiques renvoyées dans les résultats, vous pouvez déterminer les autorisations IAM accordées au principal. Vous devez appeler l'API pour chaque principal séparément.

Example

L' AWS CLI exemple suivant renvoie les politiques associées à l'utilisateur`glue_user1`.

```
aws iam list-policies-granting-service-access --arn arn:aws:iam::111122223333:user/
glue_user1 --service-namespaces glue
```

La commande renvoie des résultats similaires aux suivants.

```
{
    "PoliciesGrantingServiceAccess": [
        {
            "ServiceNamespace": "glue",
            "Policies": [
                {
                    "PolicyType": "INLINE",
                    "PolicyName": "GlueUserBasic",
                    "EntityName": "glue_user1",
                    "EntityType": "USER"
                },
                {
                    "PolicyType": "MANAGED",
```

```
                    "PolicyArn": "arn:aws:iam::aws:policy/AmazonAthenaFullAccess",
                    "PolicyName": "AmazonAthenaFullAccess"
                }
            ]
        }
    ],
    "IsTruncated": false
}
```

À l'aide du AWS Management Console

Vous pouvez également consulter ces informations sur la console AWS Identity and Access Management (IAM), dans l'onglet Access Advisor de la page Résumé de l'utilisateur ou du rôle :

1. Ouvrez la console IAM à l'adresse https://console.aws.amazon.com/iam/.
2. Dans le panneau de navigation, sélectionnez Users (Utilisateurs) ou Roles (Rôles).
3. Choisissez un nom dans la liste pour ouvrir sa page de résumé, puis cliquez sur l'onglet Access Advisor.
4. Examinez chacune des politiques afin de déterminer la combinaison de bases de données, de tables et d'actions pour laquelle chaque utilisateur est autorisé.

   N'oubliez pas d'inspecter les rôles en plus des utilisateurs au cours de ce processus, car vos tâches de traitement des données peuvent assumer des rôles pour accéder aux données.

En utilisant AWS CloudTrail

Une autre façon de déterminer vos autorisations existantes consiste à rechercher les appels d'AWS GlueAPI AWS CloudTrail pour lesquels le `additionaleventdata` champ des journaux contient une `insufficientLakeFormationPermissions` entrée. Cette entrée répertorie la base de données et la table pour lesquelles l'utilisateur doit disposer des autorisations de Lake Formation pour effectuer la même action.

Il s'agit de journaux d'accès aux données, il n'est donc pas garanti qu'ils produisent une liste complète des utilisateurs et de leurs autorisations. Nous vous recommandons de choisir une plage de temps étendue pour capturer la plupart des modèles d'accès aux données de vos utilisateurs, par exemple plusieurs semaines ou mois.

Pour plus d'informations, consultez la section Affichage des événements avec l'historique des CloudTrail événements dans le guide de AWS CloudTrail l'utilisateur.

Ensuite, vous pouvez configurer les autorisations de Lake Formation pour qu'elles correspondent aux AWS Glue autorisations. veuillez consulter Étape 2 : configurer des autorisations équivalentes pour Lake Formation.

Étape 2 : configurer des autorisations équivalentes pour Lake Formation

À l'aide des informations collectées dansÉtape 1 : Répertorier les autorisations existantes des utilisateurs et des rôles, accordez AWS Lake Formation des autorisations correspondant aux AWS Glue autorisations. Utilisez l'une des méthodes suivantes pour effectuer les subventions :

- Utilisez la console Lake Formation ou le AWS CLI.

  veuillez consulter the section called “Octroi et révocation des autorisations du catalogue de données”.
- Utilisez les opérations de `BatchGrantPermissions` l'API `GrantPermissions` or.

  veuillez consulter API d'autorisations.

Pour plus d’informations, consultez Vue d'ensemble des autorisations relatives à Lake Formation .

Après avoir configuré les autorisations de Lake Formation, passez àÉtape 3 : Donnez aux utilisateurs l'autorisation IAM d'utiliser Lake Formation.

Étape 3 : Donnez aux utilisateurs l'autorisation IAM d'utiliser Lake Formation

Pour utiliser le modèle AWS Lake Formation d'autorisations, les principaux doivent disposer d'autorisations AWS Identity and Access Management (IAM) sur les API de Lake Formation.

Créez la politique suivante dans IAM et associez-la à chaque utilisateur ayant besoin d'accéder à votre lac de données. Nommez la stratégie `LakeFormationDataAccess`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationDataAccess",
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess"
            ],
```

```
            "Resource": "*"
        }
    ]
}
```

Passez ensuite aux autorisations de Lake Formation, un emplacement de données à la fois. veuillez consulter Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations de Lake Formation.

Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations de Lake Formation

Passez aux autorisations de Lake Formation, un emplacement de données à la fois. Pour ce faire, répétez l'intégralité de cette section jusqu'à ce que vous ayez enregistré tous les chemins Amazon Simple Storage Service (Amazon S3) référencés par votre catalogue de données.

Rubriques

- Vérifiez les autorisations de Lake Formation
- Sécuriser les ressources du catalogue de données existantes
- Activez les autorisations de Lake Formation pour votre site Amazon S3

Vérifiez les autorisations de Lake Formation

Avant d'enregistrer un emplacement, effectuez une étape de vérification pour vous assurer que les bons directeurs disposent des autorisations requises pour la formation de Lake et qu'aucune autorisation de Lake Formation n'est accordée aux directeurs qui ne devraient pas en avoir. À l'aide de l'opération d'`GetEffectivePermissionsForPath`API Lake Formation, identifiez les ressources du catalogue de données qui font référence à l'emplacement Amazon S3, ainsi que les principaux détenteurs d'autorisations sur ces ressources.

L' AWS CLI exemple suivant renvoie les bases de données et les tables du catalogue de données qui font référence au compartiment Amazon S3`products`.

```
aws lakeformation get-effective-permissions-for-path --resource-arn
 arn:aws:s3:::products --profile datalake_admin
```

Notez l'`profile`option. Nous vous recommandons d'exécuter la commande en tant qu'administrateur du lac de données.

Ce qui suit est un extrait des résultats renvoyés.

```
{
        "PermissionsWithGrantOption": [
            "SELECT"
        ],
        "Resource": {
            "TableWithColumns": {
                "Name": "inventory_product",
                "ColumnWildcard": {},
                "DatabaseName": "inventory"
            }
        },
        "Permissions": [
            "SELECT"
        ],
        "Principal": {
            "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1",
            "DataLakePrincipalType": "IAM_USER"
        }
 },...
```

> **Important**
>
> Si votre catalogue de AWS Glue données est chiffré, il ne `GetEffectivePermissionsForPath` renvoie que les bases de données et les tables créées ou modifiées après la mise à disposition générale de Lake Formation.

Sécuriser les ressources du catalogue de données existantes

Ensuite, révoquez l'Superautorisation pour chaque table et `IAMAllowedPrincipals` base de données que vous avez identifiées pour l'emplacement.

> **Warning**
>
> Si vous avez mis en place une automatisation qui crée des bases de données et des tables dans le catalogue de données, les étapes suivantes peuvent entraîner l'échec des tâches d'automatisation et d'extraction, de transformation et de chargement (ETL)

en aval. Procédez uniquement après avoir modifié vos processus existants ou accordé des autorisations explicites de Lake Formation aux principaux responsables requis. Pour plus d'informations sur les autorisations de Lake Formation, consultezthe section called "Référence des autorisations de Lake Formation".

Pour révoquer **`Super`** depuis **`IAMAllowedPrincipals`** une table

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, choisissez Tables.
3. Sur la page Tables, sélectionnez le bouton radio à côté du tableau souhaité.
4. Dans le menu Actions, choisissez Révoquer.
5. Dans la boîte de dialogue Révoquer les autorisations, dans la liste des utilisateurs et des rôles IAM, faites défiler la page jusqu'à l'en-tête Groupe, puis sélectionnez IAM. AllowedPrincipals
6. Sous Autorisations relatives aux tables, assurez-vous que Super est sélectionné, puis choisissez Révoquer.

Pour révoquer l'accès **`Super`** à une **`IAMAllowedPrincipals`** base de données

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le panneau de navigation, choisissez Databases (Bases de données).
3. Sur la page Bases de données, sélectionnez le bouton radio à côté de la base de données souhaitée.
4. Dans le menu Actions, choisissez Modifier.
5. Sur la page Modifier la base de données, désactivez l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données, puis sélectionnez Enregistrer.
6. De retour sur la page Bases de données, assurez-vous que la base de données est toujours sélectionnée, puis dans le menu Actions, choisissez Révoquer.
7. Dans la boîte de dialogue Révoquer les autorisations, dans la liste des utilisateurs et des rôles IAM, faites défiler la page jusqu'à l'en-tête Groupe, puis sélectionnez IAM. AllowedPrincipals
8. Sous Autorisations de base de données, assurez-vous que Super est sélectionné, puis choisissez Révoke.

Activez les autorisations de Lake Formation pour votre site Amazon S3

Enregistrez ensuite l'emplacement Amazon S3 auprès de Lake Formation. Pour ce faire, vous pouvez utiliser le processus décrit dansAjouter un emplacement Amazon S3 à votre lac de données. Vous pouvez également utiliser l'opération d'`RegisterResource`API décrite dansAPI de vente d'informations d'identification.

> **Note**
>
> Si un établissement parent est enregistré, il n'est pas nécessaire d'enregistrer un établissement pour enfants.

Après avoir terminé ces étapes et vérifié que vos utilisateurs peuvent accéder à leurs données, vous avez passé avec succès aux autorisations Lake Formation. Passez à l'étape suivante,Étape 5 : Sécuriser les nouvelles ressources du catalogue de données.

Étape 5 : Sécuriser les nouvelles ressources du catalogue de données

Sécurisez ensuite toutes les nouvelles ressources du catalogue de données en modifiant les paramètres par défaut du catalogue de données. Désactivez les options permettant d'utiliser uniquement le contrôle d'accès AWS Identity and Access Management (IAM) pour les nouvelles bases de données et tables.

> **Warning**
>
> Si vous avez mis en place une automatisation qui crée des bases de données et des tables dans le catalogue de données, les étapes suivantes peuvent entraîner l'échec des tâches d'automatisation et d'extraction, de transformation et de chargement (ETL) en aval. Procédez uniquement après avoir modifié vos processus existants ou accordé des autorisations explicites de Lake Formation aux principaux responsables requis. Pour plus d'informations sur les autorisations de Lake Formation, consultezthe section called “Référence des autorisations de Lake Formation”.

Pour modifier les paramètres par défaut du catalogue de données

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'utilisateur administratif IAM (l'utilisateur

`Administrator` ou un autre utilisateur doté de la politique `AdministratorAccess` AWS gérée).

2. Dans le panneau de navigation, sélectionnez Settings (Paramètres).
3. Sur la page des paramètres du catalogue de données, décochez les deux cases, puis choisissez Enregistrer.

L'étape suivante consiste à accorder aux utilisateurs l'accès à des bases de données ou à des tables supplémentaires à l'avenir. veuillez consulter Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM pour l'accès futur au data lake.

Étape 6 : Donnez aux utilisateurs une nouvelle politique IAM pour l'accès futur au data lake

Pour accorder à vos utilisateurs l'accès à des bases de données ou à des tables supplémentaires du catalogue de données à l'avenir, vous devez leur donner la politique en ligne grossière AWS Identity and Access Management (IAM) décrite ci-dessous. Nommez la stratégie `GlueFullReadAccess`.

> **Important**
>
> Si vous associez cette politique à un utilisateur avant de révoquer l'accès à chaque `IAMAllowedPrincipals` base `Super` de données et table de votre catalogue de données, cet utilisateur peut consulter toutes les métadonnées de toutes les ressources pour lesquelles `Super` il est accordé à`IAMAllowedPrincipals`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "GlueFullReadAccess",
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
```

```
                "glue:GetPartitions"
            ],
            "Resource": "*"
        }
    ]
}
```

> **Note**
>
> Les politiques intégrées définies dans cette étape et dans les étapes précédentes contiennent des autorisations IAM minimales. Pour les politiques suggérées aux administrateurs de lacs de données, aux analystes de données et à d'autres personnes, consultezthe section called “Référence des personnalités de Lake Formation et des autorisations IAM”.

Ensuite, passez àÉtape 7 : Nettoyer les politiques IAM existantes.

Étape 7 : Nettoyer les politiques IAM existantes

Après avoir configuré les AWS Lake Formation autorisations et créé et attaché les politiques de contrôle d'accès AWS Identity and Access Management (IAM) grossières, effectuez la dernière étape suivante :

- Supprimez des utilisateurs, des groupes et des rôles les anciennes politiques détaillées de contrôle d'accès IAM que vous avez répliquées dans Lake Formation.

Ce faisant, vous vous assurez que ces principaux n'ont plus accès directement aux données d'Amazon Simple Storage Service (Amazon S3). Vous pouvez ensuite gérer l'accès au data lake pour ces principaux acteurs entièrement par le biais de Lake Formation.

AWS Lake Formation et points de terminaison VPC d'interface ()AWS PrivateLink

Amazon VPC est un AWS service que vous pouvez utiliser pour lancer AWS des ressources dans un réseau virtuel que vous définissez. Avec un VPC, vous contrôlez des paramètres réseau, tels que la plage d’adresses IP, les sous-réseaux, les tables de routage et les passerelles réseau.

Si vous utilisez Amazon Virtual Private Cloud (Amazon VPC) pour héberger vos AWS ressources, vous pouvez établir une connexion privée entre votre VPC et Lake Formation. Vous utilisez cette connexion pour que Lake Formation puisse communiquer avec les ressources de votre VPC sans passer par l'Internet public.

Vous pouvez établir une connexion privée entre votre VPC et créer un point de AWS Lake Formation terminaison VPC d'interface. Les points de terminaison de l'interface sont alimentés par AWS PrivateLinkune technologie qui vous permet d'accéder en privé aux API de Lake Formation sans passerelle Internet, appareil NAT, connexion VPN ou AWS Direct Connect connexion. Les instances de votre VPC n'ont pas besoin d'adresses IP publiques pour communiquer avec les API de Lake Formation. Le trafic entre votre VPC et Lake Formation ne quitte pas le réseau Amazon.

Chaque point de terminaison d'interface est représenté par une ou plusieurs interfaces réseau Elastic dans vos sous-réseaux.

Pour de plus amples informations, consultez Points de terminaison VPC (AWS PrivateLink) dans le Guide de l'utilisateur Amazon VPC.

Considérations relatives aux points d'extrémité du VPC de Lake Formation

Avant de configurer un point de terminaison VPC d'interface pour Lake Formation, assurez-vous de consulter les propriétés et les limites du point de terminaison d'interface dans le guide de l'utilisateur Amazon VPC.

Lake Formation permet d'appeler toutes ses actions d'API depuis votre VPC. Vous pouvez utiliser Lake Formation avec des points de terminaison VPC dans tout Régions AWS ce qui prend en charge à la fois les points de terminaison Lake Formation et Amazon VPC.

Création d'un point de terminaison VPC d'interface pour Lake Formation

Vous pouvez créer un point de terminaison VPC pour le service Lake Formation à l'aide de la console Amazon VPC ou du (). AWS Command Line Interface AWS CLI Pour plus d'informations, consultez Création d'un point de terminaison d'interface dans le Guide de l'utilisateur Amazon VPC.

Créez un point de terminaison VPC pour Lake Formation en utilisant le nom de service suivant :

- com.amazonaws.*region*.lakeformation

Si vous activez le DNS privé pour le point de terminaison, vous pouvez envoyer des demandes d'API à Lake Formation en utilisant son nom DNS par défaut pour la région, par exemple,`lakeformation.us-east-1.amazonaws.com`.

Pour plus d'informations, consultez Accès à un service via un point de terminaison d'interface dans le Guide de l'utilisateur Amazon VPC.

Création d'une politique de point de terminaison VPC pour Lake Formation

Lake Formation prend en charge les politiques relatives aux points de terminaison des VPC. Une politique de point de terminaison VPC est une politique de ressources AWS Identity and Access Management (IAM) que vous attachez à un point de terminaison lorsque vous créez ou modifiez le point de terminaison.

Vous pouvez associer une politique de point de terminaison à votre point de terminaison VPC qui contrôle l'accès à Lake Formation. La politique spécifie les informations suivantes :

- Le principal qui peut exécuter des actions.
- Les actions qui peuvent être effectuées.
- Les ressources sur lesquelles les actions peuvent être exécutées.

Pour plus d'informations, consultez Contrôle de l'accès aux services avec points de terminaison d'un VPC dans le Guide de l'utilisateur Amazon VPC.

Exemple : politique de point de terminaison VPC pour les actions de Lake Formation

L'exemple de politique de point de terminaison VPC suivant pour Lake Formation permet la vente d'informations d'identification à l'aide des autorisations de Lake Formation. Vous pouvez utiliser cette politique pour exécuter des requêtes à l'aide des autorisations de Lake Formation à partir d'un cluster Amazon Redshift ou d'un Amazon EMR cluster situé dans un sous-réseau privé.

```
{
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "lakeformation:GetDataAccess",
            "Resource": "*",
            "Principal": "*"
        }
    ]
```

```
}
```

> **Note**
>
> Si vous n'attachez pas de politique lorsque vous créez un point de terminaison, une politique par défaut autorisant un accès complet au service est jointe.

Pour plus d'informations, consultez les rubriques suivantes dans la documentation Amazon VPC :

- Qu'est-ce qu'Amazon VPC ?
- Création d'un point de terminaison d'interface
- Utiliser les politiques de point de terminaison VPC

Didacticiels

Les didacticiels suivants sont organisés en trois parties et fournissent des step-by-step instructions sur la façon de créer un lac de données, d'ingérer des données, de partager et de sécuriser des lacs de données à l'aide AWS Lake Formation de :

1. Créez un lac de données et ingérez des données : apprenez à créer un lac de données et à utiliser des plans pour déplacer, stocker, cataloguer, nettoyer et organiser vos données. Vous apprendrez également à configurer des tables gouvernées. Une table gouvernée est un nouveau type de table Amazon S3 qui prend en charge les transactions atomiques, cohérentes, isolées et durables (ACID).

   Avant de commencer, assurez-vous d'avoir terminé les étapes deDébuter avec Lake Formation.

   - Création d'un lac de données à partir d'une AWS CloudTrail source

     Créez et chargez votre premier lac de données en utilisant vos propres CloudTrail journaux comme source de données.

   - Création d'un lac de données à partir d'une source JDBC dans Lake Formation

     Créez un lac de données en utilisant l'un de vos magasins de données accessibles à JDBC, tel qu'une base de données relationnelle, comme source de données.

2. Sécurisation des lacs de données : apprenez à utiliser des contrôles d'accès basés sur des balises et au niveau des lignes pour sécuriser et gérer efficacement l'accès à vos lacs de données.

   - Configuration des autorisations pour les formats de stockage de tables ouvertes dans Lake Formation

     Ce didacticiel explique comment configurer des autorisations pour les formats de tables transactionnels open source (tables Apache Iceberg, Apache Hudi et Linux Foundation Delta Lake) dans Lake Formation.

   - Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation

     Apprenez à gérer l'accès aux données d'un lac de données à l'aide du contrôle d'accès basé sur des balises dans Lake Formation.

   - Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes

Apprenez à configurer des autorisations au niveau des lignes qui vous permettent de restreindre l'accès à des lignes spécifiques en fonction des politiques de conformité et de gouvernance des données de Lake Formation.

3. Partage de données : apprenez à partager vos données en toute sécurité à Comptes AWS l'aide du contrôle d'accès basé sur les balises (TBAC) et à gérer les autorisations détaillées sur les ensembles de données partagés entre eux. Comptes AWS

   - Partage d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation et de ressources nommées

     Dans ce didacticiel, vous apprendrez à partager vos données en toute sécurité à Comptes AWS l'aide de Lake Formation.

   - Partage d'un lac de données à l'aide du contrôle d'accès précis de Lake Formation

     Dans ce didacticiel, vous apprendrez à partager rapidement et facilement des ensembles de données à l'aide de Lake Formation lors de la gestion de plusieurs ensembles de données Comptes AWS avec AWS Organizations.

Rubriques

- Création d'un lac de données à partir d'une AWS CloudTrail source
- Création d'un lac de données à partir d'une source JDBC dans Lake Formation
- Configuration des autorisations pour les formats de stockage de tables ouvertes dans Lake Formation
- Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation
- Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes
- Partage d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation et de ressources nommées
- Partage d'un lac de données à l'aide du contrôle d'accès précis de Lake Formation

Création d'un lac de données à partir d'une AWS CloudTrail source

Ce didacticiel vous explique les actions à effectuer sur la console Lake Formation pour créer et charger votre premier lac de données à partir d'une AWS CloudTrail source.

Étapes de haut niveau pour créer un lac de données

1. Enregistrez un chemin Amazon Simple Storage Service (Amazon S3) en tant que lac de données.
2. Accordez à Lake Formation l'autorisation d'écrire dans le catalogue de données et dans les emplacements Amazon S3 du lac de données.
3. Créez une base de données pour organiser les tables de métadonnées dans le catalogue de données.
4. Utilisez un plan pour créer un flux de travail. Exécutez le flux de travail pour ingérer les données d'une source de données.
5. Configurez vos autorisations Lake Formation pour permettre à d'autres personnes de gérer les données du catalogue de données et du lac de données.
6. Configurez Amazon Athena pour interroger les données que vous avez importées dans votre lac de données Amazon S3.
7. Pour certains types de banques de données, configurez Amazon Redshift Spectrum pour interroger les données que vous avez importées dans votre lac de données Amazon S3.

Rubriques

- Public visé
- Prérequis
- Étape 1 : créer un utilisateur d'analyste de données
- Étape 2 : ajouter des autorisations pour lire les AWS CloudTrail journaux au rôle de flux de travail
- Étape 3 : créer un compartiment Amazon S3 pour le lac de données
- Étape 4 : enregistrer un chemin Amazon S3
- Étape 5 : accorder des autorisations de localisation des données
- Étape 6 : Création d'une base de données dans le catalogue de données
- Étape 7 : Accorder des autorisations de données
- Étape 8 : Utiliser un plan pour créer un flux de travail
- Étape 9 : Exécuter le flux de travail
- Étape 10 : Autorisez SELECT sur les tables
- Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena

Public visé

Le tableau suivant répertorie les rôles utilisés dans ce didacticiel pour créer un lac de données.

Public visé

Rôle	Description
Administrateur IAM	Possède la politique AWS gérée :`AdministratorAccess` . Peut créer des rôles IAM et des compartiments Amazon S3.
Administrateur du lac de données	Utilisateur autorisé à accéder au catalogue de données, à créer des bases de données et à accorder des autorisations Lake Formation à d'autres utilisateurs. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données.
Analyste des données	Utilisateur capable d'exécuter des requêtes sur le lac de données. Dispose uniquement des autorisations suffisantes pour exécuter des requêtes.
Rôle du flux de travail	Rôle doté des politiques IAM requises pour exécuter un flux de travail. Pour plus d'informations, consultez (Facultatif) Créez un rôle IAM pour les flux de travail.

Prérequis

Avant de commencer :

- Assurez-vous d'avoir terminé les tâches dansConfigurez AWS Lake Formation.
- Connaissez l'emplacement de vos CloudTrail journaux.
- Athena a besoin du personnage d'analyste de données pour créer un compartiment Amazon S3 pour stocker les résultats des requêtes avant d'utiliser Athena.

La connaissance de AWS Identity and Access Management (IAM) est supposée. Pour plus d'informations sur IAM, consultez le guide de l'utilisateur IAM.

Étape 1 : créer un utilisateur d'analyste de données

Cet utilisateur dispose du minimum d'autorisations pour interroger le lac de données.

1. Ouvrez la console IAM à l'adresse https://console.aws.amazon.com/iam. Connectez-vous en tant qu'utilisateur administrateur que vous avez créé Création d'un utilisateur doté d'un accès administratif ou en tant qu'utilisateur avec la politique `AdministratorAccess` AWS gérée.
2. Créez un utilisateur nommé `datalake_user` avec les paramètres suivants :
   - Activez AWS Management Console l'accès.
   - Définissez un mot de passe et n'exigez pas de réinitialisation du mot de passe.
   - Joignez la politique `AmazonAthenaFullAccess` AWS gérée.
   - Joignez la politique intégrée suivante. Nommez la stratégie `DatalakeUserBasic`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
           ],
            "Resource": "*"
        }
    ]
}
```

Étape 2 : ajouter des autorisations pour lire les AWS CloudTrail journaux au rôle de flux de travail

1. Associez la politique intégrée suivante au rôle`LakeFormationWorkflowRole`. La politique autorise la lecture de vos AWS CloudTrail journaux. Nommez la stratégie `DatalakeGetCloudTrail`.

   Pour créer le rôle `LakeFormationWorkflowRole`, consultez (Facultatif) Créez un rôle IAM pour les flux de travail.

   > **Important**
   >
   > Remplacez <your-s3-cloudtrail-bucket>par l'emplacement Amazon S3 de vos CloudTrail données.

   ```
   {
       "Version": "2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
               "Action": "s3:GetObject",
               "Resource": ["arn:aws:s3:::<your-s3-cloudtrail-bucket>/*"]
           }
       ]
   }
   ```

2. Vérifiez que trois politiques sont associées au rôle.

Étape 3 : créer un compartiment Amazon S3 pour le lac de données

Créez le compartiment Amazon S3 qui doit être l'emplacement racine de votre lac de données.

1. Ouvrez la console Amazon S3 à l'adresse https://console.aws.amazon.com/s3/ et connectez-vous en tant qu'utilisateur administrateur dans lequel vous l'avez crééeCréation d'un utilisateur doté d'un accès administratif.

2. Choisissez Create bucket, puis passez par l'assistant pour créer un bucket nommé*<yourName>*-datalake-cloudtrail, où <yourName>sont votre prénom et votre nom de famille. Par exemple : jdoe-datalake-cloudtrail.

   Pour obtenir des instructions détaillées sur la création d'un compartiment Amazon S3, consultez Création d'un compartiment.

Étape 4 : enregistrer un chemin Amazon S3

Enregistrez un chemin Amazon S3 comme emplacement racine de votre lac de données.

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
3. Choisissez Enregistrer l'emplacement, puis Parcourir.
4. Sélectionnez le *<yourName>*-datalake-cloudtrail bucket que vous avez créé précédemment, acceptez le rôle IAM par défautAWSServiceRoleForLakeFormationDataAccess, puis choisissez Enregistrer l'emplacement.

   Pour plus d'informations sur l'enregistrement des points de vente, consultezAjouter un emplacement Amazon S3 à votre lac de données.

Étape 5 : accorder des autorisations de localisation des données

Les principaux doivent disposer d'autorisations de localisation des données sur l'emplacement d'un lac de données pour créer des tables de catalogue de données ou des bases de données pointant vers cet emplacement. Vous devez accorder des autorisations de localisation des données au rôle IAM pour les flux de travail afin que le flux de travail puisse écrire sur la destination d'ingestion des données.

1. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données.
2. Choisissez Grant, puis dans la boîte de dialogue Accorder les autorisations, effectuez les sélections suivantes :

   a. Pour l'utilisateur et les rôles IAM, sélectionnezLakeFormationWorkflowRole.

   b. Pour les emplacements de stockage, choisissez votre *<yourName>*-datalake-cloudtrail compartiment.
3. Choisissez Grant (Accorder).

Pour plus d'informations sur les autorisations de localisation des données, consultezUnderlying data access control.

Étape 6 : Création d'une base de données dans le catalogue de données

Les tables de métadonnées du Lake Formation Data Catalog sont stockées dans une base de données.

1. Dans le volet de navigation, sous Catalogue de données, sélectionnez Bases de données.
2. Choisissez Créer une base de données, puis entrez le nom sous Détails de la base de donnéeslakeformation_cloudtrail.
3. Laissez les autres champs vides, puis choisissez Créer une base de données.

Étape 7 : Accorder des autorisations de données

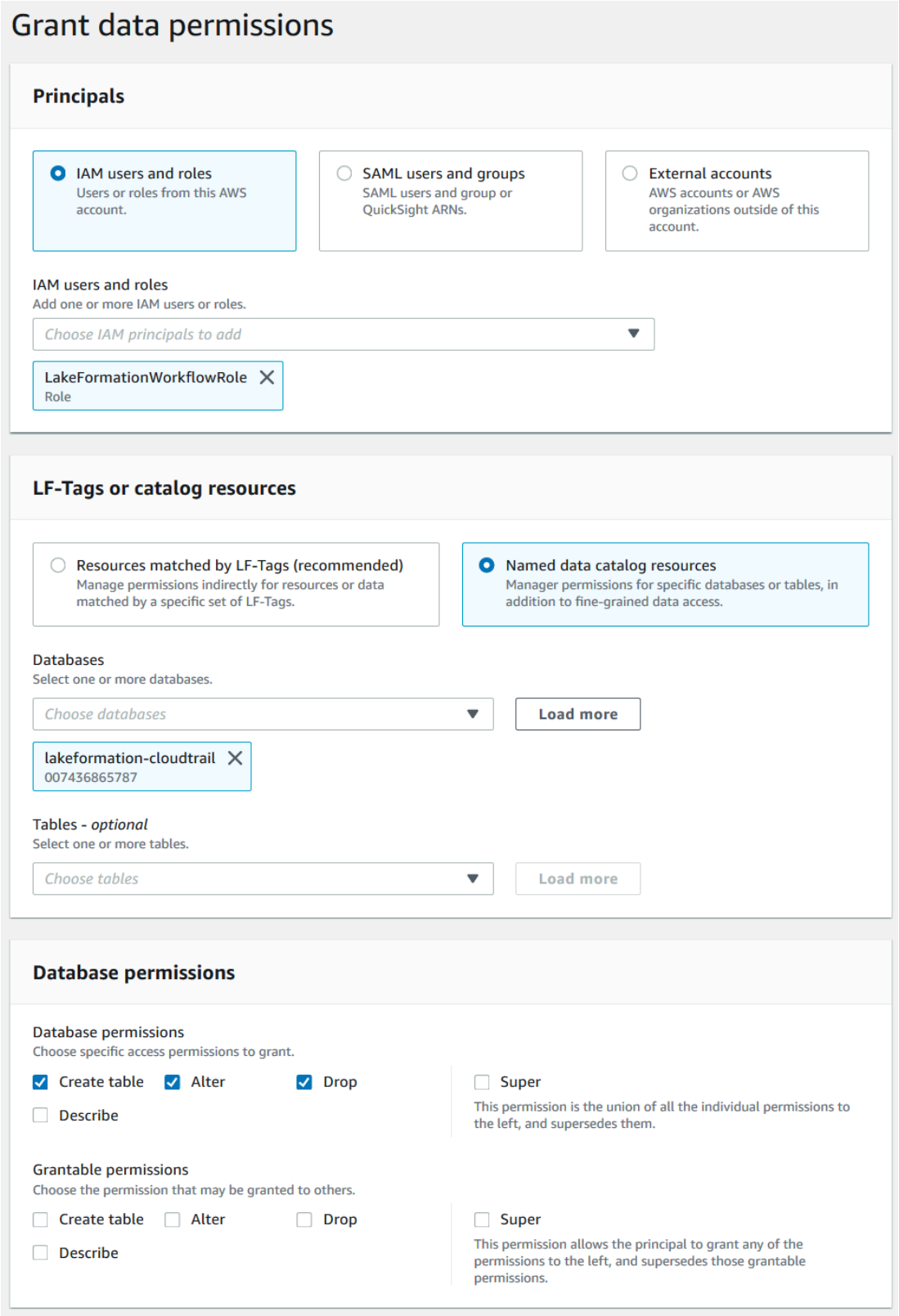
Vous devez accorder des autorisations pour créer des tables de métadonnées dans le catalogue de données. Étant donné que le flux de travail sera exécuté avec le rôleLakeFormationWorkflowRole, vous devez accorder ces autorisations au rôle.

1. Dans la console Lake Formation, dans le volet de navigation, sous Catalogue de données, sélectionnez Databases.
2. Choisissez la lakeformation_cloudtrail base de données, puis, dans la liste déroulante Actions, choisissez Grant sous le titre Permissions.
3. Dans la boîte de dialogue Accorder les autorisations relatives aux données, effectuez les sélections suivantes :

   a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez. LakeFormationWorkflowRole
   b. Sous Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées.
   c. Pour les bases de données, vous devriez voir que la lakeformation_cloudtrail base de données est déjà ajoutée.

d. Sous Autorisations de base de données, sélectionnez Créer une table, Modifier et Supprimer, puis désélectionnez Super si cette option est sélectionnée.

Votre boîte de dialogue d'autorisation des données d'octroi devrait maintenant ressembler à cette capture d'écran.



4. Choisissez Grant (Accorder).

Pour plus d'informations sur l'octroi des autorisations de Lake Formation, consultezGestion des autorisations relatives à Lake Formation.

Étape 8 : Utiliser un plan pour créer un flux de travail

Afin de lire les CloudTrail journaux, de comprendre leur structure et de créer les tables appropriées dans le catalogue de données, nous devons configurer un flux de travail composé d'un robot d'AWS Glueexploration, de tâches, de déclencheurs et de flux de travail. Les plans de Lake Formation simplifient ce processus.

Le flux de travail génère les tâches, les robots d'exploration et les déclencheurs qui découvrent et ingèrent les données dans votre lac de données. Vous créez un flux de travail basé sur l'un des plans prédéfinis de Lake Formation.

1. Dans la console Lake Formation, dans le volet de navigation, choisissez Blueprints, puis Use Blueprint.
2. Sur la page Utiliser un plan, sous Type de plan, sélectionnez. AWS CloudTrail
3. Sous Importer la source, choisissez une CloudTrail source et une date de début.
4. Sous Cible d'importation, spécifiez les paramètres suivants :

Base de données cible	`lakeformation_cloudtrail`
Emplacement de stockage cible	`s3://<yourName> -datalake-cloudtrail`
Format de données	Parquet

5. Pour la fréquence d'importation, choisissez Exécuter à la demande.
6. Sous Options d'importation, spécifiez les paramètres suivants :

Nom du flux de travail	`lakeformationcloudtrailtest`
Rôle IAM	`LakeFormationWorkflowRole`
Préfixe de table	`cloudtrailtest`

> **Note**
>
> Doit être en minuscules.

7. Choisissez Créer et attendez que la console indique que le flux de travail a été créé avec succès.

> **Tip**
>
> Avez-vous reçu le message d'erreur suivant ?
> `User: arn:aws:iam::<account-id>:user/<datalake_administrator_user> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole...`
> Si tel est le cas, vérifiez que vous avez remplacé <account-id>dans la politique en ligne l'utilisateur administrateur du lac de données par un numéro de AWS compte valide.

Étape 9 : Exécuter le flux de travail

Comme vous avez indiqué que le flux de travail est le cas run-on-demand, vous devez le démarrer manuellement.

- Sur la page Blueprints, sélectionnez le flux de travail`lakeformationcloudtrailtest`, puis dans le menu Actions, sélectionnez Démarrer.

  Au fur et à mesure que le flux de travail s'exécute, vous pouvez voir sa progression dans la colonne État de la dernière exécution. Cliquez sur le bouton d'actualisation de temps en temps.

  Le statut passe de « EN COURS » à « Découverte », puis « Importation », puis « TERMINÉ ».

  Lorsque le flux de travail est terminé :

  - Le catalogue de données comportera de nouvelles tables de métadonnées.
  - Vos CloudTrail journaux seront ingérés dans le lac de données.

  Si le flux de travail échoue, procédez comme suit :

a. Sélectionnez le flux de travail, puis dans le menu Actions, choisissez Afficher le graphique.

   Le flux de travail s'ouvre dans la AWS Glue console.

b. Assurez-vous que le flux de travail est sélectionné, puis sélectionnez l'onglet History (Historique).

c. Sous Historique, sélectionnez l'exécution la plus récente et choisissez Afficher les détails de la course.

d. Sélectionnez une tâche ou un robot d'exploration ayant échoué dans le graphique dynamique (d'exécution) et consultez le message d'erreur. Les nœuds défaillants sont rouges ou jaunes.

Étape 10 : Autorisez SELECT sur les tables

Vous devez accorder l'`SELECT`autorisation sur les nouvelles tables du catalogue de données afin que l'analyste de données puisse interroger les données vers lesquelles pointent les tables.

> **Note**
>
> Un flux de travail accorde automatiquement l'`SELECT`autorisation sur les tables qu'il crée à l'utilisateur qui l'a exécuté. Étant donné que l'administrateur du lac de données a exécuté ce flux de travail, vous devez accorder une autorisation `SELECT` à l'analyste de données.

1. Dans la console Lake Formation, dans le volet de navigation, sous Catalogue de données, sélectionnez Databases.
2. Choisissez la `lakeformation_cloudtrail` base de données, puis, dans la liste déroulante Actions, choisissez Grant sous le titre Permissions.
3. Dans la boîte de dialogue Accorder les autorisations relatives aux données, effectuez les sélections suivantes :

   a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez. `datalake_user`

   b. Sous Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées.

   c. Pour les bases de données, la `lakeformation_cloudtrail` base de données doit déjà être sélectionnée.

   d. Pour Tables, sélectionnez`cloudtrailtest-cloudtrail`.
   e. Sous Autorisations relatives aux tables et aux colonnes, choisissez Sélectionner.
4. Choisissez Grant (Accorder).

L'étape suivante est exécutée en tant qu'analyste de données.

Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena

Utilisez la Amazon Athena console pour interroger les CloudTrail données de votre lac de données.

1. Ouvrez la console Athena à l'adresse https://console.aws.amazon.com/athena/ et connectez-vous en tant qu'analyste de données, utilisateur. `datalake_user`
2. Si nécessaire, choisissez Get Started pour passer à l'éditeur de requêtes Athena.
3. Pour Data source (Source de données), choisissez AwsDataCatalog.
4. Pour Database (Base de données), sélectionnez `lakeformation_cloudtrail`.

   La liste des tables s'affiche.
5. Dans le menu déroulant (3 points disposés horizontalement) à côté du tableau`cloudtrailtest-cloudtrail`, choisissez Aperçu du tableau, puis Exécuter.

   La requête s'exécute et affiche 10 lignes de données.

   Si vous n'avez jamais utilisé Athena auparavant, vous devez d'abord configurer un emplacement Amazon S3 dans la console Athena pour stocker les résultats de la requête. Ils `datalake_user` doivent disposer des autorisations nécessaires pour accéder au compartiment Amazon S3 de votre choix.

> **Note**
>
> Maintenant que vous avez terminé le didacticiel, accordez des autorisations relatives aux données et des autorisations de localisation des données aux principaux responsables de votre organisation.

Création d'un lac de données à partir d'une source JDBC dans Lake Formation

Ce didacticiel vous explique les étapes à suivre sur la AWS Lake Formation console pour créer et charger votre premier lac de données à partir d'une source JDBC à l'aide de Lake Formation.

Rubriques

- Public visé
- Prérequis pour le didacticiel JDBC
- Étape 1 : créer un utilisateur d'analyste de données
- Étape 2 : créer une connexion dans AWS Glue
- Étape 3 : créer un compartiment Amazon S3 pour le lac de données
- Étape 4 : enregistrer un chemin Amazon S3
- Étape 5 : accorder des autorisations de localisation des données
- Étape 6 : Création d'une base de données dans le catalogue de données
- Étape 7 : Accorder des autorisations de données
- Étape 8 : Utiliser un plan pour créer un flux de travail
- Étape 9 : Exécuter le flux de travail
- Étape 10 : Autorisez SELECT sur les tables
- Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena
- Étape 12 : interroger les données du lac de données à l'aide d'Amazon Redshift Spectrum
- Étape 13 : Accorder ou révoquer les autorisations de Lake Formation à l'aide d'Amazon Redshift Spectrum

Public visé

Le tableau suivant répertorie les rôles utilisés dans ce didacticiel AWS Lake Formation JDBC.

Rôle	Description
Administrateur IAM	Utilisateur capable de créer des utilisateurs et des rôles AWS Identity and Access

Rôle	Description
	Management (IAM) ainsi que des buckets Amazon Simple Storage Service (Amazon S3). Possède la politique `AdministratorAccess` AWS gérée.
Administrateur du lac de données	Un utilisateur qui peut accéder au catalogue de données, créer des bases de données et accorder des autorisations Lake Formation à d'autres utilisateurs. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données.
Analyste des données	Un utilisateur qui peut exécuter des requêtes sur le lac de données. Dispose uniquement des autorisations suffisantes pour exécuter des requêtes.
Rôle du flux de travail	Rôle doté des politiques IAM requises pour exécuter un flux de travail.

Pour plus d'informations sur les conditions requises pour suivre le didacticiel, consultezPrérequis pour le didacticiel JDBC.

Prérequis pour le didacticiel JDBC

Avant de commencer le didacticiel AWS Lake Formation JDBC, assurez-vous d'avoir effectué les opérations suivantes :

- Effectuez les tâches définies dans Débuter avec Lake Formation.
- Choisissez un magasin de données accessible en JDBC que vous souhaitez utiliser pour le didacticiel.
- Rassemblez les informations nécessaires pour créer une AWS Glue connexion de type JDBC. Cet objet de catalogue de données inclut l'URL du magasin de données, les informations de connexion et, si le magasin de données a été créé dans un Amazon Virtual Private Cloud (Amazon VPC), des informations de configuration supplémentaires spécifiques au VPC. Pour plus d'informations,

consultez la section Définition des connexions dans le catalogue de AWS Glue données du guide du AWS Glue développeur.

Le didacticiel part du principe que vous connaissez AWS Identity and Access Management (IAM). Pour plus d'informations sur IAM, consultez le guide de l'utilisateur IAM.

Pour commencer, passez àthe section called “Étape 1 : créer un utilisateur d'analyste de données”.

Étape 1 : créer un utilisateur d'analyste de données

Au cours de cette étape, vous créez un utilisateur AWS Identity and Access Management (IAM) qui sera l'analyste de données de votre lac de données dans AWS Lake Formation.

Cet utilisateur dispose du minimum d'autorisations pour interroger le lac de données.

1. Ouvrez la console IAM à l'adresse https://console.aws.amazon.com/iam. Connectez-vous en tant qu'utilisateur administrateur que vous avez créé Création d'un utilisateur doté d'un accès administratif ou en tant qu'utilisateur avec la politique `AdministratorAccess` AWS gérée.
2. Créez un utilisateur nommé `datalake_user` avec les paramètres suivants :

   - Activez AWS Management Console l'accès.
   - Définissez un mot de passe et n'exigez pas de réinitialisation du mot de passe.
   - Joignez la politique `AmazonAthenaFullAccess` AWS gérée.
   - Joignez la politique intégrée suivante. Nommez la stratégie `DatalakeUserBasic`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
```

```
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
            ],
             "Resource": "*"
        }
    ]
}
```

Étape 2 : créer une connexion dans AWS Glue

> **Note**
>
> Ignorez cette étape si vous êtes déjà AWS Glue connecté à votre source de données JDBC.

AWS Lake Formation accède aux sources de données JDBC via une connexion. AWS Glue Une connexion est un objet du catalogue de données qui contient toutes les informations requises pour se connecter à la source de données. Vous pouvez créer une connexion à l'aide de la AWS Glue console.

Pour créer une connexion

1. Ouvrez AWS Glue la console à https://console.aws.amazon.com/glue/ l'adresse et connectez-vous en tant qu'utilisateur administrateur dans lequel vous l'avez crééeCréation d'un utilisateur doté d'un accès administratif.
2. Dans le volet de navigation, sous Data catalog (Catalogue de données), choisissez Connections (Connexions).
3. Sur la page Connectors (Connecteurs), sélectionnez Create custom connector (Créer un connecteur personnalisé).
4. Sur la page des propriétés du connecteur, entrez le **`datalake-tutorial`** nom de la connexion et choisissez JDBC comme type de connexion. Ensuite, sélectionnez Suivant.
5. Continuez à utiliser l'assistant de connexion et enregistrez la connexion.

   Pour plus d'informations sur la création d'une connexion, consultez les propriétés de connexion AWS Glue JDBC dans le manuel du AWS Glue développeur.

Étape 3 : créer un compartiment Amazon S3 pour le lac de données

Au cours de cette étape, vous créez le bucket Amazon Simple Storage Service (Amazon S3) qui doit être l'emplacement racine de votre lac de données.

1. Ouvrez la console Amazon S3 à l'adresse https://console.aws.amazon.com/s3/ et connectez-vous en tant qu'utilisateur administrateur dans lequel vous l'avez crééeCréation d'un utilisateur doté d'un accès administratif.
2. Choisissez Create bucket, puis passez par l'assistant pour créer un bucket nommé`<yourName>-datalake-tutorial`, où <yourName>sont votre prénom et votre nom de famille. Par exemple : `jdoe-datalake-tutorial`.

   Pour obtenir des instructions détaillées sur la création d'un compartiment Amazon S3, consultez Comment créer un compartiment S3 ? dans le guide de l'utilisateur d'Amazon Simple Storage Service.

Étape 4 : enregistrer un chemin Amazon S3

Au cours de cette étape, vous devez enregistrer un chemin Amazon Simple Storage Service (Amazon S3) comme emplacement racine de votre lac de données.

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
3. Choisissez Enregistrer l'emplacement, puis choisissez Parcourir.
4. Sélectionnez le `<yourName>-datalake-tutorial` bucket que vous avez créé précédemment, acceptez le rôle IAM par défaut`AWSServiceRoleForLakeFormationDataAccess`, puis choisissez Enregistrer l'emplacement.

   Pour plus d'informations sur l'enregistrement des points de vente, consultezAjouter un emplacement Amazon S3 à votre lac de données.

Étape 5 : accorder des autorisations de localisation des données

Les principaux doivent disposer d'autorisations de localisation des données sur l'emplacement d'un lac de données pour créer des tables de catalogue de données ou des bases de données pointant

vers cet emplacement. Vous devez accorder des autorisations de localisation des données au rôle IAM pour les flux de travail afin que le flux de travail puisse écrire sur la destination d'ingestion des données.

1. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Data locations.
2. Choisissez Grant, puis dans la boîte de dialogue Accorder des autorisations, procédez comme suit :
   a. Pour l'utilisateur et les rôles IAM, sélectionnez`LakeFormationWorkflowRole`.
   b. Pour les emplacements de stockage, choisissez votre *<yourName>*`-datalake-tutorial` compartiment.
3. Choisissez Grant (Accorder).

Pour plus d'informations sur les autorisations de localisation des données, consultezUnderlying data access control.

Étape 6 : Création d'une base de données dans le catalogue de données

Les tables de métadonnées du Lake Formation Data Catalog sont stockées dans une base de données.

1. Sur la console Lake Formation, dans le volet de navigation, sous Catalogue de données, sélectionnez Databases.
2. Choisissez Créer une base de données, puis entrez le nom sous Détails de la base de données`lakeformation_tutorial`.
3. Laissez les autres champs vides, puis choisissez Créer une base de données.

Étape 7 : Accorder des autorisations de données

Vous devez accorder des autorisations pour créer des tables de métadonnées dans le catalogue de données. Étant donné que le flux de travail s'exécute avec le rôle`LakeFormationWorkflowRole`, vous devez accorder ces autorisations au rôle.

1. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.

2. Choisissez Grant, puis dans la boîte de dialogue Accorder les autorisations relatives aux données, procédez comme suit :

   a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez. `LakeFormationWorkflowRole`

   b. Sous Balises LF ou ressources de catalogue, choisissez Ressources de catalogue de données nommées.

   c. Pour Bases de données, choisissez la base de données que vous avez créée précédemment,`lakeformation_tutorial`.

   d. Sous Autorisations de base de données, sélectionnez Créer une table, Modifier et Supprimer, puis désélectionnez Super si cette option est sélectionnée.

3. Choisissez Grant (Accorder).

Pour plus d'informations sur l'octroi des autorisations de Lake Formation, consultezVue d'ensemble des autorisations relatives à Lake Formation .

Étape 8 : Utiliser un plan pour créer un flux de travail

Le AWS Lake Formation flux de travail génère les AWS Glue tâches, les robots d'exploration et les déclencheurs qui découvrent et ingèrent les données dans votre lac de données. Vous créez un flux de travail basé sur l'un des plans prédéfinis de Lake Formation.

1. Sur la console Lake Formation, dans le volet de navigation, choisissez Blueprints, puis Use blueprint.
2. Sur la page Utiliser un plan, sous Type de plan, choisissez Instantané de base de données.
3. Sous Importer la source, pour Connexion à la base de données, choisissez la connexion que vous venez de créer ou choisissez une connexion existante pour votre source de données. `datalake-tutorial`
4. Dans le champ Chemin des données source, entrez le chemin à partir duquel les données seront ingérées dans le formulaire*`<database>`*/*`<schema>`*/*`<table>`*.

   Vous pouvez remplacer le caractère générique pour le pourcentage (%) par le schéma ou le tableau. <schema><database>Pour les bases de données qui prennent en charge les schémas, entrez<database>/<schema>/% pour qu'il corresponde à toutes les tables qu'elles contiennent. Oracle Database et MySQL ne prennent pas en charge le schéma dans le chemin ; entrez plutôt <database>/%. Pour Oracle Database, <database>c'est l'identifiant du système (SID).

Par exemple, si une base de données Oracle a `orcl` `orcl/%` pour SID, entrez toutes les tables auxquelles l'utilisateur indiqué dans la connexion JDCB a accès.

> **Important**
>
> Ce champ est sensible à la casse.

5. Sous Cible d'importation, spécifiez les paramètres suivants :

Base de données cible	`lakeformation_tutorial`
Emplacement de stockage cible	`s3://<yourName> -datalake-tutorial`
Format de données	(Choisissez Parquet ou CSV)

6. Pour la fréquence d'importation, choisissez Exécuter à la demande.
7. Sous Options d'importation, spécifiez les paramètres suivants :

Nom du flux de travail	`lakeformationjdbctest`
Rôle IAM	`LakeFormationWorkflowRole`
Préfixe de table	`jdbctest` **Note** Doit être en minuscules.

8. Choisissez Créer et attendez que la console indique que le flux de travail a été créé avec succès.

> **Tip**
>
> Avez-vous reçu le message d'erreur suivant ?
> `User: arn:aws:iam::<account-id>:user/<datalake_administrator_user> is not authorized to`

`perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/LakeFormationWorkflowRole...`

Si tel est le cas, vérifiez que vous avez remplacé <account-id>dans la politique en ligne l'utilisateur administrateur du lac de données par un numéro de AWS compte valide.

Étape 9 : Exécuter le flux de travail

Comme vous avez indiqué que le flux de travail est run-on-demand, vous devez le démarrer manuellement dans AWS Lake Formation.

1. Sur la console Lake Formation, sur la page Blueprints, sélectionnez le flux de travail`lakeformationjdbctest`.
2. Choisissez Actions, puis sélectionnez Démarrer.
3. Au fur et à mesure que le flux de travail s'exécute, consultez sa progression dans la colonne État de la dernière exécution. Cliquez sur le bouton d'actualisation de temps en temps.

   Le statut passe de « EN COURS » à « Découverte », puis « Importation », puis « TERMINÉ ».

   Lorsque le flux de travail est terminé :

   - Le catalogue de données contient de nouvelles tables de métadonnées.
   - Vos données sont ingérées dans le lac de données.

   Si le flux de travail échoue, procédez comme suit :

   a. Sélectionnez un flux de travail. Choisissez Actions, puis Afficher le graphique.

      Le flux de travail s'ouvre dans la AWS Glue console.

   b. Sélectionnez le flux de travail et cliquez sur l'onglet Historique.

   c. Sélectionnez l'exécution la plus récente et choisissez Afficher les détails de la course.

   d. Sélectionnez une tâche ou un robot d'exploration ayant échoué dans le graphique dynamique (d'exécution) et consultez le message d'erreur. Les nœuds défaillants sont rouges ou jaunes.

Étape 10 : Autorisez SELECT sur les tables

Vous devez accorder l'`SELECT`autorisation sur les nouvelles tables du catalogue de données AWS Lake Formation afin que l'analyste de données puisse interroger les données vers lesquelles pointent les tables.

> **Note**
>
> Un flux de travail accorde automatiquement l'`SELECT`autorisation sur les tables qu'il crée à l'utilisateur qui l'a exécuté. Étant donné que l'administrateur du lac de données a exécuté ce flux de travail, vous devez accorder une autorisation `SELECT` à l'analyste de données.

1. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.
2. Choisissez Grant, puis dans la boîte de dialogue Accorder les autorisations relatives aux données, procédez comme suit :
   a. Sous Principaux, pour Utilisateur et rôles IAM, sélectionnez. `datalake_user`
   b. Sous Balises LF ou ressources de catalogue, choisissez Ressources de catalogue de données nommées.
   c. Pour Bases de données, sélectionnez`lakeformation_tutorial`.

      La liste des tables s'affiche.
   d. Pour Tables, choisissez une ou plusieurs tables dans votre source de données.
   e. Sous Autorisations relatives aux tables et aux colonnes, choisissez Sélectionner.
3. Choisissez Grant (Accorder).

L'étape suivante est exécutée en tant qu'analyste de données.

Étape 11 : Interrogez le lac de données à l'aide de Amazon Athena

Utilisez la Amazon Athena console pour interroger les données de votre lac de données.

1. Ouvrez la console Athena à l'adresse https://console.aws.amazon.com/athena/ et connectez-vous en tant qu'analyste de données, utilisateur. `datalake_user`
2. Si nécessaire, choisissez Get Started pour passer à l'éditeur de requêtes Athena.

3. Pour Data source (Source de données), choisissez AwsDataCatalog.
4. Pour Database (Base de données), sélectionnez `lakeformation_tutorial`.

   La liste des tables s'affiche.
5. Dans le menu contextuel situé à côté de l'un des tableaux, choisissez Aperçu du tableau.

   La requête s'exécute et affiche 10 lignes de données.

Étape 12 : interroger les données du lac de données à l'aide d'Amazon Redshift Spectrum

Vous pouvez configurer Amazon Redshift Spectrum pour interroger les données que vous avez importées dans votre lac de données Amazon Simple Storage Service (Amazon S3). Créez d'abord un rôle AWS Identity and Access Management (IAM) utilisé pour lancer le cluster Amazon Redshift et pour interroger les données Amazon S3. Accordez ensuite à ce rôle les `Select` autorisations sur les tables que vous souhaitez interroger. Accordez ensuite à l'utilisateur l'autorisation d'utiliser l'éditeur de requêtes Amazon Redshift. Enfin, créez un cluster Amazon Redshift et exécutez des requêtes.

Vous créez le cluster en tant qu'administrateur et vous l'interrogez en tant qu'analyste de données.

Pour plus d'informations sur Amazon Redshift Spectrum, consultez la section Utilisation d'Amazon Redshift Spectrum pour interroger des données externes dans le manuel Amazon Redshift Database Developer Guide.

Pour configurer les autorisations permettant d'exécuter des requêtes Amazon Redshift

1. Ouvrez la console IAM à l’adresse https://console.aws.amazon.com/iam/. Connectez-vous en tant qu'utilisateur administrateur que vous avez créé Création d'un utilisateur doté d'un accès administratif (nom d'utilisateur`Administrator`) ou en tant qu'utilisateur doté de la politique `AdministratorAccess` AWS gérée.
2. Dans le panneau de navigation, choisissez Politiques.

   Si vous sélectionnez Politiques pour la première fois, la page Bienvenue dans les politiques gérées s’affiche. Sélectionnez Get started (Mise en route).
3. Sélectionnez Créer une politique.
4. Choisissez l'onglet JSON.
5. Collez le document de politique JSON suivant.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "glue:GetTable",
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetPartitions",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTags"
           ],
            "Resource": "*"
        }
    ]
}
```

6. Lorsque vous avez terminé, choisissez Review (Vérifier) pour vérifier la politique. Le programme de validation des politiques signale les éventuelles erreurs de syntaxe.
7. Sur la page Réviser la politique, entrez le nom **`RedshiftLakeFormationPolicy`** de la politique que vous créez. (Facultatif) Entrez une description. Vérifiez le récapitulatif de politique pour voir les autorisations accordées par votre politique. Sélectionnez ensuite Créer une politique pour enregistrer votre travail.
8. Dans le volet de navigation de la console IAM, sélectionnez Roles (Rôles), puis Create role (Créer un rôle).
9. Pour Sélectionner une entité de confiance, choisissez Service AWS .
10. Choisissez le service Amazon Redshift pour assumer ce rôle.
11. Choisissez le cas d'utilisation Redshift Customizable (Redshift personnalisable) pour votre service. Choisissez ensuite Suivant : Autorisations.
12. Recherchez la politique d'autorisation que vous avez créée et cochez la case à côté du nom de la politique dans la liste. `RedshiftLakeFormationPolicy`

13. Choisissez Suivant : Balises.
14. Choisissez Suivant : Vérification.
15. Pour Role name (Nom du rôle), entrez le nom **`RedshiftLakeFormationRole`**.
16. (Facultatif) Dans le champ Description du rôle, saisissez la description du nouveau rôle.
17. Passez en revue les informations du rôle, puis choisissez Create role (Créer un rôle).

Pour accorder **`Select`** des autorisations sur la table à interroger dans la base de données Lake Formation

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données, puis choisissez Grant.
3. Saisissez les informations suivantes :

   - Pour les utilisateurs et les rôles IAM, choisissez le rôle IAM que vous avez créé,. `RedshiftLakeFormationRole` Lorsque vous exécutez l'éditeur de requêtes Amazon Redshift, il utilise ce rôle IAM pour l'autorisation des données.
   - Pour Database (Base de données), sélectionnez `lakeformation_tutorial`.

     La liste des tables s'affiche.
   - Pour Table, choisissez une table dans la source de données à interroger.
   - Choisissez l'autorisation Sélectionner une table.
4. Choisissez Grant (Accorder).

Pour configurer Amazon Redshift Spectrum et exécuter des requêtes

1. Ouvrez la console Amazon Redshift à l'adresse. https://console.aws.amazon.com/redshift Connectez-vous en tant qu'utilisateur`Administrator`.
2. Choisissez Créer un cluster.
3. Sur la page Créer un cluster, entrez `redshift-lakeformation-demo` l'identifiant du cluster.
4. Pour le type de nœud, sélectionnez dc2.large.
5. Faites défiler la page vers le bas et sous Configurations de base de données, entrez ou acceptez les paramètres suivants :

- Nom d'utilisateur de l'administrateur : `awsuser`
- Mot de passe de l'utilisateur administrateur : (*`Choose a password`*)

6. Développez les autorisations du cluster et, dans la zone Rôles IAM disponibles, sélectionnez RedshiftLakeFormationRole. Ensuite, choisissez Ajouter un rôle IAM.
7. Si vous devez utiliser un port différent de la valeur par défaut de 5439, à côté de Configurations supplémentaires, désactivez l'option Utiliser les valeurs par défaut. Développez la section consacrée aux configurations de base de données et entrez un nouveau numéro de port de base de données.
8. Choisissez Créer un cluster.

   La page Clusters se charge.
9. Attendez que l'état du cluster devienne disponible. Choisissez régulièrement l'icône d'actualisation.
10. Accordez à l'analyste de données l'autorisation d'exécuter des requêtes sur le cluster. Pour ce faire, exécutez les étapes suivantes.

    a. Ouvrez la console IAM à l'adresse https://console.aws.amazon.com/iam/ et connectez-vous en tant qu'`Administrator`utilisateur.

    b. Dans le volet de navigation, choisissez Utilisateurs et associez les politiques gérées suivantes à l'utilisateur`datalake_user`.

       - `AmazonRedshiftQueryEditor`
       - `AmazonRedshiftReadOnlyAccess`
11. Déconnectez-vous de la console Amazon Redshift et reconnectez-vous en tant qu'utilisateur. `datalake_user`
12. Dans la barre d'outils verticale de gauche, cliquez sur l'icône EDITOR pour ouvrir l'éditeur de requêtes et vous connecter au cluster. Si la boîte de dialogue Connect to database apparaît, choisissez le nom du cluster`redshift-lakeformation-demo`, puis entrez le nom de la base de données**`dev`**, le nom **`awsuser`** d'utilisateur et le mot de passe que vous avez créés. Ensuite, choisissez Se connecter à la base de données.

> **Note**
>
> Si aucun paramètre de connexion ne vous est demandé et qu'un autre cluster est déjà sélectionné dans l'éditeur de requêtes, choisissez Modifier la connexion pour ouvrir la boîte de dialogue Connexion à la base de données.

13. Dans la zone de texte New Query 1, entrez et exécutez l'instruction suivante pour mapper la base de données `lakeformation_tutorial` de Lake Formation au nom du schéma Amazon Redshift : `redshift_jdbc`

> **Important**
>
> <account-id>Remplacez-le par un numéro de AWS compte valide et <region>par un nom de AWS région valide (par exemple,`us-east-1`).

```
create external schema if not exists redshift_jdbc from DATA CATALOG
 database 'lakeformation_tutorial' iam_role 'arn:aws:iam::<account-id>:role/
RedshiftLakeFormationRole' region '<region>';
```

14. Dans la liste des schémas sous Sélectionner un schéma, choisissez redshift_jdbc.

    La liste des tables s'affiche. L'éditeur de requêtes affiche uniquement les tables pour lesquelles vous avez obtenu des autorisations relatives au lac de données de Lake Formation.

15. Dans le menu contextuel situé à côté du nom d'une table, choisissez Prévisualiser les données.

    Amazon Redshift renvoie les 10 premières lignes.

    Vous pouvez désormais exécuter des requêtes sur les tables et les colonnes pour lesquelles vous avez des autorisations.

Étape 13 : Accorder ou révoquer les autorisations de Lake Formation à l'aide d'Amazon Redshift Spectrum

Amazon Redshift permet d'accorder et de révoquer les autorisations de Lake Formation sur les bases de données et les tables à l'aide d'instructions SQL modifiées. Ces instructions sont similaires aux

instructions Amazon Redshift existantes. Pour plus d'informations, consultez GRANT et REVOKE dans le manuel Amazon Redshift Database Developer Guide.

Configuration des autorisations pour les formats de stockage de tables ouvertes dans Lake Formation

AWS Lake Formation prend en charge la gestion des autorisations d'accès pour les formats de table ouverts (OTF) tels qu'Apache Iceberg, Apache Hudi et la fondation Linux Delta Lake. Dans ce didacticiel, vous allez apprendre à créer Iceberg, Hudi et Delta Lake à l'aide de tables manifestes de liens symboliques AWS Glue, à configurer des autorisations détaillées à l' AWS Glue Data Catalog aide de Lake Formation et à interroger des données à l'aide d'Amazon Athena.

> **Note**
>
> AWS les services d'analyse ne prennent pas en charge tous les formats de tables transactionnels. Pour plus d'informations, consultez Collaboration avec d'autres AWS services. Ce didacticiel couvre la création manuelle d'une nouvelle base de données et d'une table dans le catalogue de données à l'aide de AWS Glue tâches uniquement.

Ce didacticiel inclut un AWS CloudFormation modèle pour une configuration rapide. Vous pouvez le consulter et le personnaliser en fonction de vos besoins.

Rubriques

- Public visé
- Prérequis
- Étape 1 : Approvisionnez vos ressources
- Étape 2 : configurer les autorisations pour une table Iceberg
- Étape 3 : configurer les autorisations pour une table Hudi
- Étape 4 : configurer les autorisations pour une table Delta Lake
- Étape 5 : Nettoyer les AWS ressources

Public visé

Ce didacticiel est destiné aux administrateurs IAM, aux administrateurs de lacs de données et aux analystes commerciaux. Le tableau suivant répertorie les rôles utilisés dans ce didacticiel pour créer une table gouvernée à l'aide de Lake Formation.

Rôle	Description
Administrateur IAM	Utilisateur capable de créer des utilisateurs et des rôles IAM ainsi que des compartiments Amazon S3. Possède la politique `AdministratorAccess` AWS gérée.
Administrateur du lac de données	Un utilisateur qui peut accéder au catalogue de données, créer des bases de données et accorder des autorisations Lake Formation à d'autres utilisateurs. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données.
Analyste commercial	Un utilisateur qui peut exécuter des requêtes sur le lac de données. Dispose des autorisations nécessaires pour exécuter des requêtes.

Prérequis

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant auquel vous pouvez vous connecter en tant qu'utilisateur avec les autorisations appropriées. Pour plus d'informations, consultez Inscrivez-vous pour un Compte AWS et Création d'un utilisateur doté d'un accès administratif.

Le didacticiel part du principe que vous connaissez les rôles et les politiques IAM. Pour plus d'informations sur IAM, consultez le guide de l'utilisateur IAM.

Vous devez configurer les AWS ressources suivantes pour suivre ce didacticiel :

- Utilisateur administrateur du lac de données

- Données sur la formation des lacs (paramètres des lacs)
- Version 3 du moteur Amazon Athena

Pour créer un administrateur de lac de données

1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant qu'utilisateur administrateur. Vous allez créer des ressources dans la région USA Est (Virginie du Nord) pour ce didacticiel.
2. Sur la console Lake Formation, dans le volet de navigation, sous Autorisations, sélectionnez Administrative roles and tasks.
3. Sélectionnez Choisir les administrateurs sous Administrateurs du lac de données.
4. Dans la fenêtre contextuelle, Gérer les administrateurs des lacs de données, sous Utilisateurs et rôles IAM, choisissez Utilisateur administrateur IAM.
5. Choisissez Enregistrer.

Pour activer les paramètres du lac de données

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/. Dans le volet de navigation, sous Catalogue de données, sélectionnez Paramètres. Décochez les cases suivantes :
   - Utilisez uniquement le contrôle d'accès IAM pour les nouvelles bases de données.
   - Utilisez uniquement le contrôle d'accès IAM pour les nouvelles tables dans les nouvelles bases de données.
2. Dans les paramètres des versions multi-comptes, choisissez la version 3 comme version multi-comptes.
3. Choisissez Enregistrer.

Pour mettre à niveau le moteur Amazon Athena vers la version 3

1. Ouvrez la console Athena à l'adresse https://console.aws.amazon.com/athena/.
2. Sélectionnez le groupe de travail, puis le groupe de travail principal.
3. Assurez-vous que le groupe de travail dispose d'une version minimale de 3. Si ce n'est pas le cas, modifiez le groupe de travail, choisissez Manual for Upgrade query engine, puis sélectionnez la version 3.

4. Sélectionnez Enregistrer les modifications.

Étape 1 : Approvisionnez vos ressources

Cette section explique comment configurer les AWS ressources à l'aide d'un AWS CloudFormation modèle.

Pour créer vos ressources à l'aide d' AWS CloudFormation un modèle

1. Connectez-vous à la AWS CloudFormation console à l'adresse https://console.aws.amazon.com/cloudformation en tant qu'administrateur IAM dans la région USA Est (Virginie du Nord).
2. Choisissez Launch Stack.
3. Choisissez Next sur l'écran de création d'une pile.
4. Entrez un nom de pile.
5. Choisissez Suivant.
6. Sur la page suivante, choisissez Next.
7. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
8. Choisissez Créer.

   La création de la pile peut prendre jusqu'à deux minutes.

Le lancement de la pile de formation dans le cloud crée les ressources suivantes :

- lf-otf-datalake-123456789012 — Compartiment Amazon S3 pour stocker des données

> **Note**
>
> L'identifiant de compte ajouté au nom du compartiment Amazon S3 est remplacé par votre identifiant de compte.

- lf-otf-tutorial-123456789012 — Compartiment Amazon S3 pour stocker les résultats des requêtes et les scripts de travail AWS Glue
- lficebergdb — Base de données Iceberg AWS Glue
- lfhudidb — Base de données Hudi AWS Glue

- lfdeltadb — Base de données Delta AWS Glue
- native-iceberg-create — AWS Glue tâche qui crée une table Iceberg dans le catalogue de données
- native-hudi-create — AWS Glue tâche qui crée une table Hudi dans le catalogue de données
- native-delta-create — AWS Glue tâche qui crée une table Delta dans le catalogue de données
- LF-OTF- GlueServiceRole — Rôle IAM auquel vous passez AWS Glue pour exécuter les tâches. Ce rôle est associé aux politiques requises pour accéder aux ressources telles que le catalogue de données, le compartiment Amazon S3, etc.
- LF-OTF- RegisterRole — Rôle IAM pour enregistrer le site Amazon S3 auprès de Lake Formation. Ce rôle est `LF-Data-Lake-Storage-Policy` rattaché au rôle.
- lf-consumer-analystuser — Utilisateur IAM pour interroger les données à l'aide d'Athena
- lf-consumer-analystuser-credentials — Mot de passe de l'utilisateur de l'analyste de données enregistré dans AWS Secrets Manager

Une fois les créations de piles terminées, accédez à l'onglet de sortie et notez les valeurs pour :

- AthenaQueryResultLocation — Emplacement Amazon S3 pour la sortie de requête Athena
- BusinessAnalystUserCredentials — Mot de passe de l'utilisateur de l'analyste de données

  Pour récupérer la valeur du mot de passe :

  1. Choisissez la `lf-consumer-analystuser-credentials` valeur en accédant à la console Secrets Manager.
  2. Dans la section Secret value (Valeur du secret), choisissez Retrieve secret value (Récupérer la valeur du secret).
  3. Notez la valeur secrète du mot de passe.

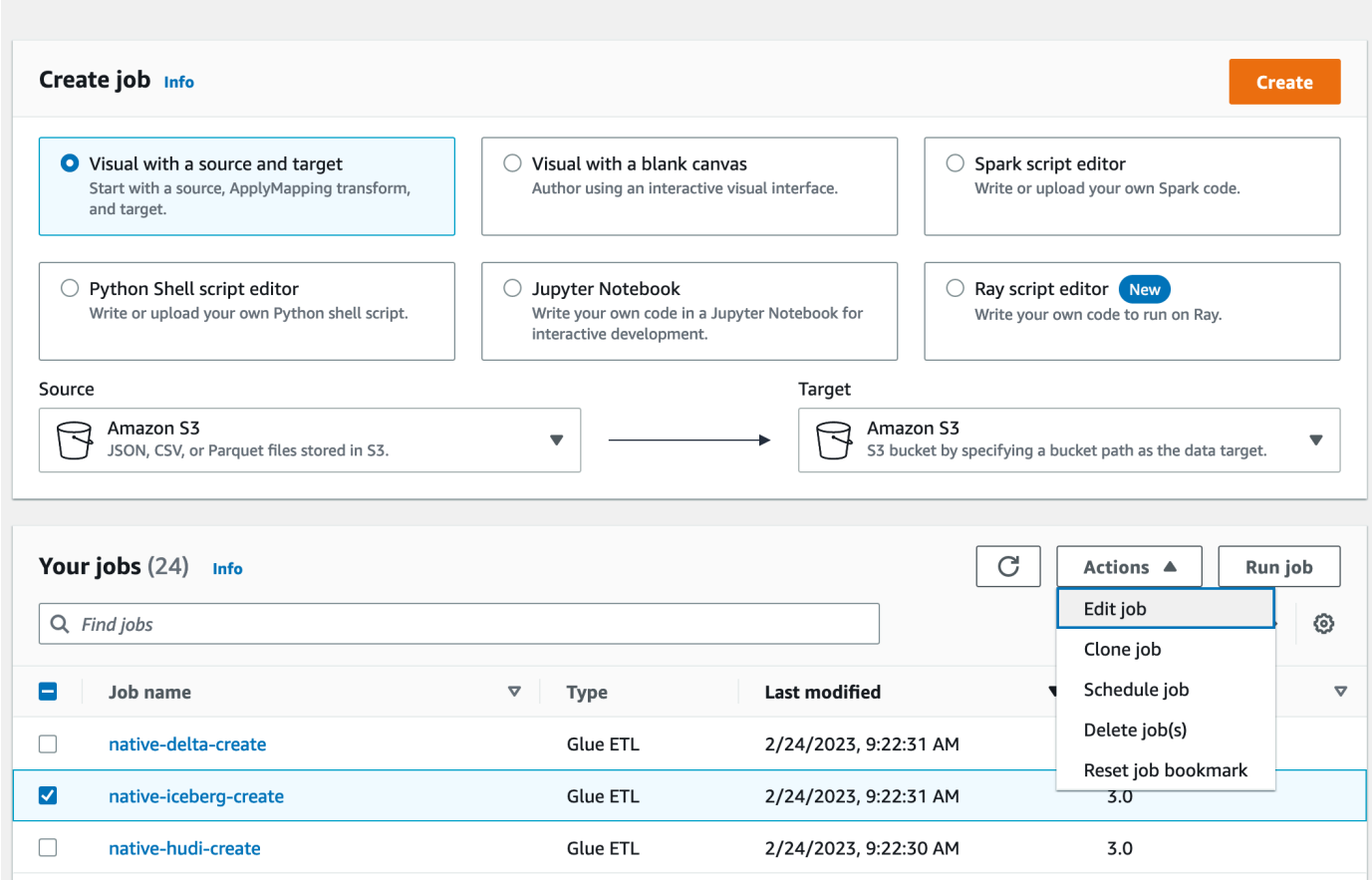
Étape 2 : configurer les autorisations pour une table Iceberg

Dans cette section, vous allez apprendre à créer une table Iceberg dans Amazon Athena AWS Glue Data Catalog, à configurer des autorisations de données et à interroger des données à AWS Lake Formation l'aide d'Amazon Athena.

Pour créer une table Iceberg

Au cours de cette étape, vous allez exécuter une AWS Glue tâche qui crée une table transactionnelle Iceberg dans le catalogue de données.

1. Ouvrez la AWS Glue console à l'adresse https://console.aws.amazon.com/glue/ dans la région USA Est (Virginie du Nord) en tant qu'utilisateur administrateur du lac de données.
2. Choisissez des tâches dans le volet de navigation de gauche.
3. Sélectionnez `native-iceberg-create`.



4. Sous Actions, sélectionnez Modifier la tâche.
5. Sous Détails du Job, développez les propriétés avancées et cochez la case à côté de Utiliser AWS Glue Data Catalog comme métastore Hive pour ajouter les métadonnées de la table dans le. AWS Glue Data Catalog Cela permet de AWS Glue Data Catalog définir le métastore pour les ressources du catalogue de données utilisées dans le cadre de la tâche et permet d'appliquer les autorisations de Lake Formation ultérieurement aux ressources du catalogue.
6. Choisissez Enregistrer.
7. Cliquez sur Exécuter. Vous pouvez consulter le statut de la tâche pendant son exécution.

   Pour plus d'informations sur les AWS Glue tâches, consultez la section Utilisation des tâches sur la AWS Glue console dans le Guide du AWS Glue développeur.

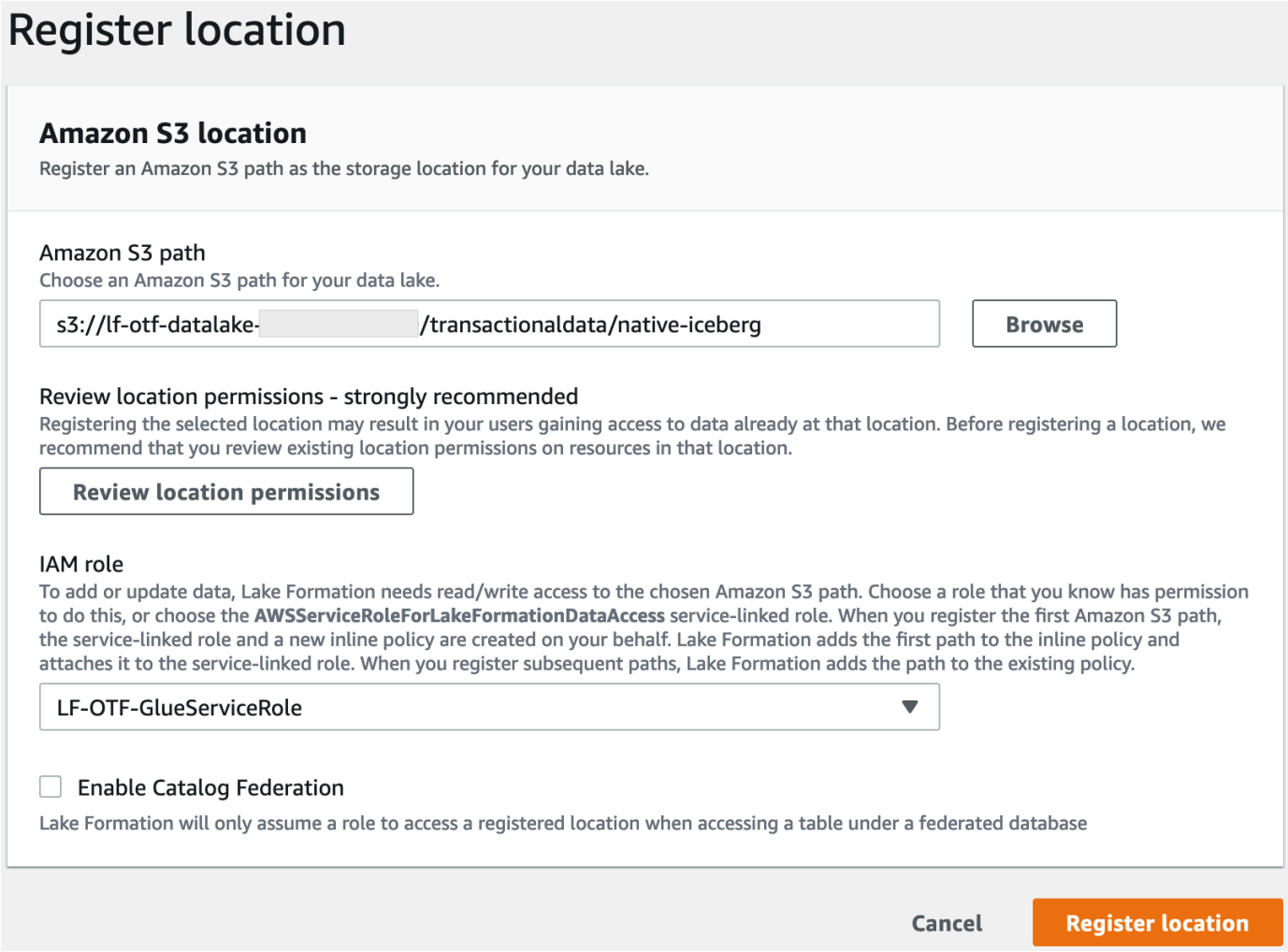
Cette tâche crée une table Iceberg nommée `product` dans la `lficebergdb` base de données. Vérifiez le tableau des produits dans la console Lake Formation.

Pour enregistrer l'emplacement des données auprès de Lake Formation

Enregistrez ensuite le chemin Amazon S3 comme emplacement de votre lac de données.

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant qu'utilisateur administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, choisissez Emplacement des données.
3. Dans le coin supérieur droit de la console, choisissez Enregistrer l'emplacement.
4. Sur la page Enregistrer l'emplacement, entrez les informations suivantes :

   - Chemin Amazon S3 : choisissez Browse et sélectionnez`lf-otf-datalake-123456789012`. Cliquez sur la flèche droite (**>**) à côté de l'emplacement racine d'Amazon S3 pour accéder à cet `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-iceberg` emplacement.
   - Rôle IAM — Choisissez `LF-OTF-RegisterRole` comme rôle IAM.
   - Choisissez Enregistrer l'emplacement.

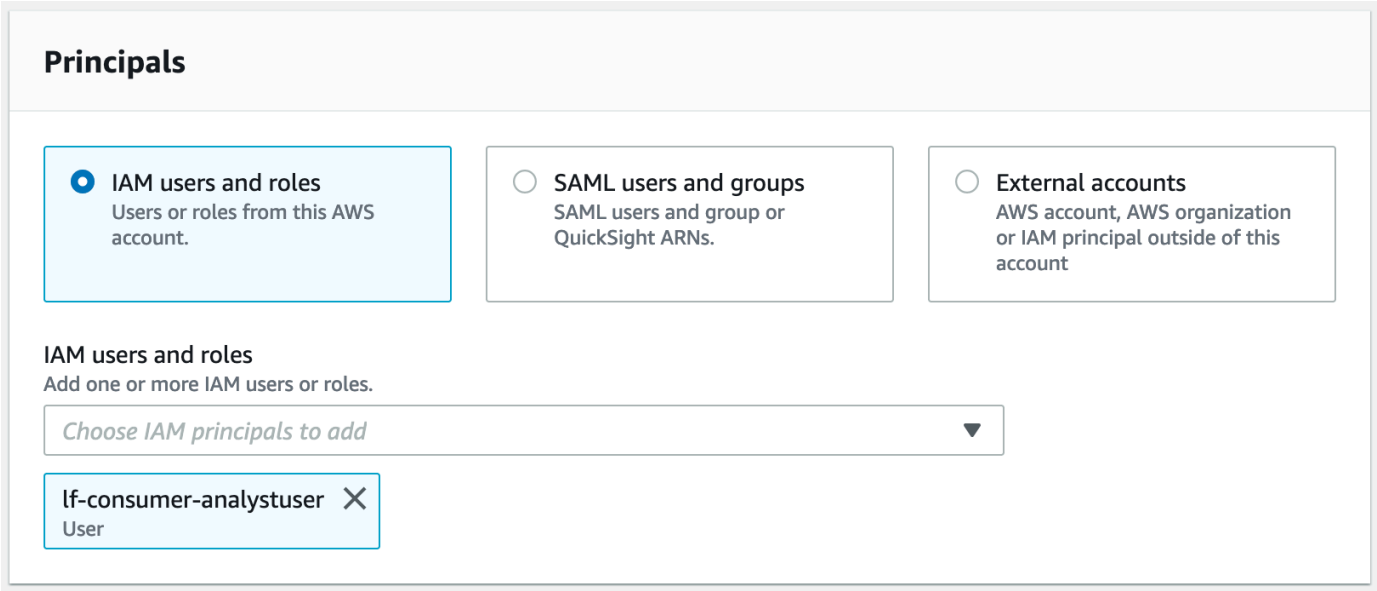


Pour plus d'informations sur l'enregistrement d'un emplacement de données auprès de Lake Formation, consultezAjouter un emplacement Amazon S3 à votre lac de données.

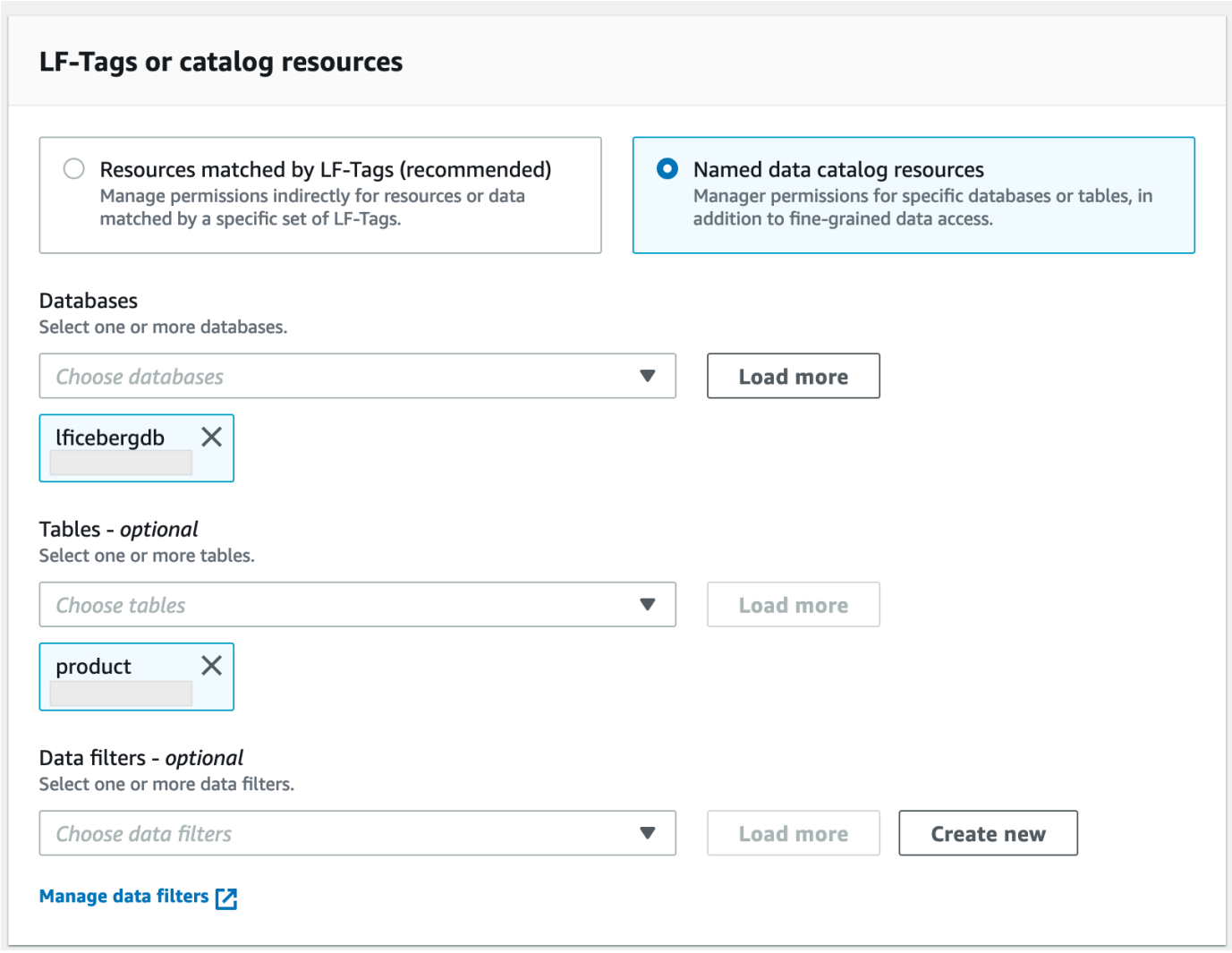
Pour accorder des autorisations à Lake Formation sur la table Iceberg

Au cours de cette étape, nous accorderons des autorisations de lac de données à l'utilisateur Business Analyst.

1. Sous Autorisations du lac de données, choisissez Grant.
2. Sur l'écran Accorder les autorisations relatives aux données, choisissez Utilisateurs et rôles IAM.
3. `lf-consumer-analystuser`Choisissez dans le menu déroulant.



4. Choisissez Ressource de catalogue de données nommée.
5. Pour les bases de données, choisissez`lficebergdb`.
6. Pour Tables, sélectionnez`product`.



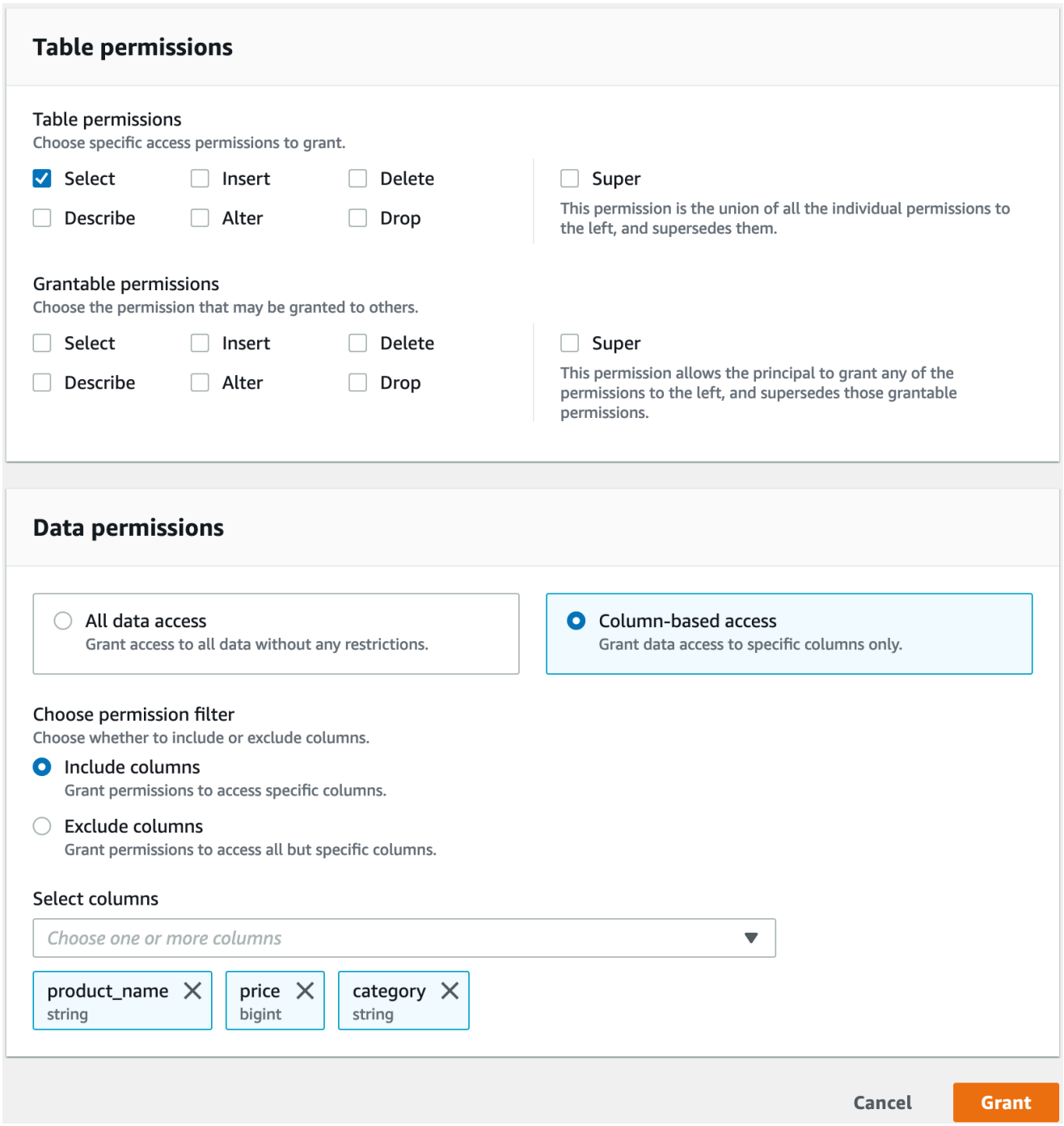
7. Vous pouvez ensuite accorder un accès basé sur les colonnes en spécifiant des colonnes.

   a. Sous Autorisations relatives aux tables, choisissez Sélectionner.

   b. Sous Autorisations relatives aux données, choisissez Accès basé sur les colonnes, choisissez Inclure les colonnes.

   c. Choisissez `product_nameprice`, et `category` colonnes.

   d. Choisissez Grant (Accorder).



Pour interroger la table Iceberg à l'aide d'Athena

Vous pouvez maintenant commencer à interroger la table Iceberg que vous avez créée avec Athena. Si c'est la première fois que vous exécutez des requêtes dans Athena, vous devez configurer l'emplacement des résultats des requêtes. Pour plus d'informations, consultez la section Spécification de l'emplacement des résultats d'une requête.

1. Déconnectez-vous en tant qu'administrateur du lac de données et connectez-vous comme `lf-consumer-analystuser` dans la région USA Est (Virginie du Nord) en utilisant le mot de passe indiqué précédemment dans le AWS CloudFormation résultat.
2. Ouvrez la console Athena à l'adresse https://console.aws.amazon.com/athena/.
3. Choisissez Paramètres, puis sélectionnez Gérer.
4. Dans le champ Emplacement du résultat de la requête, entrez le chemin d'accès au compartiment que vous avez créé dans AWS CloudFormation les sorties. Copiez la valeur de **`AthenaQueryResultLocation`** (s3 ://lf-otf-tutorial-123456789012/athena-results/) et choisissez Enregistrer.
5. Exécutez la requête suivante pour prévisualiser 10 enregistrements stockés dans la table Iceberg :

   ```
   select * from lficebergdb.product limit 10;
   ```

   Pour plus d'informations sur l'interrogation de tables Iceberg à l'aide d'Athena, consultez la section Interrogation de tables Iceberg dans le guide de l'utilisateur d'Amazon Athena.

Étape 3 : configurer les autorisations pour une table Hudi

Dans cette section, vous allez apprendre à créer une table Hudi dans le AWS Glue Data Catalog, à configurer les autorisations de données et à interroger des données à AWS Lake Formation l'aide d'Amazon Athena.

Pour créer une table Hudi

Au cours de cette étape, vous allez exécuter une AWS Glue tâche qui crée une table transactionnelle Hudi dans le catalogue de données.

1. Connectez-vous à la AWS Glue console à l'adresse https://console.aws.amazon.com/glue/ dans la région USA Est (Virginie du Nord)

   en tant qu'utilisateur administrateur du lac de données.
2. Choisissez des tâches dans le volet de navigation de gauche.
3. Sélectionnez `native-hudi-create`.
4. Sous Actions, sélectionnez Modifier la tâche.

5. Sous Détails du Job, développez les propriétés avancées et cochez la case à côté de Utiliser AWS Glue Data Catalog comme métastore Hive pour ajouter les métadonnées de la table dans le. AWS Glue Data Catalog Cela permet de AWS Glue Data Catalog définir le métastore pour les ressources du catalogue de données utilisées dans le cadre de la tâche et permet d'appliquer les autorisations de Lake Formation ultérieurement aux ressources du catalogue.
6. Choisissez Enregistrer.
7. Cliquez sur Exécuter. Vous pouvez consulter le statut de la tâche pendant son exécution.

   Pour plus d'informations sur les AWS Glue tâches, consultez la section Utilisation des tâches sur la AWS Glue console dans le Guide du AWS Glue développeur.

   Cette tâche crée une table Hudi (cow) dans la base de données : lfhudidb. Vérifiez le `product` tableau dans la console Lake Formation.

Pour enregistrer l'emplacement des données auprès de Lake Formation

Enregistrez ensuite un chemin Amazon S3 comme emplacement racine de votre lac de données.

1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant qu'utilisateur administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, choisissez Emplacement des données.
3. Dans le coin supérieur droit de la console, choisissez Enregistrer l'emplacement.
4. Sur la page Enregistrer l'emplacement, entrez les informations suivantes :

   - Chemin Amazon S3 : choisissez Browse et sélectionnez`lf-otf-datalake-123456789012`. Cliquez sur la flèche droite (>) à côté de l'emplacement racine d'Amazon S3 pour accéder à cet `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-hudi` emplacement.
   - Rôle IAM — Choisissez `LF-OTF-RegisterRole` comme rôle IAM.
   - Choisissez Enregistrer l'emplacement.

Pour accorder des autorisations de data lake sur la table Hudi

Au cours de cette étape, nous accorderons des autorisations de lac de données à l'utilisateur Business Analyst.

1. Sous Autorisations du lac de données, choisissez Grant.

2. Sur l'écran Accorder les autorisations relatives aux données, choisissez Utilisateurs et rôles IAM.
3. `lf-consumer-analystuser`depuis le menu déroulant.
4. Choisissez Ressource de catalogue de données nommée.
5. Pour les bases de données, choisissez`lfhudidb`.
6. Pour Tables, sélectionnez`product`.
7. Vous pouvez ensuite accorder un accès basé sur les colonnes en spécifiant des colonnes.

    a. Sous Autorisations relatives aux tables, choisissez Sélectionner.

    b. Sous Autorisations relatives aux données, choisissez Accès basé sur les colonnes, choisissez Inclure les colonnes.

    c. Choisissez `product_nameprice`, et `category` colonnes.

    d. Choisissez Grant (Accorder).

Pour interroger la table Hudi à l'aide d'Athena

Commencez maintenant à interroger la table Hudi que vous avez créée avec Athena. Si c'est la première fois que vous exécutez des requêtes dans Athena, vous devez configurer l'emplacement des résultats des requêtes. Pour plus d'informations, consultez la section Spécification de l'emplacement des résultats d'une requête.

1. Déconnectez-vous en tant qu'administrateur du lac de données et connectez-vous comme `lf-consumer-analystuser` dans la région USA Est (Virginie du Nord) en utilisant le mot de passe indiqué précédemment dans le AWS CloudFormation résultat.
2. Ouvrez la console Athena à l'adresse https://console.aws.amazon.com/athena/.
3. Choisissez Paramètres, puis sélectionnez Gérer.
4. Dans le champ Emplacement du résultat de la requête, entrez le chemin d'accès au compartiment que vous avez créé dans AWS CloudFormation les sorties. Copiez la valeur de **`AthenaQueryResultLocation`** (s3 ://lf-otf-tutorial-123456789012/athena-results/) et enregistrez.
5. Exécutez la requête suivante pour prévisualiser 10 enregistrements stockés dans la table Hudi :

    ```
    select * from lfhudidb.product limit 10;
    ```

    Pour plus d'informations sur l'interrogation des tables Hudi, consultez la section Interrogation des tables Hudi du guide de l'utilisateur Amazon Athena.

Étape 4 : configurer les autorisations pour une table Delta Lake

Dans cette section, vous allez apprendre à créer une table Delta Lake contenant un fichier manifeste de liens symboliques AWS Glue Data Catalog, à configurer les autorisations de données AWS Lake Formation et à interroger des données à l'aide d'Amazon Athena.

Pour créer une table Delta Lake

Au cours de cette étape, vous allez exécuter une AWS Glue tâche qui crée une table transactionnelle Delta Lake dans le catalogue de données.

1. Connectez-vous à la AWS Glue console à l'adresse https://console.aws.amazon.com/glue/ dans la région USA Est (Virginie du Nord)

   en tant qu'utilisateur administrateur du lac de données.
2. Choisissez des tâches dans le volet de navigation de gauche.
3. Sélectionnez `native-delta-create`.
4. Sous Actions, sélectionnez Modifier la tâche.
5. Sous Détails du Job, développez les propriétés avancées et cochez la case à côté de Utiliser AWS Glue Data Catalog comme métastore Hive pour ajouter les métadonnées de la table dans le. AWS Glue Data Catalog Cela permet de AWS Glue Data Catalog définir le métastore pour les ressources du catalogue de données utilisées dans le cadre de la tâche et permet d'appliquer les autorisations de Lake Formation ultérieurement aux ressources du catalogue.
6. Choisissez Enregistrer.
7. Choisissez Exécuter sous Actions.

   Cette tâche crée une table Delta Lake nommée `product` dans la `lfdeltadb` base de données. Vérifiez le `product` tableau dans la console Lake Formation.

Pour enregistrer l'emplacement des données auprès de Lake Formation

Enregistrez ensuite le chemin Amazon S3 comme emplacement racine de votre lac de données.

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant qu'utilisateur administrateur du lac de données.
2. Dans le volet de navigation, sous Enregistrer et ingérer, choisissez Emplacement des données.
3. Dans le coin supérieur droit de la console, choisissez Enregistrer l'emplacement.

4. Sur la page Enregistrer l'emplacement, entrez les informations suivantes :

   - Chemin Amazon S3 : choisissez Browse et sélectionnez`lf-otf-datalake-123456789012`. Cliquez sur la flèche droite (>) à côté de l'emplacement racine d'Amazon S3 pour accéder à cet `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-delta` emplacement.
   - Rôle IAM — Choisissez `LF-OTF-RegisterRole` comme rôle IAM.
   - Choisissez Enregistrer l'emplacement.

Pour accorder des autorisations de data lake sur la table Delta Lake

Au cours de cette étape, nous accorderons des autorisations de lac de données à l'utilisateur Business Analyst.

1. Sous Autorisations du lac de données, choisissez Grant.
2. Sur l'écran Accorder les autorisations relatives aux données, choisissez Utilisateurs et rôles IAM.
3. `lf-consumer-analystuser`depuis le menu déroulant.
4. Choisissez Ressource de catalogue de données nommée.
5. Pour les bases de données, choisissez`lfdeltadb`.
6. Pour Tables, sélectionnez`product`.
7. Vous pouvez ensuite accorder un accès basé sur les colonnes en spécifiant des colonnes.

   a. Sous Autorisations relatives aux tables, choisissez Sélectionner.

   b. Sous Autorisations relatives aux données, choisissez Accès basé sur les colonnes, choisissez Inclure les colonnes.

   c. Choisissez `product_nameprice`, et `category` colonnes.

   d. Choisissez Grant (Accorder).

Pour interroger la table Delta Lake à l'aide d'Athena

Commencez maintenant à interroger la table Delta Lake que vous avez créée avec Athena.
Si c'est la première fois que vous exécutez des requêtes dans Athena, vous devez configurer l'emplacement des résultats des requêtes. Pour plus d'informations, consultez la section Spécification de l'emplacement des résultats d'une requête.

1. Déconnectez-vous en tant qu'administrateur du lac de données et connectez-vous comme `BusinessAnalystUser` dans la région USA Est (Virginie du Nord) en utilisant le mot de passe indiqué précédemment dans le AWS CloudFormation résultat.
2. Ouvrez la console Athena à l'adresse https://console.aws.amazon.com/athena/.
3. Choisissez Paramètres, puis sélectionnez Gérer.
4. Dans le champ Emplacement du résultat de la requête, entrez le chemin d'accès au compartiment que vous avez créé dans AWS CloudFormation les sorties. Copiez la valeur de **`AthenaQueryResultLocation`** (s3 ://lf-otf-tutorial-123456789012/athena-results/) et enregistrez.
5. Exécutez la requête suivante pour prévisualiser 10 enregistrements stockés dans la table Delta Lake :

   ```
   select * from lfdeltadb.product limit 10;
   ```

   Pour plus d'informations sur l'interrogation des tables Delta Lake, consultez la section Interrogation des tables Delta Lake du guide de l'utilisateur Amazon Athena.

Étape 5 : Nettoyer les AWS ressources

Pour nettoyer des ressources

Pour éviter des frais indésirables Compte AWS, supprimez les AWS ressources que vous avez utilisées pour ce didacticiel.

1. Connectez-vous à la AWS CloudFormation console à l'adresse https://console.aws.amazon.com/cloudformation en tant qu'administrateur IAM.
2. Supprimez la pile de formation des nuages. Les tables que vous avez créées sont automatiquement supprimées avec la pile.

Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation

Des milliers de clients construisent des lacs de données à l'échelle du pétaoctet. AWS Nombre de ces clients ont l'habitude AWS Lake Formation de créer et de partager facilement leurs lacs de données au sein de l'entreprise. À mesure que le nombre de tables et d'utilisateurs augmente, les

gestionnaires de données et les administrateurs cherchent des moyens de gérer facilement et à grande échelle les autorisations sur les lacs de données. Le contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) résout ce problème en permettant aux gestionnaires de données de créer des balises LF (en fonction de leur classification et de leur ontologie de données) qui peuvent ensuite être attachées aux ressources.

Le LF-TBAC est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Dans Lake Formation, ces attributs sont appelés balises LF. Vous pouvez associer des balises LF aux ressources du catalogue de données et aux principes de Lake Formation. Les administrateurs des lacs de données peuvent attribuer et révoquer des autorisations sur les ressources de Lake Formation à l'aide de balises LF. Pour plus d'informations sur, voirContrôle d'accès basé sur des balises Lake Formation.

Ce didacticiel explique comment créer une politique de contrôle d'accès basée sur des balises Lake Formation à l'aide d'un ensemble de données AWS public. En outre, il montre comment interroger des tables, des bases de données et des colonnes associées à des politiques d'accès basées sur des balises Lake Formation.

Vous pouvez utiliser le LF-TBAC dans les cas d'utilisation suivants :

- L'administrateur du lac de données doit accorder l'accès à un grand nombre de tables et de principes.
- Vous souhaitez classer vos données en fonction d'une ontologie et accorder des autorisations en fonction de la classification
- L'administrateur du lac de données souhaite attribuer des autorisations de manière dynamique, d'une manière peu couplée

Voici les étapes de haut niveau pour configurer les autorisations à l'aide de LF-TBAC :

1. Le gestionnaire de données définit l'ontologie des balises à l'aide de deux balises LF : et. `Confidential` `Sensitive` Les données sont soumises à des `Confidential=True` contrôles d'accès plus stricts. Les données `Sensitive=True` nécessitent une analyse spécifique de la part de l'analyste.
2. Le data steward attribue différents niveaux d'autorisation à l'ingénieur de données pour créer des tables avec différentes balises LF.
3. L'ingénieur de données crée deux bases de données : `tag_database` et`col_tag_database`. Toutes les tables incluses `tag_database` sont configurées avec`Confidential=True`. Toutes

les tables du `col_tag_database` sont configurées avec`Confidential=False`. Certaines colonnes du tableau `col_tag_database` sont balisées `Sensitive=True` pour répondre à des besoins d'analyse spécifiques.

4. L'ingénieur de données accorde l'autorisation de lecture à l'analyste pour les tables présentant une condition d'expression spécifique `Confidential=True` et`Confidential=False`,`Sensitive=True`.
5. Grâce à cette configuration, l'analyste de données peut se concentrer sur l'analyse avec les bonnes données.

Rubriques

- Public visé
- Prérequis
- Étape 1 : Approvisionnez vos ressources
- Étape 2 : enregistrez l'emplacement de vos données, créez une ontologie LF-Tag et accordez des autorisations
- Étape 3 : Création de bases de données Lake Formation
- Étape 4 : Accorder les autorisations relatives aux tables
- Étape 5 : exécuter une requête dans Amazon Athena pour vérifier les autorisations
- Étape 6 : Nettoyer les AWS ressources

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Lorsqu'il s'agit de gérer AWS Glue Data Catalog et d'administrer les autorisations dans Lake Formation, les responsables des données des comptes producteurs ont une propriété fonctionnelle basée sur les fonctions qu'ils prennent en charge et peuvent accorder l'accès à divers consommateurs, organisations externes et comptes.

Le tableau suivant répertorie les rôles utilisés dans ce didacticiel :

Rôle	Description
Data Steward (administrateur)	L'`lf-data-steward` utilisateur dispose des droits d'accès suivants :

Rôle	Description
	• Accès en lecture à toutes les ressources du catalogue de données • Peut créer des balises LF et les associer au rôle d'ingénieur de données pour obtenir des autorisations pouvant être accordées à d'autres principaux
Ingénieur de données	`lf-data-engineer` l'utilisateur dispose des droits d'accès suivants : • Accès complet en lecture, écriture et mise à jour à toutes les ressources du catalogue de données • Autorisations de localisation des données dans le lac de données • Peut associer des balises LF et les associer au catalogue de données • Peut attacher des balises LF aux ressource s, ce qui permet d'accéder aux principau x en fonction des politiques créées par les gestionnaires de données
Analyste des données	L'`lf-data-analyst` utilisateur dispose des droits d'accès suivants : • Accès détaillé aux ressources partagées par les politiques d'accès basées sur les balises de Lake Formation

Prérequis

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour plus d’informations, consultez Exécution des tâches AWS de configuration initiale.

Le didacticiel part du principe que vous êtes familiarisé avec IAM. Pour plus d'informations sur IAM, consultez le guide de l'utilisateur IAM.

Étape 1 : Approvisionnez vos ressources

Ce didacticiel inclut un AWS CloudFormation modèle pour une configuration rapide. Vous pouvez le consulter et le personnaliser en fonction de vos besoins. Le modèle crée trois rôles différents (répertoriés dansPublic visé) pour effectuer cet exercice et copie le nyc-taxi-data jeu de données dans votre compartiment Amazon S3 local.

- Un compartiment Amazon S3
- Les paramètres appropriés de la Lake Formation
- Les ressources Amazon EC2 appropriées
- Trois rôles IAM avec informations d'identification

Créez vos ressources

1. Connectez-vous à la AWS CloudFormation console à l'adresse https://console.aws.amazon.com/cloudformation dans la région USA Est (Virginie du Nord).
2. Choisissez Launch Stack.
3. Choisissez Suivant.
4. Dans la section Configuration utilisateur, entrez le mot de passe pour trois rôles :`DataStewardUserPassword`, `DataEngineerUserPassword` et`DataAnalystUserPassword`.
5. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
6. Choisissez Créer.

   La création de la pile peut prendre jusqu'à cinq minutes.

> **Note**
>
> Une fois le didacticiel terminé, vous souhaiterez peut-être supprimer le stack afin d'éviter AWS CloudFormation de continuer à encourir des frais. Vérifiez que les ressources sont correctement supprimées dans le statut de l'événement pour la pile.

Étape 2 : enregistrez l'emplacement de vos données, créez une ontologie LF-Tag et accordez des autorisations

Au cours de cette étape, l'utilisateur du data steward définit l'ontologie des balises à l'aide de deux balises LF : `Confidential` et `Sensitive` donne à des principes IAM spécifiques la possibilité d'associer des balises LF nouvellement créées aux ressources.

Enregistrer un emplacement de données et définir l'ontologie LF-Tag

1. Effectuez la première étape en tant qu'utilisateur responsable de la gestion des données (`lf-data-steward`) pour vérifier les données dans Amazon S3 et le catalogue de données dans Lake Formation.

   a. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en `lf-data-steward` utilisant le mot de passe utilisé lors du déploiement de la AWS CloudFormation pile.

   b. Dans le volet de navigation, sous Autorisations, choisissez Rôles et tâches administratifs.

   c. Choisissez Ajouter dans la section Administrateurs du lac de données.

   d. Sur la page Ajouter un administrateur, pour les utilisateurs et les rôles IAM, choisissez l'utilisateur`lf-data-steward`.

   e. Choisissez Enregistrer pour l'ajouter `lf-data-steward` en tant qu'administrateur de Lake Formation.

2. Ensuite, mettez à jour les paramètres du catalogue de données pour utiliser l'autorisation Lake Formation pour contrôler les ressources du catalogue au lieu du contrôle d'accès basé sur IAM.

   a. Dans le volet de navigation, sous Administration, sélectionnez Paramètres du catalogue de données.

   b. Décochez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données.

   c. Décochez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables dans les nouvelles bases de données.

   d. Cliquez sur Sauvegarder

3. Enregistrez ensuite l'emplacement des données pour le lac de données.

   a. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.

   b. Choisissez Enregistrer l'emplacement.

c. Sur la page Enregistrer l'emplacement, pour le chemin Amazon S3, entrez`s3://lf-tagbased-demo-`*`Account-ID`*.

d. Pour le rôle IAM, laissez la valeur par défaut `AWSServiceRoleForLakeFormationDataAccess` telle quelle.

e. Choisissez Lake Formation comme mode d'autorisation.

f. Choisissez Enregistrer l'emplacement.

4. Créez ensuite l'ontologie en définissant une balise LF.

a. Sous Autorisations dans le volet de navigation, choisissez LF-Tags and permissions. .

b. Choisissez Ajouter un tag LF.

c. Pour Key (Clé), saisissez `Confidential`.

d. Dans le champ Valeurs, ajoutez `True` et`False`.

e. Choisissez Ajouter un tag LF.

f. Répétez les étapes pour créer le tag LF `Sensitive` avec la valeur. `True`

Vous avez créé toutes les balises LF nécessaires pour cet exercice.

Accorder des autorisations aux utilisateurs IAM

1. Donnez ensuite à des principes IAM spécifiques la possibilité d'associer des balises LF nouvellement créées aux ressources.

a. Sous Autorisations dans le volet de navigation, choisissez LF-Tags and permissions.

b. Dans la section Autorisations LF-Tag, choisissez Accorder des autorisations.

c. Pour le type d'autorisation, choisissez les autorisations de paire clé-valeur LF-Tag.

d. Sélectionnez les utilisateurs et les rôles IAM.

e. Pour les utilisateurs et les rôles IAM, recherchez et choisissez le `lf-data-engineer` rôle.

f. Dans la section LF-Tags, ajoutez la clé `Confidential` avec les valeurs `True` et`False`, et la clé `key Sensitive` avec valeur. `True`

g. Sous Autorisations, sélectionnez Décrire et associer pour les autorisations et les autorisations pouvant être accordées.

h. Choisissez Grant (Accorder).

2. Ensuite, accordez l'autorisation `lf-data-engineer` de créer des bases de données dans notre catalogue de données et dans le compartiment Amazon S3 sous-jacent créé par AWS CloudFormation.

   a. Sous Administration dans le volet de navigation, sélectionnez Rôles et tâches d'administration.

   b. Dans la section Créateurs de bases de données, choisissez Grant.

   c. Pour les utilisateurs et les rôles IAM, choisissez le `lf-data-engineer` rôle.

   d. Pour les autorisations du catalogue, sélectionnez Créer une base de données.

   e. Choisissez Grant (Accorder).

3. Ensuite, accordez des autorisations sur le compartiment Amazon S3 (`s3://lf-tagbased-demo-`*`Account-ID`*`)` à l'`lf-data-engineer`utilisateur.

   a. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données.

   b. Choisissez Grant (Accorder).

   c. Sélectionnez Mon compte.

   d. Pour les utilisateurs et les rôles IAM, choisissez le `lf-data-engineer` rôle.

   e. Pour les emplacements de stockage, entrez le compartiment Amazon S3 créé par le AWS CloudFormation modèle(`s3://lf-tagbased-demo-`*`Account-ID`*`).

   f. Choisissez Grant (Accorder).

4. Ensuite, accordez des autorisations `lf-data-engineer` pouvant être accordées sur les ressources associées à l'expression LF-Tag. `Confidential=True`

   a. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.

   b. Choisissez Grant (Accorder).

   c. Sélectionnez les utilisateurs et les rôles IAM.

   d. Choisissez le rôle`lf-data-engineer`.

   e. Dans la section Balises LF ou ressources du catalogue, sélectionnez Ressources associées aux balises LF.

   f. Choisissez Ajouter une paire clé-valeur LF-Tag.

   g. Ajoutez la clé `Confidential` avec les valeurs`True`.

   h. Dans la section Autorisations de base de données, sélectionnez Décrire pour les autorisations de base de données et les autorisations pouvant être accordées.

i. Dans la section Autorisations relatives aux tables, sélectionnez Décrire, sélectionner et modifier pour les autorisations relatives aux tables et aux autorisations pouvant être accordées.

j. Choisissez Grant (Accorder).

5. Ensuite, accordez des autorisations `lf-data-engineer` pouvant être accordées sur les ressources associées à l'expression LF-Tag. `Confidential=False`

a. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.

b. Choisissez Grant (Accorder).

c. Sélectionnez les utilisateurs et les rôles IAM.

d. Choisissez le rôle`lf-data-engineer`.

e. Sélectionnez les ressources associées aux balises LF.

f. Choisissez Ajouter un tag LF.

g. Ajoutez la clé `Confidential` avec la valeur`False`.

h. Dans la section Autorisations de base de données, sélectionnez Décrire pour les autorisations de base de données et les autorisations pouvant être accordées.

i. Dans la section Autorisations relatives aux tables et aux colonnes, ne sélectionnez rien.

j. Choisissez Grant (Accorder).

6. Ensuite, nous accordons des autorisations `lf-data-engineer` pouvant être accordées sur les ressources associées aux paires clé-valeur LF-Tag et. `Confidential=False Sensitive=True`

a. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations relatives aux données.

b. Choisissez Grant (Accorder).

c. Sélectionnez les utilisateurs et les rôles IAM.

d. Choisissez le rôle`lf-data-engineer`.

e. Dans la section Balises LF ou ressources du catalogue, sélectionnez Ressources associées aux balises LF.

f. Choisissez Ajouter un tag LF.

g. Ajoutez la clé `Confidential` avec la valeur`False`.

h. Choisissez Ajouter une paire clé-valeur LF-Tag.

i. Ajoutez la clé `Sensitive` avec la valeur`True`.

j. Dans la section Autorisations de base de données, sélectionnez Décrire pour les autorisations de base de données et les autorisations pouvant être accordées.

k. Dans la section Autorisations relatives aux tables, sélectionnez Décrire, sélectionner et modifier pour les autorisations relatives aux tables et aux autorisations pouvant être accordées.

l. Choisissez Grant (Accorder).

Étape 3 : Création de bases de données Lake Formation

Au cours de cette étape, vous créez deux bases de données et attachez des balises LF aux bases de données et à des colonnes spécifiques à des fins de test.

Créez vos bases de données et votre table pour un accès au niveau de la base de données

1. Créez d'abord la base de donnée`tag_database`, la table `source_data` et attachez les balises LF appropriées.

   a. Sur la console Lake Formation (https://console.aws.amazon.com/lakeformation/), sous Data Catalog, sélectionnez Databases.

   b. Choisissez Créer une base de données.

   c. Pour Name (Nom), saisissez `tag_database`.

   d. Dans Emplacement, entrez l'emplacement Amazon S3 créé par le AWS CloudFormation modèle(`s3://lf-tagbased-demo-`*`Account-ID`*`/tag_database/`).

   e. Désélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données.

   f. Choisissez Créer une base de données.

2. Ensuite, créez une nouvelle table à l'intérieur`tag_database`.

   a. Sur la page Bases de données, sélectionnez la base de donnée`tag_database`.

   b. Choisissez Afficher les tables, puis cliquez sur Créer une table.

   c. Pour Name (Nom), saisissez `source_data`.

   d. Pour Database (Base de données), choisissez la base de données `tag_database`.

   e. Pour Format de tableau, choisissez AWS Glue Tableau standard.

   f. Si les données se trouvent dans, sélectionnez le chemin spécifié dans mon compte.

g. Pour Inclure le chemin, entrez le chemin `tag_database` créé par le AWS CloudFormation modèle(`s3://lf-tagbased-demo`*`Account-ID`*`/tag_database/`).

h. Pour le format des données, sélectionnez CSV.

i. Sous Charger le schéma, entrez le tableau JSON suivant de structure de colonne pour créer un schéma :

```
[
                {
                     "Name": "vendorid",
                     "Type": "string"
                },
                {
                     "Name": "lpep_pickup_datetime",
                     "Type": "string"
                },
                {
                     "Name": "lpep_dropoff_datetime",
                     "Type": "string"

                },
                   {
                     "Name": "store_and_fwd_flag",
                     "Type": "string"
                },
                   {
                     "Name": "ratecodeid",
                     "Type": "string"

                },
                   {
                     "Name": "pulocationid",
                     "Type": "string"

                },
                {
                     "Name": "dolocationid",
                     "Type": "string"

                },
                   {
                     "Name": "passenger_count",
                     "Type": "string"
```

```
                },
                {
                     "Name": "trip_distance",
                     "Type": "string"

                },
                   {
                     "Name": "fare_amount",
                     "Type": "string"

                },
                {
                     "Name": "extra",
                     "Type": "string"

                },
                   {
                     "Name": "mta_tax",
                     "Type": "string"

                },
                {
                     "Name": "tip_amount",
                     "Type": "string"

                },
                   {
                     "Name": "tolls_amount",
                     "Type": "string"

                },
                {
                     "Name": "ehail_fee",
                     "Type": "string"

                },
                {
                     "Name": "improvement_surcharge",
                     "Type": "string"

                },
                {
                     "Name": "total_amount",
```

```
                    "Type": "string"

               },
               {
                    "Name": "payment_type",
                    "Type": "string"

               }
]
```

j. Sélectionnez Charger. Après avoir chargé le schéma, le schéma de table doit ressembler à la capture d'écran suivante :

#	Column Name	Data type
1	vendorid	string
2	lpep_pickup_datetime	string
3	lpep_dropoff_datetime	string
4	store_and_fwd_flag	string
5	ratecodeid	string
6	pulocationid	string
7	dolocationid	string
8	passenger_count	string
9	trip_distance	string
10	fare_amount	string
11	extra	string
12	mta_tax	string
13	tip_amount	string
14	tolls_amount	string
15	ehail_fee	string
16	improvement_surcharge	string
17	total_amount	string
18	payment_type	string

k. Sélectionnez Envoyer.

3. Ensuite, attachez des balises LF au niveau de la base de données.

   a. Sur la page Bases de données, recherchez et sélectionnez`tag_database`.

   b. Dans le menu Actions, choisissez Modifier les balises LF.

   c. Choisissez Attribuer un nouveau tag LF.

   d. Pour les clés assignées, choisissez le `Confidential` tag LF que vous avez créé précédemment.

   e. Dans le champ Valeurs, sélectionnez`True`.

   f. Choisissez Enregistrer.

   Ceci termine l'attribution du tag LF à la base de données tag_database.

Créez votre base de données et votre table pour un accès au niveau des colonnes

Répétez les étapes suivantes pour créer la base de données `col_tag_database` et la table`source_data_col_lvl`, et attachez des balises LF au niveau de la colonne.

1. Sur la page Bases de données, sélectionnez Créer une base de données.
2. Pour Name (Nom), saisissez `col_tag_database`.
3. Dans Emplacement, entrez l'emplacement Amazon S3 créé par le AWS CloudFormation modèle(`s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/`).
4. Désélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données.
5. Choisissez Créer une base de données.
6. Sur la page Bases de données, sélectionnez votre nouvelle base de données(`col_tag_database`).
7. Choisissez Afficher les tables, puis cliquez sur Créer une table.
8. Pour Name (Nom), saisissez `source_data_col_lvl`.
9. Dans Base de données, choisissez votre nouvelle base de données(`col_tag_database`).
10. Pour Format de tableau, choisissez AWS Glue Tableau standard.
11. Si les données se trouvent dans, sélectionnez le chemin spécifié dans mon compte.

12. Entrez le chemin Amazon S3 pour `col_tag_database(s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/)`.
13. Pour Format des données, sélectionnez`CSV`.
14. Sous`Upload schema`, entrez le schéma JSON suivant :

```
[
                {
                     "Name": "vendorid",
                     "Type": "string"


                },
                {
                     "Name": "lpep_pickup_datetime",
                     "Type": "string"


                },
                {
                     "Name": "lpep_dropoff_datetime",
                     "Type": "string"


                },
                   {
                     "Name": "store_and_fwd_flag",
                     "Type": "string"


                },
                   {
                     "Name": "ratecodeid",
                     "Type": "string"


                },
                   {
                     "Name": "pulocationid",
                     "Type": "string"


                },
```

```
                {
                     "Name": "dolocationid",
                     "Type": "string"


                },
                   {
                     "Name": "passenger_count",
                     "Type": "string"


                },
                {
                     "Name": "trip_distance",
                     "Type": "string"


                },
                   {
                     "Name": "fare_amount",
                     "Type": "string"


                },
                {
                     "Name": "extra",
                     "Type": "string"


                },
                   {
                     "Name": "mta_tax",
                     "Type": "string"


                },
                {
                     "Name": "tip_amount",
                     "Type": "string"


                },
                   {
                     "Name": "tolls_amount",
```

```
                    "Type": "string"


               },
               {
                    "Name": "ehail_fee",
                    "Type": "string"


               },
               {
                    "Name": "improvement_surcharge",
                    "Type": "string"


               },
               {
                    "Name": "total_amount",
                    "Type": "string"


               },
               {
                    "Name": "payment_type",
                    "Type": "string"


               }
]
```

15. Sélectionnez `Upload`. Après avoir chargé le schéma, le schéma de table doit ressembler à la capture d'écran suivante.

#	Column Name	Data type
1	vendorid	string
2	lpep_pickup_datetime	string
3	lpep_dropoff_datetime	string
4	store_and_fwd_flag	string
5	ratecodeid	string
6	pulocationid	string
7	dolocationid	string
8	passenger_count	string
9	trip_distance	string
10	fare_amount	string
11	extra	string
12	mta_tax	string
13	tip_amount	string
14	tolls_amount	string
15	ehail_fee	string
16	improvement_surcharge	string
17	total_amount	string
18	payment_type	string

16. Choisissez Soumettre pour terminer la création de la table.
17. Associez maintenant le `Sensitive=True` tag LF aux colonnes `vendorid` et. `fare_amount`

    a. Sur la page Tables, sélectionnez la table que vous avez créée`(source_data_col_lvl)`.
    b. Dans le menu Actions, sélectionnez Schéma.
    c. Sélectionnez la colonne `vendorid` et choisissez Modifier les balises LF.
    d. Pour Clés attribuées, choisissez Sensitive.
    e. Dans le champ Valeurs, sélectionnez Vrai.
    f. Choisissez Enregistrer.
18. Ensuite, associez le `Confidential=False` tag LF à. `col_tag_database` Cela est nécessaire pour `lf-data-analyst` pouvoir décrire la base de données `col_tag_database` lorsque vous êtes connecté depuis Amazon Athena.

    a. Sur la page Bases de données, recherchez et sélectionnez`col_tag_database`.
    b. Dans le menu Actions, choisissez Modifier les balises LF.
    c. Choisissez Attribuer un nouveau tag LF.
    d. Pour les clés attribuées, choisissez le `Confidential` tag LF que vous avez créé précédemment.
    e. Dans le champ Valeurs, sélectionnez`False`.
    f. Choisissez Enregistrer.

Étape 4 : Accorder les autorisations relatives aux tables

Accordez des autorisations aux analystes de données pour l'utilisation des bases de données `tag_database` et de la table à l'`col_tag_database`aide des balises LF `Confidential` et. `Sensitive`

1. Procédez comme suit pour accorder à l'`lf-data-analyst`utilisateur des autorisations sur les objets associés à la balise LF `Confidential=True` (Database:TAG_DATABASE) afin de disposer `Describe` de la base de données et des autorisations sur les tables. `Select`

    a. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant que`lf-data-engineer`.
    b. Sous Autorisations, sélectionnez Autorisations du lac de données.
    c. Choisissez Grant (Accorder).

d. Sous Principaux, sélectionnez Utilisateurs et rôles IAM.

e. Pour les utilisateurs et les rôles IAM, choisissez`lf-data-analyst`.

f. Sous Balises LF ou ressources du catalogue, sélectionnez Ressources associées aux balises LF.

g. Choisissez Ajouter un tag LF.

h. Pour Key, choisissez`Confidential`.

i. Dans le champ Valeurs, sélectionnez`True`.

j. Pour les autorisations de base de données, sélectionnez`Describe`.

k. Pour les autorisations relatives aux tables, choisissez Sélectionner et décrire.

l. Choisissez Grant (Accorder).

2. Répétez ensuite les étapes pour accorder des autorisations aux analystes de données pour l'expression LF-Tag for. `Confidential=False` Cette balise LF est utilisée pour décrire le `col_tag_database` et le tableau `source_data_col_lvl` lorsque vous êtes connecté `lf-data-analyst` depuis Amazon Athena.

   a. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant que`lf-data-engineer`.

   b. Sur la page Bases de données, sélectionnez la base de données`col_tag_database`.

   c. Choisissez Action et Grant.

   d. Sous Principaux, sélectionnez Utilisateurs et rôles IAM.

   e. Pour les utilisateurs et les rôles IAM, choisissez`lf-data-analyst`.

   f. Sélectionnez les ressources associées aux balises LF.

   g. Choisissez Ajouter un tag LF.

   h. Pour Key, choisissez`Confidential`.

   i. Pour Valeurs, choisissez`False`.

   j. Pour les autorisations de base de données, sélectionnez`Describe`.

   k. Pour les autorisations relatives aux tables, ne sélectionnez rien.

   l. Choisissez Grant (Accorder).

3. Répétez ensuite les étapes pour accorder des autorisations aux analystes de données pour l'expression LF-Tag pour `Confidential=False` et. `Sensitive=True` Cette balise LF est utilisée pour décrire le `col_tag_database` et le tableau `source_data_col_lvl` (au niveau des colonnes) lorsque vous êtes connecté depuis Amazon `lf-data-analyst` Athena.

a. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant que`lf-data-engineer`.

b. Sur la page Bases de données, sélectionnez la base de données`col_tag_database`.

c. Choisissez Action et Grant.

d. Sous Principaux, sélectionnez Utilisateurs et rôles IAM.

e. Pour les utilisateurs et les rôles IAM, choisissez`lf-data-analyst`.

f. Sélectionnez les ressources associées aux balises LF.

g. Choisissez Ajouter un tag LF.

h. Pour Key, choisissez`Confidential`.

i. Pour Valeurs, choisissez`False`.

j. Choisissez Ajouter un tag LF.

k. Pour Key, choisissez`Sensitive`.

l. Pour Valeurs, choisissez`True`.

m. Pour les autorisations de base de données, sélectionnez`Describe`.

n. Pour les autorisations relatives aux tables, sélectionnez `Select` et`Describe`.

o. Choisissez Grant (Accorder).

Étape 5 : exécuter une requête dans Amazon Athena pour vérifier les autorisations

Pour cette étape, utilisez Amazon Athena pour exécuter des `SELECT` requêtes sur les deux tables. `(source_data and source_data_col_lvl)` Utilisez le chemin Amazon S3 comme emplacement des résultats de la requête`(s3://lf-tagbased-demo-`*`Account-ID`*`/athena-results/)`.

1. Connectez-vous à la console Athena à l'adresse https://console.aws.amazon.com/athena/ en tant que. `lf-data-analyst`
2. Dans l'éditeur de requêtes Athena, choisissez `tag_database` dans le panneau de gauche.
3. Choisissez l'icône d'options de menu supplémentaires (trois points verticaux) à côté `source_data` et choisissez Aperçu du tableau.
4. Choisissez Exécuter la requête.

L'exécution de la requête devrait prendre quelques minutes. La requête affiche toutes les colonnes de la sortie car la balise LF est associée au niveau de la base de données et la `source_data` table en a automatiquement hérité. `LF-tag` `tag_database`

5. Exécutez une autre requête à l'aide `col_tag_database` de et`source_data_col_lvl`.

   La deuxième requête renvoie les deux colonnes étiquetées `Non-Confidential` et`Sensitive`.

6. Vous pouvez également vérifier le comportement de la politique d'accès basée sur les balises Lake Formation sur les colonnes pour lesquelles vous n'avez pas de subventions politiques. Lorsqu'une colonne non balisée est sélectionnée dans le table`source_data_col_lvl`, Athena renvoie une erreur. Par exemple, vous pouvez exécuter la requête suivante pour sélectionner des colonnes `geolocationid` non balisées :

   ```
   SELECT geolocationid FROM "col_tag_database"."source_data_col_lvl" limit 10;
   ```

Étape 6 : Nettoyer les AWS ressources

Pour éviter des frais indésirables Compte AWS, vous pouvez supprimer les AWS ressources que vous avez utilisées pour ce didacticiel.

1. Connectez-vous à la console Lake Formation en tant que `lf-data-engineer` et supprimez les bases de données `tag_database` et`col_tag_database`.
2. Ensuite, connectez-vous en tant que `lf-data-steward` et nettoyez toutes les autorisations LF-Tag, les autorisations de données et les autorisations de localisation des données accordées ci-dessus qui ont été accordées `lf-data-engineer` et. `lf-data-analyst.`
3. Connectez-vous à la console Amazon S3 en tant que propriétaire du compte à l'aide des informations d'identification IAM que vous avez utilisées pour déployer la AWS CloudFormation pile.
4. Supprimez les compartiments suivants :

   - lf-tagbased-demo-accesslogs- identifiant d'*`acct`*
   - lf-tagbased-demo- identifiant d'*`acct`*

5. Connectez-vous à AWS CloudFormation la console à l'adresse https://console.aws.amazon.com/cloudformation et supprimez la pile que vous avez créée. Attendez que le statut de la pile passe à`DELETE_COMPLETE`.

Sécurisation des lacs de données grâce au contrôle d'accès au niveau des lignes

AWS Lake Formation les autorisations au niveau des lignes vous permettent d'accéder à des lignes spécifiques d'un tableau en fonction des politiques de conformité et de gouvernance des données. Si vous avez de grandes tables stockant des milliards d'enregistrements, vous devez trouver un moyen de permettre aux différents utilisateurs et équipes d'accéder uniquement aux données qu'ils sont autorisés à consulter. Le contrôle d'accès au niveau des lignes est un moyen simple et performant de protéger les données, tout en permettant aux utilisateurs d'accéder aux données dont ils ont besoin pour effectuer leur travail. Lake Formation fournit des audits et des rapports de conformité centralisés en identifiant quels principaux ont accédé à quelles données, quand et par le biais de quels services.

Dans ce didacticiel, vous apprendrez comment fonctionnent les contrôles d'accès au niveau des lignes dans Lake Formation et comment les configurer.

Ce didacticiel inclut un AWS CloudFormation modèle permettant de configurer rapidement les ressources requises. Vous pouvez le consulter et le personnaliser en fonction de vos besoins.

Rubriques

- Public visé
- Prérequis
- Étape 1 : Approvisionnez vos ressources
- Étape 2 : Requête sans filtres de données
- Étape 3 : configurer les filtres de données et accorder des autorisations
- Étape 4 : Requête à l'aide de filtres de données
- Étape 5 : Nettoyer les AWS ressources

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Le tableau suivant répertorie les rôles et les responsabilités du propriétaire et du consommateur de données.

Rôle	Description
Administrateur IAM	Utilisateur capable de créer des utilisateurs, des rôles et des compartiments Amazon Simple Storage Service (Amazon S3). Possède la politique `AdministratorAccess` AWS gérée.
Administrateur du lac de données	Utilisateur chargé de configurer le lac de données, de créer des filtres de données et d'accorder des autorisations aux analystes de données.
Analyste des données	Un utilisateur qui peut exécuter des requêtes sur le lac de données. Les analystes de données résidant dans différents pays (pour notre cas d'utilisation, les États-Unis et le Japon) ne peuvent analyser que les avis des clients situés dans leur propre pays et, pour des raisons de conformité, ne devraient pas être en mesure de consulter les données clients situées dans d'autres pays.

Prérequis

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour plus d’informations, consultez Exécution des tâches AWS de configuration initiale.

Le didacticiel part du principe que vous connaissez IAM. Pour plus d'informations sur IAM, consultez le guide de l'utilisateur IAM.

Modifier les paramètres de Lake Formation

> **Important**
>
> Avant de lancer le AWS CloudFormation modèle, désactivez l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données/tables de Lake Formation en suivant les étapes ci-dessous :

1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ dans la région USA Est (Virginie du Nord) ou USA Ouest (Oregon).
2. Sous Catalogue de données, sélectionnez Paramètres.
3. Désélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données et Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables des nouvelles bases de données.
4. Choisissez Enregistrer.

Étape 1 : Approvisionnez vos ressources

Ce didacticiel inclut un AWS CloudFormation modèle pour une configuration rapide. Vous pouvez le consulter et le personnaliser en fonction de vos besoins. Le AWS CloudFormation modèle génère les ressources suivantes :

- Utilisateurs et politiques pour :
  - DataLakeAdmin
  - DataAnalystÉtats-Unis
  - DataAnalystJP
- Lake Formation : paramètres et autorisations du lac
- Fonction Lambda (pour les ressources AWS CloudFormation personnalisées soutenues par Lambda) utilisée pour copier des exemples de fichiers de données du compartiment Amazon S3 public vers votre compartiment Amazon S3
- Un bucket Amazon S3 qui servira de lac de données
- Une AWS Glue Data Catalog base de données, une table et une partition

Créez vos ressources

Suivez ces étapes pour créer vos ressources à l'aide du AWS CloudFormation modèle.

1. Connectez-vous à la AWS CloudFormation console à l'adresse https://console.aws.amazon.com/cloudformation dans la région USA Est (Virginie du Nord).
2. Choisissez Launch Stack.
3. Choisissez Next sur l'écran de création d'une pile.
4. Entrez un nom de pile.
5. Pour DatalakeAdminUserNameet DatalakeAdminUserPassword, entrez votre nom d'utilisateur et votre mot de passe IAM pour l'utilisateur administrateur de Data Lake.
6. Pour DataAnalystUsUserNameet DataAnalystUsUserPassword, entrez le nom d'utilisateur et le mot de passe pour le nom d'utilisateur et le mot de passe que vous souhaitez pour l'utilisateur analyste de données responsable du marché américain.
7. Pour DataAnalystJpUserNameet DataAnalystJpUserPassword, entrez le nom d'utilisateur et le mot de passe pour le nom d'utilisateur et le mot de passe que vous souhaitez pour l'utilisateur analyste de données responsable du marché japonais.
8. Pour DataLakeBucketName, entrez le nom de votre compartiment de données.
9. Pour DatabaseName, et TableNamelaissez comme valeur par défaut.
10. Choisissez Next (Suivant)
11. Sur la page suivante, choisissez Next.
12. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
13. Choisissez Créer.

    La création de la pile peut prendre une minute.

Étape 2 : Requête sans filtres de données

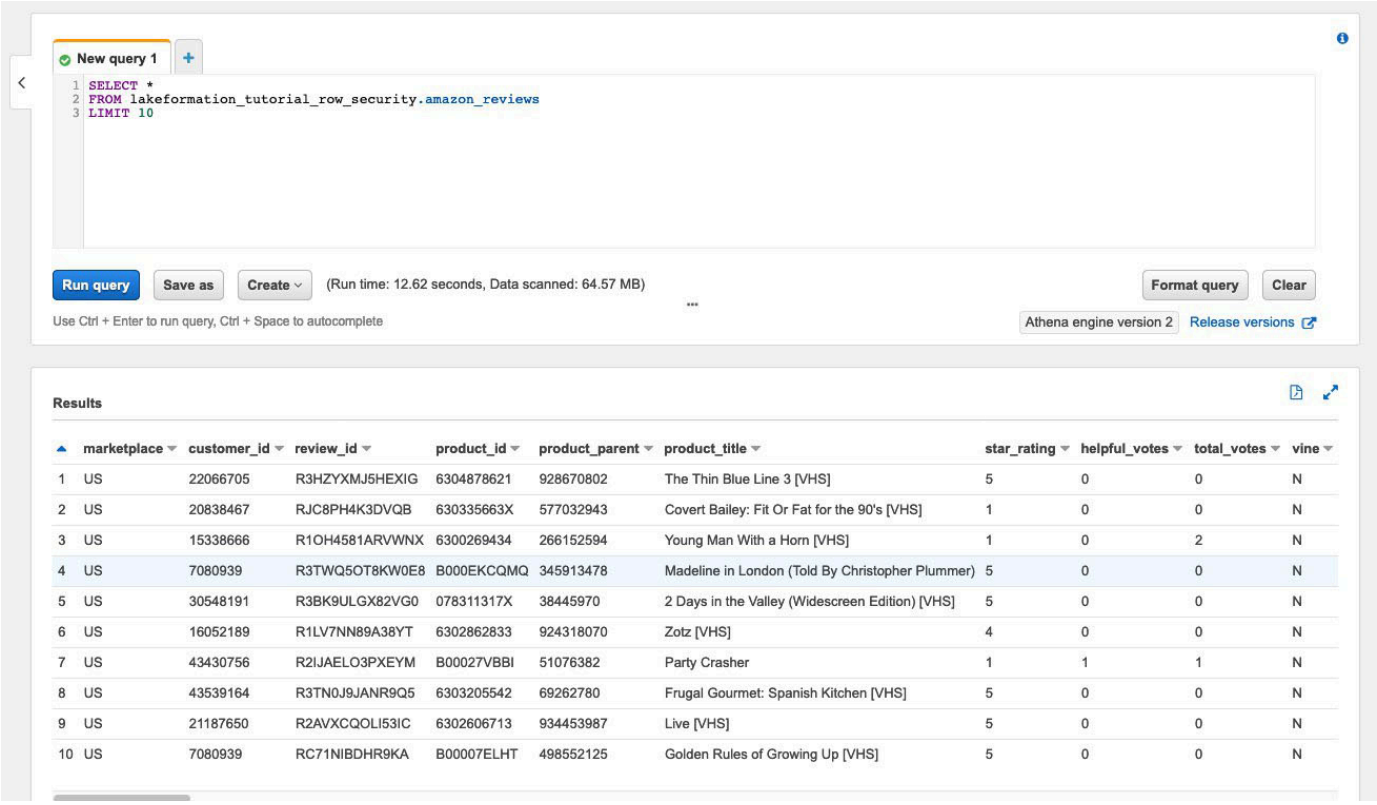
Après avoir configuré l'environnement, vous pouvez consulter le tableau des avis sur les produits. Interrogez d'abord la table sans contrôles d'accès au niveau des lignes pour vous assurer que vous pouvez voir les données. Si vous exécutez des requêtes dans Amazon Athena pour la première fois, vous devez configurer l'emplacement des résultats des requêtes.

Interrogez la table sans contrôle d'accès au niveau des lignes

1. Connectez-vous à Athena la console à l'adresse https://console.aws.amazon.com/athena/ en tant qu'`DatalakeAdmin`utilisateur et exécutez la requête suivante :

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

   La capture d'écran suivante montre le résultat de la requête. Ce tableau ne comporte qu'une seule partition. Chaque enregistrement est donc un commentaire d'évaluation pour un produit vidéo. `product_category=Video`

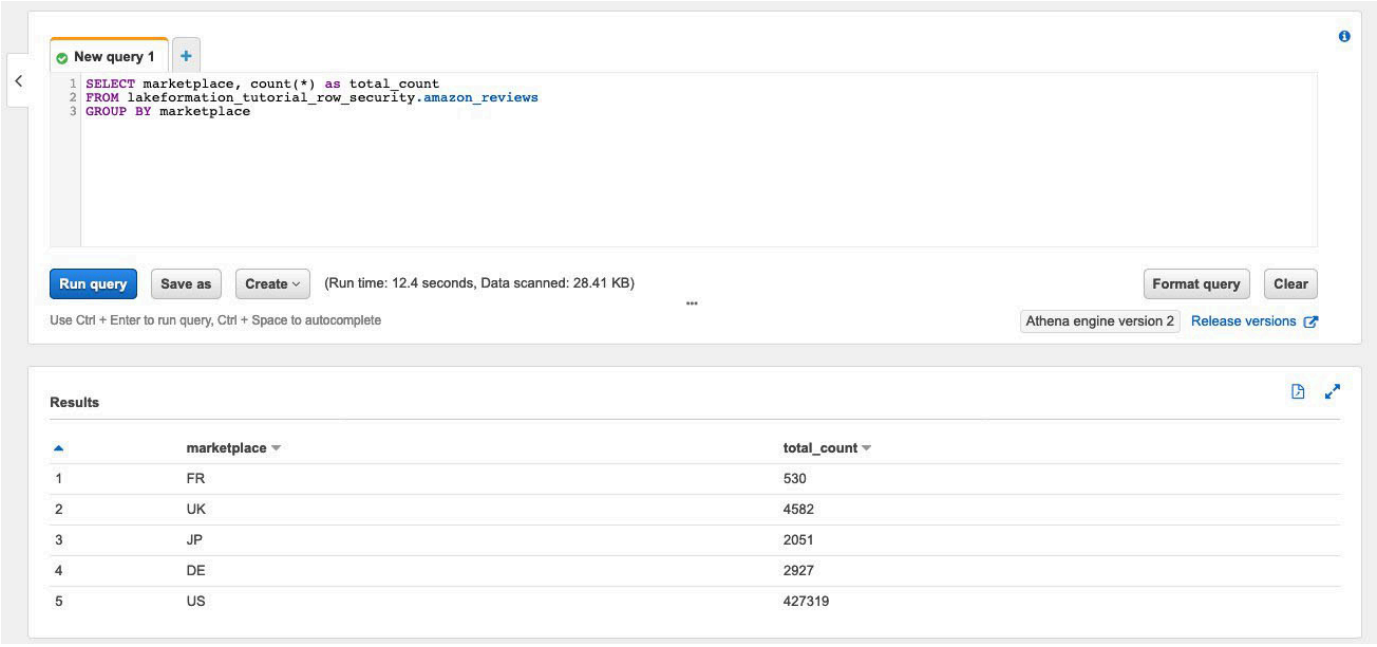


	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine
1	US	22066705	R3HZYXMJ5HEXIG	6304878621	928670802	The Thin Blue Line 3 [VHS]	5	0	0	N
2	US	20838467	RJC8PH4K3DVQB	630335663X	577032943	Covert Bailey: Fit Or Fat for the 90's [VHS]	1	0	0	N
3	US	15338666	R1OH4581ARVWNX	6300269434	266152594	Young Man With a Horn [VHS]	1	0	2	N
4	US	7080939	R3TWQ5OT8KW0E8	B000EKCQMQ	345913478	Madeline in London (Told By Christopher Plummer)	5	0	0	N
5	US	30548191	R3BK9ULGX82VG0	078311317X	38445970	2 Days in the Valley (Widescreen Edition) [VHS]	5	0	0	N
6	US	16052189	R1LV7NN89A38YT	6302862833	924318070	Zotz [VHS]	4	0	0	N
7	US	43430756	R2IJAELO3PXEYM	B00027VBBI	51076382	Party Crasher	1	1	1	N
8	US	43539164	R3TN0J9JANR9Q5	6303205542	69262780	Frugal Gourmet: Spanish Kitchen [VHS]	5	0	0	N
9	US	21187650	R2AVXCQOLI53IC	6302606713	934453987	Live [VHS]	5	0	0	N
10	US	7080939	RC71NIBDHR9KA	B00007ELHT	498552125	Golden Rules of Growing Up [VHS]	5	0	0	N

2. Exécutez ensuite une requête d'agrégation pour récupérer le nombre total d'enregistrements par`marketplace`.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
GROUP BY marketplace
```

La capture d'écran suivante montre le résultat de la requête. La `marketplace` colonne comporte cinq valeurs différentes. Au cours des étapes suivantes, vous allez configurer des filtres basés sur des lignes à l'aide de la `marketplace` colonne.



Étape 3 : configurer les filtres de données et accorder des autorisations

Ce didacticiel fait appel à deux analystes de données : l'un responsable du marché américain et l'autre du marché japonais. Chaque analyste utilise Athena pour analyser les avis des clients uniquement pour son marché spécifique. Créez deux filtres de données différents, l'un pour l'analyste responsable du marché américain et l'autre pour celui responsable du marché japonais. Accordez ensuite aux analystes leurs autorisations respectives.

Créez des filtres de données et accordez des autorisations

1. Créez un filtre pour restreindre l'accès aux US `marketplace` données.

   a. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ dans la région USA Est (Virginie du Nord) en tant qu'`DatalakeAdmin`utilisateur.

   b. Choisissez Filtres de données.

   c. Choisissez Créer un nouveau filtre.

   d. Dans Nom du filtre de données, entrez`amazon_reviews_US`.

e. Pour Base de données cible, choisissez la base de données`lakeformation_tutorial_row_security`.

f. Pour Table cible, choisissez la table`amazon_reviews`.

g. Pour l'accès au niveau des colonnes, laissez la valeur par défaut.

h. Pour Expression du filtre de ligne, entrez`marketplace='US'`.

i. Choisissez Create filter (Créer un filtre).

2. Créez un filtre pour restreindre l'accès aux `marketplace` données japonaises.

a. Sur la page Filtres de données, choisissez Créer un nouveau filtre.

b. Dans Nom du filtre de données, entrez`amazon_reviews_JP`.

c. Pour Base de données cible, choisissez la base de données`lakeformation_tutorial_row_security`.

d. Pour le tableau Target, choisissez le`table amazon_reviews`.

e. Pour l'accès au niveau des colonnes, laissez la valeur par défaut.

f. Pour Expression du filtre de ligne, entrez`marketplace='JP'`.

g. Choisissez Create filter (Créer un filtre).

3. Accordez ensuite des autorisations aux analystes de données utilisant ces filtres de données. Procédez comme suit pour accorder des autorisations à l'analyste de données américain (`DataAnalystUS`) :

a. Sous Autorisations, sélectionnez Autorisations du lac de données.

b. Sous Autorisation relative aux données, choisissez Grant.

c. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis sélectionnez le rôle. `DataAnalystUS`

d. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.

e. Pour Database (Base de données), sélectionnez `lakeformation_tutorial_row_security`.

f. Pour Tableaux (facultatif), sélectionnez. `amazon_reviews`

g. Pour les filtres de données, facultatif, sélectionnez`amazon_reviews_US`.

h. Pour les autorisations de filtrage de données, sélectionnez Sélectionner.

i. Choisissez Grant (Accorder).

4. Procédez comme suit pour accorder des autorisations à l'analyste de données japonais (`DataAnalystJP`) :

   a. Sous Autorisations, sélectionnez Autorisations du lac de données.

   b. Sous Autorisation relative aux données, choisissez Grant.

   c. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis sélectionnez le rôle. `DataAnalystJP`

   d. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.

   e. Pour Database (Base de données), sélectionnez `lakeformation_tutorial_row_security`.

   f. Pour Tableaux (facultatif), sélectionnez. `amazon_reviews`

   g. Pour les filtres de données, facultatif, sélectionnez`amazon_reviews_JP`.

   h. Pour les autorisations de filtrage de données, sélectionnez Sélectionner.

   i. Choisissez Grant (Accorder).

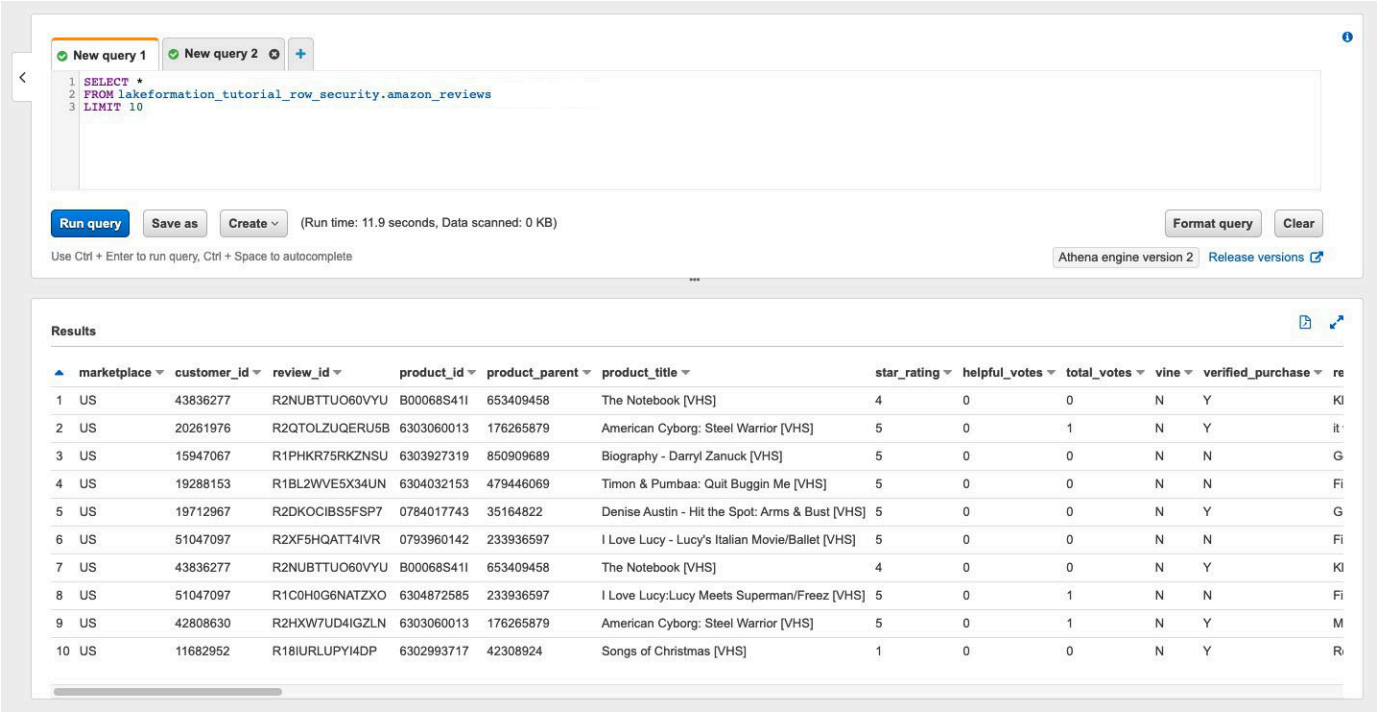
Étape 4 : Requête à l'aide de filtres de données

Avec les filtres de données attachés au tableau des avis sur les produits, exécutez quelques requêtes et découvrez comment les autorisations sont appliquées par Lake Formation.

1. Connectez-vous à la console Athena à l'adresse https://console.aws.amazon.com/athena/ en tant qu'utilisateur. `DataAnalystUS`
2. Exécutez la requête suivante pour récupérer quelques enregistrements, qui sont filtrés en fonction des autorisations au niveau des lignes que nous avons définies :

   ```
   SELECT *
   FROM lakeformation_tutorial_row_security.amazon_reviews
   LIMIT 10
   ```

   La capture d'écran suivante montre le résultat de la requête.



	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine	verified_purchase	re
1	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	KI
2	US	20261976	R2QTOLZUQERU5B	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	it
3	US	15947067	R1PHKR75RKZNSU	6303927319	850909689	Biography - Darryl Zanuck [VHS]	5	0	0	N	N	G
4	US	19288153	R1BL2WVE5X34UN	6304032153	479446069	Timon & Pumbaa: Quit Buggin Me [VHS]	5	0	0	N	N	Fi
5	US	19712967	R2DKOCIBS5FSP7	0784017743	35164822	Denise Austin - Hit the Spot: Arms & Bust [VHS]	5	0	0	N	Y	G
6	US	51047097	R2XF5HQATT4IVR	0793960142	233936597	I Love Lucy - Lucy's Italian Movie/Ballet [VHS]	5	0	0	N	N	Fi
7	US	43836277	R2NUBTTUO60VYU	B00068S41I	653409458	The Notebook [VHS]	4	0	0	N	Y	KI
8	US	51047097	R1C0H0G6NATZXO	6304872585	233936597	I Love Lucy:Lucy Meets Superman/Freez [VHS]	5	0	1	N	N	Fi
9	US	42808630	R2HXW7UD4IGZLN	6303060013	176265879	American Cyborg: Steel Warrior [VHS]	5	0	1	N	Y	M
10	US	11682952	R18IURLUPYI4DP	6302993717	42308924	Songs of Christmas [VHS]	1	0	0	N	Y	R

3. De même, exécutez une requête pour compter le nombre total d'enregistrements par site de vente.

```
SELECT marketplace , count ( * ) as total_count
FROM lakeformation_tutorial_row_security .amazon_reviews
GROUP BY marketplace
```

Le résultat de la requête n'affiche que le `marketplace` US dans les résultats. Cela est dû au fait que l'utilisateur n'est autorisé à voir que les lignes dont la valeur de `marketplace` colonne est égale à`US`.

4. Passez à l'`DataAnalystJP`utilisateur et exécutez la même requête.

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

Le résultat de la requête indique que seuls les enregistrements appartiennent au `JPmarketplace`.

5. Exécutez la requête pour compter le nombre total d'enregistrements par`marketplace`.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
```

```
GROUP BY marketplace
```

Le résultat de la requête montre uniquement la ligne appartenant au `JPmarketplace`.

Étape 5 : Nettoyer les AWS ressources

Nettoyage des ressources

Pour éviter que des frais supplémentaires ne vous soient facturés Compte AWS, vous pouvez supprimer les AWS ressources que vous avez utilisées pour ce didacticiel.

- Supprimez la pile de formation des nuages.

Partage d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation et de ressources nommées

Ce didacticiel explique comment vous pouvez configurer AWS Lake Formation pour partager en toute sécurité les données stockées dans un lac de données avec plusieurs entreprises, organisations ou unités commerciales, sans avoir à copier l'intégralité de la base de données. Il existe deux options pour partager vos bases de données et vos tables avec une autre personne Compte AWS en utilisant le contrôle d'accès entre comptes de Lake Formation :

- Contrôle d'accès basé sur des balises Lake Formation (recommandé)

  Le contrôle d'accès basé sur les balises de Lake Formation est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Dans Lake Formation, ces attributs sont appelés balises LF. Pour plus d'informations, consultez Gestion d'un lac de données à l'aide du contrôle d'accès basé sur des balises Lake Formation.
- Ressources nommées Lake Formation

  La méthode des ressources nommées Lake Formation est une stratégie d'autorisation qui définit les autorisations pour les ressources. Les ressources incluent des bases de données, des tables et des colonnes. Les administrateurs des lacs de données peuvent attribuer et révoquer des autorisations sur les ressources du Lake Formation. Pour plus d'informations, consultez Partage de données entre comptes dans Lake Formation.

Nous recommandons d'utiliser des ressources nommées si l'administrateur du lac de données préfère accorder des autorisations explicites à des ressources individuelles. Lorsque vous utilisez la méthode des ressources nommées pour accorder à Lake Formation des autorisations sur une ressource du catalogue de données à un compte externe, Lake Formation utilise AWS Resource Access Manager (AWS RAM) pour partager la ressource.

Rubriques

- Public visé
- Configurer les paramètres du catalogue de données de Lake Formation dans le compte du producteur
- Étape 1 : provisionnez vos ressources à l'aide AWS CloudFormation de modèles
- Étape 2 : Conditions préalables au partage entre comptes de Lake Formation
- Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises
- Étape 4 : Implémentation de la méthode de ressource nommée
- Étape 5 : Nettoyer les AWS ressources

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Lorsqu'il s'agit de partager des tables du catalogue de données AWS Glue et d'administrer les autorisations dans Lake Formation, les responsables des données des comptes producteurs ont la propriété fonctionnelle des fonctions qu'ils prennent en charge et peuvent accorder l'accès à divers consommateurs, organisations externes et comptes. Le tableau suivant répertorie les rôles utilisés dans ce didacticiel :

Rôle	Description
DataLakeAdminProducer	L'utilisateur IAM de l'administrateur du lac de données dispose des droits d'accès suivants : • Accès complet en lecture, écriture et mise à jour à toutes les ressources du catalogue de données

Rôle	Description
	• Possibilité d'accorder des autorisations aux ressources • Peut créer des liens de ressources pour la table partagée • Peut attacher des balises LF aux ressource s, ce qui permet d'accéder aux principau x en fonction des politiques créées par les gestionnaires de données
DataLakeAdminConsumer	L'utilisateur IAM de l'administrateur du lac de données dispose des droits d'accès suivants : • Accès complet en lecture, écriture et mise à jour à toutes les ressources du catalogue de données • Possibilité d'accorder des autorisations aux ressources • Peut créer des liens de ressources pour la table partagée • Peut attacher des balises LF aux ressource s, ce qui permet d'accéder aux principau x en fonction des politiques créées par les gestionnaires de données
DataAnalyst	L' DataAnalyst utilisateur dispose des droits d'accès suivants : • Accès détaillé aux ressources partagées par les politiques d'accès basées sur les balises de Lake Formation ou à l'aide de la méthode des ressources nommées

Configurer les paramètres du catalogue de données de Lake Formation dans le compte du producteur

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour plus d'informations, consultez Exécution des tâches AWS de configuration initiale.

Le didacticiel part du principe que vous connaissez IAM. Pour plus d'informations sur IAM, consultez le guide de l'utilisateur IAM.

Configurer les paramètres du catalogue de données de Lake Formation dans le compte du producteur

> **Note**
>
> Dans ce didacticiel, le compte contenant la table source est appelé compte producteur, et le compte qui a besoin d'accéder à la table source est appelé compte consommateur.

Lake Formation fournit son propre modèle de gestion des autorisations. Pour maintenir la rétrocompatibilité avec le modèle d'autorisation IAM, l'`Super`autorisation est accordée au groupe `IAMAllowedPrincipals` sur toutes les AWS Glue Data Catalog ressources existantes par défaut. En outre, les paramètres de contrôle d'accès Use only IAM sont activés pour les nouvelles ressources du catalogue de données. Ce didacticiel utilise un contrôle d'accès détaillé à l'aide des autorisations de Lake Formation et utilise des politiques IAM pour un contrôle d'accès détaillé. Consultez Méthodes de contrôle d'accès précis pour plus de détails. Par conséquent, avant d'utiliser un AWS CloudFormation modèle pour une configuration rapide, vous devez modifier les paramètres du Lake Formation Data Catalog dans le compte du producteur.

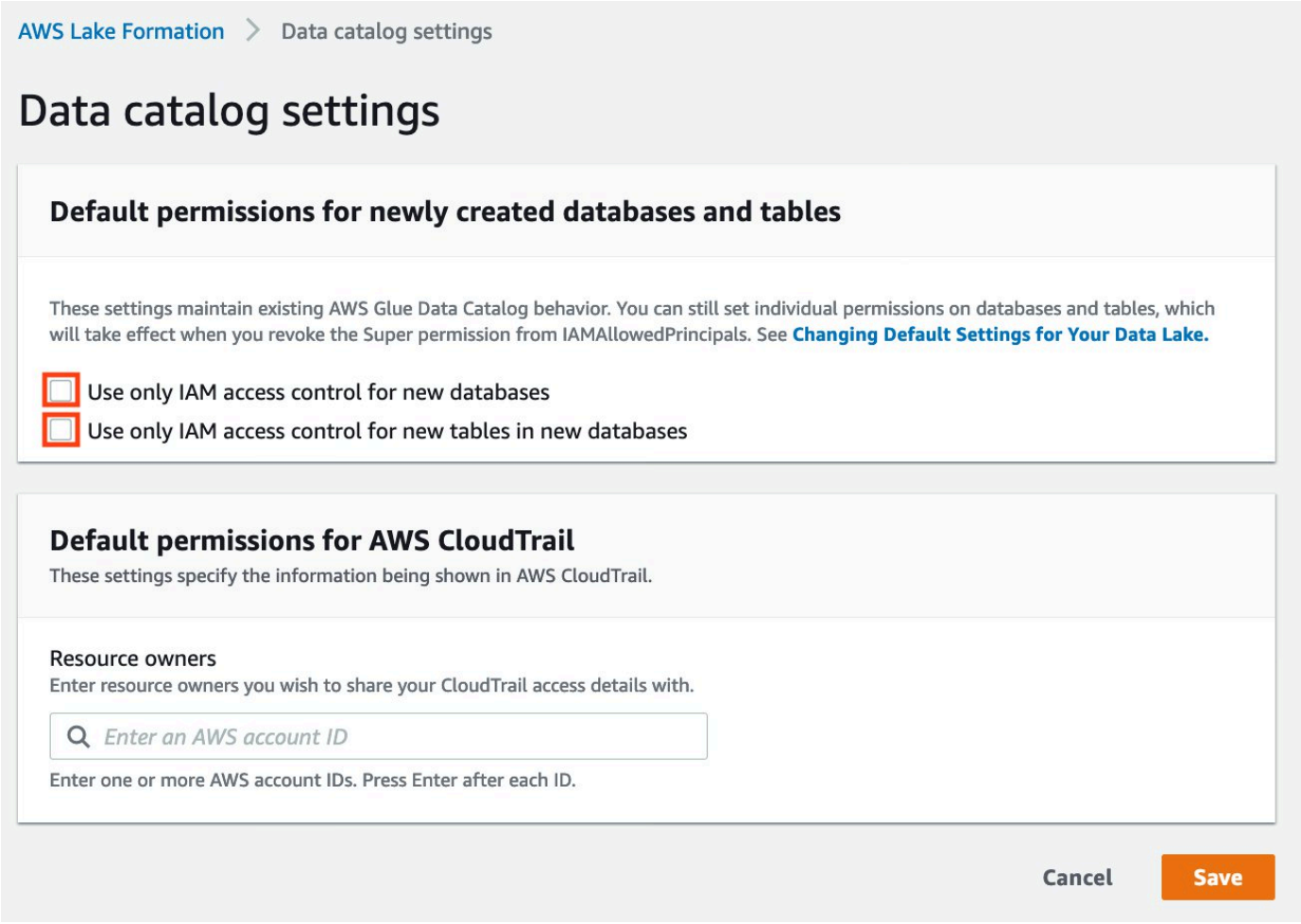
> **Important**
>
> Ce paramètre concerne toutes les bases de données et tables nouvellement créées. Nous vous recommandons donc vivement de suivre ce didacticiel dans un compte hors production ou dans un nouveau compte. De plus, si vous utilisez un compte partagé (tel que le compte de développement de votre entreprise), assurez-vous que cela n'affecte pas les ressources des autres. Si vous préférez conserver les paramètres de sécurité par défaut, vous devez effectuer une étape supplémentaire lorsque vous partagez des ressources avec d'autres comptes, au cours de laquelle vous révoquez l'autorisation Super par défaut

`IAMAllowedPrincipals` sur la base de données ou la table. Nous aborderons les détails plus loin dans ce didacticiel.

Pour configurer les paramètres du Lake Formation Data Catalog dans le compte du producteur, procédez comme suit :

1. Connectez-vous à l' AWS Management Console aide du compte producteur en tant qu'utilisateur administrateur ou en tant qu'utilisateur autorisé par l'`PutDataLakeSettings`API Lake Formation.
2. Sur la console Lake Formation, dans le volet de navigation, sous Data Catalog, sélectionnez Settings.
3. Désélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données et Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables dans les nouvelles bases de données

   Choisissez Enregistrer.



En outre, vous pouvez supprimer `CREATE_DATABASE` des autorisations pour les créateurs de base de données `IAMAllowedPrincipals` sous Rôles et tâches administratifs. Ce n'est qu'alors que vous pourrez déterminer qui peut créer une nouvelle base de données grâce aux autorisations de Lake Formation.

Étape 1 : provisionnez vos ressources à l'aide AWS CloudFormation de modèles

Le CloudFormation modèle de compte producteur génère les ressources suivantes :

- Un compartiment Amazon S3 qui servira de lac de données.
- Une fonction Lambda (pour les ressources personnalisées soutenues par AWS CloudFormation Lambda). Nous utilisons cette fonction pour copier des exemples de fichiers de données du compartiment public Amazon S3 vers votre compartiment Amazon S3.

- Utilisateurs et politiques IAM : DataLakeAdminProducer.
- Les paramètres et autorisations appropriés de Lake Formation, notamment :
  - Définition de l'administrateur du lac de données de Lake Formation dans le compte du producteur
  - Enregistrement d'un bucket Amazon S3 comme emplacement du lac de données de Lake Formation (compte du producteur)
- Une AWS Glue Data Catalog base de données, une table et une partition. Comme il existe deux options pour partager des ressources Comptes AWS, ce modèle crée deux ensembles distincts de base de données et de tables.

Le AWS CloudFormation modèle du compte client génère les ressources suivantes :

- Utilisateurs et politiques IAM :
  - DataLakeAdminConsumer
  - DataAnalyst
- Une AWS Glue Data Catalog base de données. Cette base de données permet de créer des liens vers des ressources partagées.

Créez vos ressources dans le compte du producteur

1. Connectez-vous à la AWS CloudFormation console à l'adresse https://console.aws.amazon.com/cloudformation dans la région USA Est (Virginie du Nord).
2. Choisissez Launch Stack.
3. Choisissez Suivant.
4. Pour Nom de pile, entrez un nom de pile, tel que `estack-producer`.
5. Dans la section Configuration utilisateur, entrez le nom d'utilisateur et le mot de passe pour `ProducerDatalakeAdminUserName` et `ProducerDatalakeAdminUserPassword`.
6. Pour DataLakeBucketName, entrez le nom de votre bucket de data lake. Ce nom doit être unique au monde.
7. Pour DatabaseNameet TableName, laissez les valeurs par défaut.
8. Choisissez Suivant.
9. Sur la page suivante, choisissez Next.

10. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
11. Choisissez Créer.

    La création de la pile peut prendre jusqu'à une minute.

Créez vos ressources dans le compte client

1. Connectez-vous à la AWS CloudFormation console à l'adresse https://console.aws.amazon.com/cloudformation dans la région USA Est (Virginie du Nord).
2. Choisissez Launch Stack.
3. Choisissez Suivant.
4. Pour Nom de pile, entrez un nom de pile, tel que `stack-consumer`.
5. Dans la section Configuration utilisateur, entrez le nom d'utilisateur et le mot de passe pour `ConsumerDatalakeAdminUserName` et `ConsumerDatalakeAdminUserPassword`.
6. Pour `DataAnalystUserName` et `DataAnalystUserPassword`, entrez le nom d'utilisateur et le mot de passe que vous souhaitez pour l'utilisateur IAM de l'analyste de données.
7. Pour DataLakeBucketName, entrez le nom de votre bucket de data lake. Ce nom doit être unique au monde.
8. Pour DatabaseName, conservez les valeurs par défaut.
9. Pour `AthenaQueryResultS3BucketName`, entrez le nom du compartiment Amazon S3 qui stocke les résultats des requêtes Amazon Athena. Si vous n'en avez pas, créez un compartiment Amazon S3.
10. Choisissez Suivant.
11. Sur la page suivante, choisissez Next.
12. Consultez les informations sur la dernière page et sélectionnez Je reconnais que cela AWS CloudFormation pourrait créer des ressources IAM.
13. Choisissez Créer.

    La création de la pile peut prendre jusqu'à une minute.

> **Note**
>
> Une fois le didacticiel terminé, supprimez le stack in AWS CloudFormation pour éviter d'encourir des frais. Vérifiez que les ressources sont correctement supprimées dans le statut de l'événement pour la pile.

Étape 2 : Conditions préalables au partage entre comptes de Lake Formation

Avant de partager des ressources avec Lake Formation, il existe des conditions préalables à la fois pour la méthode de contrôle d'accès basée sur les balises et pour la méthode des ressources nommées.

Conditions préalables complètes pour le contrôle d'accès basé sur des balises et le partage de données entre comptes

- Pour plus d'informations sur les exigences en matière de partage de données entre comptes, consultez la Prérequis section du chapitre Partage de données entre comptes.

  Pour partager les ressources du catalogue de données avec la version 3 ou supérieure des paramètres de version multi-comptes, le concédant doit disposer des autorisations IAM définies dans la politique AWS `AWSLakeFormationCrossAccountManager` gérée de votre compte.

  Si vous utilisez la version 1 ou la version 2 des paramètres de version multi-comptes, avant de pouvoir utiliser la méthode de contrôle d'accès basée sur les balises pour accorder l'accès aux ressources entre comptes, vous devez ajouter l'objet d'`JSON`autorisations suivant à la politique de ressources du catalogue de données du compte producteur. Cela donne au compte client l'autorisation d'accéder au catalogue de données lorsque cela `glue:EvaluatedByLakeFormationTags` est vrai. Cette condition s'applique également aux ressources pour lesquelles vous avez accordé une autorisation en utilisant les balises d'autorisation de Lake Formation sur le compte du consommateur. Cette politique est obligatoire pour tous ceux Compte AWS auxquels vous accordez des autorisations.

  La politique suivante doit être intégrée à un `Statement` élément. Nous discutons de la politique IAM complète dans la section suivante.

  ```
  {
  ```

```
    "Effect": "Allow",
    "Action": [
        "glue:*"
    ],
    "Principal": {
        "AWS": [
            "consumer-account-id"
        ]
    },
    "Resource": [
        "arn:aws:glue:region:account-id:table/*",
        "arn:aws:glue:region:account-id:database/*",
        "arn:aws:glue:region:account-id:catalog"
    ],
    "Condition": {
        "Bool": {
            "glue:EvaluatedByLakeFormationTags": true
        }
    }
}
```

Compléter les prérequis relatifs au partage entre comptes de la méthode des ressources nommées

1. S'il n'existe aucune politique de ressources du catalogue de données dans votre compte, le compte croisé Lake Formation vous permet de poursuivre comme d'habitude. Toutefois, s'il existe une politique de ressources pour le catalogue de données, vous devez y ajouter l'instruction suivante pour permettre à vos subventions entre comptes de réussir si elles sont accordées avec la méthode de ressource nommée. Si vous prévoyez d'utiliser uniquement la méthode de ressource nommée ou uniquement la méthode de contrôle d'accès basée sur des balises, vous pouvez ignorer cette étape. Dans ce didacticiel, nous évaluons les deux méthodes et nous devons ajouter la politique suivante.

   La politique suivante doit être intégrée à un `Statement` élément. Nous discutons de la politique IAM complète dans la section suivante.

   ```
   {
             "Effect": "Allow",
             "Action": [
             "glue:ShareResource"
             ],
   ```

```
            "Principal": {
              "Service":"ram.amazonaws.com"
            },
            "Resource": [
                "arn:aws:glue:region:account-id:table/*/*",
                "arn:aws:glue:region:account-id:database/*",
                "arn:aws:glue:region:account-id:catalog"
            ]
}
```

2. Ajoutez ensuite la politique de AWS Glue Data Catalog ressources à l'aide du AWS Command Line Interface (AWS CLI).

   Si vous accordez des autorisations entre comptes en utilisant à la fois la méthode de contrôle d'accès basée sur les balises et la méthode des ressources nommées, vous devez définir l'`EnableHybrid`argument sur « true » lors de l'ajout des politiques précédentes. Parce que cette option n'est actuellement pas prise en charge sur la console, vous devez utiliser l'`glue:PutResourcePolicy`API et AWS CLI.

   Créez d'abord un document de politique (tel que policy.json) et ajoutez les deux politiques précédentes. Remplacez *`consumer-account-id`*par l'*`ID de compte`* du Compte AWS bénéficiaire de la subvention, la *`région`* par la région du catalogue de données contenant les bases de données et les tables pour lesquelles vous accordez des autorisations, et l'identifiant de *`compte par l'identifiant`* du producteur Compte AWS .

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "ram.amazonaws.com"
            },
            "Action": "glue:ShareResource",
            "Resource": [
                "arn:aws:glue:region:account-id:table/*/*",
                "arn:aws:glue:region:account-id:database/*",
                "arn:aws:glue:region:account-id:catalog"
            ]
        },
        {
```

```
            "Effect": "Allow",
            "Principal": {
                "AWS": "region:account-id"
            },
            "Action": "glue:*",
            "Resource": [
                "arn:aws:glue:region:account-id:table/*/*",
                "arn:aws:glue:region:account-id:database/*",
                "arn:aws:glue:region:account-id:catalog"
            ],
            "Condition": {
                "Bool": {
                    "glue:EvaluatedByLakeFormationTags": "true"
                }
            }
        }
    ]
}
```

Entrez la AWS CLI commande suivante. Remplacez *glue-resource-policy*par les valeurs correctes (telles que file : //policy.json).

```
aws glue put-resource-policy --policy-in-json glue-resource-policy --enable-hybrid
 TRUE
```

Pour plus d'informations, consultez put-resource-policy.

Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises

Dans cette section, nous vous expliquons les étapes de haut niveau suivantes :

1. Définissez une balise LF.
2. Attribuez le tag LF à la ressource cible.
3. Accordez des autorisations LF-Tag au compte client.
4. Accordez des autorisations de données au compte client.
5. Vous pouvez éventuellement révoquer les autorisations pour la `IAMAllowedPrincipals` base de données, les tables et les colonnes.

6. Créez un lien de ressource vers la table partagée.
7. Créez une balise LF et attribuez-la à la base de données cible.
8. Accordez des autorisations relatives aux données LF-Tag au compte client.

Définir un LF-Tag

> **Note**
>
> Si vous êtes connecté à votre compte producteur, déconnectez-vous avant de suivre les étapes suivantes.

1. Connectez-vous au compte du producteur en tant qu'administrateur du lac de données à l'adresse https://console.aws.amazon.com/lakeformation/. Utilisez le numéro de compte du producteur, le nom d'utilisateur IAM (par défaut`DatalakeAdminProducer`) et le mot de passe que vous avez spécifiés lors de la création de la AWS CloudFormation pile.
2. Sur la console Lake Formation (https://console.aws.amazon.com/lakeformation/), dans le volet de navigation, sous Autorisations et sous Rôles et tâches administratifs, choisissez LF-Tags.
3. Choisissez Ajouter un tag LF.

Attribuez le tag LF à la ressource cible

Attribuez le tag LF à la ressource cible et accordez des autorisations de données à un autre compte

En tant qu'administrateur de data lake, vous pouvez associer des balises aux ressources. Si vous envisagez d'utiliser un rôle distinct, vous devrez peut-être accorder des autorisations de description et d'attachement à ce rôle distinct.

1. Dans le volet de navigation, sous Catalogue de données, sélectionnez Bases de données.
2. Sélectionnez la base de données cible (`lakeformation_tutorial_cross_account_database_tbac`) et dans le menu Actions, choisissez Modifier les balises LF.

   Dans ce didacticiel, vous allez attribuer une balise LF à une base de données, mais vous pouvez également attribuer des balises LF à des tables et à des colonnes.
3. Choisissez Attribuer un nouveau tag LF.
4. Ajoutez la clé `Confidentiality` et la valeur`public`.

5. Choisissez Enregistrer.

Accorder l'autorisation LF-Tag au compte client

Toujours dans le compte du producteur, autorisez le compte du consommateur à accéder au LF-Tag.

1. Dans le volet de navigation, sous Autorisations, Rôles et tâches administratifs, Autorisations LF-Tag, choisissez Grant.
2. Pour Principaux, sélectionnez Comptes externes.
3. Entrez l'Compte AWS ID cible.

   Comptes AWS au sein de la même organisation apparaissent automatiquement. Dans le cas contraire, vous devez saisir l' Compte AWS identifiant manuellement. Au moment d'écrire ces lignes, le contrôle d'accès basé sur les balises de Lake Formation ne permet pas d'accorder des autorisations aux organisations ou aux unités organisationnelles.
4. Pour les balises LF, choisissez la clé et les valeurs de la balise LF partagée avec le compte du consommateur (clé **`Confidentiality`** et valeur). `public`
5. Pour Autorisations, sélectionnez Décrire pour les autorisations LF-Tag.

   Les autorisations LF-Tag sont des autorisations accordées au compte du consommateur. Les autorisations pouvant être accordées sont des autorisations que le compte client peut accorder à d'autres mandants.
6. Choisissez Grant (Accorder).

   À ce stade, l'administrateur du lac de données du consommateur devrait être en mesure de trouver le tag de politique partagé via la console Lake Formation du compte consommateur, sous Autorisations, rôles et tâches administratifs, balises LF.

Accorder l'autorisation d'accès aux données du compte client

Nous allons maintenant fournir un accès aux données au compte client en spécifiant une expression LF-Tag et en accordant au compte client l'accès à toute table ou base de données correspondant à l'expression.

1. Dans le volet de navigation, sous Autorisations, Autorisations du lac de données, sélectionnez Grant.
2. Pour Principaux, choisissez Comptes externes et entrez l' Compte AWS ID cible.

3. Pour les balises LF ou les ressources du catalogue, choisissez la clé et les valeurs de la balise LF partagée avec le compte du consommateur (clé **`Confidentiality`** et valeur). `public`
4. Pour les autorisations, sous Ressources associées à des balises LF (recommandé), choisissez Ajouter une balise LF.
5. Sélectionnez la clé et la valeur de la balise partagée avec le compte client (clé `Confidentiality` et valeur`public`).
6. Pour les autorisations de base de données, sélectionnez Décrire sous Autorisations de base de données pour accorder des autorisations d'accès au niveau de la base de données.
7. L'administrateur du lac de données du consommateur doit être en mesure de trouver le tag de politique partagé via le compte du consommateur sur la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/, sous Autorisations, rôles et tâches administratifs, balises LF.
8. Sélectionnez Décrire sous Autorisations pouvant être accordées afin que le compte client puisse accorder des autorisations au niveau de la base de données à ses utilisateurs.
9. Pour les autorisations de table et de colonne, sélectionnez Sélectionner et décrire sous Autorisations de table.
10. Sélectionnez Sélectionner et décrire sous Autorisations pouvant être accordées.
11. Choisissez Grant (Accorder).

Révoquez l'autorisation pour la **`IAMAllowedPrincipals`** base de données, les tables et les colonnes (facultatif).

Au tout début de ce didacticiel, vous avez modifié les paramètres du Lake Formation Data Catalog. Si vous avez ignoré cette partie, cette étape est obligatoire. Si vous avez modifié les paramètres de votre catalogue de données sur les formations des lacs, vous pouvez ignorer cette étape.

Au cours de cette étape, nous devons révoquer l'autorisation Super par défaut `IAMAllowedPrincipals` sur la base de données ou la table. Consultez Étape 4 : Basculez vos magasins de données vers le modèle d'autorisations de Lake Formation pour plus de détails.

Avant de révoquer l'autorisation pour`IAMAllowedPrincipals`, assurez-vous d'avoir accordé aux principaux IAM existants les autorisations nécessaires par le biais de Lake Formation. Cela comprend trois étapes :

1. Ajoutez l'autorisation IAM à l'utilisateur ou au rôle IAM cible avec l'`GetDataAccess`action Lake Formation (avec la politique IAM).

2. Accordez à l'utilisateur ou au rôle IAM cible doté des autorisations relatives aux données de Lake Formation (modifier, sélectionner, etc.).
3. Révoquez ensuite les autorisations pour`IAMAllowedPrincipals`. Sinon, après avoir révoqué les autorisations pour`IAMAllowedPrincipals`, les principaux IAM existants risquent de ne plus être en mesure d'accéder à la base de données ou au catalogue de données cible.

   La révocation de l'autorisation Super `IAMAllowedPrincipals` est requise lorsque vous souhaitez appliquer le modèle d'autorisation Lake Formation (au lieu du modèle de politique IAM) pour gérer l'accès des utilisateurs au sein d'un seul compte ou entre plusieurs comptes à l'aide du modèle d'autorisation Lake Formation. Vous n'êtes pas obligé de révoquer l'autorisation `IAMAllowedPrincipals` pour les autres tables pour lesquelles vous souhaitez conserver le modèle de politique IAM traditionnel.

   À ce stade, l'administrateur du lac de données du compte client doit être en mesure de trouver la base de données et la table partagées via le compte client sur la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/, sous Catalogue de données, bases de données. Dans le cas contraire, vérifiez si les éléments suivants sont correctement configurés :

   1. La balise de politique et les valeurs correctes sont attribuées aux bases de données et aux tables cibles.
   2. L'autorisation de balise et l'autorisation de données correctes sont attribuées au compte client.
   3. Révoquez la super autorisation par défaut `IAMAllowedPrincipals` sur la base de données ou la table.

Création d'un lien de ressource vers la table partagée

Lorsqu'une ressource est partagée entre des comptes et que les ressources partagées ne sont pas placées dans le catalogue de données des comptes consommateurs. Pour les rendre disponibles et interroger les données sous-jacentes d'une table partagée à l'aide de services tels qu'Athena, nous devons créer un lien de ressource vers la table partagée. Un lien de ressource est un objet du catalogue de données qui est un lien vers une base de données ou une table locale ou partagée. Pour plus de détails, consultez Création de liens vers des ressources. En créant un lien vers une ressource, vous pouvez :

- Attribuez un nom différent à une base de données ou à une table conforme aux politiques de dénomination des ressources de votre catalogue de données.

- Utilisez des services tels qu'Athena et Redshift Spectrum pour interroger des bases de données ou des tables partagées.

Pour créer un lien vers une ressource, procédez comme suit :

1. Si vous êtes connecté à votre compte client, déconnectez-vous.
2. Connectez-vous en tant qu'administrateur du lac de données du compte client. Utilisez l'identifiant du compte client, le nom d'utilisateur IAM (par défaut DatalakeAdminConsumer) et le mot de passe que vous avez spécifiés lors de la création de la AWS CloudFormation pile.
3. Sur la console Lake Formation (https://console.aws.amazon.com/lakeformation/), dans le volet de navigation, sous Data Catalog, Databases, sélectionnez la base de données partagée`lakeformation_tutorial_cross_account_database_tbac`.

   Si la base de données n'apparaît pas, revoyez les étapes précédentes pour vérifier si tout est correctement configuré.
4. Choisissez Afficher les tables.
5. Choisissez la table partagée`amazon_reviews_table_tbac`.
6. Dans le menu Actions, choisissez Créer un lien vers une ressource.
7. Pour le nom du lien de ressource, entrez un nom (pour ce didacticiel,`amazon_reviews_table_tbac_resource_link`).
8. Sous Base de données, sélectionnez la base de données dans laquelle le lien de ressource est créé (pour cet article, la pile AWS CloudFormation n a créé la base de données`lakeformation_tutorial_cross_account_database_consumer`).
9. Choisissez Créer.

   Le lien vers la ressource apparaît sous Catalogue de données, Tables.

Créez une balise LF et attribuez-la à la base de données cible

Les balises Lake Formation se trouvent dans le même catalogue de données que les ressources. Cela signifie que les balises créées dans le compte du producteur ne peuvent pas être utilisées lors de l'octroi de l'accès aux liens vers les ressources du compte client. Vous devez créer un ensemble distinct de balises LF dans le compte client pour utiliser le contrôle d'accès basé sur les balises LF lors du partage des liens de ressources dans le compte client.

1. Définissez le tag LF dans le compte client. Pour ce didacticiel, nous utilisons des clés `Division` et des valeurs `salesmarketing`, et`analyst`.
2. Attribuez la clé `Division` et la valeur du tag LF `analyst` à la base de données`lakeformation_tutorial_cross_account_database_consumer`, dans laquelle le lien de ressource est créé.

Autoriser le consommateur à utiliser les données LF-Tag

Enfin, accordez au consommateur l'autorisation d'utiliser les données LF-Tag.

1. Dans le volet de navigation, sous Autorisations, Autorisations du lac de données, sélectionnez Grant.
2. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis choisissez l'utilisateur. `DataAnalyst`
3. Pour les balises LF ou les ressources du catalogue, choisissez Ressources associées aux balises LF (recommandé).
4. Choisissez une division clé et un analyste de valeur.
5. Pour les autorisations de base de données, sélectionnez Décrire sous Autorisations de base de données.
6. Pour les autorisations de table et de colonne, sélectionnez Sélectionner et décrire sous Autorisations de table.
7. Choisissez Grant (Accorder).
8. Répétez ces étapes pour l'utilisateur`DataAnalyst`, où se trouvent la clé LF-Tag `Confidentiality` et la valeur. `public`

   À ce stade, l'utilisateur de l'analyste de données du compte client doit être en mesure de trouver la base de données et le lien vers la ressource, et d'interroger la table partagée via la console Athena à l'adresse https://console.aws.amazon.com/athena/. Dans le cas contraire, vérifiez si les éléments suivants sont correctement configurés :

   - Le lien de ressource est créé pour la table partagée
   - Vous avez accordé à l'utilisateur l'accès au LF-Tag partagé par le compte du producteur
   - Vous avez accordé à l'utilisateur l'accès à la balise LF associée au lien de ressource et à la base de données dans lesquels le lien de ressource est créé

- Vérifiez si vous avez attribué la balise LF correcte au lien de ressource et à la base de données dans laquelle le lien de ressource est créé

Étape 4 : Implémentation de la méthode de ressource nommée

Pour utiliser la méthode de ressource nommée, nous vous expliquons les étapes de haut niveau suivantes :

1. Vous pouvez éventuellement révoquer l'autorisation pour la `IAMAllowedPrincipals` base de données, les tables et les colonnes.
2. Accordez l'autorisation d'accès aux données du compte client.
3. Acceptez un partage de ressources depuis AWS Resource Access Manager.
4. Créez un lien de ressource pour la table partagée.
5. Accordez au consommateur l'autorisation d'accéder aux données de la table partagée.
6. Accordez au consommateur l'autorisation d'accès aux données pour le lien vers la ressource.

Révoquer l'autorisation pour la **`IAMAllowedPrincipals`** base de données, les tables et les colonnes (facultatif)

- Au tout début de ce didacticiel, nous avons modifié les paramètres du Lake Formation Data Catalog. Si vous avez ignoré cette partie, cette étape est obligatoire. Pour obtenir des instructions, reportez-vous à l'étape facultative de la section précédente.

Accorder l'autorisation d'accès aux données du compte client

1. > **Note**
   >
   > Si vous êtes connecté au compte producteur en tant qu'autre utilisateur, déconnectez-vous d'abord.

   Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en utilisant l'administrateur du lac de données du compte producteur en utilisant l' Compte AWS ID, le nom d'utilisateur IAM (par défaut`DatalakeAdminProducer`) et le mot de passe spécifiés lors de la création de la AWS CloudFormation pile.

2. Sur la page Autorisations, sous Permissions du lac de données, sélectionnez Accorder.
3. Sous Principaux, choisissez Comptes externes, puis entrez un ou plusieurs Compte AWS identifiants ou identifiants d' AWS organisation. Pour plus d'informations, voir : AWS Organizations.

   Organisations auxquelles appartient le compte producteur et au Comptes AWS sein de la même organisation apparaissent automatiquement. Dans le cas contraire, entrez manuellement l'identifiant du compte ou l'identifiant de l'organisation.
4. Pour les balises LF ou les ressources du catalogue, choisissez. `Named data catalog resources`
5. Sous Bases de données, sélectionnez la base de données`lakeformation_tutorial_cross_account_database_named_resource`.
6. Choisissez Ajouter un tag LF.
7. Sous Tables, sélectionnez Toutes les tables.
8. Pour les autorisations relatives aux colonnes du tableau, choisissez Sélectionner et Décrire sous Autorisations du tableau.
9. Sélectionnez Sélectionner et décrire, sous Autorisations pouvant être accordées.
10. Facultativement, pour les autorisations relatives aux données, choisissez Accès simple basé sur les colonnes si la gestion des autorisations au niveau des colonnes est requise.
11. Choisissez Grant (Accorder).

Si vous n'avez pas révoqué l'autorisation pour`IAMAllowedPrincipals`, le message d'erreur « Échec de l'octroi des autorisations » s'affiche. À ce stade, vous devriez voir la table cible partagée AWS RAM avec le compte client sous Autorisations, Autorisations relatives aux données.

Accepter un partage de ressources depuis AWS RAM

> **Note**
>
> Cette étape est requise uniquement pour le partage Compte AWS basé, et non pour le partage basé sur l'organisation.

1. Connectez-vous à la AWS console à l'adresse https://console.aws.amazon.com/connect/ en utilisant l'administrateur du lac de données du compte consommateur en utilisant le nom

d'utilisateur IAM (par défaut DatalakeAdminConsumer) et le mot de passe spécifiés lors de la création de la AWS CloudFormation pile.

2. Sur la AWS RAM console, dans le volet de navigation, sous Shared with me, Resource shares, choisissez la ressource Lake Formation partagée. Le statut doit être En attente.
3. Choisissez Action et Grant.
4. Confirmez les détails de la ressource, puis choisissez Accepter le partage des ressources.

   À ce stade, l'administrateur du lac de données du compte client doit être en mesure de trouver la ressource partagée sur la console Lake Formation (https://console.aws.amazon.com/lakeformation/) sous Data Catalog, Databases.

Création d'un lien de ressource pour la table partagée

- Suivez les instructions de l'étape Étape 3 : mise en œuvre du partage entre comptes à l'aide de la méthode de contrôle d'accès basée sur des balises 6 pour créer un lien de ressource pour une table partagée. Nommez le lien vers la ressource`amazon_reviews_table_named_resource_resource_link`. Créez le lien vers la ressource dans la base de données`lakeformation_tutorial_cross_account_database_consumer`.

Autoriser le consommateur à accéder aux données de la table partagée

Pour autoriser le consommateur à accéder aux données de la table partagée, procédez comme suit :

1. Sur la console Lake Formation (https://console.aws.amazon.com/lakeformation/), sous Autorisations, Autorisations Data Lake, choisissez Grant.
2. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis choisissez l'utilisateur. `DataAnalyst`
3. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.
4. Sous Bases de données, sélectionnez la base de données`lakeformation_tutorial_cross_account_database_named_resource`. Si la base de données ne figure pas dans la liste déroulante, choisissez Load more.
5. Sous Tables, sélectionnez la table`amazon_reviews_table_named_resource`.
6. Pour les autorisations de table et de colonne, sélectionnez Sélectionner et décrire sous Autorisations de table.

7. Choisissez Grant (Accorder).

Accorder au consommateur l'autorisation d'accès aux données pour le lien vers la ressource

En plus d'autoriser l'utilisateur du lac de données à accéder à la table partagée, vous devez également accorder à l'utilisateur du lac de données l'autorisation d'accéder au lien de ressource.

1. Sur la console Lake Formation (https://console.aws.amazon.com/lakeformation/), sous Autorisations, Autorisations du lac de données, choisissez Grant.
2. Pour Principaux, choisissez les utilisateurs et les rôles IAM, puis choisissez l'utilisateur. `DataAnalyst`
3. Pour les balises LF ou les ressources de catalogue, choisissez Ressources de catalogue de données nommées.
4. Sous Bases de données, sélectionnez la base de données`lakeformation_tutorial_cross_account_database_consumer`. Si la base de données ne figure pas dans la liste déroulante, choisissez Load more.
5. Sous Tables, sélectionnez la table`amazon_reviews_table_named_resource_resource_link`.
6. Pour les autorisations relatives aux liens vers les ressources, sélectionnez Décrire sous Autorisations relatives aux liens vers les ressources.
7. Choisissez Grant (Accorder).

   À ce stade, l'utilisateur de l'analyste de données du compte client doit être en mesure de trouver la base de données et le lien vers la ressource, et d'interroger la table partagée via la console Athena.

   Dans le cas contraire, vérifiez si les éléments suivants sont correctement configurés :

   - Le lien de ressource est créé pour la table partagée
   - Vous avez accordé à l'utilisateur l'accès à la table partagée par le compte du producteur
   - Vous avez accordé à l'utilisateur l'accès au lien de ressource et à la base de données pour lesquels le lien de ressource est créé

Étape 5 : Nettoyer les AWS ressources

Pour éviter que des frais supplémentaires ne vous soient facturés Compte AWS, vous pouvez supprimer les AWS ressources que vous avez utilisées pour ce didacticiel.

1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ à l'aide du compte producteur et supprimez ou modifiez les éléments suivants :
   - AWS Resource Access Manager partage des ressources
   - Tags Lake Formation
   - AWS CloudFormation pile
   - Paramètres de la Lake Formation
   - AWS Glue Data Catalog
2. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ à l'aide du compte client et supprimez ou modifiez les éléments suivants :
   - Tags Lake Formation
   - AWS CloudFormation pile

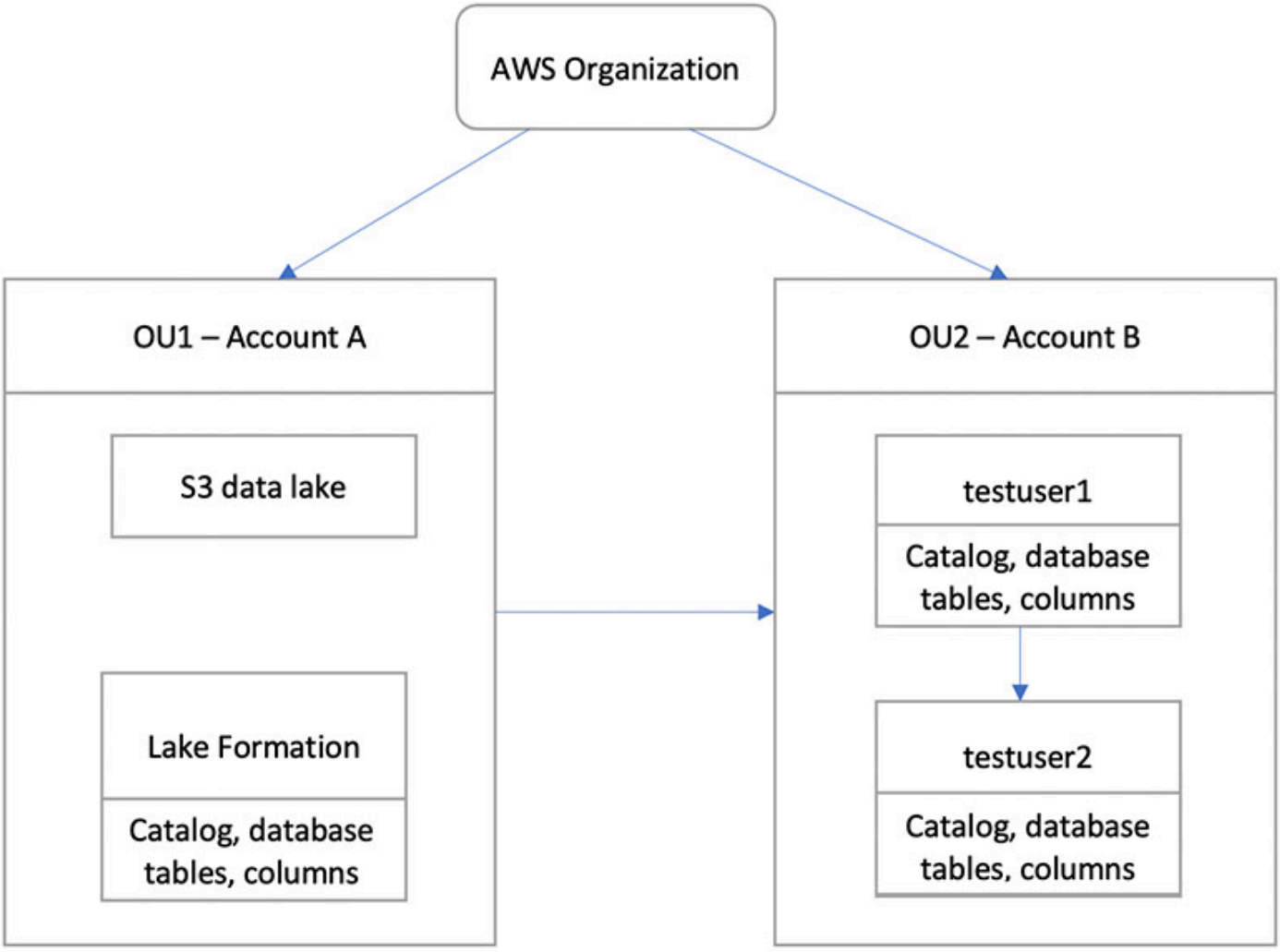
Partage d'un lac de données à l'aide du contrôle d'accès précis de Lake Formation

Ce didacticiel fournit des step-by-step instructions sur la manière de partager rapidement et facilement des ensembles de données à l'aide de Lake Formation lorsque vous en gérez plusieurs Comptes AWS avec AWS Organizations. Vous définissez des autorisations détaillées pour contrôler l'accès aux données sensibles.

Les procédures suivantes montrent également comment un administrateur du lac de données du compte A peut fournir un accès détaillé au compte B, et comment un utilisateur du compte B, agissant en tant que gestionnaire de données, peut accorder un accès détaillé à la table partagée aux autres utilisateurs de son compte. Au sein de chaque compte, les responsables des données peuvent déléguer l'accès de manière indépendante à leurs propres utilisateurs, ce qui donne de l'autonomie à chaque équipe ou secteur d'activité (LOB).

Le cas d'utilisation suppose que vous utilisez AWS Organizations pour gérer votre Comptes AWS. L'utilisateur du compte A dans une unité organisationnelle (OU1) accorde l'accès aux utilisateurs

du compte B dans l'unité organisationnelle 2. Vous pouvez utiliser la même approche lorsque vous n'utilisez pas Organizations, par exemple lorsque vous n'avez que quelques comptes. Le schéma suivant illustre le contrôle d'accès détaillé des ensembles de données dans un lac de données. Le lac de données est disponible dans le compte A. L'administrateur du lac de données du compte A fournit un accès détaillé au compte B. Le diagramme montre également qu'un utilisateur du compte B fournit un accès au niveau des colonnes de la table du lac de données du compte A à un autre utilisateur du compte B.



Rubriques

- Public visé
- Prérequis
- Étape 1 : Fournir un accès détaillé à un autre compte
- Étape 2 : fournir un accès détaillé à un utilisateur du même compte

Public visé

Ce didacticiel est destiné aux gestionnaires de données, aux ingénieurs de données et aux analystes de données. Le tableau suivant répertorie les rôles utilisés dans ce didacticiel :

Rôle	Description
Administrateur IAM	Utilisateur disposant de la politique AWS gérée :`AdministratorAccess` .
Administrateur du lac de données	Utilisateur disposant de la politique AWS gérée : `AWSLakeFormationDataAdmin` attaché au rôle.
Analyste des données	Utilisateur disposant de la politique AWS gérée : `AmazonAthenaFullAccess` joint.

Prérequis

Avant de commencer ce didacticiel, vous devez disposer d'un Compte AWS identifiant que vous pouvez utiliser pour vous connecter en tant qu'utilisateur administratif avec les autorisations appropriées. Pour plus d’informations, consultez Exécution des tâches AWS de configuration initiale.

Le didacticiel part du principe que vous connaissez IAM. Pour plus d'informations sur IAM, consultez le guide de l'utilisateur IAM.

Vous avez besoin des ressources suivantes pour ce didacticiel :

- Deux unités organisationnelles :
  - OU1 — Contient le compte A
  - OU2 — Contient le compte B
- Emplacement d'un lac de données Amazon S3 (compartiment) dans le compte A.
- Utilisateur administrateur de lac de données dans le compte A. Vous pouvez créer un administrateur de lac de données à l'aide de la console Lake Formation (https://console.aws.amazon.com/lakeformation/) ou de l'API Lake Formation. `PutDataLakeSettings`
- Lake Formation configurée dans le compte A, et l'emplacement du lac de données Amazon S3 enregistré auprès de Lake Formation dans le compte A.

- Deux utilisateurs du compte B avec les politiques gérées par IAM suivantes :
  - testuser1 — contient les politiques AWS `AWSLakeFormationDataAdmin` gérées.
  - testuser2 — La politique AWS `AmazonAthenaFullAccess` gérée est attachée.
- Une base de données testdb dans la base de données Lake Formation pour le compte B.

Étape 1 : Fournir un accès détaillé à un autre compte

Découvrez comment l'administrateur du lac de données du compte A fournit un accès détaillé au compte B.

Accorder un accès détaillé à un autre compte

1. Connectez-vous sur AWS Management Console https://console.aws.amazon.com/connect/ dans le compte A en tant qu'administrateur du lac de données.
2. Ouvrez la console Lake Formation (https://console.aws.amazon.com/lakeformation/) et choisissez Get started.
3. dans le volet de navigation, sélectionnez Databases.
4. Choisissez Create database (Créer une base de données).
5. Dans la section Détails de la base de données, sélectionnez Base de données.
6. Pour Nom, entrez un nom (pour ce didacticiel, nous utilisons`sampledb01`).
7. Assurez-vous que l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données n'est pas sélectionnée. Si cette option n'est pas sélectionnée, nous pouvons contrôler l'accès depuis Lake Formation.
8. Choisissez Créer une base de données.
9. Sur la page Bases de données, choisissez votre base de données`sampledb01`.
10. Dans le menu Actions, choisissez Grant.
11. Dans la section Accorder des autorisations, sélectionnez Compte externe.
12. Dans le Compte AWS champ ID ou ID d' AWS organisation, entrez l'ID de compte du compte B dans OU2.
13. Pour Table, choisissez la table à laquelle vous souhaitez que le compte B ait accès (pour cet article, nous utilisons la table`acc_a_area`). Vous pouvez éventuellement autoriser l'accès aux colonnes du tableau, ce que nous faisons dans cet article.
14. Pour Inclure les colonnes, choisissez les colonnes auxquelles vous souhaitez que le compte B ait accès (pour cet article, nous accordons des autorisations de type, de nom et d'identifiant).

15. Pour Colonnes, choisissez Inclure les colonnes.
16. Pour les autorisations relatives aux tables, sélectionnez Sélectionner.
17. Pour les autorisations pouvant être accordées, sélectionnez Sélectionner. Des autorisations pouvant être accordées sont requises pour que les utilisateurs administrateurs du compte B puissent accorder des autorisations à d'autres utilisateurs du compte B.
18. Choisissez Grant (Accorder).
19. Dans le volet de navigation, choisissez Tables.
20. Vous pouvez voir une connexion active dans la section « Comptes AWS et AWS organisations ayant accès ».

Création d'un lien vers une ressource

Les services intégrés tels qu'Amazon Athena ne peuvent pas accéder directement aux bases de données ou aux tables entre les comptes. Par conséquent, vous devez créer un lien de ressource afin qu'Athéna puisse accéder aux liens de ressources de votre compte vers les bases de données et les tables d'autres comptes. Créez un lien de ressource vers la table (`acc_a_area`) afin que les utilisateurs du compte B puissent interroger ses données auprès d'Athena.

1. Connectez-vous à la AWS console à l'adresse https://console.aws.amazon.com/connect/ dans le compte B en tant que`testuser1`.
2. Sur la console Lake Formation (https://console.aws.amazon.com/lakeformation/), dans le volet de navigation, sélectionnez Tables. Vous devriez voir les tables auxquelles le compte A a fourni l'accès.
3. Choisissez la table `acc_a_area`.
4. Dans le menu Actions, choisissez Créer un lien vers une ressource.
5. Pour le nom du lien de ressource, entrez un nom (pour ce didacticiel,`acc_a_area_rl`).
6. Pour Base de données, choisissez votre base de données (`testdb`).
7. Choisissez Créer.
8. Dans le volet de navigation, choisissez Tables.
9. Choisissez la table `acc_b_area_rl`.
10. Dans le menu Actions, choisissez Afficher les données.

    Vous êtes redirigé vers la console Athena, où vous devriez voir la base de données et la table.

Vous pouvez désormais exécuter une requête sur la table pour voir la valeur de la colonne pour laquelle l'accès a été accordé à testuser1 depuis le compte B.

Étape 2 : fournir un accès détaillé à un utilisateur du même compte

Cette section montre comment un utilisateur du compte B (`testuser1`), agissant en tant que gestionnaire de données, fournit à un autre utilisateur du même compte (`testuser2`) un accès détaillé au nom de colonne de la table partagée. `aac_b_area_rl`

Accorder un accès détaillé à un utilisateur du même compte

1. Connectez-vous à la AWS console à l'adresse https://console.aws.amazon.com/connect/ dans le compte B en tant que`testuser1`.
2. Sur la console Lake Formation, dans le volet de navigation, sélectionnez Tables.

   Vous pouvez accorder des autorisations sur une table via son lien de ressource. Pour ce faire, sur la page Tables, sélectionnez le lien vers la ressource`acc_b_area_rl`, puis dans le menu Actions, choisissez Grant on target.
3. Dans la section Accorder des autorisations, sélectionnez Mon compte.
4. Pour les utilisateurs et les rôles IAM, choisissez l'utilisateur`testuser2`.
5. Pour Colonne, choisissez le nom de la colonne.
6. Pour les autorisations relatives aux tables, sélectionnez Sélectionner.
7. Choisissez Grant (Accorder).

   Lorsque vous créez un lien vers une ressource, vous êtes le seul à pouvoir le consulter et y accéder. Pour permettre aux autres utilisateurs de votre compte d'accéder au lien de ressource, vous devez accorder des autorisations sur le lien de ressource lui-même. Vous devez accorder les autorisations DESCRIBE ou DROP. Sur la page Tables, sélectionnez à nouveau votre tableau et dans le menu Actions, choisissez Grant.
8. Dans la section Accorder des autorisations, sélectionnez Mon compte.
9. Pour les utilisateurs et les rôles IAM, sélectionnez l'utilisateur`testuser2`.
10. Pour les autorisations relatives aux liens vers les ressources, sélectionnez Décrire.
11. Choisissez Grant (Accorder).
12. Connectez-vous à la AWS console dans le compte B en tant que`testuser2`.

Sur la console Athena (https://console.aws.amazon.com/athena/), vous devriez voir la base de données et la table. `acc_b_area_rl` Vous pouvez désormais exécuter une requête sur la table pour voir la valeur de la colonne à laquelle `testuser2` vous avez accès.

Permissions d'intégration à Lake Formation

AWS Lake Formation utilise le AWS Glue Data Catalog pour stocker les métadonnées des données Amazon S3 sous forme de bases de données et de tables. Les tables stockent des informations sur les données sous-jacentes, notamment les informations de schéma, les informations de partition et l'emplacement des données. Les bases de données sont des ensembles de tables. Le catalogue de données contient également des liens vers des ressources, qui sont des liens vers des bases de données et des tables partagées dans des comptes externes, et sont utilisés pour l'accès entre comptes aux données du lac de données. Chaque AWS compte possède un catalogue de données par AWS région.

Lake Formation fournit un modèle d'autorisations du système de gestion de base de données relationnelle (RDBMS) permettant d'accorder ou de révoquer l'accès aux bases de données, aux tables et aux colonnes du catalogue de données contenant des données sous-jacentes dans Amazon S3.

Avant de connaître les détails du modèle d'autorisations de Lake Formation, il est utile de consulter les informations générales suivantes :

- Les lacs de données gérés par Lake Formation se trouvent dans des emplacements désignés dans Amazon Simple Storage Service (Amazon S3).
- Lake Formation gère un catalogue de données qui contient des métadonnées sur les données sources à importer dans vos lacs de données, telles que les données des journaux et des bases de données relationnelles, et sur les données de vos lacs de données dans Amazon S3. Les métadonnées sont organisées sous forme de bases de données et de tables. Les tables de métadonnées contiennent le schéma, l'emplacement, le partitionnement et d'autres informations sur les données qu'elles représentent. Les bases de métadonnées sont des ensembles de tables.
- Le Lake Formation Data Catalog est le même que celui utilisé parAWS Glue. Vous pouvez utiliser des AWS Glue robots d'exploration pour créer des tables de catalogue de données, et vous pouvez utiliser des tâches AWS Glue d'extraction, de transformation et de chargement (ETL) pour renseigner les données sous-jacentes de vos lacs de données.
- Les bases de données et les tables du catalogue de données sont appelées ressources du catalogue de données. Les tables du catalogue de données sont appelées tables de métadonnées pour les distinguer des tables des sources de données ou des données tabulaires d'Amazon S3. Les données vers lesquelles pointent les tables de métadonnées dans Amazon S3 ou dans les sources de données sont appelées données sous-jacentes.

- Un principal est un utilisateur ou un rôle, un QuickSight utilisateur ou un groupe Amazon, un utilisateur ou un groupe qui s'authentifie auprès de Lake Formation via un fournisseur SAML, ou pour le contrôle d'accès entre comptes, un identifiant de AWS compte, un identifiant d'organisation ou un identifiant d'unité organisationnelle.
- AWS Glueles robots créent des tables de métadonnées, mais vous pouvez également créer manuellement des tables de métadonnées à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI). Lorsque vous créez une table de métadonnées, vous devez spécifier un emplacement. Lorsque vous créez une base de données, l'emplacement est facultatif. Les emplacements des tables peuvent être des emplacements Amazon S3 ou des emplacements de sources de données tels qu'une base de données Amazon Relational Database Service (Amazon RDS). Les emplacements de base de données sont toujours des emplacements Amazon S3.
- Les services intégrés à Lake Formation, tels qu'Amazon Athena et Amazon Redshift, peuvent accéder au catalogue de données pour obtenir des métadonnées et vérifier l'autorisation d'exécuter des requêtes. Pour une liste complète des services intégrés, voirAWS intégrations de services avec Lake Formation.

Rubriques

- Vue d'ensemble des autorisations relatives à Lake Formation
- Référence des personnalités de Lake Formation et des autorisations IAM
- Modification des paramètres par défaut de votre lac de données
- Permissions implicites de Lake Formation
- Référence des autorisations de Lake Formation
- Intégration d'IAM Identity Center
- Ajouter un emplacement Amazon S3 à votre lac de données
- Mode d'accès hybride
- Création de tables et de bases de données du catalogue de données
- Importation de données à l'aide de flux de travail dans Lake Formation

Vue d'ensemble des autorisations relatives à Lake Formation

Il existe deux principaux types d'autorisations dans AWS Lake Formation :

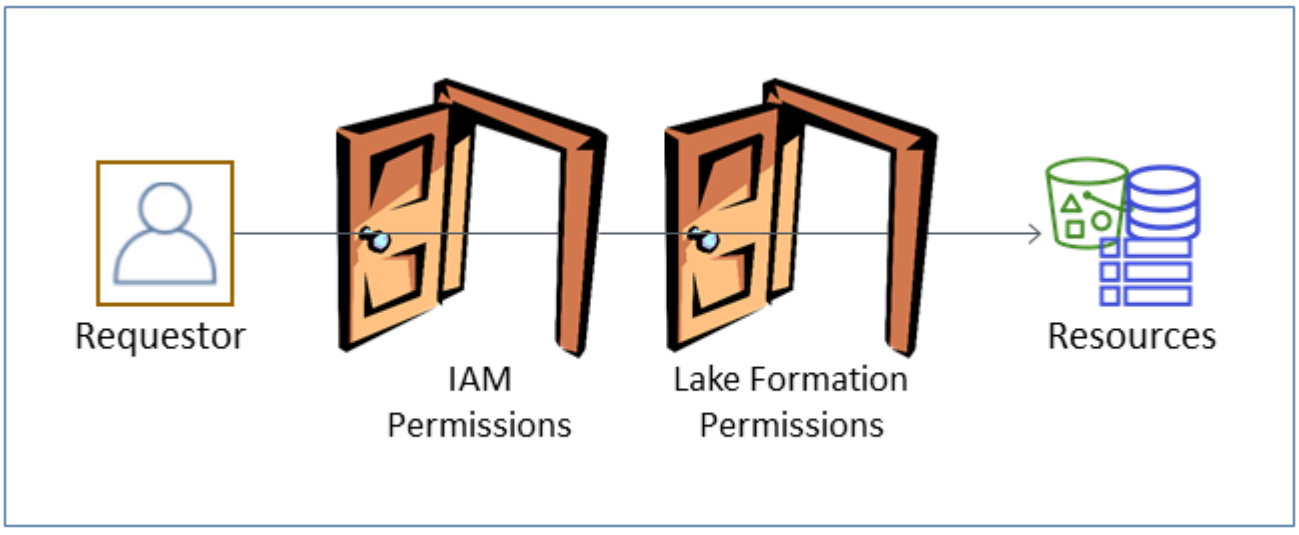
- Accès aux métadonnées : autorisations sur les ressources du catalogue de données (autorisations du catalogue de données).

  Ces autorisations permettent aux principaux de créer, de lire, de mettre à jour et de supprimer des bases de données et des tables de métadonnées dans le catalogue de données.

- Accès aux données sous-jacent : autorisations sur les sites Amazon Simple Storage Service (Amazon S3) (autorisations d'accès aux données et autorisations de localisation des données).
  - Les autorisations du lac de données permettent aux principaux de lire et d'écrire des données sur des emplacements Amazon S3 sous-jacents, des données pointées par les ressources du catalogue de données.
  - Les autorisations de localisation des données permettent aux principaux de créer et de modifier des bases de données et des tables de métadonnées qui pointent vers des emplacements Amazon S3 spécifiques.

Pour les deux zones, Lake Formation utilise une combinaison d'autorisations Lake Formation et d'autorisations AWS Identity and Access Management (IAM). Le modèle d'autorisations IAM comprend des politiques IAM. Le modèle d'autorisations de Lake Formation est implémenté sous forme de commandes GRANT/REVOKE de style DBMS, telles que. `Grant SELECT on` *`tableName`* `to` *`userName`*

Lorsqu'un principal demande d'accès aux ressources du catalogue de données ou aux données sous-jacentes, pour que la demande aboutisse, il doit passer les contrôles d'autorisation par IAM et par Lake Formation.



Les autorisations de Lake Formation contrôlent l'accès aux ressources du catalogue de données, aux sites Amazon S3 et aux données sous-jacentes de ces sites. Les autorisations IAM contrôlent l'accès à la Lake Formation, aux AWS Glue API et aux ressources. Ainsi, bien que vous ayez l'autorisation Lake Formation pour créer une table de métadonnées dans le catalogue de données (`CREATE_TABLE`), votre opération échoue si vous ne disposez pas de l'autorisation IAM sur l'`glue:CreateTable`API. (Pourquoi une `glue:` autorisation ? Parce que Lake Formation utilise le catalogue de AWS Glue données.)

> **Note**
>
> Les autorisations de Lake Formation ne s'appliquent que dans la région dans laquelle elles ont été accordées.

AWS Lake Formation exige que chaque principal (utilisateur ou rôle) soit autorisé à effectuer des actions sur les ressources gérées par Lake Formation. Le principal reçoit les autorisations nécessaires de la part de l'administrateur du lac de données ou d'un autre directeur autorisé à accorder les autorisations nécessaires à Lake Formation.

Lorsque vous accordez une autorisation de Lake Formation à un directeur, vous pouvez éventuellement accorder la possibilité de transmettre cette autorisation à un autre principal.

Vous pouvez utiliser l'API Lake Formation, le AWS Command Line Interface (AWS CLI) ou les pages Autorisations de données et Localisation des données de la console Lake Formation pour accorder et révoquer les autorisations de Lake Formation.

Méthodes de contrôle d'accès précis

Avec un lac de données, l'objectif est de disposer d'un contrôle d'accès précis aux données. Dans Lake Formation, cela signifie un contrôle d'accès précis aux ressources du catalogue de données et aux emplacements Amazon S3. Vous pouvez obtenir un contrôle d'accès précis à l'aide de l'une des méthodes suivantes.

Méthode	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
Méthode 1 :	Ouvrir	À grain fin	Il s'agit de la méthode par défaut pour la rétrocompatibilité avecAWS Glue. • Ouvert signifie que l'autorisation spéciale `Super` est accordée au groupe`IAMAllowedPrincipals` , qui `IAMAllowedPrincipals` est automatiquement créé et inclut tous les utilisateurs et rôles IAM autorisés à accéder aux ressources de votre catalogue de données par vos politiques IAM, et l'`Super`autorisation permet à un principal d'effectuer toutes les opération s de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée. Ainsi, l'accès aux ressources du catalogue de données et aux emplacements Amazon S3 est contrôlé uniquement par les politiques IAM. Pour plus d'informations, consultez Modification des paramètre s par défaut de votre lac de données et Mise à niveau AWS Glue des autorisat ions de données vers le AWS Lake Formation modèle. • La précision signifie que les politique s IAM contrôlent tous les accès aux ressources du catalogue de données et aux compartiments Amazon S3 individue ls.

Méthode	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
			Sur la console Lake Formation, cette méthode apparaît sous la forme Utiliser uniquement le contrôle d’accès IAM.

Méthode	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
Méthode 2 :	À grain fin	À gros grains	Il s'agit de la méthode recommandée. • Un accès précis implique l'octroi d'autorisations limitées de Lake Formation à des principaux responsab les individuels sur les ressources du catalogue de données, les sites Amazon S3 et les données sous-jacentes de ces sites. • Une granularité grossière signifie des autorisations plus étendues sur les opérations individuelles et sur l'accès aux sites Amazon S3. Par exemple, une politique IAM grossière peut inclure `"glue:*"` ou `"glue:Cre ate*"` plutôt laisser aux autorisat ions `"glue:CreateTables"` Lake Formation le soin de contrôler si un mandant peut ou non créer des objets de catalogue. Cela implique également de donner aux directeurs l'accès aux API dont ils ont besoin pour faire leur travail, mais de verrouiller les autres API et ressources. Par exemple, vous pouvez créer une politique IAM qui permet à un principal de créer des ressources de catalogue de données et de créer et d'exécuter des flux de travail, mais qui n'autorise pas la création de AWS Glue connexions ou de fonctions définies par

Méthode	Permissions relatives à la formation des lacs	Autorisations IAM	Commentaires
			l'utilisateur. Consultez les exemples plus loin dans cette section.

Important

Tenez compte des points suivants :

- Par défaut, les paramètres de contrôle d'accès Use only IAM de Lake Formation sont activés pour garantir la compatibilité avec le comportement existant du catalogue de AWS Glue données. Nous vous recommandons de désactiver ces paramètres une fois que vous serez passé à l'utilisation des autorisations de Lake Formation. Pour plus d’informations, consultez Modification des paramètres par défaut de votre lac de données.
- Les administrateurs de lacs de données et les créateurs de bases de données disposent d'autorisations implicites de Lake Formation que vous devez comprendre. Pour plus d’informations, consultez Permissions implicites de Lake Formation.

Contrôle d'accès aux métadonnées

Pour le contrôle d'accès aux ressources du catalogue de données, la discussion suivante suppose un contrôle d'accès précis avec les autorisations de Lake Formation et un contrôle d'accès grossier avec des politiques IAM.

Il existe deux méthodes distinctes pour accorder des autorisations à Lake Formation sur les ressources du catalogue de données :

- Contrôle d'accès aux ressources nommées : avec cette méthode, vous accordez des autorisations sur des bases de données ou des tables spécifiques en spécifiant les noms des bases de données ou des tables. Les subventions se présentent sous la forme suivante :

  Accordez des autorisations aux directeurs sur les ressources [avec option de subvention].

Avec l'option de subvention, vous pouvez autoriser le bénéficiaire à accorder les autorisations à d'autres principaux.

- Contrôle d'accès basé sur les balises : avec cette méthode, vous attribuez une ou plusieurs balises LF aux bases de données, aux tables et aux colonnes du catalogue de données, et vous accordez des autorisations sur une ou plusieurs balises LF aux principaux. Chaque balise LF est une paire clé-valeur, telle que. `department=sales` Un principal dont les balises LF correspondent aux balises LF d'une ressource de catalogue de données peut accéder à cette ressource. Cette méthode est recommandée pour les lacs de données contenant un grand nombre de bases de données et de tables. C'est expliqué en détail dansContrôle d'accès basé sur des balises Lake Formation.

Les autorisations dont dispose un principal sur une ressource sont l'union des autorisations accordées par les deux méthodes.

Le tableau suivant récapitule les autorisations de Lake Formation disponibles sur les ressources du catalogue de données. Les en-têtes des colonnes indiquent la ressource pour laquelle l'autorisation est accordée.

Catalogue	Base de données	Tableau
`CREATE_DATABASE`	`CREATE_TABLE`	`ALTER`
	`ALTER`	`DROP`
	`DROP`	`DESCRIBE`
	`DESCRIBE`	`SELECT*`
		`INSERT*`
		`DELETE*`

Par exemple, l'`CREATE_TABLE`autorisation est accordée sur une base de données. Cela signifie que le principal est autorisé à créer des tables dans cette base de données.

Les autorisations marquées d'un astérisque (*) sont accordées sur les ressources du catalogue de données, mais elles s'appliquent aux données sous-jacentes. Par exemple, l'`DROP`autorisation sur

une table de métadonnées vous permet de supprimer la table du catalogue de données. Cependant, l'`DELETE`autorisation accordée sur la même table vous permet de supprimer les données sous-jacentes de la table dans Amazon S3, à l'aide, par exemple, d'une `DELETE` instruction SQL. Avec ces autorisations, vous pouvez également consulter le tableau sur la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS GlueAPI. Ainsi, `SELECT`, `INSERT`, et `DELETE` sont à la fois des autorisations de catalogue de données et des autorisations d'accès aux données.

Lorsque vous `SELECT` accordez un accord sur un tableau, vous pouvez ajouter un filtre qui inclut ou exclut une ou plusieurs colonnes. Cela permet un contrôle d'accès précis sur les colonnes des tables de métadonnées, limitant ainsi les colonnes que les utilisateurs des services intégrés peuvent voir lorsqu'ils exécutent des requêtes. Cette fonctionnalité n'est pas disponible uniquement en utilisant les politiques IAM.

Il existe également une autorisation spéciale nommée`Super`. L'`Super`autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée. Cette autorisation peut coexister avec les autres autorisations de Lake Formation. Par exemple, vous pouvez accorder `Super``SELECT`, et `INSERT` sur une table de métadonnées. Le principal peut effectuer toutes les actions prises en charge sur la table, et lorsque vous révoquez`Super`, les `INSERT` autorisations `SELECT` et sont conservées.

Pour plus de détails sur chaque autorisation, voirRéférence des autorisations de Lake Formation.

> **Important**
>
> Pour pouvoir consulter une table de catalogue de données créée par un autre utilisateur, vous devez disposer d'au moins une autorisation Lake Formation sur la table. Si vous disposez d'au moins une autorisation sur la table, vous pouvez également voir la base de données contenant la table.

Vous pouvez accorder ou révoquer les autorisations du catalogue de données à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI). Voici un exemple de AWS CLI commande qui accorde à l'utilisateur l'`datalake_user1`autorisation de créer des tables dans la `retail` base de données.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
```

```
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Voici un exemple de politique IAM de contrôle d'accès grossière qui complète le contrôle d'accès détaillé par des autorisations Lake Formation. Il permet toutes les opérations sur n'importe quelle base de données ou table de métadonnées.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:*Database*",
                "glue:*Table*",
                "glue:*Partition*"
            ],
            "Resource": "*"
        }
    ]
}
```

L'exemple suivant est également grossier mais un peu plus restrictif. Il permet des opérations en lecture seule sur toutes les bases de données et tables de métadonnées du catalogue de données du compte et de la région désignés.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetDatabase",
                "glue:GetDatabases"
            ],
            "Resource": "arn:aws:glue:us-east-1:111122223333:*"
        }
    ]
}
```

Comparez ces politiques à la politique suivante, qui met en œuvre un contrôle d'accès précis basé sur l'IAM. Il accorde des autorisations uniquement sur un sous-ensemble de tables de la base de métadonnées de gestion de la relation client (CRM) du compte et de la région désignés.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetTables",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetDatabase",
                "glue:GetDatabases"
            ],
            "Resource": [
                "arn:aws:glue:us-east-1:111122223333:catalog",
                "arn:aws:glue:us-east-1:111122223333:database/CRM",
                "arn:aws:glue:us-east-1:111122223333:table/CRM/P*"
            ]
        }
    ]
}
```

Pour d'autres exemples de politiques de contrôle d'accès détaillées, voir. Référence des personnalités de Lake Formation et des autorisations IAM

Contrôle d'accès aux données sous-jacent

Lorsqu'un AWS service intégré demande l'accès aux données d'un site Amazon S3 dont l'accès est contrôlé par AWS Lake Formation, Lake Formation fournit des informations d'identification temporaires pour accéder aux données.

Pour permettre à Lake Formation de contrôler l'accès aux données sous-jacentes sur un site Amazon S3, vous devez enregistrer cet emplacement auprès de Lake Formation.

Après avoir enregistré un site Amazon S3, vous pouvez commencer à accorder les autorisations Lake Formation suivantes :

- Autorisations d'accès aux données (`SELECT`,`INSERT`,) et `DELETE`) sur les tables du catalogue de données qui pointent vers cet emplacement.

- Autorisations de localisation des données à cet emplacement.

Les autorisations de localisation des données de Lake Formation contrôlent la possibilité de créer des ressources de catalogue de données pointant vers des emplacements Amazon S3 particuliers. Les autorisations de localisation des données fournissent un niveau de sécurité supplémentaire aux emplacements du lac de données. Lorsque vous accordez l'`ALTER`autorisation `CREATE_TABLE` ou à un principal, vous accordez également des autorisations de localisation des données afin de limiter les emplacements pour lesquels le principal peut créer ou modifier des tables de métadonnées.

Les emplacements Amazon S3 sont des compartiments ou des préfixes situés sous un compartiment, mais pas des objets Amazon S3 individuels.

Vous pouvez accorder des autorisations de localisation des données à un directeur à l'aide de la console Lake Formation, de l'API ou du AWS CLI. La forme générale d'une subvention est la suivante :

```
grant DATA_LOCATION_ACCESS to principal on S3 location [with grant option]
```

Si vous incluez`with grant option`, le bénéficiaire peut accorder les autorisations à d'autres directeurs.

N'oubliez pas que les autorisations de Lake Formation fonctionnent toujours en combinaison avec les autorisations AWS Identity and Access Management (IAM) pour un contrôle d'accès précis. Pour les autorisations de lecture/écriture sur les données Amazon S3 sous-jacentes, les autorisations IAM sont accordées comme suit :

Lorsque vous enregistrez un emplacement, vous spécifiez un rôle IAM qui accorde des autorisations de lecture/écriture sur cet emplacement. Lake Formation assume ce rôle lorsqu'il fournit des informations d'identification temporaires aux AWS services intégrés. Un rôle typique peut être associé à la politique suivante, dans laquelle l'emplacement enregistré est le compartiment`awsexamplebucket`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
```

```
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket"
            ]
        }
    ]
}
```

Lake Formation fournit un rôle lié à un service que vous pouvez utiliser lors de votre inscription pour créer automatiquement de telles politiques. Pour plus d'informations, consultez Utilisation de rôles liés à un service pour Lake Formation.

Par conséquent, l'enregistrement d'un emplacement Amazon S3 accorde les `s3:` autorisations IAM requises pour cet emplacement, les autorisations étant spécifiées par le rôle utilisé pour enregistrer l'emplacement.

> **Important**
>
> Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Pour accéder en lecture/écriture aux données sous-jacentes, outre les autorisations de Lake Formation, les principaux ont également besoin de l'autorisation IAM suivante :

`lakeformation:GetDataAccess`

Avec cette autorisation, Lake Formation accède à la demande d'informations d'identification temporaires pour accéder aux données.

> **Note**
>
> Amazon Athena exige que l'utilisateur dispose de cette autorisation. `lakeformation:GetDataAccess` Les autres services intégrés ont besoin de leur rôle d'exécution sous-jacent pour être `lakeformation:GetDataAccess` autorisés.

Cette autorisation est incluse dans les politiques suggérées dans leRéférence des personnalités de Lake Formation et des autorisations IAM.

En résumé, pour permettre aux responsables de Lake Formation de lire et d'écrire des données sous-jacentes avec un accès contrôlé par les autorisations de Lake Formation :

- Enregistrez les sites Amazon S3 contenant les données auprès de Lake Formation.
- Les principaux responsables qui créent des tables de catalogue de données pointant vers des emplacements de données sous-jacents doivent disposer d'autorisations de localisation des données.
- Les directeurs qui lisent et écrivent les données sous-jacentes doivent disposer des autorisations d'accès aux données de Lake Formation dans les tables du catalogue de données qui pointent vers les emplacements des données sous-jacentes.
- Les directeurs qui lisent et écrivent les données sous-jacentes doivent disposer de l'autorisation `lakeformation:GetDataAccess` IAM lorsque l'emplacement des données sous-jacentes est enregistré auprès de Lake Formation.

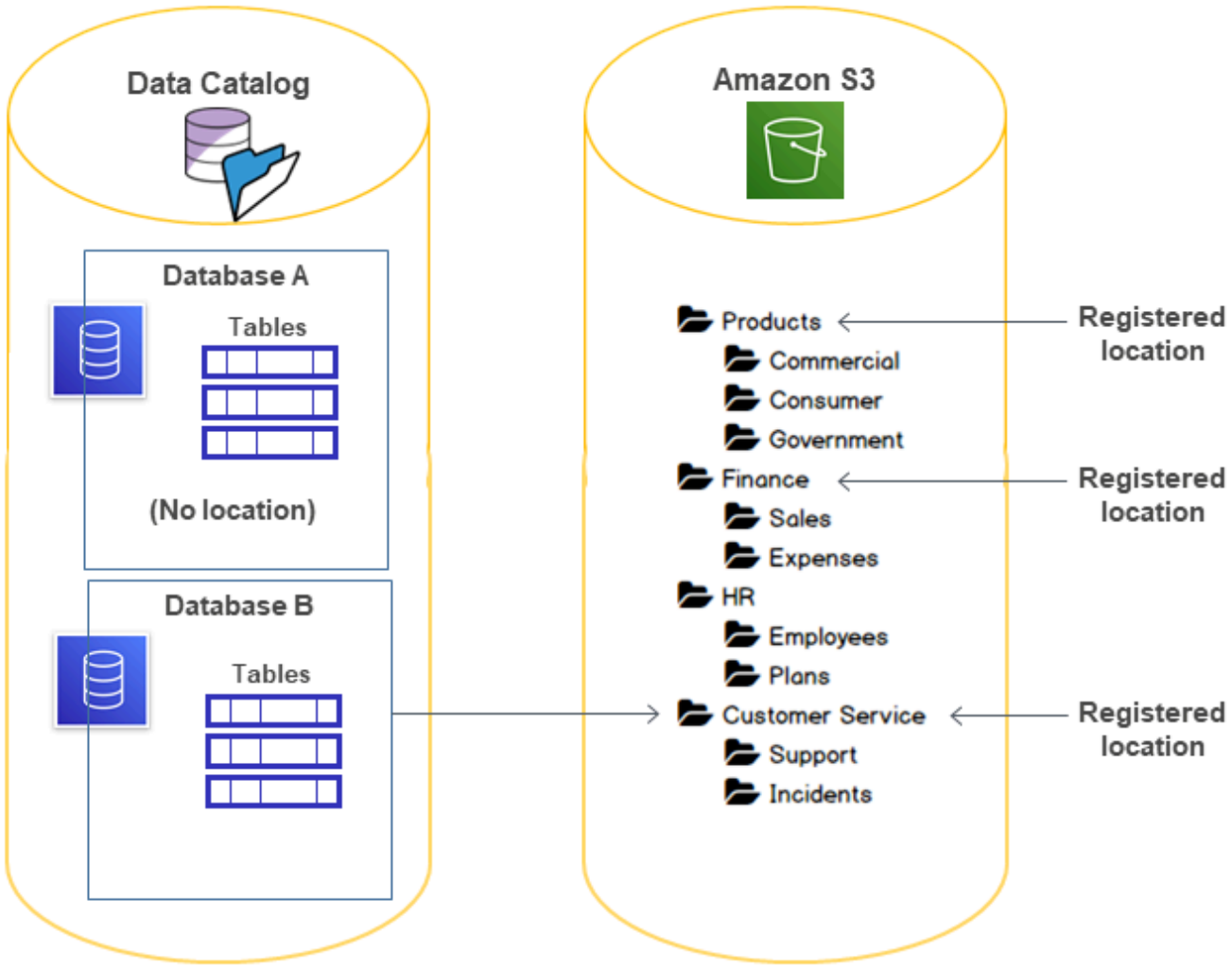
> **Note**
>
> Le modèle d'autorisations de Lake Formation n'empêche pas l'accès aux sites Amazon S3 via l'API ou la console Amazon S3 si vous y avez accès via les politiques IAM ou Amazon S3. Vous pouvez associer des politiques IAM aux principaux pour bloquer cet accès.

En savoir plus sur les autorisations de localisation des données

Les autorisations de localisation des données régissent le résultat des opérations de création et de mise à jour sur les bases de données et les tables du catalogue de données. Les règles sont les suivantes :

- Un principal doit disposer d'autorisations de localisation de données explicites ou implicites sur un emplacement Amazon S3 pour créer ou mettre à jour une base de données ou une table spécifiant cet emplacement.
- L'autorisation explicite `DATA_LOCATION_ACCESS` est accordée à l'aide de la console, de l'API ou AWS CLI.
- Des autorisations implicites sont accordées lorsqu'une base de données possède une propriété d'emplacement qui pointe vers un emplacement enregistré, que le principal dispose de l'`CREATE_TABLE`autorisation sur la base de données et que le principal essaie de créer une table à cet emplacement ou à un emplacement enfant.
- Si un principal obtient des autorisations de localisation des données sur un emplacement, il dispose des autorisations de localisation des données sur tous les emplacements enfants.
- Un principal n'a pas besoin d'autorisations de localisation des données pour effectuer des opérations de lecture/écriture sur les données sous-jacentes. Il suffit d'avoir les autorisations `SELECT` d'accès aux `INSERT` données. Les autorisations de localisation des données s'appliquent uniquement à la création de ressources de catalogue de données pointant vers l'emplacement.

Examinez le scénario illustré dans le schéma suivant.



Dans ce schéma :

- Les compartiments Amazon S3 `ProductsFinance`, et `Customer Service` sont enregistrés auprès de Lake Formation.
- `Database A`n'a aucune propriété de localisation, `Database B` mais possède une propriété de localisation qui pointe vers le `Customer Service` bucket.
- L'utilisateur `datalake_user` possède `CREATE_TABLE` les deux bases de données.
- L'utilisateur `datalake_user` a obtenu des autorisations de localisation des données uniquement sur le `Products` compartiment.

Voici les résultats obtenus lorsque l'utilisateur `datalake_user` essaie de créer une table de catalogue dans une base de données donnée à un emplacement donné.

Emplacement où **`datalake_user`** essaie de créer une table

Base de données et localisation	Succès ou échec	Raison
Base de données A à `Finance/Sales`	Échoue	Aucune autorisation de localisation des données
Base de données A à `Products`	Succès	Dispose d'une autorisation de localisation des données
Base de données A à `HR/Plans`	Succès	L'emplacement n'est pas enregistré
Base de données B à `Customer Service/Incidents`	Succès	La base de données possède une propriété de localisation à `Customer Service`

Pour plus d’informations, consultez les ressources suivantes :

- Ajouter un emplacement Amazon S3 à votre lac de données
- Référence des autorisations de Lake Formation
- Référence des personnalités de Lake Formation et des autorisations IAM

Référence des personnalités de Lake Formation et des autorisations IAM

Cette section répertorie certains personnages de Lake Formation suggérés et leurs autorisations suggérées AWS Identity and Access Management (IAM). Pour plus d'informations sur les autorisations de Lake Formation, consultezthe section called “Référence des autorisations de Lake Formation”.

AWS Lake Formation personas

Le tableau suivant répertorie les AWS Lake Formation personnages suggérés.

Lake Formation Personas

Persona	Description
Administrateur IAM (superutilisateur)	(Obligatoire) Utilisateur capable de créer des utilisateurs et des rôles IAM. Possède la politique `AdministratorAccess` AWS gérée. Possède toutes les autorisations sur toutes les ressources du Lake Formation. Possibilité d'ajouter des administrateurs de data lake. Impossible d'accorder les autorisations de Lake Formation s'il n'est pas également désigné administrateur du lac de données.
Administrateur du lac de données	(Obligatoire) Utilisateur pouvant enregistrer des sites Amazon S3, accéder au catalogue de données, créer des bases de données, créer et exécuter des flux de travail, accorder des autorisations Lake Formation à d'autres utilisateurs et consulter les AWS CloudTrail journaux. Dispose de moins d'autorisations IAM que l'administrateur IAM, mais suffisamment pour administrer le lac de données. Impossible d'ajouter d'autres administrateurs de data lake.
Administrateur en lecture seule	(Facultatif) Utilisateur qui peut consulter les informations principales, les ressources du catalogue de données, les autorisations et les AWS CloudTrail journaux, sans les autorisations nécessaires pour effectuer des mises à jour.
Ingénieur de données	(Facultatif) Utilisateur capable de créer des bases de données, de créer et d'exécuter des robots d'exploration et des flux de travail, et d'accorder des autorisations à Lake Formation sur les tables du catalogue de données créées par les robots et les flux de travail. Nous vous recommandons de faire de tous les ingénieurs de données des créateurs de bases de données. Pour plus d'informations, consultez Création d'une base de données.
Analyste des données	(Facultatif) Utilisateur qui peut exécuter des requêtes sur le lac de données en utilisant, par exemple, Amazon Athena. Dispose

Persona	Description
	uniquement des autorisations suffisantes pour exécuter des requêtes.
Rôle du flux de travail	(Obligatoire) Rôle qui exécute un flux de travail pour le compte d'un utilisateur. Vous spécifiez ce rôle lorsque vous créez un flux de travail à partir d'un plan.

AWS politiques gérées pour Lake Formation

Vous pouvez accorder les autorisations AWS Identity and Access Management (IAM) requises pour travailler en AWS Lake Formation utilisant des politiques AWS gérées et des politiques intégrées. Les politiques AWS gérées suivantes sont disponibles pour Lake Formation.

AWS politique gérée : AWSLakeFormationDataAdmin

AWSLakeFormationDataAdminla politique accorde un accès administratif AWS Lake Formation aux services connexes tels que la gestion AWS Glue des lacs de données.

Vous pouvez vous associer `AWSLakeFormationDataAdmin` à vos utilisateurs, groupes et rôles.

Détails de l'autorisation

- `CloudTrail`— Permet aux directeurs de consulter les AWS CloudTrail journaux. Cela est nécessaire pour vérifier toute erreur dans la configuration du lac de données.
- `Glue`— Permet aux principaux d'afficher, de créer et de mettre à jour des tables de métadonnées et des bases de données dans le catalogue de données. Cela inclut les opérations d'API qui commencent par `Get` `ListCreate`,`Update`,`Delete`, et`Search`. Cela est nécessaire pour gérer les métadonnées des tables des lacs de données.
- `IAM`— Permet aux principaux de récupérer des informations sur les utilisateurs IAM, les rôles et les politiques associées aux rôles. Cela est nécessaire pour que l'administrateur des données puisse examiner et répertorier les utilisateurs et les rôles IAM afin d'accorder des autorisations à Lake Formation.
- `Lake Formation`— Accorde aux administrateurs de lacs de données les autorisations nécessaires à Lake Formation pour gérer les lacs de données.
- `S3`— Permet aux principaux de récupérer des informations sur les compartiments Amazon S3 et leur emplacement afin de configurer l'emplacement des données pour les lacs de données.

```
"Statement": [
        {
            "Sid": "AWSLakeFormationDataAdminAllow",
            "Effect": "Allow",
            "Action": [
                "lakeformation:*",
                "cloudtrail:DescribeTrails",
                "cloudtrail:LookupEvents",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetConnections",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTableVersions",
                "glue:GetPartitions",
                "glue:GetTables",
                "glue:ListWorkflows",
                "glue:BatchGetWorkflows",
                "glue:DeleteWorkflow",
                "glue:GetWorkflowRuns",
                "glue:StartWorkflowRun",
                "glue:GetWorkflow",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
                "s3:GetBucketAcl",
                "iam:ListUsers",
                "iam:ListRoles",
                "iam:GetRole",
                "iam:GetRolePolicy"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AWSLakeFormationDataAdminDeny",
            "Effect": "Deny",
            "Action": [
```

```
                "lakeformation:PutDataLakeSettings"
            ],
                "Resource": "*"
        }
    ]
}
```

> **Note**
>
> La `AWSLakeFormationDataAdmin` politique n'accorde pas toutes les autorisations requises aux administrateurs de lacs de données. Des autorisations supplémentaires sont nécessaires pour créer et exécuter des flux de travail et enregistrer des sites avec le rôle lié au service`AWSServiceRoleForLakeFormationDataAccess`. Pour plus d'informations, consultez Création d'un administrateur de lac de données et Utilisation de rôles liés à un service pour Lake Formation.

AWS politique gérée : AWSLakeFormationCrossAccountManager

AWSLakeFormationCrossAccountManagerla politique fournit un accès multicompte aux AWS Glue ressources via Lake Formation et accorde un accès en lecture à d'autres services requis tels que AWS Organizations et AWS RAM.

Vous pouvez vous associer `AWSLakeFormationCrossAccountManager` à vos utilisateurs, groupes et rôles.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes.

- `Glue`— Permet aux principaux de définir ou de supprimer la politique de ressources du catalogue de données pour le contrôle d'accès.
- `Organizations`— Permet aux responsables de récupérer les informations relatives aux comptes et aux unités organisationnelles (UO) d'une organisation.
- `ram:CreateResourceShare`— Permet aux principaux de créer un partage de ressources.
- `ram:UpdateResourceShare`: permet aux principaux de modifier certaines propriétés du partage de ressources spécifié.

- `ram:DeleteResourceShare`— Permet aux principaux de supprimer le partage de ressources spécifié.
- `ram:AssociateResourceShare`— Permet aux principaux d'ajouter la liste de principes et la liste de ressources spécifiées à un partage de ressources.
- `ram:DisassociateResourceShare`— Permet aux principaux d'empêcher les principaux ou les ressources spécifiés de participer au partage de ressources spécifié.
- `ram:GetResourceShares`— Permet aux principaux de récupérer des informations sur les partages de ressources que vous possédez ou qui sont partagés avec vous.
- `ram:RequestedResourceType`— Permet aux principaux de récupérer le type de ressource (base de données, table ou catalogue).
- `AssociateResourceSharePermission`— Permet aux principaux d'ajouter ou de remplacer l' AWS RAM autorisation pour un type de ressource inclus dans un partage de ressources. Vous pouvez avoir exactement une autorisation associée à chaque type de ressource dans le partage de ressources.

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Sid": "AllowCreateResourceShare",
            "Effect": "Allow",
            "Action": [
                "ram:CreateResourceShare"
            ],
            "Resource": "*",
            "Condition": {
                "StringLikeIfExists": {
                    "ram:RequestedResourceType": [
                        "glue:Table",
                        "glue:Database",
                        "glue:Catalog"
                    ]
                }
            }
        },
        {
            "Sid": "AllowManageResourceShare",
            "Effect": "Allow",
            "Action": [
                "ram:UpdateResourceShare",
```

```
                "ram:DeleteResourceShare",
                "ram:AssociateResourceShare",
                "ram:DisassociateResourceShare",
                "ram:GetResourceShares"
            ],
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "ram:ResourceShareName": [
                        "LakeFormation*"
                    ]
                }
            }
        },
        {
            "Sid": "AllowManageResourceSharePermissions",
            "Effect": "Allow",
            "Action": [
                "ram:AssociateResourceSharePermission"
            ],
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "ram:PermissionArn": [
                        "arn:aws:ram::aws:permission/AWSRAMLFEnabled*"
                    ]
                }
            }
        },
        {
            "Sid": "AllowXAcctManagerPermissions",
            "Effect": "Allow",
            "Action": [
                "glue:PutResourcePolicy",
                "glue:DeleteResourcePolicy",
                "organizations:DescribeOrganization",
                "organizations:DescribeAccount",
                "ram:Get*",
                "ram:List*"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AllowOrganizationsPermissions",
```

```
            "Effect": "Allow",
            "Action": [
                "organizations:ListRoots",
                "organizations:ListAccountsForParent",
                "organizations:ListOrganizationalUnitsForParent"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS politique gérée : AWSGlueConsoleFullAccess

AWSGlueConsoleFullAccessla politique accorde un accès complet aux AWS Glue ressources lorsqu'une identité à laquelle la politique est attachée utilise le AWS Management Console. Si vous suivez la convention de dénomination pour les ressources spécifiées dans la politique, les utilisateurs bénéficient des capacités totales de la console. Cette politique s'applique généralement aux utilisateurs de la AWS Glue console.

Lake Formation assume également le rôle `AWSGlueServiceRole` de service permettant d'accéder aux services connexes, notamment Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3) et Amazon. AWS Glue CloudWatch

AWS managed policy:LakeFormationDataAccessServiceRolePolicy

Cette politique est attachée à un rôle lié au service nommé `ServiceRoleForLakeFormationDataAccess` qui permet au service d'effectuer des actions sur les ressources à votre demande. Vous ne pouvez pas associer cette politique à vos identités IAM.

Cette politique permet aux AWS services intégrés de Lake Formation tels qu'Amazon Redshift Amazon Athena d'utiliser le rôle lié au service pour découvrir les ressources Amazon S3.

Pour plus d'informations, veuillez consulter Utilisation de rôles liés à un service pour Lake Formation.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes.

- `s3:ListAllMyBuckets`— Renvoie la liste de tous les buckets appartenant à l'expéditeur authentifié de la demande.

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "LakeFormationDataAccessServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "s3:ListAllMyBuckets"
   ],
   "Resource": [
    "arn:aws:s3:::*"
   ]
  }
 ]
}
```

Lake Formation met à jour les politiques AWS gérées

Consultez les détails des mises à jour des politiques AWS gérées pour Lake Formation depuis que ce service a commencé à suivre ces modifications.

Modification	Description	Date
`AWSLakeFormationCrossAccountManager` Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationCrossAccountManagerpolitique en ajoutant des éléments Sid à la déclaration de politique.	mars 2024
`AWSLakeFormationDataAdmin` Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationDataAdminpolitique en ajoutant un élément Sid à la déclaration de politique et en supprimant une action redondante.	mars 2024
`LakeFormationDataAccessServiceRolePo`	Lake Formation a amélioré la LakeFormationDataAccessServiceRolePolicypolitique en ajoutant un	février 2024

Modification	Description	Date
`licy` Politique mise à jour de Lake Formation.	élément Sid à la déclaration de politique .	
`AWSLakeFormationCrossAccountManager` Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationCrossAccountManagerpolitique en ajoutant une nouvelle autorisation pour permettre le partage de données entre comptes en mode d'accès hybride.	octobre 2023
`AWSLakeFormationCrossAccountManager` Politique mise à jour de Lake Formation.	Lake Formation a amélioré la AWSLakeFormationCrossAccountManagerpolitique de création d'un seul partage de ressources par compte bénéficiaire lors du premier partage d'une ressource. Toutes les ressources partagées par la suite avec le même compte sont rattachées au même partage de ressources.	6 mai 2022
Lake Formation a commencé à suivre les changements.	Lake Formation a commencé à suivre les modifications apportées AWS à ses politiques gérées.	6 mai 2022

Autorisations suggérées par Persona

Les autorisations suggérées pour chaque personnage sont les suivantes. L'administrateur IAM n'est pas inclus car il dispose de toutes les autorisations sur toutes les ressources.

Rubriques

- Autorisations d'administrateur du lac de données
- Autorisations d'administrateur en lecture seule
- Autorisations d'ingénieur de données
- Autorisations d'analyse de données

- Autorisations relatives aux rôles du flux

Autorisations d'administrateur du lac de données

> **Important**
>
> Dans les politiques suivantes, remplacez <account-id>par un numéro de AWS compte valide et remplacez <workflow_role>par le nom d'un rôle autorisé à exécuter un flux de travail, comme défini dansAutorisations relatives aux rôles du flux.

Type de politique	Politique
AWS politiques gérées	• `AWSLakeFormationDataAdmin` • `LakeFormationDataAccessServiceRolePolicy` (politique des rôles liés au service) • `AWSGlueConsoleFullAccess` (facultatif) • `CloudWatchLogsReadOnlyAccess` (Facultatif) • `AWSLakeFormationCrossAccountManager` (Facultatif) • `AmazonAthenaFullAccess` (facultatif) Pour plus d'informations sur les politiques AWS gérées facultatives, consultezthe section called “Création d'un administrateur de lac de données”.

Politique en ligne (pour créer le rôle lié au service Lake Formation)	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": "iam:CreateServiceLinkedRole",
	"Resource": "*",
	"Condition": {
	"StringEquals": {

Type de politique	Politique

	<pre>"iam:AWSServiceName": "lakeformation.amazonaws.com"
	}
	}
	},
	{
	"Effect": "Allow",
	"Action": [
	"iam:PutRolePolicy"
	],
	"Resource": "arn:aws:iam:: *<account-id>* :role/aws-service-role/lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess"
	}
	]
	}</pre>

(Facultatif) Politique intégrée (politique de passe-rôle pour le rôle de flux de travail). Cela n'est nécessaire que si l'administrateur du lac de données crée et exécute des flux de travail.	<pre>{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow_role>* "
	]
	}
	]
	}</pre>

Type de politique	Politique

(Facultatif) Politique en ligne (si votre compte accorde ou reçoit des autorisations entre comptes Lake Formation). Cette politique permet d'accepter ou de rejeter les invitations à partager des AWS RAM ressources et d'autoriser l'octroi d'autorisations entre comptes aux organisations. `ram:EnableSharingWithAwsOrganization` est obligatoire uniquement pour les administrateurs de data lake dans le compte AWS Organizations de gestion.	<pre>{
	    "Version": "2012-10-17",
	    "Statement": [
	        {
	            "Effect": "Allow",
	            "Action": [
	                "ram:AcceptResourceShareInvitation",
	                "ram:RejectResourceShareInvitation",
	                "ec2:DescribeAvailabilityZones",
	                "ram:EnableSharingWithAwsOrganization"
	            ],
	            "Resource": "*"
	        }
	    ]
	}</pre>

Autorisations d'administrateur en lecture seule

Type de stratégie	Politique

Politique en ligne (de base)	<pre>{
	   "Version":"2012-10-17",
	   "Statement":[
	      {
	         "Effect":"Allow",
	         "Action":[
	            "lakeformation:GetEffectivePermissionsForPath",
	            "lakeformation:ListPermissions",
	            "lakeformation:ListDataCellsFilter",
	            "lakeformation:GetDataCellsFilter",
	            "lakeformation:SearchDatabasesByLFTags",
	            "lakeformation:SearchTablesByLFTags",
	            "lakeformation:GetLFTag",</pre>

Type de stratégie	Politique

```
            "lakeformation:ListLFTags",
            "lakeformation:GetResourceLFTags",
            "lakeformation:ListLakeFormationOpti
ns",
            "cloudtrail:DescribeTrails",
            "cloudtrail:LookupEvents",
            "glue:GetDatabase",
            "glue:GetDatabases",
            "glue:GetConnections",
            "glue:SearchTables",
            "glue:GetTable",
            "glue:GetTableVersions",
            "glue:GetPartitions",
            "glue:GetTables",
            "glue:GetWorkflow",
            "glue:ListWorkflows",
            "glue:BatchGetWorkflows",
            "glue:GetWorkflowRuns",
            "glue:GetWorkflow",
            "s3:ListBucket",
            "s3:GetBucketLocation",
            "s3:ListAllMyBuckets",
            "s3:GetBucketAcl",
            "iam:ListUsers",
            "iam:ListRoles",
            "iam:GetRole",
            "iam:GetRolePolicy"
         ],
         "Resource":"*"
      },
      {
         "Effect":"Deny",
         "Action":[
            "lakeformation:PutDataLakeSettings"
         ],
         "Resource":"*"
      }
   ]
}
```

Autorisations d'ingénieur de données

> **Important**
>
> Dans les politiques suivantes, remplacez <account-id>par un numéro de AWS compte valide et remplacez <workflow_role>par le nom du rôle du flux de travail.

Type de politique	Politique
AWS politique gérée	`AWSGlueConsoleFullAccess`
Politique en ligne (de base)	

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess",
                "lakeformation:GrantPermissions",
                "lakeformation:RevokePermissions",
                "lakeformation:BatchGrantPermissions",
                "lakeformation:BatchRevokePermissions",
                "lakeformation:ListPermissions",
                "lakeformation:AddLFTagsToResource",
                "lakeformation:RemoveLFTagsFromResou
rce",
                "lakeformation:GetResourceLFTags",
                "lakeformation:ListLFTags",
                "lakeformation:GetLFTag",
                "lakeformation:SearchTablesByLFTags",
                "lakeformation:SearchDatabasesByLFTa
gs",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults",
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetQueryStatistics"
            ],
            "Resource": "*"
        }
```

Type de politique	Politique
	] }
Politique intégrée (pour les opérations sur les tables gouvernées, y compris les opérations au sein des transactions)	{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction", "lakeformation:DescribeTransaction", "lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation:DeleteObjectsOnCancel"], "Resource": "*" }] }

Type de politique	Politique

Politique en ligne (pour le contrôle d'accès aux métadonnées à l'aide de la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC))	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"lakeformation:AddLFTagsToResource",
	"lakeformation:RemoveLFTagsFromResou
	rce",
	"lakeformation:GetResourceLFTags",
	"lakeformation:ListLFTags",
	"lakeformation:GetLFTag",
	"lakeformation:SearchTablesByLFTags",
	"lakeformation:SearchDatabasesByLFTags"
	],
	"Resource": "*"
	}
	]
	}

Politique intégrée (politique de passe-rôle pour le rôle de flux de travail)	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow
	_role>* "
	]
	}
	]
	}

Autorisations d'analyse de données

Type de politique	Politique
AWS politique gérée	`AmazonAthenaFullAccess`
Politique en ligne (de base)	`{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "glue:GetTable", "glue:GetTables", "glue:SearchTables", "glue:GetDatabase", "glue:GetDatabases", "glue:GetPartitions", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTags"], "Resource": "*" }] }`
(Facultatif) Politique intégrée (pour les opérations sur les tables gouvernées, y compris les opérations au sein des transactions)	`{ "Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Action": [ "lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction", "lakeformation:DescribeTransaction",`

Type de politique	Politique

	<pre>"lakeformation:ListTransactions",
	"lakeformation:GetTableObjects",
	"lakeformation:UpdateTableObjects",
	"lakeformation:DeleteObjectsOnCancel"
	],
	"Resource": "*"
	}
	]
	}</pre>

Autorisations relatives aux rôles du flux

Ce rôle dispose des autorisations requises pour exécuter un flux de travail. Vous spécifiez un rôle doté de ces autorisations lorsque vous créez un flux de travail.

> **Important**
>
> Dans les politiques suivantes, remplacez-le <region>par un identifiant de AWS région valide (par exemple`us-east-1`), <account-id>par un numéro de AWS compte valide, <workflow_role>par le nom du rôle du flux de travail et <your-s3-cloudtrail-bucket>par le chemin Amazon S3 vers vos AWS CloudTrail journaux.

Type de politique	Politique
AWS politique gérée	`AWSGlueServiceRole`

Politique en ligne (accès aux données)	<pre>{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "Lakeformation",
	"Effect": "Allow",
	"Action": [
	"lakeformation:GetDataAccess",
	"lakeformation:GrantPermissions"
	],</pre>

Type de politique	Politique

	<pre>"Resource": "*"
	}
	]
	}</pre>

Politique intégrée (politique de passe-rôle pour le rôle de flux de travail)	<pre>{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: <account-id> :role/<workflow_role> "
	]
	}
	]
	}</pre>

Politique en ligne (pour l'ingestion de données en dehors du lac de données, par exemple, AWS CloudTrail des journaux)	<pre>{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": ["s3:GetObject", "s3:ListBucket"],
	"Resource": ["arn:aws:s3::: <your-s3-cloudtrail-bucket> /*"]
	}
	]
	}</pre>

Modification des paramètres par défaut de votre lac de données

Pour maintenir la rétrocompatibilité avecAWS Glue, AWS Lake Formation possède les paramètres de sécurité initiaux suivants :

- L'`Super`autorisation est accordée au groupe `IAMAllowedPrincipals` sur toutes les ressources du catalogue de AWS Glue données existantes.
- Les paramètres « Utiliser uniquement le contrôle d'accès IAM » sont activés pour les nouvelles ressources du catalogue de données.

Ces paramètres font en sorte que l'accès aux ressources du catalogue de données et aux emplacements Amazon S3 soit contrôlé uniquement par des politiques AWS Identity and Access Management (IAM). Les autorisations individuelles de Lake Formation ne sont pas en vigueur.

Le `IAMAllowedPrincipals` groupe inclut tous les utilisateurs et rôles IAM autorisés à accéder aux ressources de votre catalogue de données par vos politiques IAM. L'`Super`autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table pour laquelle elle est accordée.

Pour modifier les paramètres de sécurité afin que l'accès aux ressources du catalogue de données (bases de données et tables) soit géré par les autorisations de Lake Formation, procédez comme suit :

1. Modifiez les paramètres de sécurité par défaut pour les nouvelles ressources. Pour obtenir des instructions, veuillez consulter Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride.
2. Modifiez les paramètres des ressources du catalogue de données existantes. Pour obtenir des instructions, veuillez consulter Mise à niveau AWS Glue des autorisations de données vers le AWS Lake Formation modèle.

Modification des paramètres de sécurité par défaut à l'aide de l'opération d'**`PutDataLakeSettings`**API Lake Formation

Vous pouvez également modifier les paramètres de sécurité par défaut à l'aide de l'opération d'PutDataLakeSettingsAPI Lake Formation. Cette action prend comme arguments un ID de catalogue facultatif et une DataLakeSettingsstructure.

Pour renforcer le contrôle d'accès aux métadonnées et aux données sous-jacentes par Lake Formation sur les nouvelles bases de données et tables, codez la `DataLakeSettings` structure comme suit.

> **Note**
>
> <AccountID>Remplacez-le par un identifiant de AWS compte valide et <Username>par un nom d'utilisateur IAM valide. Vous pouvez définir plusieurs utilisateurs en tant qu'administrateurs de data lake.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [],
        "CreateTableDefaultPermissions": []
    }
}
```

Vous pouvez également coder la structure comme suit. L'omission du `CreateTableDefaultPermissions` paramètre `CreateDatabaseDefaultPermissions` or équivaut à transmettre une liste vide.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ]
    }
}
```

Cette action révoque effectivement toutes les autorisations de Lake Formation accordées au `IAMAllowedPrincipals` groupe sur les nouvelles bases de données et tables. Lorsque vous créez une base de données, vous pouvez annuler ce paramètre.

Pour appliquer le contrôle d'accès aux métadonnées et aux données sous-jacentes uniquement par IAM sur les nouvelles bases de données et tables, codez la `DataLakeSettings` structure comme suit.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": [
                    "ALL"
                ]
            }
        ],
        "CreateTableDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": [
                    "ALL"
                ]
            }
        ]
    }
}
```

Cela permet à la `Super` Lake Formation d'autoriser le `IAMAllowedPrincipals` groupe à utiliser de nouvelles bases de données et de nouvelles tables. Lorsque vous créez une base de données, vous pouvez annuler ce paramètre.

> **Note**
>
> Dans la `DataLakeSettings` structure précédente, la seule valeur autorisée pour `DataLakePrincipalIdentifier` est`IAM_ALLOWED_PRINCIPALS`, et la seule valeur autorisée pour `Permissions` est`ALL`.

Permissions implicites de Lake Formation

AWS Lake Formation accorde les autorisations implicites suivantes aux administrateurs de lacs de données, aux créateurs de bases de données et aux créateurs de tables.

Administrateurs de data lake

- `Describe`Accédez à toutes les ressources du catalogue de données, à l'exception des ressources partagées directement depuis un autre compte avec un autre principal. Cet accès ne peut pas être révoqué par un administrateur.
- Disposez d'autorisations de localisation des données partout dans le lac de données.
- Peut accorder ou révoquer l'accès à toutes les ressources du catalogue de données à n'importe quel principal (y compris lui-même). Cet accès ne peut pas être révoqué par un administrateur.
- Peut créer des bases de données dans le catalogue de données.
- Peut accorder l'autorisation de créer une base de données à un autre utilisateur.

> **Note**
>
> Les administrateurs de data lake peuvent enregistrer des sites Amazon S3 uniquement s'ils disposent des autorisations IAM pour le faire. Les politiques proposées aux administrateurs des lacs de données dans ce guide accordent ces autorisations. En outre, les administrateurs des lacs de données ne disposent pas des autorisations implicites leur permettant de supprimer des bases de données ou de modifier/supprimer des tables créées par d'autres. Ils peuvent toutefois s'octroyer l'autorisation de le faire.

Pour plus d'informations sur les administrateurs de data lake, consultezCréation d'un administrateur de lac de données.

créateurs de bases de données

- Disposez de toutes les autorisations de base de données sur les bases de données qu'ils créent, d'autorisations sur les tables qu'ils créent dans la base de données et vous pouvez accorder aux autres principaux du même AWS compte l'autorisation de créer des tables dans la base de données. Un créateur de base de données qui dispose également de la politique `AWSLakeFormationCrossAccountManager` AWS gérée peut accorder des autorisations sur la base de données à d'autres AWS comptes ou organisations.

  Les administrateurs de data lake peuvent utiliser la console ou l'API Lake Formation pour désigner les créateurs de bases de données.

  > **Note**
  >
  > Les créateurs de base de données ne disposent pas implicitement d'autorisations sur les tables créées par d'autres utilisateurs dans la base de données.

Pour plus d'informations, consultez Création d'une base de données.

Créateurs de tables

- Disposent de toutes les autorisations sur les tables qu'ils créent.
- Ils peuvent accorder des autorisations sur toutes les tables qu'ils créent aux principaux d'un même AWS compte.
- Ils peuvent accorder des autorisations sur toutes les tables qu'ils créent à d'autres AWS comptes ou organisations s'ils disposent de la politique `AWSLakeFormationCrossAccountManager` AWS gérée.
- Peut afficher les bases de données contenant les tables qu'ils créent.

Référence des autorisations de Lake Formation

Pour effectuer des AWS Lake Formation opérations, les directeurs ont besoin à la fois des autorisations Lake Formation et des autorisations AWS Identity and Access Management (IAM). Vous accordez généralement des autorisations IAM à l'aide de politiques de contrôle d'accès grossières, comme décrit dans. the section called “Vue d'ensemble des autorisations relatives à Lake Formation ” Vous pouvez accorder des autorisations à Lake Formation à l'aide de la console, de l'API ou du AWS Command Line Interface (AWS CLI).

Pour savoir comment accorder ou révoquer les autorisations de Lake Formation, consultez the section called “Octroi et révocation des autorisations du catalogue de données” etthe section called “Octroi d'autorisations de localisation des données”.

> **Note**
>
> Les exemples présentés dans cette section montrent comment accorder des autorisations aux principaux d'un même AWS compte. Pour des exemples de subventions entre comptes, voirthe section called “Partage de données entre comptes”.

Permissions de Lake Formation par type de ressource

Voici les autorisations valides de Lake Formation disponibles pour chaque type de ressource :

Ressource	Autorisation
`Database`	`ALL (Super)`
	`ALTER`
	`CREATE_TABLE`
	`DESCRIBE`
	`DROP`
`Table`	`ALL (Super)`
	`ALTER`
	`DELETE`
	`DESCRIBE`
	`DROP`
	`INSERT`
	`SELECT`

Ressource	Autorisation
View	ALL (Super)
	SELECT
	DESCRIBE
	DROP
Data Catalog	CREATE_DATABASE
Amazon S3 location	DATA_LOCATION_ACCESS
LF-Tags	DROP
	ALTER
LF-Tag values	ASSOCIATE
	DESCRIBE
	GrantWithLFTagExpression
LF-Tag policy - Database	ALL (Super)
	ALTER
	CREATE_TABLE
	DESCRIBE
	DROP
LF-Tag policy - Table	ALL (Super)
	ALTER
	DESCRIBE
	DELETE

Ressource	Autorisation
	DROP
	INSERT
	SELECT
Resource link - Database or Table	DESCRIBE
	DROP
Table with data filters	DESCRIBE
	DROP
	SELECT
Table with column filter	SELECT

Rubriques

- Lake Formation accorde et AWS CLI révoque des commandes
- Permissions de Lake Formation

Lake Formation accorde et AWS CLI révoque des commandes

Chaque description d'autorisation présentée dans cette section inclut des exemples d'octroi de l'autorisation à l'aide d'une AWS CLI commande. Voici les synopsis de la Lake Formation grant-permissions et revoke-permissions AWS CLI des commandes.

```
grant-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

```
revoke-permissions
[--catalog-id <value>]
--principal <value>
--resource <value>
--permissions <value>
[--permissions-with-grant-option <value>]
[--cli-input-json <value>]
[--generate-cli-skeleton <value>]
```

Pour une description détaillée de ces commandes, voir grant-permissions et revoke-permissions dans la référence des commandes.AWS CLI Cette section fournit des informations supplémentaires sur `--principal` cette option.

La valeur de l'`--principal`option est l'une des suivantes :

- Amazon Resource Name (ARN) pour un utilisateur ou un rôle AWS Identity and Access Management (IAM)
- ARN pour un utilisateur ou un groupe qui s'authentifie via un fournisseur SAML, tel que Microsoft Active Directory Federation Service (AD FS)
- ARN pour un QuickSight utilisateur ou un groupe Amazon
- Pour les autorisations entre comptes, un identifiant de AWS compte, un identifiant d'organisation ou un identifiant d'unité organisationnelle

Vous trouverez ci-dessous la syntaxe et des exemples pour tous les `--principal` types.

Le principal est un utilisateur IAM

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1
```

Le directeur est un rôle IAM

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/workflowrole
```

Le principal est un utilisateur qui s'authentifie via un fournisseur SAML

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:user/<user-name>
```

Exemples :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
idp1:user/datalake_user1
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
AthenaLakeFormationOkta:user/athena-user@example.com
```

Le principal est un groupe qui s'authentifie via un fournisseur SAML

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:group/<group-name>
```

Exemples :

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
idp1:group/data-scientists
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
AthenaLakeFormationOkta:group/my-group
```

Le principal est un utilisateur QuickSight d'Amazon Enterprise Edition

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-id>:user/<namespace>/<user-name>
```

> **Note**
>
> Pour<namespace>, vous devez spécifier`default`.

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-east-1:111122223333:user/default/bi_user1
```

Principal est un groupe Amazon QuickSight Enterprise Edition

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-id>:group/<namespace>/<group-name>
```

> **Note**
>
> Pour<namespace>, vous devez spécifier`default`.

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-east-1:111122223333:group/default/data_scientists
```

Le principal est un AWS compte

Syntaxe :

```
--principal DataLakePrincipalIdentifier=<account-id>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=111122223333
```

Le principal est une organisation

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-id>:organization/<organization-id>
```

Exemple :

```
--principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-abcdefghijkl
```

Le directeur est une unité organisationnelle

Syntaxe :

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-id>:ou/<organization-id>/<organizational-unit-id>
```

Exemple :

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:ou/o-abcdefghijkl/ou-ab00-cdefghij
```

Le principal est un utilisateur ou un groupe d'identités IAM Identity Center

Exemple : utilisateur

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserID>
```

Exemple : Groupe :

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::group/<GroupID>
```

Le principal est un groupe IAM - **`IAMAllowedPrincipals`**

Lake Formation attribue des `Super` autorisations sur toutes les bases de données et tables du catalogue de données à un groupe appelé `IAMAllowedPrincipals` par défaut. Si cette autorisation de groupe existe sur une base de données ou une table, tous les principaux de votre compte auront accès à la ressource par le biais des politiques principales IAM pour. AWS

Glue Il assure une rétrocompatibilité lorsque vous commencez à utiliser les autorisations de Lake Formation pour sécuriser les ressources du catalogue de données qui étaient auparavant protégées par les politiques IAM pour AWS Glue.

Lorsque vous utilisez Lake Formation pour gérer les autorisations pour les ressources de votre catalogue de données, vous devez d'abord révoquer l'`IAMAllowedPrincipals`autorisation sur les ressources ou activer les principes et les ressources en mode d'accès hybride pour que les autorisations de Lake Formation fonctionnent.

Exemple :

```
--principal DataLakePrincipalIdentifier=IAM_Allowed_Principals
```

Le principal est un groupe IAM - **`ALLIAMPrincipals`**

Lorsque vous autorisez un `ALLIAMPrincipals` groupe sur une ressource de catalogue de données, chaque principal du compte a accès à la ressource de catalogue de données à l'aide des autorisations Lake Formation et des autorisations IAM.

Exemple :

```
--principal DataLakePrincipalIdentifier=123456789012:IAMPrincipals
```

Permissions de Lake Formation

Cette section contient les autorisations de Lake Formation disponibles que vous pouvez accorder aux directeurs.

`ALTER`

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
`ALTER`	`DATABASE`	`glue:UpdateDatabase`
`ALTER`	`TABLE`	`glue:UpdateTable`
`ALTER`	`LF-Tag`	`lakeformation:UpdateLFTag`

Un directeur disposant de cette autorisation peut modifier les métadonnées d'une base de données ou d'une table dans le catalogue de données. Pour les tables, vous pouvez modifier le schéma des colonnes et ajouter des paramètres de colonne. Vous ne pouvez pas modifier les colonnes des données sous-jacentes vers lesquelles pointe une table de métadonnées.

Si la propriété modifiée est un emplacement enregistré auprès d'Amazon Simple Storage Service (Amazon S3), le principal doit disposer des autorisations de localisation des données sur le nouvel emplacement.

Example

L'exemple suivant accorde l'`ALTER`autorisation à l'utilisateur `datalake_user1` sur la base de données `retail` dans le AWS compte 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "ALTER" --resource '{ "Database": {"Name":"retail"}}'
```

Example

L'exemple suivant accorde `ALTER` à l'utilisateur le droit d'`datalake_user1`accéder à la table `inventory` de la base de données`retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

CREATE_DATABASE

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
`CREATE_DATABASE`	Catalogue de données	`glue:CreateDatabase`

Un directeur disposant de cette autorisation peut créer une base de données de métadonnées ou un lien de ressource dans le catalogue de données. Le principal peut également créer des tables dans la base de données.

Example

L'exemple suivant accorde une autorisation CREATE_DATABASE à l'utilisateur datalake_user1 du AWS compte 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Lorsqu'un principal crée une base de données dans le catalogue de données, aucune autorisation sur les données sous-jacentes n'est accordée. Les autorisations de métadonnées supplémentaires suivantes sont accordées (ainsi que la possibilité d'accorder ces autorisations à d'autres personnes) :

- CREATE_TABLEdans la base de données
- Base de données ALTER
- Base de données DROP

Lors de la création d'une base de données, le principal peut éventuellement spécifier un emplacement Amazon S3. Selon que le principal dispose ou non d'autorisations de localisation des données, celles-ci CREATE_DATABASE peuvent ne pas être suffisantes pour créer des bases de données dans tous les cas. Il est important de garder à l'esprit les trois cas suivants.

Créer un cas d'utilisation de base de données	Autorisations nécessaires
La propriété de localisation n'est pas spécifiée.	CREATE_DATABASE est suffisant.
La propriété de localisation est spécifiée, et l'emplacement n'est pas géré par Lake Formation (il n'est pas enregistré).	CREATE_DATABASE est suffisant.
La propriété de localisation est spécifiée, et l'emplacement est géré par Lake Formation (est enregistré).	CREATE_DATABASE est obligatoire, ainsi que les autorisations de localisation des données à l'emplacement spécifié.

CREATE_TABLE

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
CREATE_TABLE	DATABASE	glue:CreateTable

Un directeur disposant de cette autorisation peut créer une table de métadonnées ou un lien de ressource dans le catalogue de données de la base de données spécifiée.

Example

L'exemple suivant accorde à l'utilisateur l'datalake_user1autorisation de créer des tables dans la retail base de données sous le AWS compte 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Lorsqu'un directeur crée une table dans le catalogue de données, toutes les autorisations de Lake Formation sur la table sont accordées au principal, avec la possibilité d'accorder ces autorisations à d'autres personnes.

Subventions entre comptes

Si un compte propriétaire de base de données accorde une autorisation CREATE_TABLE à un compte destinataire et qu'un utilisateur du compte destinataire crée avec succès une table dans la base de données du compte propriétaire, les règles suivantes s'appliquent :

- L'utilisateur et les administrateurs du lac de données du compte destinataire disposent de toutes les autorisations relatives à Lake Formation. Ils peuvent accorder des autorisations sur la table aux autres principaux de leur compte. Ils ne peuvent pas accorder d'autorisations aux principaux sur le compte du propriétaire ou sur tout autre compte.
- Les administrateurs du lac de données du compte propriétaire peuvent accorder des autorisations sur la table aux autres principaux de leur compte.

Autorisations de localisation des données

Lorsque vous tentez de créer une table pointant vers un emplacement Amazon S3, selon que vous disposez ou non des autorisations de localisation des données, ces `CREATE_TABLE` autorisations peuvent ne pas être suffisantes pour créer une table. Il est important de garder à l'esprit les trois cas suivants.

Création d'un cas d'utilisation d'une table	Autorisations nécessaires
L'emplacement spécifié n'est pas géré par Lake Formation (il n'est pas enregistré).	`CREATE_TABLE` est suffisant.
L'emplacement spécifié est géré par Lake Formation (il est enregistré), et la base de données qui le contient ne possède aucune propriété de localisation ou possède une propriété de localisation qui n'est pas un préfixe Amazon S3 de l'emplacement de la table.	`CREATE_TABLE` est obligatoire, ainsi que les autorisations de localisation des données à l'emplacement spécifié.
L'emplacement spécifié est géré par Lake Formation (il est enregistré), et la base de données contenant possède une propriété de localisation qui pointe vers un emplacement enregistré et qui est un préfixe Amazon S3 de l'emplacement de la table.	`CREATE_TABLE` est suffisant.

`DATA_LOCATION_ACCESS`

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
`DATA_LOCATION_ACCESS`	Emplacement Amazon S3	(Autorisations Amazon S3 sur l'emplacement, qui doivent être spécifiées par le rôle utilisé pour enregistrer l'emplacement.)

Il s'agit de la seule autorisation de localisation des données. Un principal disposant de cette autorisation peut créer une base de données ou une table de métadonnées pointant vers l'emplacement Amazon S3 spécifié. L'emplacement doit être enregistré. Un directeur qui dispose d'autorisations de localisation des données sur un emplacement possède également des autorisations de localisation sur les emplacements des enfants.

Example

L'exemple suivant accorde des autorisations de localisation des données `s3://products/retail` à l'utilisateur du AWS compte `datalake_user1` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::products/retail"}}'
```

`DATA_LOCATION_ACCESS`n'est pas nécessaire pour interroger ou mettre à jour les données sous-jacentes. Cette autorisation s'applique uniquement à la création de ressources de catalogue de données.

Pour plus d'informations sur les autorisations de localisation des données, consultezUnderlying data access control.

DELETE

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
`DELETE`	`TABLE`	(Aucune autorisation IAM supplémentaire n'est requise si l'emplacement est enregistré.)

Un principal disposant de cette autorisation peut supprimer les données sous-jacentes à l'emplacement Amazon S3 indiqué dans le tableau. Le directeur peut également consulter le tableau sur la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS GlueAPI.

Example

L'exemple suivant accorde l'`DELETE`autorisation à l'utilisateur `datalake_user1` sur la table de la base de données `inventory` dans le AWS compte `retail` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DELETE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Cette autorisation s'applique uniquement aux données d'Amazon S3, et non aux données d'autres magasins de données tels qu'Amazon Relational Database Service (Amazon RDS).

`DESCRIBE`

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin

`DESCRIBE`	Lien vers les ressources du tableau
	
	Lien vers les ressources de la base
		`glue:GetTable`
		
		`glue:GetDatabase`

`DESCRIBE`	`DATABASE`	`glue:GetDatabase`
`DESCRIBE`	`TABLE`	`glue:GetTable`

`DESCRIBE`	`LF-Tag`	`glue:GetTable`
		
		`glue:GetDatabase`
		
		`lakeformation:GetResourceLFTags`
		
		`lakeformation:ListLFTags`
		
		`lakeformation:GetLFTag`

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin

		`lakeformation:SearchTablesByLFTags`
		
		`lakeformation:SearchDatabasesByLFTags`

Un principal disposant de cette autorisation peut consulter la base de données, la table ou le lien de ressource spécifié. Aucune autre autorisation de catalogue de données n'est accordée implicitement, et aucune autorisation d'accès aux données n'est accordée implicitement. Les bases de données et les tables apparaissent dans les éditeurs de requêtes des services intégrés, mais aucune requête ne peut être effectuée à leur encontre à moins que d'autres autorisations de Lake Formation (par exemple,`SELECT`) ne soient accordées.

Par exemple, un utilisateur qui possède une `DESCRIBE` base de données peut voir la base de données et toutes les métadonnées de la base de données (description, emplacement, etc.). Toutefois, l'utilisateur ne peut pas savoir quelles tables contient la base de données et ne peut pas supprimer, modifier ou créer des tables dans la base de données. De même, un utilisateur qui possède `DESCRIBE` une table peut voir la table et ses métadonnées (description, schéma, emplacement, etc.), mais ne peut pas supprimer, modifier ou exécuter des requêtes sur la table.

Voici quelques règles supplémentaires pour `DESCRIBE` :

- Si un utilisateur dispose d'autres autorisations Lake Formation sur une base de données, une table ou un lien de ressource, elles `DESCRIBE` sont implicitement accordées.
- Si un utilisateur ne dispose `SELECT` que d'un sous-ensemble de colonnes pour un tableau (partiel`SELECT`), il est limité à l'affichage de ces colonnes.
- Vous ne pouvez pas accorder d'autorisation `DESCRIBE` à un utilisateur qui a effectué une sélection partielle sur une table. À l'inverse, vous ne pouvez pas spécifier de listes d'inclusion ou d'exclusion de colonnes pour les `DESCRIBE` tables accordées sur.

Example

L'exemple suivant accorde l'`DESCRIBE`autorisation à l'utilisateur `datalake_user1` sur le lien de ressource de la table dans la base de données `inventory-link` dans le AWS compte `retail` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory-link"}}'
```

DROP

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
`DROP`	`DATABASE`	`glue:DeleteDatabase`
`DROP`	`TABLE`	`glue:DeleteTable`
`DROP`	`LF-Tag`	`lakeformation:DeleteLFTag`

`DROP`	Lien vers les ressources de la base
	
	Lien vers les ressources du tableau
		`glue:DeleteDatabase`
		
		`glue:DeleteTable`

Un principal disposant de cette autorisation peut supprimer une base de données, une table ou un lien de ressource dans le catalogue de données. Vous ne pouvez pas accorder l'autorisation DROP sur une base de données à un compte ou à une organisation externe.

> **Warning**
>
> La suppression d'une base de données entraîne la suppression de toutes les tables de la base de données.

Example

L'exemple suivant accorde l'DROPautorisation à l'utilisateur `datalake_user1` sur la base de données `retail` dans le AWS compte 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Database": {"Name":"retail"}}'
```

Example

L'exemple suivant accorde des DROP autorisations à l'utilisateur `datalake_user1` sur la table de `inventory` la base de données`retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Example

L'exemple suivant accorde DROP à l'utilisateur `datalake_user1` sur la table un lien de ressource `inventory-link` dans la base de données`retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-
link"}}'
```

INSERT

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
INSERT	TABLE	(Aucune autorisation IAM supplémentaire n'est requise si l'emplacement est enregistré.)

Un principal disposant de cette autorisation peut insérer, mettre à jour et lire les données sous-jacentes à l'emplacement Amazon S3 indiqué dans le tableau. Le directeur peut également consulter le tableau dans la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS GlueAPI.

Example

L'exemple suivant accorde l'`INSERT`autorisation à l'utilisateur `datalake_user1` sur la table de la base de données `inventory` dans le AWS compte `retail` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "INSERT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Cette autorisation s'applique uniquement aux données d'Amazon S3, et non aux données d'autres magasins de données tels qu'Amazon RDS.

SELECT

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
`SELECT`	• `TABLE`	(Aucune autorisation IAM supplémentaire n'est requise si l'emplacement est enregistré.)

Un directeur disposant de cette autorisation peut consulter une table dans le catalogue de données et interroger les données sous-jacentes dans Amazon S3 à l'emplacement spécifié par la table. Le directeur peut consulter le tableau dans la console Lake Formation et récupérer des informations sur le tableau à l'aide de l'AWS GlueAPI. Si le filtrage des colonnes a été appliqué lorsque cette autorisation a été accordée, le principal peut consulter les métadonnées uniquement pour les colonnes incluses et ne peut interroger les données qu'à partir des colonnes incluses.

> **Note**
>
> Il est de la responsabilité du service d'analyse intégré d'appliquer le filtrage des colonnes lors du traitement d'une requête.

Example

L'exemple suivant accorde l'`SELECT`autorisation à l'utilisateur `datalake_user1` sur la table de la base de données `inventory` dans le AWS compte `retail` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Cette autorisation s'applique uniquement aux données d'Amazon S3, et non aux données d'autres magasins de données tels qu'Amazon RDS.

Vous pouvez filtrer (restreindre l'accès à) des colonnes spécifiques à l'aide d'une liste d'inclusion ou d'exclusion facultative. Une liste d'inclusion indique les colonnes accessibles. Une liste d'exclusion indique les colonnes qui ne sont pas accessibles. En l'absence de liste d'inclusion ou d'exclusion, toutes les colonnes du tableau sont accessibles.

Les résultats de `glue:GetTable` renvoient uniquement les colonnes que l'appelant est autorisé à consulter. Les services intégrés tels qu'Amazon Athena et Amazon Redshift respectent les listes d'inclusion et d'exclusion des colonnes.

Example

L'exemple suivant octroie une `SELECT` autorisation à l'utilisateur figurant `datalake_user1` sur la table `inventory` à l'aide d'une liste d'inclusion.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail",
 "Name":"inventory", "ColumnNames": ["prodcode","location","period","withdrawals"]}}'
```

Example

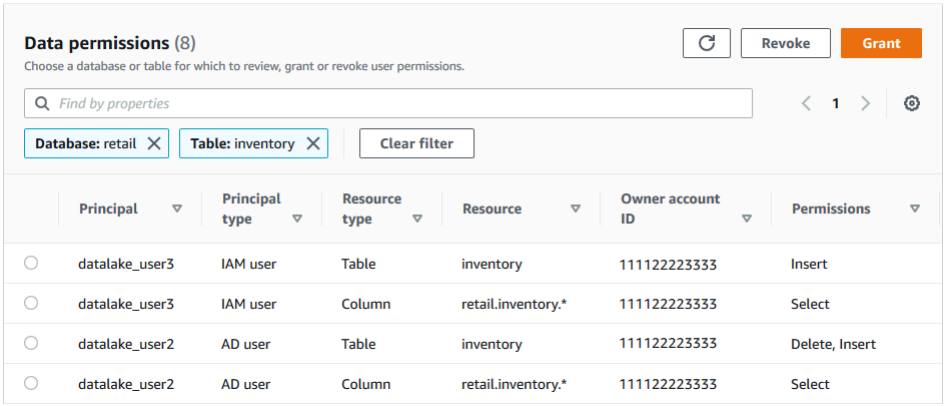
L'exemple suivant donne des autorisations SELECT sur la `inventory` table à l'aide d'une liste d'exclusion.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail",
 "Name":"inventory", "ColumnWildcard": {"ExcludedColumnNames": ["intkey",
 "prodcode"]}}}'
```

Les restrictions suivantes s'appliquent à l'SELECTautorisation :

- Lors de l'octroiSELECT, vous ne pouvez pas inclure l'option d'octroi si le filtrage des colonnes est appliqué.
- Vous ne pouvez pas restreindre le contrôle d'accès aux colonnes qui sont des clés de partition.
- Un principal disposant de l'SELECTautorisation sur un sous-ensemble de colonnes d'une table ne peut pas obtenir l'INSERTautorisation ALTERDROP,DELETE, ou sur cette table. De même, un principal disposant de l'INSERTautorisation ALTER DROPDELETE,, ou sur une table ne peut pas SELECT obtenir l'autorisation de filtrage des colonnes.

L'SELECTautorisation apparaît toujours sur la page Autorisations relatives aux données de la console Lake Formation sous forme de ligne séparée. L'image suivante montre que cela SELECT est accordé aux utilisateurs `datalake_user2` et `datalake_user3` sur toutes les colonnes du `inventory` tableau.



Super

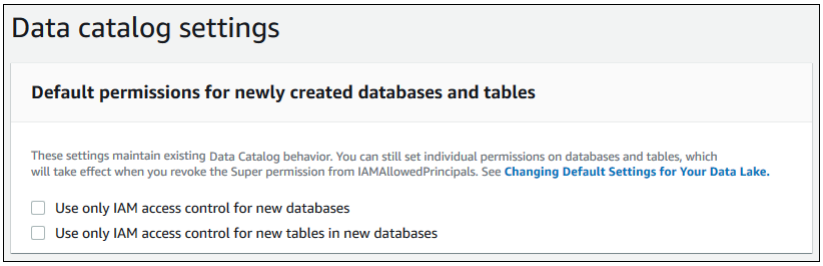
Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin
Super	DATABASE	glue:*Database*
Super	TABLE	glue:*Table*, glue:*Partition*

Cette autorisation permet au principal d'effectuer toutes les opérations de Lake Formation prises en charge sur la base de données ou la table. Vous ne pouvez pas accorder d'autorisation Super sur une base de données à un compte externe.

Cette autorisation peut coexister avec les autres autorisations de Lake Formation. Par exemple, vous pouvez accorder les INSERT autorisations SuperSELECT, et sur une table de métadonnées. Le principal peut ensuite effectuer toutes les opérations prises en charge sur la table. Lorsque vous révoquezSuper, les INSERT autorisations SELECT et sont conservées et le principal ne peut effectuer que des opérations de sélection et d'insertion.

Au lieu de l'octroyer Super à un directeur individuel, vous pouvez l'accorder au groupeIAMAllowedPrincipals. Le IAMAllowedPrincipals groupe est créé automatiquement et inclut tous les utilisateurs et rôles IAM autorisés à accéder aux ressources de votre catalogue de données par vos politiques IAM. Lorsqu'il Super est accordé IAMAllowedPrincipals pour une ressource de catalogue de données, l'accès à la ressource est effectivement contrôlé uniquement par les politiques IAM.

Vous pouvez Super obtenir l'autorisation d'accéder automatiquement IAMAllowedPrincipals aux nouvelles ressources du catalogue en tirant parti des options de la page Paramètres de la console Lake Formation.



- Pour l'accorder `Super` à `IAMAllowedPrincipals` toutes les nouvelles bases de données, sélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles bases de données.
- `IAMAllowedPrincipals`Pour l'accorder `Super` à toutes les nouvelles tables des nouvelles bases de données, sélectionnez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables des nouvelles bases de données.

> **Note**
>
> Cette option entraîne la sélection par défaut de la case Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données dans la boîte de dialogue Créer une base de données. Il ne fait rien de plus que cela. C'est la case à cocher de la boîte de dialogue Créer une base de données qui permet d'accorder `Super` à`IAMAllowedPrincipals`.

Ces options de la page Paramètres sont activées par défaut. Pour plus d'informations, consultez les ressources suivantes :

- the section called "Modification des paramètres par défaut de votre lac de données"
- the section called "Mise à niveau des autorisations de AWS Glue données pour le modèle Lake Formation"

ASSOCIATE

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin

`ASSOCIATE`	`LF-Tag`	`glue:GetDatabase`
		`glue:GetTable`
		`lakeformation:AddLFTagsToResource"`
		`lakeformation:RemoveLFTagsFromResource"`

Autorisation	Accordé sur cette ressource	Le bénéficiaire a également besoin

		`lakeformation:GetResourceLFTags`
		`lakeformation:ListLFTags`
		`lakeformation:GetLFTag`
		`lakeformation:SearchTablesByLFTags`
		`lakeformation:SearchDatabasesByLFTags`

Un directeur disposant de cette autorisation sur une balise LF peut attribuer la balise LF à une ressource de catalogue de données. Accorder `ASSOCIATE` implicitement des subventions`DESCRIBE`.

Example

Cet exemple accorde à l'utilisateur `datalake_user1` l'`ASSOCIATE`autorisation d'utiliser le tag LF avec la clé. `module` Il autorise l'affichage et l'attribution de toutes les valeurs pour cette clé, comme indiqué par l'astérisque (*).

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Intégration d'IAM Identity Center

Vous pouvez ainsi vous connecter à des fournisseurs d'identité (IdPs) et gérer de manière centralisée l'accès des utilisateurs et des groupes à travers les services AWS d'analyse. AWS IAM Identity Center Vous pouvez intégrer des fournisseurs d'identité tels qu'Okta, Ping et Microsoft Entra ID (anciennement Azure Active Directory) à IAM Identity Center pour que les utilisateurs de votre

organisation puissent accéder aux données via une expérience d'authentification unique. IAM Identity Center prend également en charge la connexion d'autres fournisseurs d'identité tiers.

Pour plus d'informations, consultez la section Fournisseurs d'identité pris en charge dans le Guide de AWS IAM Identity Center l'utilisateur.

Vous pouvez AWS Lake Formation le configurer en tant qu'application activée dans IAM Identity Center, et les administrateurs de data lake peuvent accorder des autorisations détaillées aux utilisateurs et aux groupes autorisés sur les ressources. AWS Glue Data Catalog

Les utilisateurs de votre organisation peuvent se connecter à n'importe quelle application compatible avec Identity Center en utilisant le fournisseur d'identité de votre organisation et interroger des ensembles de données en appliquant les autorisations de Lake Formation. Grâce à cette intégration, vous pouvez gérer l'accès aux AWS services sans créer plusieurs rôles IAM.

> **Note**
>
> La propagation fiable des identités permet aux utilisateurs et aux groupes auxquels ils appartiennent déjà d'accéder aux données de tous les services AWS d'analyse. Grâce à la propagation sécurisée des identités, un utilisateur peut se connecter à une application, qui peut transmettre l'identité de l'utilisateur dans les demandes d'accès aux données des AWS services. Il n'est pas nécessaire d'effectuer des configurations de fournisseur d'identité ou de rôles IAM spécifiques à un service. Pour plus d'informations, consultez la section Propagation d'identités fiables entre les applications dans le Guide de AWS IAM Identity Center l'utilisateur.

Pour connaître les limitations, veuillez consulter Limites de l'intégration d'IAM Identity Center.

Rubriques

- Prérequis
- Connecter Lake Formation à l'IAM Identity Center
- Mise à jour d'une intégration à IAM Identity Center
- Suppression d'une connexion Lake Formation avec IAM Identity Center
- Octroi d'autorisations aux utilisateurs et aux groupes

Prérequis

Les conditions préalables à l'intégration d'IAM Identity Center à Lake Formation sont les suivantes.

1. Activer le centre d'identité IAM : l'activation d'IAM Identity Center est une condition préalable à la prise en charge de l'authentification et de la propagation des identités.
2. Choisissez votre source d'identité : après avoir activé IAM Identity Center, vous devez disposer d'un fournisseur d'identité pour gérer les utilisateurs et les groupes. Vous pouvez utiliser le répertoire intégré Identity Center comme source d'identité ou utiliser un IdP externe, tel que Microsoft Entra ID ou Okta.

   Pour plus d'informations, voir Gérer votre source d'identité et Se connecter à un fournisseur d'identité externe dans le guide de AWS IAM Identity Center l'utilisateur.
3. Créer un rôle IAM — Le rôle qui crée la connexion IAM Identity Center nécessite des autorisations pour créer et modifier la configuration de l'application dans Lake Formation et IAM Identity Center, conformément à la politique en ligne suivante.

   Vous devez ajouter des autorisations conformément aux meilleures pratiques IAM. Les autorisations spécifiques sont détaillées dans les procédures qui suivent. Pour plus d'informations, consultez Getting Started with IAM Identity Center.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:CreateLakeFormationIdentityCenterConfiguration",
                "sso:CreateApplication",
                "sso:PutApplicationAssignmentConfiguration",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope",
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

Si vous partagez des ressources du catalogue de données avec des organisations externes Comptes AWS ou avec des organisations, vous devez disposer des autorisations AWS Resource Access Manager (AWS RAM) pour créer des partages de ressources. Pour plus d'informations sur les autorisations requises pour partager des ressources, consultez la section Conditions requises pour le partage de données entre comptes.

Les politiques intégrées suivantes contiennent les autorisations spécifiques requises pour afficher, mettre à jour et supprimer les propriétés de l'intégration de Lake Formation à IAM Identity Center.

- Utilisez la politique intégrée suivante pour autoriser un rôle IAM à visualiser une intégration de Lake Formation avec IAM Identity Center.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Utilisez la politique intégrée suivante pour autoriser un rôle IAM à mettre à jour une intégration de Lake Formation avec IAM Identity Center. La politique inclut également les autorisations facultatives requises pour partager des ressources avec des comptes externes.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
```

```
                "lakeformation:UpdateLakeFormationIdentityCenterConfiguration",
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication",
                "sso:UpdateApplication",
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Utilisez la politique intégrée suivante pour autoriser un rôle IAM à supprimer une intégration de Lake Formation avec IAM Identity Center.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DeleteLakeFormationIdentityCenterConfiguration",
                "sso:DeleteApplication",
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Pour les autorisations IAM requises pour accorder ou révoquer des autorisations de lac de données pour les utilisateurs et les groupes IAM Identity Center, consultez. Autorisations IAM requises pour accorder ou révoquer les autorisations de Lake Formation

Description des autorisations

- `lakeformation:CreateLakeFormationIdentityCenterConfiguration`— Crée la configuration iDC de Lake Formation.

- `lakeformation:DescribeLakeFormationIdentityCenterConfiguration`— Décrit une configuration iDC existante.
- `lakeformation:DeleteLakeFormationIdentityCenterConfiguration`— Permet de supprimer une configuration iDC de Lake Formation existante.
- `lakeformation:UpdateLakeFormationIdentityCenterConfiguration`— Utilisé pour modifier la configuration d'une Lake Formation existante.
- `sso:CreateApplication` : sert à créer une application IAM Identity Center.
- `sso:DeleteApplication` : sert à supprimer une application IAM Identity Center.
- `sso:UpdateApplication` : sert à mettre à jour une application IAM Identity Center.
- `sso:PutApplicationGrant` : sert à modifier les informations relatives à l'émetteur de jetons approuvé.
- `sso:PutApplicationAuthenticationMethod`— Accorde un accès d'authentification à Lake Formation.
- `sso:GetApplicationGrant` : sert à répertorier les informations relatives à l'émetteur de jetons approuvé.
- `sso:DeleteApplicationGrant` : supprime les informations relatives à l'émetteur de jetons approuvé.
- `sso:PutApplicationAccessScope`— Ajoute ou met à jour la liste des cibles autorisées pour une étendue d'accès à l'IAM Identity Center pour une application.
- `sso:PutApplicationAssignmentConfiguration`— Utilisé pour configurer la manière dont les utilisateurs accèdent à une application.

Connecter Lake Formation à l'IAM Identity Center

Avant de pouvoir utiliser IAM Identity Center pour gérer les identités afin d'accorder l'accès aux ressources du catalogue de données à l'aide de Lake Formation, vous devez suivre les étapes suivantes. Vous pouvez créer l'intégration IAM Identity Center à l'aide de la console Lake Formation ou AWS CLI.

AWS Management Console

Pour connecter Lake Formation à l'IAM Identity Center

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.

2. Dans le volet de navigation de gauche, sélectionnez Intégration à IAM Identity Center.

Create IAM Identity Center Integration

Enable IAM Identify Center and then create Lake Formation - IAM Identity Center integration to manage identities from IAM Identity Center (external IdPs like Azure AD or Okta Universal Directory). Learn more

▼ **How it works**

Enable IAM Identity Center
Enable IAM Identity Center for your account or organization and select an identity provider.

Create Lake Formation integration
Integrate Lake Formation with IAM Identity Center to permit Lake Formation to access users from your selected identity provider.

Grant permissions
Grant permissions to users on Data Catalog databases and tables using fine-grained Lake Formation permissions.

Connect Lake Formation to IAM Identity Center

Connect to organization instance of IAM Identity Center

Manage access to Lake Formation by assigning users and groups from the Identity Center directory for your organization. Learn more

Recommended

Connect to account instance of IAM Identity Center

Manage access to Lake Formation by assigning existing or creating dedicated users and groups from your Identity Center directory. Learn more

instance of IAM Identity Center
Manage access to Lake Formation by assigning users and groups from your Identity Center directory.
arn:aws:sso:::instance/ssoins-6987513bf5410c2f

Add AWS account and organization IDs

Add AWS accounts and organizations whose users need access to Lake Formation managed resources.

AWS Accounts and AWS organizations
Enter one or more AWS account IDs and AWS organization IDs. Press Enter after each ID.

Choose AWS account, AWS organization ID

▶ **Lake Formation application integration - *optional***
Add application IDs that can access S3 data locations registered with Lake Formation on behalf of the user.

After this step, you can't edit the connection. You can edit AWS accounts, organizations, and applications. If you want to modify the connection, delete it and create a new connection.

3. (Facultatif) Entrez un ou plusieurs Compte AWS identifiants, identifiants d'organisation et/ou identifiants d'unité organisationnelle valides pour permettre aux comptes externes d'accéder aux ressources du catalogue de données. Lorsque des utilisateurs ou des groupes d'IAM Identity Center tentent d'accéder aux ressources du catalogue de données géré par Lake Formation, Lake Formation assume un rôle IAM pour autoriser l'accès aux métadonnées. Si le rôle IAM appartient à un compte externe qui n'a pas de politique de AWS Glue ressources ni de partage de AWS RAM ressources, les utilisateurs et les groupes de l'IAM Identity Center ne pourront pas accéder à la ressource même s'ils disposent des autorisations de Lake Formation.

   Lake Formation utilise le service AWS Resource Access Manager (AWS RAM) pour partager la ressource avec des comptes et des organisations externes. AWS RAM envoie une invitation au compte du bénéficiaire pour qu'il accepte ou rejette le partage des ressources.

   Pour plus d'informations, consultez Acceptation d'une invitation de partage de ressources de AWS RAM.

   > **Note**
   >
   > Lake Formation permet aux rôles IAM issus de comptes externes d'agir en tant que rôles de support au nom des utilisateurs et des groupes d'IAM Identity Center pour accéder aux ressources du catalogue de données, mais les autorisations ne peuvent être accordées que sur les ressources du catalogue de données du compte propriétaire. Si vous essayez d'accorder des autorisations aux utilisateurs et aux groupes d'IAM Identity Center sur les ressources du catalogue de données d'un compte externe, Lake Formation génère le message d'erreur suivant : « Les autorisations entre comptes ne sont pas prises en charge pour le principal ».

4. (Facultatif) Sur l'écran d'intégration Create Lake Formation, spécifiez les ARN des applications tierces qui peuvent accéder aux données des sites Amazon S3 enregistrés auprès de Lake Formation. Lake Formation fournit des informations d'identification temporaires limitées sous forme de AWS STS jetons aux sites Amazon S3 enregistrés en fonction des autorisations effectives, afin que les applications autorisées puissent accéder aux données pour le compte des utilisateurs.
5. Sélectionnez Submit (Envoyer).

   Une fois que l'administrateur de Lake Formation a terminé les étapes et créé l'intégration, les propriétés de l'IAM Identity Center apparaissent dans la console Lake Formation. L'exécution

de ces tâches fait de Lake Formation une application compatible avec IAM Identity Center. Les propriétés de la console incluent l'état de l'intégration. L'état de l'intégration indique `Success` quand elle est terminée. Ce statut indique si la configuration du centre d'identité IAM est terminée.

AWS CLI

- L'exemple suivant montre comment créer une intégration de Lake Formation avec IAM Identity Center. Vous pouvez également spécifier le `Status` (`ENABLED`,`DISABLED`) des applications.

```
aws lakeformation create-lake-formation-identity-center-configuration \
    --catalog-id <123456789012> \
    --instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
    --share-recipients '[{"DataLakePrincipalIdentifier": "<123456789012>"},
                         {"DataLakePrincipalIdentifier": "<555555555555>"}]' \
    --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"],
 "Status": "ENABLED"}'
```

- L'exemple suivant montre comment visualiser une intégration de Lake Formation avec IAM Identity Center.

```
aws lakeformation describe-lake-formation-identity-center-configuration
 --catalog-id <123456789012>
```

Mise à jour d'une intégration à IAM Identity Center

Après avoir créé la connexion, vous pouvez ajouter des applications tierces pour l'intégration d'IAM Identity Center afin de les intégrer à Lake Formation et d'accéder aux données Amazon S3 pour le compte des utilisateurs. Vous pouvez également supprimer des applications existantes de l'intégration d'IAM Identity Center. Vous pouvez ajouter ou supprimer des applications à l'aide de la console Lake Formation et en utilisant UpdateLakeFormationIdentityCenterConfigurationOperation.
AWS CLI

> **Note**
>
> Après avoir créé l'intégration IAM Identity Center, vous ne pouvez pas mettre à jour l'instanceARN.

AWS Management Console

Pour mettre à jour une connexion IAM Identity Center existante avec Lake Formation

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Dans le volet de navigation de gauche, sélectionnez Intégration à IAM Identity Center.
3. Sélectionnez Ajouter sur la page d'intégration d'IAM Identity Center.
4. Entrez un ou plusieurs Compte AWS identifiants, identifiants d'organisation et/ou identifiants d'unité organisationnelle valides pour permettre aux comptes externes d'accéder aux ressources du catalogue de données.
5. Sur l'écran Ajouter des applications, entrez les identifiants des applications tierces que vous souhaitez intégrer à Lake Formation.
6. Sélectionnez Ajouter.

AWS CLI

Vous pouvez ajouter ou supprimer des applications tierces pour l'intégration d'IAM Identity Center en exécutant la AWS CLI commande suivante. Lorsque vous définissez l'état du filtrage externe surENABLED, cela permet à l'IAM Identity Center de fournir une gestion des identités permettant aux applications tierces d'accéder aux données gérées par Lake Formation. Vous pouvez également activer ou désactiver l'intégration d'IAM Identity Center en définissant le statut de l'application.

```
aws lakeformation update-lake-formation-identity-center-configuration \
 --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"], "Status":
 "ENABLED"}'\
 --share-recipients '[{"DataLakePrincipalIdentifier": "<444455556666>"}
                     {"DataLakePrincipalIdentifier": "<777788889999>"}]' \
 --application-status ENABLED
```

Suppression d'une connexion Lake Formation avec IAM Identity Center

Si vous souhaitez supprimer une intégration IAM Identity Center existante, vous pouvez le faire à l'aide de la console Lake Formation ou d'une [DeleteLakeFormationIdentityCenterConfiguration](#)opération. AWS CLI

AWS Management Console

Pour supprimer une connexion IAM Identity Center existante avec Lake Formation

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Dans le volet de navigation de gauche, sélectionnez Intégration à IAM Identity Center.
3. Sélectionnez Supprimer sur la page d'intégration d'IAM Identity Center.
4. Sur l'écran Confirmer l'intégration, confirmez l'action, puis sélectionnez Supprimer.

AWS CLI

Vous pouvez supprimer l'intégration d'IAM Identity Center en exécutant la AWS CLI commande suivante.

```
aws lakeformation delete-lake-formation-identity-center-configuration \
    --catalog-id <123456789012>
```

Octroi d'autorisations aux utilisateurs et aux groupes

L'administrateur de votre lac de données peut accorder des autorisations aux utilisateurs et aux groupes d'IAM Identity Center sur les ressources du catalogue de données (bases de données, tables et vues) afin de faciliter l'accès aux données. Pour accorder ou révoquer des autorisations de lac de données, le concédant a besoin d'autorisations pour les actions suivantes de l'IAM Identity Center.

- DescribeUser
- DescribeGroup
- DescribeInstance

Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS CLI.

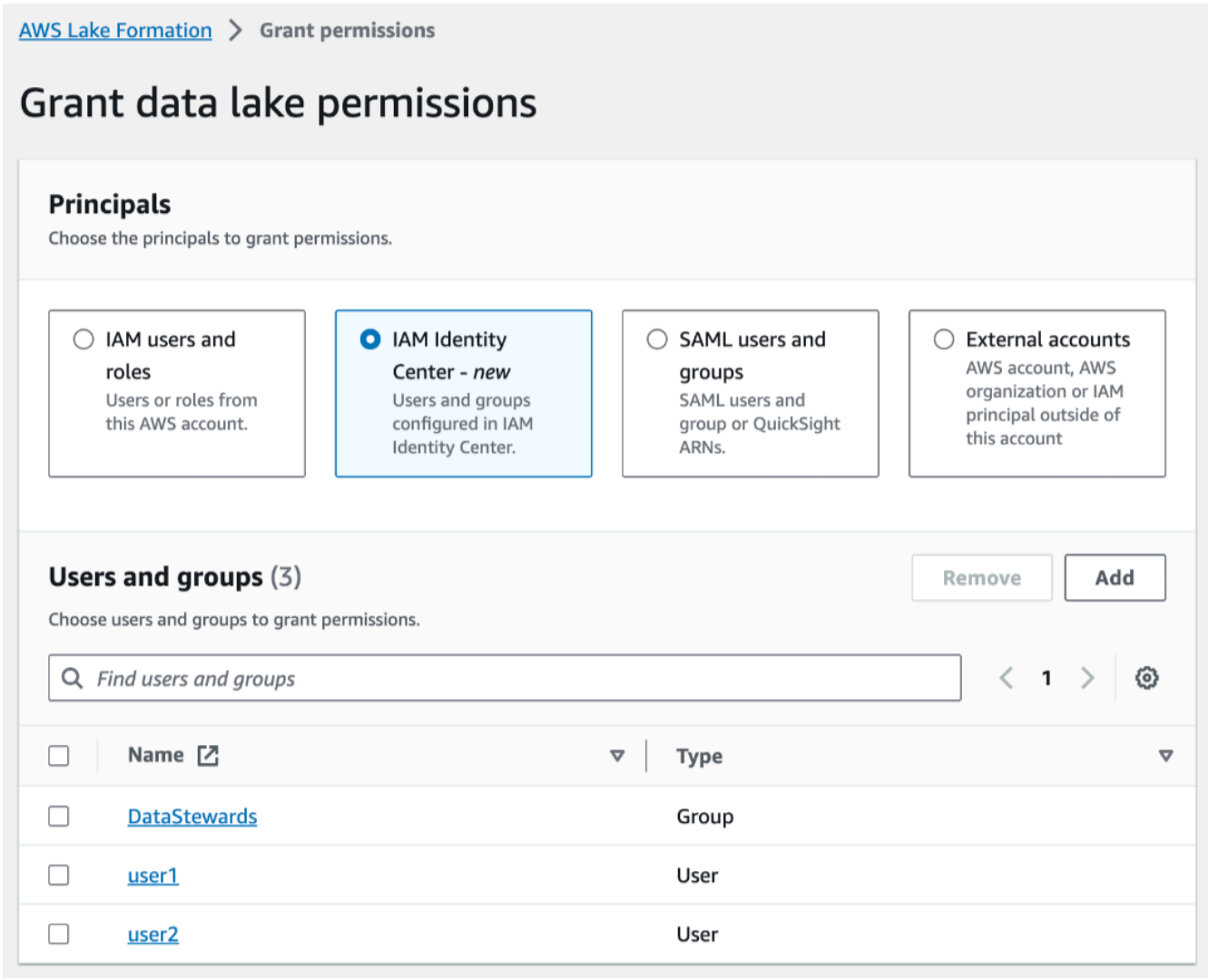
Pour plus d'informations sur l'octroi d'autorisations, consultezthe section called “Octroi et révocation des autorisations du catalogue de données”.

> **Note**
>
> Vous ne pouvez accorder des autorisations que sur les ressources de votre compte. Pour attribuer des autorisations en cascade aux utilisateurs et aux groupes sur les ressources partagées avec vous, vous devez utiliser AWS RAM des partages de ressources.

AWS Management Console

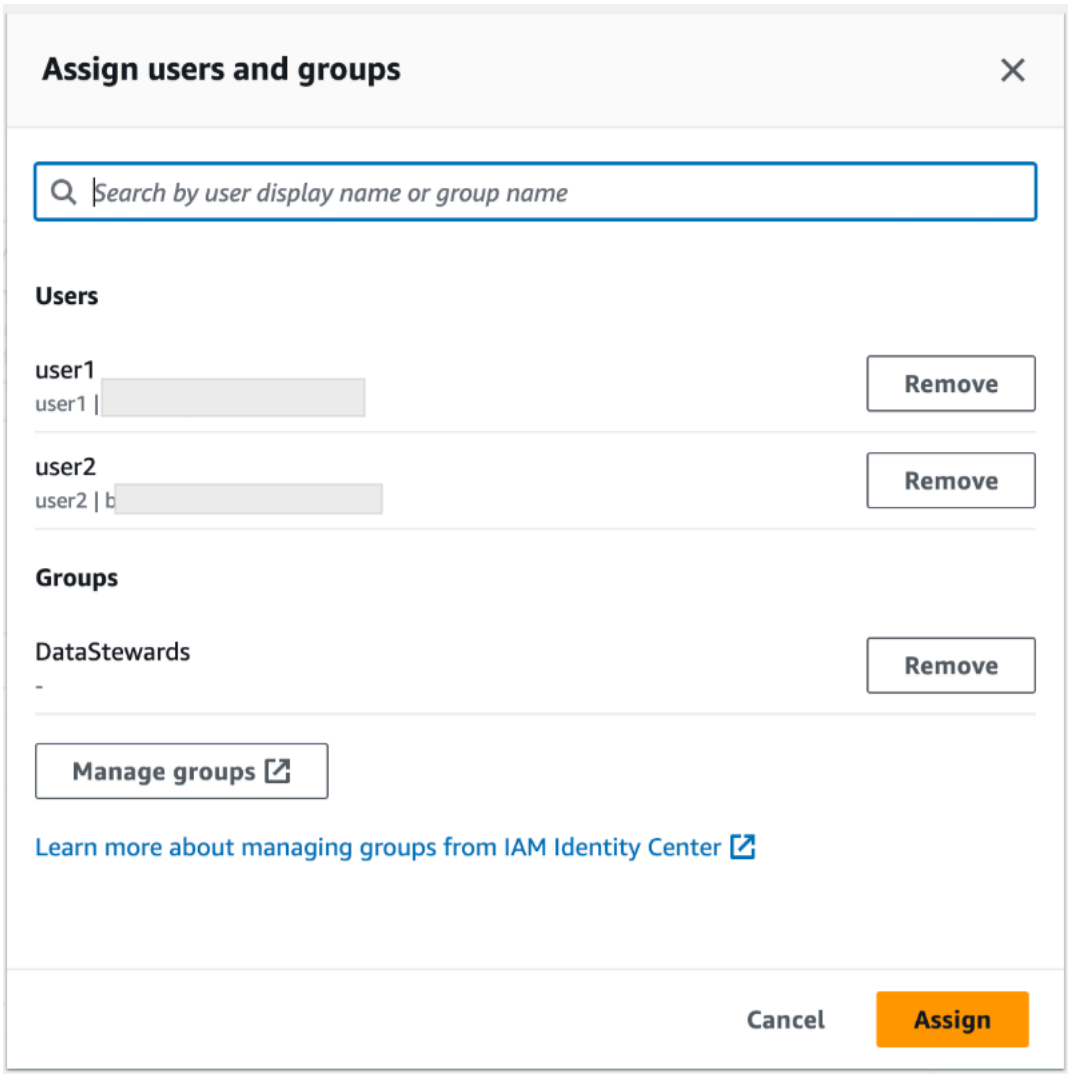
Pour accorder des autorisations aux utilisateurs et aux groupes

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Sélectionnez Autorisations du lac de données sous Autorisations dans la console Lake Formation.
3. Sélectionnez Grant.
4. Sur la page Accorder les autorisations du lac de données, sélectionnez Utilisateurs et groupes SSM.
5. Sélectionnez Ajouter pour choisir les utilisateurs et les groupes auxquels accorder des autorisations.



6. Sur l'écran Attribuer des utilisateurs et des groupes, choisissez les utilisateurs et/ou les groupes auxquels vous souhaitez accorder des autorisations.

   Sélectionnez Attribuer.



7. Choisissez ensuite la méthode d'octroi des autorisations.

   Pour obtenir des instructions sur l'octroi d'autorisations à l'aide de la méthode des ressources nommées, consultezOctroi d'autorisations de data lake à l'aide de la méthode de ressource nommée.

   Pour obtenir des instructions sur l'octroi d'autorisations à l'aide de balises LF, consultez. Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC

8. Choisissez les ressources du catalogue de données pour lesquelles vous souhaitez accorder des autorisations.
9. Choisissez les autorisations de catalogue de données à accorder.
10. Sélectionnez Grant.

AWS CLI

L'exemple suivant montre comment accorder l'`SELECT`autorisation utilisateur d'IAM Identity Center sur une table.

```
aws lakeformation grant-permissions \
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserId> \
--permissions "SELECT" \
--resource '{ "Table": { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Pour effectuer une récupération `UserId` depuis IAM Identity Center, voir le GetUserIdfonctionnement dans le manuel de référence de l'API IAM Identity Center.

Ajouter un emplacement Amazon S3 à votre lac de données

Pour ajouter un emplacement Amazon Simple Storage Service (Amazon S3) comme espace de stockage dans votre lac de données, vous devez enregistrer l'emplacement auprès de. AWS Lake Formation Vous pouvez ensuite utiliser les autorisations de Lake Formation pour un contrôle d'accès précis aux AWS Glue Data Catalog objets pointant vers cet emplacement, ainsi qu'aux données sous-jacentes de cet emplacement.

Lake Formation permet également d'enregistrer un emplacement de données en mode d'accès hybride et vous offre la flexibilité d'activer de manière sélective les autorisations Lake Formation pour les bases de données et les tables de votre catalogue de données. Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

Pour plus d'informations sur la configuration du mode d'accès hybride, voir Mode d'accès hybride

Lorsque vous enregistrez un emplacement, ce chemin Amazon S3 et tous les dossiers situés sous ce chemin sont enregistrés.

Supposons, par exemple, que vous disposiez d'une organisation des chemins Amazon S3 telle que la suivante :

`/mybucket/accounting/sales/`

Si vous vous inscrivez`S3://mybucket/accounting`, le `sales` dossier est également enregistré et placé sous Gestion de Lake Formation.

Pour plus d'informations sur l'enregistrement des points de vente, consultezUnderlying data access control.

> **Note**
>
> Les autorisations Lake Formation sont recommandées pour les données structurées (organisées dans des tableaux avec des lignes et des colonnes). Si vos données contiennent des données non structurées basées sur des objets, pensez à utiliser l'autorisation IAM pour Amazon S3 afin de gérer l'accès aux données.

Rubriques

- Exigences relatives aux rôles utilisés pour enregistrer des sites
- Enregistrement d'un emplacement Amazon S3
- Enregistrement d'un emplacement Amazon S3 chiffré
- Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte
- Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes
- Annulation de l'enregistrement d'un site Amazon S3

Exigences relatives aux rôles utilisés pour enregistrer des sites

Vous devez spécifier un rôle AWS Identity and Access Management (IAM) lorsque vous enregistrez un emplacement Amazon Simple Storage Service (Amazon S3). AWS Lake Formation assume ce rôle lors de l'accès aux données à cet emplacement.

Vous pouvez utiliser l'un des types de rôles suivants pour enregistrer un emplacement :

- Le rôle lié au service Lake Formation. Ce rôle accorde les autorisations requises sur l'emplacement. L'utilisation de ce rôle est le moyen le plus simple d'enregistrer l'emplacement. Pour plus d’informations, consultez Utilisation de rôles liés à un service pour Lake Formation.
- Un rôle défini par l'utilisateur. Utilisez un rôle défini par l'utilisateur lorsque vous devez accorder plus d'autorisations que le rôle lié à un service n'en fournit.

  Vous devez utiliser un rôle défini par l'utilisateur dans les cas suivants :

- Lors de l'enregistrement d'une position dans un autre compte.

  Pour plus d’informations, consultez the section called “Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte” et the section called “Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes”.

- Si vous avez utilisé une AWS clé CMK gérée (`aws/s3`) pour chiffrer l'emplacement Amazon S3.

  Pour plus d’informations, consultez Enregistrement d'un emplacement Amazon S3 chiffré.

- Si vous prévoyez d'accéder à l'emplacement via Amazon EMR.

  Si vous avez déjà enregistré un emplacement avec le rôle lié au service et que vous souhaitez commencer à y accéder avec Amazon EMR, vous devez annuler l'enregistrement du point de vente et le réenregistrer avec un rôle défini par l'utilisateur. Pour plus d’informations, consultez the section called “Annulation de l'enregistrement d'un site Amazon S3”.

Utilisation de rôles liés à un service pour Lake Formation

AWS Lake Formation utilise un rôle AWS Identity and Access Management lié à un service (IAM). Un rôle lié à un service est un type unique de rôle IAM directement lié à Lake Formation. Le rôle lié au service est prédéfini par Lake Formation et inclut toutes les autorisations dont le service a besoin pour appeler d'autres AWS services en votre nom.

Un rôle lié à un service facilite la configuration de Lake Formation, car il n'est pas nécessaire de créer un rôle et d'ajouter manuellement les autorisations nécessaires. Lake Formation définit les autorisations associées à son rôle lié aux services et, sauf indication contraire, seule Lake Formation peut assumer ses rôles. Les autorisations définies comprennent la politique d’approbation et la politique d’autorisation. De plus, cette politique d’autorisation ne peut pas être attachée à une autre entité IAM.

Ce rôle lié à un service fait confiance aux services suivants pour assumer le rôle :

- `lakeformation.amazonaws.com`

Lorsque vous utilisez un rôle lié à un service dans le compte A pour enregistrer un emplacement Amazon S3 appartenant au compte B, la politique de compartiment Amazon S3 (une politique basée sur les ressources) du compte B doit accorder des autorisations d'accès au rôle lié au service dans le compte A.

> **Note**
>
> Les politiques de contrôle des services (SCP) n'affectent pas les rôles liés aux services. Pour plus d'informations, consultez la section Politiques de contrôle des services (SCP) dans le guide de l'AWS Organizations utilisateur.

Autorisations de rôle liées à un service pour Lake Formation

Lake Formation utilise le rôle lié au service nommé. `AWSServiceRoleForLakeFormationDataAccess` Ce rôle fournit un ensemble d'autorisations Amazon Simple Storage Service (Amazon S3) qui permettent au service intégré Lake Formation (Amazon Athena tel que) d'accéder aux emplacements enregistrés. Lorsque vous enregistrez l'emplacement d'un lac de données, vous devez fournir un rôle disposant des autorisations de lecture/écriture Amazon S3 requises pour cet emplacement. Au lieu de créer un rôle avec les autorisations Amazon S3 requises, vous pouvez utiliser ce rôle lié à un service.

La première fois que vous nommez le rôle lié au service comme le rôle avec lequel enregistrer un chemin, le rôle lié au service et une nouvelle politique IAM sont créés en votre nom. Lake Formation ajoute le chemin à la politique en ligne et l'associe au rôle lié au service. Lorsque vous enregistrez les chemins suivants avec le rôle lié au service, Lake Formation ajoute le chemin à la politique existante.

Lorsque vous êtes connecté en tant qu'administrateur du lac de données, enregistrez l'emplacement d'un lac de données. Ensuite, dans la console IAM, recherchez le rôle `AWSServiceRoleForLakeFormationDataAccess` et consultez les politiques qui lui sont associées.

Par exemple, une fois que vous avez enregistré l'emplacements`3://my-kinesis-test/logs`, Lake Formation crée la politique en ligne suivante et l'attache à`AWSServiceRoleForLakeFormationDataAccess`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3",
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
```

```
                "s3:DeleteObject",
                "s3:AbortMultipartUpload",
                "s3:ListMultipartUploadParts"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test/logs/*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3ListBucket",
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test"
            ]
        }
    ]
}
```

Création d'un rôle lié à un service pour Lake Formation

Vous n’avez pas besoin de créer manuellement un rôle lié à un service. Lorsque vous enregistrez un site Amazon S3 avec Lake Formation dans l' AWS Management Console AWS API AWS CLI, Lake Formation crée le rôle lié au service pour vous.

> **Important**
>
> Ce rôle lié à un service peut apparaître dans votre compte si vous avez effectué une action dans un autre service qui utilise les fonctions prises en charge par ce rôle. Pour de plus amples informations, veuillez consulter Un nouveau rôle est apparu dans mon compte IAM.

Si vous supprimez ce rôle lié à un service et que vous avez ensuite besoin de le recréer, vous pouvez utiliser la même procédure pour recréer le rôle dans votre compte. Lorsque vous enregistrez un site Amazon S3 auprès de Lake Formation, Lake Formation crée à nouveau le rôle lié au service pour vous.

Vous pouvez également utiliser la console IAM pour créer un rôle lié à un service avec le cas d'utilisation de Lake Formation. Dans l'API AWS CLI ou dans l' AWS API, créez un rôle lié à un

service avec le nom du `lakeformation.amazonaws.com` service. Pour plus d'informations, consultez Création d'un rôle lié à un service dans le Guide de l'utilisateur IAM. Si vous supprimez ce rôle lié à un service, vous pouvez utiliser ce même processus pour créer le rôle à nouveau.

Modification d'un rôle lié à un service pour Lake Formation

Lake Formation ne vous permet pas de modifier le rôle `AWSServiceRoleForLakeFormationDataAccess` lié au service. Une fois que vous avez créé un rôle lié à un service, vous ne pouvez pas changer le nom du rôle, car plusieurs entités peuvent faire référence à ce rôle. Néanmoins, vous pouvez modifier la description du rôle à l'aide d'IAM. Pour plus d'informations, consultez Modification d'un rôle lié à un service dans le Guide de l'utilisateur IAM.

Supprimer un rôle lié à un service pour Lake Formation

Si vous n'avez plus besoin d'utiliser une fonctionnalité ou un service qui nécessite un rôle lié à un service, nous vous recommandons de supprimer ce rôle. De cette façon, vous n'avez aucune entité inutilisée qui n'est pas surveillée ou gérée activement. Cependant, vous devez nettoyer les ressources de votre rôle lié à un service avant de pouvoir les supprimer manuellement.

> **Note**
>
> Si le service Lake Formation utilise le rôle lorsque vous essayez de supprimer les ressources, la suppression risque d'échouer. Si cela se produit, patientez quelques minutes et réessayez.

Pour supprimer les ressources de la Formation du Lac utilisées par la Formation du Lac

- Si vous avez utilisé le rôle lié à un service pour enregistrer des sites Amazon S3 auprès de Lake Formation, avant de supprimer le rôle lié au service, vous devez désenregistrer l'emplacement et le réenregistrer à l'aide d'un rôle personnalisé.

Pour supprimer manuellement le rôle lié à un service à l'aide d'IAM

Utilisez la console IAM, le AWS CLI, ou l' AWS API pour supprimer le rôle lié au `AWSServiceRoleForLakeFormationDataAccess` service. Pour plus d'informations, veuillez consulter Suppression d'un rôle lié à un service dans le Guide de l'utilisateur IAM.

Les conditions requises pour un rôle défini par l'utilisateur sont les suivantes :

- Lors de la création du nouveau rôle, sur la page Créer un rôle de la console IAM, sélectionnez AWS service, puis sous Choisir un cas d'utilisation, choisissez Lake Formation.

  Si vous créez le rôle en utilisant un chemin différent, assurez-vous que le rôle entretient une relation de confiance avec`lakeformation.amazonaws.com`. Pour plus d'informations, consultez la section Modification d'une politique d'approbation des rôles (console).

- Le rôle doit entretenir des relations de confiance avec les entités suivantes :
  - `glue.amazonaws.com`
  - `lakeformation.amazonaws.com`

  Pour plus d'informations, consultez la section Modification d'une politique d'approbation des rôles (console).

- Le rôle doit avoir une politique intégrée qui accorde des autorisations de lecture/écriture à Amazon S3 sur le site. Voici une politique typique.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket"
            ]
        }
    ]
}
```

- Ajoutez la politique de confiance suivante au rôle IAM pour permettre au service Lake Formation d'assumer le rôle et de vendre des informations d'identification temporaires aux moteurs d'analyse intégrés.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": [
                "sts:AssumeRole"
            ]
        }
    ]
}
```

- L'administrateur du lac de données qui enregistre l'emplacement doit disposer de l'`iam:PassRole`autorisation associée au rôle.

  Voici une politique intégrée qui accorde cette autorisation. Remplacez <account-id>par un numéro de AWS compte valide et remplacez <role-name>par le nom du rôle.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::<account-id>:role/<role-name>"
            ]
        }
```

```
    ]
}
```

- Pour permettre à Lake Formation d'ajouter des journaux dans CloudWatch les journaux et de publier des métriques, ajoutez la politique en ligne suivante.

> **Note**
>
> L'écriture dans CloudWatch Logs entraîne des frais.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Sid1",
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:PutLogEvents"
            ],
            "Resource": [
                 "arn:aws:logs:<region>:<account-id>:log-group:/aws-lakeformation-
acceleration/*",
                 "arn:aws:logs:<region>:<account-id>:log-group:/aws-lakeformation-
acceleration/*:log-stream:*"
            ]
        }
    ]
}
```

Enregistrement d'un emplacement Amazon S3

Vous devez spécifier un rôle AWS Identity and Access Management (IAM) lorsque vous enregistrez un emplacement Amazon Simple Storage Service (Amazon S3). Lake Formation assume ce rôle lorsqu'elle accorde des informations d'identification temporaires aux AWS services intégrés qui accèdent aux données à cet endroit.

> **Important**
>
> Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Vous pouvez utiliser la AWS Lake Formation console, l'API Lake Formation ou AWS Command Line Interface (AWS CLI) pour enregistrer un emplacement Amazon S3.

Avant de commencer

Passez en revue les exigences relatives au rôle utilisé pour enregistrer l'emplacement.

Pour enregistrer un emplacement (console)

> **Important**
>
> Les procédures suivantes supposent que le site Amazon S3 se trouve dans le même AWS compte que le catalogue de données et que les données du site ne sont pas chiffrées. Les autres sections de ce chapitre traitent de l'enregistrement entre comptes et de l'enregistrement des emplacements chiffrés.

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation `lakeformation:RegisterResource` IAM.
2. Dans le volet de navigation, sous Register and Ingest, sélectionnez Data lake locations.
3. Choisissez Register location, puis Browse pour sélectionner un chemin Amazon Simple Storage Service (Amazon S3).
4. (Facultatif, mais fortement recommandé) Sélectionnez Vérifier les autorisations de localisation pour afficher la liste de toutes les ressources existantes dans l'emplacement Amazon S3 sélectionné et leurs autorisations.

L'enregistrement de l'emplacement sélectionné peut permettre aux utilisateurs de votre Lake Formation d'accéder aux données déjà présentes à cet emplacement. La consultation de cette liste vous permet de garantir la sécurité des données existantes.

5. Pour le rôle IAM, choisissez le rôle `AWSServiceRoleForLakeFormationDataAccess` lié au service (par défaut) ou un rôle IAM personnalisé répondant aux exigences de. the section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”

   Vous pouvez mettre à jour un emplacement enregistré ou d'autres informations uniquement lorsque vous l'enregistrez à l'aide d'un rôle IAM personnalisé. Pour modifier un point de vente enregistré à l'aide d'un rôle lié à un service, vous devez le désenregistrer, puis l'enregistrer à nouveau.

6. Choisissez l'option Enable Data Catalog Federation pour autoriser Lake Formation à assumer un rôle et à vendre des informations d'identification temporaires aux AWS services intégrés afin d'accéder aux tables des bases de données fédérées. Si un emplacement est enregistré auprès de Lake Formation et que vous souhaitez utiliser le même emplacement pour une table dans une base de données fédérée, vous devez enregistrer le même emplacement à l'aide de l'option Enable Data Catalog Federation.

7. Choisissez le mode d'accès hybride pour ne pas activer les autorisations de Lake Formation par défaut. Lorsque vous enregistrez l'emplacement Amazon S3 en mode d'accès hybride, vous pouvez activer les autorisations de Lake Formation en optant pour les principes pour les bases de données et les tables situées sous cet emplacement.

   Pour plus d'informations sur la configuration du mode d'accès hybride, consultezMode d'accès hybride.

8. Sélectionnez Enregistrer l'emplacement.

Pour enregistrer un point de vente (AWS CLI)

1. Enregistrez un nouvel emplacement avec Lake Formation

   Cet exemple utilise un rôle lié à un service pour enregistrer l'emplacement. Vous pouvez plutôt utiliser l'`--role-arn`argument pour indiquer votre propre rôle.

   <s3-path>Remplacez-le par un chemin Amazon S3 valide, un numéro de AWS compte associé à un compte valide et <s3-access-role>par un rôle IAM autorisé à enregistrer un emplacement de données.

> **Note**
>
> Vous ne pouvez pas modifier les propriétés d'un point de vente enregistré s'il est enregistré à l'aide d'un rôle lié à un service.

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --use-service-linked-role
```

L'exemple suivant utilise un rôle personnalisé pour enregistrer l'emplacement.

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>
```

2. Pour mettre à jour une position enregistrée auprès de Lake Formation

   Vous ne pouvez modifier un emplacement enregistré que s'il est enregistré à l'aide d'un rôle IAM personnalisé. Pour un emplacement enregistré avec un rôle lié à un service, vous devez le désenregistrer et l'enregistrer à nouveau. Pour plus d'informations, consultez the section called "Annulation de l'enregistrement d'un site Amazon S3".

   ```
   aws lakeformation update-resource \
    --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>\
    --resource-arn arn:aws:s3:::<s3-path>
   ```

   ```
   aws lakeformation update-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --use-service-linked-role
   ```

3. Enregistrer un emplacement de données en mode d'accès hybride avec fédération

   ```
   aws lakeformation register-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
    --hybrid-access-enabled
   ```

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --with-federation
```

```
aws lakeformation update-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --hybrid-access-enabled
```

Pour plus d'informations, consultez la section Fonctionnement de l'RegisterResourceAPI.

> **Note**
>
> Une fois que vous avez enregistré un emplacement Amazon S3, toute AWS Glue table pointant vers cet emplacement (ou l'un de ses emplacements enfants) renvoie la valeur du `IsRegisteredWithLakeFormation` paramètre comme `true` dans l'`GetTable`appel. Il existe une limite connue selon laquelle les opérations de l'API du catalogue de données, telles que `GetTables` la mise à jour ou non de la valeur du `IsRegisteredWithLakeFormation` paramètre, renvoient la valeur par défaut, qui est fausse. `SearchTables` Il est recommandé d'utiliser l'`GetTable`API pour afficher la valeur correcte du `IsRegisteredWithLakeFormation` paramètre.

Enregistrement d'un emplacement Amazon S3 chiffré

Lake Formation s'intègre à AWS Key Management Service(AWS KMS) pour vous permettre de configurer plus facilement d'autres services intégrés pour chiffrer et déchiffrer les données sur les sites Amazon Simple Storage Service (Amazon S3).

Ils Clés gérées par AWS sont tous deux gérés par AWS KMS keys le client et pris en charge. Actuellement, le chiffrement/déchiffrement côté client n'est pris en charge qu'avec Athena.

Vous devez spécifier un rôle AWS Identity and Access Management (IAM) lorsque vous enregistrez un emplacement Amazon S3. Pour les sites Amazon S3 chiffrés, le rôle doit être autorisé à chiffrer

et à déchiffrer les données avec le AWS KMS key, ou la politique de clé KMS doit accorder des autorisations sur la clé du rôle.

> **Important**
>
> Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Le moyen le plus simple d'enregistrer l'emplacement consiste à utiliser le rôle lié au service Lake Formation. Ce rôle accorde les autorisations de lecture/écriture requises sur l'emplacement. Vous pouvez également utiliser un rôle personnalisé pour enregistrer l'emplacement, à condition qu'il réponde aux exigences dethe section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”.

> **Important**
>
> Si vous avez utilisé un Clé gérée par AWS (`aws/s3`) pour chiffrer la position Amazon S3, vous ne pouvez pas utiliser le rôle lié au service Lake Formation. Vous devez utiliser un rôle personnalisé et ajouter des autorisations IAM sur la clé du rôle. Les détails sont fournis plus loin dans cette section.

Les procédures suivantes expliquent comment enregistrer un emplacement Amazon S3 chiffré à l'aide d'une clé gérée par le client ou d'un Clé gérée par AWS.

- Enregistrement d'un emplacement chiffré à l'aide d'une clé gérée par le client
- Enregistrement d'un emplacement crypté avec un Clé gérée par AWS

Avant de commencer

Passez en revue les exigences relatives au rôle utilisé pour enregistrer l'emplacement.

Pour enregistrer un emplacement Amazon S3 chiffré à l'aide d'une clé gérée par le client

> **Note**
>
> Si la clé KMS ou l'emplacement Amazon S3 ne se trouvent pas dans le même AWS compte que le catalogue de données, suivez the section called “Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes” plutôt les instructions indiquées.

1. Ouvrez la AWS KMS console à l'adresse https://console.aws.amazon.com/kms et connectez-vous en tant qu'utilisateur administratif AWS Identity and Access Management (IAM) ou en tant qu'utilisateur pouvant modifier la politique de clé KMS utilisée pour chiffrer l'emplacement.
2. Dans le volet de navigation, sélectionnez Clés gérées par le client, puis choisissez le nom de la clé KMS souhaitée.
3. Sur la page de détails des clés KMS, choisissez l'onglet Politique clé, puis effectuez l'une des opérations suivantes pour ajouter votre rôle personnalisé ou le rôle lié au service Lake Formation en tant qu'utilisateur clé KMS :

   - Si la vue par défaut s'affiche (avec les sections Administrateurs clés, Suppression des clés, Utilisateurs clés et Autres AWS comptes), dans la section Utilisateurs clés, ajoutez votre rôle personnalisé ou le rôle lié au service Lake Formation. `AWSServiceRoleForLakeFormationDataAccess`
   - Si la politique clé (JSON) s'affiche, modifiez la politique pour ajouter votre rôle personnalisé ou le rôle lié au service Lake Formation `AWSServiceRoleForLakeFormationDataAccess` à l'objet « Autoriser l'utilisation de la clé », comme indiqué dans l'exemple suivant.

     > **Note**
     >
     > Si cet objet est manquant, ajoutez-le avec les autorisations indiquées dans l'exemple. L'exemple utilise le rôle lié à un service.

```
...
{
    "Sid": "Allow use of the key",
    "Effect": "Allow",
    "Principal": {
        "AWS": [
```

```
                    "arn:aws:iam::111122223333:role/aws-service-role/
lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAccess",
                    "arn:aws:iam::111122223333:user/keyuser"
                ]
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:DescribeKey"
            ],
            "Resource": "*"
        },
        ...
```

4. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation `lakeformation:RegisterResource` IAM.
5. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
6. Choisissez Register location, puis Browse pour sélectionner un chemin Amazon Simple Storage Service (Amazon S3).
7. (Facultatif, mais fortement recommandé) Choisissez Vérifier les autorisations de localisation pour afficher la liste de toutes les ressources existantes dans l'emplacement Amazon S3 sélectionné ainsi que leurs autorisations.

   L'enregistrement de l'emplacement sélectionné peut permettre aux utilisateurs de votre Lake Formation d'accéder aux données déjà présentes à cet emplacement. La consultation de cette liste vous permet de garantir la sécurité des données existantes.
8. Pour le rôle IAM, choisissez soit le rôle `AWSServiceRoleForLakeFormationDataAccess` lié au service (par défaut), soit votre rôle personnalisé qui répond aux. the section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”
9. Choisissez Enregistrer l'emplacement.

Pour de plus amples informations sur le rôle lié à un service, veuillez consulter Autorisations de rôle liées à un service pour Lake Formation.

Pour enregistrer une position Amazon S3 chiffrée à l'aide d'un Clé gérée par AWS

> **Important**
>
> Si l'emplacement Amazon S3 ne figure pas dans le même AWS compte que le catalogue de données, suivez the section called “Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes” plutôt les instructions indiquées dans le document.

1. Créez un rôle IAM à utiliser pour enregistrer l'emplacement. Assurez-vous qu'il répond aux exigences répertoriées dansthe section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”.
2. Ajoutez la politique intégrée suivante au rôle. Il accorde des autorisations sur la clé du rôle. La `Resource` spécification doit indiquer le nom de ressource Amazon (ARN) du Clé gérée par AWS. Vous pouvez obtenir l'ARN à partir de la AWS KMS console. Pour obtenir le bon ARN, assurez-vous de vous connecter à la AWS KMS console avec le même AWS compte et la même région Clé gérée par AWS que ceux utilisés pour chiffrer l'emplacement.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "<Clé gérée par AWS ARN>"
    }
  ]
}
```

3. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation `lakeformation:RegisterResource` IAM.
4. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.

5. Choisissez Register location, puis Browse pour sélectionner un chemin Amazon S3.
6. (Facultatif, mais fortement recommandé) Choisissez Vérifier les autorisations de localisation pour afficher la liste de toutes les ressources existantes dans l'emplacement Amazon S3 sélectionné ainsi que leurs autorisations.

   L'enregistrement de l'emplacement sélectionné peut permettre aux utilisateurs de votre Lake Formation d'accéder aux données déjà présentes à cet emplacement. La consultation de cette liste vous permet de garantir la sécurité des données existantes.
7. Pour le rôle IAM, choisissez le rôle que vous avez créé à l'étape 1.
8. Choisissez Enregistrer l'emplacement.

Enregistrement d'un emplacement Amazon S3 sur un autre AWS compte

AWS Lake Formation vous permet d'enregistrer des sites AWS Amazon Simple Storage Service (Amazon S3) sur plusieurs comptes. Par exemple, s'il AWS Glue Data Catalog se trouve dans le compte A, un utilisateur du compte A peut enregistrer un compartiment Amazon S3 dans le compte B.

L'enregistrement d'un compartiment Amazon S3 dans le AWS compte B à l'aide d'un rôle AWS Identity and Access Management (IAM) dans le AWS compte A nécessite les autorisations suivantes :

- Le rôle du compte A doit accorder des autorisations sur le compartiment du compte B.
- La politique de compartiment du compte B doit accorder des autorisations d'accès au rôle dans le compte A.

> **Important**
>
> Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.
> Vous ne pouvez pas utiliser le rôle lié au service Lake Formation pour enregistrer une position sur un autre compte. Vous devez plutôt utiliser un rôle défini par l'utilisateur. Le rôle doit répondre aux exigences dethe section called “Exigences relatives aux rôles utilisés pour

enregistrer des sites". Pour de plus amples informations sur le rôle lié à un service, veuillez consulter Autorisations de rôle liées à un service pour Lake Formation.

Avant de commencer

Passez en revue les exigences relatives au rôle utilisé pour enregistrer l'emplacement.

Pour enregistrer une position dans un autre AWS compte

> **Note**
>
> Si l'emplacement est crypté, suivez the section called "Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes" plutôt les instructions indiquées.

La procédure suivante suppose qu'un mandant du compte 1111-2222-3333, qui contient le catalogue de données, souhaite enregistrer le compartiment Amazon S3, qui se trouve dans le compte `awsexamplebucket1` 1234-5678-9012.

1. Dans le compte 1111-2222-3333, connectez-vous à la console IAM AWS Management Console et ouvrez-la à l'adresse. https://console.aws.amazon.com/iam/
2. Créez un nouveau rôle ou consultez un rôle existant qui répond aux exigences dethe section called "Exigences relatives aux rôles utilisés pour enregistrer des sites". Assurez-vous que le rôle accorde des autorisations à Amazon S3 sur`awsexamplebucket1`.
3. Ouvrez la console Amazon S3 sur https://console.aws.amazon.com/s3/. Connectez-vous avec le compte 1234-5678-9012.
4. Dans la liste Nom du compartiment, choisissez le nom du compartiment,`awsexamplebucket1`.
5. Choisissez Autorisations.
6. Sur la page Autorisations, choisissez Bucket Policy.
7. Dans l'éditeur de politique Bucket, collez la politique suivante. Remplacez <role-name>par le nom de votre rôle.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
```

```
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::awsexamplebucket1"
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::awsexamplebucket1/*"
        }
    ]
}
```

8. Choisissez Enregistrer.
9. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous au compte 1111-2222-3333 en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant des autorisations suffisantes pour enregistrer des emplacements.
10. Dans le volet de navigation, sous Administration, sélectionnez Data lake locations.
11. Sur la page des emplacements des Data Lake, choisissez Enregistrer l'emplacement.
12. Sur la page Enregistrer l'emplacement, pour le chemin Amazon S3, entrez le nom du compartiments3://awsexamplebucket1.

> **Note**
>
> Vous devez saisir le nom du bucket car les buckets multi-comptes n'apparaissent pas dans la liste lorsque vous choisissez Browse.

13. Pour le rôle IAM, choisissez votre rôle.
14. Choisissez Enregistrer l'emplacement.

Enregistrement d'un emplacement Amazon S3 chiffré sur plusieurs AWS comptes

AWS Lake Formation s'intègre à AWS Key Management Service(AWS KMS) pour vous permettre de configurer plus facilement d'autres services intégrés pour chiffrer et déchiffrer les données sur les sites Amazon Simple Storage Service (Amazon S3).

Les deux clés sont gérées par le client et Clés gérées par AWS sont prises en charge. Le chiffrement/déchiffrement côté client n'est pas pris en charge.

> **Important**
>
> Évitez d'enregistrer un compartiment Amazon S3 sur lequel les paiements par les demandeurs sont activés. Pour les buckets enregistrés auprès de Lake Formation, le rôle utilisé pour enregistrer le bucket est toujours considéré comme le demandeur. Si un autre AWS compte accède au bucket, l'accès aux données est facturé au propriétaire du bucket si le rôle appartient au même compte que le propriétaire du bucket.

Cette section explique comment enregistrer un emplacement Amazon S3 dans les circonstances suivantes :

- Les données de l'emplacement Amazon S3 sont chiffrées à l'aide d'une clé KMS créée dans AWS KMS.
- L'emplacement Amazon S3 n'est pas enregistré dans le même AWS compte que le AWS Glue Data Catalog.
- La clé KMS se trouve ou non dans le même AWS compte que le catalogue de données.

L'enregistrement d' AWS KMS un compartiment Amazon S3 chiffré dans le AWS compte B à l'aide d'un rôle AWS Identity and Access Management (IAM) dans le AWS compte A nécessite les autorisations suivantes :

- Le rôle du compte A doit accorder des autorisations sur le compartiment du compte B.
- La politique de compartiment du compte B doit accorder des autorisations d'accès au rôle dans le compte A.
- Si la clé KMS se trouve dans le compte B, la politique clé doit accorder l'accès au rôle dans le compte A, et le rôle dans le compte A doit accorder des autorisations sur la clé KMS.

Dans la procédure suivante, vous allez créer un rôle dans le AWS compte qui contient le catalogue de données (compte A dans la discussion précédente). Vous utilisez ensuite ce rôle pour enregistrer l'emplacement. Lake Formation assume ce rôle lors de l'accès aux données sous-jacentes dans Amazon S3. Le rôle assumé dispose des autorisations requises sur la clé KMS. Par conséquent, vous n'avez pas à accorder d'autorisations sur la clé KMS aux principaux accédant aux données sous-jacentes via des tâches ETL ou des services intégrés tels que Amazon Athena.

> **Important**
>
> Vous ne pouvez pas utiliser le rôle lié au service Lake Formation pour enregistrer une position sur un autre compte. Vous devez plutôt utiliser un rôle défini par l'utilisateur. Le rôle doit répondre aux exigences dethe section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”. Pour de plus amples informations sur le rôle lié à un service, veuillez consulter Autorisations de rôle liées à un service pour Lake Formation.

Avant de commencer

Passez en revue les exigences relatives au rôle utilisé pour enregistrer l'emplacement.

Pour enregistrer une position Amazon S3 chiffrée sur plusieurs AWS comptes

1. Dans le même AWS compte que le catalogue de données, connectez-vous à la console IAM AWS Management Console et ouvrez-la à https://console.aws.amazon.com/iam/ l'adresse.
2. Créez un nouveau rôle ou consultez un rôle existant qui répond aux exigences dethe section called “Exigences relatives aux rôles utilisés pour enregistrer des sites”. Assurez-vous que le rôle inclut une politique qui accorde des autorisations à Amazon S3 sur le site.
3. Si la clé KMS ne se trouve pas dans le même compte que le catalogue de données, ajoutez au rôle une politique intégrée qui accorde les autorisations requises sur la clé KMS. Voici un exemple de politique . Remplacez <cmk-region>et *< cmk-account-id >* par la région et le numéro de compte de la clé KMS. Remplacez <key-id>par l'identifiant de la clé.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
        "Effect": "Allow",
        "Action": [
            "kms:Encrypt",
```

```
            "kms:Decrypt",
            "kms:ReEncrypt*",
            "kms:GenerateDataKey*",
            "kms:DescribeKey"
         ],
        "Resource": "arn:aws:kms:<cmk-region>:<cmk-account-id>:key/<key-id>"
        }
    ]
}
```

4. Sur la console Amazon S3, ajoutez une politique de compartiment accordant les autorisations Amazon S3 requises au rôle. Voici un exemple de stratégie de compartiment. Remplacez *< catalog-account-id >* par le numéro de AWS compte du catalogue de données, <role-name>par le nom de votre rôle et <bucket-name>par le nom du bucket.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::<catalog-account-id>:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::<bucket-name>"
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::<catalog-account-id>:role/<role-name>"
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::<bucket-name>/*"
        }
    ]
}
```

5. Dans AWS KMS, ajoutez le rôle en tant qu'utilisateur de la clé KMS.

a. Ouvrez la AWS KMS console à l'adresse https://console.aws.amazon.com/kms. Connectez-vous ensuite en tant qu'administrateur ou en tant qu'utilisateur pouvant modifier la politique de clé de la clé KMS utilisée pour chiffrer l'emplacement.

b. Dans le volet de navigation, choisissez Customer managed keys, puis choisissez le nom de la clé KMS.

c. Sur la page des détails de la clé KMS, sous l'onglet Stratégie clé, si la vue JSON de la politique clé ne s'affiche pas, choisissez Basculer vers la vue politique.

d. Dans la section Politique clé, choisissez Modifier et ajoutez le nom de ressource Amazon (ARN) du rôle à l'`Allow use of the key`objet, comme indiqué dans l'exemple suivant.

> **Note**
>
> Si cet objet est manquant, ajoutez-le avec les autorisations indiquées dans l'exemple.

```
...
{
    "Sid": "Allow use of the key",
    "Effect": "Allow",
    "Principal": {
        "AWS": [
            "arn:aws:iam::<catalog-account-id>:role/<role-name>"
        ]
    },
    "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
    ],
    "Resource": "*"
},
...
```

Pour plus d'informations, voir Autoriser les utilisateurs d'autres comptes à utiliser une clé KMS dans le Guide du AWS Key Management Service développeur.

6. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous au AWS compte Data Catalog en tant qu'administrateur du lac de données.
7. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
8. Choisissez Enregistrer l'emplacement.
9. Sur la page Enregistrer l'emplacement, pour le chemin Amazon S3, entrez le chemin de localisation sous la forme**s3://*<bucket>*/*<prefix>***. <bucket>Remplacez-le par le nom du compartiment et <prefix>par le reste du chemin correspondant à l'emplacement.

   > **Note**
   >
   > Vous devez saisir le chemin car les compartiments multicomptes n'apparaissent pas dans la liste lorsque vous choisissez Parcourir.

10. Pour le rôle IAM, choisissez le rôle à l'étape 2.
11. Choisissez Enregistrer l'emplacement.

Annulation de l'enregistrement d'un site Amazon S3

Vous pouvez annuler l'enregistrement d'un site Amazon Simple Storage Service (Amazon S3) si vous ne souhaitez plus qu'il soit géré par Lake Formation. L'annulation de l'enregistrement d'un emplacement n'affecte pas les autorisations de localisation des données de Lake Formation accordées pour cet emplacement. Vous pouvez réenregistrer un emplacement que vous avez désenregistré, et les autorisations de localisation des données restent en vigueur. Vous pouvez utiliser un autre rôle pour réenregistrer l'emplacement.

Pour désenregistrer une position (console)

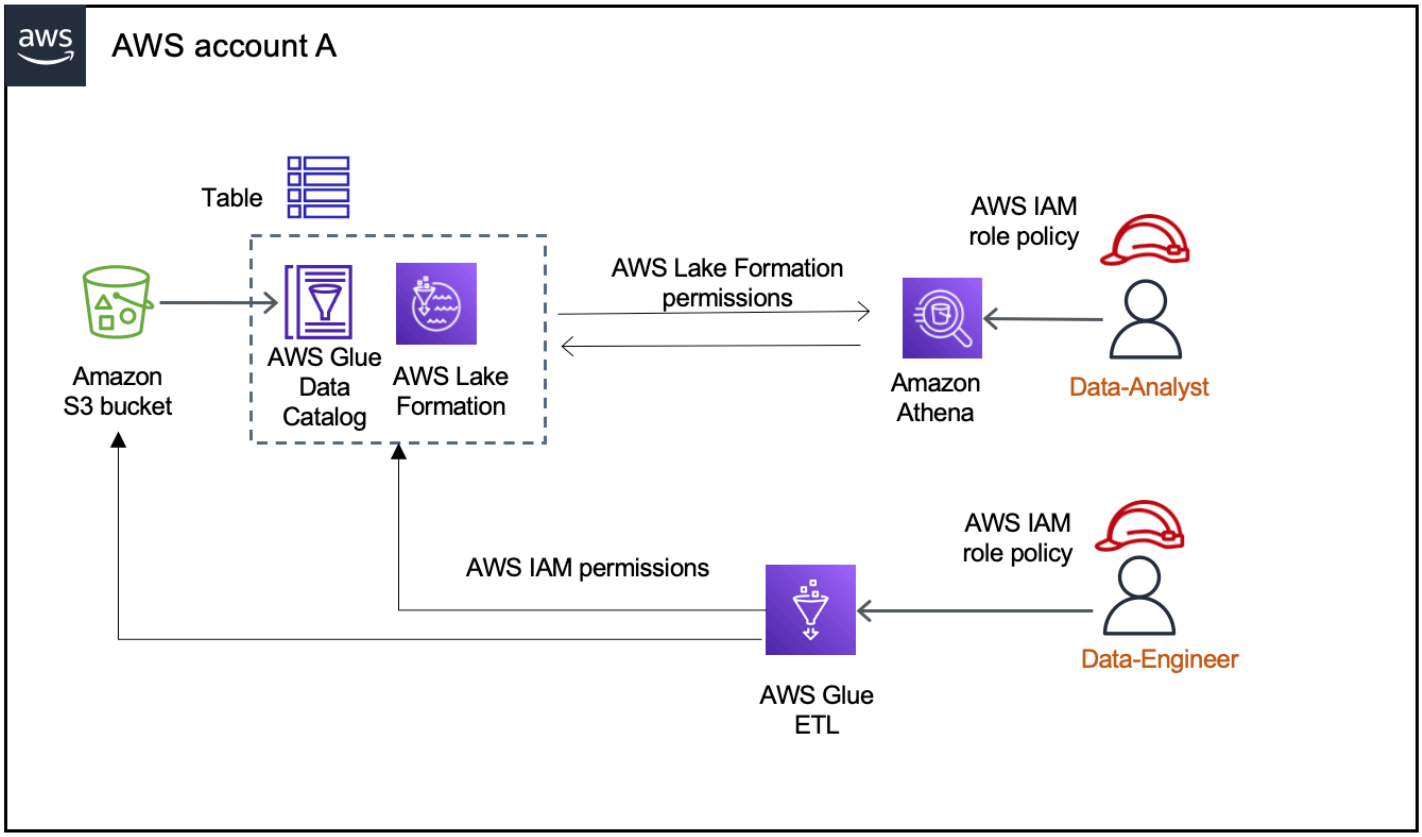
1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant de l'autorisation `lakeformation:RegisterResource` IAM.
2. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
3. Sélectionnez un emplacement, puis dans le menu Actions, choisissez Supprimer.
4. Lorsque vous êtes invité à confirmer, choisissez Supprimer.

Mode d'accès hybride

AWS Lake Formation le mode d'accès hybride prend en charge deux voies d'autorisation vers les mêmes AWS Glue Data Catalog bases de données et tables.
Dans le premier parcours, Lake Formation vous permet de sélectionner des principes spécifiques et de leur accorder des autorisations Lake Formation pour accéder aux bases de données et aux tables en vous inscrivant. La deuxième voie permet à tous les autres principaux d'accéder à ces ressources via les politiques principales IAM par défaut pour Amazon S3 et AWS Glue les actions.

Lorsque vous enregistrez un site Amazon S3 auprès de Lake Formation, vous avez la possibilité d'appliquer les autorisations de Lake Formation à toutes les ressources de cet emplacement ou d'utiliser le mode d'accès hybride. Le mode d'accès hybride applique uniquement `CREATE_TABLE` `CREATE_PARTITION` les `UPDATE_TABLE` autorisations par défaut. Lorsqu'un site Amazon S3 est en mode hybride, vous pouvez activer les autorisations de Lake Formation en optant pour les principes pour les bases de données et les tables situées sous cet emplacement.

Ainsi, le mode d'accès hybride offre la flexibilité nécessaire pour activer Lake Formation de manière sélective pour les bases de données et les tables de votre catalogue de données pour un ensemble spécifique d'utilisateurs sans interrompre l'accès pour les autres utilisateurs ou charges de travail existants.



Pour les considérations et les restrictions, consultez Considérations et limites relatives au mode d'accès hybride .

Termes et définitions

Voici les définitions des ressources du catalogue de données en fonction de la façon dont vous configurez les autorisations d'accès :

Ressource Lake Formation

Une ressource enregistrée auprès de Lake Formation. Les utilisateurs ont besoin des autorisations de Lake Formation pour accéder à la ressource.

AWS Glue ressource

Une ressource qui n'est pas enregistrée auprès de Lake Formation. Les utilisateurs n'ont besoin que d'autorisations IAM pour accéder à la ressource, car celle-ci dispose d'autorisations de `IAMAllowedPrincipals` groupe. Les autorisations de Lake Formation ne sont pas appliquées.

Pour plus d'informations sur les autorisations de `IAMAllowedPrincipals` groupe, consultezAutorisations relatives aux métadonnées.

Ressource hybride

Une ressource enregistrée en mode d'accès hybride. En fonction des utilisateurs accédant à la ressource, celle-ci passe de manière dynamique à une ressource de Lake Formation ou à une AWS Glue ressource.

Cas d'utilisation courants du mode d'accès hybride

Vous pouvez utiliser le mode d'accès hybride pour fournir un accès dans le cadre de scénarios de partage de données à compte unique ou entre comptes :

Scénarios de compte unique

- Convertir une AWS Glue ressource en ressource hybride : dans ce scénario, vous n'utilisez pas actuellement Lake Formation, mais vous souhaitez adopter les autorisations Lake Formation pour les bases de données et les tables du catalogue de données. Lorsque vous enregistrez l'emplacement Amazon S3 en mode d'accès hybride, vous pouvez accorder des autorisations Lake Formation aux utilisateurs qui optent pour des bases de données et des tables spécifiques pointant vers cet emplacement.
- Conversion d'une ressource Lake Formation en ressource hybride — Actuellement, vous utilisez les autorisations Lake Formation pour contrôler l'accès à une base de données de catalogue de données, mais vous souhaitez fournir l'accès à de nouveaux principaux en utilisant les autorisations IAM pour Amazon S3 et AWS Glue sans interrompre les autorisations Lake Formation existantes.

  Lorsque vous mettez à jour l'enregistrement d'un emplacement de données en mode d'accès hybride, les nouveaux principaux peuvent accéder à la base de données du catalogue de données pointant vers l'emplacement Amazon S3 en utilisant les politiques d'autorisation IAM sans interrompre les autorisations Lake Formation des utilisateurs existants.

  Avant de mettre à jour l'enregistrement de la localisation des données pour activer le mode d'accès hybride, vous devez d'abord activer les principaux qui accèdent actuellement à la ressource avec les autorisations de Lake Formation.
  Cela permet d'éviter toute interruption potentielle du flux de travail en cours.
  Vous devez également `Super` autoriser le `IAMAllowedPrincipal` groupe à accéder aux tables de la base de données.

Scénarios de partage de données entre comptes

- Partagez AWS Glue des ressources à l'aide du mode d'accès hybride : dans ce scénario, le compte producteur possède des tables dans une base de données qui sont actuellement partagées avec un compte client conformément aux politiques d'autorisation IAM pour Amazon S3 et aux AWS Glue actions. L'emplacement des données de la base de données n'est pas enregistré auprès de Lake Formation.

  Avant d'enregistrer l'emplacement des données en mode d'accès hybride, vous devez mettre à jour les paramètres de version du compte Cross vers la version 4. La version 4 fournit les nouvelles politiques AWS RAM d'autorisation requises pour le partage entre comptes lorsque le `IAMAllowedPrincipal` groupe dispose d'une `Super` autorisation sur la ressource. Pour les ressources disposant d'autorisations de `IAMAllowedPrincipal` groupe, vous pouvez accorder des autorisations de Lake Formation à des comptes externes et les autoriser à utiliser les autorisations de Lake Formation. L'administrateur du lac de données du compte destinataire peut accorder des autorisations Lake Formation aux principaux du compte et les autoriser à appliquer les autorisations Lake Formation.
- Partagez les ressources de Lake Formation en mode d'accès hybride — Actuellement, le compte producteur contient des tables dans une base de données qui sont partagées avec un compte consommateur appliquant les autorisations de Lake Formation. L'emplacement des données de la base de données est enregistré auprès de Lake Formation.

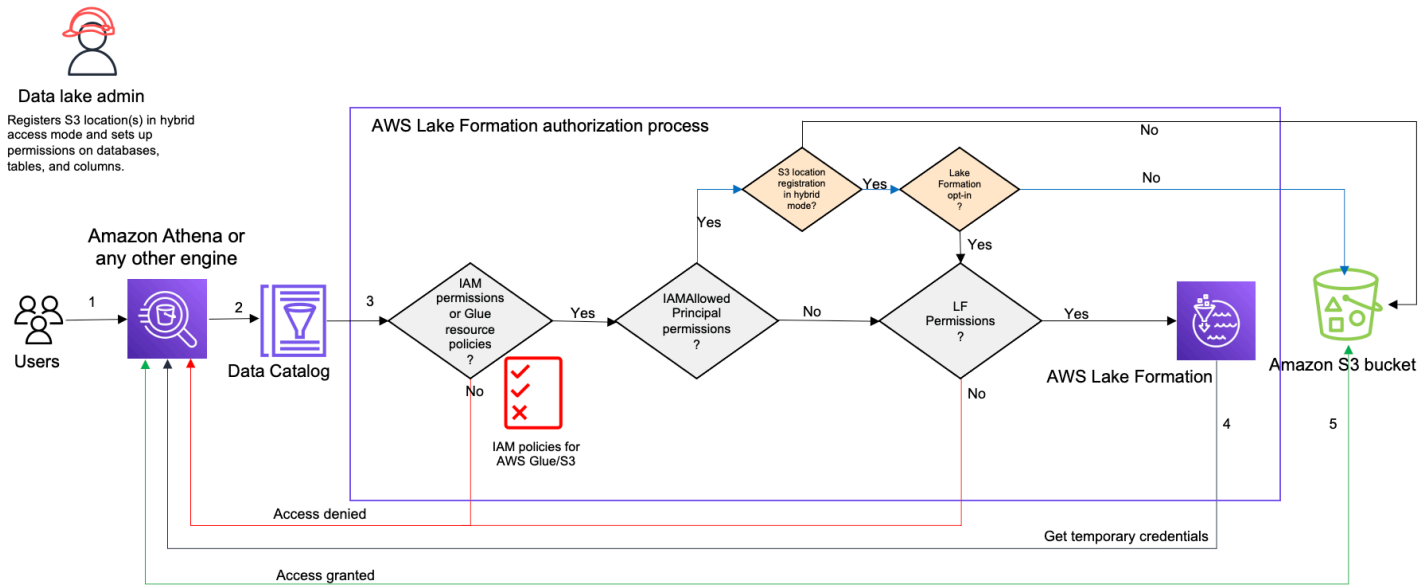
  Dans ce cas, vous pouvez mettre à jour l'enregistrement de l'emplacement Amazon S3 en mode d'accès hybride, et partager les données d'Amazon S3 et les métadonnées de Data Catalog en utilisant les politiques de compartiment Amazon S3 et les politiques de ressources du catalogue de données avec les principaux du compte client. Vous devez réoctroyer les autorisations Lake Formation existantes et accepter les principales avant de mettre à jour l'enregistrement de la position Amazon S3. Vous devez également `Super` autoriser le `IAMAllowedPrincipals` groupe à accéder aux tables de la base de données.

Rubriques

- Comment fonctionne le mode d'accès hybride
- Configuration du mode d'accès hybride : scénarios courants
- Supprimer les principes et les ressources du mode d'accès hybride
- Affichage des principes et des ressources en mode d'accès hybride
- Ressources supplémentaires

Comment fonctionne le mode d'accès hybride

Le schéma suivant montre comment fonctionne l'autorisation de Lake Formation en mode d'accès hybride lorsque vous interrogez les ressources du catalogue de données.



Avant d'accéder aux données de votre lac de données, un administrateur de lac de données ou un utilisateur disposant d'autorisations administratives définit des politiques utilisateur individuelles pour les tables du catalogue de données afin d'autoriser ou de refuser l'accès aux tables de votre catalogue de données. Ensuite, un directeur autorisé à effectuer une `RegisterResource` opération enregistre l'emplacement de la table sur Amazon S3 auprès de Lake Formation en mode d'accès hybride. L'administrateur accorde des autorisations Lake Formation à des utilisateurs spécifiques sur les bases de données et les tables du catalogue de données et les autorise à utiliser les autorisations Lake Formation pour ces bases de données et tables en mode d'accès hybride.

1. Soumet une requête : un directeur soumet une requête ou un script ETL à l'aide d'un service intégré tel qu'Amazon Athena, AWS Glue Amazon EMR ou Amazon Redshift Spectrum.
2. Demande des données : le moteur d'analyse intégré identifie la table demandée et envoie la demande de métadonnées au catalogue de données (`GetTable`,`GetDatabase`).
3. Vérifie les autorisations - Le catalogue de données vérifie les autorisations d'accès du principal demandeur auprès de Lake Formation.
   a. Si aucune autorisation de `IAMAllowedPrincipals` groupe n'est attachée à la table, les autorisations de Lake Formation sont appliquées.

b. Si le principal a choisi d'utiliser les autorisations de Lake Formation en mode d'accès hybride et que des autorisations de `IAMAllowedPrincipals` groupe sont associées à la table, les autorisations de Lake Formation sont appliquées. Le moteur de requête applique les filtres qu'il a reçus de Lake Formation et renvoie les données à l'utilisateur.

c. Si l'emplacement de la table n'est pas enregistré auprès de Lake Formation et que le directeur n'a pas choisi d'utiliser les autorisations de Lake Formation en mode d'accès hybride, le catalogue de données vérifie si des autorisations de `IAMAllowedPrincipals` groupe sont associées à la table. Si cette autorisation existe sur la table, tous les principaux du compte obtiennent `Super` ou obtiennent `All` des autorisations sur la table.

4. Obtenir des informations d'identification — Le catalogue de données vérifie et indique au moteur si l'emplacement de la table est enregistré auprès de Lake Formation ou non. Si les données sous-jacentes sont enregistrées auprès de Lake Formation, le moteur d'analyse demande à Lake Formation des informations d'identification temporaires pour accéder aux données du compartiment Amazon S3.

5. Obtenir des données — Si le principal est autorisé à accéder aux données de la table, Lake Formation fournit un accès temporaire au moteur d'analyse intégré. À l'aide de l'accès temporaire, le moteur d'analyse extrait les données d'Amazon S3 et effectue le filtrage nécessaire, tel que le filtrage par colonne, ligne ou cellule. Lorsque le moteur a terminé d'exécuter la tâche, il renvoie les résultats à l'utilisateur. Ce processus est appelé distributeur d'informations d'identification. Pour plus d'informations, voirIntégration à Lake Formation.

6. Si l'emplacement des données de la table n'est pas enregistré auprès de Lake Formation, le deuxième appel du moteur d'analyse est directement envoyé à Amazon S3. La politique de compartiment Amazon S3 et la politique utilisateur IAM concernées sont évaluées pour l'accès aux données. Chaque fois que vous utilisez des politiques IAM, veillez à respecter les bonnes pratiques IAM. Pour plus d'informations, consultez la section Bonnes pratiques en matière de sécurité dans IAM dans le guide de l'utilisateur d'IAM.

Configuration du mode d'accès hybride : scénarios courants

Comme pour les autorisations de Lake Formation, vous disposez généralement de deux types de scénarios dans lesquels vous pouvez utiliser le mode d'accès hybride pour gérer l'accès aux données : fournir un accès aux principaux au sein d'un seul Compte AWS et fournir un accès à un externe Compte AWS ou à un principal.

Cette section fournit des instructions pour configurer le mode d'accès hybride dans les scénarios suivants :

Gérez les autorisations en mode d'accès hybride en un Compte AWS

- Conversion d'une AWS Glue ressource en ressource hybride — Vous fournissez actuellement l'accès aux tables d'une base de données à tous les principaux de votre compte à l'aide des autorisations IAM pour Amazon S3, AWS Glue mais vous souhaitez adopter Lake Formation pour gérer les autorisations de manière progressive.

- Conversion d'une ressource de la Lake Formation en une ressource hybride — Vous utilisez actuellement Lake Formation pour gérer l'accès aux tables d'une base de données pour tous les principaux de votre compte, mais vous souhaitez utiliser Lake Formation uniquement pour des principes spécifiques. Vous souhaitez donner accès à de nouveaux principaux en utilisant les autorisations IAM pour AWS Glue Amazon S3 sur la même base de données et les mêmes tables.

Gérez les autorisations en mode d'accès hybride sur Compte AWS s

- Partage d'une AWS Glue ressource à l'aide du mode d'accès hybride— À l'heure actuelle, vous n'utilisez pas Lake Formation pour gérer les autorisations associées à une table, mais vous souhaitez appliquer les autorisations Lake Formation pour permettre aux principaux d'accéder à un autre compte.
- Partage d'une ressource de Lake Formation à l'aide du mode d'accès hybride— Vous utilisez Lake Formation pour gérer l'accès à une table, mais vous souhaitez fournir l'accès aux principaux d'un autre compte en utilisant les autorisations IAM pour AWS Glue Amazon S3 sur la même base de données et les mêmes tables.

Configuration du mode d'accès hybride — Étapes de haut niveau

1. Enregistrez l'emplacement des données Amazon S3 auprès de Lake Formation en sélectionnant le mode d'accès hybride.
2. Les principaux doivent être `DATA_LOCATION` autorisés à accéder à l'emplacement d'un lac de données pour créer des tables ou des bases de données du catalogue de données pointant vers cet emplacement.
3. Définissez le paramètre de version multi-comptes sur Version 4.

4. Accordez des autorisations détaillées à des utilisateurs ou à des rôles IAM spécifiques sur les bases de données et les tables. Dans le même temps, assurez-vous de définir `Super` des `All` autorisations pour le `IAMAllowedPrincipals` groupe sur la base de données et sur toutes les tables ou sur certaines d'entre elles.
5. Choisissez les principes et les ressources. Les autres principaux du compte peuvent continuer à accéder aux bases de données et aux tables en utilisant les politiques d'autorisation IAM pour AWS Glue les actions Amazon S3.
6. Nettoyez éventuellement les politiques d'autorisation IAM pour Amazon S3 pour les principaux utilisateurs qui ont choisi d'utiliser les autorisations de Lake Formation.

Conditions préalables à la configuration du mode d'accès hybride

Les conditions préalables à la configuration du mode d'accès hybride sont les suivantes :

> **Note**
>
> Nous recommandons à un administrateur de Lake Formation d'enregistrer l'emplacement Amazon S3 en mode d'accès hybride et d'opter pour les principes et les ressources.

1. Accordez l'autorisation de localisation des données (`DATA_LOCATION_ACCESS`) pour créer des ressources de catalogue de données pointant vers les sites Amazon S3. Les autorisations de localisation des données contrôlent la possibilité de créer des bases de données et des tables de catalogue de données pointant vers des emplacements Amazon S3 particuliers.
2. Pour partager les ressources du catalogue de données avec un autre compte en mode d'accès hybride (sans supprimer les autorisations de `IAMAllowedPrincipals` groupe associées à la ressource), vous devez mettre à jour les paramètres de version multi-comptes vers la version 4. Pour mettre à jour la version à l'aide de la console Lake Formation, choisissez Version 4 sous Paramètres de version multi-comptes sur la page des paramètres du catalogue de données.

   Vous pouvez également utiliser la `put-data-lake-settings` AWS CLI commande pour définir le `CROSS_ACCOUNT_VERSION` paramètre sur la version 4 :

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
 file://settings
{
"DataLakeAdmins": [
```

```
        {
"DataLakePrincipalIdentifier": "arn:aws:iam::<111122223333>:user/<user-name>"
        }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "Parameters": {
"CROSS_ACCOUNT_VERSION": "4"
    }
}
```

3.

Pour accorder des autorisations entre comptes en mode d'accès hybride, le concédant doit disposer des autorisations IAM requises pour AWS Glue les services et les services. AWS RAM La politique AWS gérée `AWSLakeFormationCrossAccountManager` accorde les autorisations requises.
Pour permettre le partage de données entre comptes en mode d'accès hybride, nous avons mis à jour la politique `AWSLakeFormationCrossAccountManager` gérée en ajoutant deux nouvelles autorisations IAM :

- RAM : ListResourceSharePermissions
- RAM : AssociateResourceSharePermission

> **Note**
>
> Si vous n'utilisez pas la politique AWS gérée pour le rôle de donateur, ajoutez les politiques ci-dessus à vos politiques personnalisées.

Conversion d'une AWS Glue ressource en ressource hybride

Suivez ces étapes pour enregistrer un site Amazon S3 en mode d'accès hybride et intégrer de nouveaux utilisateurs de Lake Formation sans interrompre l'accès aux données des utilisateurs existants du catalogue de données.

Description du scénario - L'emplacement des données n'est pas enregistré auprès de Lake Formation, et l'accès des utilisateurs à la base de données et aux tables du catalogue de données est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions.

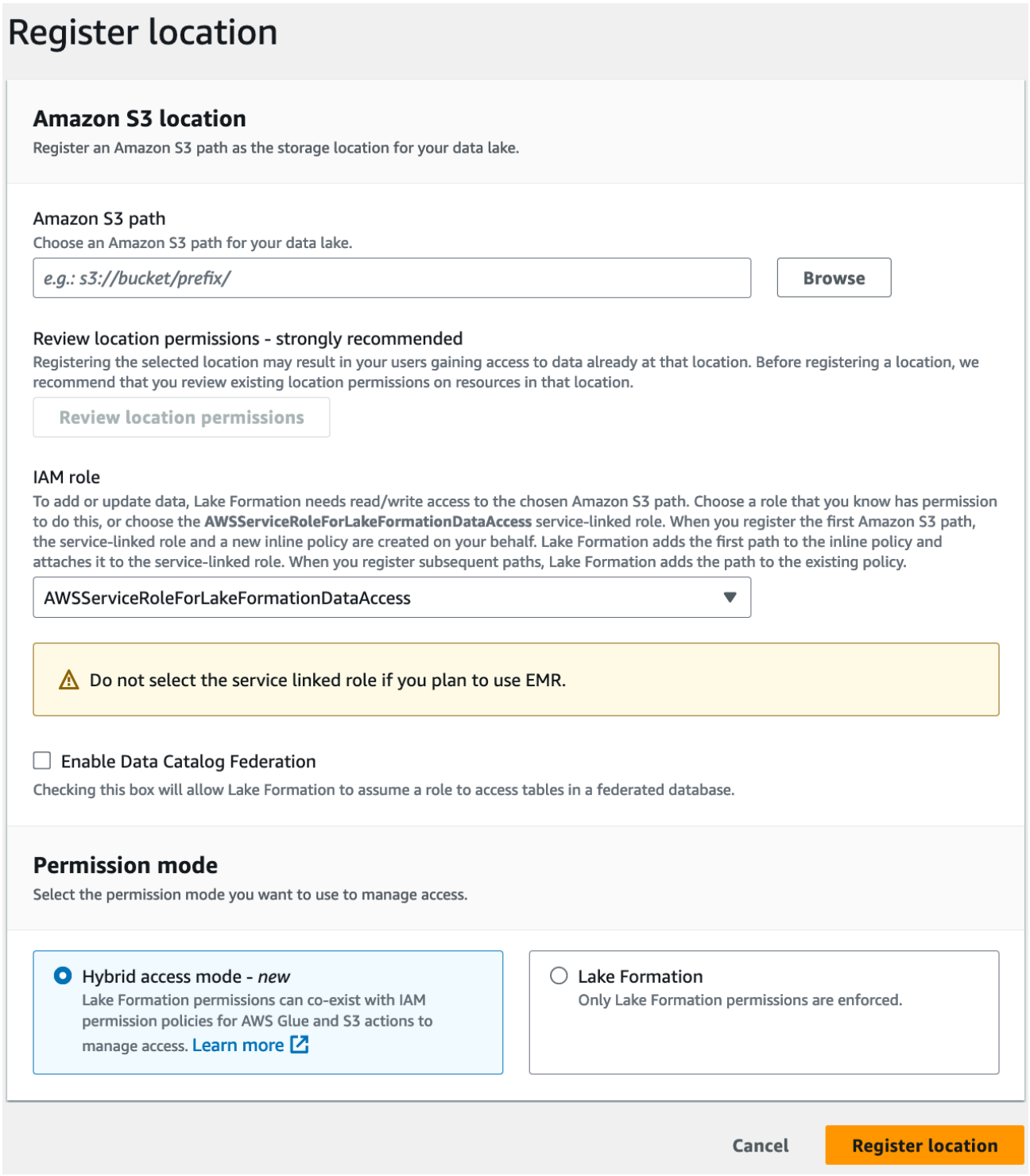
Le `IAMAllowedPrincipals` groupe dispose par défaut d'`Super`autorisations sur toutes les tables de la base de données.

Pour activer le mode d'accès hybride pour un emplacement de données non enregistré auprès de Lake Formation

1. Enregistrez un emplacement Amazon S3 permettant le mode d'accès hybride.

   Console

   1. Connectez-vous à la console Lake Formation en tant qu'administrateur du lac de données.
   2. Dans le volet de navigation, sélectionnez Emplacements des lacs de données sous Administration.
   3. Choisissez Enregistrer l'emplacement.



4. Dans la fenêtre Enregistrer l'emplacement, choisissez le chemin Amazon S3 que vous souhaitez enregistrer auprès de Lake Formation.
5. Pour le rôle IAM, choisissez le rôle `AWSServiceRoleForLakeFormationDataAccess` lié au service (par défaut) ou un rôle IAM personnalisé rôle qui répond aux exigences de[Exigences relatives aux rôles utilisés pour enregistrer des sites](#).

6. Choisissez le mode d'accès hybride pour appliquer des politiques précises de contrôle d'accès à Lake Formation aux principes d'adhésion ainsi qu'aux bases de données et aux tables du catalogue de données pointant vers l'emplacement enregistré.

   Choisissez Lake Formation pour permettre à Lake Formation d'autoriser les demandes d'accès à l'emplacement enregistré.

7. Choisissez Enregistrer l'emplacement.

AWS CLI

Voici un exemple d'enregistrement d'un emplacement de données auprès de Lake Formation HybridAccessEnabled avec:true/false. La valeur par défaut du `HybridAccessEnabled` paramètre est false. Remplacez le chemin, le nom du rôle et l'identifiant du AWS compte Amazon S3 par des valeurs valides.

```
aws lakeformation register-resource --cli-input-json file:file path
json:
    {
        "ResourceArn": "arn:aws:s3:::s3-path",
        "UseServiceLinkedRole": false,
        "RoleArn": "arn:aws:iam::<123456789012>:role/<role-name>",
        "HybridAccessEnabled": true
    }
```

2. Accordez des autorisations et autorisez les principaux à utiliser les autorisations de Lake Formation pour les ressources en mode d'accès hybride

Avant d'activer les principes et les ressources en mode d'accès hybride, vérifiez que les `All` autorisations `Super` ou autorisations de `IAMAllowedPrincipals` regroupement existent sur les bases de données et les tables dont l'emplacement est enregistré auprès de Lake Formation en mode d'accès hybride.

> **Note**
>
> Vous ne pouvez pas accorder d'autorisation au `IAMAllowedPrincipals` groupe `All tables` dans une base de données. Vous devez sélectionner chaque table séparément dans le menu déroulant et accorder des autorisations. En outre, lorsque vous créez de nouvelles tables dans la base de données, vous pouvez choisir de les utiliser `Use only IAM access control for new tables in new databases` dans les paramètres

du catalogue de données. Cette option accorde automatiquement `Super` l'autorisation au `IAMAllowedPrincipals` groupe lorsque vous créez de nouvelles tables dans la base de données.

Console

1. Sur la console Lake Formation, sous Catalogue de données, sélectionnez Databases ou Tables.
2. Sélectionnez une base de données ou une table dans la liste, puis choisissez Grant dans le menu Actions.
3. Choisissez des principes pour accorder des autorisations sur la base de données, les tables et les colonnes à l'aide d'une méthode de ressource nommée ou de balises LF.

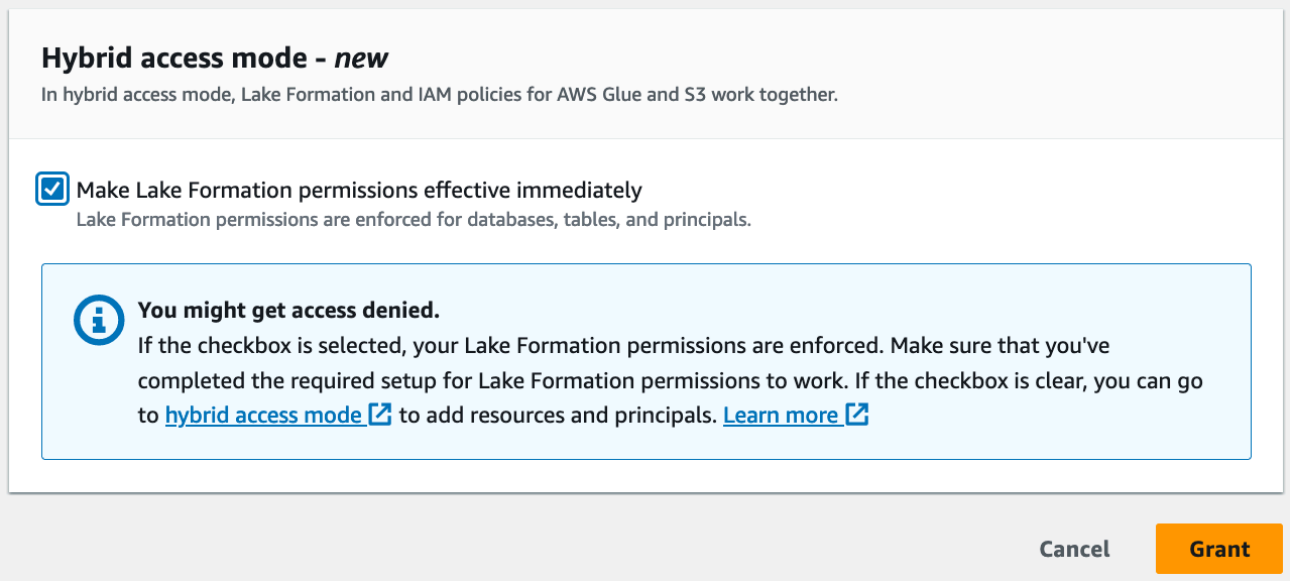
   Vous pouvez également choisir les autorisations du lac de données, sélectionner les principaux auxquels accorder les autorisations dans la liste, puis choisir Grant.

   Pour plus de détails sur l'octroi d'autorisations de données, consultezOctroi et révocation d'autorisations sur les ressources du catalogue de données.

   > **Note**
   >
   > Si vous accordez à un principal l'autorisation de créer une table, vous devez également accorder des autorisations de localisation des données (`DATA_LOCATION_ACCESS`) au principal. Cette autorisation n'est pas nécessaire pour mettre à jour les tables.
   > Pour plus d'informations, consultez Octroi d'autorisations de localisation des données.

4. Lorsque vous utilisez la méthode des ressources nommées pour accorder des autorisations, l'option permettant d'activer les principes et les ressources est disponible dans la section inférieure de la page d'autorisation des données d'octroi.

   Choisissez Rendre les autorisations Lake Formation effectives immédiatement pour activer les autorisations Lake Formation pour les principaux et les ressources.



5. Choisissez Grant (Accorder).

   Lorsque vous activez le principal A sur la table A qui pointe vers un emplacement de données, cela permet au principal A d'accéder à l'emplacement de cette table en utilisant les autorisations de Lake Formation si l'emplacement des données est enregistré en mode hybride.

AWS CLI

Voici un exemple d'activation d'un principal et d'une table en mode d'accès hybride. Remplacez le nom du rôle, l'identifiant du AWS compte, le nom de la base de données et le nom de la table par des valeurs valides.

```
aws lakeformation create-lake-formation-opt-in --cli-input-json file://file path
json:
  {
        "Principal": {
            "DataLakePrincipalIdentifier":
 "arn:aws:iam::<123456789012>:role/<hybrid-access-role>"
        },
        "Resource": {
            "Table": {
                "CatalogId": "<123456789012>",
                "DatabaseName": "<hybrid_test>",
                "Name": "<hybrid_test_table>"
            }
```

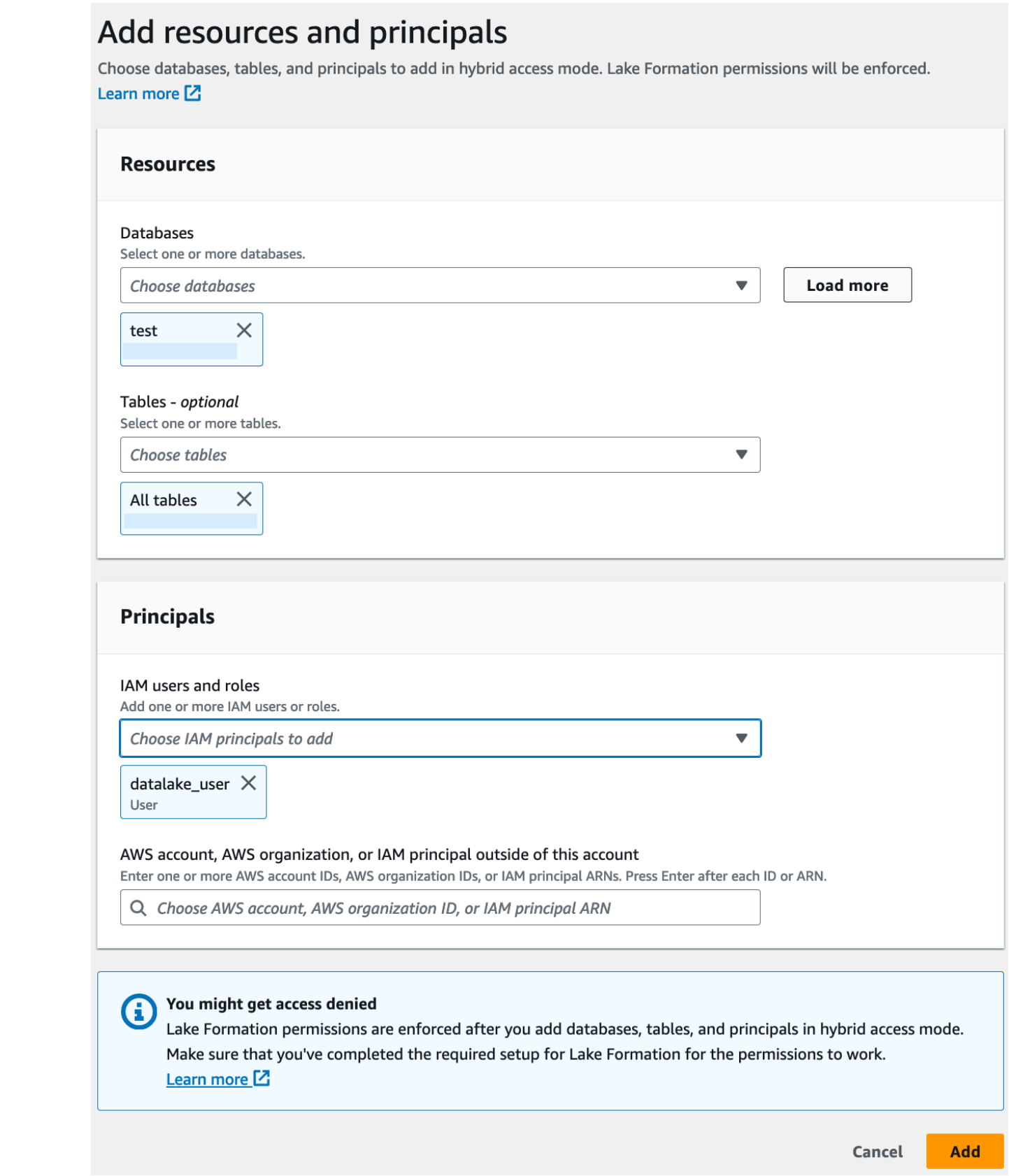
```
        }
    }
```

a. (Optional) Si vous choisissez les balises LF pour octroyer des autorisations, vous pouvez activer les principes pour utiliser les autorisations de Lake Formation lors d'une étape séparée. Vous pouvez le faire en choisissant le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche.

b. Dans la section inférieure de la page du mode d'accès hybride, choisissez Ajouter pour ajouter des ressources et des principes au mode d'accès hybride.

c. Sur la page Ajouter des ressources et des principes, sélectionnez les bases de données et les tables enregistrées en mode d'accès hybride. Choisissez les principaux qui accepteront d'utiliser les autorisations de Lake Formation en mode d'accès hybride.

   Vous pouvez choisir d'autoriser l'accès `All tables` dans une base de données.


Add resources and principals
Choose databases, tables, and principals to add in hybrid access mode. Lake Formation permissions will be enforced.
Learn more
Resources
Databases
Select one or more databases.
Choose databases
Load more
test
Tables - optional
Select one or more tables.
Choose tables
All tables
Principals
IAM users and roles
Add one or more IAM users or roles.
Choose IAM principals to add
datalake_user
User
AWS account, AWS organization, or IAM principal outside of this account
Enter one or more AWS account IDs, AWS organization IDs, or IAM principal ARNs. Press Enter after each ID or ARN.
Choose AWS account, AWS organization ID, or IAM principal ARN
You might get access denied
Lake Formation permissions are enforced after you add databases, tables, and principals in hybrid access mode.
Make sure that you've completed the required setup for Lake Formation for the permissions to work.
Learn more
Cancel
Add

Conversion d'une ressource de la Lake Formation en une ressource hybride

Dans les cas où vous utilisez actuellement les autorisations Lake Formation pour les bases de données et les tables de votre catalogue de données, vous pouvez modifier les propriétés d'enregistrement des emplacements pour activer le mode d'accès hybride. Cela vous permet de fournir aux nouveaux principaux l'accès aux mêmes ressources en utilisant les politiques d'autorisation IAM pour Amazon S3 et les AWS Glue actions sans interrompre les autorisations Lake Formation existantes.

Description du scénario - Les étapes suivantes supposent que vous avez un emplacement de données enregistré auprès de Lake Formation et que vous avez défini des autorisations pour les principaux sur les bases de données, les tables ou les colonnes pointant vers cet emplacement. Si l'emplacement a été enregistré avec un rôle lié à un service, vous ne pouvez pas mettre à jour les paramètres de localisation et activer le mode d'accès hybride. Le `IAMAllowedPrincipals` groupe dispose par défaut de super autorisations sur la base de données et toutes ses tables.

> **Important**
>
> Ne mettez pas à jour l'enregistrement d'un emplacement en mode d'accès hybride sans avoir activé les principaux qui accèdent aux données dans cet emplacement.

Activation du mode d'accès hybride pour un emplacement de données enregistré auprès de Lake Formation

1.
> **Warning**
>
> Nous ne recommandons pas de convertir un emplacement de données géré par Lake Formation en mode d'accès hybride pour éviter d'interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

   Ajoutez les principaux existants qui ont les autorisations de Lake Formation.

   1. Répertoriez et passez en revue les autorisations que vous avez accordées aux principaux sur les bases de données et les tables. Pour plus d'informations, consultez Affichage des autorisations de base de données et de tables dans Lake Formation.

2. Choisissez le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche, puis choisissez Ajouter.
3. Sur la page Ajouter des principes et des ressources, choisissez les bases de données et les tables de l'emplacement de données Amazon S3 que vous souhaitez utiliser en mode d'accès hybride. Choisissez les principaux qui disposent déjà des autorisations de Lake Formation.
4. Choisissez Ajouter pour activer les principes permettant d'utiliser les autorisations de Lake Formation en mode d'accès hybride.

2. Mettez à jour l'enregistrement du bucket ou du préfixe Amazon S3 en choisissant l'option Mode d'accès hybride.

   Console

   1. Connectez-vous à la console Lake Formation en tant qu'administrateur du lac de données.
   2. Dans le volet de navigation, sous Enregistrer et ingérer, sélectionnez Data lake locations.
   3. Sélectionnez un emplacement, puis dans le menu Actions, choisissez Modifier.
   4. Choisissez le mode d'accès hybride.
   5. Choisissez Enregistrer.
   6. Sous Catalogue de données, sélectionnez la base de données ou la table et accordez `Super All` des autorisations au groupe virtuel appelé`IAMAllowedPrincipals`.
   7. Vérifiez que l'accès de vos utilisateurs actuels de Lake Formation n'est pas interrompu lorsque vous avez mis à jour les propriétés d'enregistrement de la localisation. Connectez-vous à la console Athena en tant que responsable de Lake Formation et exécutez un exemple de requête sur une table pointant vers l'emplacement mis à jour.

      De même, vérifiez l'accès des AWS Glue utilisateurs qui utilisent les politiques d'autorisation IAM pour accéder à la base de données et aux tables.

   AWS CLI

   Voici un exemple d'enregistrement d'un emplacement de données auprès de Lake Formation HybridAccessEnabled avec:true/false. La valeur par défaut du `HybridAccessEnabled` paramètre est false. Remplacez le chemin, le nom du rôle et l'identifiant du AWS compte Amazon S3 par des valeurs valides.

   ```
   aws lakeformation update-resource --cli-input-json file://file path
   json:
   ```

```
{
    "ResourceArn": "arn:aws:s3:::<s3-path>",
    "RoleArn": "arn:aws:iam::<123456789012>:role/<test>",
    "HybridAccessEnabled": true
}
```

Partage d'une AWS Glue ressource à l'aide du mode d'accès hybride

Partagez des données avec une autre personne Compte AWS ou un responsable d'un autre en Compte AWS appliquant les autorisations de Lake Formation sans interrompre l'accès basé sur l'IAM des utilisateurs existants du catalogue de données.

Description du scénario - Le compte producteur dispose d'une base de données de catalogue de données dont l'accès est contrôlé à l'aide des principales politiques et AWS Glue actions IAM pour Amazon S3. L'emplacement des données de la base de données n'est pas enregistré auprès de Lake Formation. Le `IAMAllowedPrincipals` groupe dispose par défaut d'`Super`autorisations sur la base de données et sur toutes ses tables.

Octroi d'autorisations entre comptes Lake Formation en mode d'accès hybride

1. Configuration du compte producteur

    1. Connectez-vous à la console Lake Formation à l'aide d'un rôle autorisé par `lakeformation:PutDataLakeSettings` IAM.
    2. Accédez aux paramètres du catalogue de données et choisissez `Version 4` les paramètres de version pour plusieurs comptes.

        Si vous utilisez actuellement la version 1 ou 2, consultez Mise à jour des paramètres de version de partage de données entre comptes les instructions relatives à la mise à jour vers la version 3.

        Aucune modification de la politique d'autorisation n'est requise lors de la mise à niveau de la version 3 vers la version 4.
    3. Enregistrez l'emplacement Amazon S3 de la base de données ou de la table que vous prévoyez de partager en mode d'accès hybride.

4. Vérifiez que l'`Super`autorisation d'accès au `IAMAllowedPrincipals` groupe existe sur les bases de données et les tables dont vous avez enregistré l'emplacement des données en mode d'accès hybride à l'étape ci-dessus.
5. Accordez des autorisations Lake Formation à AWS des organisations, à des unités organisationnelles (UO) ou directement auprès d'un responsable IAM sur un autre compte.
6. Si vous accordez des autorisations directement à un directeur IAM, activez le principal depuis le compte client pour appliquer les autorisations Lake Formation en mode d'accès hybride en activant l'option Rendre les autorisations Lake Formation effectives immédiatement.

   Si vous accordez des autorisations entre comptes à un autre AWS compte, lorsque vous activez le compte, les autorisations de Lake Formation ne sont appliquées qu'aux administrateurs de ce compte. L'administrateur du lac de données du compte destinataire doit répartir les autorisations en cascade et sélectionner les principaux du compte pour appliquer les autorisations de Lake Formation aux ressources partagées en mode d'accès hybride.

   Si vous choisissez l'option Ressources correspondant à des balises LF pour accorder des autorisations entre comptes, vous devez d'abord terminer l'étape d'octroi des autorisations. Vous pouvez activer les principes et les ressources pour le mode d'accès hybride lors d'une étape séparée en choisissant le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche de la console Lake Formation. Choisissez ensuite Ajouter pour ajouter les ressources et les principes auxquels vous souhaitez appliquer les autorisations de Lake Formation.

2. Configuration du compte client

   1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant qu'administrateur de lac de données.
   2. Accédez à https://console.aws.amazon.com/ram et acceptez l'invitation de partage de ressources. L'onglet Partagé avec moi de la AWS RAM console affiche la base de données et les tables partagées avec votre compte.
   3. Créez un lien de ressource vers la base de données et/ou la table partagée dans Lake Formation.
   4. Accordez `Describe` l'autorisation sur le lien vers la ressource et l'`Grant on target`autorisation (sur la ressource partagée d'origine) aux responsables IAM de votre compte (consommateur).
   5. Accordez aux responsables de votre compte des autorisations relatives à Lake Formation sur la base de données ou la table partagée avec vous. Choisissez les principes et les ressources

nécessaires pour appliquer les autorisations de Lake Formation en mode d'accès hybride en activant l'option Rendre les autorisations Lake Formation effectives immédiatement.

6. Testez les autorisations du principal sur la Lake Formation en exécutant des exemples de requêtes Athena. Testez l'accès existant de vos AWS Glue utilisateurs à l'aide des principales politiques et AWS Glue actions IAM pour Amazon S3.

   (Facultatif) Supprimez la politique de compartiment Amazon S3 pour l'accès aux données et les politiques principales IAM AWS Glue et d'accès aux données Amazon S3 pour les principaux que vous avez configurés pour utiliser les autorisations de Lake Formation.

Partage d'une ressource de Lake Formation à l'aide du mode d'accès hybride

Autorisez les nouveaux utilisateurs du catalogue de données d'un compte externe à accéder aux bases de données et aux tables du catalogue de données à l'aide de politiques basées sur l'IAM sans interrompre les autorisations de partage entre comptes existantes de Lake Formation.

Description du scénario - Le compte producteur possède une base de données gérée par Lake Formation et des tables partagées avec un compte externe (consommateur) au niveau du compte ou au niveau principal de l'IAM. L'emplacement des données de la base de données est enregistré auprès de Lake Formation. Le `IAMAllowedPrincipals` groupe ne dispose pas d'`Super`autorisations sur la base de données et ses tables.

Accorder un accès multicompte aux nouveaux utilisateurs du catalogue de données via des politiques basées sur l'IAM sans interrompre les autorisations existantes de Lake Formation

1. Configuration du compte producteur

   1. Connectez-vous à la console Lake Formation à l'aide d'un rôle qui`lakeformation:PutDataLakeSettings`.
   2. Sous Paramètres du catalogue de données, choisissez `Version 4` les paramètres de version multi-comptes.

      Si vous utilisez actuellement la version 1 ou 2, consultez Mise à jour des paramètres de version de partage de données entre comptes les instructions relatives à la mise à jour vers la version 3.

      Aucune modification de la politique d'autorisation n'est requise pour passer de la version 3 à la version 4.

3. Répertoriez les autorisations que vous avez accordées aux principaux sur les bases de données et les tables. Pour plus d'informations, consultez Affichage des autorisations de base de données et de tables dans Lake Formation.
4. Réaccordez les autorisations de comptes croisées de Lake Formation existantes en optant pour les principes et les ressources.

   > **Note**
   >
   > Avant de mettre à jour un enregistrement de localisation de données en mode d'accès hybride afin d'accorder des autorisations entre comptes, vous devez réaccorder au moins un partage de données entre comptes par compte. Cette étape est nécessaire pour mettre à jour les autorisations AWS RAM gérées associées au partage de AWS RAM ressources.
   > En juillet 2023, Lake Formation a mis à jour les autorisations AWS RAM gérées utilisées pour le partage de bases de données et de tables :
   >
   > - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueAllTablesReadWriteForDatabase`(politique de partage au niveau de la base de données)
   > - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueTableReadWrite`(politique de partage au niveau de la table)
   >
   > Les autorisations entre comptes accordées avant juillet 2023 ne sont pas assorties de ces AWS RAM autorisations mises à jour.
   > Si vous avez accordé des autorisations multi-comptes directement aux principaux, vous devez les accorder de nouveau individuellement aux principaux. Si vous ignorez cette étape, les principaux accédant à la ressource partagée risquent de recevoir une erreur de combinaison illégale.

5. Accédez à https://console.aws.amazon.com/ram.
6. L'onglet Shared by me de la AWS RAM console affiche les noms de base de données et de tables que vous avez partagés avec un compte ou un principal externe.

   Assurez-vous que les autorisations associées à la ressource partagée possèdent le bon ARN.
7. Vérifiez que le `Associated` statut des ressources du AWS RAM partage est valide. Si le statut est défini comme tel`Associating`, attendez qu'ils passent à `Associated` l'état. Si

le statut devient le mêmeFailed, arrêtez-vous et contactez l'équipe de service de Lake Formation.

8. Choisissez le mode d'accès hybride sous Autorisations dans la barre de navigation de gauche, puis choisissez Ajouter.
9. La page Ajouter des principes et des ressources affiche les bases de données et/ou les tables, ainsi que les principes auxquels ils ont accès. Vous pouvez effectuer les mises à jour requises en ajoutant ou en supprimant des principes et des ressources.
10. Choisissez les principes dotés des autorisations Lake Formation pour la base de données et les tables que vous souhaitez passer en mode d'accès hybride. Choisissez les bases de données et les tables.
11. Choisissez Ajouter pour activer les principes permettant d'appliquer les autorisations de Lake Formation en mode d'accès hybride.
12. Accordez l'Superautorisation au groupe virtuel IAMAllowedPrincipals sur votre base de données et sur les tables sélectionnées.
13. Modifiez l'enregistrement du site Amazon S3 Lake Formation en mode d'accès hybride.
14. Accordez des autorisations aux AWS Glue utilisateurs du compte externe (consommateur) en utilisant les politiques d'autorisation IAM pour les AWS Glue actions Amazon S3.

2. Configuration du compte client

   1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ en tant qu'administrateur de lac de données.
   2. Accédez à https://console.aws.amazon.com/ram et acceptez l'invitation de partage de ressources. L'onglet Ressources partagées avec moi de la AWS RAM page affiche les noms de base de données et de tables partagés avec votre compte.

      Pour le AWS RAM partage, assurez-vous que l'autorisation attachée possède le bon ARN de l' AWS RAM invitation partagée. Vérifiez si les ressources du AWS RAM partage sont en Associated état. Si le statut est défini comme telAssociating, attendez qu'ils passent à Associated l'état. Si le statut devient le mêmeFailed, arrêtez-vous et contactez l'équipe de service de Lake Formation.
   3. Créez un lien de ressource vers la base de données et/ou la table partagée dans Lake Formation.
   4. Accordez Describe l'autorisation sur le lien vers la ressource et l'Grant on targetautorisation (sur la ressource partagée d'origine) aux responsables IAM de votre compte (consommateur).

5. Configurez ensuite les autorisations de Lake Formation pour les principaux utilisateurs de votre compte sur la base de données ou la table partagée.

   Dans la barre de navigation de gauche, sous Autorisations, choisissez le mode d'accès hybride.

6. Choisissez Ajouter dans la section inférieure de la page du mode d'accès hybride pour activer les principes et la base de données ou la table partagée avec vous depuis le compte du producteur.

7. Accordez des autorisations aux AWS Glue utilisateurs de votre compte en utilisant les politiques d'autorisation IAM pour les AWS Glue actions Amazon S3.

8. Testez les autorisations et AWS Glue autorisations des utilisateurs de Lake Formation en exécutant des exemples de requêtes distincts sur la table à l'aide d'Athena

   (Facultatif) Nettoyez les politiques d'autorisation IAM pour Amazon S3 pour les principaux qui sont en mode d'accès hybride.

Supprimer les principes et les ressources du mode d'accès hybride

Procédez comme suit pour supprimer les bases de données, les tables et les principes du mode d'accès hybride.

Console

1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Sous Autorisations, choisissez le mode d'accès hybride.
3. Sur la page Mode d'accès hybride, cochez la case à côté du nom de la base de données ou de la table et choisissez`Remove`.
4. Un message d'avertissement vous invite à confirmer l'action. Sélectionnez Remove (Supprimer).

   Lake Formation n'applique plus les autorisations pour ces ressources, et l'accès à ces ressources sera contrôlé à l'aide de l'IAM et AWS Glue des autorisations. Cela peut empêcher l'utilisateur d'avoir accès à cette ressource s'il ne dispose pas des autorisations IAM appropriées.

AWS CLI

L'exemple suivant montre comment supprimer des ressources du mode d'accès hybride.

```
aws lakeformation delete-lake-formation-opt-in --cli-input-json file://file path

json:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::<123456789012>:role/role name"
    },
    "Resource": {
        "Table": {
            "CatalogId": "<123456789012>",
            "DatabaseName": "<database name>",
            "Name": "<table name>"
          }
    }
}
```

Affichage des principes et des ressources en mode d'accès hybride

Procédez comme suit pour afficher les bases de données, les tables et les principaux en mode d'accès hybride.

Console

1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Sous Autorisations, choisissez le mode d'accès hybride.
3. La page Mode d'accès hybride affiche les ressources et les principaux qui sont actuellement en mode d'accès hybride.

AWS CLI

L'exemple suivant montre comment répertorier tous les principes et ressources optionnels qui sont en mode d'accès hybride.

```
aws lakeformation list-lake-formation-opt-ins
```

L'exemple suivant montre comment répertorier l'opt-in pour une paire principale-ressource spécifique.

```
aws lakeformation list-lake-formation-opt-ins --cli-input-json file://file path

json:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::<account-id>:role/<role name>"
    },
    "Resource": {
        "Table": {
            "CatalogId": "<account-id>",
            "DatabaseName": "<database name>",
            "Name": "<table name>"
          }
    }
}
```

Ressources supplémentaires

Dans le billet de blog suivant, nous vous expliquons comment intégrer les autorisations Lake Formation en mode d'accès hybride pour certains utilisateurs alors que la base de données est déjà accessible aux autres utilisateurs via les autorisations IAM et Amazon S3. Nous passerons en revue les instructions pour configurer le mode d'accès hybride au sein d'un AWS compte et entre deux comptes.

- Présentation du mode d'accès hybride pour sécuriser l'accès AWS Glue Data Catalog à l'aide de Lake Formation et des politiques IAM et Amazon S3.

Création de tables et de bases de données du catalogue de données

AWS Lake Formation utilise le catalogue de AWS Glue données pour stocker les métadonnées relatives aux lacs de données, aux sources de données, aux transformations et aux cibles. Les métadonnées relatives aux sources de données et aux cibles se présentent sous la forme de bases de données et de tables. Les tables stockent des informations sur les données sous-jacentes, notamment les informations de schéma, les informations de partition et l'emplacement des données. Les bases de données sont des ensembles de tables. Le catalogue de données contient également des liens vers des ressources, qui sont des liens vers des bases de données et des tables partagées dans des comptes externes, et sont utilisés pour l'accès entre comptes aux données du lac de données.

Chaque AWS compte possède un catalogue de données par AWS région.

Rubriques

- Création d’une base de données
- Création de tables
- Utilisation des vues

Création d’une base de données

Les tables de métadonnées du catalogue de données sont stockées dans des bases de données. Vous pouvez créer autant de bases de données que vous le souhaitez, et vous pouvez accorder différentes autorisations Lake Formation pour chaque base de données.

Les bases de données peuvent avoir une propriété d'emplacement facultative. Cet emplacement se trouve généralement dans un site Amazon Simple Storage Service (Amazon S3) enregistré auprès de Lake Formation. Lorsque vous spécifiez un emplacement, les principaux n'ont pas besoin d'autorisations de localisation des données pour créer des tables de catalogue de données qui pointent vers des emplacements au sein de l'emplacement de la base de données. Pour plus d’informations, consultez Underlying data access control.

Pour créer une base de données à l'aide de la console Lake Formation, vous devez être connecté en tant qu'administrateur de lac de données ou créateur de base de données. Un créateur de base de données est un directeur qui a obtenu l'`CREATE_DATABASE`autorisation de Lake

Formation. Vous pouvez consulter la liste des créateurs de bases de données sur la page Rôles et tâches administratifs de la console Lake Formation. Pour consulter cette liste, vous devez disposer de l'autorisation `lakeformation:ListPermissions` IAM et être connecté en tant qu'administrateur de lac de données ou en tant que créateur de base de données avec l'option d'octroi sur l'`CREATE_DATABASE`autorisation.

Pour créer une base de données

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/ et connectez-vous en tant qu'administrateur de lac de données ou créateur de base de données.
2. Dans le volet de navigation, sous Catalogue de données, sélectionnez Bases de données.
3. Choisissez Créer une base de données.
4. Dans la boîte de dialogue Créer une base de données, entrez un nom de base de données, un emplacement facultatif et une description facultative.
5. Sélectionnez éventuellement Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données.

   Pour de plus amples informations sur cette option, veuillez consulter the section called “Modification des paramètres par défaut de votre lac de données”.
6. Choisissez Créer une base de données.

Création de tables

AWS Lake Formation les tables de métadonnées contiennent des informations sur les données du lac de données, notamment des informations sur le schéma, les informations de partition et l'emplacement des données. Ces tables sont stockées dans le catalogue AWS Glue de données. Vous les utilisez pour accéder aux données sous-jacentes du lac de données et pour gérer ces données avec les autorisations de Lake Formation. Les tables sont stockées dans les bases de données du catalogue de données.

Il existe plusieurs méthodes pour créer des tables de catalogue de données :

- Lancez un crawler dedans. AWS Glue Consultez la section Définition des robots d'exploration dans le guide du AWS Glue développeur.
- Créez et exécutez un flux de travail. veuillez consulter the section called “Importation de données à l'aide de workflows”.

- Créez une table manuellement à l'aide de la console Lake Formation, de AWS Glue l'API ou AWS Command Line Interface (AWS CLI).
- Créez un tableau à l'aide de Amazon Athena.
- Créez un lien de ressource vers une table dans un compte externe. veuillez consulter the section called “Création de liens vers des ressources”.

Création de tables Apache Iceberg

AWS Lake Formation prend en charge la création de tables Apache Iceberg qui utilisent le format de données Apache Parquet dans AWS Glue Data Catalog les données résidant dans Amazon S3. Une table du catalogue de données est la définition des métadonnées qui représente les données d'un magasin de données. Par défaut, Lake Formation crée des tables Iceberg v2. Pour connaître la différence entre les tables v1 et v2, consultez la section Modifications de version de format dans la documentation Apache Iceberg.

Apache Iceberg est un format de table ouvert pour les jeux de données analytiques très volumineux. Iceberg permet de modifier facilement votre schéma, également connu sous le nom d'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. Iceberg fournit également un support pour le versionnement des données, ce qui permet aux utilisateurs de suivre les modifications apportées aux données au fil du temps. Cela active la fonction de voyage dans le temps, qui permet aux utilisateurs d'accéder aux versions historiques des données, de les interroger et d'analyser les modifications apportées aux données entre les mises à jour et les suppressions.

Vous pouvez utiliser la console Lake Formation ou l'`CreateTable`opération de l' AWS Glue API pour créer une table Iceberg dans le catalogue de données. Pour plus d'informations, consultez CreateTable action (Python : create_table).

Lorsque vous créez une table Iceberg dans le catalogue de données, vous devez spécifier le format de la table et le chemin du fichier de métadonnées dans Amazon S3 pour pouvoir effectuer des lectures et des écritures.

Vous pouvez utiliser Lake Formation pour sécuriser votre table Iceberg à l'aide d'autorisations de contrôle d'accès précises lorsque vous enregistrez l'emplacement des données Amazon S3 auprès de celui-ci. AWS Lake Formation Pour les données source dans Amazon S3 et les métadonnées qui ne sont pas enregistrées auprès de Lake Formation, l'accès est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions. Pour plus d’informations, consultez Gestion des autorisations relatives à Lake Formation.

> **Note**
>
> Le catalogue de données ne prend pas en charge la création de partitions ni l'ajout de propriétés de table Iceberg.

Rubriques

- Prérequis
- Création d'une table Iceberg

Prérequis

Pour créer des tables Iceberg dans le catalogue de données et configurer les autorisations d'accès aux données de Lake Formation, vous devez remplir les conditions suivantes :

1. Autorisations requises pour créer des tables Iceberg sans les données enregistrées auprès de Lake Formation.

   Outre les autorisations requises pour créer une table dans le catalogue de données, le créateur de la table doit disposer des autorisations suivantes :

   - `s3:PutObject`sur la ressource arn:aws:s3 : :1 {bucketName}
   - `s3:GetObject`sur la ressource arn:aws:s3 : :1 {bucketName}
   - `s3:DeleteObject`sur la ressource arn:aws:s3 : :1 {bucketName}
2. Autorisations requises pour créer des tables Iceberg avec des données enregistrées auprès de Lake Formation :

   Pour utiliser Lake Formation afin de gérer et de sécuriser les données de votre lac de données, enregistrez votre site Amazon S3 contenant les données pour les tables auprès de Lake Formation. Cela permet à Lake Formation de vendre des informations d'identification à AWS des services d'analyse tels qu'Athena, Redshift Spectrum et Amazon EMR pour accéder aux données. Pour plus d'informations sur l'enregistrement d'un site Amazon S3, consultezAjouter un emplacement Amazon S3 à votre lac de données.

   Un directeur qui lit et écrit les données sous-jacentes enregistrées auprès de Lake Formation doit disposer des autorisations suivantes :

   - `lakeformation:GetDataAccess`

- `DATA_LOCATION_ACCESS`

  Un directeur qui possède des autorisations de localisation des données sur un emplacement possède également des autorisations de localisation sur tous les sites enfants.

  Pour plus d'informations sur les autorisations de localisation des données, consultezContrôle d'accès aux données sous-jacent.

Pour activer le compactage, le service doit assumer un rôle IAM autorisé à mettre à jour les tables dans le catalogue de données. Pour plus d'informations, consultez Conditions préalables requises pour l'optimisation des tables .

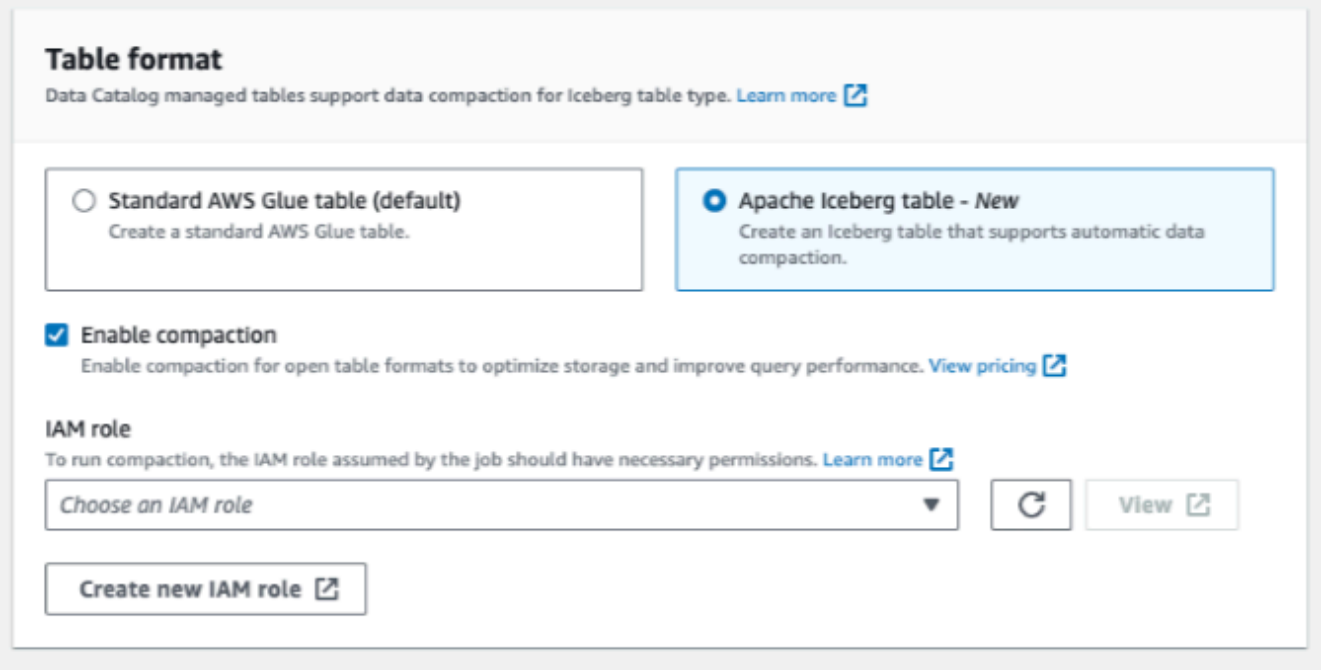
Création d'une table Iceberg

Vous pouvez créer des tables Iceberg v1 et v2 à l'aide de la console Lake Formation ou AWS Command Line Interface comme indiqué sur cette page. Vous pouvez également créer des tables Iceberg à l'aide de AWS Glue la console ou AWS Glue crawler. Pour plus d'informations, consultez la section Data Catalog and Crawlers dans le manuel du AWS Glue développeur.

Pour créer une table Iceberg

Console

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Sous Catalogue de données, choisissez Tables, puis utilisez le bouton Créer une table pour spécifier les attributs suivants :
   - Nom de la table : entrez le nom de la table. Si vous utilisez Athena pour accéder aux tables, suivez ces conseils de dénomination dans le guide de l'utilisateur d'Amazon Athena.
   - Base de données : Choisissez une base de données existante ou créez-en une nouvelle.
   - Description : description de la table. Vous pouvez écrire une description vous aidant à comprendre le contenu de la table.
   - Format de tableau : pour Format de tableau, choisissez Apache Iceberg.



- Activer le compactage : choisissez Activer le compactage pour compacter les petits objets Amazon S3 du tableau en objets plus grands.
- Rôle IAM : pour exécuter le compactage, le service assume un rôle IAM en votre nom. Vous pouvez choisir un rôle IAM à l'aide de la liste déroulante. Assurez-vous que le rôle dispose des autorisations requises pour activer le compactage.

  Pour en savoir plus sur les autorisations requises, consultez Conditions préalables requises pour l'optimisation des tables .
- Emplacement : Spécifiez le chemin d'accès au dossier dans Amazon S3 qui stocke la table de métadonnées. Iceberg a besoin d'un fichier de métadonnées et d'un emplacement dans le catalogue de données pour pouvoir effectuer des lectures et des écritures.
- Schéma : choisissez Ajouter des colonnes pour ajouter des colonnes et les types de données des colonnes. Vous avez la possibilité de créer une table vide et de mettre à jour le schéma ultérieurement. Le catalogue de données prend en charge les types de données Hive. Pour plus d'informations, consultez la section Types de données Hive.

  Iceberg vous permet de faire évoluer le schéma et la partition après avoir créé la table. Vous pouvez utiliser les requêtes Athena pour mettre à jour le schéma de table et les requêtes Spark pour mettre à jour les partitions.

AWS CLI

```
aws glue create-table \
    --database-name iceberg-db \
    --region us-west-2 \
    --open-table-format-input '{
      "IcebergInput": {
           "MetadataOperation": "CREATE",
           "Version": "2"
        }
      }' \
    --table-input '{"Name":"test-iceberg-input-demo",
           "TableType": "EXTERNAL_TABLE",
           "StorageDescriptor":{
              "Columns":[
                  {"Name":"col1", "Type":"int"},
                  {"Name":"col2", "Type":"int"},
                  {"Name":"col3", "Type":"string"}
               ],
              "Location":"s3://DOC_EXAMPLE_BUCKET_ICEBERG/"
           }
       }'
```

Optimisation des tables Iceberg

Les lacs de données Amazon S3 utilisant des formats de table ouverts tels qu'Apache Iceberg stockent les données sous forme d'objets Amazon S3. La présence de milliers de petits objets Amazon S3 dans une table de lac de données augmente la surcharge de métadonnées sur les tables Iceberg et affecte les performances de lecture. Pour améliorer les performances de lecture des services AWS d'analyse tels que Amazon EMR Amazon Athena et les tâches AWS Glue ETL, AWS Glue Data Catalog fournit un compactage géré (un processus qui compacte de petits objets Amazon S3 en objets plus grands) pour les tables Iceberg dans Data Catalog. Vous pouvez utiliser la console, AWS Glue la console ou l' AWS API Lake Formation pour activer ou désactiver le compactage des tables Iceberg individuelles figurant dans le catalogue de données. AWS CLI

L'optimiseur de table surveille en permanence les partitions des tables et lance le processus de compactage lorsque le seuil est dépassé pour le nombre de fichiers et leur taille. Une table Iceberg peut être compactée si la taille du fichier est spécifiée lors de l'écriture. target-file-size-bytes la propriété est comprise entre 128 Mo et 512 Mo. Dans le catalogue de données, le processus de

compactage démarre si la table contient plus de cinq fichiers, chacun représentant une durée d'écriture inférieure à 75 %. target-file-size-bytes propriété.

Par exemple, vous avez une table dont le seuil de taille de fichier est défini à 512 Mo lors de l'écriture. target-file-size-bytes propriété (dans la plage prescrite de 128 Mo à 512 Mo), et la table contient 10 fichiers. Si 6 des 10 fichiers mesurent moins de 384 Mo (0,75 x 512) chacun, le catalogue de données déclenche le compactage.

Le catalogue de données effectue le compactage sans interférer avec les requêtes simultanées. Le catalogue de données prend en charge le compactage des données uniquement pour les tables au format Parquet.

Pour connaître les types de données, les formats de compression et les limitations pris en charge, consultezFormats pris en charge et limites pour le compactage géré des données .

Rubriques

- Conditions préalables requises pour l'optimisation des tables
- Activation du compactage
- Désactivation du compactage
- Affichage des détails de compactage
- Afficher Amazon CloudWatch les métriques
- Suppression d'un optimiseur

Conditions préalables requises pour l'optimisation des tables

L'optimiseur de table assume les autorisations du rôle AWS Identity and Access Management (IAM) que vous spécifiez lorsque vous activez le compactage d'une table. Le rôle IAM doit être autorisé à lire les données et à mettre à jour les métadonnées dans le catalogue de données. Vous pouvez créer un rôle IAM et y attacher les stratégies en ligne suivantes :

- Ajoutez la stratégie en ligne suivante qui accorde à Amazon S3 des autorisations de lecture/écriture sur l'emplacement pour les données qui ne sont pas enregistrées auprès de Lake Formation. Cette politique inclut également des autorisations pour mettre à jour le tableau dans le catalogue de données, ainsi que AWS Glue pour autoriser l'ajout de journaux dans les Amazon CloudWatch journaux et la publication de métriques. Pour les données sources dans Amazon S3 qui ne sont pas enregistrées auprès de Lake Formation, l'accès est déterminé par les stratégies d'autorisation IAM pour Amazon S3 et les actions AWS Glue .

Dans les stratégies en ligne suivantes, remplacez le `bucket-name` par le nom de votre compartiment Amazon S3, `aws-account-id` et `region` par un numéro valide du compte AWS et une région du catalogue de données, `database_name` par le nom de votre base de données et `table_name` par le nom de la table.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::<bucket-name>/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::<bucket-name>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "glue:UpdateTable",
                "glue:GetTable"
            ],
            "Resource": [
                "arn:aws:glue:<region>:<aws-account-id>:table/<database-name>/<table-
name>",
                "arn:aws:glue:<region>:<aws-account-id>:database/<database-name>",
                "arn:aws:glue:<region>:<aws-account-id>:catalog"
            ]
        },
        {
```

```
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": "arn:aws:logs:<region>:<aws-account-id>:log-group:/aws-glue/
iceberg-compaction/logs:*"
        }
    ]
}
```

- Utilisez la stratégie suivante pour activer le compactage des données enregistrées auprès de Lake Formation.

  Si aucune autorisation de `IAM_ALLOWED_PRINCIPALS` groupe n'est accordée au rôle de compactage sur la table, le rôle nécessite Lake Formation `ALTER` `INSERT` et `DELETE` des autorisations sur la table. `DESCRIBE`

  Pour plus d'informations sur l'enregistrement d'un bucket Amazon S3 auprès de Lake Formation, consultezAjouter un emplacement Amazon S3 à votre lac de données.

```
{
 "Version": "2012-10-17",
 "Statement": [
    {
        "Effect": "Allow",
        "Action": [
            "lakeformation:GetDataAccess"
        ],
        "Resource": "*"
    },
    {
        "Effect": "Allow",
        "Action": [
        "glue:UpdateTable",
        "glue:GetTable"
        ],

        "Resource": [
            "arn:aws:glue:<region>:<aws-account-
id>:table/<databaseName>/<tableName>",
            "arn:aws:glue:<region>:<aws-account-id>:database/<database-name>",
```

```
            "arn:aws:glue:<region>:<aws-account-id>:catalog"
        ]
    },
    {
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogGroup",
            "logs:CreateLogStream",
            "logs:PutLogEvents"
        ],
        "Resource": "arn:aws:logs:<region>:<aws-account-id>:log-group:/aws-glue/
iceberg-compaction/logs:*"
        }
    ]
 }
```

- (Facultatif) Pour compacter des tables Iceberg avec des données contenues dans des compartiments Amazon S3 chiffrés à l'aide du chiffrement côté serveur, le rôle de compactage nécessite des autorisations pour déchiffrer les objets Amazon S3 et générer une nouvelle clé de données pour écrire des objets dans les compartiments chiffrés. Ajoutez la politique suivante à la AWS KMS clé souhaitée. Nous prenons uniquement en charge le chiffrement au niveau du compartiment.

```
{
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::<aws-account-id>:role/<compaction-role-name>"
    },
    "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey"
    ],
    "Resource": "*"
}
```

- (Facultatif) Pour l'emplacement des données enregistré auprès de Lake Formation, le rôle utilisé pour enregistrer l'emplacement nécessite des autorisations pour déchiffrer les objets Amazon S3 et générer une nouvelle clé de données pour écrire des objets dans les compartiments chiffrés. Pour plus d’informations, consultez Enregistrement d'un emplacement Amazon S3 chiffré.

- (Facultatif) Si la AWS KMS clé est stockée dans un autre AWS compte, vous devez inclure les autorisations suivantes pour le rôle de compactage.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:GenerateDataKey"
            ],
            "Resource": ["arn:aws:kms:<REGION>:<KEY_OWNER_ACCOUNT_ID>:key/<KEY_ID>" ]
        }
    ]
}
```

- Le rôle que vous utilisez pour exécuter le compactage doit disposer de l'autorisation `iam:PassRole` correspondante.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::<account-id>:role/<compaction-role-name>"
            ]
        }
    ]
}
```

- Ajoutez la politique de confiance suivante au rôle afin que le AWS Glue service assume le rôle IAM pour exécuter le processus de compactage.

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "glue.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

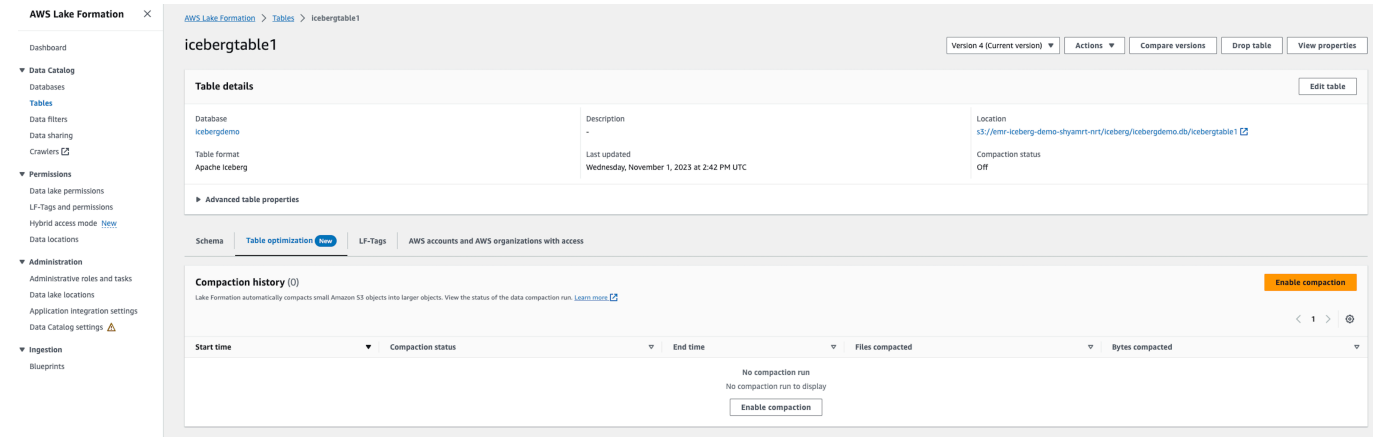
Activation du compactage

Vous pouvez utiliser la console, AWS Glue la console ou l' AWS API Lake Formation pour activer le compactage de vos tables Apache Iceberg dans le catalogue de données. AWS CLI Pour les nouvelles tables, vous pouvez choisir Apache Iceberg comme format de table et activer le compactage lors de la création de la table. Le compactage est désactivé par défaut pour les nouvelles tables.

Console

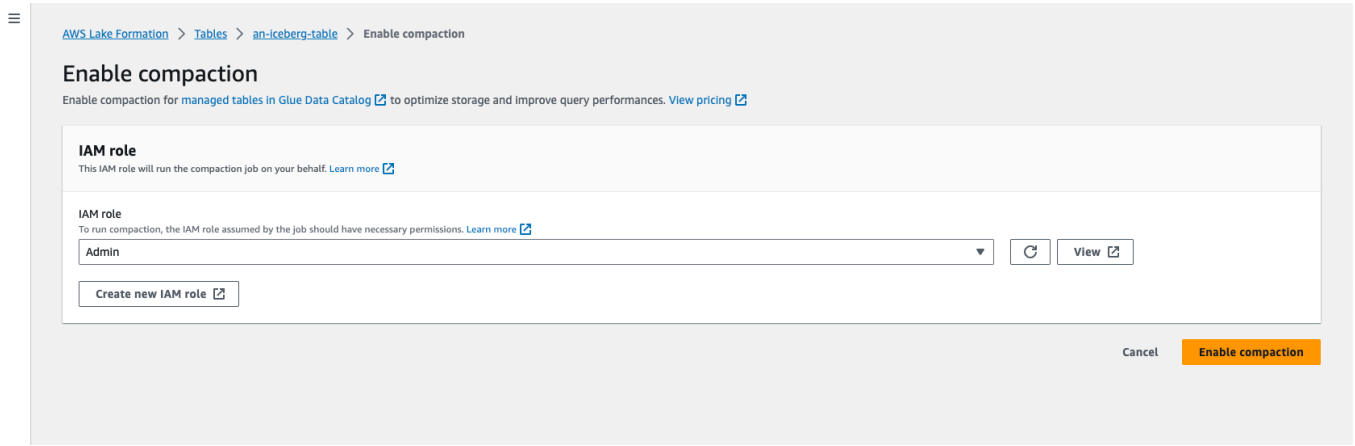
Pour activer le compactage

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/ et connectez-vous en tant qu'administrateur de lac de données, créateur de table ou utilisateur ayant obtenu les `lakeformation:GetDataAccess` autorisations `glue:UpdateTable` et sur la table.
2. Dans le panneau de navigation, sous Catalogue de données, choisissez Tables.
3. Sur la page Tables, choisissez une table au format ouvert pour laquelle vous souhaitez activer le compactage, puis dans le menu Actions, choisissez Activer le compactage.
4. Vous pouvez également activer le compactage en sélectionnant la table et en ouvrant la page Détails de la table. Choisissez l'onglet Optimisation des tables dans la partie inférieure de la page, puis sélectionnez Activer le compactage.



5. Sélectionnez ensuite un rôle IAM existant dans le menu déroulant avec les autorisations indiquées dans la section Conditions préalables requises pour l'optimisation des tables .

   Lorsque vous choisissez l'option Créer un nouveau rôle IAM, le service crée un rôle personnalisé avec des autorisations requises pour exécuter le compactage.



   Suivez les étapes ci-dessous pour mettre à jour un rôle IAM existant :

   a. Pour mettre à jour la stratégie d'autorisation pour le rôle IAM, dans la console IAM, accédez au rôle IAM utilisé pour exécuter le compactage.

   b. Dans la section Ajouter des autorisations, choisissez Créer une stratégie. Dans la fenêtre du navigateur nouvellement ouverte, créez une nouvelle stratégie à utiliser avec votre rôle.

   c. Sur la page Créer une politique, choisissez l'onglet JSON. Copiez le code JSON affiché dans les conditions préalables dans le champ de l'éditeur de politiques.

AWS CLI

L'exemple suivant montre comment activer le compactage. Remplacez l'identifiant de compte par un identifiant de AWS compte valide. Remplacez le nom de la base de données et le nom de la table par un nom réel de la table Iceberg et le nom de la base de données. Remplacez le `roleArn` par le nom de AWS ressource (ARN) du rôle IAM et le nom du rôle IAM disposant des autorisations requises pour exécuter le compactage.

```
aws glue create-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --table-optimizer-configuration
 '{"roleArn":"arn:aws:iam::123456789012:role/compaction_role", "enabled":'true'}' \
  --type compaction
```

AWS API

Appelez une opération `CreateTableOptimizer` pour activer le compactage d'une table.

Après avoir activé le compactage, l'onglet Optimisation des tables affiche les détails de compactage suivants (après environ 15 à 20 minutes) :

L’heure de début

Heure à laquelle le processus de compactage a commencé dans la Lake Formation. La valeur est un horodatage selon le fuseau UTC.

L’heure de fin

Heure à laquelle le processus de compactage s'est terminé dans le catalogue de données. La valeur est un horodatage selon le fuseau UTC.

Statut

État du cycle de compactage. Les valeurs sont la réussite ou l'échec.

Fichiers compactés

Nombre total de fichiers compactés.

Octets compactés

Nombre total d'octets compactés.

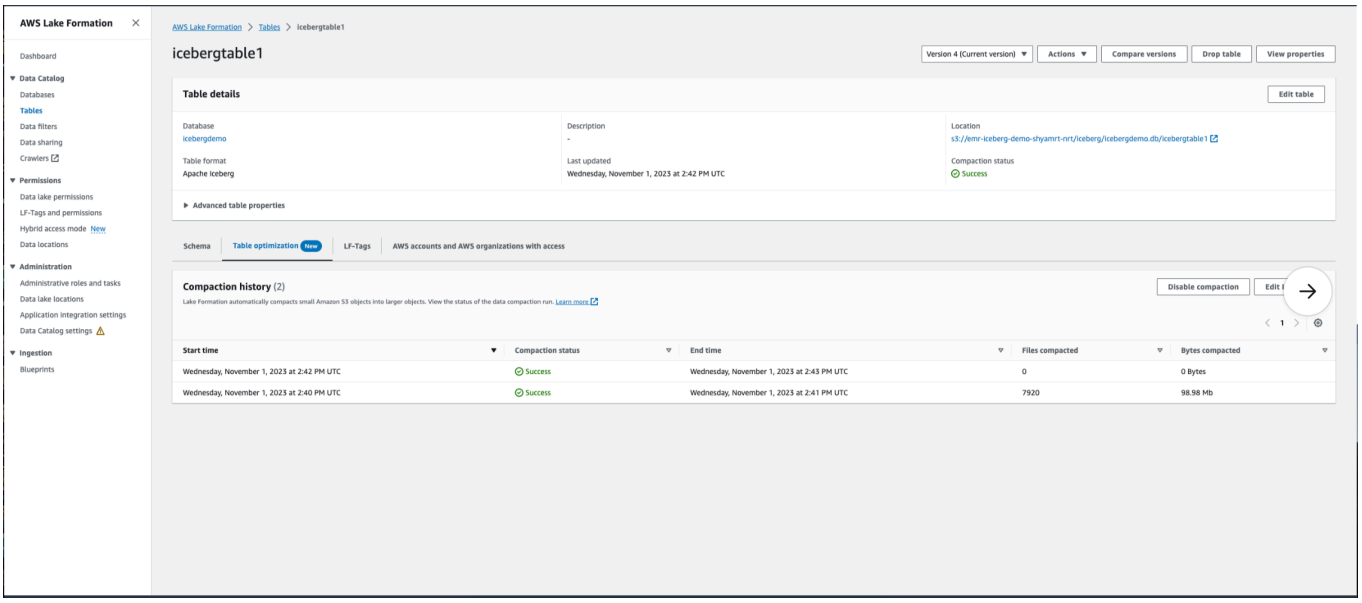
Désactivation du compactage

Vous pouvez désactiver le compactage automatique pour une table Apache Iceberg particulière à l'aide de AWS Glue la console ou. AWS CLI

Console

1. Choisissez Catalogue de données, puis choisissez Tables. Dans la liste des tables, choisissez la table au format ouvert dont vous souhaitez désactiver le compactage.
2. Vous pouvez choisir une table Iceberg, puis choisir Désactiver le compactage sous Actions.

   Vous pouvez également désactiver le compactage de la table en choisissant Désactiver le compactage dans la partie inférieure de la page Détails des tables.



3. Choisissez Désactiver le compactage dans le message de confirmation. Vous pouvez réactiver le compactage ultérieurement.

   Une fois que vous avez confirmé, le compactage est désactivé et l'état de compactage de la table revient à `Off`.

AWS CLI

Dans l'exemple suivant, remplacez l'ID de compte par un ID de AWS compte valide. Remplacez le nom de la base de données et le nom de la table par un nom réel de la table Iceberg et le nom de la base de données. Remplacez le `roleArn` par le nom de AWS ressource (ARN) du rôle IAM et le nom réel du rôle IAM disposant des autorisations requises pour exécuter le compactage.

```
aws glue update-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --table-optimizer-configuration
 '{"roleArn":"arn:aws:iam::123456789012:role/compaction_role", "enabled":'false'}'\
  --type compaction
```

AWS API

UpdateTableOptimizer Opération d'appel pour désactiver le compactage d'une table spécifique.

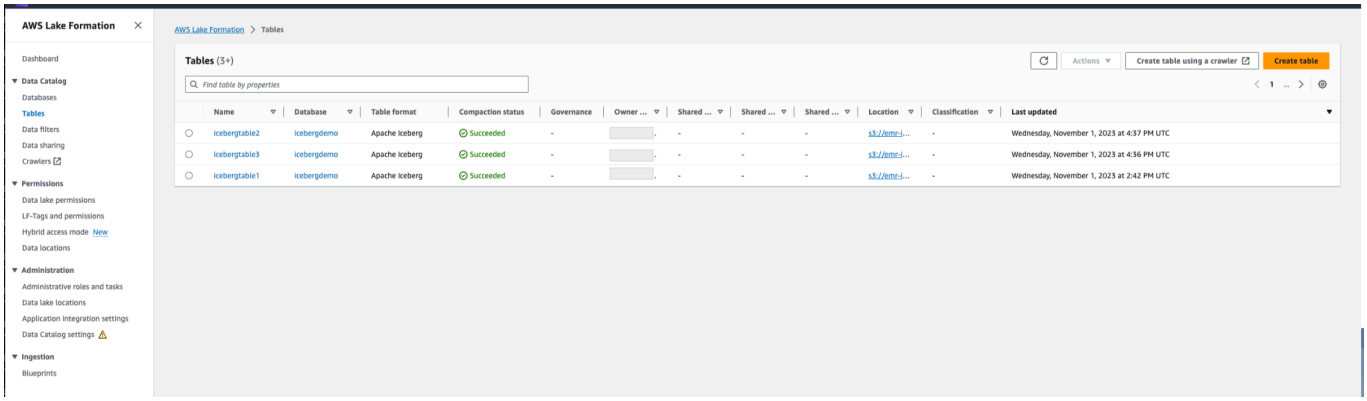
Affichage des détails de compactage

Vous pouvez consulter l'état de compactage d'Apache Iceberg dans la console Lake Formation ou à l'aide des AWS CLI opérations de l' AWS API.

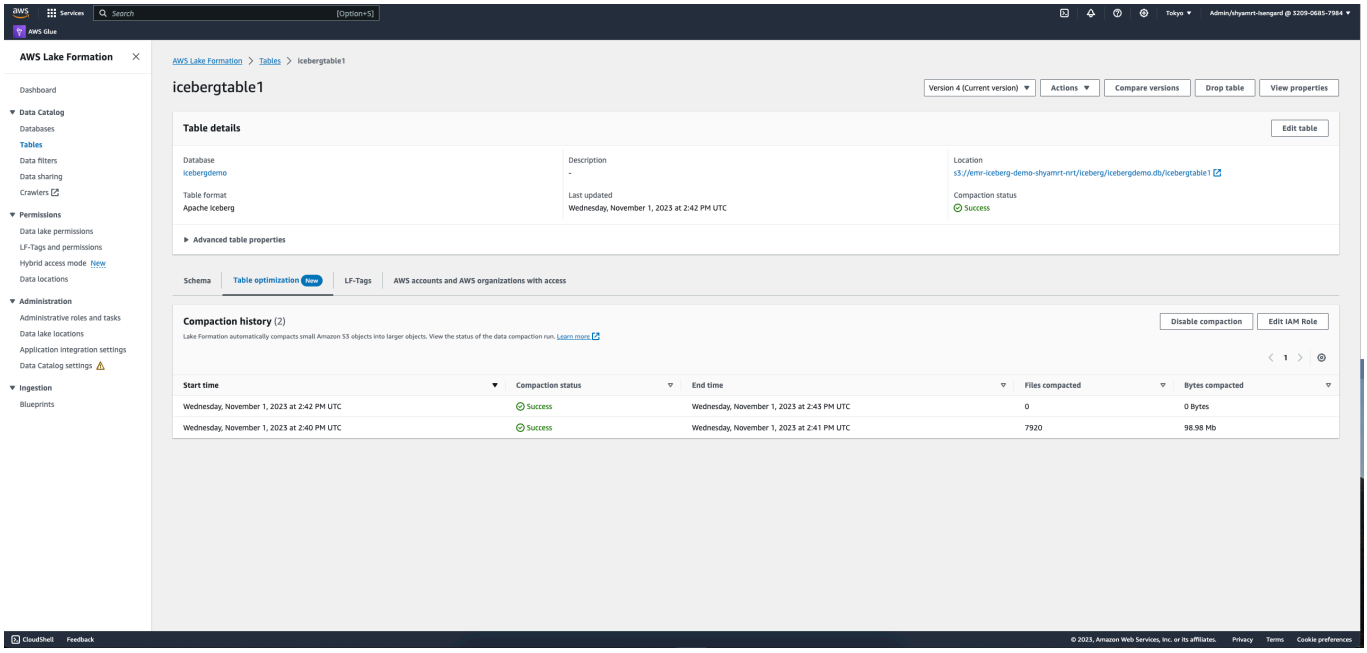
Console

Pour afficher l'état de compactage des tables Iceberg (console)

- Vous pouvez consulter l'état de compactage des tables Iceberg sur la console Lake Formation en choisissant Tables sous Catalogue de données. Le champ État de compactage affiche l'état d'exécution du compactage. Vous pouvez afficher le format de la table et l'état de compactage à l'aide des préférences de table.



- Pour afficher l'historique des opérations de compactage pour une table spécifique, choisissez Tables sous AWS Glue Data Catalog, puis choisissez une table pour afficher les détails de la table. L'onglet Optimisation des tables affiche l'historique du compactage de la table.



AWS CLI

Vous pouvez consulter les détails du compactage à l'aide AWS CLI de.

Dans les exemples suivants, remplacez l'ID de compte par un ID de AWS compte valide, le nom de la base de données et le nom de la table par le nom réel de la table Iceberg.

- Pour obtenir les détails de la dernière exécution du compactage d'une table

```
aws get-table-optimizer \
  --catalog-id 123456789012 \
```

```
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

- Utilisez l'exemple suivant pour récupérer l'historique d'un optimiseur pour une table spécifique.

```
aws list-table-optimizer-runs \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

- L'exemple suivant montre comment récupérer les détails de l'exécution du compactage et de la configuration de plusieurs optimiseurs. Vous pouvez spécifier un maximum de 20 optimiseurs.

```
aws glue batch-get-table-optimizer \
--entries '[{"catalogId":"123456789012", "databaseName":"iceberg_db",
 "tableName":"iceberg_table", "type":"compaction"}]'
```

AWS API

- Utilisez l'opération `GetTableOptimizer` pour récupérer les détails de la dernière exécution d'un optimiseur.
- Utilisez l'opération `ListTableOptimizerRuns` pour récupérer l'historique d'un optimiseur donné sur une table spécifique. Vous pouvez spécifier 20 optimiseurs en un seul appel d'API.
- Utilisez l'opération `BatchGetTableOptimizer` pour récupérer les détails de configuration pour plusieurs optimiseurs de votre compte. Cette opération ne prend pas en charge les appels entre comptes.

Afficher Amazon CloudWatch les métriques

Une fois le compactage effectué avec succès, le service crée des Amazon CloudWatch métriques sur les performances de la tâche de compactage. Vous pouvez accéder aux CloudWatch métriques et choisir Metrics, All metrics. Vous pouvez filtrer les métriques en fonction de l'espace de noms spécifique (par exemple AWS Glue), du nom de la table ou du nom de la base de données.

Pour de plus amples informations, consultez Affichage des métriques disponibles dans le Guide de l'utilisateur Amazon CloudWatch .

- Nombre d'octets compactés
- Nombre de fichiers compactés
- Nombre de DPU alloués à la tâche
- Durée de la tâche (heures)

Suppression d'un optimiseur

Vous pouvez supprimer un optimiseur et les métadonnées associées à la table à l'aide d' AWS CLI une opération d' AWS API.

Exécutez la AWS CLI commande suivante pour supprimer l'historique de compactage d'une table.

```
aws glue delete-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

Utilisez l'opération `DeleteTableOptimizer` pour supprimer un optimiseur pour une table.

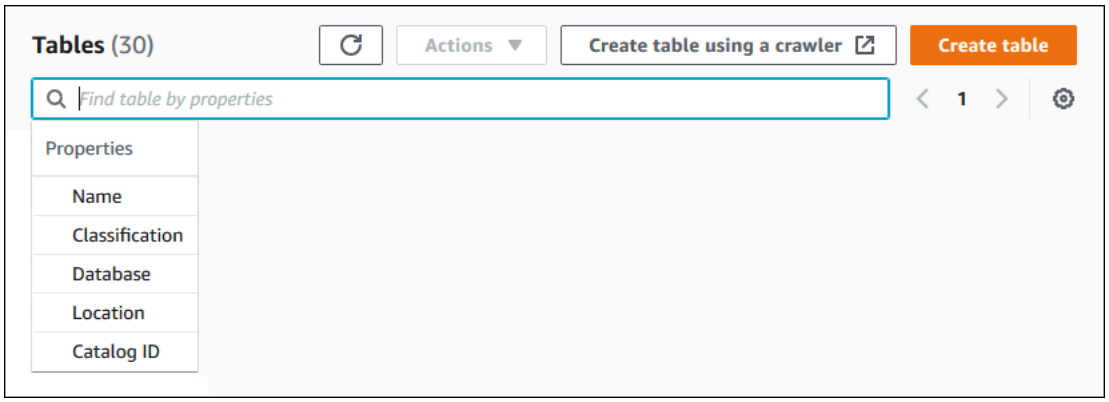
Recherche de tables

Vous pouvez utiliser la AWS Lake Formation console pour rechercher des tables du catalogue de données par nom, emplacement, base de données contenant, etc. Les résultats de recherche n'affichent que les tables pour lesquelles vous êtes autorisé à accéder à Lake Formation.

Pour rechercher des tables (console)

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Dans le volet de navigation, choisissez Tables.
3. Positionnez le curseur dans le champ de recherche en haut de la page. Le champ contient le texte d'espace réservé Rechercher le tableau par propriétés.

Le menu Propriétés apparaît et indique les différentes propriétés de table à partir desquelles effectuer une recherche.



4. Effectuez l'une des actions suivantes :

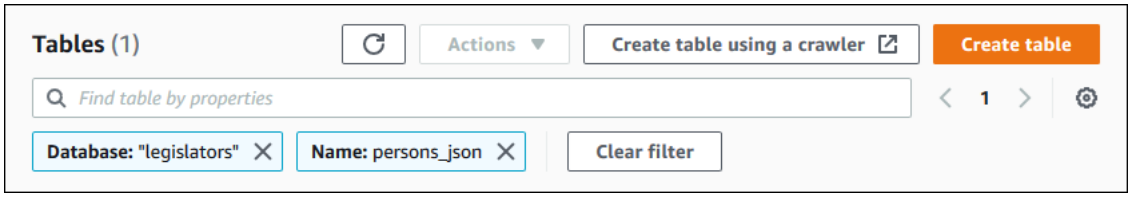
- Recherche par base de données contenant.

  1. Choisissez Base de données dans le menu Propriétés, puis choisissez une base de données dans le menu Bases de données qui apparaît ou tapez un nom de base de données et appuyez sur Entrée.

     Les tables sur lesquelles vous avez des autorisations dans la base de données sont répertoriées.

  2. (Facultatif) Pour réduire la liste à une seule table de la base de données, repositionnez le curseur dans le champ de recherche, choisissez Nom dans le menu Propriétés, puis choisissez un nom de table dans le menu Tables qui apparaît ou tapez un nom de table et appuyez sur Entrée.

     La table unique est répertoriée, et le nom de la base de données et le nom de la table apparaissent sous forme de vignettes sous le champ de recherche.



     Pour ajuster le filtre, fermez l'une des vignettes ou choisissez Effacer le filtre.

- Effectuez une recherche par autres propriétés.

  1. Choisissez une propriété de recherche dans le menu Propriétés.

Pour effectuer une recherche par numéro de AWS compte, choisissez Numéro de catalogue dans le menu Propriétés, entrez un identifiant de AWS compte valide (par exemple, 111122223333), puis appuyez sur Entrée.

Pour effectuer une recherche par lieu, choisissez Emplacement dans le menu Propriétés, puis sélectionnez un lieu dans le menu Emplacements qui apparaît. Toutes les tables situées à l'emplacement racine de l'emplacement sélectionné (par exemple, Amazon S3) sont renvoyées.

Partage des tables et des bases de données du catalogue de données entre les AWS comptes

Vous pouvez partager les ressources du catalogue de données (bases de données et tables) avec AWS des comptes externes en accordant à Lake Formation des autorisations sur les ressources aux comptes externes. Les utilisateurs peuvent ensuite exécuter des requêtes et des tâches qui joignent et interrogent des tables sur plusieurs comptes. Sous réserve de certaines restrictions, lorsque vous partagez une ressource de catalogue de données avec un autre compte, les principaux associés de ce compte peuvent utiliser cette ressource comme si elle figurait dans leur catalogue de données.

Vous ne partagez pas de ressources avec des responsables spécifiques dans des AWS comptes externes ; vous partagez les ressources avec un compte ou une AWS organisation. Lorsque vous partagez une ressource avec une AWS organisation, vous la partagez avec tous les comptes à tous les niveaux de cette organisation. L'administrateur du lac de données de chaque compte externe doit ensuite accorder des autorisations sur les ressources partagées aux principaux de son compte.

Pour plus d’informations, consultez Partage de données entre comptes dans Lake Formation et Octroi et révocation d'autorisations sur les ressources du catalogue de données.

Voir aussi :

- Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci
- Prérequis

Utilisation des vues

Cette fonction est disponible en version préliminaire et susceptible d'être modifiée. Pour plus d'informations, consultez la section Bêtas et aperçus du document Conditions de service AWS.

Dans AWS Glue Data Catalog, une vue est une table virtuelle dont le contenu est défini par une requête qui fait référence à une ou plusieurs tables. Vous pouvez créer une vue qui référence jusqu'à 10 tables à l'aide d'éditeurs SQL pour Amazon Athena, Amazon Redshift ou Amazon EMR. Les tables de référence sous-jacentes d'une vue peuvent appartenir à la même base de données ou à différentes bases de données au sein de la même base de données Compte AWS.

SQL est un langage de programmation utilisé pour interroger des tables, et chaque moteur AWS d'analyse utilise sa propre variante de SQL, ou dialecte SQL. Le catalogue de données permet de créer des vues à l'aide de différents dialectes SQL, à condition que chaque dialecte fasse référence au même ensemble de tables, de colonnes et de types de données. En définissant un schéma de vue commun et un objet de métadonnées que vous pouvez interroger à partir de plusieurs moteurs, les vues du catalogue de données vous permettent d'utiliser des vues uniformes sur l'ensemble de votre lac de données.

Lorsque vous gérez des vues dans le catalogue de données, vous pouvez AWS Lake Formation accorder des autorisations détaillées via la méthode des ressources nommées ou à l'aide de balises LF, et les partager entre les AWS organisations et les Comptes AWS unités organisationnelles. Vous pouvez également partager les vues du catalogue de données entre elles Régions AWS. Cela permet aux utilisateurs de fournir un accès aux données Régions AWS sans dupliquer la source de données.

Pour plus d'informations sur le partage de données entre comptes et l'accès aux données entre régions, voir :

- Partage de données entre comptes dans Lake Formation
- Accès aux tables dans toutes les régions

Vous pouvez utiliser les vues du catalogue de données pour :

- Créez et gérez les autorisations sur un schéma de vue unique. Cela vous permet d'éviter le risque d'autorisations incohérentes sur les vues dupliquées créées dans plusieurs moteurs.
- Accordez des autorisations aux utilisateurs sur une vue qui référence plusieurs tables sans accorder d'autorisations directement sur les tables de référence sous-jacentes.

Pour connaître les limites, voir Considérations et limites relatives aux affichages du catalogue de données

Rubriques

- Conditions préalables à la création de vues
- Création de vues
- Octroi d'autorisations sur les vues du catalogue de données

Conditions préalables à la création de vues

- Pour créer des vues dans Data Catalog, vous devez enregistrer les emplacements de données Amazon S3 sous-jacents des tables de référence auprès de Lake Formation.

  Pour plus de détails sur l'enregistrement des données auprès de Lake Formation, voirAjouter un emplacement Amazon S3 à votre lac de données.
- Le définisseur de vue doit être un rôle IAM. Les autres identités IAM ne peuvent pas créer de vues de catalogue de données.
- Le rôle IAM qui définit la vue doit disposer des autorisations suivantes :
  - `SELECT`Permission complète de Lake Formation avec `Grantable` option sur toutes les tables de référence.
  - Une politique de confiance permettant à Lake Formation et AWS Glue aux services d'assumer ce rôle.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                 ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
```

```
}
```

- Le but : PassRole l'autorisation pour AWS Glue Lake Formation.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerPassRole1",
            "Action": [
                "iam:PassRole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": [
                        "glue.amazonaws.com",
                        "lakeformation.amazonaws.com"
                      ]
                }
            }
        }
    ]
}
```

- AWS Glue et autorisations de Lake Formation.

```
{
    "Version": "2012-10-17",
                 "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "Glue:GetDatabase",
                "Glue:GetDatabases",
                "Glue:CreateTable",
                "Glue:GetTable",
                "Glue:UpdateTable",
                "Glue:DeleteTable",
                "Glue:GetTables",
                "Glue:SearchTables",
```

```
                "Glue:BatchGetPartition",
                "Glue:GetPartitions",
                "Glue:GetPartition",
                "Glue:GetTableVersion",
                "Glue:GetTableVersions",
                "lakeFormation:GetDataAccess",
                "lakeFormation:GetTemporaryTableCredentials",
                "lakeFormation:GetTemporaryGlueTableCredentials",
                "lakeFormation:GetTemporaryUserCredentialsWithSAML"
            ],
            "Resource": "*"
        }
    ]
}
```

- Vous ne pouvez pas créer de vues si la base de données dans laquelle la vue est créée dispose `Super` d'une `ALL` autorisation accordée au `IAMAllowedPrincipals` groupe. Pour révoquer l'`Super`autorisation d'un `IAMAllowedPrincipals` groupe sur une base de données, consultezÉtape 4 : Basculez vos magasins de données vers le modèle d'autorisations de Lake Formation.

  Si vos paramètres de lac de données existants ne vous permettent pas de définir un `IAMAllowedPrincipals` groupe `CreateTableDefaultPermissions` vide, vous pouvez créer une nouvelle base de données et coder le paramètre du lac de données à l'aide de la structure suivante.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        CreateTableDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": []
            }
        ]
```

```
}
```

Création de vues

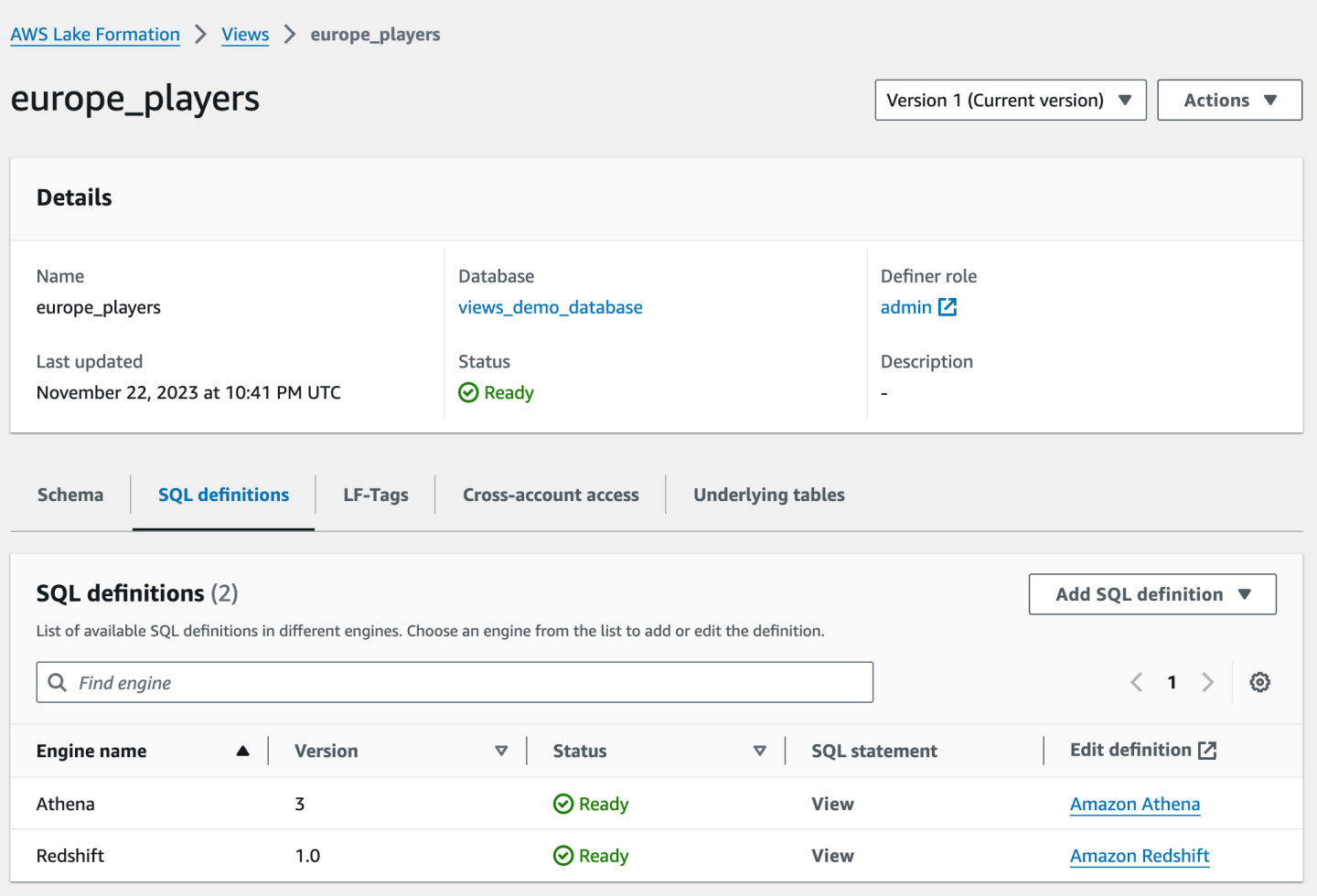
Vous pouvez utiliser les éditeurs SQL pour Athena, Amazon Redshift ou Amazon EMR pour créer des vues dans le. AWS Glue Data Catalog

Pour plus d'informations sur la syntaxe de création et de gestion des vues du catalogue de données, voir :

- Utilisation des AWS Glue Data Catalog vues dans le guide de l'utilisateur d'Amazon Athena.
- Création de vues AWS Glue Data Catalog dans le manuel du développeur de base de données Amazon Redshift.
- Utilisation des AWS Glue Data Catalog vues figurant dans le guide de gestion Amazon EMR.

Après avoir créé une vue du catalogue de données, les détails de la vue apparaissent dans la console Lake Formation.

1. Choisissez Views sous Data Catalog dans la console Lake Formation.
2. La liste des vues disponibles apparaît sur la page des vues.
3. Choisissez une vue dans la liste et la page de détails affiche les attributs de la vue.



Schema

Choisissez une `Column` ligne, puis sélectionnez Modifier les balises LF pour mettre à jour les valeurs des balises ou attribuer de nouvelles balises LF.

Définitions SQL

Vous pouvez consulter la liste des définitions SQL disponibles. Sélectionnez Ajouter une définition SQL, puis choisissez un moteur de requête pour ajouter une définition SQL. Choisissez un moteur de requête (Athena ou Amazon Redshift) `Edit definition` dans la colonne pour mettre à jour une définition SQL.

Étiquettes LF

Choisissez Modifier les balises LF pour modifier les valeurs d'une balise ou attribuer de nouvelles balises. Vous pouvez utiliser les balises LF pour accorder des autorisations sur les vues.

Accès intercomptes

Vous pouvez consulter la Comptes AWS liste des organisations et unités organisationnelles (UO) que vous avez partagées dans la vue du catalogue de données.

Tables sous-jacentes

Les tables sous-jacentes référencées dans la définition SQL utilisée pour créer la vue sont affichées sous cet onglet.

Octroi d'autorisations sur les vues du catalogue de données

Après avoir créé des vues, vous pouvez accorder des autorisations de lac de données sur les vues aux principaux responsables des organisations et des unités organisationnelles. Comptes AWS Pour plus d'informations sur l'octroi d'autorisations, consultez Octroi d'autorisations sur les vues à l'aide de la méthode de ressource nommée .

Importation de données à l'aide de flux de travail dans Lake Formation

Avec AWS Lake Formation, vous pouvez importer vos données à l'aide de flux de travail. Un flux de travail définit la source de données et le calendrier d'importation des données dans votre lac de données. Il s'agit d'un conteneur pour AWS Glue les robots d'exploration, les tâches et les déclencheurs utilisés pour orchestrer les processus de chargement et de mise à jour du lac de données.

Rubriques

- Plans et flux de travail dans Lake Formation
- Création d'un flux de travail
- Exécution d'un flux de travail

Plans et flux de travail dans Lake Formation

Un flux de travail encapsule une activité complexe d'extraction, de transformation et de chargement (ETL) multi-tâches. Les flux de travail génèrent des AWS Glue robots, des tâches et des déclencheurs pour orchestrer le chargement et la mise à jour des données. Lake Formation exécute

et suit un flux de travail en tant qu'entité unique. Vous pouvez configurer un flux de travail pour qu'il s'exécute à la demande ou selon un calendrier.

Les flux de travail que vous créez dans Lake Formation sont visibles dans la AWS Glue console sous la forme d'un graphe acyclique dirigé (DAG). Chaque nœud DAG est une tâche, un robot d'exploration ou un déclencheur. Pour suivre les progrès et résoudre les problèmes, vous pouvez suivre l'état de chaque nœud du flux de travail.

Lorsqu'un flux de travail Lake Formation est terminé, l'utilisateur qui l'a exécuté reçoit l'`SELECT`autorisation Lake Formation sur les tables du catalogue de données créées par le flux de travail.

Vous pouvez également créer des flux de travail dansAWS Glue. Cependant, dans la mesure où Lake Formation vous permet de créer un flux de travail à partir d'un plan, la création de flux de travail est beaucoup plus simple et automatisée dans Lake Formation. Lake Formation fournit les types de plans suivants :

- Instantané de base de données : charge ou recharge les données de toutes les tables dans le lac de données à partir d'une source JDBC. Vous pouvez exclure certaines données de la source selon un modèle d'exclusion.
- Base de données incrémentielle : charge uniquement les nouvelles données dans le lac de données à partir d'une source JDBC, en fonction des signets définis précédemment. Vous spécifiez les tables individuelles à inclure dans la base de données source JDBC. Pour chaque tableau, vous choisissez les colonnes des signets et l'ordre de tri des favoris afin de suivre les données précédemment chargées. La première fois que vous exécutez un plan de base de données incrémentiel sur un ensemble de tables, le flux de travail charge toutes les données des tables et définit des signets pour la prochaine exécution du plan de base de données incrémentiel. Vous pouvez donc utiliser un plan de base de données incrémentiel au lieu du plan de capture de base de données pour charger toutes les données, à condition de spécifier chaque table de la source de données en tant que paramètre.
- Fichier journal : charge en bloc des données à partir de sources de fichiers journaux AWS CloudTrail, notamment les journaux Elastic Load Balancing et les journaux Application Load Balancer.

Utilisez le tableau suivant pour déterminer s'il convient d'utiliser un instantané de base de données ou un plan de base de données incrémentiel.

Utilisez un instantané de base de données lorsque...	Utiliser une base de données incrémentielle lorsque...
• L'évolution du schéma est flexible. (Les colonnes sont renommées, les colonnes précédentes sont supprimées et de nouvelles colonnes sont ajoutées à leur place.) • Une cohérence complète est nécessaire entre la source et la destination.	• L'évolution du schéma est progressive. (Il n'y a que des ajouts successifs de colonnes.) • Seules les nouvelles lignes sont ajoutées ; les lignes précédentes ne sont pas mises à jour.

> **Note**
>
> Les utilisateurs ne peuvent pas modifier les plans et les flux de travail créés par Lake Formation.

Création d'un flux de travail

Avant de commencer, assurez-vous d'avoir accordé au rôle les autorisations de données et les autorisations de localisation des données requises`LakeFormationWorkflowRole`. Le flux de travail peut ainsi créer des tables de métadonnées dans le catalogue de données et écrire des données vers des emplacements cibles dans Amazon S3. Pour plus d'informations, consultez (Facultatif) Créez un rôle IAM pour les flux de travail et Vue d'ensemble des autorisations relatives à Lake Formation .

> **Note**
>
> Lake Formation utilise `GetTemplateInstanceGetTemplateInstances`, et effectue des `InstantiateTemplate` opérations pour créer des flux de travail à partir de plans. Ces opérations ne sont pas accessibles au public et ne sont utilisées qu'en interne pour créer des ressources en votre nom. Vous recevez des CloudTrail événements pour créer des flux de travail.

Pour créer un flux de travail à partir d'un plan

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant d'autorisations d'ingénieur de données. Pour plus d'informations, consultez Référence des personnalités de Lake Formation et des autorisations IAM.
2. Dans le volet de navigation, choisissez Blueprints, puis choisissez Use Blueprint.
3. Sur la page Utiliser un plan, choisissez une vignette pour sélectionner le type de plan.
4. Sous Source d'importation, spécifiez la source de données.

   Si vous effectuez une importation à partir d'une source JDBC, spécifiez les éléments suivants :

   - Connexion à la base de données —Choisissez une connexion dans la liste. Créez des connexions supplémentaires à l'aide de la AWS Glue console. Le nom d'utilisateur et le mot de passe JDBC de la connexion déterminent les objets de base de données auxquels le flux de travail a accès.
   - Chemin des données source : entrez<database>/<schema>/<table>ou<database>/<table>, selon le produit de base de données. Oracle Database et MySQL ne prennent pas en charge le schéma dans le chemin. Vous pouvez remplacer le caractère pourcentage (%) par *<schema>* ou *<table>*. Par exemple, pour une base de données Oracle dont l'identifiant système (SID) est égal `orcl/%` à`orcl`, entrez pour importer toutes les tables auxquelles l'utilisateur nommé dans la connexion a accès.

     > **Important**
     >
     > Ce champ distingue les majuscules et minuscules. Le flux de travail échouera s'il existe une incompatibilité majuscules/minuscules pour l'un des composants.

     Si vous spécifiez une base de données MySQL, AWS Glue ETL utilise le pilote JDBC Mysql5 par défaut, donc MySQL8 n'est pas supporté nativement. Vous pouvez modifier le script de tâche ETL pour utiliser un `customJdbcDriverS3Path` paramètre tel que décrit dans la section JDBC ConnectionType Values du manuel du AWS Glue développeur afin d'utiliser un autre pilote JDBC compatible avec MySQL8.

   Si vous effectuez une importation à partir d'un fichier journal, assurez-vous que le rôle que vous spécifiez pour le flux de travail (le « rôle de flux de travail ») dispose des autorisations IAM

requises pour accéder à la source de données. Par exemple, pour importer AWS CloudTrail des journaux, l'utilisateur doit disposer des `cloudtrail:LookupEvents` autorisations `cloudtrail:DescribeTrails` et pour consulter la liste des CloudTrail journaux lors de la création du flux de travail, et le rôle du flux de travail doit disposer d'autorisations sur l' CloudTrail emplacement dans Amazon S3.

5. Effectuez l'une des actions suivantes :

   - Pour le type de plan instantané de base de données, identifiez éventuellement un sous-ensemble de données à importer en spécifiant un ou plusieurs modèles d'exclusion. Ces modèles d'exclusion sont des modèles de style Unix. `glob` Ils sont stockés en tant que propriété des tables créées par le flux de travail.

     Pour plus de détails sur les modèles d'exclusion disponibles, consultez la section Include et d'exclusion des modèles dans le manuel du AWS Glue développeur.

   - Pour le type de plan de base de données incrémentiel, spécifiez les champs suivants. Ajoutez une ligne pour chaque table à importer.

     Nom de la table

     Tableau à importer. Tout doit être en minuscules.

     Clés de signet

     Liste de noms de colonnes séparés par des virgules qui définissent les clés des signets. Si ce champ est vide, la clé primaire est utilisée pour déterminer les nouvelles données. Le cas de chaque colonne doit correspondre au majuscule défini dans la source de données.

     > **Note**
     >
     > La clé primaire est considérée comme la clé de signet par défaut uniquement si elle augmente ou diminue de manière séquentielle (sans interruption). Si vous souhaitez utiliser la clé primaire comme clé de signet et qu'elle comporte des lacunes, vous devez nommer la colonne de clé primaire comme clé de signet.

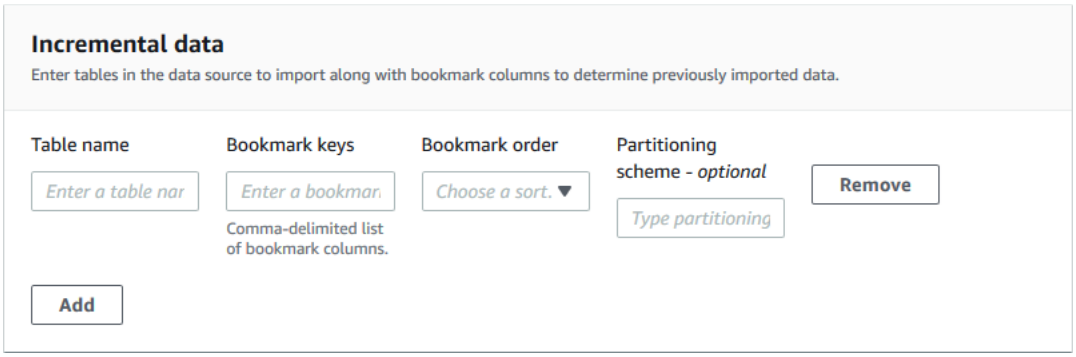
     Ajouter une commande à vos favoris

     Lorsque vous choisissez Ascending, les lignes dont les valeurs sont supérieures aux valeurs enregistrées dans les favoris sont identifiées comme de nouvelles lignes. Lorsque

vous choisissez Décroissant, les lignes dont les valeurs sont inférieures aux valeurs enregistrées dans les favoris sont identifiées comme de nouvelles lignes.

Schéma de partitionnement

(Facultatif) Liste des colonnes clés de partitionnement, délimitées par des barres obliques (/). Exemple : `year/month/day`.



Pour plus d'informations, consultez la section Suivi des données traitées à l'aide des signets de tâches dans le Guide du AWS Glue développeur.

6. Sous Cible d'importation, spécifiez la base de données cible, l'emplacement Amazon S3 cible et le format des données.

   Assurez-vous que le rôle de flux de travail dispose des autorisations Lake Formation requises sur la base de données et sur l'emplacement cible Amazon S3.

   > **Note**
   >
   > À l'heure actuelle, les plans ne prennent pas en charge le chiffrement des données sur la cible.

7. Choisissez une fréquence d'importation.

   Vous pouvez spécifier une `cron` expression à l'aide de l'option Personnalisée.

8. Sous Options d'importation :

   a. Entrez un nom de flux de travail.

   b. Pour rôle, choisissez le rôle `LakeFormationWorkflowRole` que vous avez créé dans(Facultatif) Créez un rôle IAM pour les flux de travail.

   c. Spécifiez éventuellement un préfixe de table. Le préfixe est ajouté aux noms des tables du catalogue de données créées par le flux de travail.

9. Choisissez Créer et attendez que la console indique que le flux de travail a été créé avec succès.

> **Tip**
>
> Avez-vous reçu le message d'erreur suivant ?
> `User: arn:aws:iam::<account-id>:user/<username> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/<rolename>...`
> Si tel est le cas, vérifiez que vous l'avez <account-id>remplacé par un numéro de AWS compte valide dans toutes les polices.

> **Voir aussi :**
>
> - Plans et flux de travail dans Lake Formation

Exécution d'un flux de travail

Vous pouvez exécuter un flux de travail à l'aide de la console Lake Formation, de la AWS Glue console, de l'interface de ligne de AWS Glue commande (AWS CLI) ou de l'API.

Pour exécuter un flux de travail (console Lake Formation)

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur disposant d'autorisations d'ingénieur de données. Pour plus d'informations, consultez Référence des personnalités de Lake Formation et des autorisations IAM.
2. Dans le panneau de navigation, sélectionnez Blueprints (Modèles).
3. Sur la page Blueprints, sélectionnez le flux de travail. Ensuite, dans le menu Actions, choisissez Démarrer.
4. Au fur et à mesure que le flux de travail s'exécute, consultez sa progression dans la colonne État de la dernière exécution. Cliquez sur le bouton d'actualisation de temps en temps.

   Le statut passe de « EN COURS » à « Découverte », puis « Importation », puis « TERMINÉ ».

   Lorsque le flux de travail est terminé :

- Le catalogue de données contient de nouvelles tables de métadonnées.
- Vos données sont ingérées dans le lac de données.

Si le flux de travail échoue, procédez comme suit :

a. Sélectionnez un flux de travail. Choisissez Actions, puis Afficher le graphique.

   Le flux de travail s'ouvre dans la AWS Glue console.

b. Assurez-vous que le flux de travail est sélectionné, puis sélectionnez l'onglet History (Historique).

c. Sous Historique, sélectionnez l'exécution la plus récente et choisissez Afficher les détails de la course.

d. Sélectionnez une tâche ou un robot d'exploration ayant échoué dans le graphique dynamique (d'exécution) et consultez le message d'erreur. Les nœuds défaillants sont rouges ou jaunes.

Voir aussi :

- Plans et flux de travail dans Lake Formation

Gestion des autorisations relatives à Lake Formation

Lake Formation fournit des contrôles d'accès centralisés aux données de votre lac de données. Vous pouvez définir des règles basées sur des politiques de sécurité pour vos utilisateurs et applications par rôle dans Lake Formation, et l'intégration permet d' AWS Identity and Access Management authentifier ces utilisateurs et ces rôles. Une fois les règles définies, Lake Formation applique vos contrôles d'accès au niveau de la table et de la granularité au niveau des colonnes pour les utilisateurs d'Amazon Redshift Spectrum et Amazon Athena.

Rubriques

- Octroi d'autorisations de localisation des données
- Octroi et révocation d'autorisations sur les ressources du catalogue de données
- Exemple de scénario d'autorisations
- Filtrage des données et sécurité au niveau des cellules dans Lake Formation
- Affichage des autorisations de base de données et de tables dans Lake Formation
- Révocation de l'autorisation à l'aide de la console Lake Formation
- Partage de données entre comptes dans Lake Formation
- Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci
- Création de liens vers des ressources
- Accès aux tables dans toutes les régions

Octroi d'autorisations de localisation des données

Les autorisations de localisation des données AWS Lake Formation permettent aux principaux de créer et de modifier les ressources du catalogue de données qui pointent vers des sites Amazon S3 enregistrés désignés. Les autorisations de localisation des données s'ajoutent aux autorisations relatives aux données de Lake Formation pour sécuriser les informations contenues dans votre lac de données.

Lake Formation n'utilise pas le service AWS Resource Access Manager (AWS RAM) pour octroyer des autorisations de localisation des données. Vous n'avez donc pas besoin d'accepter des invitations à partager des ressources pour obtenir des autorisations de localisation des données.

Vous pouvez accorder des autorisations de localisation des données à l'aide de la console Lake Formation, de l'API ou AWS Command Line Interface (AWS CLI).

> **Note**
>
> Pour qu'une subvention soit acceptée, vous devez d'abord enregistrer l'emplacement des données auprès de Lake Formation.

> **Voir aussi :**
>
> - Underlying data access control

Rubriques

- Octroi d'autorisations de localisation des données (même compte)
- Octroi d'autorisations de localisation des données (compte externe)
- Octroi d'autorisations sur un emplacement de données partagé avec votre compte

Octroi d'autorisations de localisation des données (même compte)

Suivez ces étapes pour accorder des autorisations de localisation des données aux principaux de votre AWS compte. Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

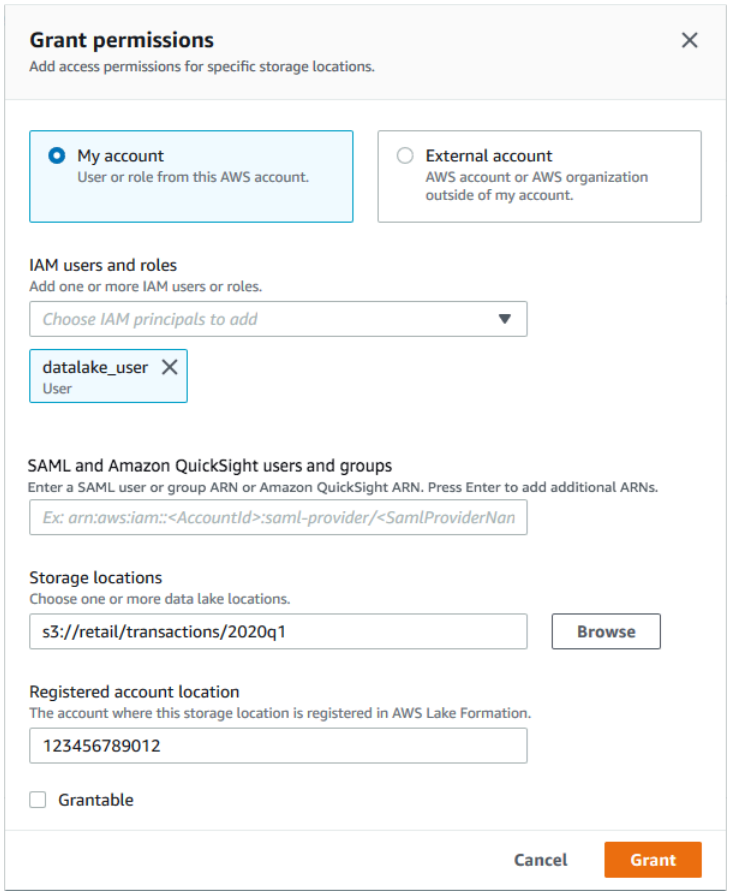
Pour accorder des autorisations de localisation des données (même compte, même console)

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant que principal ayant accordé des autorisations sur l'emplacement de données souhaité.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données.
3. Choisissez Grant (Accorder).
4. Dans la boîte de dialogue Octroyer des autorisations, assurez-vous que la vignette Mon compte est sélectionnée. Fournissez ensuite les informations suivantes :

    - Pour les utilisateurs et les rôles IAM, choisissez un ou plusieurs principaux.

- Pour les QuickSight utilisateurs et les groupes SAML et Amazon, entrez un ou plusieurs Amazon Resource Names (ARN) pour les utilisateurs ou groupes fédérés via SAML ou des ARN pour les utilisateurs ou les groupes Amazon. QuickSight

  Entrez un ARN à la fois, puis appuyez sur Entrée après chaque ARN. Pour plus d'informations sur la façon de construire les ARN, consultezLake Formation accorde et AWS CLI révoque des commandes.
- Pour les emplacements de stockage, choisissez Browse, puis choisissez un emplacement de stockage Amazon Simple Storage Service (Amazon S3). L'emplacement doit être enregistré auprès de Lake Formation. Choisissez à nouveau Parcourir pour ajouter un autre emplacement. Vous pouvez également saisir le lieu, mais assurez-vous de le faire précéder de. `s3://`
- Dans Emplacement du compte enregistré, entrez le numéro du AWS compte sur lequel le point de vente est enregistré. Il s'agit par défaut de votre identifiant de compte. Dans un scénario multicompte, les administrateurs du lac de données d'un compte destinataire peuvent spécifier ici le compte propriétaire lorsqu'ils accordent l'autorisation de localisation des données aux autres principaux du compte destinataire.
- (Facultatif) Pour permettre aux principaux sélectionnés d'accorder des autorisations de localisation des données sur l'emplacement sélectionné, sélectionnez Accordable.



5. Choisissez Grant (Accorder).

Pour accorder des autorisations de localisation des données (même compte, AWS CLI)

- Exécutez une `grant-permissions` commande et `DATA_LOCATION_ACCESS` accordez-la au principal, en spécifiant le chemin Amazon S3 comme ressource.

  Example

  L'exemple suivant accorde des autorisations de localisation des données `s3://retail` à l'utilisateur`datalake_user1`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail"}}'
```

Example

L'exemple suivant accorde des autorisations de localisation des données `s3://retail` à `ALLIAMPrincipals` un groupe.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "111122223333"}}'
```

Voir aussi :

- Référence des autorisations de Lake Formation

Octroi d'autorisations de localisation des données (compte externe)

Suivez ces étapes pour accorder des autorisations de localisation des données à un AWS compte ou à une organisation externe.

Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

Avant de commencer

Assurez-vous que toutes les conditions d'accès entre comptes sont satisfaites. Pour plus d'informations, consultez Prérequis.

Pour accorder des autorisations de localisation des données (compte externe, console)

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Emplacements des données, puis choisissez Grant.
3. Dans la boîte de dialogue Octroyer des autorisations, choisissez la vignette Compte externe.
4. Saisissez les informations suivantes :

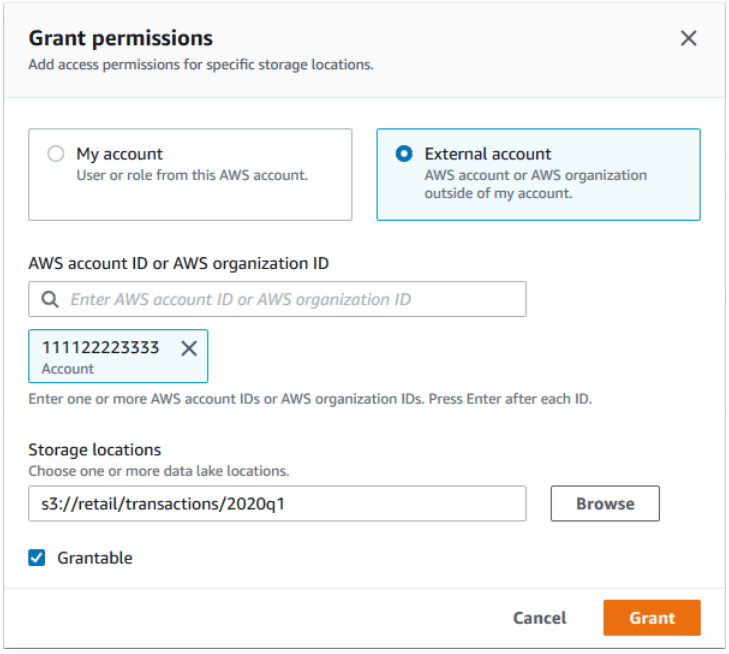
- Pour l'ID de AWS compte ou AWS l'ID d'organisation, entrez des numéros de AWS compte, des identifiants d'organisation ou des identifiants d'unité organisationnelle valides.

  Appuyez sur Entrée après chaque identifiant.

  Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

  Un identifiant d'unité organisationnelle se compose de « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (l'identifiant de la racine qui contient l'unité d'organisation). Cette chaîne est suivie d'un deuxième « - » (trait d'union) et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

- Sous Emplacements de stockage, choisissez Browse, puis choisissez un emplacement de stockage Amazon Simple Storage Service (Amazon S3). L'emplacement doit être enregistré auprès de Lake Formation.



5. Sélectionnez Grantable.
6. Choisissez Grant (Accorder).

Pour accorder des autorisations de localisation des données (compte externe, AWS CLI)

- Pour accorder des autorisations à un AWS compte externe, entrez une commande similaire à la suivante.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333  --permissions "DATA_LOCATION_ACCESS"
 --permissions-with-grant-option "DATA_LOCATION_ACCESS" --resource
 '{ "DataLocation": {"CatalogId":"123456789012","ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

Cette commande octroie `DATA_LOCATION_ACCESS` avec l'option grant le compte 1111-2222-3333 sur le site Amazon S3`s3://retail/transactions/2020q1`, qui appartient au compte 1234-5678-9012.

Pour accorder des autorisations à une organisation, entrez une commande similaire à la suivante.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "DATA_LOCATION_ACCESS" --permissions-
with-grant-option "DATA_LOCATION_ACCESS" --resource '{"DataLocation":
 {"CatalogId":"123456789012","ResourceArn":"arn:aws:s3::retail/
transactions/2020q1"}}'
```

Cette commande accorde `DATA_LOCATION_ACCESS` une option d'autorisation à l'organisation `o-abcdefghijkl` sur le site Amazon S3`s3://retail/transactions/2020q1`, qui appartient au compte 1234-5678-9012.

Pour accorder des autorisations à un mandant sur un AWS compte externe, entrez une commande similaire à la suivante.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3::retail/transactions/2020q1", "CatalogId":
 "123456789012"}}'
```

Cette commande est accordée DATA_LOCATION_ACCESS à un mandant du compte 1111-2222-3333 sur le site Amazon S3s3://retail/transactions/2020q1, qui appartient au compte 1234-5678-9012.

Example

L'exemple suivant accorde des autorisations de localisation des données s3://retail pour les ALLIAMPrincipals regrouper dans un compte externe.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "123456789012"}}'
```

Voir aussi :

- Référence des autorisations de Lake Formation

Octroi d'autorisations sur un emplacement de données partagé avec votre compte

Une fois qu'une ressource de catalogue de données est partagée avec votre AWS compte, en tant qu'administrateur du lac de données, vous pouvez accorder des autorisations sur la ressource aux autres principaux de votre compte. Si l'ALTERautorisation est accordée sur une table partagée et que la table pointe vers un emplacement Amazon S3 enregistré, vous devez également accorder des autorisations de localisation des données sur cet emplacement. De même, si l'ALTERautorisation CREATE_TABLE ou est accordée sur une base de données partagée et que la base de données possède une propriété d'emplacement qui pointe vers un emplacement enregistré, vous devez également accorder des autorisations de localisation des données sur cet emplacement.

Pour accorder des autorisations de localisation des données sur un emplacement partagé à un responsable de votre compte, votre compte doit avoir obtenu l'DATA_LOCATION_ACCESSautorisation sur le lieu avec l'option d'octroi. Lorsque vous accordez ensuite DATA_LOCATION_ACCESS l'autorisation à un autre principal sur votre compte, vous devez inclure l'ID du catalogue de données

(ID de AWS compte) du compte propriétaire. Le compte propriétaire est le compte qui a enregistré l'emplacement.

Vous pouvez utiliser la AWS Lake Formation console, l'API ou le AWS Command Line Interface (AWS CLI pour accorder des autorisations de localisation des données.

Pour accorder des autorisations sur un emplacement de données partagé avec votre compte (console)

- Suivez les étapes de Octroi d'autorisations de localisation des données (même compte).

  Pour les emplacements de stockage, vous devez saisir les emplacements. Dans Emplacement du compte enregistré, entrez le AWS numéro de compte du propriétaire.

Pour accorder des autorisations sur un emplacement de données partagé avec votre compte (AWS CLI)

- Entrez l'une des commandes suivantes pour accorder des autorisations à un utilisateur ou à un rôle.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"CatalogId":"<owner-account-ID>","ResourceArn":"arn:aws:s3:::<s3-location>"}}'
```

Octroi et révocation d'autorisations sur les ressources du catalogue de données

Vous pouvez accorder des autorisations de lac de données aux principaux AWS Lake Formation afin que ceux-ci puissent créer et gérer les ressources du catalogue de données et accéder aux données sous-jacentes. Vous pouvez accorder des autorisations Data Lake sur les bases de données, les tables et les vues. Lorsque vous accordez des autorisations sur des tables, vous pouvez limiter l'accès à des colonnes ou à des lignes de table spécifiques pour un contrôle d'accès encore plus précis.

Vous pouvez accorder des autorisations sur des tables et des vues individuelles, ou avec une seule opération d'autorisation, vous pouvez accorder des autorisations sur toutes les tables et vues d'une base de données. Si vous accordez des autorisations sur toutes les tables d'une base de données, vous les accordez implicitement sur la `DESCRIBE` base de données. La base de données apparaît ensuite sur la page Bases de données de la console et est renvoyée par l'opération `GetDatabases` d'API.

Vous pouvez accorder des autorisations en utilisant la méthode des ressources nommées ou la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC).

Vous pouvez accorder des autorisations aux directeurs d'un même compte ou d'une organisation Compte AWS ou à un compte externe. Lorsque vous accordez des subventions à des comptes ou à des organisations externes, vous partagez des ressources que vous possédez avec ces comptes ou organisations. Les responsables de ces comptes ou organisations peuvent ensuite accéder aux ressources du catalogue de données dont vous êtes propriétaire et aux données sous-jacentes.

> **Note**
>
> Actuellement, la méthode LF-TBAC prend en charge l'octroi d'autorisations entre comptes aux principaux, aux organisations et aux unités organisationnelles (Comptes AWS UO) IAM.

Lorsque vous accordez des autorisations à des comptes ou à des organisations externes, vous devez inclure l'option d'octroi. Seul l'administrateur du lac de données du compte externe peut accéder aux ressources partagées jusqu'à ce qu'il accorde des autorisations sur les ressources partagées aux autres principaux du compte externe.

Vous pouvez accorder des autorisations au catalogue de données à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

> **Note**
>
> Lorsque vous supprimez une ressource de catalogue de données, toutes les autorisations associées à cette ressource ne sont plus valides. Le fait de recréer la même ressource avec le même nom ne récupérera pas les autorisations de Lake Formation. Les utilisateurs devront à nouveau configurer de nouvelles autorisations.

Voir aussi :

- Partage des tables et des bases de données du catalogue de données entre les AWS comptes
- Contrôle d'accès aux métadonnées
- Référence des autorisations de Lake Formation

Autorisations IAM requises pour accorder ou révoquer les autorisations de Lake Formation

Tous les principaux, y compris l'administrateur du lac de données, ont besoin des autorisations AWS Identity and Access Management (IAM) suivantes pour accorder ou révoquer les autorisations de catalogue de AWS Lake Formation données ou les autorisations de localisation des données avec l'API Lake Formation ou le : AWS CLI

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`
- `glue:GetTable`ou `glue:GetDatabase` pour une table ou une base de données à laquelle vous accordez des autorisations à l'aide de la méthode de ressource nommée.

Note

Les administrateurs des lacs de données disposent d'autorisations implicites pour accorder et révoquer les autorisations relatives à Lake Formation. Mais ils ont toujours besoin des autorisations IAM sur l'octroi de l'autorisation Lake Formation et de la révocation des opérations d'API.

Les rôles IAM dotés d'une politique `AWSLakeFormationDataAdmin` AWS gérée ne peuvent pas ajouter de nouveaux administrateurs de lacs de données, car cette politique contient un refus explicite du fonctionnement de l'API Lake Formation. `PutDataLakeSetting`

La politique IAM suivante est recommandée aux directeurs qui ne sont pas des administrateurs de lacs de données et qui souhaitent accorder ou révoquer des autorisations à l'aide de la console Lake Formation.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:ListPermissions",
                "lakeformation:GrantPermissions",
                "lakeformation:BatchGrantPermissions",
                "lakeformation:RevokePermissions",
                "lakeformation:BatchRevokePermissions",
                "glue:GetDatabases",
                "glue:SearchTables",
                "glue:GetTables",
                "glue:GetDatabase",
                "glue:GetTable",
                "iam:ListUsers",
                "iam:ListRoles",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "sso:DescribeInstance"
            ],
            "Resource": "*"
        }
    ]
}
```

Toutes les `iam:` autorisations `glue:` et autorisations de cette politique sont disponibles dans la politique AWS géréeAWSGlueConsoleFullAccess.

Pour accorder des autorisations à l'aide du contrôle d'accès basé sur les balises Lake Formation (LF-TBAC), les principaux ont besoin d'autorisations IAM supplémentaires. Pour plus d’informations, consultez Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation et Référence des personnalités de Lake Formation et des autorisations IAM.

Autorisations entre comptes

Les utilisateurs qui souhaitent accorder des autorisations entre comptes Lake Formation à l'aide de la méthode des ressources nommées doivent également disposer des autorisations définies dans la politique `AWSLakeFormationCrossAccountManager` AWS gérée.

Les administrateurs des lacs de données ont besoin des mêmes autorisations pour accorder des autorisations entre comptes, ainsi que de l'autorisation AWS Resource Access Manager (AWS RAM) pour autoriser l'octroi d'autorisations aux organisations. Pour plus d'informations, consultez Autorisations d'administrateur du lac de données.

L'utilisateur administratif

Un directeur disposant d'autorisations administratives (par exemple, dans le cadre de la politique `AdministratorAccess` AWS gérée) est autorisé à accorder des autorisations à Lake Formation et à créer des administrateurs de lacs de données. Pour refuser à un utilisateur ou à un rôle l'accès aux opérations de l'administrateur de Lake Formation, joignez ou ajoutez à sa politique une `Deny` déclaration concernant les opérations d'API de l'administrateur.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "lakeformation:GetDataLakeSettings",
                "lakeformation:PutDataLakeSettings"
            ],
            "Effect": "Deny",
            "Resource": [
                "*"
            ]
        }
    ]
}
```

Important

Pour empêcher les utilisateurs de s'ajouter en tant qu'administrateurs à l'aide d'un script d'extraction, de transformation et de chargement (ETL), assurez-vous que l'accès à ces opérations d'API est refusé à tous les utilisateurs et rôles non administrateurs. La politique `AWSLakeFormationDataAdmin` AWS gérée contient un refus explicite du fonctionnement

de l'API Lake Formation, `PutDataLakeSetting` qui empêche les utilisateurs d'ajouter de nouveaux administrateurs de lacs de données.

Octroi d'autorisations de data lake à l'aide de la méthode de ressource nommée

Vous pouvez utiliser la méthode des ressources nommées pour accorder à Lake Formation des autorisations sur des bases de données, des tables et des vues spécifiques du catalogue de données. Vous pouvez accorder des autorisations à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Rubriques

- Octroi d'autorisations de base de données à l'aide de la méthode de ressource
- Octroi d'autorisations de table à l'aide de la méthode de ressource nommée
- Octroi d'autorisations sur les vues à l'aide de la méthode de ressource nommée

Octroi d'autorisations de base de données à l'aide de la méthode de ressource

Les étapes suivantes expliquent comment accorder des autorisations de base de données à l'aide de la méthode des ressources nommées.

Console

Utilisez la page Accorder les autorisations du data lake sur la console Lake Formation. La page est divisée selon les sections suivantes :

- Principaux : utilisateurs, rôles, utilisateurs et groupes IAM Identity Center, utilisateurs et groupes SAML, AWS comptes, organisations ou unités organisationnelles auxquels accorder les autorisations.
- Balises LF ou ressources du catalogue : bases de données, tables, vues ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder dans le cadre de la Lake Formation.

> **Note**
>
> Pour accorder des autorisations sur un lien vers une ressource de base de données, consultezOctroi d'autorisations relatives aux liens vers.

1. Ouvrez la page des autorisations du lac de données Grant.

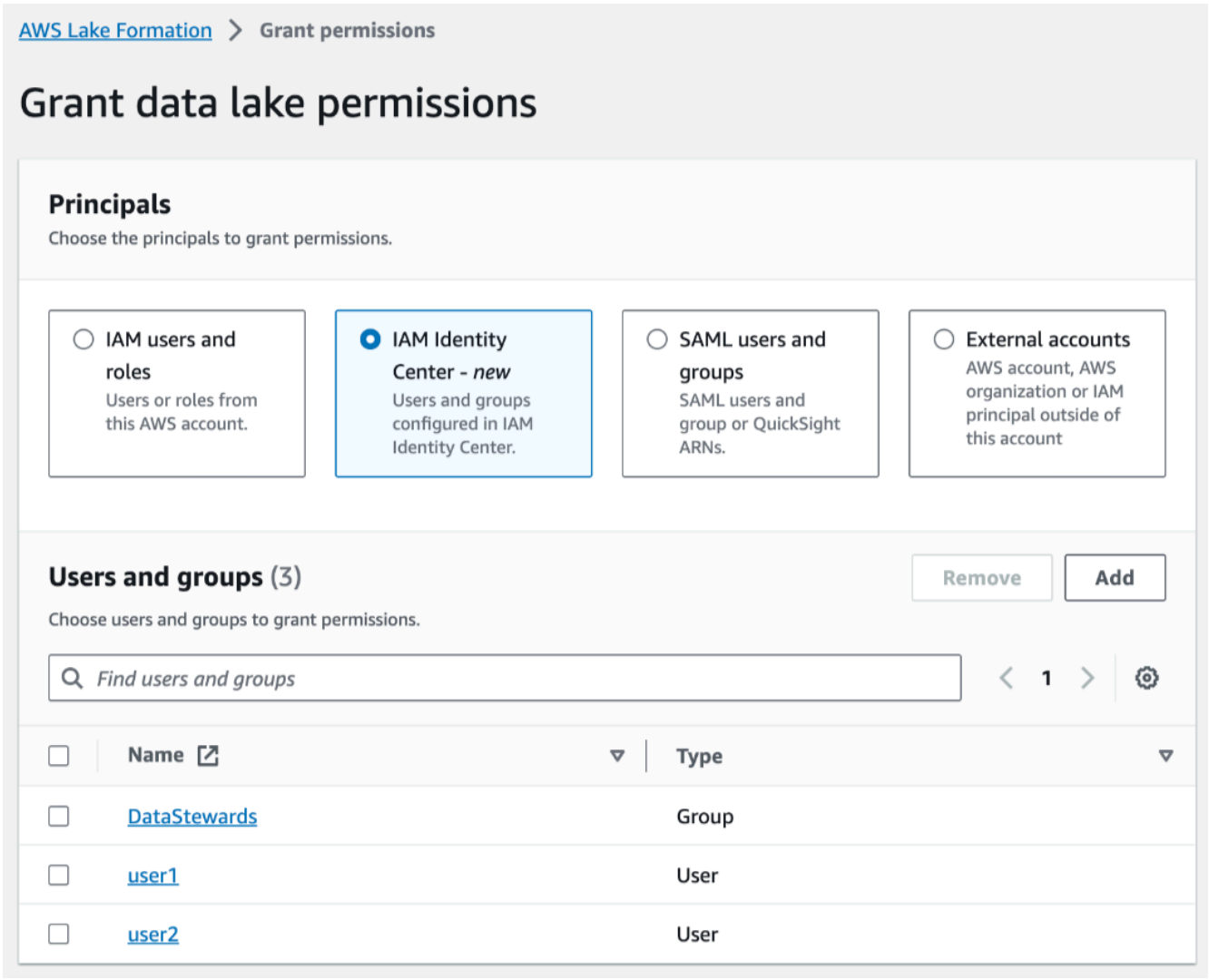
   Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/ et connectez-vous en tant qu'administrateur du lac de données, créateur de base de données ou utilisateur IAM disposant des autorisations Grantable sur la base de données.

   Effectuez l’une des actions suivantes :

   - Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données. Choisissez ensuite Grant.
   - Dans le volet de navigation, sélectionnez Bases de données sous Catalogue de données. Ensuite, sur la page Bases de données, choisissez une base de données, puis dans le menu Actions, sous Autorisations, choisissez Grant.

   > **Note**
   >
   > Vous pouvez accorder des autorisations sur une base de données via son lien de ressource. Pour ce faire, sur la page Bases de données, choisissez un lien vers une ressource, puis dans le menu Actions, choisissez Grant on target. Pour plus d’informations, consultez Mode de fonctionnement des liens des ressources dans Lake Formation.

2. Ensuite, dans la section Principaux, choisissez un type de principal, puis spécifiez les principaux auxquels accorder les autorisations.



Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou groupes dans la liste Utilisateurs et groupes. Sélectionnez Ajouter pour ajouter d'autres utilisateurs ou groupes.

Utilisateurs et groupes SAML

Pour les QuickSight utilisateurs et les groupes SAML et Amazon, entrez un ou plusieurs Amazon Resource Names (ARN) pour les utilisateurs ou les groupes fédérés via SAML, ou des ARN pour les utilisateurs ou les groupes Amazon. QuickSight Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire les ARN, consultezLake Formation accorde et AWS CLI révoque des commandes.

> **Note**
>
> L'intégration de Lake Formation à Amazon n' QuickSight est prise en charge que pour Amazon QuickSight Enterprise Edition.

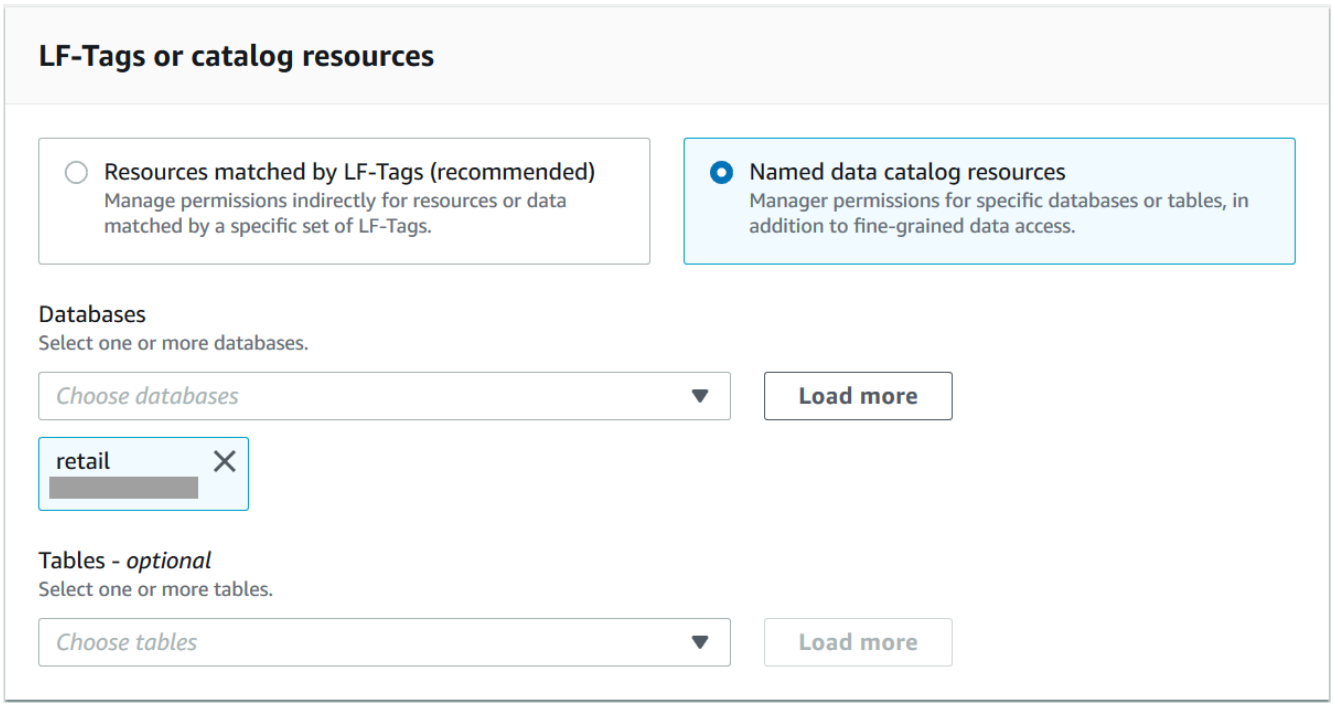
Comptes externes

Pour Compte AWS, AWS organisation ou IAM Principal, entrez un ou plusieurs ID de AWS compte, ID d'organisation, ID d'unité organisationnelle ou ARN valides pour l'utilisateur ou le rôle IAM. Appuyez sur Entrée après chaque identifiant.

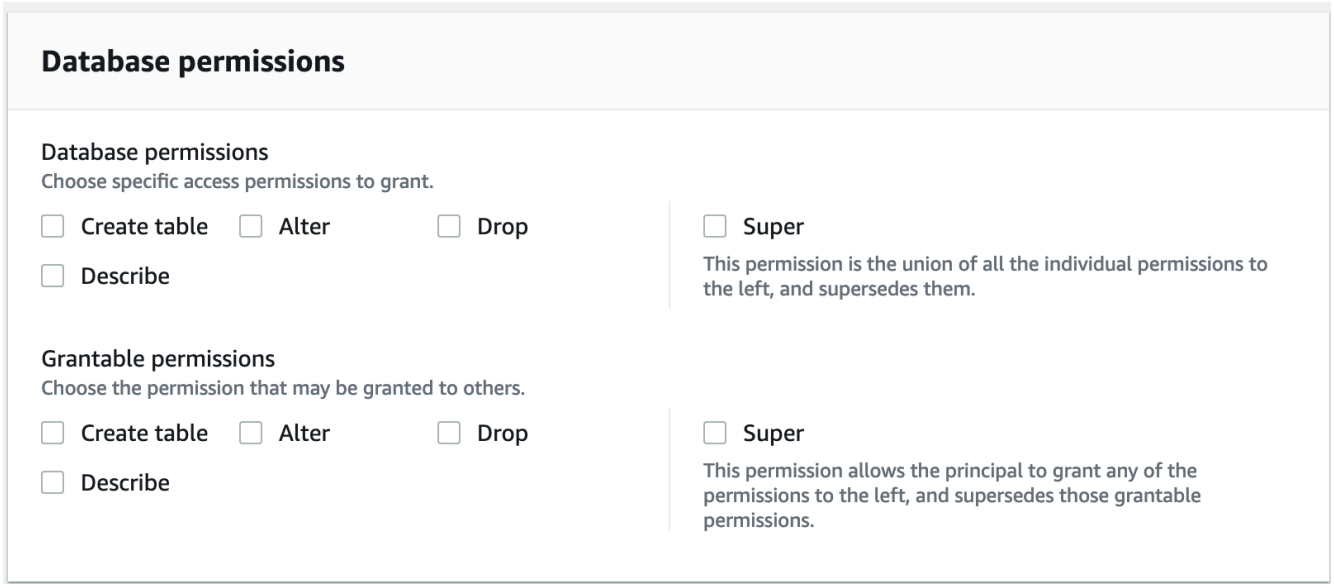
Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

3. Dans la section Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées.



4. Choisissez une ou plusieurs bases de données dans la liste des bases de données. Vous pouvez également choisir un ou plusieurs tableaux et/ou filtres de données.
5. Dans la section Autorisations, sélectionnez les autorisations et les autorisations pouvant être accordées. Sous Autorisations de base de données, sélectionnez une ou plusieurs autorisations à accorder.



> **Note**
>
> Après avoir accordé `Create Table` ou `Alter` sur une base de données dotée d'une propriété d'emplacement pointant vers un emplacement enregistré, veillez à accorder également des autorisations de localisation des données sur cet emplacement aux principaux. Pour plus d'informations, consultez Octroi d'autorisations de localisation des données.

6. (Facultatif) Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS Cette option n'est pas prise en charge lorsque vous accordez des autorisations à un directeur IAM à partir d'un compte externe.
7. Choisissez Grant (Accorder).

AWS CLI

Vous pouvez accorder des autorisations de base de données en utilisant la méthode de ressource nommée et le AWS Command Line Interface (AWS CLI).

Pour accorder des autorisations de base de données à l'aide du AWS CLI

- Exécutez une `grant-permissions` commande et spécifiez une base de données ou le catalogue de données comme ressource, en fonction de l'autorisation accordée.

  Dans les exemples suivants, remplacez-le <account-id>par un identifiant de AWS compte valide.

  Example — Subvention pour créer une base de données

  Cet exemple accorde des autorisations `CREATE_DATABASE` à l'utilisateur`datalake_user1`. Étant donné que la ressource pour laquelle cette autorisation est accordée est le catalogue de données, la commande spécifie une `CatalogResource` structure vide comme `resource` paramètre.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Example — Autorisation de créer des tables dans une base de données désignée

L'exemple suivant octroie la `CREATE_TABLE` base de données `retail` à l'utilisateur`datalake_user1`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Example — Subvention à un AWS compte externe avec l'option Grant

L'exemple suivant octroie `CREATE_TABLE` l'option grant sur la base de données `retail` au compte externe 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE"
 --permissions-with-grant-option "CREATE_TABLE" --resource '{ "Database":
 {"Name":"retail"}}'
```

Example — Subvention à une organisation

L'exemple suivant octroie `ALTER` à l'organisation l'option grant sur la base `issues` de données`o-abcdefghijkl`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "ALTER" --permissions-with-grant-option "ALTER" --
resource '{ "Database": {"Name":"issues"}}'
```

Example - Accordez **`ALLIAMPrincipals`** à sur le même compte

L'exemple suivant accorde `CREATE_TABLE` l'autorisation d'accéder à la base `retail` de données à tous les principaux d'un même compte. Cette option permet à chaque principal du compte de créer une table dans la base de données et de créer un lien vers une ressource de table permettant aux moteurs de requêtes intégrés d'accéder aux bases de données et aux tables partagées. Cette option est particulièrement utile lorsqu'un directeur reçoit une subvention entre comptes et n'est pas autorisé à créer des liens vers des ressources. Dans ce scénario, l'administrateur du lac de données peut créer une base de données d'espaces

réservés et accorder des `CREATE_TABLE` autorisations au `ALLIAMPrincipal` groupe, permettant ainsi à chaque responsable IAM du compte de créer des liens de ressources dans la base de données d'espaces réservés.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"temp","CatalogId":"111122223333"}}'
```

Example - Subvention **`ALLIAMPrincipals`** à un compte externe

L'exemple suivant accorde des autorisations `CREATE_TABLE` d'accès à la base `retail` de données à tous les principaux d'un compte externe. Cette option permet à tous les principaux du compte de créer une table dans la base de données.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"retail","CatalogId":"123456789012"}}'
```

> **Note**
>
> Après avoir accordé `CREATE_TABLE` ou `ALTER` sur une base de données dotée d'une propriété d'emplacement pointant vers un emplacement enregistré, veillez à accorder également des autorisations de localisation des données sur cet emplacement aux principaux. Pour plus d'informations, consultez Octroi d'autorisations de localisation des données.

> **Consultez aussi**
>
> - Référence des autorisations de Lake Formation
> - Octroi d'autorisations sur une base de données ou une table partagée avec votre compte
> - Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci

Octroi d'autorisations de table à l'aide de la méthode de ressource nommée

Vous pouvez utiliser la console Lake Formation ou AWS CLI accorder des autorisations Lake Formation sur les tables du catalogue de données. Vous pouvez accorder des autorisations sur des tables individuelles, ou avec une seule opération d'autorisation, vous pouvez accorder des autorisations sur toutes les tables d'une base de données.

Si vous accordez des autorisations sur toutes les tables d'une base de données, vous les accordez implicitement sur la `DESCRIBE` base de données. La base de données apparaît ensuite sur la page Bases de données de la console et est renvoyée par l'opération `GetDatabases` d'API.

Lorsque vous choisissez `SELECT` l'autorisation à accorder, vous avez la possibilité d'appliquer un filtre de colonne, un filtre de ligne ou un filtre de cellule.

Console

Les étapes suivantes expliquent comment accorder des autorisations de table à l'aide de la méthode de ressource nommée et de la page Accorder des autorisations de lac de données sur la console Lake Formation. La page est divisée en sections suivantes :

- Principaux : utilisateurs, rôles, AWS comptes, organisations ou unités organisationnelles auxquels accorder des autorisations.
- Balises LF ou ressources du catalogue : bases de données, tables ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder dans le cadre de la Lake Formation.

> **Note**
>
> Pour accorder des autorisations sur le lien d'une ressource de table, consultezOctroi d'autorisations relatives aux liens vers.

1. Ouvrez la page des autorisations du lac de données Grant.

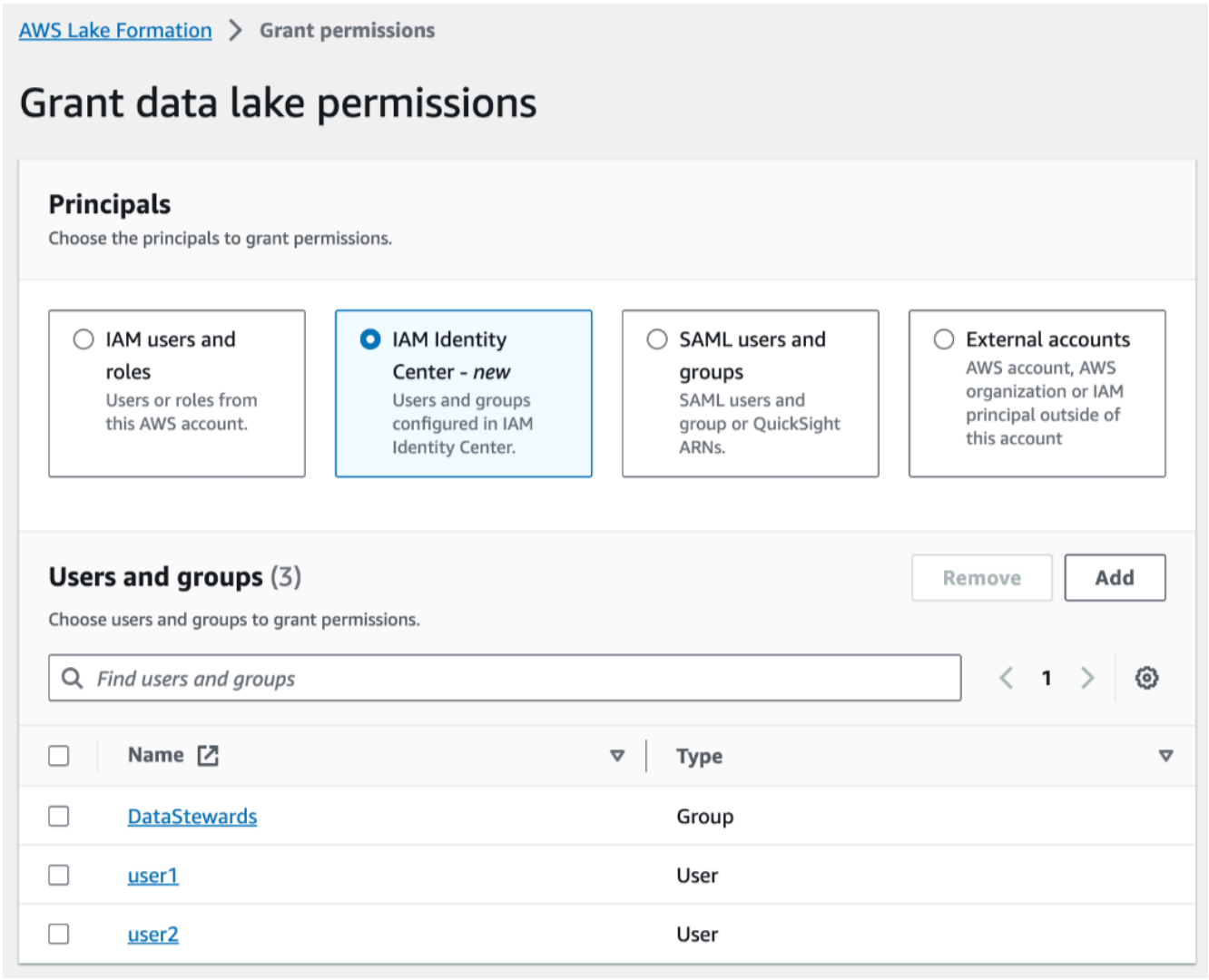
   Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/ et connectez-vous en tant qu'administrateur du lac de données, créateur de la table ou utilisateur ayant obtenu des autorisations sur la table avec l'option grant.

   Effectuez l’une des actions suivantes :

- Dans le volet de navigation, sélectionnez Autorisations du lac de données sous Autorisations. Choisissez ensuite Grant.
- Dans le volet de navigation, choisissez Tables. Ensuite, sur la page Tables, choisissez un tableau, puis dans le menu Actions, sous Autorisations, choisissez Accorder.

> **Note**
>
> Vous pouvez accorder des autorisations sur une table via son lien de ressource. Pour ce faire, sur la page Tables, choisissez un lien vers une ressource, puis dans le menu Actions, choisissez Grant on target. Pour plus d'informations, consultez Mode de fonctionnement des liens des ressources dans Lake Formation.

2. Ensuite, dans la section Principaux, choisissez un type de principal et spécifiez les principaux auxquels accorder les autorisations.



Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou groupes dans la liste Utilisateurs et groupes.

Utilisateurs et groupes SAML

Pour les QuickSight utilisateurs et les groupes SAML et Amazon, entrez un ou plusieurs Amazon Resource Names (ARN) pour les utilisateurs ou les groupes fédérés via SAML, ou des ARN pour les utilisateurs ou les groupes Amazon. QuickSight Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire les ARN, consultezLake Formation accorde et AWS CLI révoque des commandes.

> **Note**
>
> L'intégration de Lake Formation à Amazon QuickSight n'est prise en charge que pour Amazon QuickSight Enterprise Edition.

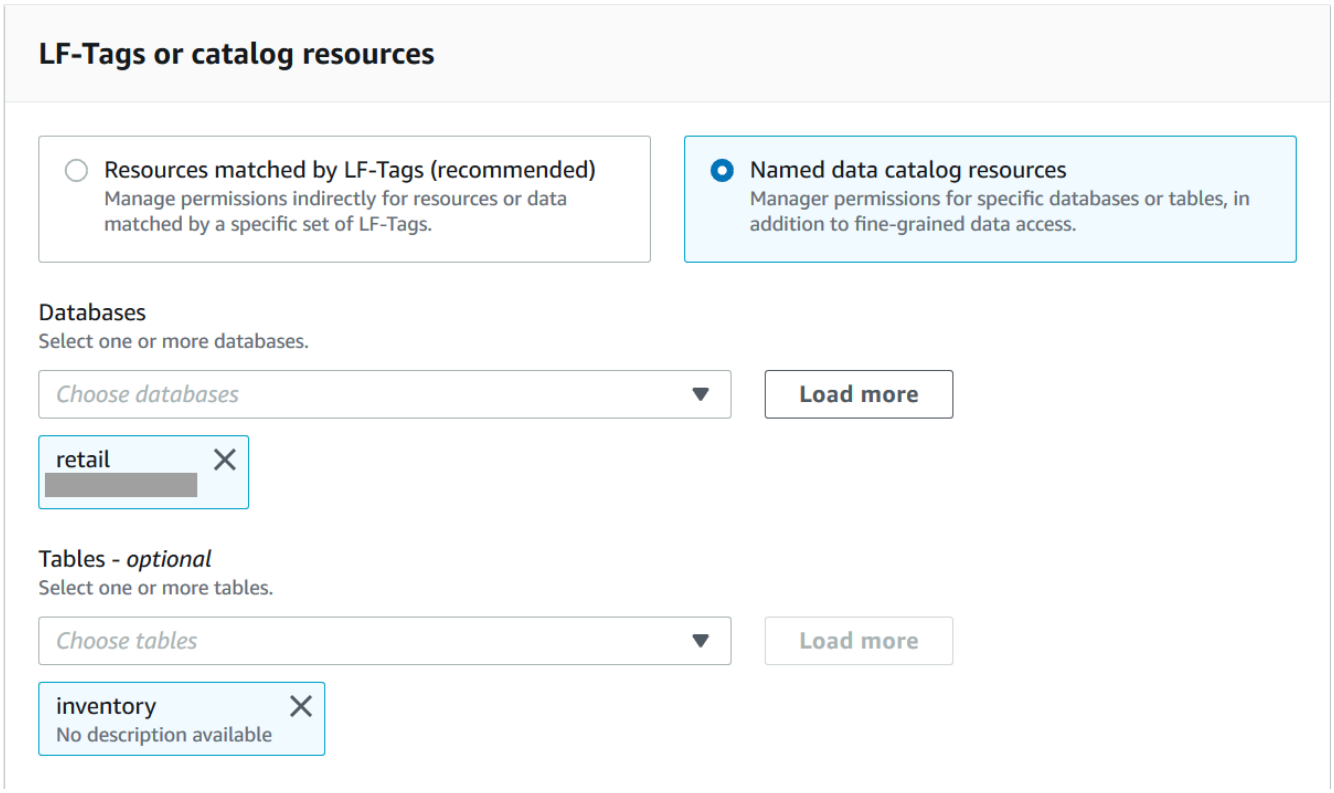
Comptes externes

Pour Compte AWS , AWS organisation ou directeur IAM, entrez un ou plusieurs Compte AWS identifiants, identifiants d'organisation, identifiants d'unité organisationnelle ou ARN valides pour l'utilisateur ou le rôle IAM. Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

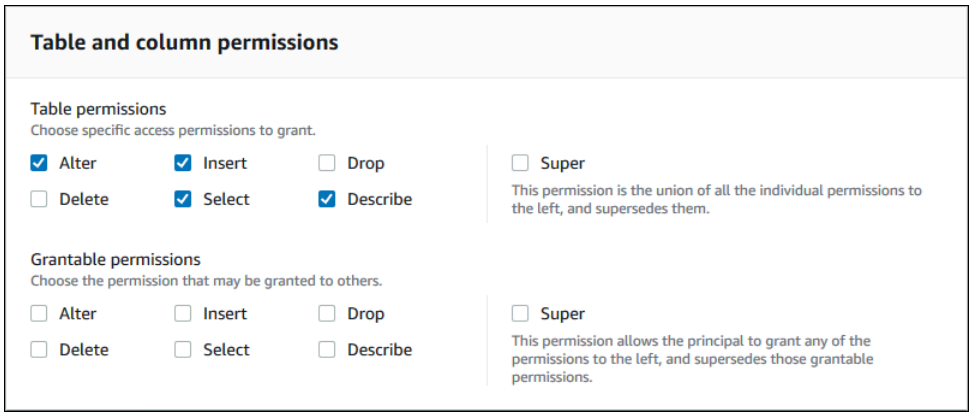
L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième caractère « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

3. Dans la section Balises LF ou ressources du catalogue, choisissez une base de données. Choisissez ensuite une ou plusieurs tables, ou toutes les tables.

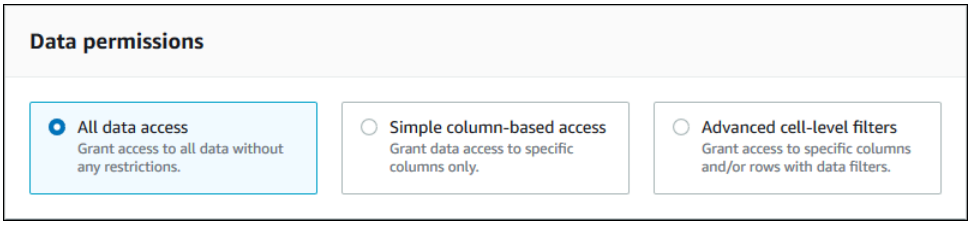


4. Spécifiez les autorisations sans filtrage des données

   Dans la section Autorisations, sélectionnez les autorisations de table à accorder, et sélectionnez éventuellement les autorisations pouvant être accordées.



   Si vous autorisez Select, la section Autorisations relatives aux données apparaît sous la section Autorisations relatives aux tables et aux colonnes, l'option Accès à toutes les données étant sélectionnée par défaut. Acceptez la valeur par défaut.



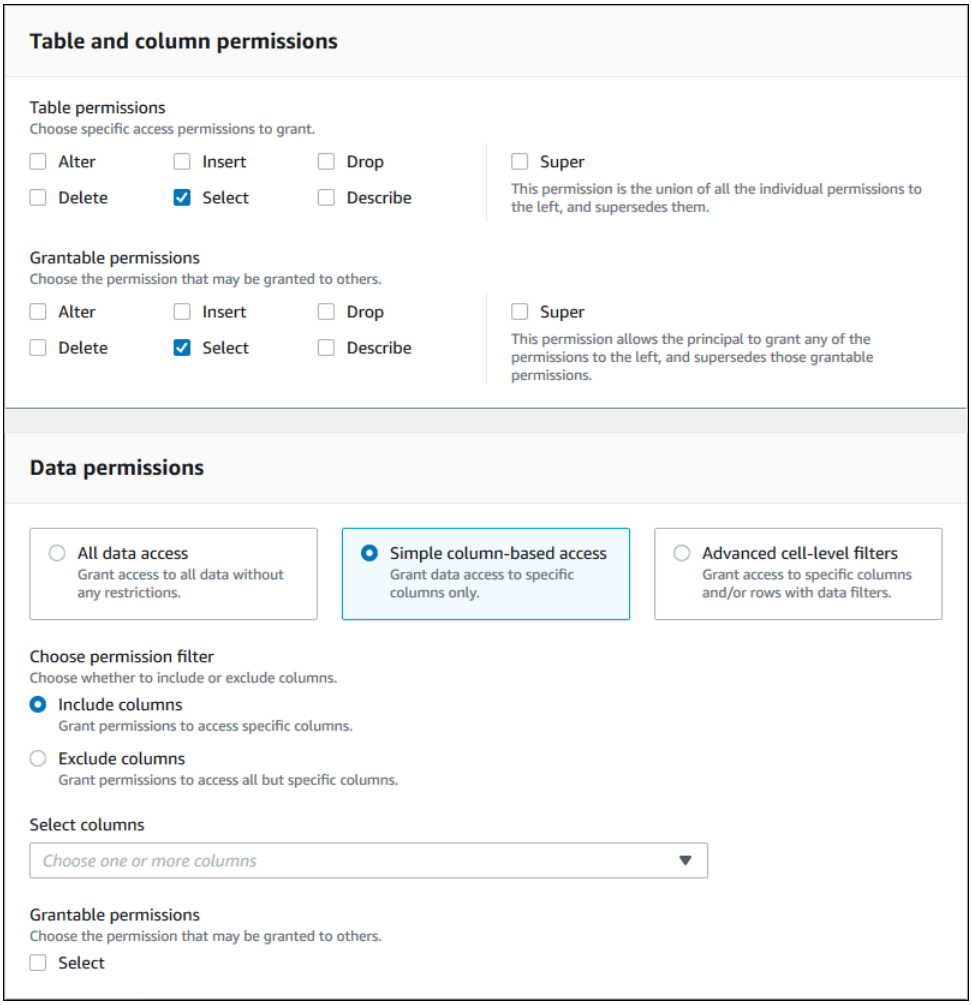
5. Choisissez Grant (Accorder).
6. Spécifiez l'autorisation de sélection avec filtrage des données

   Sélectionnez l'autorisation Select. Ne sélectionnez aucune autre autorisation.

   La section Autorisations relatives aux données apparaît sous la section Autorisations relatives aux tables et aux colonnes.
7. Effectuez l’une des actions suivantes :

   - Appliquez uniquement un filtrage par colonne simple.

     1. Choisissez Accès simple basé sur des colonnes.



2. Choisissez d'inclure ou d'exclure des colonnes, puis choisissez les colonnes à inclure ou à exclure.

   Seules les listes d'inclusion sont prises en charge lors de l'octroi d'autorisations à un AWS compte ou à une organisation externe.

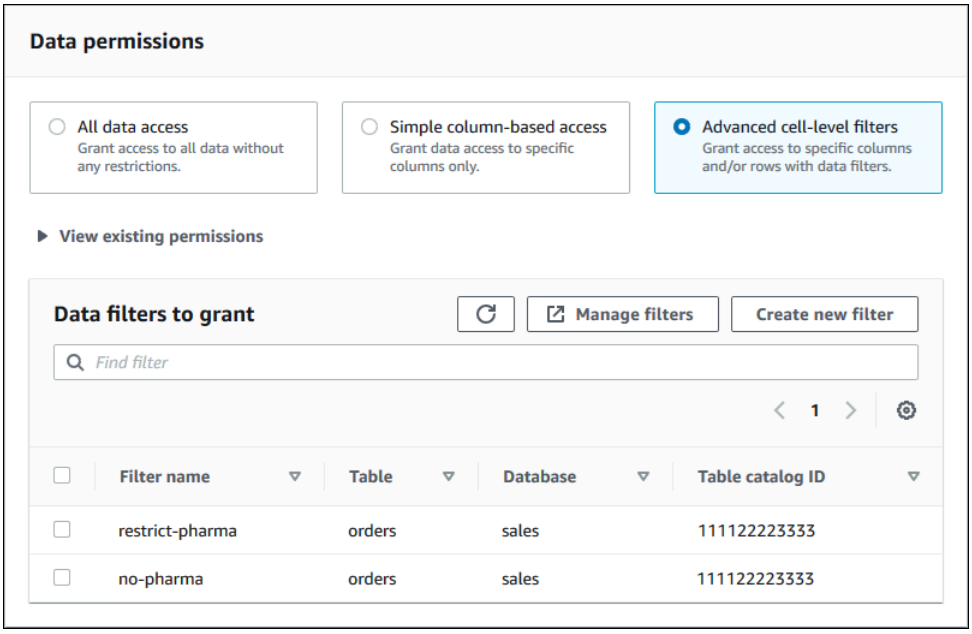
3. (Facultatif) Sous Autorisations pouvant être accordées, activez l'option d'octroi pour l'autorisation Select.

   Si vous incluez l'option de subvention, le bénéficiaire de la subvention ne peut accorder des autorisations que sur les colonnes que vous lui accordez.

> **Note**
>
> Vous pouvez également appliquer le filtrage des colonnes uniquement en créant un filtre de données qui spécifie un filtre de colonne et spécifie toutes les lignes comme filtre de ligne. Cependant, cela nécessite d'autres étapes.

- Appliquez un filtrage par colonne, ligne ou cellule.

  1. Choisissez Filtres avancés au niveau des cellules.



  2. (Facultatif) Agrandir Afficher les autorisations existantes.
  3. (Facultatif) Choisissez Créer un nouveau filtre.
  4. (Facultatif) Pour afficher les détails des filtres répertoriés, ou pour créer de nouveaux filtres ou supprimer des filtres existants, choisissez Gérer les filtres.

     La page Filtres de données s'ouvre dans une nouvelle fenêtre de navigateur.

     Lorsque vous avez terminé sur la page Filtres de données, retournez à la page Accorder des autorisations et, si nécessaire, actualisez la page pour afficher les nouveaux filtres de données que vous avez créés.
  5. Sélectionnez un ou plusieurs filtres de données à appliquer à la subvention.

> **Note**
>
> Si la liste ne contient aucun filtre de données, cela signifie qu'aucun filtre de données n'a été créé pour la table sélectionnée.

8. Choisissez Grant (Accorder).

AWS CLI

Vous pouvez accorder des autorisations de table en utilisant la méthode de ressource nommée et le AWS Command Line Interface (AWS CLI).

Pour accorder des autorisations de table à l'aide du AWS CLI

- Exécutez une `grant-permissions` commande et spécifiez une table comme ressource.

Example — Subvention sur une seule table, sans filtrage

L'exemple suivant accorde `SELECT` et `ALTER` à l'utilisateur `datalake_user1` dans le AWS compte 1111-2222-3333 sur la table de la base de données. `inventory` `retail`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

> **Note**
>
> Si vous accordez l'`ALTER`autorisation sur une table dont les données sous-jacentes se trouvent dans un emplacement enregistré, veillez à accorder également des autorisations de localisation des données sur cet emplacement aux principaux. Pour plus d'informations, consultez Octroi d'autorisations de localisation des données.

Example — Subvention sur toutes les tables avec l'option Grant : aucun filtrage

L'exemple suivant accorde des autorisations `SELECT` avec l'option grant sur toutes les tables de la base de données`retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --permissions-with-grant-option "SELECT" --resource '{ "Table":
 { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Example — Subvention avec filtrage simple par colonne

L'exemple suivant accorde des SELECT autorisations sur un sous-ensemble de colonnes de la table. `persons` Il utilise un simple filtrage par colonne.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --resource '{ "TableWithColumns": {"DatabaseName":"hr",
 "Name":"persons", "ColumnNames":["family_name", "given_name", "gender"]}}'
```

Example — Subvention avec filtre de données

Cet exemple accorde des SELECT autorisations sur la `orders` table et applique le filtre de `restrict-pharma` données.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

Le contenu du fichier est le suivant`grant-params.json`.

```
{
    "Principal": {"DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"},
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
    "Permissions": ["SELECT"],
    "PermissionsWithGrantOption": ["SELECT"]
}
```

Consultez aussi

- Vue d'ensemble des autorisations relatives à Lake Formation
- Filtrage des données et sécurité au niveau des cellules dans Lake Formation
- Référence des personnalités de Lake Formation et des autorisations IAM
- Octroi d'autorisations relatives aux liens vers
- Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci

Octroi d'autorisations sur les vues à l'aide de la méthode de ressource nommée

Les étapes suivantes expliquent comment accorder des autorisations sur les vues à l'aide de la méthode des ressources nommées et de la page Accorder des autorisations aux lacs de données. La page est divisée selon les sections suivantes :

- Principaux : utilisateurs, rôles, utilisateurs et groupes IAM Identity Center Comptes AWS, organisations ou unités organisationnelles auxquels accorder les autorisations.
- Balises LF ou ressources du catalogue : bases de données, tables, vues ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder au lac de données.

Ouvrez la page des autorisations du lac de données Grant

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/ et connectez-vous en tant qu'administrateur du lac de données, créateur de base de données ou utilisateur IAM disposant des autorisations Grantable sur la base de données.
2. Effectuez l’une des actions suivantes :

   - Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données. Choisissez ensuite Grant.
   - Dans le volet de navigation, sélectionnez Views sous Catalogue de données. Ensuite, sur la page Vues, choisissez une vue, puis dans le menu Actions, sous Autorisations, choisissez Accorder.

> **Note**
>
> Vous pouvez accorder des autorisations sur une vue via son lien de ressource. Pour ce faire, sur la page Vues, choisissez un lien vers une ressource, puis dans le menu Actions, choisissez Grant on target. Pour plus d'informations, consultez Mode de fonctionnement des liens des ressources dans Lake Formation.

Spécifiez les principes

Dans la section Principaux, choisissez un type de principal, puis spécifiez les principaux auxquels accorder les autorisations.

Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

IAM Identity Center

Choisissez un ou plusieurs utilisateurs ou groupes dans la liste Utilisateurs et groupes.

Utilisateurs et groupes SAML

Pour les QuickSight utilisateurs et les groupes SAML et Amazon, entrez un ou plusieurs Amazon Resource Names (ARN) pour les utilisateurs ou les groupes fédérés via SAML, ou des ARN pour les utilisateurs ou les groupes Amazon. QuickSight Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire les ARN, consultezLake Formation accorde et AWS CLI révoque des commandes.

> **Note**
>
> L'intégration de Lake Formation à Amazon n' QuickSight est prise en charge que pour Amazon QuickSight Enterprise Edition.

Comptes externes

Pour Compte AWS, AWS organisation ou IAM Principal, entrez un ou plusieurs ID de AWS compte, ID d'organisation, ID d'unité organisationnelle ou ARN valides pour l'utilisateur ou le rôle IAM. Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

consultez aussi

- Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci

Spécifiez les vues

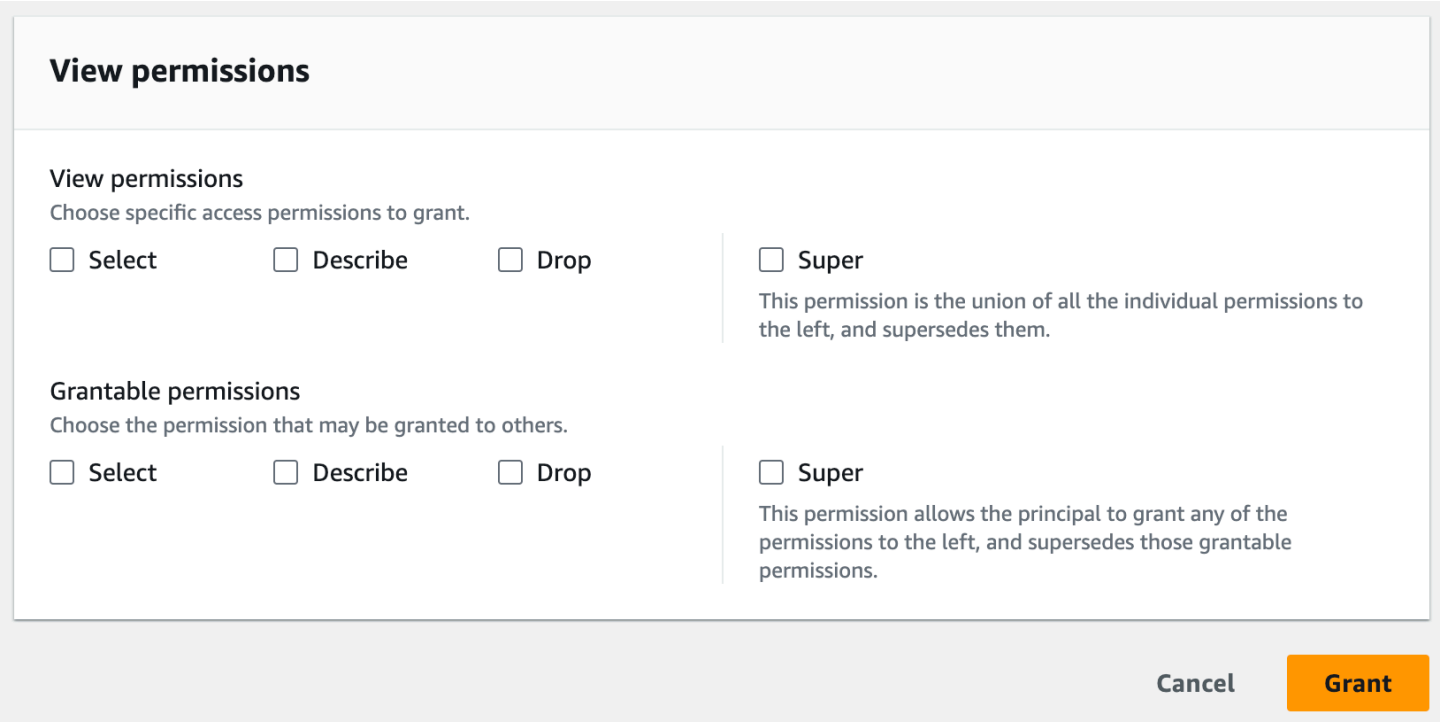
Dans la section LF-Tags ou ressources du catalogue, choisissez une ou plusieurs vues pour lesquelles vous souhaitez accorder des autorisations.

1. Choisissez Ressources de catalogue de données nommées.
2. Choisissez une ou plusieurs vues dans la liste des vues. Vous pouvez également choisir une ou plusieurs bases de données, tables et/ou filtres de données.

   L'octroi d'autorisations de lac de données au `All views` sein d'une base de données donnera au bénéficiaire des autorisations sur toutes les tables et vues de la base de données.

Spécifiez les autorisations

Dans la section Autorisations, sélectionnez les autorisations et les autorisations pouvant être accordées.



1. Sous Afficher les autorisations, sélectionnez une ou plusieurs autorisations à accorder.
2. (Facultatif) Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux dans son domaine. Compte AWS Cette option n'est pas prise en charge lorsque vous accordez des autorisations à un directeur IAM à partir d'un compte externe.
3. Choisissez Grant (Accorder).

consultez aussi

- Référence des autorisations de Lake Formation
- Octroi d'autorisations sur une base de données ou une table partagée avec votre compte

Contrôle d'accès basé sur des balises Lake Formation

Le contrôle d'accès basé sur les balises de Lake Formation (LF-TBAC) est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Dans Lake Formation, ces attributs sont appelés balises LF. Vous pouvez associer des balises LF aux ressources du catalogue de données et accorder des autorisations aux responsables de Lake Formation sur ces ressources à

l'aide de ces balises LF. Lake Formation autorise les opérations sur ces ressources lorsque la valeur de balise du principal correspond à la valeur de l'étiquette de ressource. Le LF-TBAC est utile dans les environnements qui se développent rapidement et dans les situations où la gestion des politiques devient fastidieuse.

La méthode LF-TBAC est recommandée pour accorder des autorisations à Lake Formation lorsqu'il existe un grand nombre de ressources du catalogue de données. La méthode LF-TBAC est plus évolutive que la méthode des ressources nommées et nécessite moins de frais de gestion des autorisations.

> **Note**
>
> Les balises IAM sont différentes des balises LF. Ces étiquettes ne sont pas interchangeables. Les balises LF sont utilisées pour accorder des autorisations à Lake Formation et les balises IAM sont utilisées pour définir les politiques IAM.

Comment fonctionne le contrôle d'accès basé sur des balises Lake Formation

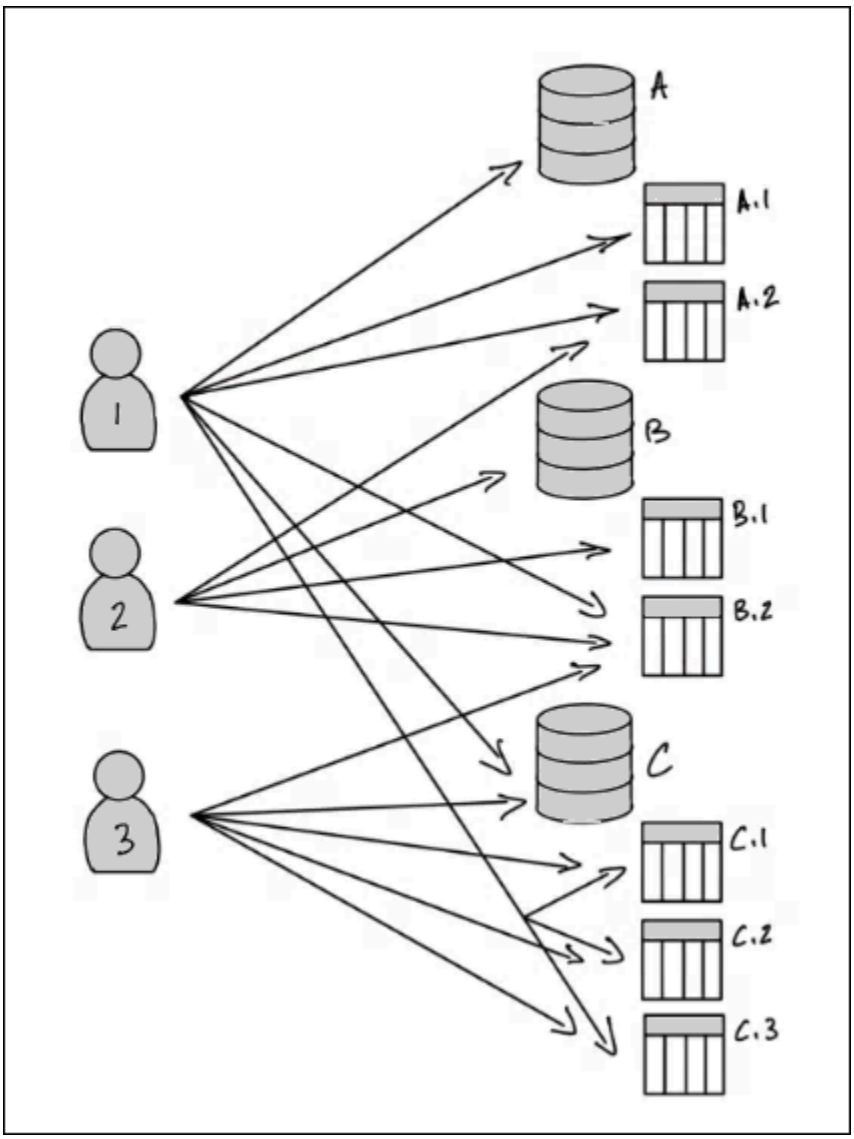
Chaque balise LF est une paire clé-valeur, telle que ou. `department=sales` `classification=restricted` Une clé peut avoir plusieurs valeurs définies, telles que`department=sales,marketing,engineering,finance`.

Pour utiliser la méthode LF-TBAC, les administrateurs de lacs de données et les ingénieurs de données effectuent les tâches suivantes.

Tâche	Détails de la tâche
1. Définissez les propriétés et les relations des balises LF.	-
2. Créez les créateurs de balises LF dans Lake Formation.	Ajouter des créateurs de LF-Tag
3. Créez le tag LF dans Lake Formation.	Création de balises LF
4. Attribuez des balises LF aux ressources du catalogue de données.	Affectation de balises LF aux ressources du catalogue de données

Tâche	Détails de la tâche
5. Accordez des autorisations à d'autres principaux pour attribuer des balises LF aux ressources, éventuellement avec l'option d'octroi.	Octroi, révocation et listage des autorisations relatives aux valeurs des balises LF
6. Accordez des expressions LF-Tag aux principaux, éventuellement avec l'option grant.	Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC
7. (Recommandé) Après avoir vérifié que les principaux ont accès aux bonnes ressources par le biais de la méthode LF-TBAC, révoquez les autorisations accordées à l'aide de la méthode des ressources nommées.	-

Imaginons le cas où vous devez accorder des autorisations à trois principaux sur trois bases de données et sept tables.



Pour obtenir les autorisations indiquées dans le schéma précédent en utilisant la méthode des ressources nommées, vous devez accorder 17 autorisations, comme suit (en pseudo-code).

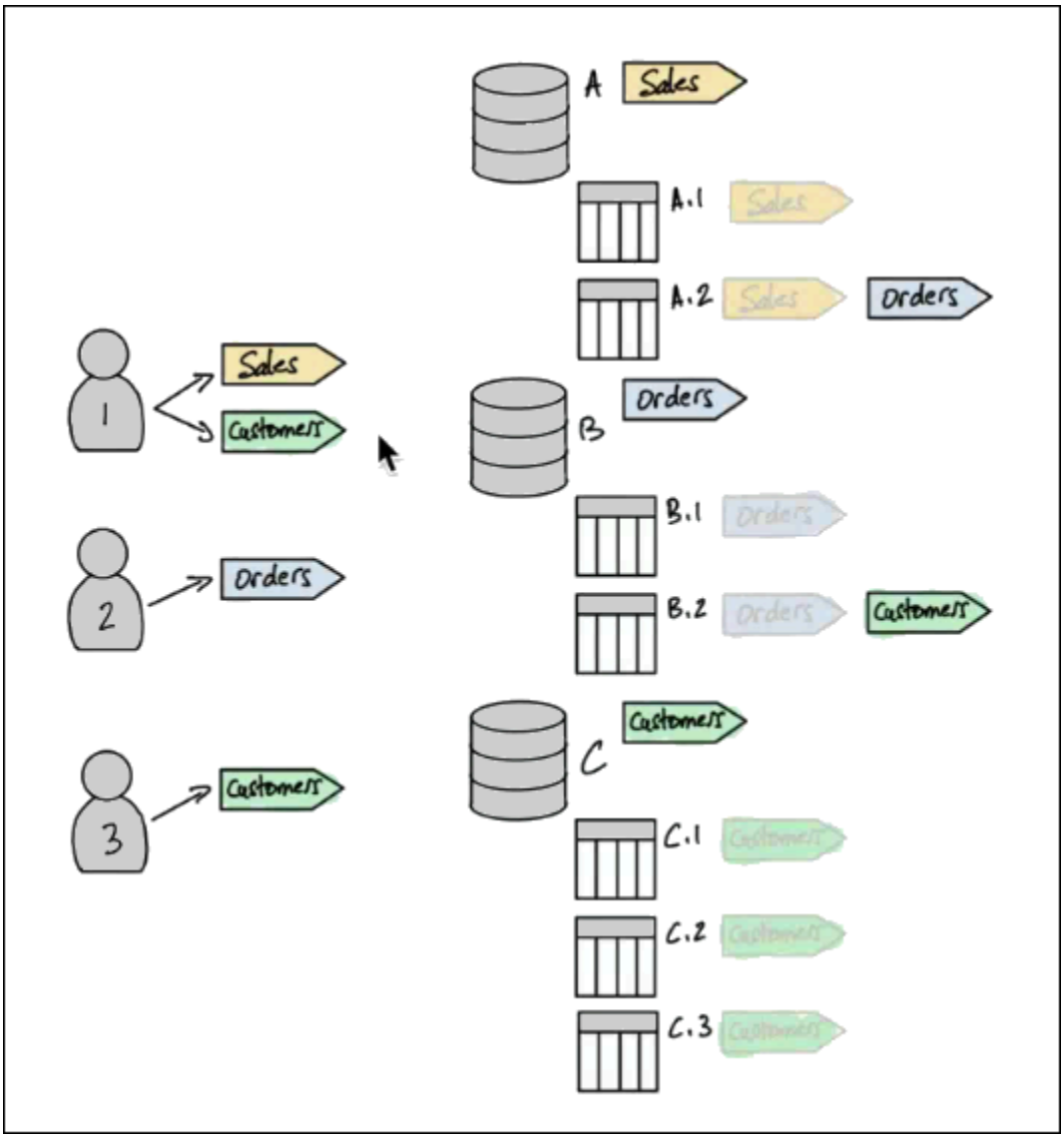
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

Réfléchissez maintenant à la manière dont vous accorderiez des autorisations en utilisant le LF-TBAC. Le schéma suivant indique que vous avez attribué des balises LF aux bases de données et aux tables, et que vous avez accordé des autorisations sur les balises LF aux principaux.

Dans cet exemple, les balises LF représentent des zones du lac de données qui contiennent des analyses pour différents modules d'une suite d'applications de planification des ressources d'entreprise (ERP). Vous pouvez contrôler l'accès aux données d'analyse pour les différents modules. Toutes les balises LF ont la clé `module` et les valeurs possibles `SalesOrders`, et. `Customers` Voici un exemple de balise LF :

```
module=Sales
```

Le diagramme montre uniquement les valeurs des balises LF.



Attributions de balises aux ressources et à l'héritage du catalogue de données

Les tables héritent des balises LF des bases de données et les colonnes héritent des balises LF des tables. Les valeurs héritées peuvent être remplacées. Dans le schéma précédent, les balises LF grisées sont héritées.

En raison de l'héritage, l'administrateur du lac de données doit uniquement attribuer les cinq balises LF suivantes aux ressources (en pseudo-code).

```
ASSIGN TAGS module=Sales TO database A
ASSIGN TAGS module=Orders TO table A.2
ASSIGN TAGS module=Orders TO database B
ASSIGN TAGS module=Customers TO table B.2
ASSIGN TAGS module=Customers TO database C
```

Étiquetez les subventions aux directeurs

Après avoir attribué des balises LF aux bases de données et aux tables, l'administrateur du lac de données ne doit accorder que quatre balises LF aux principaux, comme suit (en pseudo-code).

```
GRANT TAGS module=Sales TO Principal 1
GRANT TAGS module=Customers TO Principal 1
GRANT TAGS module=Orders TO Principal 2
GRANT TAGS module=Customers TO Principal 3
```

Désormais, un principal doté du `module=Sales` tag LF peut accéder aux ressources du catalogue de données avec le `module=Sales` tag LF (par exemple, la base de données A), un principal doté du tag `module=Customers` LF peut accéder aux ressources avec le `module=Customers` tag LF, etc.

Les commandes d'autorisation précédentes sont incomplètes. En effet, bien qu'ils indiquent par le biais de balises LF les ressources du catalogue de données sur lesquelles les principaux sont autorisés, ils n'indiquent pas exactement quelles autorisations de Lake Formation (telles que`SELECT`,`ALTER`) les principaux ont sur ces ressources. Par conséquent, les commandes de pseudo-code suivantes sont une représentation plus précise de la manière dont les autorisations Lake Formation sont accordées sur les ressources du catalogue de données via des balises LF.

```
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Sales TO Principal 1
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Sales TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 1
```

```
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Orders TO Principal 2
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Orders TO Principal 2
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 3
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 3
```

Assemblage : autorisations résultantes sur les ressources

Compte tenu des balises LF attribuées aux bases de données et aux tables dans le schéma précédent, et des balises LF accordées aux principaux dans le diagramme, le tableau suivant répertorie les autorisations de Lake Formation dont disposent les principaux sur les bases de données et les tables.

Principal	Autorisations accordées par le biais de balises LF
Principal 1	• `CREATE_TABLE` sur la base de données A • `SELECT`, `INSERT` sur le tableau A.1 • `SELECT`, `INSERT` sur le tableau B.2 • `CREATE_TABLE` sur la base de données C • `SELECT`, `INSERT` sur le tableau C.1 • `SELECT`, `INSERT` sur le tableau C.2 • `SELECT`, `INSERT` sur le tableau C.3
Principal 2	• `SELECT`, `INSERT` sur le tableau A.2 • `CREATE_TABLE` sur la base de données B • `SELECT`, `INSERT` sur le tableau B.1 • `SELECT`, `INSERT` sur le tableau B.2
Principal 3	• `SELECT`, `INSERT` sur le tableau B.2 • `CREATE_TABLE` sur la base de données C • `SELECT`, `INSERT` sur le tableau C.1 • `SELECT`, `INSERT` sur le tableau C.2 • `SELECT`, `INSERT` sur le tableau C.3

Conclusion

Dans cet exemple simple, à l'aide de cinq opérations d'attribution et de huit opérations de subvention, l'administrateur du lac de données a pu spécifier 17 autorisations. Lorsqu'il existe des dizaines de bases de données et des centaines de tables, l'avantage de la méthode LF-TBAC par rapport à la méthode des ressources nommées devient évident. Dans le cas hypothétique de la nécessité d'accorder à chaque principal l'accès à chaque ressource, et où `n(P)` sont le nombre de principaux et `n(R)` le nombre de ressources :

- Avec la méthode des ressources nommées, le nombre de subventions requises est de `n(P)` ×`n(R)`.
- Avec la méthode LF-TBAC, en utilisant un seul tag LF, le total du nombre de subventions aux directeurs et d'affectations aux ressources est de +. `n(P) n(R)`

Consultez aussi

- Gestion des balises LF pour le contrôle d'accès aux métadonnées
- Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC

Rubriques

- Gestion des balises LF pour le contrôle d'accès aux métadonnées
- Octroi, révocation et listage des autorisations relatives aux valeurs des balises LF

Gestion des balises LF pour le contrôle d'accès aux métadonnées

Pour utiliser la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) afin de sécuriser les ressources du catalogue de données (bases de données, tables et colonnes), vous créez des balises LF, vous les attribuez aux ressources et vous accordez des autorisations de balises LF aux principaux.

Avant de pouvoir attribuer des balises LF aux ressources du catalogue de données ou accorder des autorisations aux principaux, vous devez définir des balises LF. Seul un administrateur de lac de données ou un directeur disposant des autorisations de création de balises LF peut créer des balises LF.

Créateurs de LF-Tag

Le créateur de balises LF est un administrateur principal non administrateur autorisé à créer et à gérer des balises LF. Les administrateurs de data lake peuvent ajouter des créateurs de balises LF à l'aide de la console Lake Formation ou de la CLI. Les créateurs de balises LF ont des autorisations implicites de Lake Formation pour mettre à jour et supprimer les balises LF, pour attribuer des balises LF aux ressources et pour accorder des autorisations de balise LF et des autorisations de valeur de balise LF à d'autres directeurs.

Grâce aux rôles de créateur de balises LF, les administrateurs des lacs de données peuvent déléguer des tâches de gestion des balises, telles que la création et la mise à jour des clés et des valeurs des balises, à des personnes qui ne sont pas des administrateurs principaux. Les administrateurs de lacs de données peuvent également accorder aux créateurs de balises LF des autorisations susceptibles d'être accordées`Create LF-Tag`. Ensuite, le créateur de balises LF peut accorder l'autorisation de créer des balises LF à d'autres personnes principales.

Vous pouvez accorder deux types d'autorisations sur les LF-Tags :

- Autorisations LF-Tag - `Create LF-Tag`Alter, et. `Drop` Ces autorisations sont requises pour créer, mettre à jour et supprimer des balises LF.

  Les administrateurs du lac de données et les créateurs de balises LF disposent implicitement de ces autorisations sur les balises LF qu'ils créent et peuvent accorder ces autorisations explicitement aux principaux pour gérer les balises dans le lac de données.
- Autorisations de paire clé-valeur LF-Tag -`Assign`, et. `Describe` `Grant with LF-Tag expressions` Ces autorisations sont nécessaires pour attribuer des balises LF aux bases de données, aux tables et aux colonnes du catalogue de données, et pour accorder des autorisations sur les ressources aux principaux utilisant le contrôle d'accès basé sur les balises Lake Formation. Les créateurs de balises LF reçoivent implicitement ces autorisations lors de la création de balises LF.

Après avoir reçu l'`Create LF-Tag`autorisation et créé avec succès les balises LF, le créateur de balises LF peut attribuer des balises LF aux ressources et accorder des autorisations de balises LF (`Create LF-Tag`, `AlterDrop`, et) à d'autres personnes non administratives pour gérer les balises dans le lac de données. Vous pouvez gérer les balises LF à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface ()AWS CLI.

> **Note**
>
> Les administrateurs des lacs de données disposent des autorisations implicites de Lake Formation pour créer, mettre à jour et supprimer des balises LF, attribuer des balises LF aux ressources et accorder des autorisations LF-Tag aux principaux.

Pour connaître les meilleures pratiques et les considérations, voir Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation

Rubriques

- Ajouter des créateurs de LF-Tag
- Création de balises LF
- Mise à jour des balises LF
- Supprimer des balises LF
- Répertorier les balises LF
- Affectation de balises LF aux ressources du catalogue de données
- Afficher les balises LF attribuées à une ressource
- Afficher les ressources auxquelles un tag LF est attribué
- Cycle de vie d'un LF-Tag
- Comparaison entre le contrôle d'accès basé sur les balises Lake Formation et le contrôle d'accès basé sur les attributs IAM

> **Consultez aussi**
>
> - Octroi, révocation et listage des autorisations relatives aux valeurs des balises LF
> - Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC
> - Contrôle d'accès basé sur des balises Lake Formation

Ajouter des créateurs de LF-Tag

Par défaut, les administrateurs des lacs de données peuvent créer, mettre à jour et supprimer des balises LF, attribuer des balises aux ressources du catalogue de données et accorder des

autorisations de balises aux principaux. Si vous souhaitez déléguer les opérations de création et de gestion des balises à des administrateurs non administrateurs, l'administrateur du lac de données peut créer des rôles de créateur de balises LF et accorder à Lake Formation l'`Create LF-Tag`autorisation d'accéder à ces rôles. Avec une `Create LF-Tag` autorisation pouvant être accordée, les créateurs de balises LF peuvent déléguer les tâches de création et de maintenance des balises à d'autres personnes non administratives.

> **Note**
>
> Les autorisations accordées entre comptes ne peuvent inclure que `Describe` des `Associate` autorisations. Vous ne pouvez pas accorder `Create LF-Tag`Drop,`Alter`, et `Grant with LFTag expressions` d'autorisations aux principaux d'un autre compte.

Rubriques

- Autorisations IAM requises pour créer des balises LF
- Ajouter des créateurs de LF-Tag

> **Consultez aussi**
>
> - Octroi, révocation et listage des autorisations relatives aux valeurs des balises LF
> - Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC
> - Contrôle d'accès basé sur des balises Lake Formation

Autorisations IAM requises pour créer des balises LF

Vous devez configurer les autorisations pour permettre à un directeur de Lake Formation de créer des balises LF. Ajoutez la déclaration suivante à la politique d'autorisation pour le principal qui doit être un créateur de balises LF.

> **Note**
>
> Bien que les administrateurs des lacs de données disposent des autorisations implicites de Lake Formation pour créer, mettre à jour et supprimer des balises LF, pour attribuer

des balises LF aux ressources et pour accorder des balises LF aux principaux, les administrateurs des lacs de données ont également besoin des autorisations IAM suivantes.

Pour plus d'informations, consultez Référence des personnalités de Lake Formation et des autorisations IAM.

```
{
"Sid": "Transformational",
"Effect": "Allow",
    "Action": [
        "lakeformation:AddLFTagsToResource",
        "lakeformation:RemoveLFTagsFromResource",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:CreateLFTag",
        "lakeformation:GetLFTag",
        "lakeformation:UpdateLFTag",
        "lakeformation:DeleteLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
     ]
 }
```

Les principaux qui attribuent des balises LF aux ressources et accordent des balises LF aux principaux doivent avoir les mêmes autorisations, à l'exception des autorisations, et. `CreateLFTag` `UpdateLFTag` `DeleteLFTag`

Ajouter des créateurs de LF-Tag

Un créateur de balise LF peut créer une balise LF, mettre à jour la clé et les valeurs de balise, supprimer des balises, associer des balises aux ressources du catalogue de données et accorder des autorisations sur les ressources du catalogue de données aux principaux à l'aide de la méthode LF-TBAC. Le créateur du LF-Tag peut également accorder ces autorisations aux principaux.

Vous pouvez créer des rôles de créateur de balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

console

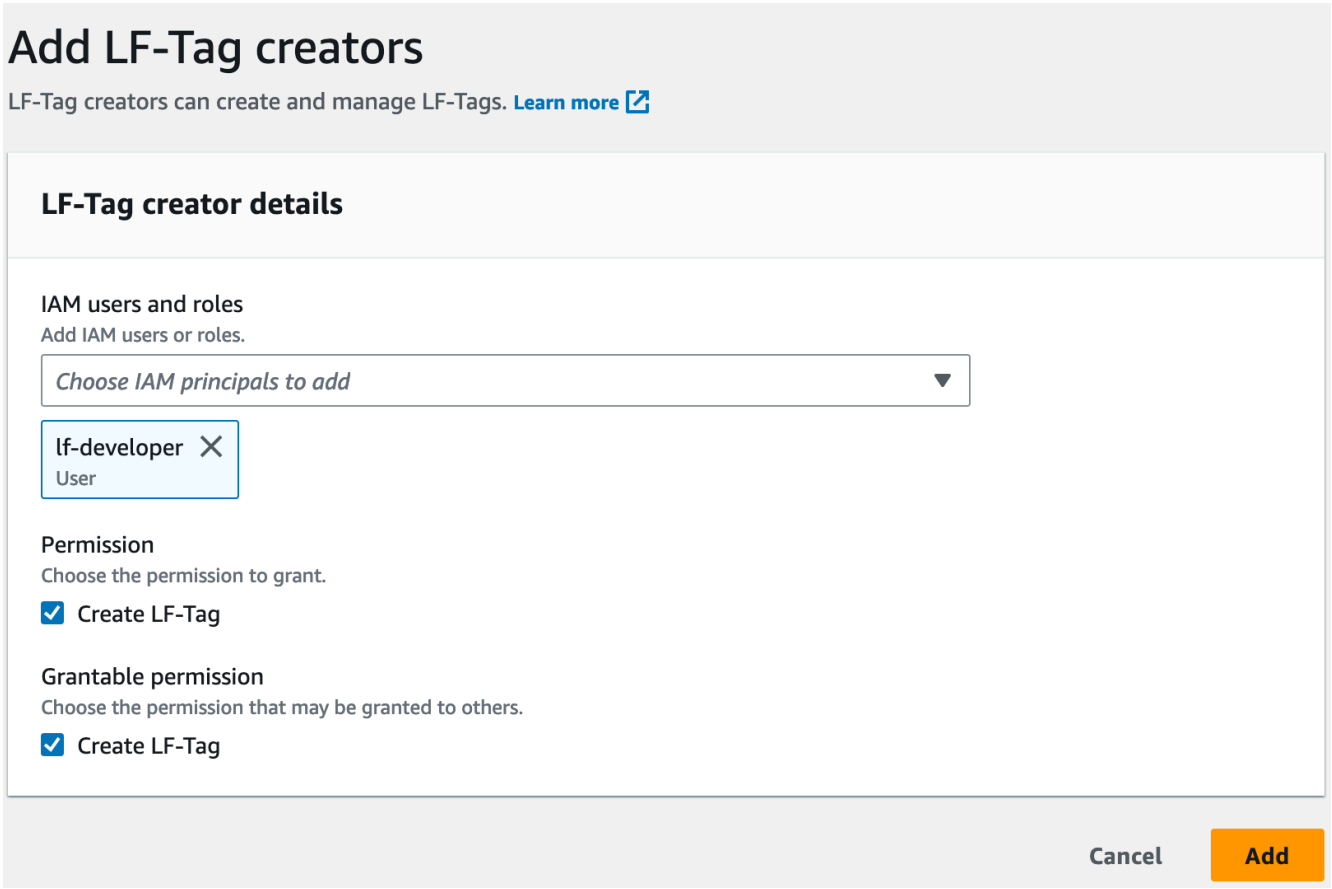
Pour ajouter un créateur de balises LF

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du datalake.

2. Dans le volet de navigation, sous Permissions, sélectionnez LF-Tags and permissions.

   Sur la page des balises LF et des autorisations, choisissez la section des créateurs de balises LF, puis choisissez Ajouter des créateurs de balises LF.



3. Sur la page Ajouter des créateurs de balises LF, choisissez un rôle ou un utilisateur IAM disposant des autorisations requises pour créer des balises LF.
4. Activez `Create LF-Tag` la case à cocher d'autorisation.
5. (Facultatif) Pour permettre aux principaux sélectionnés d'accorder des `Create LF-Tag` autorisations aux principaux, choisissez Autorisation accordable. `Create LF-Tag`
6. Choisissez Ajouter.

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
    },
    "Resource": {
        "Catalog": {}
    },
    "Permissions": [
        "CreateLFTag"
    ],
    "PermissionsWithGrantOption": [
        "CreateLFTag"
    ]
}
```

Les autorisations disponibles pour un rôle de créateur de balises LF sont les suivantes :

Autorisation	Description
`Drop`	Un directeur disposant de cette autorisation sur un tag LF peut supprimer un tag LF du lac de données. Le principal obtient une `Describe` autorisation implicite sur toutes les valeurs de balise d'une ressource LF-Tag.
`Alter`	Un directeur disposant de cette autorisation sur une balise LF peut ajouter ou supprimer une valeur de balise à une balise LF. Le principal obtient une `Alter` autorisation implicite sur toutes les valeurs de balise d'une balise LF.
`Describe`	Un directeur disposant de cette autorisation sur un LF-Tag peut voir le LF-Tag et ses valeurs lorsqu'il attribue des LF-Tags à des ressources ou accorde des autorisations sur des LF-Tags. Vous pouvez accorder une autorisation `Describe` sur toutes les valeurs clés ou sur des valeurs spécifiques.

Autorisation	Description
`Associate`	Un directeur disposant de cette autorisation sur une balise LF peut attribuer la balise LF à une ressource de catalogue de données. Accorder `Associate` implicitement des subventions`Describe`.
`Grant with LF-Tag expression`	Un directeur disposant de cette autorisation sur une balise LF peut accorder des autorisations sur les ressources d'un catalogue de données à l'aide de la clé et des valeurs de la balise LF. Accorder `Grant with LF-Tag expression` implicitement des subventions`Describe`.

Ces autorisations peuvent être accordées. Un directeur qui a obtenu ces autorisations avec l'option d'octroi peut les accorder à d'autres principaux.

Création de balises LF

Toutes les balises LF doivent être définies dans Lake Formation avant de pouvoir être utilisées. Une balise LF se compose d'une clé et d'une ou de plusieurs valeurs possibles pour la clé.

Une fois que l'administrateur du lac de données a configuré les autorisations IAM et Lake Formation requises pour le rôle de créateur de balises LF, le principal peut créer une balise LF. Le créateur de la balise LF obtient l'autorisation implicite de mettre à jour ou de supprimer toute valeur de balise de la balise LF et de supprimer la balise LF.

Vous pouvez créer des balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

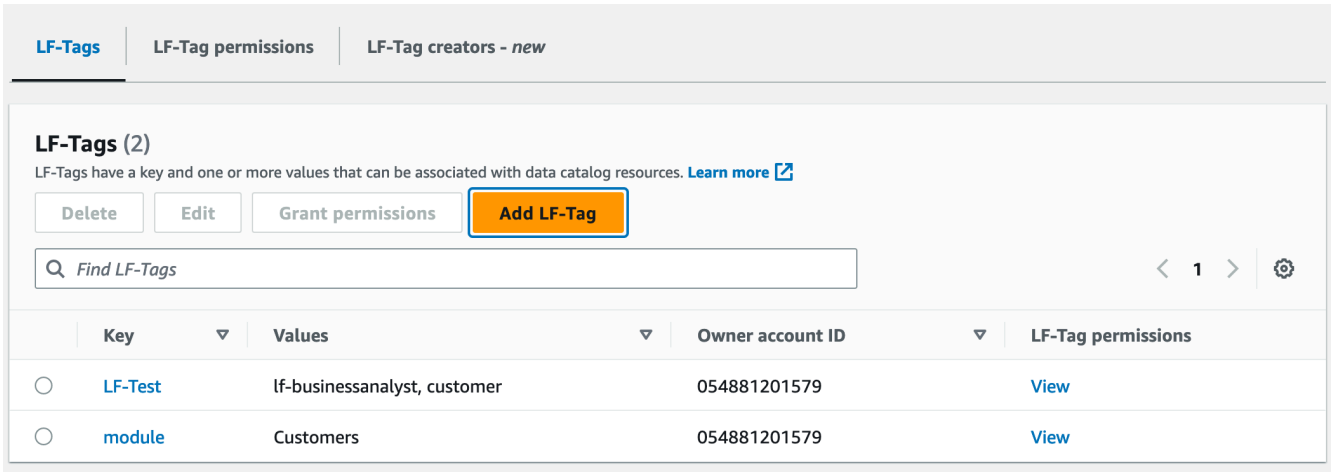
Console

Pour créer un LF-Tag

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant que principal avec les autorisations de création de balises LF ou en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Balises LF et autorisations, choisissez LF-Tags.

   La page LF-Tags apparaît.



3. Choisissez Ajouter un tag LF.
4. Dans la boîte de dialogue Ajouter une balise LF, entrez une clé et une ou plusieurs valeurs.

   Chaque clé doit avoir au moins une valeur. Pour saisir plusieurs valeurs, entrez une liste séparée par des virgules puis appuyez sur Entrée, ou entrez une valeur à la fois et choisissez Ajouter après chacune d'elles. Le nombre maximum de valeurs autorisées est de 1 000.
5. Choisissez Ajouter une balise.

AWS CLI

Pour créer un LF-Tag

- Entrez une `create-lf-tag` commande.

  L'exemple suivant crée une balise LF avec une clé `module` et des valeurs `Customers` et. `Orders`

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

En tant que créateur de balise, le principal obtient l'`Alter`autorisation d'utiliser cette balise LF et peut mettre à jour ou supprimer toute valeur de balise de cette balise LF. Le créateur principal du tag LF peut également `Alter` autoriser un autre principal à mettre à jour et à supprimer les valeurs des balises sur ce tag LF.

Mise à jour des balises LF

Vous mettez à jour une balise LF sur laquelle vous avez l'`Alter`autorisation en ajoutant ou en supprimant des valeurs clés autorisées. Vous ne pouvez pas modifier la touche LF-Tag. Pour modifier la clé, supprimez le tag LF et ajoutez-en un avec la clé requise. Outre l'`Alter`autorisation, vous avez également besoin de l'autorisation `lakeformation:UpdateLFTag` IAM pour mettre à jour les valeurs.

Lorsque vous supprimez une valeur de balise LF, aucune vérification n'est effectuée pour vérifier la présence de cette valeur de balise LF sur aucune ressource du catalogue de données. Si la valeur de balise LF supprimée est associée à une ressource, elle n'est plus visible pour cette ressource, et les principaux auxquels des autorisations ont été accordées sur cette paire clé-valeur ne disposent plus de ces autorisations.

Avant de supprimer une valeur de balise LF, vous pouvez éventuellement utiliser la `remove-lf-tags-from-resource`commande pour supprimer la balise LF des ressources du catalogue de données contenant la valeur que vous souhaitez supprimer, puis reétiqueter la ressource avec les valeurs que vous souhaitez conserver.

Seuls les administrateurs du lac de données, le créateur du tag LF et les principaux détenteurs d'`Alter`autorisations sur le tag LF peuvent mettre à jour un tag LF.

Vous pouvez mettre à jour une balise LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

Console

Pour mettre à jour un LF-Tag (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données, créateur du tag LF ou principal `Alter` autorisé à utiliser le tag LF.
2. Dans le volet de navigation, sous Balises LF et autorisations, choisissez LF-Tags.
3. Sur la page Balises LF, sélectionnez une balise LF, puis choisissez Modifier.
4. Dans la boîte de dialogue Modifier le tag LF, ajoutez ou supprimez des valeurs du tag LF.

Pour ajouter plusieurs valeurs, dans le champ Valeurs, entrez une liste séparée par des virgules et appuyez sur Entrée, ou entrez une valeur à la fois ou choisissez Ajouter après chacune d'elles.

5. Choisissez Enregistrer.

AWS CLI

Pour mettre à jour un LF-Tag ()AWS CLI

- Entrez une `update-lf-tag` commande. Fournissez l'un des arguments suivants ou les deux :
  - `--tag-values-to-add`
  - `--tag-values-to-delete`

Example

L'exemple suivant remplace la valeur `vp` par la valeur de `vice-president` la clé LF-Tag. `level`

```
aws lakeformation update-lf-tag --tag-key level --tag-values-to-add vice-president
--tag-values-to-delete vp
```

Supprimer des balises LF

Vous pouvez supprimer les balises LF qui ne sont plus utilisées. Aucune vérification n'est effectuée pour vérifier la présence de la balise LF sur une ressource de catalogue de données. Si le tag LF supprimé est associé à une ressource, il n'est plus visible pour la ressource, et les principaux auxquels des autorisations ont été accordées sur ce tag LF ne le sont plus.

Avant de supprimer une balise LF, vous pouvez éventuellement utiliser la `remove-lf-tags-from-resource`commande pour supprimer la balise LF de toutes les ressources.

Seuls les administrateurs du lac de données, le créateur de la balise LF ou un principal `Drop` autorisé à utiliser la balise LF peuvent supprimer une balise LF. Outre l'`Drop`autorisation, le principal doit également disposer de l'autorisation `lakeformation:DeleteLFTag` IAM pour supprimer un tag LF.

Vous pouvez supprimer une balise LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

Console

Pour supprimer un tag LF (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données.
2. Dans le volet de navigation, sous Balises LF et autorisations, choisissez LF-Tags.
3. Sur la page Balises LF, sélectionnez une balise LF, puis choisissez Supprimer.
4. Dans l'environnement Delete tag ? boîte de dialogue, pour confirmer la suppression, entrez la valeur de la clé LF-Tag dans le champ désigné, puis choisissez Supprimer.

AWS CLI

Pour supprimer un tag LF ()AWS CLI

- Entrez une `delete-lf-tag` commande. Indiquez la clé du tag LF à supprimer.

  Example

  L'exemple suivant supprime le tag LF avec la clé. `region`

  ```
  aws lakeformation delete-lf-tag --tag-key region
  ```

Répertorier les balises LF

Vous pouvez répertorier les balises LF pour lesquelles vous avez les `Associate` autorisations `Describe` ou les autorisations. Les valeurs répertoriées avec chaque clé LF-Tag sont les valeurs pour lesquelles vous avez des autorisations.

Le créateur de balises LF dispose d'autorisations implicites pour voir les balises LF qu'il a créées.

Les administrateurs du data lake peuvent voir toutes les balises LF définies dans le AWS compte local et toutes les balises LF pour lesquelles les `Associate` autorisations `Describe` et les

autorisations ont été accordées au compte local à partir de comptes externes. L'administrateur du lac de données peut voir toutes les valeurs de toutes les balises LF.

Vous pouvez répertorier les balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

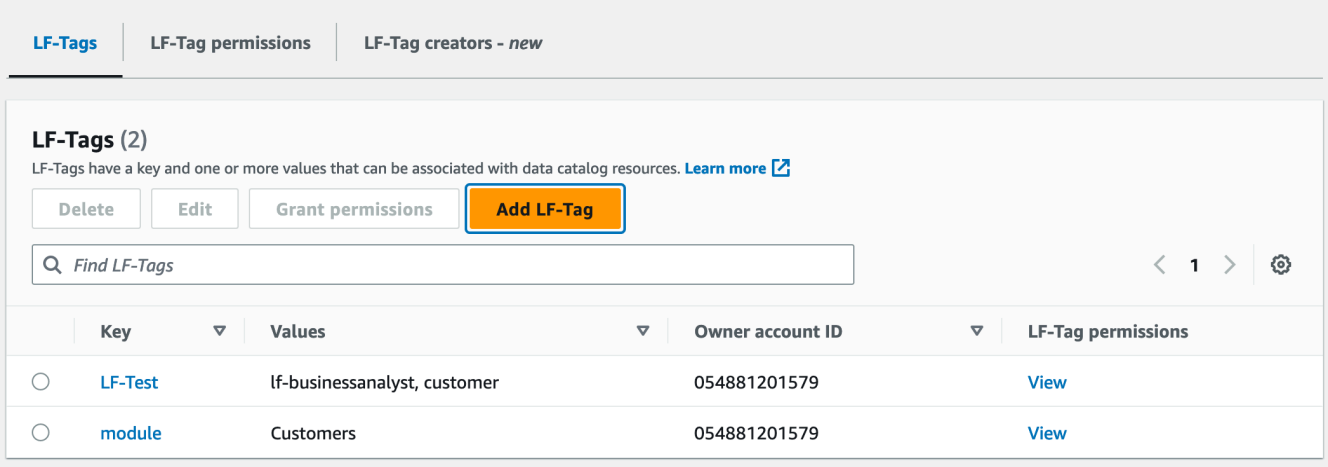
Console

Pour répertorier les balises LF (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant que créateur de balises LF, en tant qu'administrateur du lac de données ou en tant que principal ayant obtenu des autorisations sur les balises LF et disposant de l'autorisation IAM. `lakeformation:ListLFTags`
2. Dans le volet de navigation, sous Balises LF et autorisations, choisissez LF-Tags.

   La page LF-Tags apparaît.



   Consultez la colonne ID du compte propriétaire pour déterminer les balises LF qui ont été partagées avec votre compte à partir d'un compte externe.

AWS CLI

Pour répertorier les balises LF ()AWS CLI

- Exécutez la commande suivante en tant qu'administrateur du lac de données ou en tant que principal ayant obtenu des autorisations sur les balises LF et disposant de l'autorisation `lakeformation:ListLFTags` IAM.

```
aws lakeformation list-lf-tags
```

La sortie est similaire à ce qui suit.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ]
}
```

Pour voir également les balises LF attribuées à partir de comptes externes, incluez l'option de commande. `--resource-share-type ALL`

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

La sortie est similaire à ce qui suit. Notez la `NextToken` clé, qui indique qu'il y en a d'autres à répertorier.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ],
    "NextToken": "eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ=="
}
```

Répétez la commande et ajoutez l'`--next-token`argument pour afficher les balises LF locales restantes et les balises LF accordées par des comptes externes. Les balises LF provenant de comptes externes se trouvent toujours sur une page séparée.

```
aws lakeformation list-lf-tags --resource-share-type ALL
--next-token eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ==
```

```
{
    "LFTags": [
        {
            "CatalogId": "123456789012",
            "TagKey": "region",
            "TagValues": [
```

```
                "central",
                "south"
            ]
        }
    ]
}
```

API

Vous pouvez utiliser les SDK disponibles pour Lake Formation pour répertorier les balises que le demandeur est autorisé à consulter.

```
import boto3

client = boto3.client('lakeformation')
...

response = client.list_lf_tags(
    CatalogId='string',
    ResourceShareType='ALL',
    MaxResults=50'
)
```

Cette commande renvoie un `dict` objet dont la structure est la suivante :

```
{
    'LFTags': [
        {
            'CatalogId': 'string',
            'TagKey': 'string',
            'TagValues': [
                'string',
            ]
        },
    ],
    'NextToken': 'string'
}
```

Pour plus d'informations sur les autorisations requises, consultez Référence des personnalités de Lake Formation et des autorisations IAM.

Affectation de balises LF aux ressources du catalogue de données

Vous pouvez attribuer des balises LF aux ressources du catalogue de données (bases de données, tables et colonnes) pour contrôler l'accès à ces ressources. Seuls les principaux auxquels des balises LF correspondantes sont attribuées (et les principaux auxquels l'accès est accordé avec la méthode de ressource nommée) peuvent accéder aux ressources.

Si une table hérite d'une balise LF d'une base de données ou si une colonne hérite d'une balise LF d'une table, vous pouvez remplacer la valeur héritée en affectant une nouvelle valeur à la clé de balise LF.

Le nombre maximum de balises LF que vous pouvez attribuer à une ressource est de 50.

Rubriques

- Exigences relatives à la gestion des balises attribuées aux ressources
- Attribuer des balises LF à une colonne de tableau
- Attribuer des balises LF à une ressource de catalogue de données
- Mettre à jour les balises LF d'une ressource
- Supprimer le tag LF d'une ressource

Exigences relatives à la gestion des balises attribuées aux ressources

Pour attribuer une balise LF à une ressource du catalogue de données, vous devez :

- Ayez l'`ASSOCIATE`autorisation de Lake Formation sur le LF-Tag.
- Ayez l'`lakeformation:AddLFTagsToResource`autorisation IAM.
- Avoir de la colle : GetDatabase autorisation sur une base de données Glue.
- Soyez le propriétaire (créateur) de la ressource, `Super` détenez l'autorisation Lake Formation sur la ressource avec l'`GRANT`option, ou disposez des autorisations suivantes avec l'`GRANT`option :
  - Pour les bases de données du même AWS compte : `DESCRIBE`,`CREATE_TABLE`,`ALTER`, et `DROP`
  - Pour les bases de données d'un compte externe :`DESCRIBE`, `CREATE_TABLE` et `ALTER`
  - Pour les tables (et les colonnes) : `DESCRIBE`,`ALTER`,`DROP`,`INSERT`,`SELECT`, et `DELETE`

De plus, le tag LF et la ressource à laquelle il est attribué doivent se trouver dans le même AWS compte.

Pour supprimer une balise LF d'une ressource de catalogue de données, vous devez répondre à ces exigences et disposer de l'autorisation `lakeformation:RemoveLFTagsFromResource` IAM.

Attribuer des balises LF à une colonne de tableau

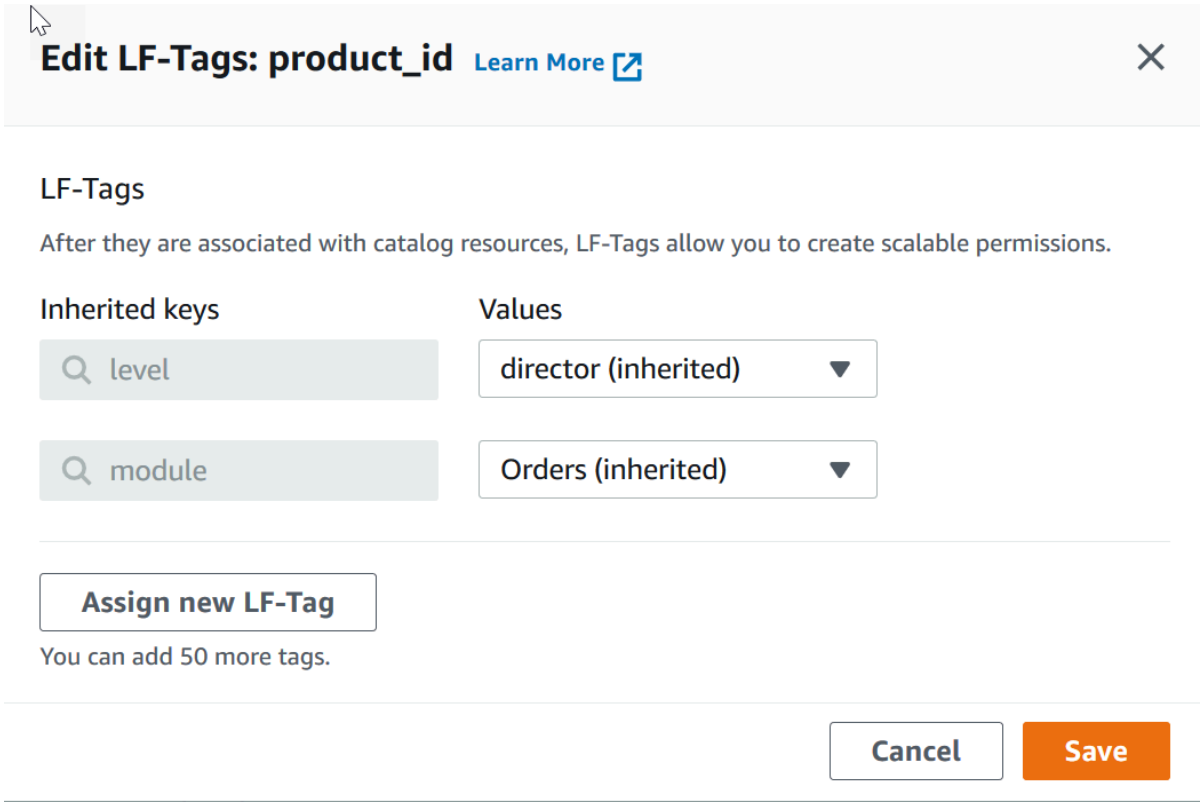
Pour attribuer des balises LF à une colonne de tableau (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

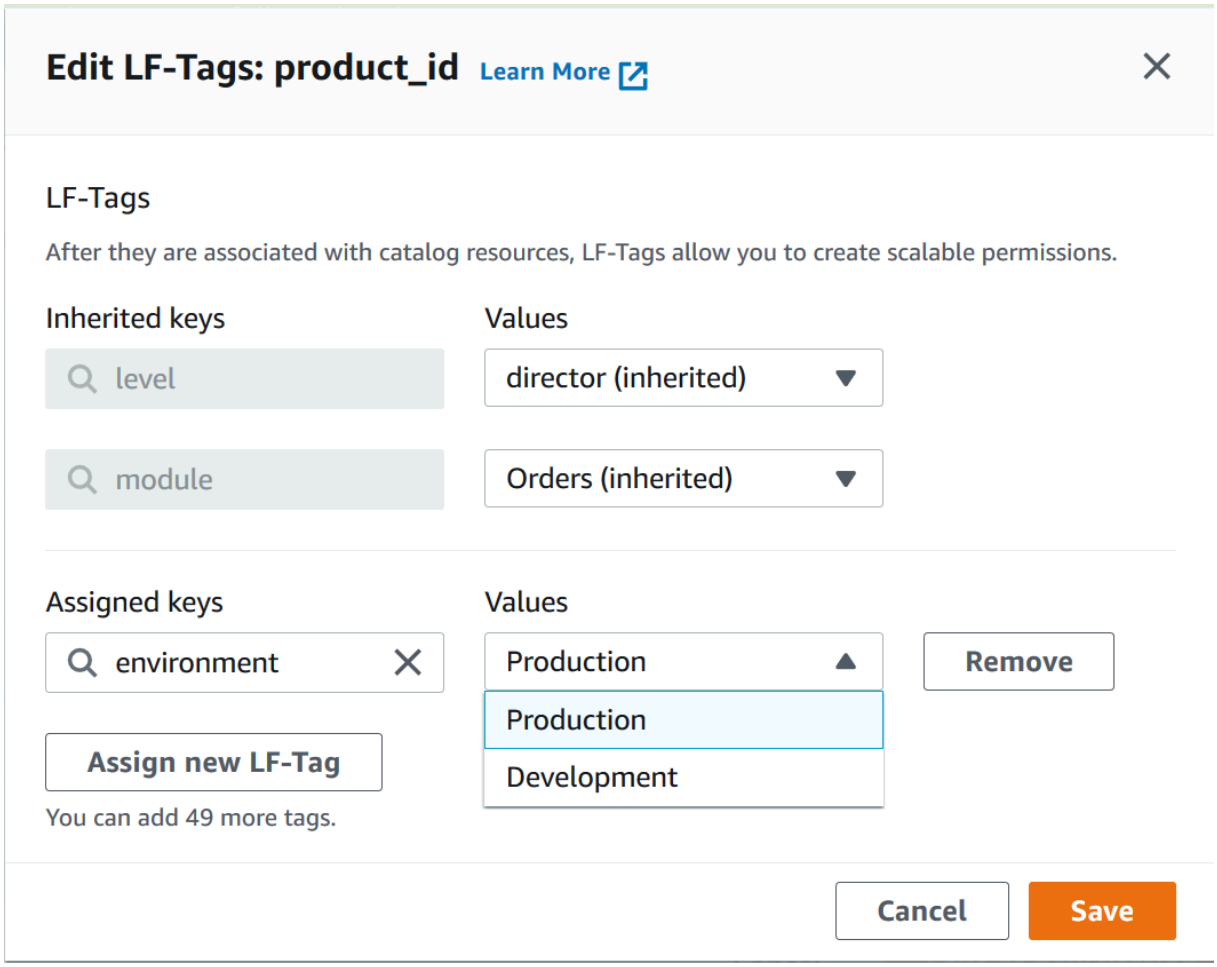
   Connectez-vous en tant qu'utilisateur répondant aux exigences répertoriées ci-dessus.
2. Dans le volet de navigation, choisissez Tables.
3. Choisissez un nom de table (et non le bouton d'option situé à côté du nom de la table).
4. Sur la page des détails de la table, dans la section Schéma, choisissez Modifier le schéma.
5. Sur la page Modifier le schéma, sélectionnez une ou plusieurs colonnes, puis choisissez Modifier les balises.

   > **Note**
   >
   > Si vous avez l'intention d'ajouter ou de supprimer des colonnes et d'enregistrer une nouvelle version, faites-le d'abord. Modifiez ensuite les balises LF.

   La boîte de dialogue Modifier les balises LF apparaît et affiche toutes les balises LF héritées du tableau.



6. (Facultatif) Dans la liste des valeurs située à côté d'un champ Clés héritées, choisissez une valeur qui remplacera la valeur héritée.
7. (Facultatif) Choisissez Attribuer un nouveau tag LF. Ensuite, pour Clés attribuées, choisissez une clé, et pour Valeurs, choisissez une valeur pour la clé.



8. (Facultatif) Choisissez à nouveau Attribuer un nouveau tag LF pour ajouter un autre tag LF.
9. Choisissez Enregistrer.

Attribuer des balises LF à une ressource de catalogue de données

Console

Pour attribuer des balises LF à une base de données ou à une table de catalogue de données

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'utilisateur répondant aux exigences répertoriées précédemment.
2. Dans le volet de navigation, sous Catalogue de données, effectuez l'une des opérations suivantes :

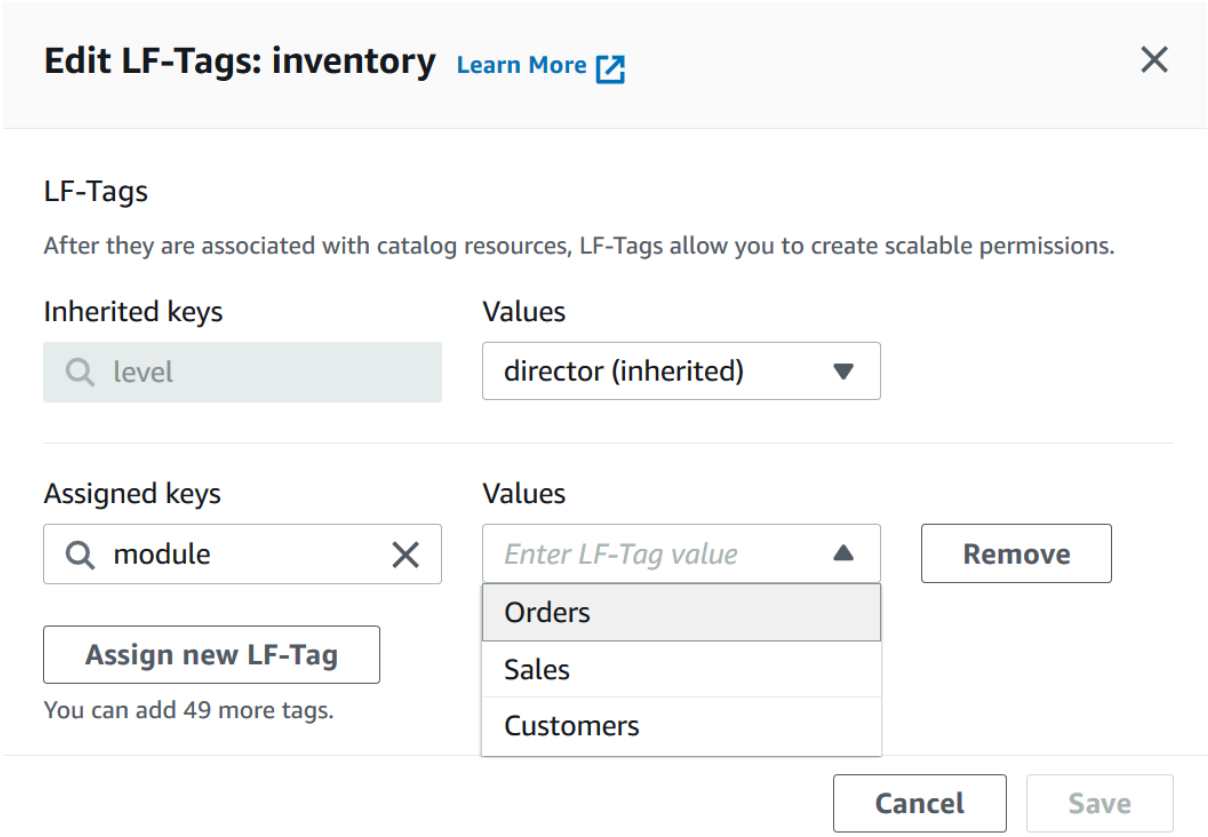
   - Pour attribuer des balises LF aux bases de données, choisissez Databases.

- Pour attribuer des balises LF aux tables, choisissez Tables.

3. Choisissez une base de données ou une table, puis dans le menu Actions, choisissez Modifier les balises.

   La boîte de dialogue Modifier les balises LF : ***nom de ressource s'affiche***.

   Si une table hérite des balises LF de la base de données qui la contient, la fenêtre affiche les balises LF héritées. Sinon, le texte « Aucune balise LF héritée n'est associée à la ressource » s'affiche.



4. (Facultatif) Si une table possède des balises LF héritées, dans la liste des valeurs située à côté d'un champ Clés héritées, vous pouvez choisir une valeur qui remplacera la valeur héritée.
5. Pour attribuer de nouvelles balises LF, procédez comme suit :

   a. Choisissez Attribuer un nouveau tag LF.

   b. Dans le champ Clés assignées, choisissez une clé LF-Tag, et dans le champ Valeurs, choisissez une valeur.

c. (Facultatif) Choisissez à nouveau Assigner un nouveau tag LF pour attribuer un tag LF supplémentaire.

6. Choisissez Enregistrer.

AWS CLI

Pour attribuer des balises LF à une ressource de catalogue de données

- Exécutez la commande `add-lf-tags-to-resource`.

  L'exemple suivant attribue la balise LF à la table `module=orders` de la base de données`orders`. `erp` Il utilise la syntaxe du raccourci pour l'`--lf-tags`argument. La `CatalogID` propriété pour `--lf-tags` est facultative. S'il n'est pas fourni, l'ID de catalogue de la ressource (dans ce cas, la table) est supposé.

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"orders"}}' --lf-tags
 CatalogId=111122223333,TagKey=module,TagValues=orders
```

  Le résultat suivant est le résultat si la commande aboutit.

```
{
    "Failures": []
}
```

  L'exemple suivant affecte deux balises LF à la `sales` table et utilise la syntaxe JSON pour l'argument. `--lf-tags`

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"sales"}}' --lf-tags '[{"TagKey":
 "module","TagValues": ["sales"]},{"TagKey": "environment","TagValues":
 ["development"]}]'
```

  L'exemple suivant affecte la balise LF `level=director` à la `total` colonne de la table. `sales`

```
aws lakeformation add-lf-tags-to-resource --resource '{ "TableWithColumns":
 {"DatabaseName":"erp", "Name":"sales", "ColumnNames":["total"]}}' --lf-tags
 TagKey=level,TagValues=director
```

Mettre à jour les balises LF d'une ressource

Pour mettre à jour une balise LF pour une ressource de catalogue de données ()AWS CLI

- Utilisez la `add-lf-tags-to-resource` commande, comme décrit dans la procédure précédente.

  L'ajout d'un tag LF avec la même clé qu'un tag LF existant, mais avec une valeur différente met à jour la valeur existante.

Supprimer le tag LF d'une ressource

Pour supprimer une balise LF pour une ressource de catalogue de données ()AWS CLI

- Exécutez la commande `remove-lf-tags-from-resource`.

  Si une table possède une valeur de balise LF qui remplace la valeur héritée de la base de données parent, la suppression de cette balise LF de la table restaure la valeur héritée. Ce comportement s'applique également à une colonne qui remplace les valeurs clés héritées de la table.

  L'exemple suivant supprime le tag LF `level=director` de la `total` colonne du `sales` tableau. La `CatalogID` propriété pour `--lf-tags` est facultative. S'il n'est pas fourni, l'ID de catalogue de la ressource (dans ce cas, la table) est supposé.

  ```
  aws lakeformation remove-lf-tags-from-resource
  --resource ' { "TableWithColumns":
  { "DatabaseName": "erp",  "Name": "sales",  "ColumnNames":[ "total"]}}'
  --lf-tags  CatalogId=111122223333,TagKey=level,TagValues=director
  ```

Afficher les balises LF attribuées à une ressource

Vous pouvez afficher les balises LF attribuées à une ressource de catalogue de données. Vous devez disposer de l'`ASSOCIATE`autorisation `DESCRIBE` or sur un LF-Tag pour le visualiser.

Console

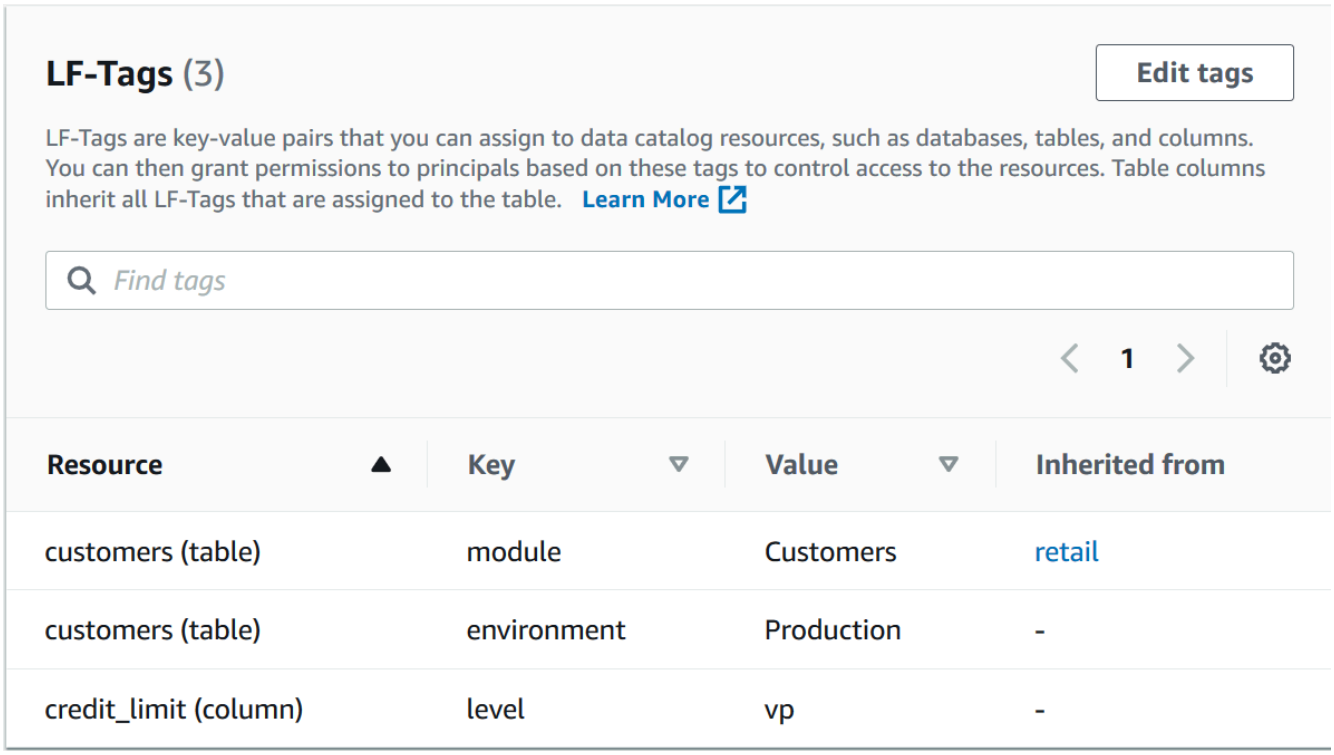
Pour afficher les balises LF attribuées à une ressource (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données, propriétaire de la ressource ou en tant qu'utilisateur ayant obtenu les autorisations de Lake Formation sur la ressource.
2. Dans le volet de navigation, sous le titre Catalogue de données, effectuez l'une des opérations suivantes :

   - Pour afficher les balises LF attribuées à une base de données, sélectionnez Databases.
   - Pour afficher les balises LF attribuées à une table, choisissez Tables.
3. Sur la page Tables ou bases de données, choisissez le nom de la base de données ou de la table. Ensuite, sur la page de détails, faites défiler la page vers le bas jusqu'à la section LF-Tags.

   La capture d'écran suivante montre les balises LF attribuées à une `customers` table contenue dans la `retail` base de données. La `module` balise LF est héritée de la base de données. Le `level=vp` tag LF est attribué à la `credit_limit` colonne.



AWS CLI

Pour afficher les balises LF attribuées à une ressource ()AWS CLI

- Utilisez une commande similaire à la suivante.

```
aws lakeformation get-resource-lf-tags --show-assigned-lf-tags --
resource '{ "Table": {"CatalogId":"111122223333", "DatabaseName":"erp",
 "Name":"sales"}}'
```

La commande renvoie le résultat suivant.

```
{
    "TableTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "sales"
            ]
        },
        {
```

```
            "CatalogId": "111122223333",
            "TagKey": "environment",
            "TagValues": [
                "development"
            ]
        }
    ],
    "ColumnTags": [
        {
            "Name": "total",
            "Tags": [
                {
                    "CatalogId": "111122223333",
                    "TagKey": "level",
                    "TagValues": [
                        "director"
                    ]
                }
            ]
        }
    ]
}
```

Cette sortie affiche uniquement les balises LF attribuées de manière explicite et non héritées. Si vous souhaitez voir toutes les balises LF sur toutes les colonnes, y compris les balises LF héritées, omettez cette option. `--show-assigned-lf-tags`

Afficher les ressources auxquelles un tag LF est attribué

Vous pouvez afficher toutes les ressources du catalogue de données auxquelles une clé LF-Tag particulière est attribuée. Pour ce faire, vous devez disposer des autorisations Lake Formation suivantes :

- `Describe`ou `Associate` sur le LF-Tag.
- `Describe`ou toute autre autorisation de Lake Formation sur la ressource.

En outre, vous devez disposer des autorisations AWS Identity and Access Management (IAM) suivantes :

- `lakeformation:SearchDatabasesByLFTags`

- `lakeformation:SearchTablesByLFTags`

Console

Pour afficher les ressources auxquelles un tag LF est attribué (console)

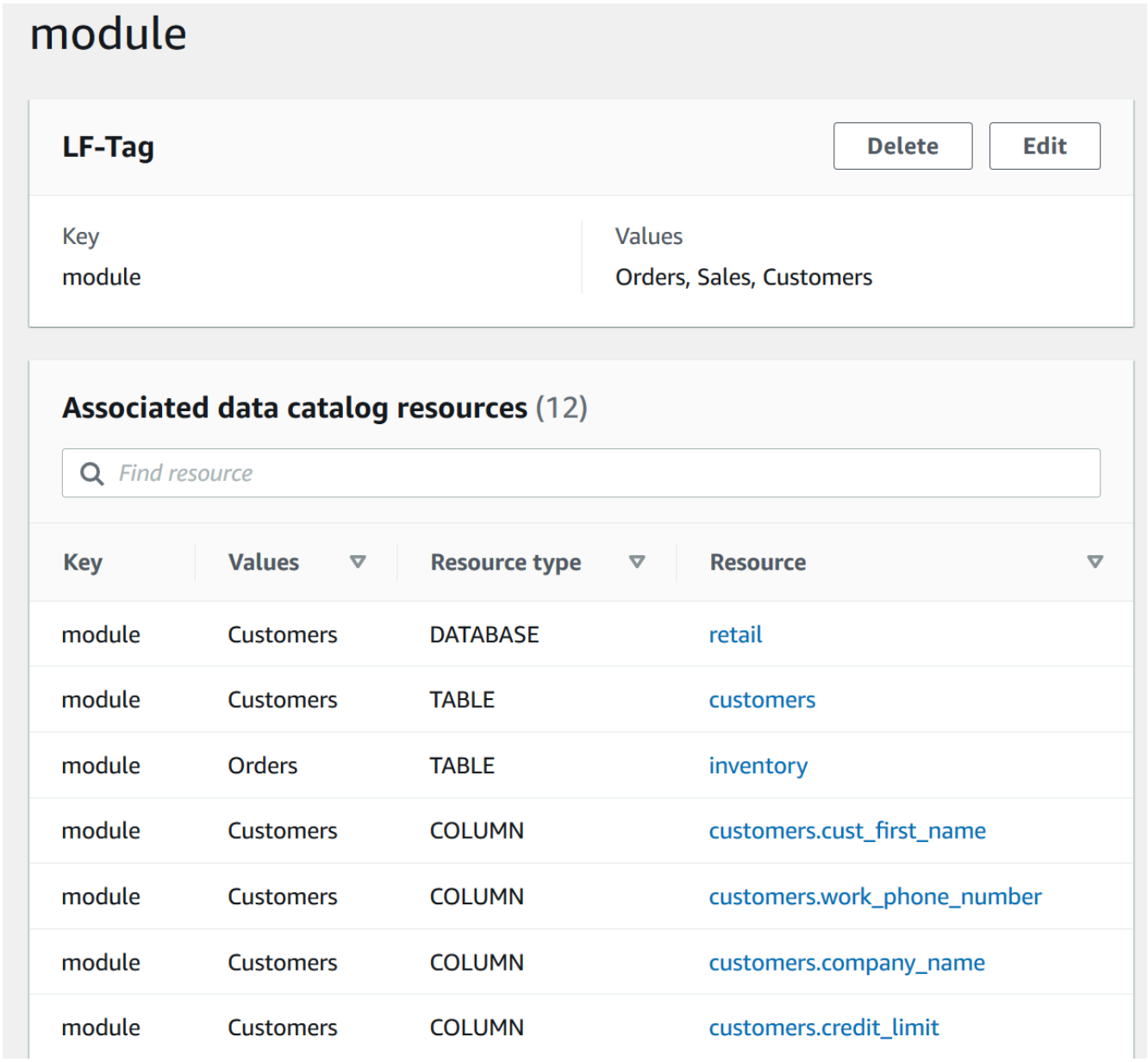
1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur répondant aux exigences répertoriées précédemment.

2. Dans le volet de navigation, sous Permissions et balises LF et autorisations, choisissez LF-Tags.

3. Choisissez une touche LF-Tag (et non le bouton d'option à côté du nom de la clé).

   La page de détails du LF-Tag affiche une liste des ressources auxquelles le LF-Tag a été attribué.



AWS CLI

Pour afficher les ressources auxquelles un tag LF est attribué

- Exécutez une `search-databases-by-lf-tags` commande `search-tables-by-lf-tags` or.

Example

L'exemple suivant répertorie les tables et les colonnes auxquelles le `level=vp` tag LF est attribué. Pour chaque table ou colonne répertoriée, toutes les balises LF attribuées à la table ou à la colonne sont affichées, et pas seulement l'expression de recherche.

```
aws lakeformation search-tables-by-lf-tags --expression
 TagKey=level,TagValues=vp
```

Pour plus d'informations sur les autorisations requises, consultez Référence des personnalités de Lake Formation et des autorisations IAM.

Cycle de vie d'un LF-Tag

1. Michael, le créateur du LF-Tag, crée un LF-Tag. `module=Customers`
2. Michael octroie `Associate` le LF-Tag à l'ingénieur de données Eduardo. Accorder `Associate` implicitement des subventions`Describe`.
3. Michael accorde une subvention `Super` sur la table `Custs` à Eduardo avec l'option de subvention, afin qu'Eduardo puisse attribuer des balises LF à la table. Pour plus d'informations, consultez Affectation de balises LF aux ressources du catalogue de données.
4. Eduardo attribue le tag LF à `module=customers` la table. `Custs`
5. Michael accorde la subvention suivante à l'ingénieure de données Sandra (en pseudo-code).

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. Sandra accorde la subvention suivante à l'analyste de données Maria.

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Maria peut désormais exécuter des requêtes sur la `Custs` table.

Consultez aussi

- Contrôle d'accès aux métadonnées

Comparaison entre le contrôle d'accès basé sur les balises Lake Formation et le contrôle d'accès basé sur les attributs IAM

Le contrôle d’accès par attributs (ABAC) est une stratégie d’autorisation qui définit des autorisations en fonction des attributs. Dans AWS, ces attributs sont appelés balises. Vous pouvez associer des balises aux ressources IAM, notamment aux entités IAM (utilisateurs ou rôles) et aux AWS ressources. Vous pouvez créer une seule politique ABAC ou un petit nombre de politiques pour vos principaux IAM. Ces politiques ABAC sont conçues pour autoriser des opérations lorsque la balise du principal correspond à celle de la ressource. L'ABAC est utile dans les environnements qui connaissent une croissance rapide et pour les cas où la gestion des politiques devient fastidieuse.

Les équipes de sécurité et de gouvernance du cloud utilisent IAM pour définir des politiques d'accès et des autorisations de sécurité pour toutes les ressources, y compris les compartiments Amazon S3, les instances Amazon EC2 et toutes les ressources auxquelles vous pouvez faire référence avec un ARN. Les politiques IAM définissent des autorisations générales (grossières) sur les ressources de votre lac de données, par exemple, pour autoriser ou refuser l'accès au niveau du compartiment ou du préfixe Amazon S3 ou au niveau de la base de données. Pour plus d'informations sur IAM ABAC, voir À quoi sert ABAC ? AWS dans le guide de l'utilisateur IAM.

Par exemple, vous pouvez créer trois rôles avec la clé de balise `project-access`. Définissez la valeur de la balise du premier rôle sur `Dev`, celle du deuxième sur `Marketing`, et celle du troisième sur `Support`. Attribuez des balises avec la valeur appropriée aux ressources. Vous pouvez alors utiliser une seule politique qui autorise l'accès lorsque le rôle et la ressource sont balisés avec la même valeur pour `project-access`.

Les équipes de gouvernance des données utilisent Lake Formation pour définir des autorisations précises pour des ressources de lacs de données spécifiques. Les balises LF sont attribuées aux ressources du catalogue de données (bases de données, tables et colonnes) et sont accordées aux principaux. Un principal dont les balises LF correspondent aux balises LF d'une ressource peut accéder à cette ressource. Les autorisations de Lake Formation sont secondaires aux autorisations IAM. Par exemple, si les autorisations IAM n'autorisent pas un utilisateur à accéder à un lac de données, Lake Formation n'accorde l'accès à aucune ressource de ce lac de données à cet utilisateur, même si le principal et la ressource ont des balises LF identiques.

Le contrôle d'accès basé sur des balises Lake Formation (LF-TBAC) fonctionne avec IAM ABAC pour fournir des niveaux d'autorisations supplémentaires pour vos données et ressources de Lake Formation.

- Les autorisations TBAC de Lake Formation évoluent avec l'innovation. L'administrateur n'a plus besoin de mettre à jour les politiques existantes pour autoriser l'accès aux nouvelles ressources. Supposons, par exemple, que vous utilisiez une stratégie IAM ABAC avec la `project-access` balise pour fournir un accès à des bases de données spécifiques au sein de Lake Formation. À l'aide du LF-TBAC, le tag LF `Project=SuperApp` est attribué à des tables ou à des colonnes spécifiques, et le même tag LF est accordé à un développeur pour ce projet. Grâce à IAM, le développeur peut accéder à la base de données, et les autorisations LF-TBAC lui permettent d'accéder davantage à des tables ou à des colonnes spécifiques au sein de tables. Si une nouvelle table est ajoutée au projet, l'administrateur de Lake Formation n'a qu'à attribuer la balise à la nouvelle table pour que le développeur puisse y accéder.
- Lake Formation TBAC nécessite moins de politiques IAM. Comme vous utilisez les politiques IAM pour accorder un accès de haut niveau aux ressources de Lake Formation et le TBAC de Lake Formation pour gérer un accès aux données plus précis, vous créez moins de politiques IAM.
- Grâce au Lake Formation TBAC, les équipes peuvent changer et se développer rapidement. En effet, les autorisations d'accès aux nouvelles ressources sont automatiquement accordées en fonction des attributs. Par exemple, si un nouveau développeur rejoint le projet, il est facile de lui accorder l'accès en associant le rôle IAM à l'utilisateur, puis en lui attribuant les balises LF requises. Il n'est pas nécessaire de modifier la politique IAM pour prendre en charge un nouveau projet ou pour créer de nouveaux LF-Tags.
- Des autorisations plus précises sont possibles grâce au Lake Formation TBAC. Les politiques IAM accordent l'accès aux ressources de haut niveau, telles que les bases de données ou les tables du catalogue de données. À l'aide de Lake Formation TBAC, vous pouvez autoriser l'accès à des tables ou à des colonnes spécifiques contenant des valeurs de données spécifiques.

> **Note**
>
> Les balises IAM sont différentes des balises LF. Ces étiquettes ne sont pas interchangeables. Les balises LF sont utilisées pour accorder des autorisations à Lake Formation et les balises IAM sont utilisées pour définir les politiques IAM.

Octroi, révocation et listage des autorisations relatives aux valeurs des balises LF

Vous pouvez accorder les `Alter` autorisations sur `Drop` les balises LF aux principaux afin de gérer les expressions de valeur des balises LF. Vous pouvez également accorder `DescribeAssociate`, et des `Grant with LF-Tag expressions` autorisations sur les balises LF aux principaux pour

qu'ils puissent visualiser les balises LF et les attribuer aux ressources du catalogue de données (bases de données, tables et colonnes). Lorsque des balises LF sont attribuées aux ressources du catalogue de données, vous pouvez utiliser la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC) pour sécuriser ces ressources. Pour plus d'informations, consultez Contrôle d'accès basé sur des balises Lake Formation.

Vous pouvez accorder ces autorisations avec l'option d'octroi afin que les autres principaux puissent les accorder. Les `Associate` autorisations `Grant with LF-Tag expressionsDescribe`, et sont expliquées dansAjouter des créateurs de LF-Tag.

Vous pouvez accorder les `Associate` autorisations `Describe` et sur un LF-Tag à un compte externe AWS . L'administrateur du lac de données de ce compte peut ensuite accorder ces autorisations aux autres principaux du compte. Les principaux auxquels l'administrateur du lac de données du compte externe accorde l'`Associate`autorisation peuvent ensuite attribuer des balises LF aux ressources du catalogue de données que vous avez partagées avec leur compte.

Lorsque vous accordez à un compte externe, vous devez inclure l'option de subvention.

Vous pouvez accorder des autorisations sur les balises LF à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface ()AWS CLI.

Rubriques

- Répertorier les autorisations LF-Tag à l'aide de la console
- Octroi d'autorisations LF-Tag à l'aide de la console
- Octroyer, révoquer et répertorier les autorisations LF-Tag à l'aide du AWS CLI

Pour plus d'informations, consultez Gestion des balises LF pour le contrôle d'accès aux métadonnées et Contrôle d'accès basé sur des balises Lake Formation.

Répertorier les autorisations LF-Tag à l'aide de la console

Vous pouvez utiliser la console Lake Formation pour consulter les autorisations accordées sur les LF-Tags. Vous devez être un créateur de balises LF, un administrateur de lac de données ou avoir l'`Associate`autorisation `Describe` ou l'autorisation sur une balise LF pour le voir.

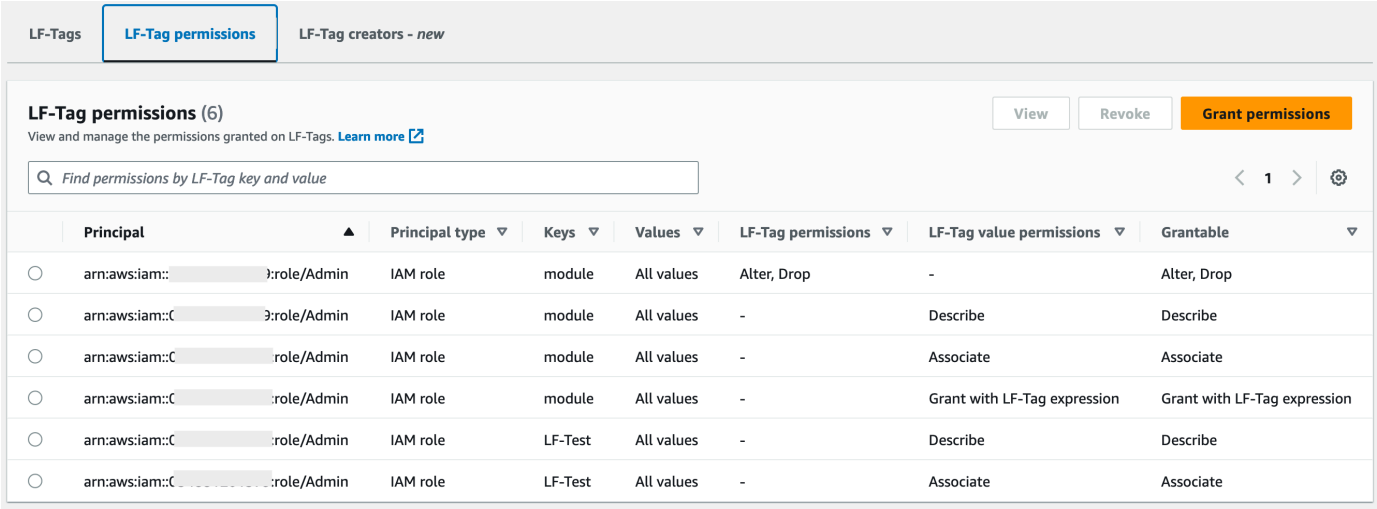
Pour répertorier les autorisations LF-Tag (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

Connectez-vous en tant que créateur de balises LF, administrateur du lac de données ou en tant qu'utilisateur à qui les `Drop` `Describe` autorisations relatives aux balises LF ont été accordées. `Alter` `Associate`

2. Dans le volet de navigation, sous Autorisations, choisissez LF-Tags et autorisations, puis choisissez la section Autorisations LF-Tag.

   La section des autorisations LF-Tag présente un tableau contenant le principal, les clés de balise, les valeurs et les autorisations.



Octroi d'autorisations LF-Tag à l'aide de la console

Les étapes suivantes expliquent comment accorder des autorisations sur les balises LF à l'aide de la page Accorder des autorisations sur les balises LF de la console Lake Formation. La page est divisée en sections suivantes :

- Types d'autorisation : type d'autorisation à accorder.
- Principaux : utilisateurs, rôles ou AWS comptes auxquels accorder des autorisations.
- LF-Tags — Les LF-Tags auxquels accorder des autorisations.
- Autorisations : autorisations à accorder.

Ouvrez la page Accorder les autorisations LF-Tag

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

Connectez-vous en tant que créateur du tag LF, administrateur du lac de données ou en tant qu'utilisateur. Les autorisations LF-Tag ou paire clé-valeur LF-Tag sur les balises LF ont été accordées avec cette option. `Grant`

2. Dans le volet de navigation, choisissez LF-Tags et permissions, choisissez la section LF-Tag permissions.
3. Choisissez Grant permissions (Accorder des autorisations).

Spécifiez le type d'autorisation

Dans la section Type d'autorisations, choisissez un type d'autorisation.

Autorisations LF-Tag

Choisissez les autorisations LF-Tag pour permettre aux principaux de mettre à jour les valeurs des LF-Tag ou de supprimer des LF-Tag.

Autorisations relatives aux paires clé-valeur LF-Tag

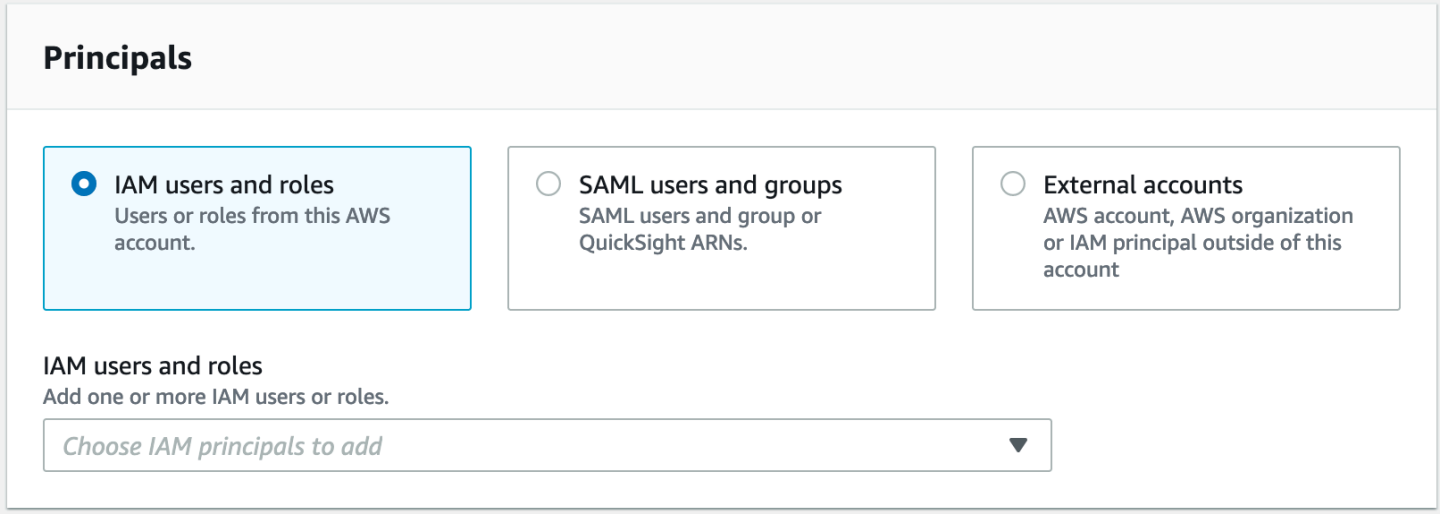
Choisissez les autorisations relatives à la paire clé-valeur LF-Tag pour permettre aux principaux d'attribuer des balises LF aux ressources du catalogue de données, d'afficher les balises LF et d'accorder aux principaux des autorisations basées sur les balises LF sur les ressources du catalogue de données.

Les options disponibles dans les sections suivantes dépendent du type d'autorisations.

Spécifiez les principes

> **Note**
>
> Vous ne pouvez pas accorder d'autorisations LF-Tag (`AlteretDrop`) à des comptes externes ou à des directeurs d'un autre compte.

Dans la section Principaux, choisissez un type de principal et spécifiez les principaux auxquels accorder des autorisations.



Utilisateurs et rôles IAM

Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

Utilisateurs et groupes SAML

Pour les QuickSight utilisateurs et les groupes SAML et Amazon, entrez un ou plusieurs Amazon Resource Names (ARN) pour les utilisateurs ou les groupes fédérés via SAML, ou des ARN pour les utilisateurs ou les groupes Amazon. QuickSight Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire les ARN, consultezLake Formation accorde et AWS CLI révoque des commandes.

> **Note**
>
> L'intégration de Lake Formation à Amazon QuickSight n'est prise en charge que pour Amazon QuickSight Enterprise Edition.

Comptes externes

Pour le AWS compte, entrez un ou plusieurs identifiants de AWS compte valides. Appuyez sur Entrée après chaque identifiant.

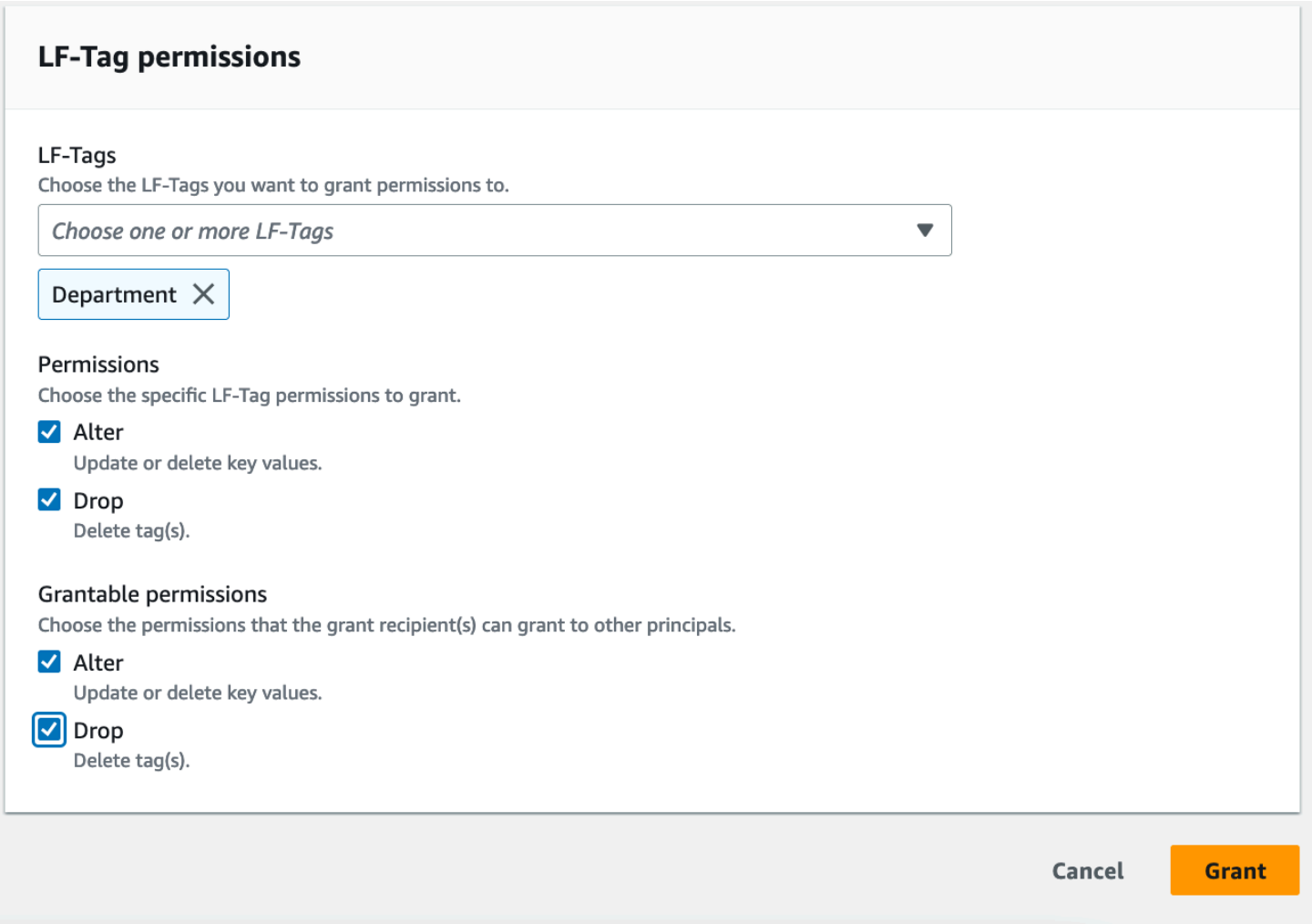
Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

Pour le principal IAM, entrez l'ARN de l'utilisateur ou du rôle IAM.

Spécifiez les balises LF

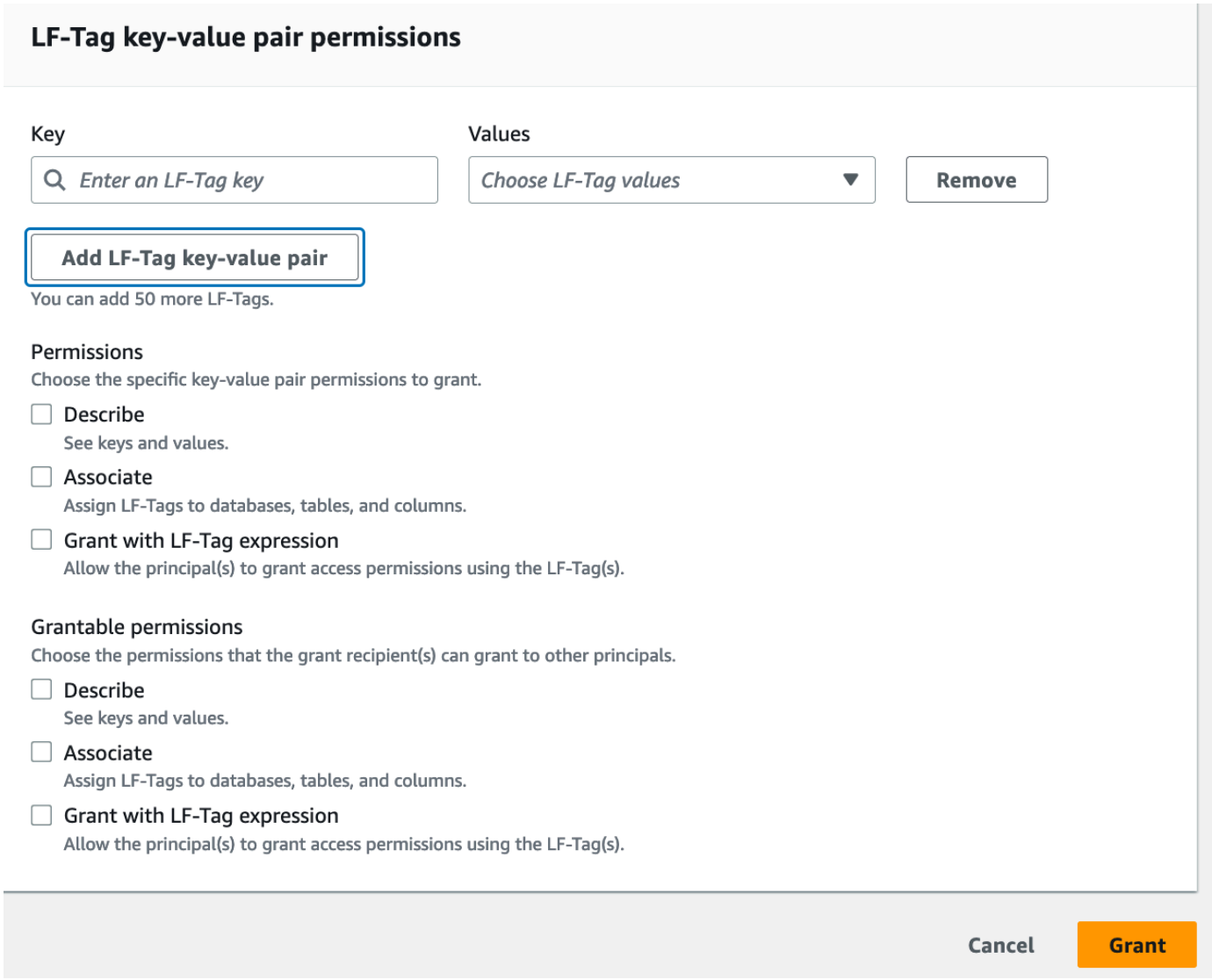
Pour accorder des autorisations sur les balises LF, dans la section Autorisations des balises LF, spécifiez les balises LF auxquelles accorder des autorisations.



- Choisissez un ou plusieurs LF-Tag à l'aide de la liste déroulante.

Spécifiez les paires clé-valeur LF-Tag

1. Pour accorder des autorisations sur les paires clé-valeur LF-Tag, (vous devez d'abord choisir les autorisations de paire clé-valeur LF-Tag comme type d'autorisation), choisissez Ajouter une paire clé-valeur LF-Tag pour afficher la première ligne de champs permettant de spécifier la clé et les valeurs LF-Tag.



2. Positionnez le curseur dans le champ Clé, commencez éventuellement à taper pour affiner la liste de sélection, puis sélectionnez une touche LF-Tag.
3. Dans la liste des valeurs, sélectionnez une ou plusieurs valeurs, puis appuyez sur Tab ou cliquez ou appuyez en dehors du champ pour enregistrer les valeurs sélectionnées.

> **Note**
>
> Si l'une des lignes de la liste des valeurs est sélectionnée, appuyez sur Entrée pour sélectionner ou désactiver la case à cocher.

Les valeurs sélectionnées apparaissent sous forme de vignettes sous la liste des valeurs. Choisissez le ✖ pour supprimer une valeur. Choisissez Supprimer pour supprimer le LF-tag dans son intégralité.

4. Pour ajouter un autre tag LF, choisissez à nouveau Ajouter un tag LF et répétez les deux étapes précédentes.

Spécifiez les autorisations

Cette section affiche les autorisations LF-Tag ou les autorisations à valeur LF-Tag en fonction du type d'autorisation que vous avez choisi à l'étape précédente.

Selon le type d'autorisation que vous avez choisi d'accorder, sélectionnez les autorisations LF-Tag ou les autorisations par paire clé-valeur LF-Tag, et les autorisations pouvant être accordées.

1. Sous Autorisations LF-Tag, sélectionnez les autorisations à accorder.

   Si vous accordez Drop and Alter, vous accordez implicitement Describe.

   Vous devez accorder les autorisations Alter and Drop sur toutes les valeurs des balises.
2. Sous Autorisations à valeur clé-valeur LT-Tag, sélectionnez les autorisations à accorder.

   L'associé octroie implicitement l'autorisation à Describe. Choisissez Grant avec expression LF-Tag pour permettre au bénéficiaire de la subvention d'accorder ou de révoquer les autorisations d'accès aux ressources du catalogue de données à l'aide de la méthode LF-TBAC.
3. (Facultatif) Sous Autorisations pouvant être accordées, sélectionnez les autorisations que le bénéficiaire de la subvention peut accorder aux autres principaux de son compte. AWS
4. Choisissez Grant (Accorder).

Octroyer, révoquer et répertorier les autorisations LF-Tag à l'aide du AWS CLI

Vous pouvez accorder, révoquer et répertorier les autorisations sur les balises LF en utilisant le AWS Command Line Interface ().AWS CLI

Pour répertorier les autorisations LF-Tag ()AWS CLI

- Entrez une `list-permissions` commande. Vous devez être le créateur du tag LF, un administrateur du lac de données ou avoir l'`Grant with LF-Tag permissions`autorisation`Drop`,`Alter`, `DescribeAssociate`, sur un tag LF pour le voir.

  La commande suivante demande toutes les balises LF pour lesquelles vous avez des autorisations.

```
aws lakeformation list-permissions --resource-type LF_TAG
```

  Voici un exemple de sortie destiné à un administrateur de lac de données, qui voit toutes les balises LF accordées à tous les principaux. Les utilisateurs non administrateurs ne voient que les balises LF qui leur sont accordées. Les autorisations LF-Tag accordées à partir d'un compte externe apparaissent sur une page de résultats distincte. Pour les voir, répétez la commande et fournissez à l'`--next-token`argument le jeton renvoyé lors de l'exécution de la commande précédente.

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_admin"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "environment",
                    "TagValues": [
                        "*"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": [
                "ASSOCIATE"
            ]
        },
```

```
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "DESCRIBE"
            ],
            "PermissionsWithGrantOption": []
        },
...
    ],
    "NextToken": "eyJzaG91bGRRdWVy...Wlzc2lvbnMiOnRydWV9"
}
```

Vous pouvez répertorier toutes les subventions pour une clé LF-Tag spécifique. La commande suivante renvoie toutes les autorisations accordées sur le LF-Tag`module`.

```
aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Vous pouvez également répertorier les valeurs de balise LF accordées à un principal spécifique pour une balise LF spécifique. Lorsque vous fournissez l'`--principal`argument, vous devez le `--resource` fournir. Par conséquent, la commande ne peut effectivement demander que les valeurs accordées à un principal spécifique pour une clé LF-Tag spécifique. La commande suivante montre comment procéder pour la touche principale `datalake_user1` et la touche LF-Tag. `module`

```
aws lakeformation list-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
```

```
datalake_user1 --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Voici un exemple de sortie.

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": []
        }
    ]
}
```

Pour accorder des autorisations sur les balises LF ()AWS CLI

1. Utilisez une commande similaire à la suivante. Cet exemple accorde à l'utilisateur `datalake_user1` l'`Associate`autorisation d'utiliser le tag LF avec la clé. `module` Il autorise l'affichage et l'attribution de toutes les valeurs pour cette clé, comme indiqué par l'astérisque (*).

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

L'octroi de l'`Associate`autorisation octroie implicitement l'`Describe`autorisation.

L'exemple suivant accorde `Associate` au AWS compte externe 1234-5678-9012 sur le LF-Tag avec la clé, avec l'option d'octroi. `module` Il accorde des autorisations pour afficher et attribuer uniquement les valeurs `sales` et`orders`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=123456789012 --permissions "ASSOCIATE"
 --permissions-with-grant-option "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
```

2. L'octroi de l'`GrantWithLFTagExpression`autorisation octroie implicitement l'`Describe`autorisation.

   L'exemple suivant accorde `GrantWithLFTagExpression` à un utilisateur sur le tag LF avec la clé`module`, avec l'option d'octroi. Il accorde des autorisations pour consulter et octroyer des autorisations sur les ressources du catalogue de données en utilisant uniquement les valeurs `sales` et`orders`.

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "GrantWithLFTagExpression"
    --permissions-with-grant-option "GrantWithLFTagExpression" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
   ```

3. L'exemple suivant accorde des `Drop` autorisations à un utilisateur sur le tag LF avec la clé`module`, avec l'option d'octroi. Il accorde l'autorisation de supprimer le tag LF. Pour supprimer une balise LF, vous devez disposer d'autorisations sur toutes les valeurs de cette clé.

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "DROP"
    --permissions-with-grant-option "DROP" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

4. L'exemple suivant accorde des `Alter` autorisations à l'utilisateur sur le tag LF avec la clé`module`, avec l'option d'octroi. Il accorde l'autorisation de supprimer le tag LF. Pour mettre à jour une balise LF, vous devez disposer d'autorisations sur toutes les valeurs de cette clé.

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "ALTER"
   ```

```
 --permissions-with-grant-option "ALTER" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Pour révoquer les autorisations sur les balises LF ()AWS CLI

- Utilisez une commande similaire à la suivante. Cet exemple révoque l'`Associate`autorisation sur le tag LF avec la clé `module` de l'utilisateur. `datalake_user1`

```
aws lakeformation revoke-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Octroi d'autorisations de data lake à l'aide de la méthode LF-TBAC

Vous pouvez accorder les autorisations `DESCRIBE` et `ASSOCIATE` Lake Formation sur les balises LF aux principaux afin qu'ils puissent visualiser les balises LF et les attribuer aux ressources du catalogue de données (bases de données, tables, vues et colonnes). Lorsque des balises LF sont attribuées aux ressources du catalogue de données, vous pouvez utiliser la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC) pour sécuriser ces ressources. Pour plus d'informations, consultez Contrôle d'accès basé sur des balises Lake Formation.

Dans un premier temps, seul l'administrateur du lac de données peut accorder ces autorisations. Si l'administrateur du lac de données accorde ces autorisations avec l'option d'octroi, d'autres principaux peuvent les accorder. Les `ASSOCIATE` autorisations `DESCRIBE` et sont expliquées dansMeilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation.

Vous pouvez accorder les `ASSOCIATE` autorisations `DESCRIBE` et sur un LF-Tag à un compte externe AWS . L'administrateur du lac de données de ce compte peut ensuite accorder ces autorisations aux autres principaux du compte. Les principaux auxquels l'administrateur du lac de données du compte externe accorde l'`ASSOCIATE`autorisation peuvent ensuite attribuer des balises LF aux ressources du catalogue de données que vous avez partagées avec leur compte.

Lorsque vous accordez à un compte externe, vous devez inclure l'option de subvention.

Vous pouvez accorder des autorisations sur les balises LF à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface ()AWS CLI.

Rubriques

- Octroi d'autorisations au catalogue de données

Consultez aussi

- Octroi, révocation et listage des autorisations relatives aux valeurs des balises LF
- Gestion des balises LF pour le contrôle d'accès aux métadonnées
- Contrôle d'accès basé sur des balises Lake Formation

Octroi d'autorisations au catalogue de données

Utilisez la console Lake Formation ou accordez AWS CLI à Lake Formation des autorisations sur les bases de données, les tables, les vues et les colonnes du catalogue de données à l'aide de la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC).

Console

Les étapes suivantes expliquent comment accorder des autorisations à l'aide de la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) et de la page Accorder des autorisations aux lacs de données sur la console Lake Formation. La page est divisée selon les sections suivantes :

- Principaux — Les utilisateurs, les rôles et les autorisations Comptes AWS auxquelles accorder des autorisations.
- Balises LF ou ressources du catalogue : bases de données, tables ou liens de ressources sur lesquels accorder des autorisations.
- Autorisations — Les autorisations à accorder dans le cadre de la Lake Formation.

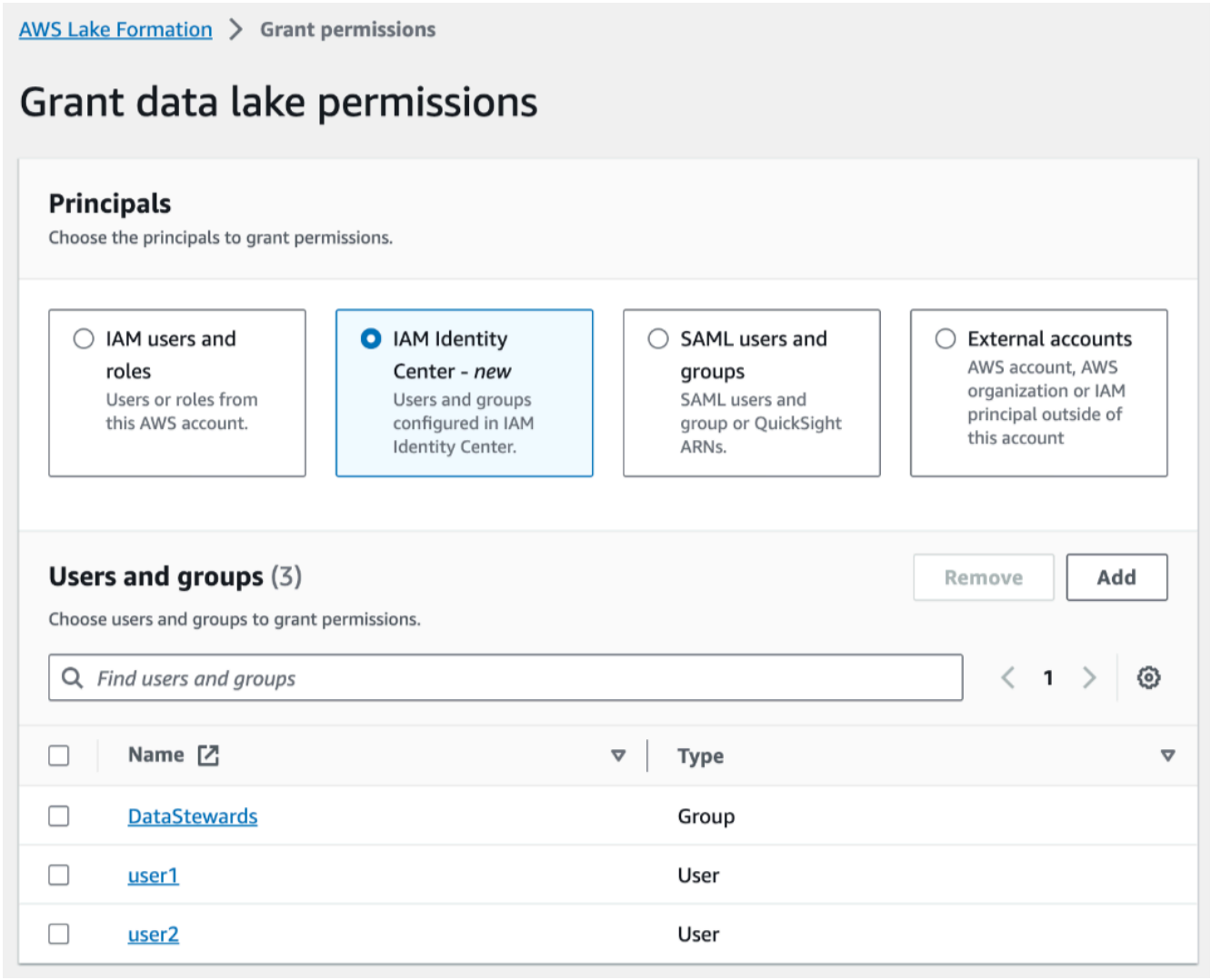
1. Ouvrez la page des autorisations du lac de données Grant.

   Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/ et connectez-vous en tant qu'administrateur de lac de données ou en tant qu'utilisateur ayant obtenu les autorisations de Lake Formation sur les ressources du catalogue de données via LF-TBAC avec l'option grant.

Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données. Choisissez ensuite Grant.

2. Spécifiez les principes.

   Dans la section Principaux, choisissez un type de principal, puis spécifiez les principaux auxquels accorder des autorisations.



   Utilisateurs et rôles IAM

   Choisissez un ou plusieurs utilisateurs ou rôles dans la liste des utilisateurs et des rôles IAM.

   IAM Identity Center

   Choisissez un ou plusieurs utilisateurs ou dans la liste Utilisateurs et groupes.

Utilisateurs et groupes SAML

Pour les QuickSight utilisateurs et les groupes SAML et Amazon, entrez un ou plusieurs Amazon Resource Names (ARN) pour les utilisateurs ou les groupes fédérés via SAML, ou des ARN pour les utilisateurs ou les groupes Amazon. QuickSight Appuyez sur Entrée après chaque ARN.

Pour plus d'informations sur la façon de construire les ARN, consultezLake Formation accorde et AWS CLI révoque des commandes.

> **Note**
>
> L'intégration de Lake Formation à Amazon QuickSight n'est prise en charge que pour Amazon QuickSight Enterprise Edition.

Comptes externes

Pour Comptes AWS AWS l'organisation ou le principal IAM, entrez un ou plusieurs Compte AWS identifiants, identifiants d'organisation, identifiants d'unité organisationnelle ou ARN valides pour l'utilisateur ou le rôle IAM. Appuyez sur Entrée après chaque identifiant.

Un identifiant d'organisation se compose de « o- » suivi de 10 à 32 lettres minuscules ou chiffres.

L'identifiant d'une unité organisationnelle commence par « ou- » suivi de 4 à 32 lettres minuscules ou chiffres (identifiant de la racine contenant l'unité organisationnelle). Cette chaîne est suivie d'un deuxième tiret « - » et de 8 à 32 lettres minuscules ou chiffres supplémentaires.

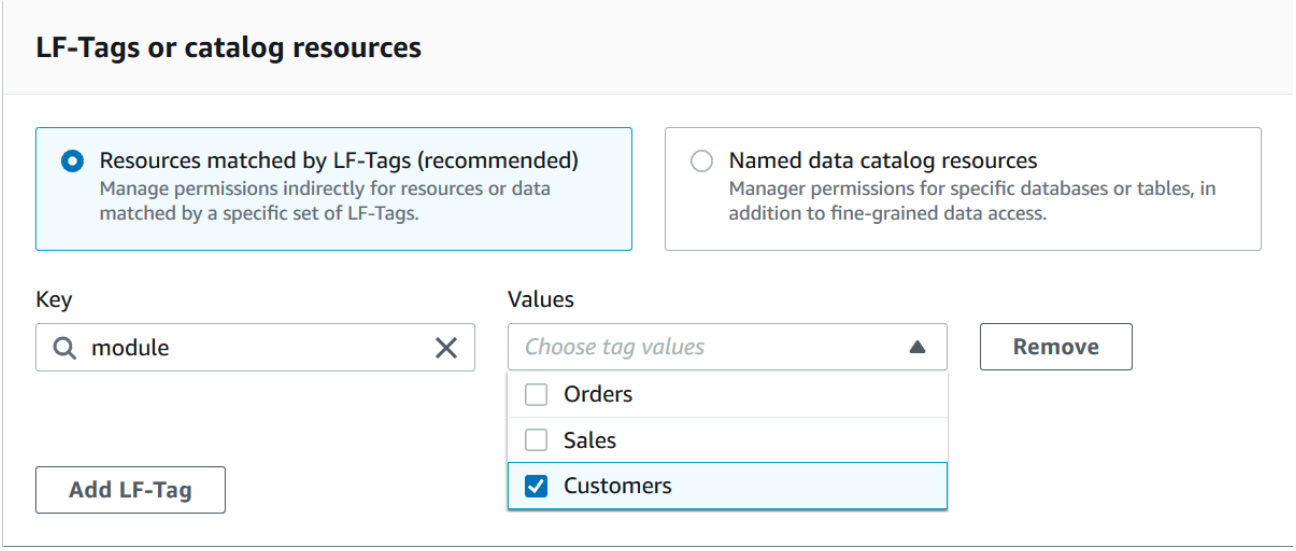
3. Spécifiez les balises LF.

Assurez-vous que l'option Resources matched by LF-Tags est sélectionnée. Choisissez Ajouter un tag LF.

1. Choisissez une clé et des valeurs LF-Tag.

Si vous choisissez plusieurs valeurs, vous créez une expression LF-Tag à l'aide d'un `OR` opérateur. Cela signifie que si l'une des valeurs de balise LF correspond à une balise LF

attribuée à une ressource de catalogue de données, des autorisations sur cette ressource vous sont accordées.



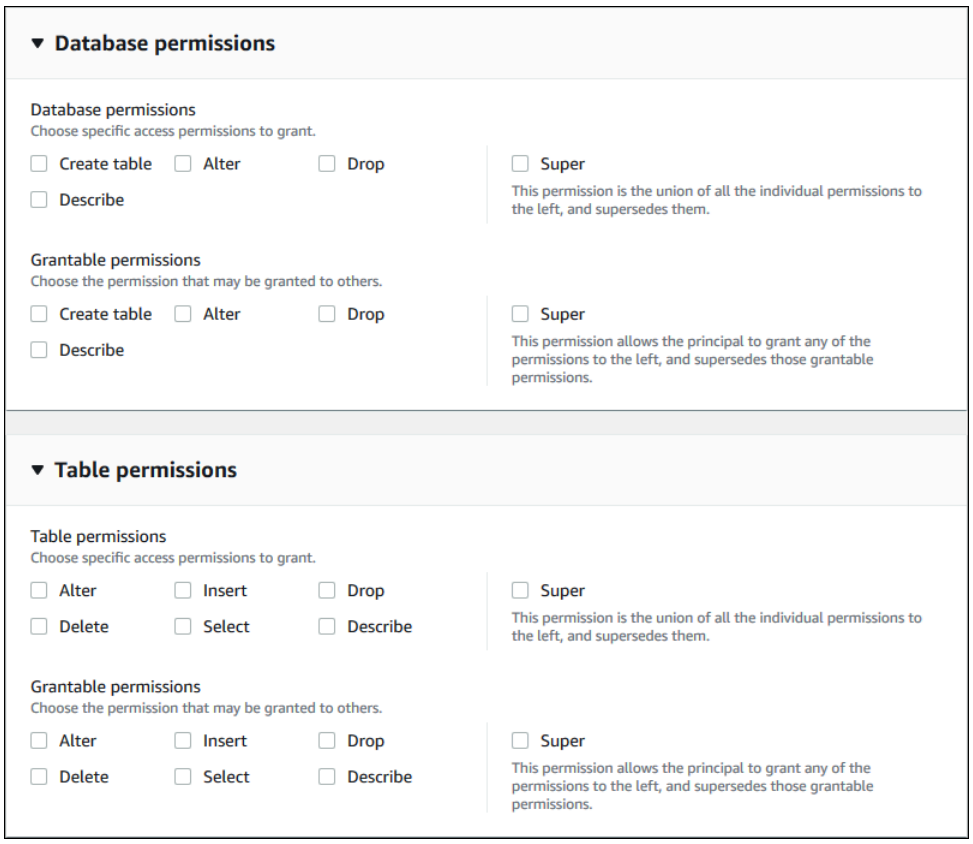
2. (Facultatif) Choisissez à nouveau Ajouter un tag LF pour en spécifier un autre.

   Si vous spécifiez plusieurs balises LF, vous créez une expression de balise LF à l'aide d'un opérateur. AND Le principal n'obtient des autorisations sur une ressource de catalogue de données que si une balise LF correspondante a été attribuée à la ressource pour chaque balise LF de l'expression LF-Tag.

4. Spécifiez les autorisations.

   Spécifiez les autorisations que vous souhaitez accorder au principal pour faire correspondre les ressources du catalogue de données. Les ressources correspondantes sont les ressources auxquelles des balises LF ont été attribuées qui correspondent à l'une des expressions de balise LF accordées au principal.

   Vous pouvez spécifier les autorisations à accorder sur les bases de données correspondantes, les tables correspondantes et les vues correspondantes.



Sous Autorisations de base de données, sélectionnez les autorisations de base de données à accorder au principal sur les bases de données correspondantes.

Sous Autorisations relatives aux tables, sélectionnez les autorisations de table ou de vue à accorder au principal sur les tables et les vues correspondantes.

Vous pouvez également choisir `Select`, `Describe`, et `Drop` les autorisations dans le tableau, les autorisations à appliquer aux vues.

5. Choisissez Grant (Accorder).

AWS CLI

Vous pouvez utiliser la méthode AWS Command Line Interface (AWS CLI) et la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) pour accorder à Lake Formation des autorisations sur les bases de données, les tables et les colonnes du catalogue de données.

Octroi d'autorisations de data lake à l'aide de la méthode AWS CLI et de la méthode LF-TBAC

- Utilisez la commande `grant-permissions`.

Example

L'exemple suivant accorde l'expression LF-tag « `module=*` » (toutes les valeurs de la clé `module` LF-Tag) à l'utilisateur. `datalake_user1` Cet utilisateur aura l'`CREATE_TABLE`autorisation d'accéder à toutes les bases de données correspondantes, c'est-à-dire aux bases de données auxquelles a été attribuée la balise LF avec la clé`module`, quelle que soit la valeur.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "CREATE_TABLE" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
[{"TagKey":"module","TagValues":["*"]}]}}'
```

Example

L'exemple suivant accorde l'expression LF-tag « `(level=director) AND (region=west OR region=south)` » à l'utilisateur. `datalake_user1` Cet utilisateur aura les `DROP` autorisations `SELECTALTER`, et avec l'option d'autorisation sur les tables correspondantes, c'est-à-dire les tables auxquelles les deux `level=director` et (`region=westouregion=south`) ont été assignés.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "SELECT" "ALTER" "DROP" --permissions-
with-grant-option "SELECT" "ALTER" "DROP" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"TABLE","Expression": [{"TagKey":
 "level","TagValues": ["director"]},{"TagKey": "region","TagValues": ["west",
 "south"]}]}}'
```

Example

L'exemple suivant accorde l'expression LF-tag « `module=orders` » au AWS compte 1234-5678-9012. L'administrateur du lac de données de ce compte peut ensuite accorder l'expression `module=orders` « » aux principaux de son compte. Ces principaux auront alors l'`CREATE_TABLE`autorisation de faire correspondre les bases de données détenues par le compte 1111-2222-3333 et partagées avec le compte 1234-5678-9012 en utilisant soit la méthode des ressources nommées, soit la méthode LF-TBAC.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=123456789012 --permissions "CREATE_TABLE" --
permissions-with-grant-option "CREATE_TABLE" --resource '{ "LFTagPolicy":
 {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
[{"TagKey":"module","TagValues":["orders"]}]}}'
```

Exemple de scénario d'autorisations

Le scénario suivant montre comment configurer des autorisations pour sécuriser l'accès aux données dans AWS Lake Formation.

Shirley est administratrice de données. Elle souhaite créer un lac de données pour son entreprise AnyCompany. Actuellement, toutes les données sont stockées dans Amazon S3. John est directeur marketing et a besoin d'un accès écrit aux informations d'achat des clients (contenues dans`s3://customerPurchases`). Diego, analyste marketing, rejoint John cet été. John doit être en mesure d'autoriser Diego à effectuer des requêtes sur les données sans impliquer Shirley.

Mateo, du service financier, a besoin d'accéder aux données comptables (par exemple,`s3://transactions`). Il souhaite interroger les données des transactions dans les tables d'une base de données (`Finance_DB`) utilisée par l'équipe financière. Son manager, Arnav, peut lui donner accès au`Finance_DB`. Bien qu'il ne devrait pas être en mesure de modifier les données comptables, il doit être capable de convertir les données dans un format (schéma) adapté aux prévisions. Ces données seront stockées dans un compartiment séparé (`s3://financeForecasts`) qu'il pourra modifier.

Pour résumer :

- Shirley est l'administratrice du lac de données.
- John a besoin `CREATE_DATABASE` d'une `CREATE_TABLE` autorisation pour créer de nouvelles bases de données et de nouvelles tables dans le catalogue de données.
- John a également besoin `SELECT` d'`INSERTDELETE`autorisations et d'autorisations sur les tables qu'il crée.
- Diego a besoin `SELECT` d'une autorisation sur la table pour exécuter des requêtes.

Les employés de AnyCompany exécutent les actions suivantes pour configurer les autorisations. Les opérations d'API présentées dans ce scénario présentent une syntaxe simplifiée pour plus de clarté.

1. Shirley enregistre le chemin Amazon S3 contenant les informations d'achat des clients auprès de Lake Formation.

```
RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
```

2. Shirley accorde à John l'accès au chemin Amazon S3 contenant les informations d'achat des clients.

```
GrantPermissions(John, S3Location("s3://customerPurchases"),
 [DATA_LOCATION_ACCESS]) )
```

3. Shirley autorise John à créer des bases de données.

```
GrantPermissions(John, catalog, [CREATE_DATABASE])
```

4. John crée la base de données`John_DB`. John a automatiquement `CREATE_TABLE` l'autorisation d'accéder à cette base de données parce qu'il l'a créée.

```
CreateDatabase(John_DB)
```

5. John crée le tableau `John_Table` pointant vers`s3://customerPurchases`. Comme il a créé la table, il dispose de toutes les autorisations sur celle-ci et peut accorder des autorisations sur celle-ci.

```
CreateTable(John_DB, John_Table)
```

6. John autorise son analyste, Diego, à accéder à la table`John_Table`.

```
 GrantPermissions(Diego, John_Table, [SELECT])
```

7. John autorise son analyste, Diego, à accéder au`s3://customerPurchases/London/`. Shirley étant déjà inscrite`s3://customerPurchases`, ses sous-dossiers sont enregistrés auprès de Lake Formation.

```
 GrantDataLakePrivileges( 123456789012/datalake, Diego, [DATA_LOCATION_ACCESS], [],
 S3Location("s3://customerPurchases/London/") )
```

8. John autorise son analyste, Diego, à créer des tables dans la base de données`John_DB`.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, John_DB, [CREATE_TABLE],
[] )
```

9. Diego crée une table dans `John_DB` at `s3://customerPurchases/London/` et obtient automatiquement`ALTER`,,`DROP`, `SELECT``INSERT`, et `DELETE` les autorisations.

```
CreateTable( 123456789012/datalake, John_DB, Diego_Table )
```

Filtrage des données et sécurité au niveau des cellules dans Lake Formation

Lorsque vous accordez des autorisations à Lake Formation sur une table du catalogue de données, vous pouvez inclure des spécifications de filtrage des données afin de restreindre l'accès à certaines données dans les résultats des requêtes et les moteurs intégrés à Lake Formation. Lake Formation utilise le filtrage des données pour garantir la sécurité au niveau des colonnes, au niveau des lignes et au niveau des cellules. Vous pouvez définir et appliquer des filtres de données sur des colonnes imbriquées si vos données sources contiennent des structures imbriquées.

Rubriques

- Vue d'ensemble du filtrage des données
- Filtres de données dans Lake Formation
- Support partiQL dans les expressions de filtre de ligne
- Remarques et restrictions relatives au filtrage au niveau des colonnes
- Autorisations requises pour interroger des tables avec filtrage au niveau des cellules
- Gestion des filtres de données

Vue d'ensemble du filtrage des données

Grâce aux fonctionnalités de filtrage des données de Lake Formation, vous pouvez mettre en œuvre les niveaux de sécurité des données suivants.

Sécurité au niveau des colonnes

L'octroi d'autorisations sur une table du catalogue de données avec sécurité au niveau des colonnes (filtrage des colonnes) permet aux utilisateurs de n'afficher que les colonnes spécifiques et les

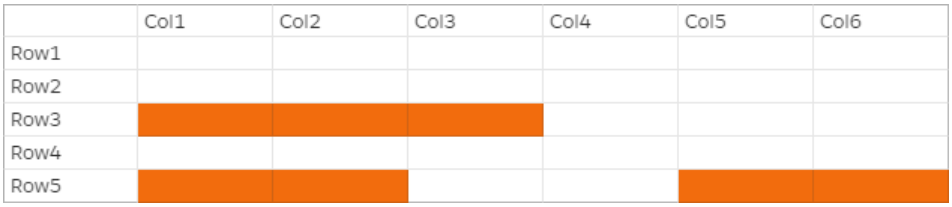
colonnes imbriquées auxquelles ils ont accès dans la table. Prenons l'`persons`exemple d'une table utilisée dans plusieurs applications pour une grande entreprise de communication multirégionale. L'octroi d'autorisations sur les tables du catalogue de données avec filtrage par colonne peut empêcher les utilisateurs qui ne travaillent pas dans le service des ressources humaines de voir les informations personnelles identifiables (PII) telles qu'un numéro de sécurité sociale ou une date de naissance. Vous pouvez également définir des politiques de sécurité et n'autoriser l'accès qu'à des sous-structures partielles de colonnes imbriquées.

Sécurité au niveau des lignes

L'octroi d'autorisations sur une table du catalogue de données avec sécurité au niveau des lignes (filtrage des lignes) permet aux utilisateurs de n'afficher que les lignes de données spécifiques auxquelles ils ont accès dans la table. Le filtrage est basé sur les valeurs d'une ou de plusieurs colonnes. Vous pouvez inclure des structures de colonnes imbriquées lorsque vous définissez des expressions de filtre de ligne. Par exemple, si les différents bureaux régionaux de l'entreprise de communication disposent de leur propre service des ressources humaines, vous pouvez limiter les dossiers personnels que les employés des ressources humaines peuvent consulter uniquement aux dossiers des employés de leur région.

Sécurité au niveau des cellules

La sécurité au niveau des cellules combine le filtrage des lignes et le filtrage des colonnes pour créer un modèle d'autorisations extrêmement flexible. Si vous visualisez les lignes et les colonnes d'un tableau sous forme de grille, en utilisant la sécurité au niveau des cellules, vous pouvez restreindre l'accès à des éléments individuels (cellules) de la grille n'importe où dans les deux dimensions. En d'autres termes, vous pouvez restreindre l'accès aux différentes colonnes en fonction de la ligne. Ceci est illustré par le schéma suivant, dans lequel les colonnes restreintes sont ombrées.



En reprenant l'exemple de la table des personnes, vous pouvez créer un filtre de données au niveau de la cellule qui restreint l'accès à la colonne des adresses postales si la colonne du pays est définie sur « Royaume-Uni », mais autorise l'accès à la colonne des adresses postales si la colonne du pays est définie sur « États-Unis ».

Les filtres s'appliquent uniquement aux opérations de lecture. Par conséquent, vous ne pouvez accorder que l'autorisation à la `SELECT` Lake Formation avec des filtres.

Sécurité au niveau des cellules sur les colonnes imbriquées

Lake Formation vous permet de définir et d'appliquer des filtres de données avec une sécurité au niveau des cellules sur des colonnes imbriquées. Cependant, les moteurs d'analyse intégrés tels qu'Amazon Athena, Amazon EMR et Amazon Redshift Spectrum permettent d'exécuter des requêtes sur des tables imbriquées gérées par Lake Formation avec une sécurité au niveau des lignes et des colonnes.

Pour connaître les limitations, veuillez consulter Limites du filtrage des données.

Filtres de données dans Lake Formation

Vous pouvez implémenter la sécurité au niveau des colonnes, des lignes et des cellules en créant des filtres de données. Vous sélectionnez un filtre de données lorsque vous accordez l'autorisation `SELECT` Lake Formation sur les tables. Si votre table contient des structures de colonnes imbriquées, vous pouvez définir un filtre de données en incluant ou en excluant les colonnes enfants et définir des expressions de filtre au niveau des lignes sur les attributs imbriqués.

Chaque filtre de données appartient à une table spécifique de votre catalogue de données. Un filtre de données inclut les informations suivantes :

- Nom du filtre
- Les identifiants de catalogue de la table associée au filtre
- Nom de la table
- Nom de la base de données contenant la table
- Spécification des colonnes : liste de colonnes et de colonnes imbriquées (avec `struct` types de données) à inclure ou à exclure dans les résultats de la requête.
- Expression de filtre de ligne : expression qui spécifie les lignes à inclure dans les résultats de la requête. Sous réserve de certaines restrictions, l'expression possède la syntaxe d'une `WHERE` clause dans le langage partiQL. Pour spécifier toutes les lignes, choisissez Accès à toutes les lignes sous Accès au niveau des lignes dans la console ou Utiliser `AllRowsWildcard` dans les appels d'API.

  Pour plus d'informations sur les éléments pris en charge dans les expressions de filtre de ligne, consultezSupport partiQL dans les expressions de filtre de ligne.

Le niveau de filtrage obtenu dépend de la façon dont vous renseignez le filtre de données.

- Lorsque vous spécifiez le caractère générique « all columns » (« toutes les colonnes ») et que vous fournissez une expression de filtre de ligne, vous établissez uniquement une sécurité au niveau de la ligne (filtrage des lignes).
- Lorsque vous incluez ou excluez des colonnes spécifiques et des colonnes imbriquées, et que vous spécifiez « toutes les lignes » à l'aide du caractère générique pour toutes les lignes, vous établissez une sécurité au niveau des colonnes (filtrage des colonnes) uniquement.
- Lorsque vous incluez ou excluez des colonnes spécifiques et que vous fournissez également une expression de filtre de ligne, vous établissez une sécurité au niveau de la cellule (filtrage des cellules).

La capture d'écran suivante de la console Lake Formation montre un filtre de données qui effectue un filtrage au niveau des cellules. Pour les requêtes portant sur la `orders` table, cela restreint l'accès à la `customer_name` colonne et les résultats de la requête ne renvoient que les lignes où la `product_type` colonne contient « pharma ».

Create data filter

Data filter name
Enter a name that describes this data access filter.

restrict-pharma

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.

Target database
Select the database that contains the target table.

Choose databases Load more

sales
054881201579

Target table
Select the table for which the data filter will be created.

Choose tables Load more

orders
054881201579

Column-level access
Choose whether this filter should have column-level restrictions.

- Access to all columns
  Filter won't have any column restrictions.
- Include columns
  Filter will only allow access to specific columns.
- Exclude columns
  Filter will allow access to all but specific columns.

Select columns

Choose one or more columns

customer_name
string

Notez l'utilisation de guillemets simples pour entourer la chaîne littérale,. `'pharma'`

Vous pouvez utiliser la console Lake Formation pour créer ce filtre de données, ou vous pouvez fournir l'objet de requête suivant à l'opération `CreateDataCellsFilter` d'API.

```
{
     "Name": "restrict-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type='pharma'"},
     "ColumnWildcard": {
         "ExcludedColumnNames": ["customer_name"]
     }
}
```

Vous pouvez créer autant de filtres de données que nécessaire pour une table. Pour ce faire, vous avez besoin d'une `SELECT` autorisation avec l'option d'octroi sur une table. Par défaut, les administrateurs de Data Lake sont autorisés à créer des filtres de données sur toutes les tables de ce compte. Vous n'utilisez généralement qu'un sous-ensemble des filtres de données possibles lorsque vous accordez des autorisations sur la table à un principal. Par exemple, vous pouvez créer un deuxième filtre de données pour la `orders` table qui est un filtre de row-security-only données. En vous référant à la capture d'écran précédente, vous pouvez choisir l'option Accès à toutes les colonnes et inclure une expression de filtre de ligne de`product_type<>pharma`. Le nom de ce filtre de données peut êtreno-`pharma`. Il restreint l'accès à toutes les lignes dont la `product_type` colonne est définie sur « pharma ».

L'objet de demande pour l'opération `CreateDataCellsFilter` d'API pour ce filtre de données est le suivant.

```
{
     "Name": "no-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type<>'pharma'"},
     "ColumnNames": ["customer_id", "customer_name", "order_num"
          "product_id", "purchase_date", "product_type",
          "product_manufacturer", "quantity", "price"]
}
```

Vous pouvez ensuite accorder une autorisation `SELECT` sur le `orders` tableau contenant le filtre de `restrict-pharma` données à un utilisateur administratif, et `SELECT` sur le `orders` tableau contenant le filtre de `no-pharma` données à des utilisateurs non administratifs. Pour les utilisateurs du secteur de la santé, vous accorderiez `SELECT` sur le `orders` tableau un accès complet à toutes les lignes et colonnes (aucun filtre de données), ou peut-être un autre filtre de données restreignant l'accès aux informations tarifaires.

Vous pouvez inclure ou exclure des colonnes imbriquées lorsque vous spécifiez la sécurité au niveau des colonnes et au niveau des lignes dans un filtre de données. Dans l'exemple suivant, l'accès au `product.offer` champ est spécifié à l'aide de noms de colonnes qualifiés (entre guillemets). Ceci est important pour les champs imbriqués afin d'éviter que des erreurs ne se produisent lorsque les noms de colonnes contiennent des caractères spéciaux et pour maintenir la rétrocompatibilité avec les définitions de sécurité de niveau supérieur des colonnes.

```
{
     "Name": "example_dcf",
     "DatabaseName": "example_db",
     "TableName": "example_table",
     "TableCatalogId": "111122223333",
     "RowFilter": { "FilterExpression": "customer.customerName <> 'John'" },
     "ColumnNames": ["customer", "\"product\".\"offer\""]
}
```

Consultez aussi

- Gestion des filtres de données

Support partiQL dans les expressions de filtre de ligne

Vous pouvez créer des expressions de filtre de ligne à l'aide d'un sous-ensemble de types de données, d'opérateurs et d'agrégations PartiQL. Lake Formation n'autorise aucune fonction PartiQL standard ou définie par l'utilisateur dans l'expression du filtre. Vous pouvez utiliser des opérateurs de comparaison pour comparer des colonnes à des constantes (par exemple,`views >= 10000`), mais vous ne pouvez pas comparer des colonnes à d'autres colonnes.

Une expression de filtre Row peut être une expression simple ou une expression composite. La longueur totale de l'expression doit être inférieure à 2 048 caractères.

Expression simple

Une expression simple aura le format suivant : `<column name > <comparison operator ><value >`

- Nom de colonne

  Il peut s'agir d'une colonne de données de niveau supérieur, d'une colonne de partition ou d'une colonne imbriquée présente dans le schéma de table et doit appartenir à la Types de données pris en charge liste ci-dessous.
- Opérateur de comparaison

  Les opérateurs pris en charge sont les suivants : `=, >, <, >=, <=, <>,!=, BETWEEN, IN, LIKE, NOT, IS [NOT] NULL`
- Toutes les comparaisons de chaînes et les correspondances de `LIKE` modèles distinguent les majuscules et minuscules. Vous ne pouvez pas utiliser l'opérateur IS [NOT] NULL sur les colonnes de partition.
- Valeur de colonne

  La valeur de colonne doit correspondre au type de données du nom de colonne.

Expression composite

Une expression composite sera au format :`( <simple expression >) <AND/OR >(<simple expression >)`. Les expressions composites peuvent être combinées ultérieurement à l'aide d'opérateurs logiques`AND/OR`.

Types de données pris en charge

Les filtres de ligne qui font référence à une AWS Glue Data Catalog table contenant un type de données non pris en charge provoqueront une erreur. Les types de données pris en charge pour les colonnes et les constantes de table sont les suivants, qui sont mappés aux types de Amazon Redshift données :

- `STRING, CHAR, VARCHAR`
- `INT, LONG, BIGINT, FLOAT, DECIMAL, DOUBLE`
- `BOOLEAN`
- `STRUCT`

Pour plus d'informations sur les types de données dans Amazon Redshift, consultez la section Types de données dans le guide du développeur de base de données Amazon Redshift.

Expressions de filtre de ligne

Example

Voici des exemples d'expressions de filtre de ligne valides pour une table comportant des colonnes : `country (String), id (Long), year (partition column of type Integer), month (partition column of type Integer)`

- `year > 2010 and country != 'US'`
- `(year > 2010 and country = 'US') or (month < 8 and id > 23)`
- `(country between 'Z' and 'U') and (year = 2018)`
- `(country like '%ited%') and (year > 2000)`

Example

Voici des exemples valides d'expressions de filtre de ligne pour une table comportant des colonnes imbriquées : `year > 2010 and customer.customerId <> 1`

Les champs imbriqués sous les colonnes de partition ne doivent pas être référencés lors de la définition d'expressions imbriquées au niveau des lignes.

Les constantes de chaîne doivent être placées entre guillemets simples.

Mots-clés réservés

Si votre expression de filtre de ligne contient des mots clés partiQL, vous recevrez une erreur d'analyse car les noms des colonnes peuvent entrer en conflit avec les mots clés. Dans ce cas, évitez les noms de colonnes en utilisant des guillemets doubles. Voici quelques exemples de mots clés réservés : « premier », « dernier », « asc », « manquant ». Consultez la spécification partiQL pour obtenir la liste des mots clés réservés.

Référence PartiQL

Pour plus d'informations sur PartiQL, consultez. https://partiql.org/

Remarques et restrictions relatives au filtrage au niveau des colonnes

Il existe trois méthodes pour définir le filtrage des colonnes :

- En utilisant des filtres de données, comme décrit précédemment.
- En utilisant un filtrage par colonne simple ou un filtrage par colonnes imbriquées.
- En utilisant des TAGs.

Le filtrage simple des colonnes indique simplement une liste de colonnes à inclure ou à exclure. La console Lake Formation, l'API et l'API prennent en AWS CLI charge un filtrage simple par colonne. Pour obtenir un exemple, consultez Grant with Simple Column Filtering.

Les remarques et restrictions suivantes s'appliquent au filtrage des colonnes :

- AWS Glue Les tâches ETL prennent en charge le filtrage des colonnes uniquement à l'aide de filtres de données (sécurité au niveau des cellules). La tâche échoue si un simple filtrage par colonne est appliqué à une table à laquelle la tâche fait référence. Si vous souhaitez uniquement filtrer les colonnes, accordez l'accès aux tables à l'aide de filtres de données et entrez `true` l'expression du filtre de ligne dans la console ou utilisez-la `AllRowsWildcard` dans vos appels d'API.
- Pour octroyer `SELECT` avec l'option d'autorisation et le filtrage des colonnes, vous devez utiliser une liste d'inclusion et non une liste d'exclusion. Sans l'option d'autorisation, vous pouvez utiliser des listes d'inclusion ou d'exclusion.
- Pour accorder une autorisation `SELECT` sur une table avec filtrage par colonne, vous devez avoir été autorisé `SELECT` sur la table avec l'option d'attribution et sans aucune restriction de ligne. Vous devez avoir accès à toutes les lignes.
- Si vous accordez une subvention `SELECT` avec l'option de subvention et le filtrage des colonnes à un principal de votre compte, ce principal doit spécifier le filtrage des colonnes pour les mêmes colonnes ou pour un sous-ensemble des colonnes accordées lorsqu'il accorde à un autre principal. Si vous accordez `SELECT` à un compte externe à l'aide de l'option d'attribution et du filtrage des colonnes, l'administrateur du lac de données du compte externe peut accorder une autorisation `SELECT` sur toutes les colonnes à un autre principal de son compte. Cependant, même dans toutes `SELECT` les colonnes, ce principal n'aura de visibilité que sur les colonnes accordées au compte externe.
- Vous ne pouvez pas appliquer le filtrage des colonnes aux clés de partition.
- Un principal `SELECT` autorisé autorisé à accéder à un sous-ensemble de colonnes d'une table ne peut pas obtenir l'`INSERT`autorisation `ALTER`,`DROP`,`DELETE`, ou sur cette table. Pour un directeur disposant de l'`INSERT`autorisation `ALTER`,`DROP`,`DELETE`, ou sur une table, si vous accordez l'`SELECT`autorisation avec le filtrage des colonnes, cela n'a aucun effet.

Les remarques et restrictions suivantes s'appliquent au filtrage des colonnes imbriquées :

- Vous pouvez inclure ou exclure cinq niveaux de champs imbriqués dans un filtre de données.

  Example

  Col1.Col1_1.Col1_1_1.Col1_1_1_1.Col1_1_1_1_1_1_1
- Vous ne pouvez pas appliquer de filtrage de colonnes à des champs imbriqués dans des colonnes de partition.
- Si le schéma de votre table contient un nom de colonne de premier niveau (« client »). » adresse ») qui a le même modèle de représentation de champs imbriqués dans un filtre de données (une colonne imbriquée avec un nom de colonne de premier niveau `customer` et un nom de champ imbriqué `address` est spécifiée comme `"customer"."address"` dans un filtre de données), vous ne pouvez pas spécifier explicitement l'accès à une colonne de niveau supérieur ou à un champ imbriqué car les deux sont représentés selon le même modèle dans les listes d'inclusion/exclusion. Cela est ambigu, et Lake Formation ne peut pas être résolu si vous spécifiez la colonne de niveau supérieur ou le champ imbriqué.
- Si le nom d'une colonne ou d'un champ imbriqué de niveau supérieur contient un guillemet double, vous devez inclure un deuxième guillemet double lorsque vous spécifiez l'accès à un champ imbriqué dans la liste d'inclusion et d'exclusion d'un filtre de cellules de données.

  Example

  Exemple de nom de colonne imbriqué entre guillemets : `a.b.double"quote`

  Example

  Exemple de représentation de colonnes imbriquées dans un filtre de données : `"a"."b"."double""quote"`

Autorisations requises pour interroger des tables avec filtrage au niveau des cellules

Les autorisations AWS Identity and Access Management (IAM) suivantes sont requises pour exécuter des requêtes sur des tables avec un filtrage au niveau des cellules.

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults"
            ],
            "Resource": "*"
        }
    ]
}
```

Pour plus d'informations sur les autorisations de Lake Formation, consultezRéférence des personnalités de Lake Formation et des autorisations IAM.

Gestion des filtres de données

Pour implémenter la sécurité au niveau des colonnes, des lignes et des cellules, vous pouvez créer et gérer des filtres de données. Chaque filtre de données appartient à une table du catalogue de données. Vous pouvez créer plusieurs filtres de données pour une table, puis utiliser un ou plusieurs d'entre eux lorsque vous accordez des autorisations sur la table. Vous pouvez également définir et appliquer des filtres de données sur des colonnes imbriquées dont les `struct` types de données permettent aux utilisateurs d'accéder uniquement aux sous-structures des colonnes imbriquées.

Vous avez besoin `SELECT` d'une autorisation avec l'option d'octroi pour créer ou afficher un filtre de données. Pour permettre aux principaux utilisateurs de votre compte de consulter et d'utiliser un filtre de données, vous pouvez leur accorder l'`DESCRIBE`autorisation.

> **Note**
>
> Lake Formation n'est pas favorable à l'octroi d'une `Describe` autorisation sur un filtre de données partagé depuis un autre compte.

Vous pouvez gérer les filtres de données à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Pour plus d'informations sur les filtres de données, voir Filtres de données dans Lake Formation

Création d'un filtre de données

Vous pouvez créer un ou plusieurs filtres de données pour chaque table du catalogue de données.

Pour créer un filtre de données pour une table du catalogue de données (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

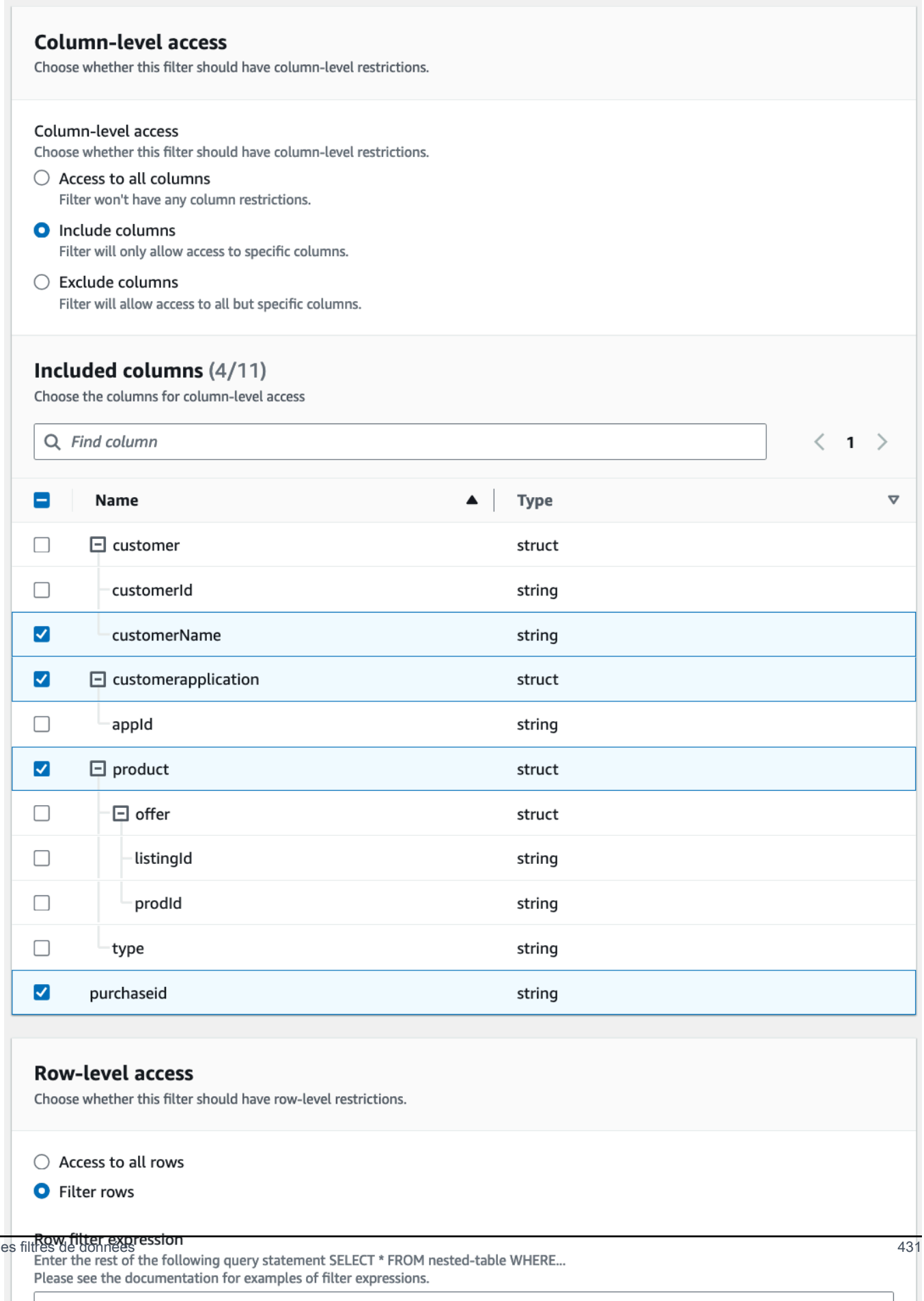
   Connectez-vous en tant qu'administrateur du lac de données, propriétaire de la table cible ou principal disposant d'une autorisation Lake Formation sur la table cible.
2. Dans le volet de navigation, sous Catalogue de données, sélectionnez Filtres de données.
3. Sur la page Filtres de données, choisissez Créer un nouveau filtre.
4. Dans la boîte de dialogue Créer un filtre de données, entrez les informations suivantes :

   - Nom du filtre de données
   - Base de données cible : spécifiez la base de données qui contient la table.
   - Tableau des cibles
   - Accès au niveau des colonnes : laissez ce paramètre défini sur Accès à toutes les colonnes pour spécifier le filtrage des lignes uniquement. Choisissez Inclure les colonnes ou Exclure les colonnes pour définir le filtrage des colonnes ou des cellules, puis spécifiez les colonnes à inclure ou à exclure.

     Colonnes imbriquées : si vous appliquez le filtre à une table contenant des colonnes imbriquées, vous pouvez spécifier explicitement les sous-structures des colonnes de structure imbriquées dans un filtre de données.

     Lorsque vous accordez l'autorisation SELECT à un principal sur ce filer, le principal exécutant la requête suivante ne verra que les données pour `customer.customerName` et non`customer.customerId`.

     ```
     SELECT "customer" FROM "example_db"."example_table";
     ```


Column-level access
Choose whether this filter should have column-level restrictions.
Column-level access
Choose whether this filter should have column-level restrictions.
Access to all columns
Filter won't have any column restrictions.
Include columns
Filter will only allow access to specific columns.
Exclude columns
Filter will allow access to all but specific columns.
Included columns (4/11)
Choose the columns for column-level access
Find column
1
Name
Type
customer
struct
customerId
string
customerName
string
customerapplication
struct
appId
string
product
struct
offer
struct
listingId
string
prodId
string
type
string
purchaseid
string
Row-level access
Choose whether this filter should have row-level restrictions.
Access to all rows
Filter rows
Row filter expression
Enter the rest of the following query statement SELECT * FROM nested-table WHERE...
Please see the documentation for examples of filter expressions.
customer.customerName <> 'John'

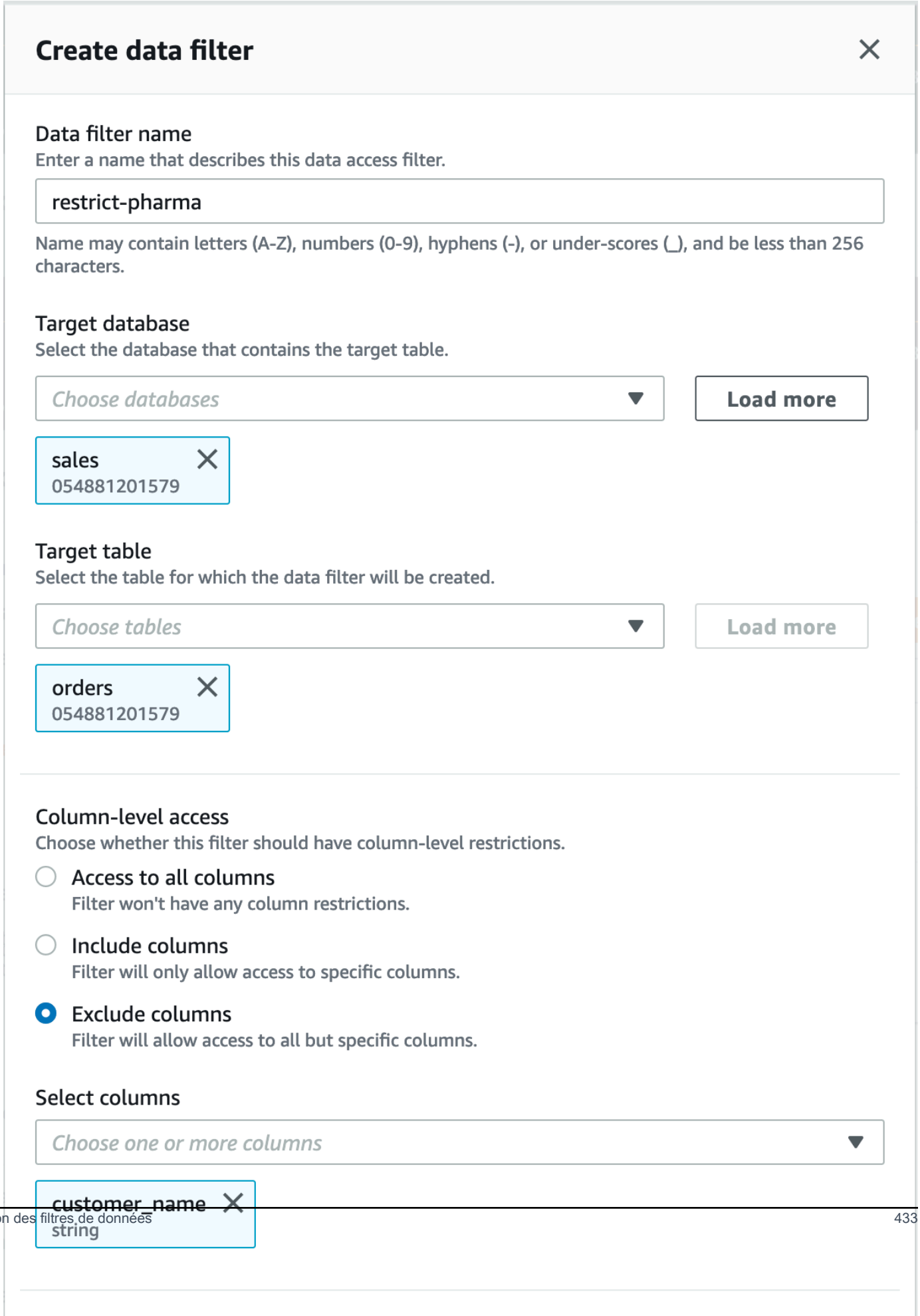
Lorsque vous accordez des autorisations à la `customer` colonne, le principal reçoit l'accès à la colonne et aux champs imbriqués situés sous la colonne (`customerNameetcustomerID`).

- Expression de filtre de ligne : entrez une expression de filtre pour spécifier le filtrage des lignes ou des cellules. Pour les types de données et les opérateurs pris en charge, consultezSupport partiQL dans les expressions de filtre de ligne. Choisissez Accès à toutes les lignes pour accorder l'accès à toutes.

  Vous pouvez inclure des structures de colonnes partielles issues de colonnes imbriquées dans une expression de filtre de ligne afin de filtrer les lignes contenant une valeur spécifique.

  Lorsqu'un principal est autorisé à accéder à une table avec une expression `Select * from example_nestedtable where customer.customerName <>'John'` de filtre de ligne et que l'accès au niveau des colonnes est défini sur Accès à toutes les colonnes, les résultats de la requête n'affichent que les lignes dont la valeur est `customerName <>'John'` vraie.

La capture d'écran suivante montre un filtre de données qui implémente le filtrage des cellules. Dans les requêtes portant sur la `orders` table, il refuse l'accès à la `customer_name` colonne et affiche uniquement les lignes contenant « pharma » dans la `product_type` colonne.


Create data filter
Data filter name
Enter a name that describes this data access filter.
restrict-pharma
Name may contain letters (A-Z), numbers (0-9), hyphens (-), or under-scores (_), and be less than 256 characters.
Target database
Select the database that contains the target table.
Choose databases
Load more
sales
054881201579
Target table
Select the table for which the data filter will be created.
Choose tables
Load more
orders
054881201579
Column-level access
Choose whether this filter should have column-level restrictions.
Access to all columns
Filter won't have any column restrictions.
Include columns
Filter will only allow access to specific columns.
Exclude columns
Filter will allow access to all but specific columns.
Select columns
Choose one or more columns
customer_name
string

5. Choisissez Create filter (Créer un filtre).

Pour créer un filtre de données avec des politiques de filtre cellulaire sur un champ imbriqué

Cette section utilise l'exemple de schéma suivant pour montrer comment créer un filtre de cellules de données :

```
[
    { name: "customer", type: "struct<customerId:string,customerName:string>" },
    { name: "customerApplication", type: "struct<appId:string>" },
    { name: "product", type:
 "struct<offer:struct<prodId:string,listingId:string>,type:string>" },
    { name: "purchaseId", type: "string" },
]
```

1. Sur la page Créer un filtre de données, entrez le nom du filtre de données.
2. Ensuite, utilisez le menu déroulant pour choisir le nom de la base de données et le nom de la table.
3. Dans la section Accès au niveau des colonnes, choisissez Colonnes incluses, puis sélectionnez une colonne imbriquée (). `customer.customerName`
4. Dans la section Accès au niveau des lignes, choisissez l'option Accès à toutes les lignes.
5. Choisissez Create filter (Créer un filtre).

   Lorsque vous `SELECT` autorisez ce filtre, le principal a accès à toutes les lignes de la `customerName` colonne.
6. Définissez ensuite un autre filtre de données pour la même base de données/table.
7. Dans la section Accès au niveau des colonnes, choisissez Colonnes incluses, puis sélectionnez une autre colonne imbriquée (). `customer.customerid`
8. Dans la section Accès au niveau des lignes, choisissez Filtrer les lignes, puis entrez une expression de filtre de ligne ()`customer.customerid <> 5`.
9. Choisissez Create filter (Créer un filtre).

   Lorsque vous `SELECT` autorisez ce filtre, le principal a accès à toutes les lignes des `customerId` champs`customerName`, à l'exception de la cellule dont la valeur est 5 dans la `customerId` colonne.

Octroi d'autorisations de filtrage de données

Vous pouvez accorder les `SELECT` autorisations `DESCRIBE` et `DROP` Lake Formation sur les filtres de données aux principaux.

Dans un premier temps, vous êtes le seul à pouvoir visualiser les filtres de données que vous créez pour une table. Pour permettre à un autre principal de consulter un filtre de données et d'accorder des autorisations au catalogue de données avec le filtre de données, vous devez soit :

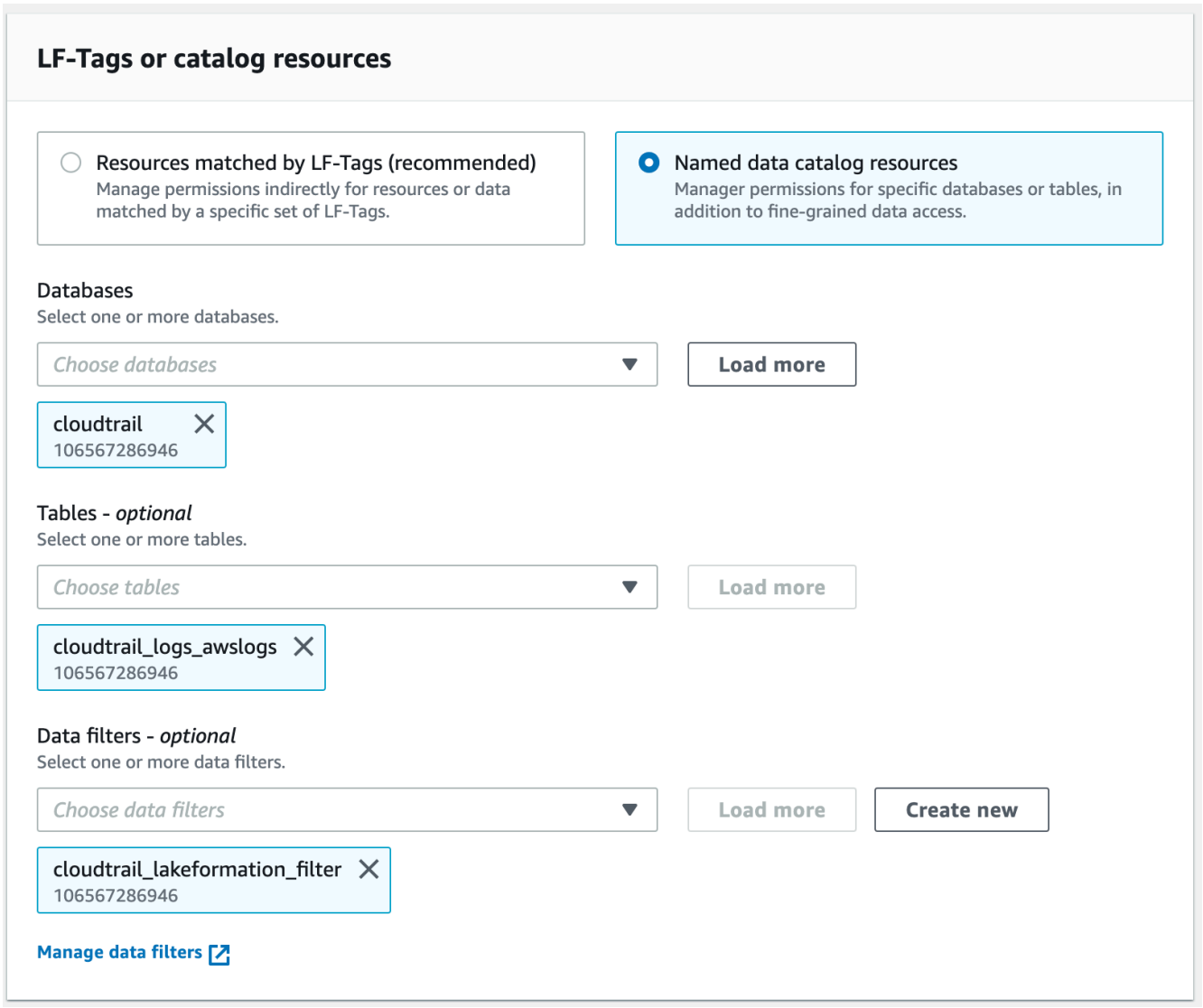
- Accordez `SELECT` sur un tableau au directeur avec l'option de subvention, et appliquez le filtre de données à la subvention.
- Accordez l'`DROP`autorisation `DESCRIBE` or sur le filtre de données au principal.

Vous pouvez accorder l'`SELECT`autorisation à un AWS compte externe. L'administrateur du lac de données de ce compte peut ensuite accorder cette autorisation aux autres principaux du compte. Lorsque vous accordez des autorisations à un compte externe, vous devez inclure l'option d'octroi afin que l'administrateur du compte externe puisse transférer en cascade les autorisations aux autres utilisateurs de son compte. Lorsque vous accordez à un mandant de votre compte, l'option de subvention est facultative.

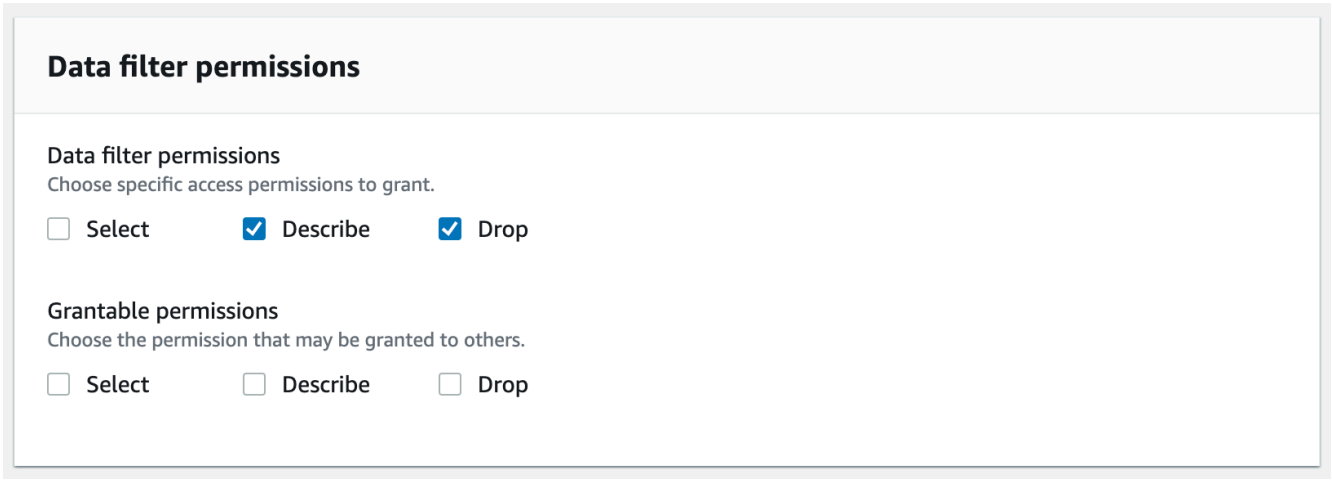
Vous pouvez accorder et révoquer des autorisations sur les filtres de données à l'aide de la AWS Lake Formation console, de l'API ou du AWS Command Line Interface (AWS CLI).

Console

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données.
3. Sur la page Autorisations, dans la section Autorisations relatives aux données, choisissez Grant.
4. Sur la page Accorder les autorisations relatives aux données, choisissez les principaux auxquels accorder les autorisations.
5. Dans la section Balises LF ou ressources de catalogue, sélectionnez Ressources de catalogue de données nommées. Choisissez ensuite la base de données, la table et le filtre de données pour lesquels vous souhaitez accorder des autorisations.



6. Dans la section Autorisations du filtre de données, choisissez les autorisations que vous souhaitez accorder aux principaux sélectionnés.



AWS CLI

- Entrez une `grant-permissions` commande. Spécifiez `DataCellsFilter` pour l'`resource`argument, et spécifiez `DESCRIBE` ou `DROP` pour l'`Permissions`argument et, éventuellement, pour l'`PermissionsWithGrantOption`argument.

  L'exemple suivant octroie l'option grant `DESCRIBE` à l'utilisateur `datalake_user1` sur le filtre de donnéesrestrict-pharma, qui appartient à la `orders` table de la `sales` base de données dans le AWS compte 1111-2222-3333.

  ```
  aws lakeformation grant-permissions --cli-input-json file://grant-params.json
  ```

  Le contenu du fichier est le suivant`grant-params.json`.

  ```
  {
      "Principal": {"DataLakePrincipalIdentifier":
   "arn:aws:iam::111122223333:user/datalake_user1"},
      "Resource": {
          "DataCellsFilter": {
              "TableCatalogId": "111122223333",
              "DatabaseName": "sales",
              "TableName": "orders",
              "Name": "restrict-pharma"
          }
      },
      "Permissions": ["DESCRIBE"],
      "PermissionsWithGrantOption": ["DESCRIBE"]
  }
  ```

Octroi des autorisations de données fournies par les filtres de données

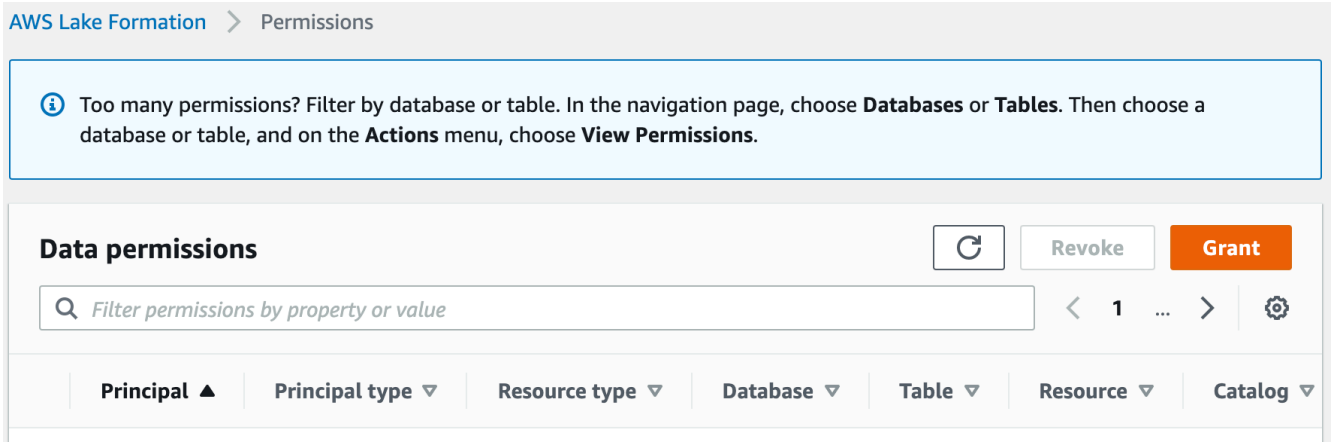
Les filtres de données représentent un sous-ensemble de données d'une table. Pour fournir un accès aux données aux principaux, `SELECT` des autorisations doivent être accordées à ces principaux. Avec cette autorisation, les directeurs peuvent :

- Afficher le nom réel de la table dans la liste des tables partagées avec leur compte.
- Créez des filtres de données sur la table partagée et accordez des autorisations à leurs utilisateurs sur ces filtres de données.

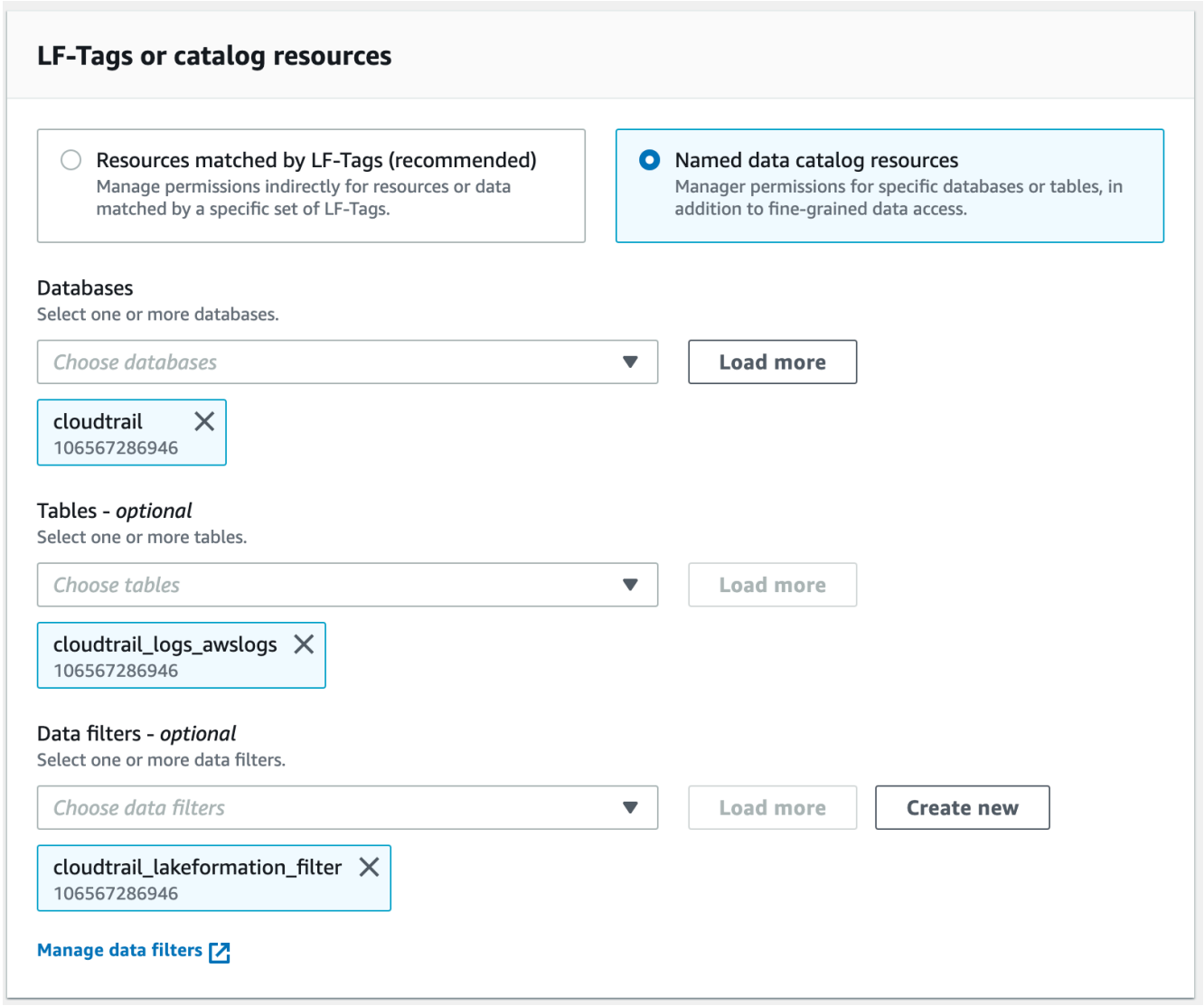
Console

Pour accorder des autorisations SELECT

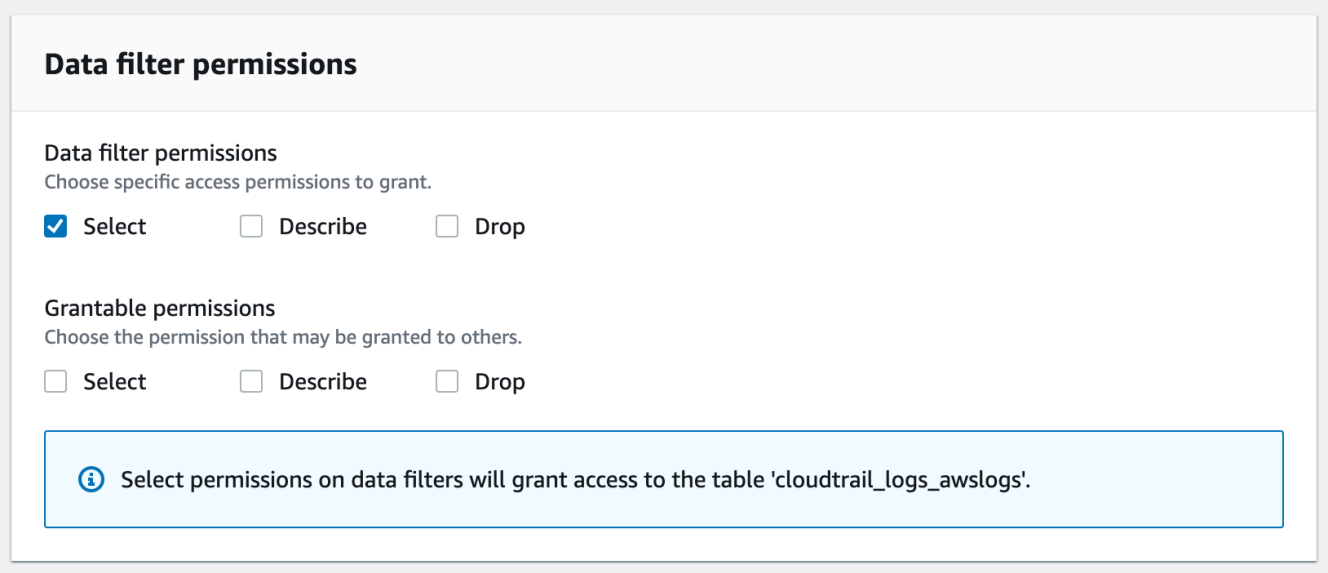
1. Accédez à la page Autorisations de la console Lake Formation, puis choisissez Grant.



2. Sélectionnez les principaux auxquels vous souhaitez donner accès, puis sélectionnez Ressources de catalogue de données nommées.



3. Pour donner accès aux données représentées par le filtre, choisissez Sélectionner sous Autorisations du filtre de données.



CLI

Entrez une `grant-permissions` commande. Spécifiez `DataCellsFilter` l'argument ressource et spécifiez `SELECT` l'argument Permissions.

L'exemple suivant octroie `SELECT` l'option grant à l'utilisateur `datalake_user1` sur le filtre de données`restrict-pharma`, qui appartient à la `orders` table de la `sales` base de données dans Compte AWS `1111-2222-3333`.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

Le contenu du fichier est le suivant`grant-params.json`.

```
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
    },
    "Resource": {
        "DataCellsFilter": {
            "TableCatalogId": "111122223333",
            "DatabaseName": "sales",
            "TableName": "orders",
            "Name": "restrict-pharma"
        }
    },
```

```
    "Permissions": ["SELECT"]
}
```

Affichage des filtres de données

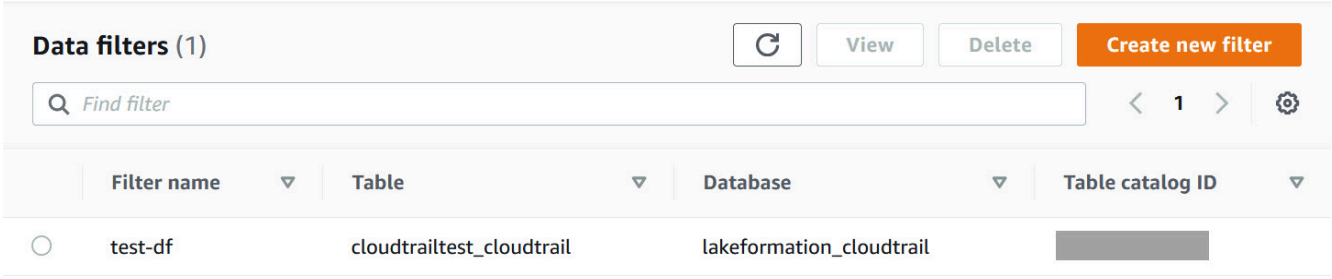
Vous pouvez utiliser la console Lake Formation ou AWS CLI l'API Lake Formation pour afficher les filtres de données.

Pour afficher les filtres de données, vous devez être un administrateur de Data Lake ou disposer des autorisations requises sur les filtres de données.

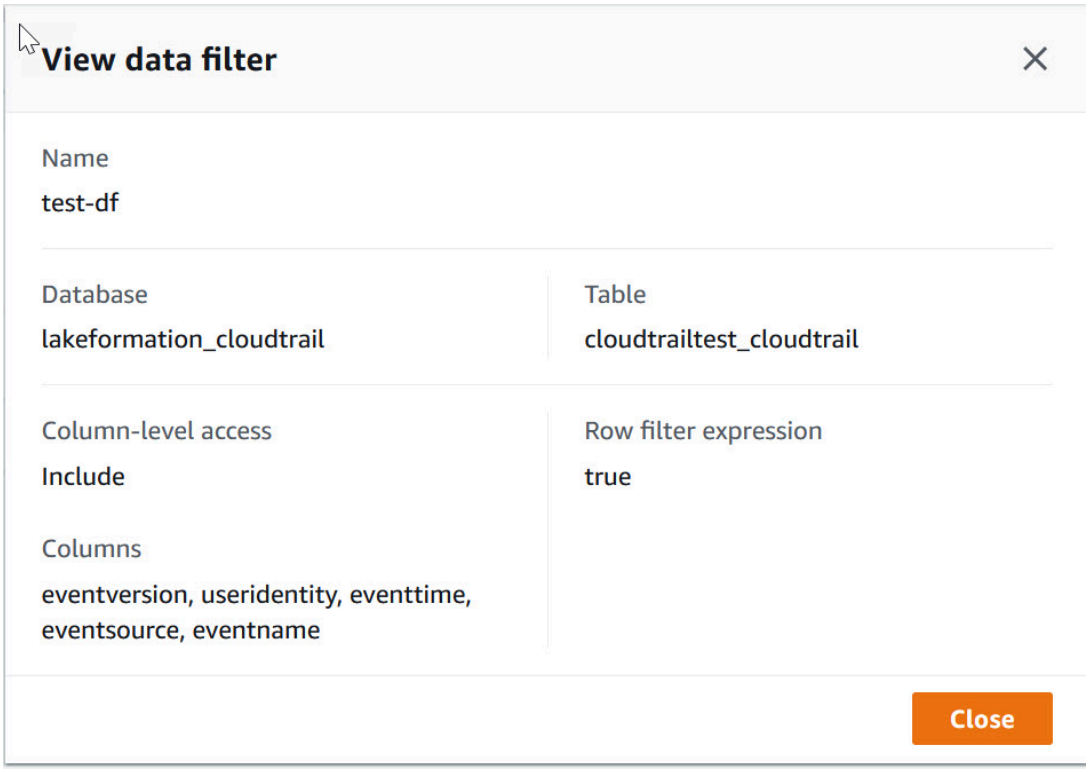
Console

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Dans le volet de navigation, sous Catalogue de données, sélectionnez Filtres de données.

   La page affiche les filtres de données auxquels vous avez accès.



3. Pour afficher les détails du filtre de données, choisissez le filtre de données, puis cliquez sur Afficher. Une nouvelle fenêtre apparaît avec les informations détaillées du filtre de données.



AWS CLI

Entrez une `list-data-cells-filter` commande et spécifiez une ressource de table.

L'exemple suivant répertorie les filtres de données de la `cloudtrailtest_cloudtrail` table.

```
aws lakeformation list-data-cells-filter --table '{ "CatalogId":"123456789012",
"DatabaseName":"lakeformation_cloudtrail", "Name":"cloudtrailtest_cloudtrail"}'
```

API/SDK

Utilisez l'`ListDataCellsFilter`API et spécifiez une ressource de table.

L'exemple suivant utilise Python pour répertorier les 20 premiers filtres de données de la `myTable` table.

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
```

```
    },
    MaxResults=20
)
```

Autorisations de filtrage des données de liste

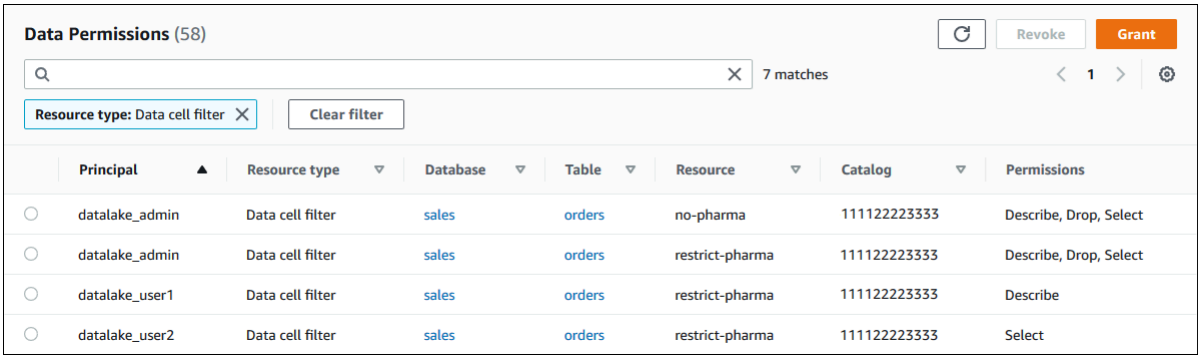
Vous pouvez utiliser la console Lake Formation pour consulter les autorisations accordées sur les filtres de données.

Pour consulter les autorisations sur un filtre de données, vous devez être un administrateur de Data Lake ou disposer des autorisations requises sur le filtre de données.

Console

1. Connectez-vous à la console Lake Formation AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations relatives aux données.
3. Sur la page Autorisations relatives aux données, cliquez ou appuyez dans le champ de recherche, puis dans le menu Propriétés, choisissez Type de ressource.
4. Dans le menu Type de ressource, choisissez Type de ressource : filtre de cellules de données.

   Les filtres de données pour lesquels vous avez des autorisations sont répertoriés. Il se peut que vous deviez faire défiler la page horizontalement pour voir les colonnes Permissions et Octroyable.



AWS CLI

- Entrez une `list-permissions` commande. Spécifiez `DataCellsFilter` pour l'`resource`argument, et spécifiez `DESCRIBE` ou `DROP` pour l'`Permissions`argument et, éventuellement, pour l'`PermissionsWithGrantOption`argument.

  L'exemple suivant répertorie les `DESCRIBE` autorisations avec l'option d'octroi sur le filtre de données`restrict-pharma`. Les résultats sont limités aux autorisations accordées pour le principal `datalake_user1` et la `orders` table de la `sales` base de données dans le AWS compte 1111-2222-3333.

  ```
  aws lakeformation list-permissions --cli-input-json file://list-params.json
  ```

  Le contenu du fichier est le suivant`grant-params.json`.

  ```
  {
      "Principal": {"DataLakePrincipalIdentifier":
   "arn:aws:iam::111122223333:user/datalake_user1"},
      "Resource": {
          "DataCellsFilter": {
              "TableCatalogId": "111122223333",
              "DatabaseName": "sales",
              "TableName": "orders",
              "Name": "restrict-pharma"
          }
      },
      "Permissions": ["DESCRIBE"],
      "PermissionsWithGrantOption": ["DESCRIBE"]
  }
  ```

Affichage des autorisations de base de données et de tables dans Lake Formation

Vous pouvez consulter les autorisations de Lake Formation accordées sur une base de données ou une table de catalogue de données. Vous pouvez le faire en utilisant la console Lake Formation, l'API ou le AWS Command Line Interface (AWS CLI).

À l'aide de la console, vous pouvez consulter les autorisations à partir des pages Bases de données ou Tables, ou à partir de la page Autorisations relatives aux données.

> **Note**
>
> Si vous n'êtes ni administrateur de base de données ni propriétaire de ressource, vous pouvez consulter les autorisations accordées aux autres principaux sur la ressource uniquement si vous disposez d'une autorisation Lake Formation sur la ressource avec l'option d'attribution.
> Outre les autorisations requises pour Lake Formation, vous avez besoin des autorisations AWS Identity and Access Management (IAM)`glue:GetDatabases`,`glue:GetDatabase`, `glue:GetTablesglue:GetTable`, et`glue:ListPermissions`.

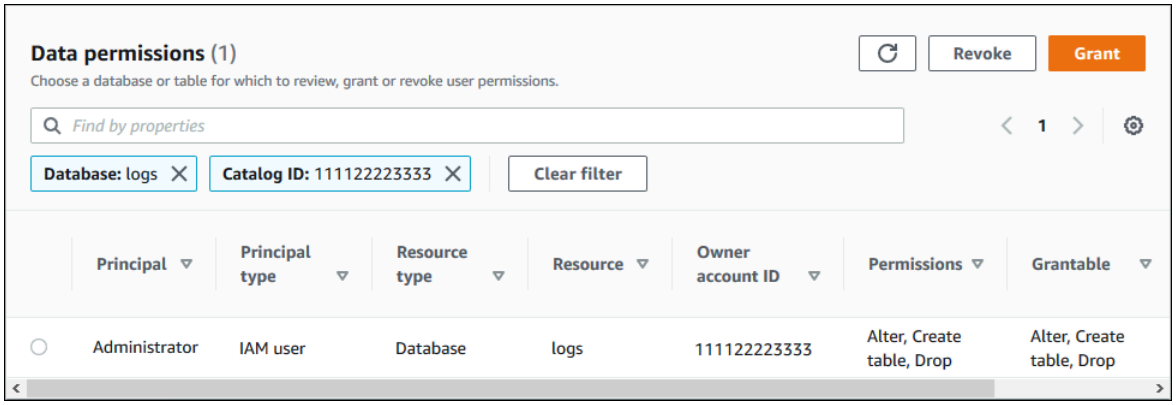
Pour afficher les autorisations sur une base de données (console, à partir de la page Bases de données)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données, créateur de base de données ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la base de données avec l'option grant.
2. Dans le panneau de navigation, choisissez Databases (Bases de données).
3. Choisissez une base de données, puis dans le menu Actions, choisissez Afficher les autorisations.

   > **Note**
   >
   > Si vous choisissez un lien de ressource de base de données, Lake Formation affiche les autorisations sur le lien de ressource, et non sur la base de données cible du lien de ressource.

   La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la base de données. Le nom de la base de données et l'ID de catalogue (ID de AWS compte) du propriétaire de la base de données apparaissent sous forme d'étiquettes sous la zone de recherche. Les vignettes indiquent qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette base de données. Vous pouvez ajuster le filtre en fermant une vignette ou en choisissant Effacer le filtre.



Pour afficher les autorisations sur une base de données (console, à partir de la page Autorisations relatives aux données)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données, créateur de base de données ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la base de données avec l'option grant.
2. Dans le volet de navigation, sélectionnez Autorisations relatives aux données.
3. Positionnez le curseur dans le champ de recherche en haut de la page, puis dans le menu Propriétés qui apparaît, sélectionnez Base de données.
4. Dans le menu Bases de données qui apparaît, choisissez une base de données.

> **Note**
>
> Si vous choisissez un lien de ressource de base de données, Lake Formation affiche les autorisations sur le lien de ressource, et non sur la base de données cible du lien de ressource.

La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la base de données. Le nom de la base de données apparaît sous forme de vignette sous la zone de recherche. La vignette indique qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette base de données. Vous pouvez supprimer le filtre en fermant la vignette ou en choisissant Effacer le filtre.

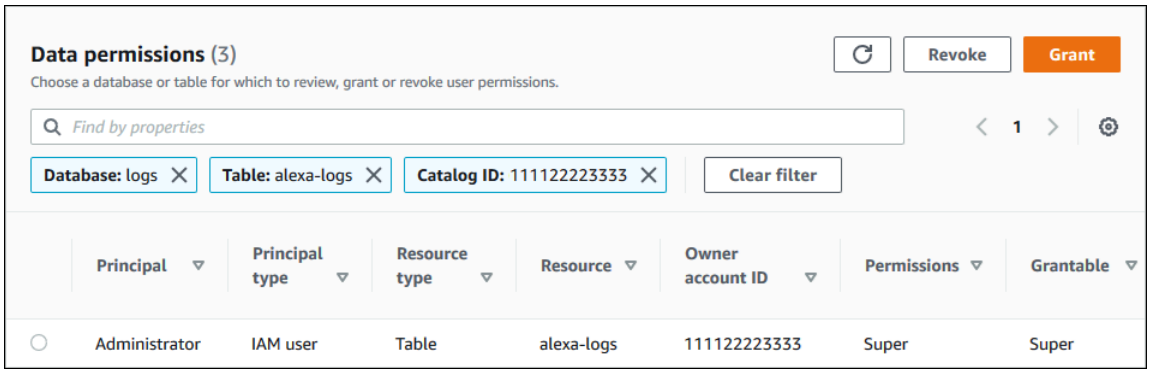
Pour afficher les autorisations sur une table (console, à partir de la page Tables)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données, créateur de la table ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la table avec l'option d'attribution.
2. Dans le volet de navigation, choisissez Tables.
3. Choisissez un tableau, puis dans le menu Actions, choisissez Afficher les autorisations.

   > **Note**
   >
   > Si vous choisissez un lien de ressource de table, Lake Formation affiche les autorisations sur le lien de ressource, et non sur la table cible du lien de ressource.

   La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la table. Le nom de la table, le nom de la base de données contenant la table et l'ID de catalogue (ID de AWS compte) du propriétaire de la table apparaissent sous forme d'étiquettes sous la zone de recherche. Les étiquettes indiquent qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette table. Vous pouvez ajuster le filtre en fermant une étiquette ou en choisissant Effacer le filtre.



Pour afficher les autorisations sur une table (console, à partir de la page Autorisations relatives aux données)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données, créateur de la table ou en tant qu'utilisateur disposant d'une autorisation Lake Formation sur la table avec l'option d'attribution.

2. Dans le volet de navigation, sélectionnez Autorisations relatives aux données.
3. Positionnez le curseur dans le champ de recherche en haut de la page, puis dans le menu Propriétés qui apparaît, sélectionnez Base de données.
4. Dans le menu Bases de données qui apparaît, choisissez une base de données.

   > **Important**
   >
   > Si vous souhaitez consulter les autorisations sur une table partagée avec votre AWS compte à partir d'un compte externe, vous devez choisir la base de données du compte externe qui contient la table, et non un lien de ressource vers la base de données.

   La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la base de données.
5. Positionnez à nouveau le curseur dans la zone de recherche, puis dans le menu Propriétés qui apparaît, sélectionnez Tableau.
6. Dans le menu Tables qui apparaît, choisissez un tableau.

   La page Autorisations relatives aux données répertorie toutes les autorisations de Lake Formation pour la table. Le nom de la table et le nom de la base de données contenant la table apparaissent sous forme de vignettes sous la zone de recherche. Les vignettes indiquent qu'un filtre a été appliqué pour répertorier les autorisations uniquement pour cette table. Vous pouvez ajuster le filtre en fermant une vignette ou en choisissant Effacer le filtre.

Pour afficher les autorisations sur une table (AWS CLI)

- Entrez une `list-permissions` commande.

  L'exemple suivant répertorie les autorisations sur une table partagée depuis un compte externe. La `CatalogId` propriété est l'ID de AWS compte du compte externe, et le nom de la base de données fait référence à la base de données du compte externe qui contient la table.

```
aws lakeformation list-permissions  --resource-type TABLE --resource '{ "Table":
 {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"}}'
```

Révocation de l'autorisation à l'aide de la console Lake Formation

Vous pouvez utiliser la console pour révoquer tous les types d'autorisations relatives à Lake Formation : autorisations relatives au catalogue de données, autorisations relatives aux balises de politique, autorisations de filtrage des données et autorisations de localisation.

Pour révoquer les autorisations de Lake Formation sur une ressource (console)

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur ayant obtenu des autorisations avec l'option d'octroi sur la ressource.
2. Dans le volet de navigation, sous Autorisations, sélectionnez Autorisations du lac de données, balises LF et autorisations, ou Emplacements des données.
3. Sélectionnez l'autorisation ou l'emplacement, puis choisissez Révoquer.
4. Dans la boîte de dialogue qui s'ouvre, choisissez Révoquer.

Partage de données entre comptes dans Lake Formation

Les fonctionnalités multi-comptes de Lake Formation permettent aux utilisateurs de partager en toute sécurité des lacs de données distribués entre plusieurs AWS organisations ou directement avec les responsables IAM d'un autre compte Comptes AWS, offrant ainsi un accès détaillé aux métadonnées du catalogue de données et aux données sous-jacentes. Les grandes entreprises en utilisent généralement plusieurs Comptes AWS, et bon nombre de ces comptes peuvent avoir besoin d'accéder à un lac de données géré par un seul Compte AWS. Les utilisateurs et les tâches AWS Glue d'extraction, de transformation et de chargement (ETL) peuvent interroger et joindre des tables sur plusieurs comptes tout en bénéficiant de la protection des données au niveau des tables et des colonnes de Lake Formation.

Lorsque vous accordez des autorisations sur une ressource du catalogue de données à un compte externe ou directement à un responsable IAM d'un autre compte, Lake Formation utilise le service AWS Resource Access Manager (AWS RAM) pour partager la ressource. Si le compte du bénéficiaire appartient à la même organisation que le compte du donateur, la ressource partagée est immédiatement accessible au bénéficiaire. Si le compte du bénéficiaire n'appartient pas à la même organisation, AWS RAM envoie une invitation au compte du bénéficiaire pour qu'il accepte ou rejette la subvention de ressources. Ensuite, pour rendre la ressource partagée disponible, l'administrateur

du lac de données du compte bénéficiaire doit utiliser la AWS RAM console ou AWS CLI accepter l'invitation.

Lake Formation prend en charge le partage des ressources du catalogue de données avec des comptes externes en mode d'accès hybride. Le mode d'accès hybride offre la flexibilité d'activer de manière sélective les autorisations de Lake Formation pour les bases de données et les tables de votre AWS Glue Data Catalog.
Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

Pour plus d’informations, consultez Mode d'accès hybride.

Partage direct entre comptes

Les principaux autorisés peuvent partager des ressources de manière explicite avec un directeur IAM sur un compte externe. Cette fonctionnalité est utile lorsqu'un propriétaire de compte souhaite contrôler les utilisateurs du compte externe qui peuvent accéder aux ressources. Les autorisations que recevra le directeur de l'IAM seront une combinaison de subventions directes et de subventions au niveau du compte, qui seront répercutées en cascade sur les principaux. L'administrateur du lac de données du compte bénéficiaire peut consulter les autorisations directes entre comptes, mais ne peut pas révoquer les autorisations. Le principal qui reçoit la part de ressources ne peut pas partager la ressource avec d'autres principaux.

Méthodes de partage des ressources du catalogue de données

Avec une seule opération de subvention Lake Formation, vous pouvez accorder des autorisations entre comptes sur les ressources du catalogue de données suivantes.

- Une base de données
- Un tableau individuel (avec filtrage des colonnes optionnel)
- Quelques tables sélectionnées
- Toutes les tables d'une base de données (en utilisant le caractère générique Toutes les tables)

Il existe deux options pour partager vos bases de données et vos tables avec un autre compte Compte AWS ou avec les principaux IAM d'un autre compte.

- Contrôle d'accès basé sur des balises Lake Formation (LF-TBAC) (recommandé)

Le contrôle d'accès basé sur les balises de Lake Formation est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Vous pouvez utiliser le contrôle d'accès basé sur des balises pour partager les ressources du catalogue de données (bases de données, tables et colonnes) avec des principaux IAM, des Comptes AWS Organisations et des unités organisationnelles (UO) externes. Dans Lake Formation, ces attributs sont appelés balises LF. Pour plus d'informations, voir Gestion d'un lac de données à l'aide du contrôle d'accès basé sur les balises Lake Formation.

> **Note**
>
> La méthode LF-TBAC d'octroi des autorisations de catalogue de données est utilisée AWS Resource Access Manager pour les autorisations entre comptes.
> Lake Formation prend désormais en charge l'octroi d'autorisations entre comptes aux Organisations et aux unités organisationnelles à l'aide de la méthode LF-TBAC.
> Pour activer cette fonctionnalité, vous devez mettre à jour les paramètres de version des comptes Cross vers la version 3.
> Pour plus d’informations, consultez Mise à jour des paramètres de version de partage de données entre comptes.

- Ressources nommées Lake Formation

  La méthode de partage de données entre comptes Lake Formation à l'aide de ressources nommées vous permet d'accorder des autorisations Lake Formation avec une option d'octroi sur les tables et les bases de données du catalogue de données à des responsables externes Comptes AWS, à des organisations ou à des unités organisationnelles IAM. L'opération de subvention partage automatiquement ces ressources.

> **Note**
>
> Vous pouvez également autoriser le AWS Glue robot d'exploration à accéder à un magasin de données dans un autre compte à l'aide des informations d'identification de Lake Formation. Pour plus d'informations, consultez la section Exploration entre comptes dans le Guide du AWS Glue développeur.

Les services intégrés tels qu'Athena et Amazon Redshift Spectrum nécessitent des liens de ressources pour pouvoir inclure des ressources partagées dans les requêtes. Pour plus d'informations sur les liens vers des ressources, consultezMode de fonctionnement des liens des ressources dans Lake Formation.

Pour les considérations et les limites, voirMeilleures pratiques et considérations relatives au partage de données entre comptes.

Rubriques

- Prérequis
- Mise à jour des paramètres de version de partage de données entre comptes
- Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des responsables IAM
- Octroi d'autorisations sur une base de données ou une table partagée avec votre compte
- Octroi d'autorisations relatives aux liens vers
- Accès aux données sous-jacentes d'une table partagée
- Journalisation entre comptes CloudTrail
- Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation
- Afficher toutes les subventions entre comptes à l'aide de l'opération GetResourceShares API

Rubriques en relation

- Vue d'ensemble des autorisations relatives à Lake Formation
- Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci
- Création de liens vers des ressources
- Résolution des problèmes d'accès entre comptes

Prérequis

Avant que votre AWS compte puisse partager les ressources du catalogue de données (bases de données et tables) avec un autre compte ou des principaux d'un autre compte, et avant de pouvoir

accéder aux ressources partagées avec votre compte, les conditions préalables suivantes doivent être remplies.

Exigences générales en matière de partage de données entre comptes

- Pour partager des bases de données et des tables du catalogue de données en mode d'accès hybride, vous devez mettre à jour les paramètres de version entre comptes vers la version 4.
- Avant d'accorder des autorisations entre comptes sur une ressource de catalogue de données, vous devez révoquer toutes les autorisations de Lake Formation accordées au `IAMAllowedPrincipals` groupe pour cette ressource. Si le principal appelant dispose d'autorisations intercomptes pour accéder à une ressource et que `IAMAllowedPrincipals` cette autorisation existe sur la ressource, Lake Formation lance la requête`AccessDeniedException`.

  Cette exigence s'applique uniquement lorsque vous enregistrez l'emplacement des données sous-jacentes en mode Lake Formation. Si vous enregistrez l'emplacement des données en mode hybride, les autorisations de `IAMAllowedPrincipals` groupe peuvent exister sur la base de données ou la table partagée.
- Pour les bases de données contenant des tables que vous souhaitez partager, vous devez empêcher que les nouvelles tables soient associées par défaut `Super` à`IAMAllowedPrincipals`. Sur la console Lake Formation, modifiez la base de données et désactivez Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données ou entrez la AWS CLI commande suivante en la `database` remplaçant par le nom de la base de données. Si l'emplacement des données sous-jacent est enregistré en mode d'accès hybride, il n'est pas nécessaire de modifier ce paramètre par défaut. En mode d'accès hybride, Lake Formation vous permet d'appliquer de manière sélective les autorisations Lake Formation et les politiques d'autorisations IAM pour Amazon S3 et AWS Glue sur la même ressource.

  ```
  aws glue update-database --name database --database-input
   '{"Name":"database","CreateTableDefaultPermissions":[]}'
  ```
- Pour accorder des autorisations entre comptes, le concédant doit disposer des autorisations AWS Identity and Access Management (IAM) requises sur et sur le AWS Glue service. AWS RAM La politique AWS gérée `AWSLakeFormationCrossAccountManager` accorde les autorisations requises.

Les administrateurs de lacs de données des comptes recevant des partages de ressources par le biais de ces comptes AWS RAM doivent appliquer la politique supplémentaire suivante. Il permet à l'administrateur d'accepter AWS RAM des invitations de partage de ressources. Cela permet également à l'administrateur d'activer le partage des ressources avec les organisations.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ram:AcceptResourceShareInvitation",
                "ram:RejectResourceShareInvitation",
                "ec2:DescribeAvailabilityZones",
                "ram:EnableSharingWithAwsOrganization"
            ],
            "Resource": "*"
        }
    ]
}
```

- Si vous souhaitez partager des ressources du catalogue de données avec AWS Organizations ou des unités organisationnelles, le partage avec les organisations doit être activé dans AWS RAM.

  Pour plus d'informations sur la manière d'activer le partage avec les organisations, voir Activer le partage avec AWS les organisations dans le Guide de AWS RAM l'utilisateur.

  Vous devez être `ram:EnableSharingWithAwsOrganization` autorisé à activer le partage avec les organisations.

- Pour partager des ressources directement avec le principal IAM d'un autre compte, vous devez mettre à jour les paramètres de version entre comptes vers la version 3. Ce paramètre est disponible sur la page des paramètres du catalogue de données. Si vous utilisez la version 1, consultez les instructions pour mettre à jour le paramètreMise à jour des paramètres de version de partage de données entre comptes.

- Vous ne pouvez pas partager les ressources du catalogue de données chiffrées avec une clé gérée par le AWS Glue service avec un autre compte. Vous ne pouvez partager que les ressources du catalogue de données chiffrées avec la clé de chiffrement du client, et le compte recevant le partage des ressources doit disposer des autorisations sur la clé de chiffrement du catalogue de données pour déchiffrer les objets.

Partage de données entre comptes selon les exigences du LF-TBAC

- Pour partager les ressources du catalogue de données AWS Organizations et les unités organisationnelles (UO), vous devez mettre à jour les paramètres de version multi-comptes vers la version 3.
- Pour partager les ressources du catalogue de données avec la version 3 des paramètres de version entre comptes, le concédant doit disposer des autorisations IAM définies dans la politique AWS `AWSLakeFormationCrossAccountManager` gérée de votre compte.
- Si vous utilisez la version 1 ou la version 2 des paramètres de version entre comptes, vous devez disposer d'une politique de ressources du catalogue de données (`glue:PutResourcePolicy`) qui active le LF-TBAC. Pour plus d'informations, consultez Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation.
- Si vous utilisez actuellement une politique de ressources du catalogue de AWS Glue données pour partager des ressources et que vous souhaitez accorder des autorisations entre comptes à l'aide de la version 3 des paramètres de version entre comptes, vous devez ajouter l'`glue:ShareResource`autorisation dans les paramètres du catalogue de données à l'aide de l'opération d'`glue:PutResourcePolicy`API, comme indiqué dans la Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation section. Cette politique n'est pas requise si votre compte n'a accordé aucune autorisation entre comptes en utilisant la politique de ressources du catalogue de AWS Glue données (`glue:PutResourcePolicy`autorisation d'utilisation des versions 1 et 2) pour accorder un accès entre comptes.

```
{
      "Effect": "Allow",
      "Action": [
        "glue:ShareResource"
      ],
      "Principal": {"Service": [
        "ram.amazonaws.com"
      ]},
      "Resource": [
        "arn:aws:glue:<region>:<account-id>:table/*/*",
        "arn:aws:glue:<region>:<account-id>:database/*",
        "arn:aws:glue:<region>:<account-id>:catalog"
      ]
    }
```

- Si votre compte a effectué des partages entre comptes en utilisant la politique de ressources du catalogue de AWS Glue données et que vous utilisez actuellement la méthode des ressources nommées ou le LF-TBAC avec la version 3 des paramètres entre comptes pour partager des ressources, qui utilise AWS RAM pour partager des ressources, vous devez définir l'`EnableHybrid`argument sur `'true'` lorsque vous appelez l'opération d'API. `glue:PutResourcePolicy` Pour plus d'informations, consultez Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation.

Configuration requise pour chaque compte accédant à la ressource partagée

- Si vous partagez des ressources avec Comptes AWS, au moins un utilisateur du compte client doit être un administrateur de data lake pour consulter les ressources partagées. Pour plus d'informations sur la création d'un administrateur de lac de données, consultezCréation d'un administrateur de lac de données.

  L'administrateur du lac de données peut accorder des autorisations de Lake Formation sur les ressources partagées aux autres principaux du compte. Les autres principaux ne peuvent pas accéder aux ressources partagées tant que l'administrateur du lac de données ne leur a pas accordé d'autorisations sur les ressources.
- Les services intégrés tels qu'Athena et Redshift Spectrum nécessitent des liens de ressources pour pouvoir inclure des ressources partagées dans les requêtes. Les directeurs doivent créer un lien de ressource dans leur catalogue de données vers une ressource partagée par une autre Compte AWS personne. Pour plus d'informations sur les liens vers des ressources, consultezMode de fonctionnement des liens des ressources dans Lake Formation.
- Lorsqu'une ressource est partagée directement avec un principal IAM, pour interroger la table à l'aide d'Athena, le principal doit créer un lien vers une ressource. Pour créer un lien vers une ressource, le directeur a besoin de la Lake Formation `CREATE_TABLE` `glue:CreateTable` ou `CREATE_DATABASE` de l'autorisation `glue:CreateDatabase` IAM.

  Si le compte producteur partage une table différente dans la même base de données avec le même principal ou un autre principal, ce principal peut immédiatement interroger la table.

> **Note**
>
> Pour l'administrateur du lac de données et pour les principaux auxquels l'administrateur du lac de données a accordé des autorisations, les ressources partagées apparaissent dans le

catalogue de données comme s'il s'agissait de ressources locales (détenues). Les tâches d'extraction, de transformation et de chargement (ETL) peuvent accéder aux données sous-jacentes des ressources partagées.
Pour les ressources partagées, les pages Tables et Bases de données de la console Lake Formation affichent l'ID de compte du propriétaire.
Lorsque les données sous-jacentes d'une ressource partagée sont accessibles, les événements du CloudTrail journal sont générés à la fois dans le compte du destinataire de la ressource partagée et dans le compte du propriétaire de la ressource. Les CloudTrail événements peuvent contenir l'ARN du principal qui a accédé aux données, mais uniquement si le compte du destinataire choisit d'inclure l'ARN principal dans les journaux. Pour plus d'informations, consultez Journalisation entre comptes CloudTrail .

Mise à jour des paramètres de version de partage de données entre comptes

Met à AWS Lake Formation jour de temps à autre les paramètres de partage de données entre comptes afin de distinguer les modifications apportées à l' AWS RAM utilisation et de prendre en charge les mises à jour apportées à la fonctionnalité de partage de données entre comptes. Lorsque Lake Formation effectue cette opération, il crée une nouvelle version des paramètres de version du compte Cross.

Principales différences entre les paramètres des versions multi-comptes

Pour plus d'informations sur le fonctionnement du partage de données entre comptes selon les différents paramètres de version entre comptes, consultez les sections suivantes.

Note

Pour partager des données avec un autre compte, le donateur doit avoir `AWSLakeFormationCrossAccountManager` géré les autorisations liées à la politique IAM. Il s'agit d'une condition préalable pour toutes les versions.
La mise à jour des paramètres de version de plusieurs comptes n'a aucune incidence sur les autorisations dont dispose le destinataire sur les ressources partagées. Cela s'applique lors de la mise à jour de la version 1 vers la version 2, de la version 2 vers la version 3 et de la version 1 vers la version 3. Consultez les considérations répertoriées ci-dessous lors de la mise à jour des versions.

Version 1

Méthode de ressource nommée : mappe chaque autorisation de Lake Formation accordée entre comptes à un partage de AWS RAM ressources. L'utilisateur (rôle du concédant ou principal) n'a pas besoin d'autorisations supplémentaires.

Méthode LF-TBAC : les autorisations accordées entre comptes sur la Lake Formation ne sont pas AWS RAM utilisées pour partager des données. L'utilisateur doit disposer d'une `glue:PutResourcePolicy` autorisation.

Avantages de la mise à jour des versions : Version initiale, non applicable.

Considérations relatives à la mise à jour des versions : Version initiale - non applicable

Version 2

Méthode des ressources nommées : optimise le nombre de partages de AWS RAM ressources en mappant plusieurs autorisations accordées entre comptes avec un seul partage de AWS RAM ressources. L'utilisateur n'a pas besoin d'autorisations supplémentaires.

Méthode LF-TBAC : les autorisations accordées entre comptes sur la Lake Formation ne sont pas AWS RAM utilisées pour partager des données. L'utilisateur doit disposer d'une `glue:PutResourcePolicy` autorisation.

Avantages de la mise à jour des versions : configuration multi-comptes évolutive grâce à une utilisation optimale de la AWS RAM capacité.

Considérations à prendre en compte lors de la mise à jour des versions : les utilisateurs qui souhaitent accorder des autorisations multi-comptes à Lake Formation doivent disposer des autorisations définies dans la politique `AWSLakeFormationCrossAccountManager` AWS gérée. Dans le cas contraire, vous devez disposer `ram:AssociateResourceShare` des `ram:DisassociateResourceShare` autorisations nécessaires pour partager correctement des ressources avec un autre compte.

Version 3

Méthode des ressources nommées : optimise le nombre de partages de AWS RAM ressources en mappant plusieurs autorisations accordées entre comptes avec un seul partage de AWS RAM ressources. L'utilisateur n'a pas besoin d'autorisations supplémentaires.

Méthode LF-TBAC : Lake Formation utilise AWS RAM pour les subventions entre comptes. L'utilisateur doit ajouter l'ShareResource instruction glue : à

l'`glue:PutResourcePolicy`autorisation. Le destinataire doit accepter les invitations à partager des ressources provenant de AWS RAM.

Avantages de la mise à jour des versions : prend en charge les fonctionnalités suivantes :

- Permet de partager des ressources de manière explicite avec un directeur IAM dans un compte externe.

  Pour plus d'informations, consultez Octroi et révocation d'autorisations sur les ressources du catalogue de données.
- Permet les partages entre comptes à l'aide de la méthode LF-TBAC pour les Organisations ou les unités organisationnelles (UO).
- Supprime les frais liés à la gestion de AWS Glue politiques supplémentaires pour les subventions entre comptes.

Considérations relatives à la mise à jour des versions : Lorsque vous utilisez la méthode LF-TBAC pour partager des ressources, si le concédant utilise une version inférieure à la version 3 et que le destinataire utilise la version 3 ou supérieure, le concédant reçoit le message d'erreur suivant : « Demande de subvention entre comptes non valide. Le compte client a opté pour la version multi-comptes : v3. Veuillez passer `CrossAccountVersion` `DataLakeSetting` à la version minimale v3 (Service : AmazonDataCatalog ; Code d'état : 400 ; Code d'erreur : InvalidInputException) ». Toutefois, si le concédant utilise la version 3 et que le destinataire utilise la version 1 ou la version 2, les subventions entre comptes utilisant des balises LF sont traitées avec succès.

Les subventions entre comptes accordées à l'aide de la méthode des ressources nommées sont compatibles entre les différentes versions. Même si le compte du concédant utilise une ancienne version (version 1 ou 2) et que le compte du bénéficiaire utilise une version plus récente (version 3 ou supérieure), la fonctionnalité d'accès entre comptes fonctionne parfaitement sans aucun problème de compatibilité ni erreur.

Pour partager des ressources directement avec les responsables IAM sur un autre compte, seul le donateur doit utiliser la version 3.

Les subventions entre comptes accordées à l'aide de la méthode LF-TBAC nécessitent que les utilisateurs disposent d'une politique de AWS Glue Data Catalog ressources dans le compte. Lorsque vous passez à la version 3, le LF-TBAC autorise des utilisations. AWS RAM Pour que les subventions AWS RAM basées sur plusieurs comptes soient couronnées de succès, vous devez ajouter la `glue:ShareResource` déclaration à vos politiques de ressources de catalogue de

données existantes, comme indiqué dans la Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation section.

Version 4

Le concédant a besoin de la version 4 ou supérieure pour partager les ressources du catalogue de données en mode d'accès hybride.

Optimisez le partage AWS RAM des ressources

Les nouvelles versions (version 2 et supérieures) des subventions entre comptes utilisent de manière optimale la AWS RAM capacité afin de maximiser l'utilisation entre comptes. Lorsque vous partagez une ressource avec un responsable externe Compte AWS ou un directeur IAM, Lake Formation peut créer un nouveau partage de ressources ou associer la ressource à un partage existant. En s'associant à des actions existantes, Lake Formation réduit le nombre d'invitations à partager des ressources qu'un consommateur doit accepter.

Activez AWS RAM les partages via TBAC ou partagez les ressources directement avec les principaux

Pour partager des ressources directement avec les responsables IAM d'un autre compte ou pour activer les partages entre comptes TBAC vers des Organisations ou des unités organisationnelles, vous devez mettre à jour les paramètres de version entre comptes vers la version 3. Pour plus d'informations sur les limites de AWS RAM ressources, consultezMeilleures pratiques et considérations relatives au partage de données entre comptes.

Autorisations requises pour mettre à jour les paramètres des versions entre comptes

Si un fournisseur d'autorisations entre comptes a `AWSLakeFormationCrossAccountManager` géré les autorisations de politique IAM, aucune configuration d'autorisation supplémentaire n'est requise pour le rôle ou le principal du concédant d'autorisations entre comptes. Toutefois, si le concédant entre comptes n'utilise pas la politique gérée, le rôle ou le principal du concédant doit disposer des autorisations IAM suivantes accordées pour que la nouvelle version de la subvention entre comptes soit couronnée de succès.

```
{
  "Version": "2012-10-17",
  "Statement": [
```

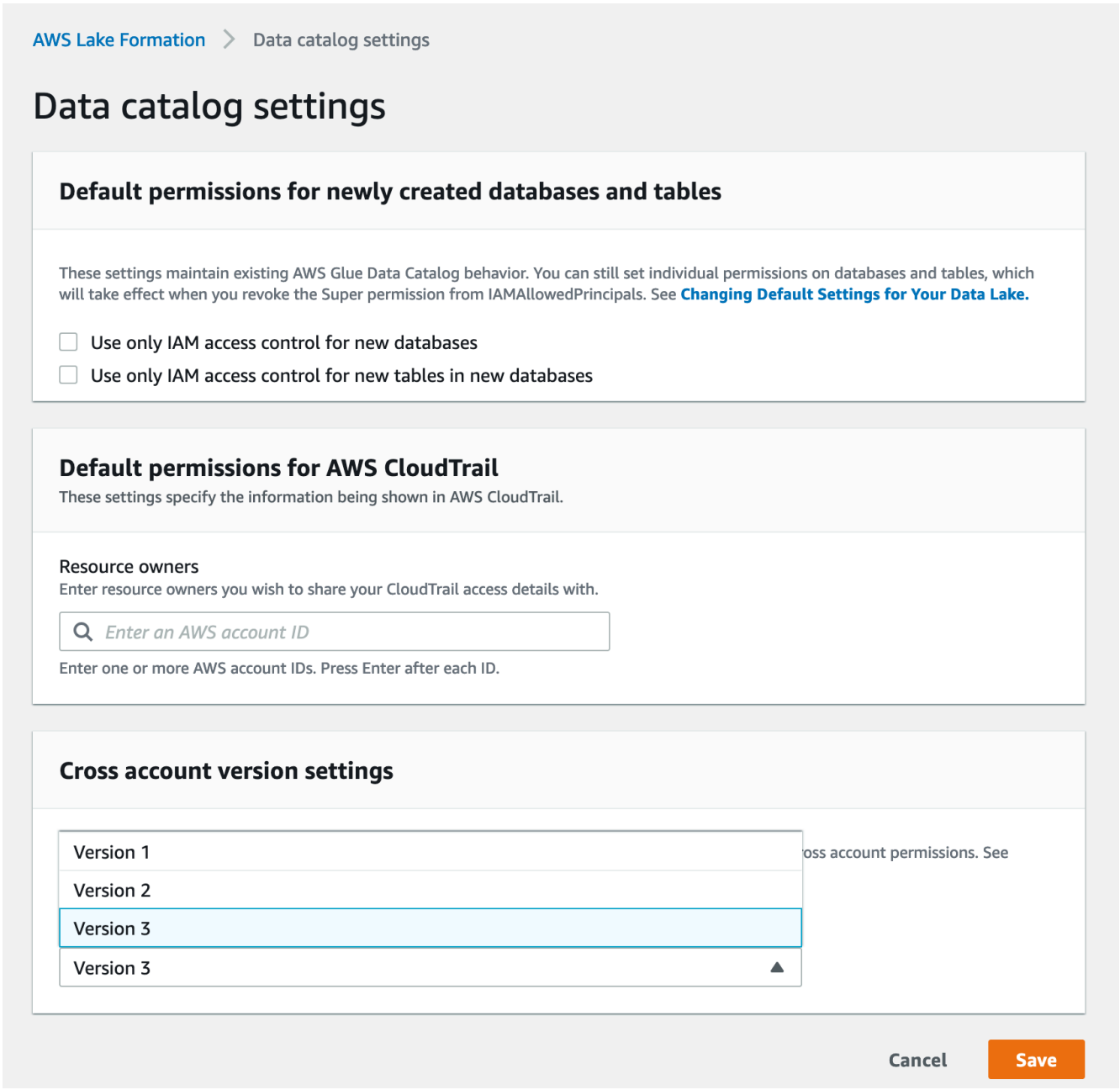
```
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
         "ram:AssociateResourceShare",
         "ram:DisassociateResourceShare",
         "ram:GetResourceShares"
       ],
     "Resource": "*",
     "Condition": {
       "StringLike": {
         "ram:ResourceShareName": "LakeFormation*"
        }
       }
    }
  ]
}
```

Pour activer la nouvelle version

Procédez comme suit pour mettre à jour les paramètres de version de plusieurs comptes via la AWS Lake Formation console ou le AWS CLI.

Console

1. Choisissez la version 2, la version 3 ou la version 4 dans les paramètres de version multi-comptes sur la page des paramètres du catalogue de données. Si vous sélectionnez la version 1, Lake Formation utilisera le mode de partage des ressources par défaut.



2. Choisissez Enregistrer.

AWS Command Line Interface (AWS CLI)

Utilisez la `put-data-lake-settings` AWS CLI commande pour définir le `CROSS_ACCOUNT_VERSION` paramètre. Les valeurs acceptées sont 1, 2, 3 et 4.

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
 file://settings
{
```

```
    "DataLakeAdmins": [
        {
            "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/test"
        }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "Parameters": {
        "CROSS_ACCOUNT_VERSION": "3"
    }
}
```

Important

Une fois que vous aurez choisi la version 2 ou la version 3, toutes les nouvelles subventions de ressources nommées passeront par le nouveau mode d'attribution entre comptes. Pour utiliser de manière optimale la AWS RAM capacité de vos partages entre comptes existants, nous vous recommandons de révoquer les autorisations accordées avec l'ancienne version et de les réattribuer dans le nouveau mode.

Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des responsables IAM

Cette section contient des instructions sur la façon d'activer les autorisations entre comptes sur les tables et les bases de données du catalogue de données pour un AWS compte externe, un responsable IAM, une organisation ou une unité organisationnelle. L'opération de subvention partage automatiquement ces ressources.

Rubriques

- Partage de données à l'aide du contrôle d'accès basé sur des balises
- Partage de données entre comptes à l'aide de la méthode des ressources nommées

Partage de données à l'aide du contrôle d'accès basé sur des balises

Configuration requise sur le compte du producteur/concédant

1. Définissez une balise LF. Pour obtenir des instructions sur la création d'une balise LF, consultez. Création de balises LF
2. Attribuez le tag LF à la ressource cible. Pour plus d'informations, consultez Affectation de balises LF aux ressources du catalogue de données.
3. Accordez l'autorisation LF-Tag au compte externe. Pour plus d'informations, consultez Octroi d'autorisations LF-Tag à l'aide de la console.

   À ce stade, l'administrateur du lac de données des consommateurs devrait être en mesure de trouver le tag de politique partagé via la console Lake Formation du compte bénéficiaire, sous Autorisations, rôles et tâches administratifs, balises LF.
4. Accordez l'autorisation des données au compte externe/bénéficiaire.
   a. Dans le volet de navigation, sous Autorisations, Autorisations du lac de données, sélectionnez Grant.
   b. Pour Principaux, choisissez External accounts, puis entrez l' Compte AWS ID cible ou le rôle IAM du principal ou le Amazon Resource Name (ARN) pour le principal (ARN principal).
   c. Pour les balises LF ou les ressources du catalogue, choisissez la clé et les valeurs de la balise LF partagée avec le compte du consommateur (clé **`Confidentiality`** et valeur). `public`
   d. Pour les autorisations, sous Ressources associées à des balises LF (recommandé), choisissez Ajouter une balise LF.
   e. Sélectionnez la clé et la valeur du tag partagé avec le compte du bénéficiaire (clé `Confidentiality` et valeur`public`).
   f. Pour les autorisations de base de données, sélectionnez Décrire sous Autorisations de base de données pour accorder des autorisations d'accès au niveau de la base de données.
   g. L'administrateur du lac de données du consommateur doit être en mesure de trouver le tag de politique partagé via le compte du consommateur sur la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/, sous Autorisations, rôles et tâches administratifs, balises LF.
   h. Sélectionnez Décrire sous Autorisations pouvant être accordées afin que le compte client puisse accorder des autorisations au niveau de la base de données à ses utilisateurs.

Étant donné que l'administrateur du lac de données doit accorder des autorisations sur les ressources partagées aux principaux du compte bénéficiaire, les autorisations entre comptes doivent toujours être accordées avec l'option d'octroi.

> **Note**
>
> Les directeurs qui reçoivent des subventions directes entre comptes n'auront pas l'option Autorisations accordables.

i. Pour les autorisations de table et de colonne, sélectionnez Sélectionner et décrire sous Autorisations de table.

j. Sélectionnez Sélectionner et décrire sous Autorisations pouvant être accordées.

k. Choisissez Grant (Accorder).

Configuration requise sur le compte bénéficiaire/bénéficiaire

1. Lorsque vous partagez une ressource avec un autre compte, elle appartient toujours au compte du producteur et n'est pas visible dans la console Athena. Pour que la ressource soit visible dans la console Athena, vous devez créer un lien de ressource pointant vers la ressource partagée. Pour obtenir des instructions sur la création d'un lien vers une ressource, consultez Création d'un lien de ressource vers une table de catalogue de données partagée et Création d'un lien de ressource vers une base de données de catalogue de données partagée
2. Vous devez créer un ensemble distinct de balises LF dans le compte client pour utiliser le contrôle d'accès basé sur les balises LF lors du partage des liens de ressources. Créez et attribuez les balises LF requises à la base de données/aux tables partagées et aux liens vers les ressources.
3. Accordez des autorisations sur ces balises LF aux principaux IAM du compte du bénéficiaire.

Partage de données entre comptes à l'aide de la méthode des ressources nommées

Vous pouvez accorder des autorisations directement aux principaux d'un autre AWS compte, ou à un compte externe Comptes AWS ou AWS Organizations. Accorder des autorisations de Lake Formation à des organisations ou à des unités organisationnelles revient à accorder l'autorisation Compte AWS à tous les membres de cette organisation ou unité organisationnelle.

Lorsque vous accordez des autorisations à des comptes ou à des organisations externes, vous devez inclure l'option Autorisations pouvant être accordées. Seul l'administrateur du lac de données

du compte externe peut accéder aux ressources partagées jusqu'à ce qu'il accorde des autorisations sur les ressources partagées aux autres principaux du compte externe.

> **Note**
>
> L'option d'autorisations pouvant être accordées n'est pas prise en charge lors de l'octroi d'autorisations directement aux principaux IAM à partir de comptes externes.

Suivez les instructions Octroi d'autorisations de base de données à l'aide de la méthode de ressource pour accorder des autorisations entre comptes à l'aide de la méthode de ressource nommée.

Octroi d'autorisations sur une base de données ou une table partagée avec votre compte

Une fois qu'une ressource de catalogue de données appartenant à un autre AWS compte est partagée avec votre AWS compte, en tant qu'administrateur du lac de données, vous pouvez accorder des autorisations sur la ressource partagée aux autres principaux de votre compte. Vous ne pouvez toutefois pas accorder d'autorisations sur la ressource à d'autres AWS comptes ou organisations.

Vous pouvez utiliser la AWS Lake Formation console, l'API ou le AWS Command Line Interface (AWS CLI) pour accorder les autorisations.

Pour accorder des autorisations sur une base de données partagée (méthode de ressource nommée, console)

- Suivez les instructions de la section Octroi d'autorisations de base de données à l'aide de la méthode de ressource. Dans la liste des bases de données, sous Tags LF ou ressources du catalogue, assurez-vous de sélectionner la base de données dans le compte externe, et non un lien de ressource pour la base de données.

  Si la base de données ne figure pas dans la liste des bases de données, assurez-vous d'avoir accepté l'invitation de partage de ressources AWS Resource Access Manager (AWS RAM) pour la base de données. Pour plus d’informations, consultez Acceptation d'une invitation de partage de ressources de AWS RAM.

De plus, pour les `ALTER` autorisations `CREATE_TABLE` et, suivez les Octroi d'autorisations de localisation des données (même compte) instructions fournies et assurez-vous de saisir l'identifiant du compte propriétaire dans le champ Emplacement du compte enregistré.

Pour accorder des autorisations sur une table partagée (méthode de ressource nommée, console)

- Suivez les instructions de la section Octroi d'autorisations de table à l'aide de la méthode de ressource nommée. Dans la liste des bases de données, sous Tags LF ou ressources du catalogue, assurez-vous de sélectionner la base de données dans le compte externe, et non un lien de ressource pour la base de données.

  Si le tableau ne figure pas dans la liste des tableaux, assurez-vous d'avoir accepté l'invitation de partage de AWS RAM ressources pour le tableau. Pour plus d'informations, consultez Acceptation d'une invitation de partage de ressources de AWS RAM.

  De plus, pour `ALTER` obtenir l'autorisation, suivez les Octroi d'autorisations de localisation des données (même compte) instructions fournies et assurez-vous de saisir l'identifiant du compte propriétaire dans le champ Emplacement du compte enregistré.

Pour accorder des autorisations sur des ressources partagées (méthode LF-TBAC, console)

- Suivez les instructions de la section Octroi d'autorisations au catalogue de données . Dans la section Balises LF ou ressources du catalogue, accordez l'expression de balise LF exacte que le compte externe a accordée à votre compte, ou un sous-ensemble de cette expression.

  Par exemple, si un compte externe a accordé l'expression LF-tag `module=customers AND environment=production` à votre compte avec l'option d'attribution, en tant qu'administrateur du lac de données, vous pouvez accorder cette même expression, `module=customers` ou `environment=production` à un mandant de votre compte. Vous ne pouvez accorder que les mêmes autorisations ou un sous-ensemble des autorisations de Lake Formation (par exemple,`SELECT`,`ALTER`, etc.) qui ont été accordées aux ressources via l'expression LF-Tag.

Pour accorder des autorisations sur une table partagée (méthode de ressource nommée, AWS CLI)

- Utilisez une commande similaire à la suivante. Dans cet exemple :

  - L'identifiant de votre AWS compte est 1111-2222-3333.

- Le compte propriétaire de la table et qui l'a attribuée à votre compte est le 1234-5678-9012.
- L'`SELECT`autorisation est accordée à l'utilisateur sur la table `pageviews` partagée`datalake_user1`. Cet utilisateur est le principal de votre compte.
- La `pageviews` table se trouve dans la `analytics` base de données, qui appartient au compte 1234-5678-9012.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"CatalogId":"123456789012",
 "DatabaseName":"analytics", "Name":"pageviews"}}'
```

Notez que le compte propriétaire doit être spécifié dans la `CatalogId` propriété de l'`resource`argument.

Octroi d'autorisations relatives aux liens vers

Procédez comme suit pour accorder AWS Lake Formation des autorisations sur un ou plusieurs liens de ressources à un responsable de votre AWS compte.

Une fois que vous avez créé un lien vers une ressource, vous êtes le seul à pouvoir le consulter et y accéder. (Cela suppose que l'option Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données n'est pas activée pour la base de données.) Pour permettre aux autres utilisateurs de votre compte d'accéder au lien de la ressource, accordez au moins l'`DESCRIBE`autorisation.

> **Important**
>
> L'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la base de données ou la table cible (liée). Vous devez accorder des autorisations à la cible séparément.

Vous pouvez accorder des autorisations à l'aide de la console Lake Formation, de l'API ou du AWS Command Line Interface (AWS CLI).

console

Pour accorder des autorisations relatives aux liens vers des ressources à l'aide de la console Lake Formation

1. Effectuez l’une des actions suivantes :

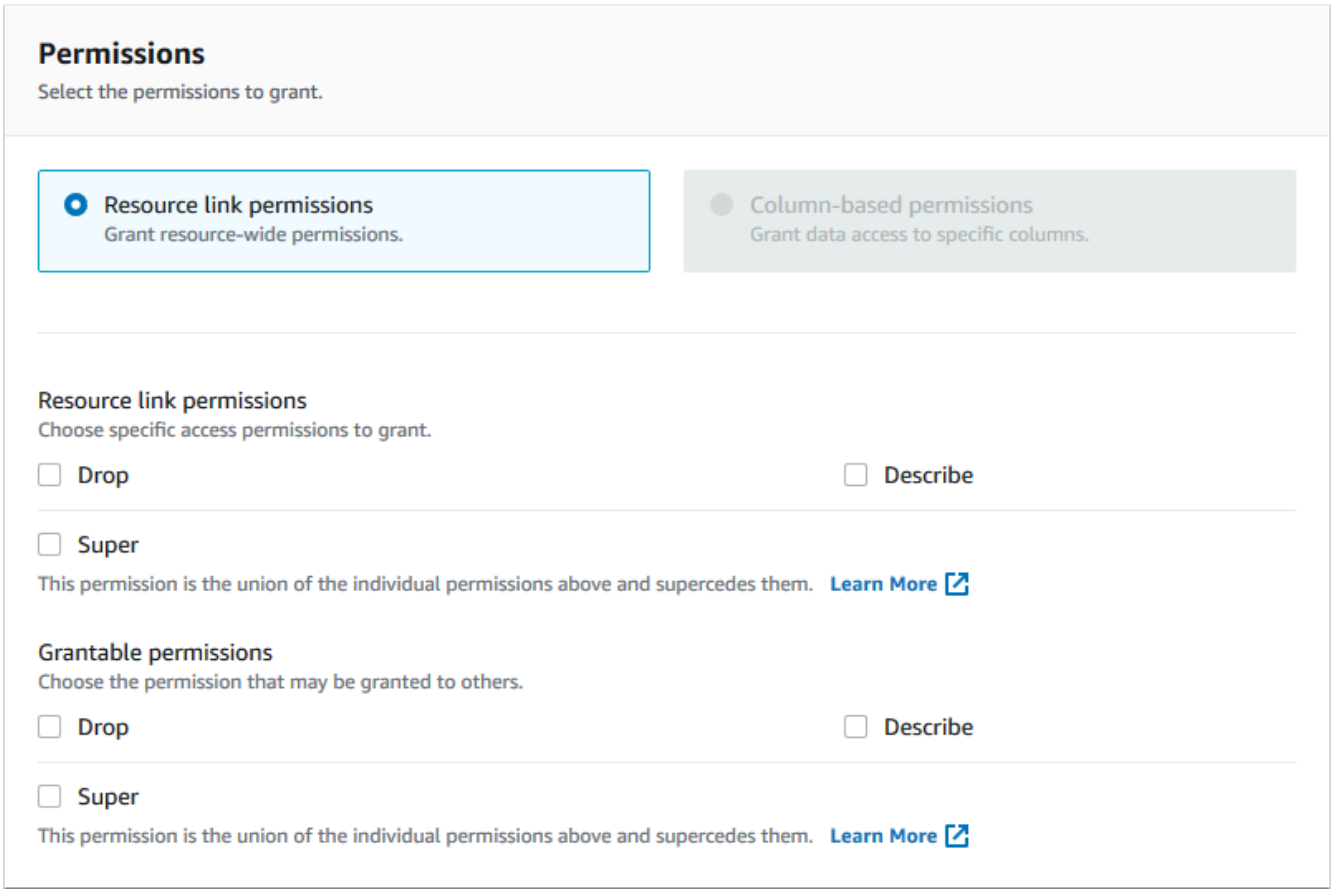
   - Pour les liens vers des ressources de base de données, suivez les étapes Octroi d'autorisations de base de données à l'aide de la méthode de ressource décrites dans. pour effectuer les opérations suivantes :

     1. Ouvrez la page des autorisations du lac de données Grant.
     2. Spécifiez les bases de données. Spécifiez un ou plusieurs liens de ressources de base de données.
     3. Spécifiez les principes.

   - Pour les liens vers les ressources des tables, suivez Octroi d'autorisations de table à l'aide de la méthode de ressource nommée les étapes décrites ci-dessous :

     1. Ouvrez la page des autorisations du lac de données Grant.
     2. Spécifiez les tables. Spécifiez un ou plusieurs liens de ressources de table.
     3. Spécifiez les principes.

2. Sous Autorisations, sélectionnez les autorisations à accorder. Sélectionnez éventuellement les autorisations pouvant être accordées.



3. Choisissez Grant (Accorder).

AWS CLI

Pour accorder des autorisations de lien vers des ressources à l'aide de AWS CLI

- Exécutez la `grant-permissions` commande en spécifiant un lien de ressource comme ressource.

  Example

  Cet exemple accorde `DESCRIBE` à l'utilisateur `datalake_user1` sur la table un lien de ressource `incidents-link` dans la base de données du AWS compte `issues` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"issues",
 "Name":"incidents-link"}}'
```

Voir aussi :

- Création de liens vers des ressources
- Référence des autorisations de Lake Formation

Accès aux données sous-jacentes d'une table partagée

Supposons que le AWS compte A partage une table du catalogue de données SELECT avec le compte B, par exemple, en octroyant au compte B. Pour que le principal du compte B puisse lire les données sous-jacentes de la table partagée, les conditions suivantes doivent être remplies :

- L'administrateur du lac de données du compte B doit accepter le partage. (Cela n'est pas nécessaire si les comptes A et B appartiennent à la même organisation ou si la subvention a été accordée selon la méthode de contrôle d'accès basée sur les balises Lake Formation.)
- L'administrateur du lac de données doit réaccorder au principal l'SELECTautorisation Lake Formation accordée par le compte A sur la table partagée.
- Le principal doit disposer des autorisations IAM suivantes sur la table, la base de données qui la contient et le compte A Data Catalog.

Note

Dans la politique IAM suivante :

- Remplacez <account-id-A>par le AWS numéro de compte du compte A.
- Remplacez <region>par une région valide.
- Remplacez <database>par le nom de la base de données du compte A qui contient la table partagée.
- Remplacez <table>par le nom de la table partagée.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
          "Effect": "Allow",
          "Action": [
```

```
            "glue:GetTable",
            "glue:GetTables",
            "glue:GetPartition",
            "glue:GetPartitions",
            "glue:BatchGetPartition",
            "glue:GetDatabase",
            "glue:GetDatabases"
           ],
           "Resource": [
            "arn:aws:glue:<region>:<account-id-A>:table/<database>/<table>",
            "arn:aws:glue:<region>:<account-id-A>:database/<database>",
            "arn:aws:glue:<region>:<account-id-A>:catalog"
           ]
        },
        {
          "Effect": "Allow",
          "Action": [
            "lakeformation:GetDataAccess"
           ],
          "Resource": [
            "*"
           ],
          "Condition": {
            "StringEquals": {
              "lakeformation:GlueARN":"arn:aws:glue:<region>:<account-id-
A>:table/<database>/<table>"
            }
        }
    }
   ]
}
```

Voir aussi :

- Acceptation d'une invitation de partage de ressources de AWS RAM

Journalisation entre comptes CloudTrail

Lake Formation fournit une piste d'audit centralisée de tous les accès entre comptes aux données de votre lac de données. Lorsqu'un AWS compte destinataire accède aux données d'une table partagée,

Lake Formation copie l' CloudTrail événement dans les CloudTrail journaux du compte propriétaire. Les événements copiés incluent des requêtes portant sur les données par des services intégrés tels qu' Amazon Athena Amazon Redshift Spectrum, et des accès aux AWS Glue données par des tâches.

CloudTrail les événements relatifs aux opérations entre comptes sur les ressources du catalogue de données sont copiés de la même manière.

En tant que propriétaire de ressources, si vous activez la journalisation au niveau des objets dans Amazon S3, vous pouvez exécuter des requêtes associant des événements S3 à CloudTrail des événements Lake Formation CloudTrail afin de déterminer les comptes qui ont accédé à vos compartiments S3.

Rubriques

- Inclure les identités principales dans les journaux intercomptes CloudTrail
- Consultation des CloudTrail journaux pour l'accès entre comptes Amazon S3

Inclure les identités principales dans les journaux intercomptes CloudTrail

Par défaut, les CloudTrail événements intercomptes ajoutés aux journaux du destinataire de la ressource partagée et copiés dans les journaux du propriétaire de la ressource contiennent uniquement l'identifiant AWS principal du compte externe, et non le nom de ressource Amazon (ARN) lisible par l'homme du principal (ARN principal). Lorsque vous partagez des ressources dans des limites fiables, par exemple au sein d'une même organisation ou d'une même équipe, vous pouvez choisir d'inclure l'ARN principal dans les CloudTrail événements. Les comptes propriétaires des ressources peuvent ensuite suivre les principaux des comptes des destinataires qui accèdent à leurs propres ressources.

> **Important**
>
> En tant que destinataire de ressources partagées, pour voir l'ARN principal dans les événements de vos propres CloudTrail journaux, vous devez choisir de partager l'ARN principal avec le compte du propriétaire.
> Si l'accès aux données se fait via un lien de ressource, deux événements sont enregistrés dans le compte du destinataire de la ressource partagée : un pour l'accès au lien de ressource et un pour l'accès à la ressource cible. L'événement pour l'accès au lien de ressource inclut l'ARN principal. L'événement pour l'accès à la ressource cible n'inclut pas

l'ARN principal sans l'opt-in. L'événement d'accès au lien de ressource n'est pas copié sur le compte du propriétaire.

Ce qui suit est un extrait d'un CloudTrail événement multicompte par défaut (sans opt-in). Le compte effectuant l'accès aux données est le 1111-2222-3333. Il s'agit du journal affiché à la fois dans le compte d'appel et dans le compte du propriétaire de la ressource. Lake Formation remplit les journaux dans les deux comptes dans le cas de comptes croisés.

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```

En tant que consommateur de ressources partagées, lorsque vous choisissez d'inclure l'ARN principal, l'extrait devient le suivant. Le `lakeFormationPrincipal` champ représente le rôle final ou l'utilisateur exécutant la requête via Amazon Athena, Amazon Redshift Spectrum ou des jobs. AWS Glue

```
{
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
```

```
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```

Pour choisir d'inclure les principaux ARN dans les journaux entre comptes CloudTrail

1. Ouvrez la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'`Administrator`utilisateur ou en tant qu'utilisateur avec la politique `Administrator Access` IAM.
2. Dans le panneau de navigation, sélectionnez Settings (Paramètres).
3. Sur la page des paramètres du catalogue de données, dans la AWS CloudTrail section Autorisations par défaut pour, pour les propriétaires de ressources, entrez un ou plusieurs ID de compte de propriétaire de AWS ressources.

   Appuyez sur Entrée après chaque identifiant de compte.
4. Choisissez Enregistrer.

   Désormais, les CloudTrail événements entre comptes stockés dans les journaux du destinataire de la ressource partagée et du propriétaire de la ressource contiennent l'ARN principal.

Consultation des CloudTrail journaux pour l'accès entre comptes Amazon S3

En tant que propriétaire de ressources partagées, vous pouvez interroger les CloudTrail journaux S3 pour déterminer les comptes qui ont accédé à vos compartiments Amazon S3 (à condition que vous ayez activé la journalisation au niveau des objets dans Amazon S3). Cela s'applique uniquement aux sites S3 que vous avez enregistrés auprès de Lake Formation. Si les consommateurs de ressources partagées choisissent d'inclure les principaux Rans dans les CloudTrail journaux de Lake Formation, vous pouvez déterminer les rôles ou les utilisateurs qui ont accédé aux compartiments.

Lorsque vous exécutez des requêtes avec Amazon Athena, vous pouvez joindre les CloudTrail événements Lake Formation et les CloudTrail événements S3 sur la propriété du nom de session. Les requêtes peuvent également filtrer les événements Lake Formation

sureventName="GetDataAccess", et les événements S3 sur eventName="Get Object" oueventName="Put Object".

Ce qui suit est un extrait d'un CloudTrail événement inter-comptes de Lake Formation au cours duquel les données d'un emplacement S3 enregistré ont été consultées.

```
{
  "eventSource": "lakeformation.amazonaws.com",
  "eventName": "GetDataAccess",
  ..............
  ..............
  "additionalEventData": {
    "requesterService": "GLUE_JOB",
    "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
    "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-B8JSAjo5QA"
   }
}
```

La valeur de la lakeFormationRoleSessionName AWSLF-00-GL-111122223333-B8JSAjo5QA clé peut être jointe au nom de session dans la principalId clé de l' CloudTrail événement S3. Ce qui suit est un extrait de l' CloudTrail événement S3. Il indique l'emplacement du nom de session.

```
{
   "eventSource": "s3.amazonaws.com",
   "eventName": "Get Object"
   ..............
   ..............
   "principalId": "AROAQSOX5XXUR7D6RMYLR:AWSLF-00-GL-111122223333-B8JSAjo5QA",
   "arn": "arn:aws:sets::111122223333:assumed-role/Deformationally/AWSLF-00-
GL-111122223333-B8JSAjo5QA",
   "session Context": {
     "session Issuer": {
       "type": "Role",
       "principalId": "AROAQSOX5XXUR7D6RMYLR",
       "arn": "arn:aws:iam::111122223333:role/aws-service-role/
lakeformation.amazonaws.com/Deformationally",
       "accountId": "111122223333",
       "user Name": "Deformationally"
     },
   ..............
   ..............
```

```
}
```

Le nom de session est formaté comme suit :

```
AWSLF-<version-number>-<query-engine-code>-<account-id->-<suffix>
```

`version-number`

Version de ce format, actuellement`00`. Si le format du nom de session change, la prochaine version sera la suivante`01`.

`query-engine-code`

Indique l'entité qui a accédé aux données. Les valeurs actuelles sont les suivantes :

`GL`	AWS GlueTâche ETL
`AT`	Athena
`RE`	Amazon Redshift Spectrum

`account-id`

L'identifiant du AWS compte qui a demandé les informations d'identification à Lake Formation.

`suffix`

Chaîne générée aléatoirement.

Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation

Il est possible d'accorder un accès entre comptes aux ressources du catalogue de données et aux données sous-jacentes en utilisant l'un AWS Glue ou AWS Lake Formation l'autre des deux.

DansAWS Glue, vous accordez des autorisations entre comptes en créant ou en mettant à jour une politique de ressources du catalogue de données. Dans Lake Formation, vous accordez des autorisations entre comptes en utilisant le modèle d'`GRANT/REVOKE`autorisations Lake Formation et le fonctionnement de l'`Grant Permissions`API.

> **Tip**
>
> Nous vous recommandons de vous fier uniquement aux autorisations de Lake Formation pour sécuriser votre lac de données.

Vous pouvez consulter les subventions multicomptes de Lake Formation à l'aide de la console Lake Formation ou de la console AWS Resource Access Manager (AWS RAM). Toutefois, ces pages de console n'affichent pas les autorisations entre comptes accordées par la politique de ressources du catalogue de AWS Glue données. De même, vous pouvez consulter les autorisations entre comptes dans la politique de ressources du catalogue de données à l'aide de la page Paramètres de la AWS Glue console, mais cette page n'affiche pas les autorisations entre comptes accordées via Lake Formation.

Pour vous assurer de ne manquer aucune subvention lorsque vous consultez et gérez les autorisations entre comptes, Lake Formation vous AWS Glue demande d'effectuer les actions suivantes pour indiquer que vous êtes au courant et que vous autorisez les subventions croisées par Lake Formation et. AWS Glue

Lorsque vous accordez des autorisations entre comptes à l'aide de la politique de ressources du catalogue de AWS Glue données

Si votre compte (compte du donateur ou compte du producteur) n'a accordé aucune subvention entre comptes destinée AWS RAM à partager les ressources, vous pouvez enregistrer une politique de ressources du catalogue de données comme d'habitude dans. AWS Glue Toutefois, si des subventions impliquant AWS RAM des partages de ressources ont déjà été accordées, vous devez effectuer l'une des opérations suivantes pour garantir le succès de l'enregistrement de la politique de ressources :

- Lorsque vous enregistrez la politique de ressources sur la page Paramètres de la AWS Glue console, celle-ci émet une alerte indiquant que les autorisations définies dans la politique s'ajouteront à celles accordées à l'aide de la console Lake Formation. Vous devez choisir Proceed pour enregistrer la politique.
- Lorsque vous enregistrez la politique de ressources à l'aide de l'opération `glue:PutResourcePolicy` API, vous devez définir le `EnableHybrid` champ sur `TRUE` « » (type = chaîne). L'exemple de code suivant montre comment procéder en Python.

```
import boto3
```

```
import json

REGION = 'us-east-2'
PRODUCER_ACCOUNT_ID = '123456789012'
CONSUMER_ACCOUNT_IDs = ['111122223333']

glue = glue_client = boto3.client('glue')

policy = {
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Cataloguers",
            "Effect": "Allow",
            "Action": [
                "glue:*"
            ],
            "Principal": {
                "AWS": CONSUMER_ACCOUNT_IDs
            },
            "Resource": [
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:catalog",
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:database/*",
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:table/*/*"
            ]
        }
    ]
}

policy = json.dumps(policy)
glue.put_resource_policy(PolicyInJson=policy, EnableHybrid='TRUE')
```

Pour plus d'informations, consultez PutResourcePolicy Action (Python : put_resource_policy) dans le manuel du développeur.AWS Glue

Lorsque vous accordez des autorisations entre comptes à l'aide de la méthode des ressources nommées de Lake Formation

S'il n'y a aucune politique de ressources du catalogue de données sur votre compte (compte producteur), les subventions croisées de Lake Formation que vous accordez se poursuivent comme d'habitude. Toutefois, s'il existe une politique de ressources pour le catalogue de données, vous devez y ajouter l'instruction suivante pour permettre à vos subventions entre comptes de réussir si

elles sont accordées avec la méthode de ressource nommée. <region>Remplacez-le par un nom de région valide et <account-id>par votre numéro de AWS compte (numéro de compte producteur).

```
{
  "Effect": "Allow",
  "Action": [
    "glue:ShareResource"
  ],
  "Principal": {"Service": [
    "ram.amazonaws.com"
  ]},
  "Resource": [
    "arn:aws:glue:<region>:<account-id>:table/*/*",
    "arn:aws:glue:<region>:<account-id>:database/*",
    "arn:aws:glue:<region>:<account-id>:catalog"
  ]
}
```

Sans cette déclaration supplémentaire, la subvention Lake Formation est acceptée, mais elle est bloquée et le compte du bénéficiaire ne peut pas accéder à la ressource accordée. AWS RAM

Important

Lorsque vous utilisez la méthode de contrôle d'accès basé sur les balises Lake Formation (LF-TBAC) pour accorder des autorisations entre comptes, vous devez disposer d'une politique de ressources du catalogue de données avec au moins les autorisations spécifiées dans. Prérequis

Voir aussi :

- Contrôle d'accès aux métadonnées(pour une discussion sur la méthode des ressources nommées par rapport à la méthode de contrôle d'accès basée sur les balises Lake Formation (LF-TBAC)).
- Affichage des tables et des bases de données partagées du catalogue de données
- Utilisation des paramètres du catalogue de données sur la AWS Glue console dans le manuel du AWS Glue développeur

- Octroi d'un accès entre comptes dans le guide du AWS Glue développeur (pour des exemples de politiques relatives aux ressources du catalogue de données)

Afficher toutes les subventions entre comptes à l'aide de l'opération GetResourceShares API

Si votre entreprise accorde des autorisations entre comptes en utilisant à la fois une politique de AWS Glue Data Catalog ressources et des subventions de Lake Formation, le seul moyen de consulter toutes les autorisations entre comptes en un seul endroit est d'utiliser l'opération `glue:GetResourceShares` API.

Lorsque vous accordez des autorisations à Lake Formation sur plusieurs comptes à l'aide de la méthode des ressources nommées, AWS Resource Access Manager (AWS RAM) crée une politique de ressources AWS Identity and Access Management (IAM) et l'enregistre dans votre AWS compte. La politique accorde les autorisations requises pour accéder à la ressource. AWS RAM crée une politique de ressources distincte pour chaque subvention intercomptes. Vous pouvez consulter toutes ces politiques à l'aide de l'opération `glue:GetResourceShares` API.

> **Note**
>
> Cette opération renvoie également la politique de ressources du catalogue de données. Toutefois, si vous avez activé le chiffrement des métadonnées dans les paramètres du catalogue de données et que vous n'êtes pas autorisé à utiliser la AWS KMS clé, l'opération ne renverra pas la politique de ressources du catalogue de données.

Pour voir toutes les subventions entre comptes

- Entrez la AWS CLI commande suivante.

```
aws glue get-resource-policies
```

Voici un exemple de politique de ressources qui AWS RAM crée et stocke lorsque vous accordez des autorisations sur une table `t` dans la base de données `db1` au AWS compte 1111-2222-3333.

```
{
```

```
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
         "glue:GetTable",
         "glue:GetTables",
         "glue:GetTableVersion",
         "glue:GetTableVersions",
         "glue:GetPartition",
         "glue:GetPartitions",
         "glue:BatchGetPartition",
         "glue:SearchTables"
       ],
      "Principal": {"AWS": [
        "111122223333"
      ]},
      "Resource": [
        "arn:aws:glue:<region>:111122223333:table/db1/t"
     ]
    }
  ]
}
```

Voir aussi :

- GetResourceShares Action (Python : get_resource_policies) dans le manuel du développeur AWS Glue

Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci

Pour l'administrateur du lac de données et pour les principaux auxquels des autorisations ont été accordées, les ressources partagées avec votre AWS compte apparaissent dans le catalogue de données comme s'il s'agissait de ressources de votre compte. La console affiche le compte propriétaire de la ressource.

Vous pouvez consulter les ressources partagées avec votre compte à l'aide de la console Lake Formation. Vous pouvez également utiliser la console AWS Resource Access Manager (AWS RAM)

pour afficher à la fois les ressources partagées avec votre compte et les ressources que vous avez partagées avec d'autres AWS comptes en utilisant la méthode des ressources nommées.

Important

Lorsqu'une personne utilise la méthode de ressource nommée pour accorder des autorisations entre comptes sur une ressource du catalogue de données à votre compte ou à votre AWS organisation, Lake Formation utilise le service AWS Resource Access Manager (AWS RAM) pour partager la ressource. Si votre compte appartient à la même AWS organisation que le compte donateur, la ressource partagée est immédiatement disponible. Toutefois, si votre compte n'appartient pas à la même organisation, AWS RAM envoie une invitation à votre compte pour accepter ou refuser le partage des ressources. Ensuite, pour rendre la ressource partagée disponible, l'administrateur du lac de données de votre compte doit utiliser la AWS RAM console ou la CLI pour accepter l'invitation.
La console Lake Formation affiche une alerte si une invitation de partage de AWS RAM ressources attend d'être acceptée. Seuls les utilisateurs autorisés à consulter les AWS RAM invitations reçoivent l'alerte.

Voir aussi :

- Partage des tables et des bases de données du catalogue de données entre les AWS comptes
- Partage de données entre comptes dans Lake Formation
- Accès aux données sous-jacentes d'une table partagée
- Contrôle d'accès aux métadonnées(pour plus d'informations sur la méthode des ressources nommées par rapport à la méthode LF-TBAC pour le partage des ressources.)

Rubriques

- Acceptation d'une invitation de partage de ressources de AWS RAM
- Affichage des tables et des bases de données partagées du catalogue de données

Acceptation d'une invitation de partage de ressources de AWS RAM

Si une ressource du catalogue de données est partagée avec votre AWS compte et que celui-ci n'appartient pas à la même AWS organisation que le compte de partage, vous n'avez pas accès à la ressource partagée tant que vous n'avez pas accepté une invitation de partage de ressources de AWS Resource Access Manager (AWS RAM). En tant qu'administrateur du lac de données, vous devez d'abord AWS RAM demander les invitations en attente, puis accepter l'invitation.

Vous pouvez utiliser la AWS RAM console, l'API ou AWS Command Line Interface (AWS CLI) pour consulter et accepter les invitations.

Pour consulter et accepter une invitation de partage de ressources depuis AWS RAM (console)

1. Assurez-vous de disposer des autorisations AWS Identity and Access Management (IAM) requises pour consulter et accepter les invitations de partage de ressources.

   Pour plus d'informations sur les politiques IAM suggérées pour les administrateurs de data lake, consultezthe section called “Autorisations d'administrateur du lac de données”.
2. Suivez les instructions de la section Acceptation et rejet des invitations du guide de l'AWS RAM utilisateur.

Pour consulter et accepter une invitation de partage de ressources depuis AWS RAM (AWS CLI)

1. Assurez-vous de disposer des autorisations AWS Identity and Access Management (IAM) requises pour consulter et accepter les invitations de partage de ressources.

   Pour plus d'informations sur les politiques IAM suggérées pour les administrateurs de data lake, consultezthe section called “Autorisations d'administrateur du lac de données”.
2. Entrez la commande suivante pour afficher les invitations de partage de ressources en attente.

   ```
   aws ram get-resource-share-invitations
   ```

   La sortie doit ressembler à ce qui suit.

   ```
   {
       "resourceShareInvitations": [
           {
   ```

```
            "resourceShareInvitationArn": "arn:aws:ram:us-
east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
a4e72eec1d9f",
            "resourceShareName": "111122223333-123456789012-uswuU",
            "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
            "senderAccountId": "111122223333",
            "receiverAccountId": "123456789012",
            "invitationTimestamp": 1589576601.79,
            "status": "PENDING"
        }
    ]
}
```

Notez le statut de`PENDING`.

3. Copiez la valeur de la `resourceShareInvitationArn` clé dans le presse-papiers.
4. Collez la valeur dans la commande suivante, en la remplaçant<invitation-arn>, puis entrez la commande.

```
aws ram accept-resource-share-invitation --resource-share-invitation-
arn <invitation-arn>
```

La sortie doit ressembler à ce qui suit.

```
{
    "resourceShareInvitations": [
        {
            "resourceShareInvitationArn": "arn:aws:ram:us-
east-1:111122223333:resource-share-invitation/a93aa60a-1bd9-46e8-96db-
a4e72eec1d9f",
            "resourceShareName": "111122223333-123456789012-uswuU",
            "resourceShareArn": "arn:aws:ram:us-east-1:111122223333:resource-
share/2a4ab5fb-d859-4751-84f7-8760b35fc1fe",
            "senderAccountId": "111122223333",
            "receiverAccountId": "123456789012",
            "invitationTimestamp": 1589576601.79,
            "status": "ACCEPTED"
        }
    ]
}
```

Notez le statut de`ACCEPTED`.

Affichage des tables et des bases de données partagées du catalogue de données

Vous pouvez consulter les ressources partagées avec votre compte à l'aide de la console Lake Formation ou de la AWS CLI. Vous pouvez également utiliser la console AWS Resource Access Manager (AWS RAM) ou la CLI pour afficher à la fois les ressources partagées avec votre compte et les ressources que vous avez partagées avec d'autres AWS comptes.

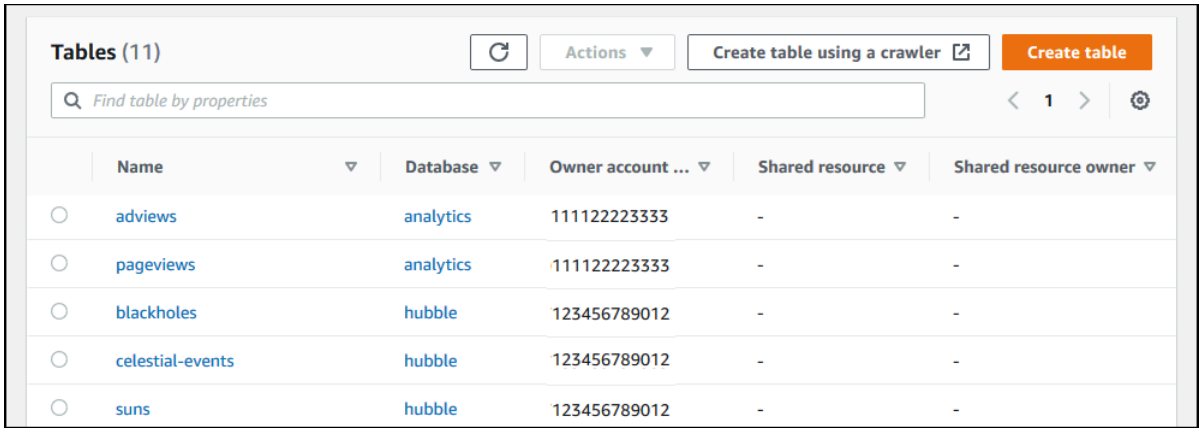
Pour afficher les ressources partagées à l'aide de la console Lake Formation

1. Ouvrez la console Lake Formation à l’adresse https://console.aws.amazon.com/lakeformation/.

   Connectez-vous en tant qu'administrateur du lac de données ou en tant qu'utilisateur ayant obtenu des autorisations sur une table partagée.
2. Pour consulter les ressources partagées avec votre AWS compte, effectuez l'une des opérations suivantes :

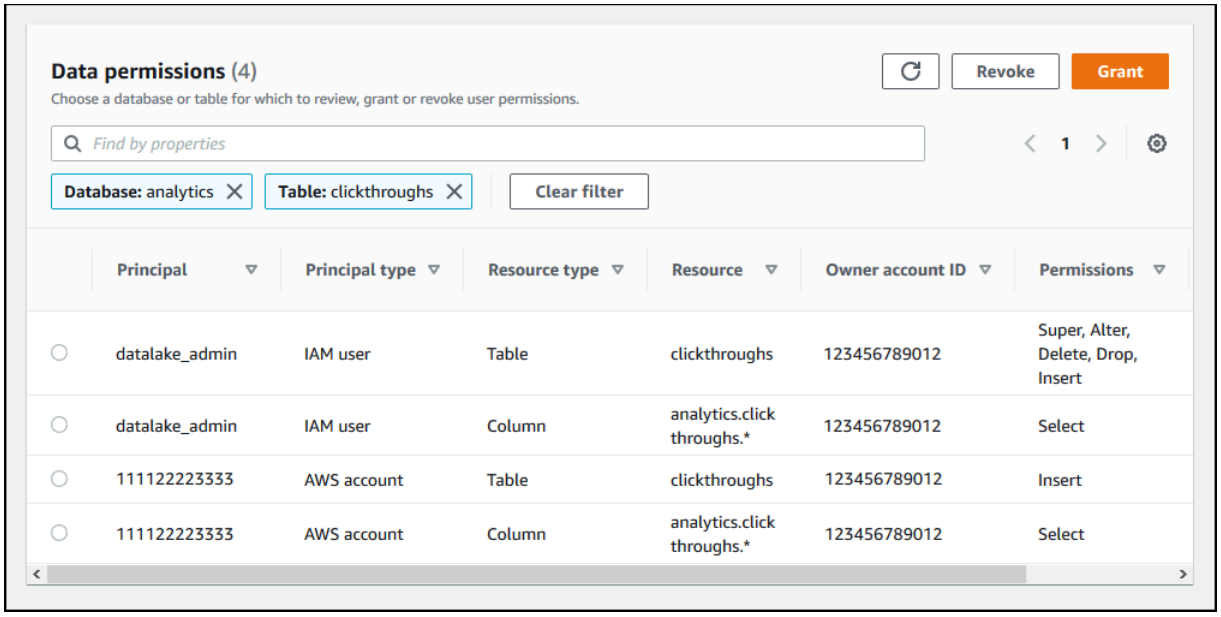
   - Pour afficher les tables partagées avec votre compte, dans le volet de navigation, sélectionnez Tables.
   - Pour afficher les bases de données partagées avec votre compte, dans le volet de navigation, sélectionnez Bases de données.

   La console affiche une liste de bases de données ou de tables à la fois présentes dans votre compte et partagées avec votre compte. Pour les ressources partagées avec votre compte, la console affiche l'identifiant du AWS compte du propriétaire dans la colonne ID du compte du propriétaire (la troisième colonne de la capture d'écran suivante).



3. Pour afficher les ressources que vous avez partagées avec d'autres AWS comptes ou organisations, dans le volet de navigation, sélectionnez Autorisations relatives aux données.

   Les ressources que vous avez partagées sont répertoriées sur la page Autorisations relatives aux données avec le numéro de compte externe indiqué dans la colonne Principal, comme indiqué dans l'image suivante.



Pour afficher les ressources partagées à l'aide de la AWS RAM console

1. Assurez-vous de disposer des autorisations AWS Identity and Access Management (IAM) requises pour consulter les ressources partagées à l'aide AWS RAM de.

   Au minimum, vous devez avoir l'autorisation`ram:ListResources`. Cette autorisation est incluse dans la politique gérée par AWS. `AWSLakeFormationCrossAccountManager`

2. Connectez-vous à la AWS RAM console AWS Management Console et ouvrez-la à l'adresse https://console.aws.amazon.com/ram.
3. Effectuez l'une des actions suivantes :

   - Pour voir les ressources que vous avez partagées, dans le volet de navigation, sous Shared by me, sélectionnez Shared resources.
   - Pour voir les ressources partagées avec vous, dans le volet de navigation, sous Partagé avec moi, choisissez Ressources partagées.

Création de liens vers des ressources

Les liens de ressources sont des objets du catalogue de données qui sont des liens vers des bases de données et des tables de métadonnées, généralement vers des bases de données partagées et des tables d'autres AWS comptes. Ils aident à permettre l'accès entre comptes aux données du lac de données dans toutes les AWS régions.

> **Note**
>
> Lake Formation permet d'interroger les tables du catalogue de données dans toutes les AWS régions. Vous pouvez accéder aux bases de données et aux tables du catalogue de données depuis n'importe quelle AWS région en créant des liens de ressources dans ces régions qui pointent vers des bases de données et des tables partagées dans différentes régions.

Rubriques

- Mode de fonctionnement des liens des ressources dans Lake Formation
- Création d'un lien de ressource vers une table de catalogue de données partagée
- Création d'un lien de ressource vers une base de données de catalogue de données partagée
- Gestion des liens de ressources dans les AWS Glue API

Mode de fonctionnement des liens des ressources dans Lake Formation

Un lien de ressource est un objet du catalogue de données qui est un lien vers une base de données ou une table locale ou partagée. Après avoir créé un lien de ressource vers une base de données ou une table, vous pouvez utiliser le nom du lien de ressource partout où vous utiliseriez le nom de la base de données ou de la table. Outre les tables que vous possédez ou les tables partagées avec

vous, les liens vers les ressources des tables sont renvoyés par la page Tables de la console Lake Formation `glue:GetTables()` et apparaissent sous forme d'entrées sur celle-ci. Les liens vers les ressources vers les bases de données agissent de la même manière.

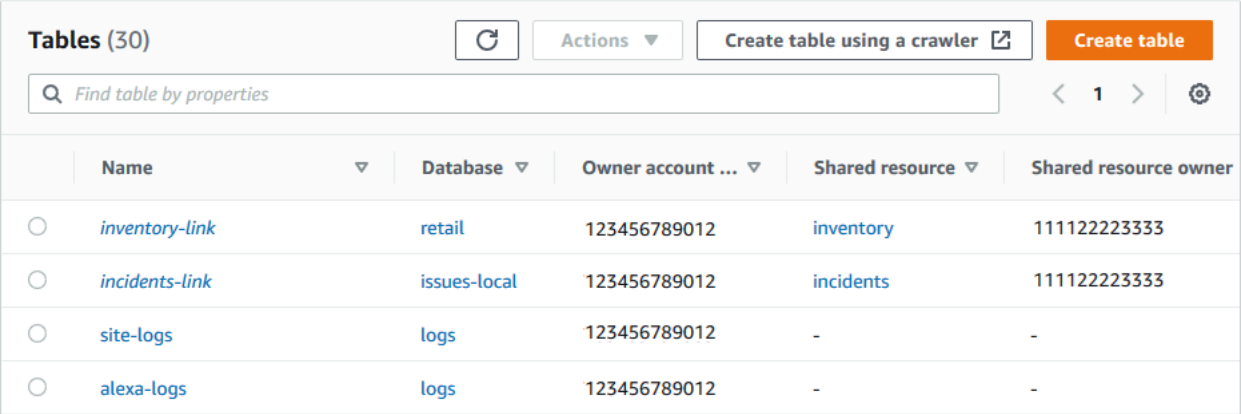
La création d'un lien de ressource vers une base de données ou une table vous permet d'effectuer les opérations suivantes :

- Attribuez un nom différent à une base de données ou à une table dans votre catalogue de données. Cela est particulièrement utile si différents AWS comptes partagent des bases de données ou des tables portant le même nom, ou si plusieurs bases de données de votre compte possèdent des tables portant le même nom.
- Accédez aux bases de données et aux tables du catalogue de données depuis n'importe quelle AWS région en créant des liens de ressources dans ces régions pointant vers la base de données et les tables d'une autre région. Vous pouvez exécuter des requêtes dans n'importe quelle région avec ces liens de ressources à l'aide d'Athena, Amazon EMR et exécuter des tâches AWS Glue ETL Spark, sans copier les données source ni les métadonnées dans Glue Data Catalog.
- Utilisez AWS des services intégrés tels qu' Amazon Athena Amazon Redshift Spectrum pour exécuter des requêtes qui accèdent à des bases de données ou à des tables partagées. Certains services intégrés ne peuvent pas accéder directement aux bases de données ou aux tables d'un compte à l'autre. Ils peuvent toutefois accéder aux liens de ressources de votre compte vers les bases de données et les tables d'autres comptes.

> **Note**
>
> Il n'est pas nécessaire de créer un lien de ressource pour référencer une base de données ou une table partagée dans les scripts AWS Glue d'extraction, de transformation et de chargement (ETL). Toutefois, pour éviter toute ambiguïté lorsque plusieurs AWS comptes partagent une base de données ou une table portant le même nom, vous pouvez soit créer et utiliser un lien de ressource, soit spécifier l'ID du catalogue lorsque vous appelez des opérations ETL.

L'exemple suivant montre la page Tables de la console Lake Formation, qui répertorie deux liens vers des ressources. Les noms des liens vers les ressources sont toujours affichés en italique. Chaque lien de ressource est affiché avec le nom et le propriétaire de la ressource partagée associée. Dans cet exemple, un administrateur de lac de données du AWS compte 1111-2222-3333 a partagé les

tables `inventory` et `incidents` avec le compte 1234-5678-9012. Un utilisateur de ce compte a ensuite créé des liens de ressources vers ces tables partagées.



Les remarques et restrictions relatives aux liens vers les ressources sont les suivantes :

- Des liens de ressources sont nécessaires pour permettre aux services intégrés tels qu'Athena et Redshift Spectrum d'interroger les données sous-jacentes des tables partagées. Les requêtes de ces services intégrés sont construites en fonction des noms des liens vers les ressources.
- En supposant que le paramètre Utiliser uniquement le contrôle d'accès IAM pour les nouvelles tables de cette base de données soit désactivé pour la base de données qui la contient, seul le principal qui a créé un lien de ressource peut le consulter et y accéder. Pour permettre aux autres utilisateurs de votre compte d'accéder à un lien vers une ressource, accordez-leur l'`DESCRIBE`autorisation correspondante. Pour permettre à d'autres personnes de supprimer un lien vers une ressource, `DROP` accordez-lui l'autorisation. Les administrateurs du data lake peuvent accéder à tous les liens vers les ressources du compte. Pour supprimer un lien de ressource créé par un autre principal, l'administrateur du lac de données doit d'abord s'accorder l'`DROP`autorisation sur le lien de ressource. Pour plus d’informations, consultez Référence des autorisations de Lake Formation.

  > **Important**
  >
  > L'octroi d'autorisations sur un lien de ressource n'accorde pas d'autorisations sur la base de données ou la table cible (liée). Vous devez accorder des autorisations à la cible séparément.

- Pour créer un lien vers une ressource, vous avez besoin de la Lake Formation `CREATE_TABLE` ou de l'`CREATE_DATABASE`autorisation `glue:CreateTable` ou `glue:CreateDatabase` AWS Identity and Access Management (IAM).

- Vous pouvez créer des liens vers des ressources locales (détenues) du catalogue de données, ainsi que vers des ressources partagées avec votre AWS compte.
- Lorsque vous créez un lien vers une ressource, aucune vérification n'est effectuée pour vérifier si la ressource partagée cible existe ou si vous disposez d'autorisations entre comptes sur la ressource. Cela vous permet de créer le lien vers la ressource et la ressource partagée dans n'importe quel ordre.
- Si vous supprimez un lien vers une ressource, la ressource partagée associée n'est pas supprimée. Si vous supprimez une ressource partagée, les liens vers cette ressource ne sont pas supprimés.
- Il est possible de créer des chaînes de liens entre les ressources. Cependant, cela n'a aucun intérêt, car les API ne suivent que le premier lien de ressource.

Voir aussi :

- Octroi et révocation d'autorisations sur les ressources du catalogue de données

Création d'un lien de ressource vers une table de catalogue de données partagée

Vous pouvez créer un lien de ressource vers une table partagée dans n'importe quelle AWS région à l'aide de la AWS Lake Formation console, de l'API ou AWS Command Line Interface (AWS CLI).

Pour créer un lien de ressource vers une table partagée (console)

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant que directeur `CREATE_TABLE` autorisé à accéder à la base de données pour contenir le lien vers la ressource dans la base de données.
2. Dans le panneau de navigation, choisissez Tables, puis Create table (Créer une table).
3. Sur la page Créer une table, choisissez la vignette Lien vers une ressource, puis fournissez les informations suivantes :

   Nom du lien vers la ressource

   Entrez un nom qui respecte les mêmes règles qu'un nom de table. Le nom peut être identique à celui de la table partagée cible.

Base de données

La base de données du catalogue de données local qui doit contenir le lien vers la ressource.

Région du propriétaire de la table partagée

Si vous créez le lien de ressource dans une autre région, sélectionnez la région de la table partagée cible.

Table partagée

Sélectionnez une table partagée dans la liste ou entrez un nom de table locale (détenue) ou partagée.

La liste contient toutes les tables partagées avec votre compte. Notez la base de données et l'ID du compte propriétaire répertoriés dans chaque table. Si vous ne voyez aucune table dont vous savez qu'elle a été partagée avec votre compte, vérifiez les points suivants :

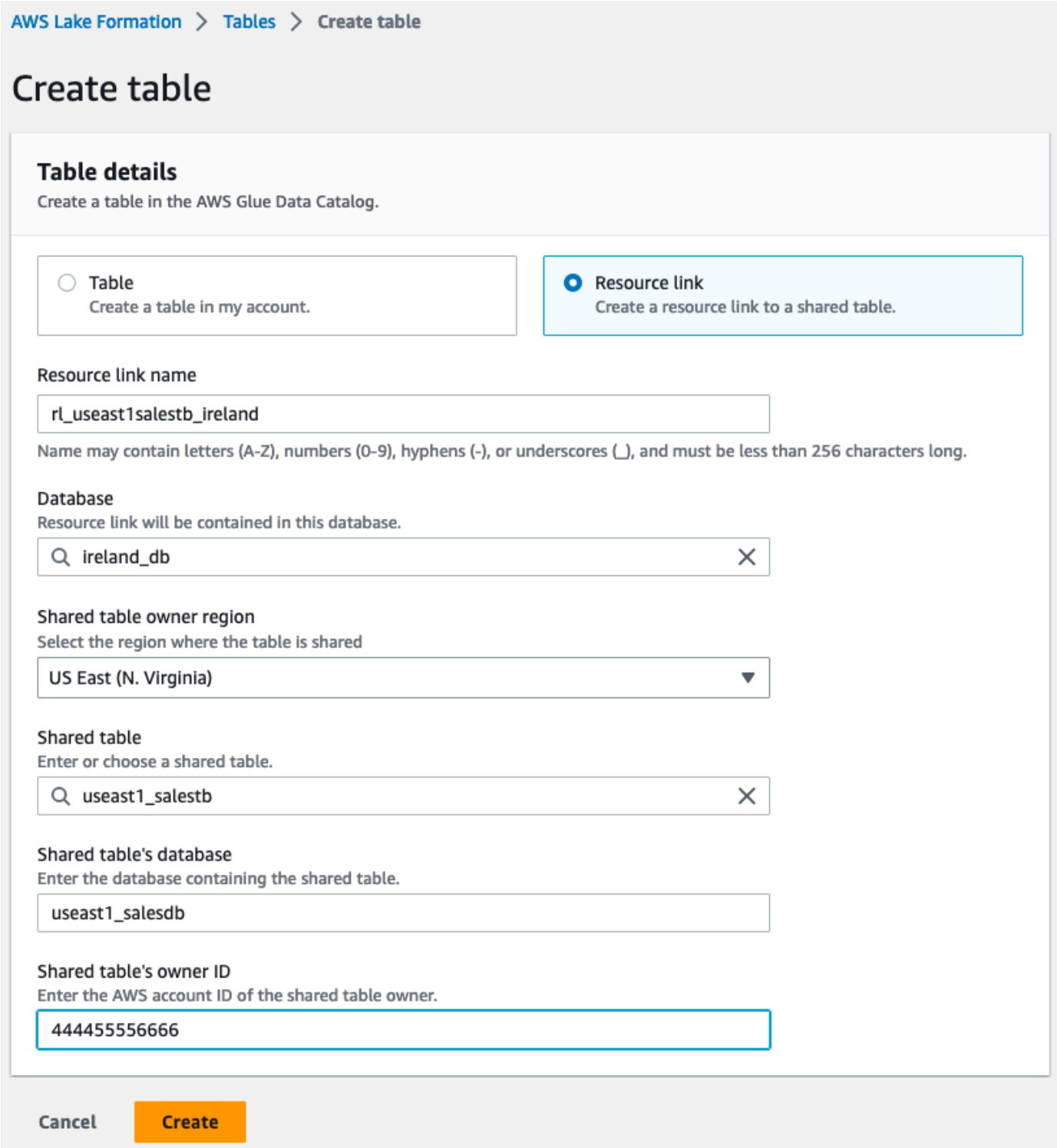
- Si vous n'êtes pas un administrateur de lac de données, vérifiez que l'administrateur du lac de données vous a accordé les autorisations de Lake Formation sur la table.
- Si vous êtes administrateur d'un lac de données et que votre compte n'appartient pas à la même AWS organisation que le compte émetteur, assurez-vous d'avoir accepté l'invitation de partage de ressources AWS Resource Access Manager (AWS RAM) pour la table. Pour plus d’informations, consultez Acceptation d'une invitation de partage de ressources de AWS RAM.

Base de données de tables partagées

Si vous avez sélectionné une table partagée dans la liste, ce champ est renseigné avec la base de données de la table partagée dans le compte externe. Sinon, entrez une base de données locale (pour un lien de ressource vers une table locale) ou la base de données de la table partagée dans le compte externe.

Propriétaire de la table partagée

Si vous avez sélectionné une table partagée dans la liste, ce champ est renseigné avec l'ID de compte du propriétaire de la table partagée. Sinon, entrez votre identifiant de AWS compte (pour un lien de ressource vers une table locale) ou l'identifiant du AWS compte qui a partagé la table.



4. Choisissez Créer pour créer le lien vers la ressource.

   Vous pouvez ensuite afficher le nom du lien vers la ressource dans la colonne Nom de la page Tables.

5. (Facultatif) Accordez l'DESCRIBEautorisation de Lake Formation sur le lien de la ressource aux principaux qui doivent être en mesure de voir le lien et d'accéder à la table cible.

Pour créer un lien de ressource vers une table partagée dans la même région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-table --database-name myissues --table-input
 '{"Name":"my_customers","TargetTable":
{"CatalogId":"111122223333","DatabaseName":"issues","Name":"customers"}}'
```

   Cette commande crée un lien de ressource nommé `my_customers` vers la table partagée`customers`, qui se trouve dans la base de données `issues` sous le AWS compte 1111-2222-3333. Le lien vers la ressource est stocké dans la base de données locale`myissues`.

2. (Facultatif) Accordez l'DESCRIBEautorisation de Lake Formation sur le lien de la ressource aux principaux qui doivent être en mesure de voir le lien et d'accéder à la table cible.

Pour créer un lien de ressource vers une table partagée dans une autre région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-table –region eu-west-1 –cli-input-json '{
    "CatalogId": "111122223333",
    "DatabaseName": "ireland_db",
    "TableInput": {
        "Name": "rl_useast1salestb_ireland",
        "TargetTable": {
            "CatalogId": "444455556666",
            "DatabaseName": "useast1_salesdb",
            "Region": "us-east-1",
            "Name":"useast1_salestb"
        }
    }
}‘
```

   Cette commande crée un lien de ressource nommé `rl_useast1salestb_ireland` dans la région Europe (Irlande) vers la table partagée`useast1_salestb`, qui se trouve dans la base de données `useast1_salesdb` du AWS compte 444455556666 dans la région USA Est (Virginie du Nord). Le lien vers la ressource est stocké dans la base de données locale`ireland_db`.

2. Accordez l'`DESCRIBE`autorisation de la Lake Formation aux principaux qui doivent être en mesure de voir le lien et d'accéder à la cible du lien via le lien.

> Voir aussi :
>
> - Mode de fonctionnement des liens des ressources dans Lake Formation
> - `DESCRIBE`

Création d'un lien de ressource vers une base de données de catalogue de données partagée

Vous pouvez créer un lien de ressource vers une base de données partagée à l'aide de la AWS Lake Formation console, de l'API ou AWS Command Line Interface (AWS CLI).

Pour créer un lien de ressource vers une base de données partagée (console)

1. Ouvrez la AWS Lake Formation console à l'adresse https://console.aws.amazon.com/lakeformation/. Connectez-vous en tant qu'administrateur du lac de données ou en tant que créateur de base de données.

   Un créateur de base de données est un directeur qui a obtenu l'`CREATE_DATABASE`autorisation de Lake Formation.
2. Dans le volet de navigation, choisissez Databases, puis Create database.
3. Sur la page Créer une base de données, choisissez la vignette Lien vers une ressource, puis fournissez les informations suivantes :

   Nom du lien vers la ressource

   Entrez un nom qui respecte les mêmes règles qu'un nom de base de données. Le nom peut être identique à celui de la base de données partagée cible.

   Région propriétaire de la base de données partagée

   Si vous créez le lien de ressource dans une autre région, sélectionnez la région de la base de données partagée cible.

Base de données partagée

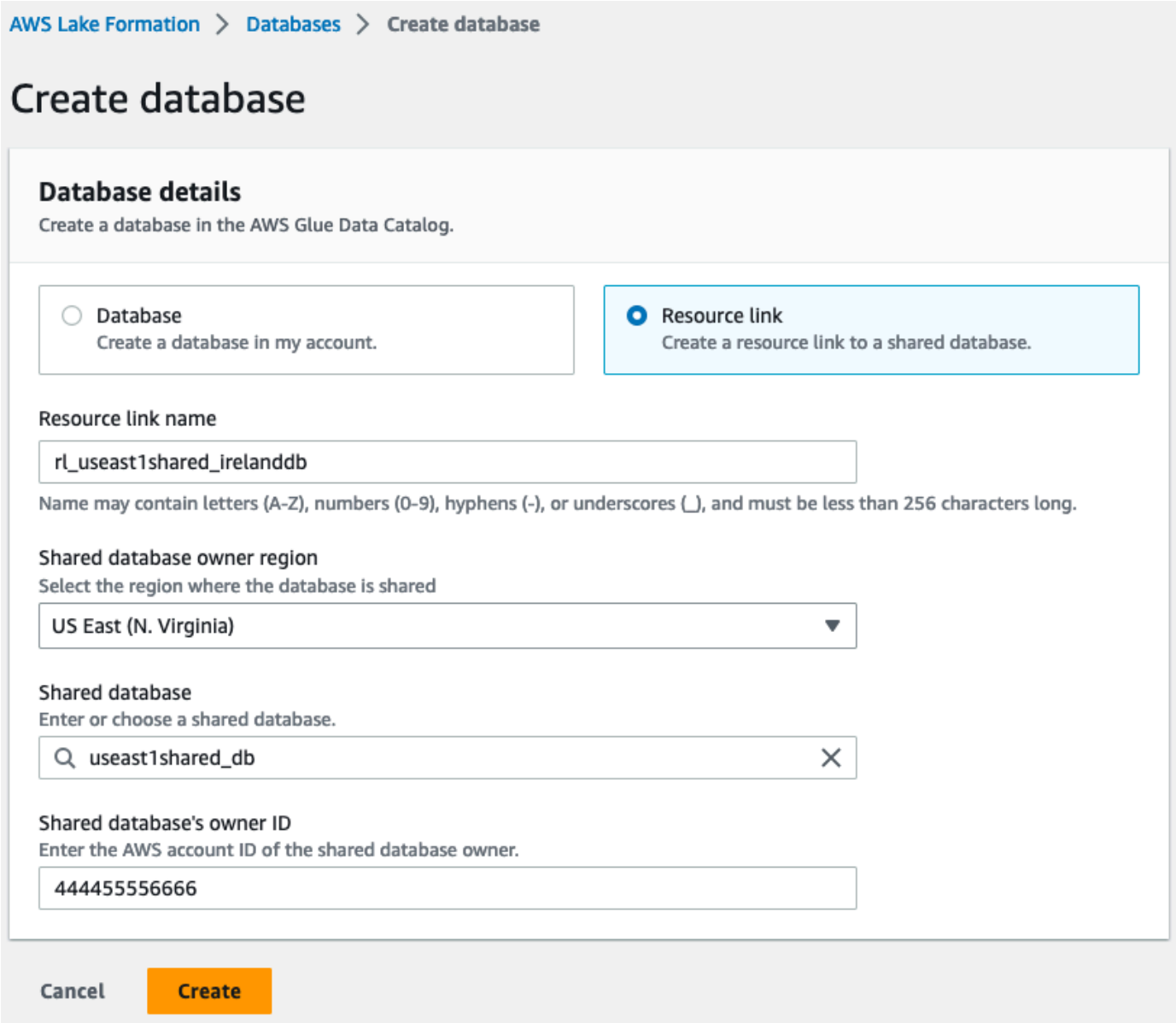
Choisissez une base de données dans la liste ou entrez un nom de base de données locale (détenue) ou partagée.

La liste contient toutes les bases de données partagées avec votre compte. Notez l'ID du compte propriétaire répertorié dans chaque base de données. Si vous ne trouvez aucune base de données dont vous savez qu'elle a été partagée avec votre compte, vérifiez les points suivants :

- Si vous n'êtes pas un administrateur de lac de données, vérifiez que l'administrateur du lac de données vous a accordé les autorisations de Lake Formation sur la base de données.
- Si vous êtes administrateur d'un lac de données et que votre compte n'appartient pas à la même AWS organisation que le compte octroyant, assurez-vous d'avoir accepté l'invitation de partage de ressources AWS Resource Access Manager (AWS RAM) pour la base de données. Pour plus d’informations, consultez Acceptation d'une invitation de partage de ressources de AWS RAM.

Propriétaire de la base de données partagée

Si vous avez sélectionné une base de données partagée dans la liste, ce champ est renseigné avec l'ID de compte du propriétaire de la base de données partagée. Sinon, entrez votre identifiant de AWS compte (pour un lien de ressource vers une base de données locale) ou l'identifiant du AWS compte qui a partagé la base de données.



4. Choisissez Créer pour créer le lien vers la ressource.

   Vous pouvez ensuite afficher le nom du lien vers la ressource dans la colonne Nom de la page Bases de données.

5. (Facultatif) Accordez l'`DESCRIBE`autorisation de la Lake Formation sur le lien de la ressource aux responsables de la région Europe (Irlande) qui doivent être en mesure de consulter le lien et d'accéder à la base de données cible.

Pour créer un lien de ressource vers une base de données partagée dans la même région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
{"CatalogId":"111122223333","DatabaseName":"issues"}}'
```

   Cette commande crée un lien de ressource nommé `myissues` vers la base de données partagée`issues`, qui se trouve dans le AWS compte 1111-2222-3333.

2. (Facultatif) Accordez l'`DESCRIBE`autorisation de la Lake Formation aux principaux sur le lien de ressource qui doivent être en mesure de voir le lien et d'accéder à la base de données cible.

Pour créer un lien de ressource vers une base de données partagée dans une autre région (AWS CLI)

1. Utilisez une commande similaire à la suivante.

```
aws glue create-database --region eu-west-1 --cli-input-json '{
    "CatalogId": "111122223333",
    "DatabaseInput": {
      "Name": "rl_useast1shared_irelanddb",
      "TargetDatabase": {
          "CatalogId": "444455556666",
          "DatabaseName": "useast1shared_db",
          "Region": "us-east-1"
       }
    }
}'
```

   Cette commande crée un lien de ressource nommé `rl_useast1shared_irelanddb` dans le AWS compte 111122223333 dans la région Europe (Irlande) vers la base de données partagée`useast1shared_db`, qui se trouve dans le AWS compte 444455556666 dans la région USA Est (Virginie du Nord).

2. Accordez l'`DESCRIBE`autorisation de la Lake Formation aux responsables de la région Europe (Irlande) qui doivent être en mesure de voir le lien et d'accéder à la cible du lien via le lien.

Voir aussi :

- Mode de fonctionnement des liens des ressources dans Lake Formation
- `DESCRIBE`

Gestion des liens de ressources dans les AWS Glue API

Les tableaux suivants expliquent comment les API du catalogue de AWS Glue données gèrent les liens vers les ressources des bases de données et des tables. Pour toutes les opérations `Get*` d'API, seules les bases de données et les tables sur lesquelles l'appelant est autorisé sont renvoyées. En outre, lorsque vous accédez à une base de données ou à une table cible via un lien de ressource, vous devez disposer à la fois des autorisations AWS Identity and Access Management (IAM) et de Lake Formation sur la cible et sur le lien de ressource. L'autorisation de Lake Formation requise pour les liens de ressources est`DESCRIBE`. Pour plus d'informations, consultez `DESCRIBE`.

Opérations de l'API de base

Opération API	Gestion des liens vers les ressources
`CreateDatabase`	Si la base de données est un lien de ressource, crée le lien de ressource vers la base de données cible désignée.
`UpdateDatabase`	Si la base de données désignée est un lien de ressource, suivez le lien et mettez à jour la base de données cible. Si le lien de ressource doit être modifié pour être lié à une autre base de données, vous devez le supprimer et en créer une nouvelle.
`DeleteDatabase`	Supprime le lien vers la ressource. Il ne supprime pas la base de données liée (cible).
`GetDatabase`	Si l'appelant dispose d'autorisations sur la cible, suivez le lien pour renvoyer les propriétés de la cible. Dans le cas contraire, elle renvoie les propriétés du lien.
`GetDatabases`	Renvoie une liste de bases de données, y compris des liens vers des ressources. Pour chaque lien de ressource du jeu de résultats, l'opération suit le lien pour obtenir les propriétés de la cible du lien.

Opération API	Gestion des liens vers les ressources
	Vous devez spécifier `ResourceShareType = ALL` pour voir les bases de données partagées avec votre compte.

Opérations de l'API Table

Opération API	Gestion des liens vers les ressources
`CreateTable`	Si la base de données est un lien de ressource, suit le lien de base de données et crée une table dans la base de données cible. Si la table est un lien de ressource, l'opération crée le lien de ressource dans la base de données désignée. La création d'un lien vers une ressource de table via un lien vers une ressource de base de données n'est pas prise en charge.
`UpdateTable`	Si la table ou la base de données désignée est un lien de ressource , met à jour la table cible. Si la table et la base de données sont des liens de ressources, l'opération échoue.
`DeleteTable`	Si la base de données désignée est un lien de ressource, suit le lien et supprime le lien de ressource de table ou de table dans la base de données cible. Si la table est un lien de ressource, l'opération supprime le lien de ressource de table dans la base de données désignée. La suppression d'un lien vers une ressource de table ne supprime pas la table cible.
`BatchDeleteTable`	Identique à `DeleteTable` .
`GetTable`	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie le lien de ressource de table ou de table depuis la base de données cible. Sinon, si la table est un lien vers une ressource, l'opération suit le lien et renvoie les propriétés de la table cible.
`GetTables`	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie les tables et les liens de ressource s de table depuis la base de données cible. Si la base de données

Opération API	Gestion des liens vers les ressources
	cible est une base de données partagée provenant d'un autre AWS compte, l'opération renvoie uniquement les tables partagées de cette base de données. Il ne suit pas les liens vers les ressource s de la table dans la base de données cible. Sinon, si la base de données désignée est une base de données locale (détenue), l'opération renvoie toutes les tables de la base de données locale et suit chaque lien de ressource de table pour renvoyer les propriétés de la table cible.
`SearchTables`	Renvoie les tables et les liens vers les ressources des tables. Il ne suit pas les liens pour renvoyer les propriétés de la table cible. Vous devez spécifier `ResourceShareType = ALL` pour voir les tables partagées avec votre compte.
`GetTableVersion`	Identique à `GetTable`.
`GetTableVersions`	Identique à `GetTable`.
`DeleteTableVersion`	Identique à `DeleteTable` .
`BatchDeleteTableVersion`	Identique à `DeleteTable` .

Opérations de l'API de partition

Opération API	Gestion des liens vers les ressources
`CreatePartition`	Si la base de données désignée est un lien de ressource, suit le lien de base de données et crée une partition dans la table désignée de la base de données cible. Si la table est un lien de ressource, l'opération suit le lien de ressource et crée la partition dans la table cible. La création d'une partition via un lien de ressource de table et un lien de ressource de base de données n'est pas prise en charge.
`BatchCreatePartition`	Identique à `CreatePartition` .

Opération API	Gestion des liens vers les ressources
`UpdatePartition`	Si la base de données désignée est un lien de ressource, suit le lien de base de données et met à jour la partition dans la table désignée dans la base de données cible. Si la table est un lien de ressource, l'opération suit le lien de ressource et met à jour la partition dans la table cible. La mise à jour d'une partition via un lien de ressource de table et un lien de ressource de base de données n'est pas prise en charge.
`DeletePartition`	Si la base de données désignée est un lien de ressource, suit le lien de base de données et supprime la partition dans la table désignée dans la base de données cible. Si la table est un lien de ressource , l'opération suit le lien de ressource et supprime la partition dans la table cible. La suppression d'une partition via un lien de ressource de table et un lien de ressource de base de données n'est pas prise en charge.
`BatchDeletePartition`	Identique à `DeletePartition` .
`GetPartition`	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie les informations de partition à partir de la table désignée. Sinon, si la table est un lien de ressource , l'opération suit le lien et renvoie les informations de partition. Si la table et la base de données sont toutes deux des liens de ressources, elles renvoient un jeu de résultats vide.
`GetPartitions`	Si la base de données désignée est un lien de ressource, suit le lien de base de données et renvoie les informations de partition pour toutes les partitions de la table désignée. Sinon, si la table est un lien de ressource, l'opération suit le lien et renvoie les informations de partition. Si la table et la base de données sont toutes deux des liens de ressources, elles renvoient un jeu de résultats vide.
`BatchGetPartition`	Identique à `GetPartition` .

Fonctions définies par l'utilisateur et opérations d'API

Opération API	Gestion des liens vers les ressources
(Toutes les opérations d'API)	Si la base de données est un lien de ressource, suit le lien de ressource et exécute l'opération sur la base de données cible.

Voir aussi :

- Mode de fonctionnement des liens des ressources dans Lake Formation

Accès aux tables dans toutes les régions

Lake Formation permet d'interroger les tables du catalogue de données dans toutes les AWS régions. Vous pouvez accéder aux données d'une région depuis d'autres régions à l'aide d'Amazon Athena, Amazon EMR et AWS Glue ETL en créant des liens de ressources dans d'autres régions pointant vers les bases de données et les tables sources. Grâce à l'accès aux tables entre régions, vous pouvez accéder aux données entre les régions sans copier les données sous-jacentes ou les métadonnées dans le catalogue de données.

Par exemple, vous pouvez partager une base de données ou une table d'un compte producteur avec un compte client de la région A. Après avoir accepté l'invitation de partage de ressources dans la région A, l'administrateur du lac de données du compte client peut créer des liens vers la ressource partagée dans la région B. L'administrateur du compte client peut accorder des autorisations sur la ressource partagée aux principaux IAM de ce compte dans la région A et peut accorder des autorisations de lien de ressource dans la région B. En utilisant le lien de ressource, les principaux du compte client peuvent interrogez les données partagées de la région B.

Vous pouvez également héberger la source de données Amazon S3 dans la région A dans un compte producteur et enregistrer l'emplacement des données dans un compte central dans la région B. Vous pouvez créer des ressources de catalogue de données dans le compte central, configurer les autorisations de Lake Formation et partager des données avec les consommateurs de votre compte ou avec des comptes externes dans la région B. La fonctionnalité inter-régions permet aux utilisateurs d'accéder à ces tables du catalogue de données depuis la région C à l'aide de liens vers des ressources.

Grâce à cette fonctionnalité, vous pouvez interroger des bases de données fédérées dans les métastores Apache Hive d'une région à l'autre, et également joindre des tables de la région locale à des tables d'une autre région lors de l'exécution de requêtes.

Lake Formation prend en charge les fonctionnalités suivantes avec un accès aux tables interrégional :

- Contrôle d'accès basé sur des balises LF
- Autorisations de contrôle d'accès détaillées
- Opérations d'écriture sur la base de données ou la table partagée avec les autorisations appropriées
- Partage de données entre comptes au niveau du compte et directement avec les responsables IAM

Les utilisateurs non administrateurs dotés `Create_Database` d'`Create_Table`autorisations peuvent créer des liens de ressources interrégionaux.

> **Note**
>
> Vous pouvez créer des liens de ressources entre régions dans n'importe quelle région et accéder aux données sans appliquer les autorisations de Lake Formation. Pour les données source dans Amazon S3 qui ne sont pas enregistrées auprès de Lake Formation, l'accès est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions.

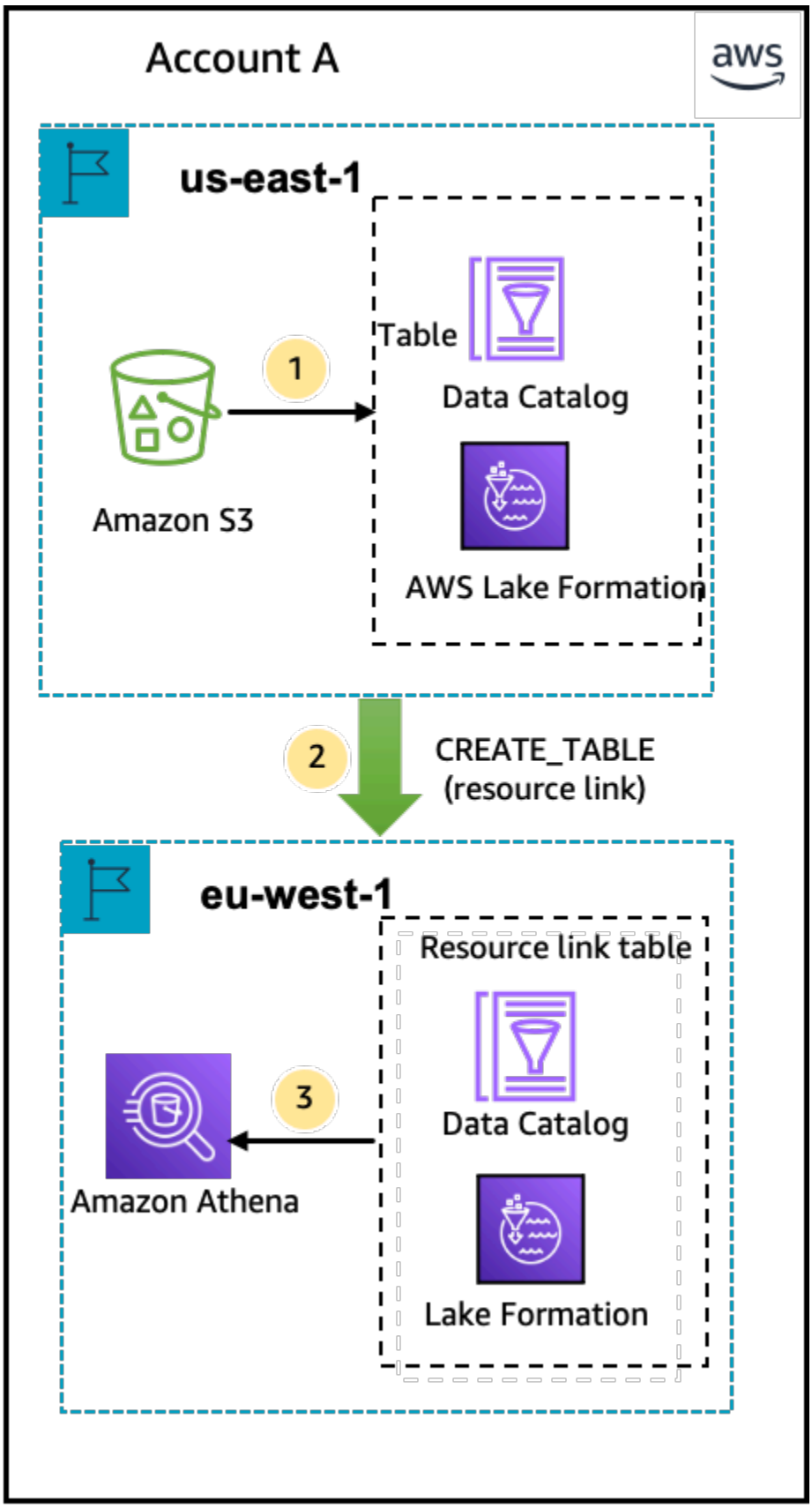
Pour connaître les limitations, veuillez consulter Limites d'accès aux données entre régions.

Flux de travail

Les diagrammes suivants montrent les flux de travail permettant d'accéder aux données entre AWS les régions à partir du même AWS compte et d'un compte externe.

Flux de travail pour accéder aux tables partagées au sein d'un même AWS compte

Dans le schéma ci-dessous, les données sont partagées avec un utilisateur du même AWS compte dans la région USA Est (Virginie du Nord), et l'utilisateur interroge les données partagées depuis la région Europe (Irlande).


Account A
aws
us-east-1
Table
Data Catalog
1
Amazon S3
AWS Lake Formation
2
CREATE_TABLE
(resource link)
eu-west-1
Resource link table
Data Catalog
3
Amazon Athena
Lake Formation

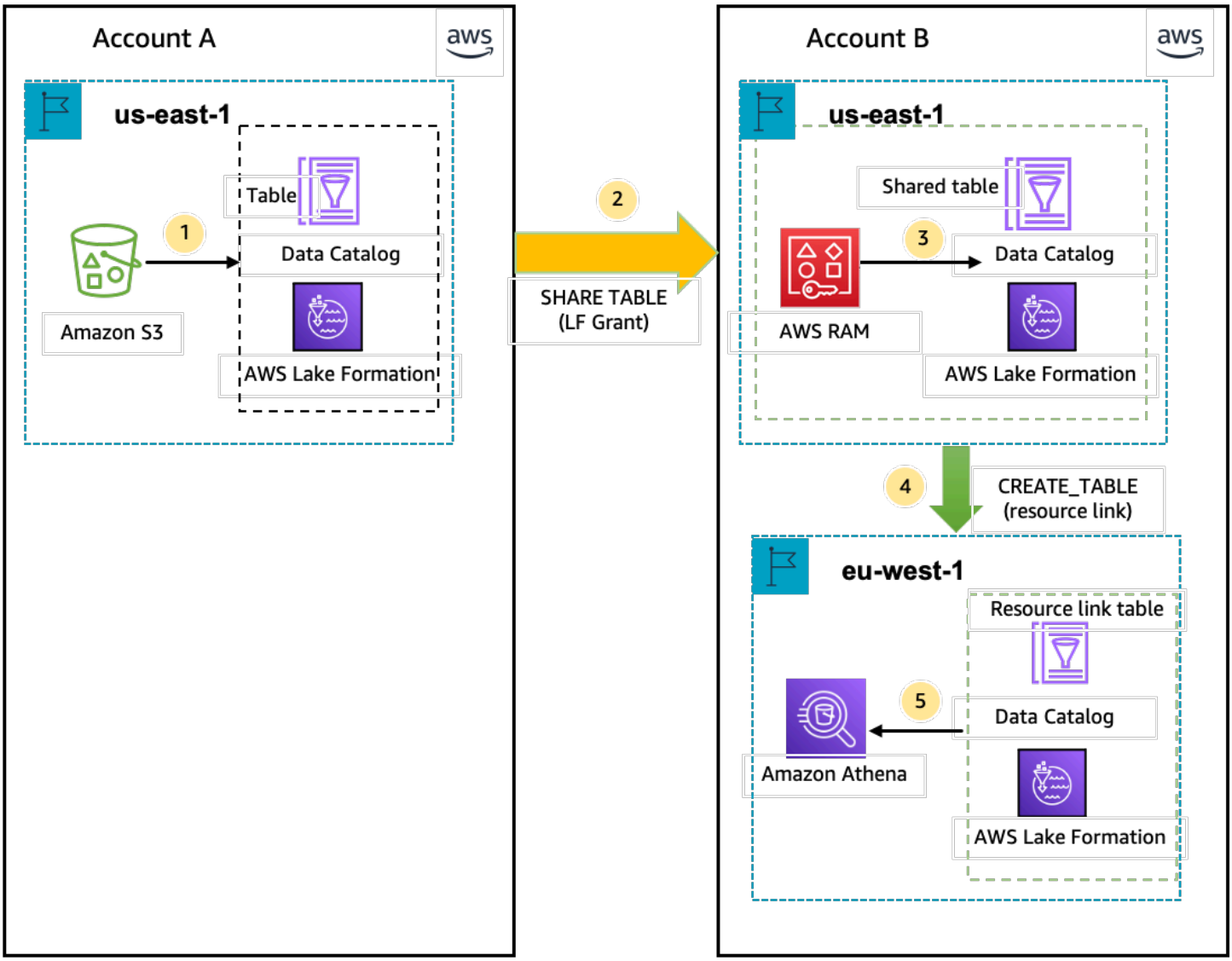
L'administrateur du lac de données effectue les activités suivantes (étapes 1 et 2) :

1. Un administrateur de lac de données crée un AWS compte auprès des bases de données et des tables du catalogue de données et enregistre un emplacement de données Amazon S3 auprès de Lake Formation dans la région de l'est des États-Unis (Virginie du Nord).

   Accorde une `Select` autorisation sur une ressource du catalogue de données (tableau des produits dans le diagramme) à un principal (utilisateur) du même compte.
2. Crée un lien de ressource dans la région Europe (Irlande) pointant vers la table source dans la région USA Est (Virginie du Nord). Accorde l'`DESCRIBE`autorisation sur le lien de ressource de la région Europe (Irlande) vers le principal.
3. L'utilisateur interroge le tableau depuis la région Europe (Irlande) à l'aide d'Athena.

Flux de travail pour accéder aux tables partagées avec un AWS compte externe

Dans le schéma ci-dessous, le compte producteur (compte A) héberge le compartiment Amazon S3, enregistre l'emplacement des données et partage une table du catalogue de données avec un compte consommateur (compte B) dans la région USA Est (Virginie du Nord) et un utilisateur du compte consommateur (compte B) interroge le tableau depuis la région Europe (Irlande).



1. Un administrateur de lac de données crée un AWS compte (compte producteur) avec les ressources du catalogue de données et un emplacement de données Amazon S3 enregistré auprès de Lake Formation dans la région de l'est des États-Unis (Virginie du Nord).
2. L'administrateur du lac de données du compte producteur partage une table du catalogue de données avec un compte client.
3. L'administrateur du lac de données du compte client accepte l'invitation à partager des données dans la région USA Est (Virginie du Nord) et accorde l'`Select`autorisation d'utiliser la table partagée à un mandant de la même région.
4. L'administrateur du lac de données du compte client crée un lien de ressource dans la région Europe (Irlande) pointant vers la table partagée cible dans la région USA Est (Virginie du Nord) et

accorde à l'utilisateur l'`DESCRIBE`autorisation d'utiliser le lien de ressource depuis la région Europe (Irlande).

5. L'utilisateur interroge les données de la région Europe (Irlande) à l'aide d'Athena.

Configuration de l'accès aux tables entre régions

Pour accéder aux données d'une autre région, vous devez d'abord configurer les bases de données et les tables du catalogue de données dans la région dans laquelle vous enregistrez votre emplacement de données Amazon S3. Vous pouvez partager les bases de données et les tables du catalogue de données avec les principaux de votre compte ou d'un autre compte. Vous devez ensuite créer des administrateurs de lacs de données qui peuvent créer des liens de ressources pointant vers l'emplacement des données partagées cible dans les régions où les utilisateurs interrogent les données.

Pour interroger des données partagées au sein d'un même compte depuis une autre région

Dans cette section, la table partagée cible Region est appelée Région A et les utilisateurs exécutent des requêtes depuis la Région B.

1. Configuration du compte dans la région A (où vous créez et partagez les données)

   Un administrateur de lac de données doit effectuer les actions suivantes :

   a. Enregistrez un emplacement de données Amazon S3.

      Pour plus d'informations, consultez Ajouter un emplacement Amazon S3 à votre lac de données.

   b. Créez des bases de données et des tables dans le compte. Cela peut également être fait par un utilisateur non administratif autorisé à créer des bases de données et des tables.

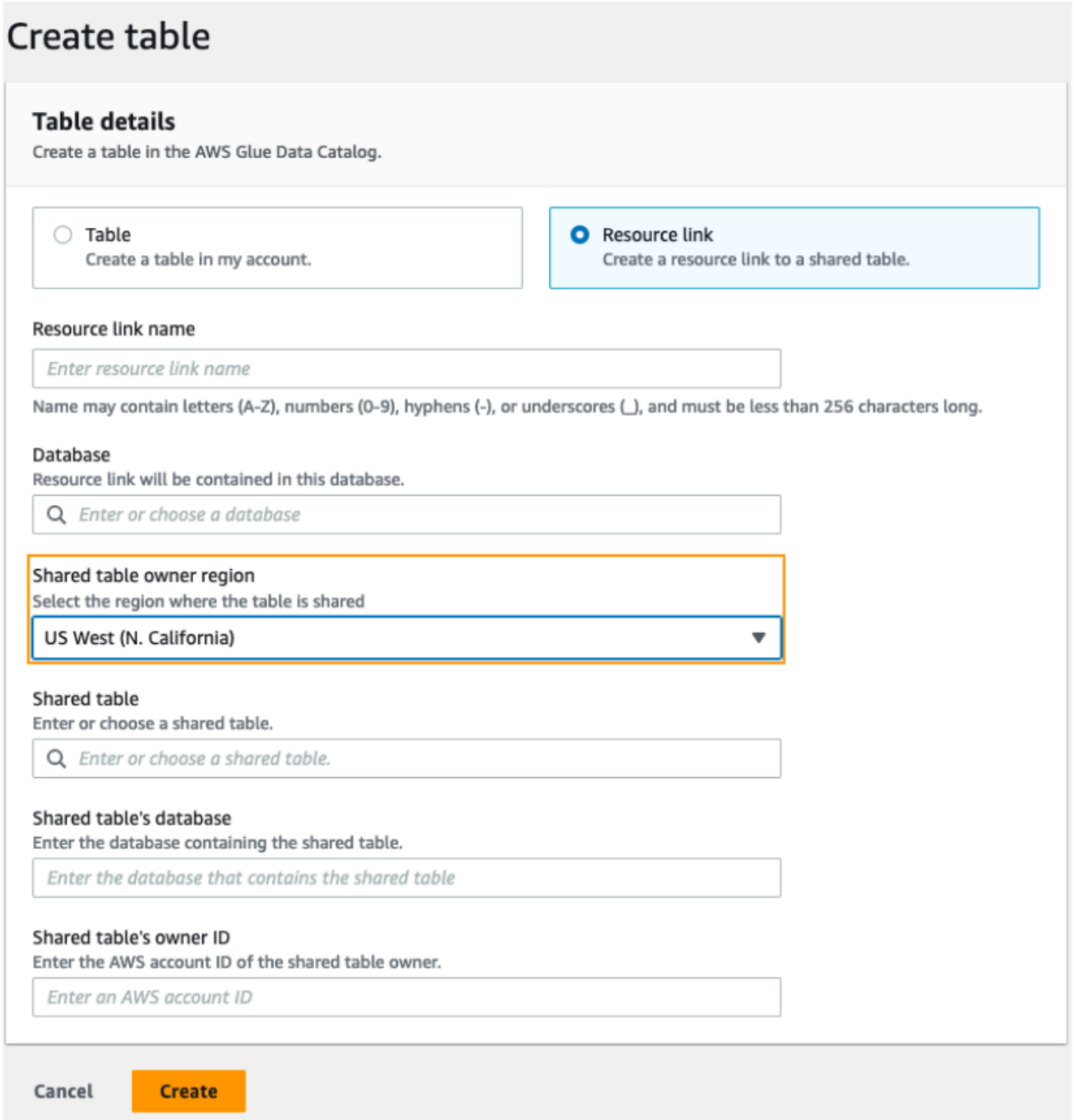
   c. Accordez des autorisations de données sur une table aux principaux avec`Grantable permissions`.

      Pour plus d'informations, veuillez consulter Octroi et révocation d'autorisations sur les ressources du catalogue de données.

2. Configuration du compte dans la région B (où vous accédez aux données)

   Un administrateur de lac de données doit effectuer les actions suivantes :

a. Créez un lien de ressource dans la région B pointant vers la table partagée cible dans la région A. Spécifiez la région propriétaire de la table partagée sur l'écran Créer une table.



Pour obtenir des instructions sur la création de liens de ressources vers des bases de données et des tables, consultezCréation de liens vers des ressources.

b. Accordez `Describe` l'autorisation aux responsables de l'IAM sur le lien de ressource dans la région B.

Pour plus d'informations sur l'octroi d'autorisations sur les liens vers des ressources, consultezOctroi d'autorisations relatives aux liens vers.

Les responsables IAM de la région B peuvent interroger la table cible via le lien en utilisant Athena.

Pour accéder aux données entre comptes d'une autre région

1. Configuration du compte producteur/concédant

   Un administrateur de lac de données doit effectuer les actions suivantes :

   a. Configurez le compte producteur/concédant dans la région A.

   b. Enregistrez un emplacement de données Amazon S3 dans la région A.

   c. Créez des bases de données et des tables. Cela peut être fait par un utilisateur non administrateur autorisé à créer des tables.

   d. Accordez des autorisations de données au compte du consommateur/bénéficiaire sur une table de la région A avec. `Grantable permissions`

      Pour plus d'informations, consultez Partage de tables et de bases de données du catalogue de données entre des comptes externes Comptes AWS ou avec des responsables IAM.

2. Configuration du compte consommateur/bénéficiaire

   Un administrateur de lac de données doit effectuer les actions suivantes :

   a. Acceptez l'invitation de partage de ressources depuis AWS RAM la région A.

   b. Créez un lien de ressource dans la région B pointant vers la table partagée. La région B est l'endroit où les utilisateurs voudront interroger la table.

   c. Accordez des autorisations de données sur la table partagée aux principaux IAM de la région A.

      > **Note**
      >
      > Vous devez accorder des autorisations à la table partagée dans la même région où la table a été partagée.

   d. Accordez des autorisations aux directeurs sur le lien de ressource dans la région B.

Les responsables du compte client de la région B interrogent ensuite la table partagée depuis la région B à l'aide d'Athena.

Partage de données dans AWS Lake Formation

Vous pouvez utiliser la fonctionnalité de partage de AWS Lake Formation données pour accorder et gérer des autorisations sur les données stockées dans des emplacements autres qu'Amazon S3, et sur les métadonnées stockées dans des emplacements autres que le AWS Glue Data Catalog. Grâce à la fonctionnalité de partage de données, vous pouvez configurer et gérer les autorisations sur les ensembles de données dans Amazon Redshift sans migrer les données vers Amazon S3. Vous pouvez également utiliser la fonction de fédération du catalogue de données pour vous connecter à des métastores externes.

Ensuite, vous pouvez utiliser Lake Formation pour gérer les données et les autorisations d'accès dans un catalogue de données central en définissant des politiques de contrôle d'accès précises. Les administrateurs de data lake peuvent accorder des autorisations à d'autres principaux IAM au sein du compte ou entre comptes sur les ressources du catalogue de données. Les responsables d'IAM peuvent interroger les données partagées à l'aide d'Amazon Redshift Spectrum et d'Amazon Athena.

Lake Formation propose les méthodes suivantes pour partager des données et gérer les autorisations sur des ensembles de données externes et des métastores externes :

- Intégration de Lake Formation au partage de données Amazon Redshift : utilisez Lake Formation pour gérer de manière centralisée les autorisations d'accès aux bases de données, aux tables, aux colonnes et aux lignes des partages de données Amazon Redshift et pour restreindre l'accès des utilisateurs aux objets d'un partage de données.
- Connexion AWS Glue Data Catalog à des métastores externes : connectez les à des métastores externes AWS Glue Data Catalog pour gérer les autorisations d'accès aux ensembles de données dans Amazon S3 à l'aide de Lake Formation. Aucune migration de métadonnées vers le AWS Glue Data Catalog n'est nécessaire.
- Intégrer Lake Formation à AWS Data Exchange — Lake Formation prend en charge l'octroi de licences d'accès à vos données via AWS Data Exchange. Si vous souhaitez obtenir une licence pour vos données sur la Lake Formation, consultez la section Contenu AWS Data Exchange du guide de AWS Data Exchange l'utilisateur.

Rubriques

- Gestion des autorisations pour les données dans un partage de données Amazon Redshift
- Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes

Gestion des autorisations pour les données dans un partage de données Amazon Redshift

Avec AWS Lake Formation, vous pouvez gérer les données en toute sécurité dans un partage de données d'Amazon Redshift. Amazon Redshift est un service d'entrepôt de données entièrement géré de plusieurs pétaoctets dans le cloud. AWS Grâce à la fonctionnalité de partage de données, Amazon Redshift vous permet de partager des données entre différentes entités. Comptes AWS Pour plus d'informations sur le partage de données Amazon Redshift, consultez Présentation du partage de données dans Amazon Redshift.

Dans Amazon Redshift, l'administrateur du cluster de producteurs crée un partage de données et le partage avec l'administrateur du lac de données. Pour step-by-step obtenir des instructions sur la création d'un administrateur de lac de données, consultezCréation d'un administrateur de lac de données.

Une fois que vous (administrateur du lac de données) avez accepté le partage de données, vous devez créer une AWS Glue Data Catalog base de données pour le partage de données spécifique. Cela vous permet de contrôler l'accès à ce site à l'aide des autorisations de Lake Formation. Lake Formation fait correspondre chaque partage de données à une base de données du catalogue de données correspondante. Elles apparaissent sous forme de bases de données fédérées dans le catalogue de données.

Une base de données est qualifiée de base de données fédérée lorsqu'elle pointe vers une entité extérieure au catalogue de données. Les tables et les vues du partage de données Amazon Redshift sont répertoriées sous forme de tables individuelles dans le catalogue de données. Vous pouvez partager la base de données fédérée avec des principaux IAM et des utilisateurs SAML sélectionnés au sein du même compte ou d'un autre compte avec Lake Formation. Vous pouvez également inclure des expressions de filtre de ligne et de colonne pour restreindre l'accès à certaines données. Pour plus d’informations, consultez Vue d'ensemble du filtrage des données.

Pour permettre aux utilisateurs d'accéder à un partage de données Amazon Redshift, vous devez effectuer les opérations suivantes :

1. Mettez à jour les paramètres du catalogue de données pour activer les autorisations de Lake Formation.
2. Acceptez l'invitation de partage de données envoyée par l'administrateur du cluster de producteurs Amazon Redshift et enregistrez le partage de données dans Lake Formation.

Une fois cette étape terminée, vous pouvez gérer le partage de données dans le Lake Formation Data Catalog.

3. Créez une base de données fédérée et définissez les autorisations sur cette base de données.
4. Accordez des autorisations aux utilisateurs sur les bases de données et les tables. Vous pouvez partager l'intégralité de la base de données ou un sous-ensemble de tables avec des utilisateurs du même compte ou d'un autre compte.

Pour connaître les limitations, veuillez consulter Limites du partage de données Amazon Redshift.

Rubriques

- Conditions préalables à la configuration des autorisations sur les partages de données Amazon Redshift
- Configuration des autorisations pour les partages de données Amazon Redshift
- Interrogation de bases de données fédérées

Conditions préalables à la configuration des autorisations sur les partages de données Amazon Redshift

Mettre à jour les paramètres par défaut du catalogue de données

Pour activer les autorisations Lake Formation pour les ressources du catalogue de données, nous vous recommandons de désactiver les paramètres par défaut du catalogue de données dans Lake Formation. Pour plus d'informations, consultez Modifier le modèle d'autorisation par défaut ou utiliser le mode d'accès hybride.

Mettre à jour les autorisations

Outre les autorisations d'administrateur du lac de données (`AWSLakeFormationDataAdmin`), les autorisations suivantes sont également requises pour accepter un partage de données Amazon Redshift dans Lake Formation :

- `glue:PassConnection on aws:redshift`
- `redshift:AssociateDataShareConsumer`
- `redshift:DescribeDataSharesForConsumer`
- `redshift:DescribeDataShares`

L'utilisateur IAM de l'administrateur du lac de données dispose implicitement des autorisations suivantes.

- accès_location_aux données
- créer_une base de données
- Formation lacustre : Enregistrer une ressource

Configuration des autorisations pour les partages de données Amazon Redshift

Cette rubrique décrit les étapes à suivre pour accepter une invitation de partage de données, créer une base de données fédérée et accorder des autorisations. Vous pouvez utiliser la console Lake Formation ou le AWS Command Line Interface (AWS CLI). Les exemples présentés dans cette rubrique montrent le cluster de producteurs, le catalogue de données et le consommateur de données dans le même compte.

Pour en savoir plus sur les fonctionnalités multicomptes de Lake Formation, voirPartage de données entre comptes dans Lake Formation.

Pour configurer des autorisations pour un partage de données

1. Passez en revue une invitation de partage de données et acceptez-la.

   Console

   1. Connectez-vous à la console Lake Formation en tant qu'administrateur de lac de données à l'adresse https://console.aws.amazon.com/lakeformation/. Accédez à la page Partage de données.
   2. Passez en revue les partages de données auxquels vous êtes autorisé à accéder. La colonne État indique votre statut de participation actuel pour le partage de données. Le statut En attente indique que vous avez été ajouté à un partage de données, mais que vous ne l'avez pas encore accepté ou que vous avez rejeté l'invitation.
   3. Pour répondre à une invitation de partage de données, sélectionnez le nom du partage de données et choisissez Vérifier l'invitation. Dans Accepter ou rejeter le partage de données, passez en revue les détails de l'invitation. Choisissez Accepter pour accepter l'invitation ou Refuser pour refuser l'invitation. Vous n'avez pas accès au partage de données si vous rejetez l'invitation.

AWS CLI

Les exemples suivants montrent comment afficher, accepter et enregistrer l'invitation. Remplacez l' Compte AWS identifiant par un Compte AWS identifiant valide. Remplacez le `data-share-arn` par le véritable Amazon Resource Name (ARN) qui fait référence au partage de données.

1. Afficher une invitation en attente.

```
aws redshift describe-data-shares \
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
```

2. Acceptez un partage de données.

```
 aws redshift associate-data-share-consumer \
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
 --consumer-arn 'arn:aws:glue:us-east-1:111122223333:catalog
```

3. Enregistrez le partage de données dans le compte Lake Formation. Utilisez l'opération RegisterResourceAPI pour enregistrer le partage de données dans Lake Formation. `DataShareArn`est le paramètre d'entrée pour`ResourceArn`.

> **Note**
>
> Il s'agit d'une étape obligatoire.

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds'
```

2. Créez une base de données.

Après avoir accepté une invitation de partage de données, vous devez créer une base de données qui pointe vers la base de données Amazon Redshift associée au partage de données. Vous devez être un administrateur de data lake pour créer une base de données.

Console

1. Sélectionnez le partage de données dans le volet Invitations et choisissez Définir les détails de la base de données.
2. Dans Définir les détails de la base de données, entrez un nom et un identifiant uniques pour le partage de données. Vous utilisez cet identifiant pour mapper le partage de données en interne dans la hiérarchie des métadonnées (DBName.Schema.Table).
3. Choisissez Next pour accorder des autorisations à d'autres utilisateurs sur la base de données et les tables partagées.

AWS CLI

Utilisez l'exemple de code suivant pour créer une base de données qui pointe vers la base de données Amazon Redshift partagée avec Lake Formation à l'aide du. AWS CLI

```
aws glue create-database --cli-input-json \

'{
 "CatalogId": "111122223333",
 "DatabaseInput": {
  "Name": "tahoedb",
  "FederatedDatabase": {
       "Identifier": "arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/federatedds",
       "ConnectionName": "aws:redshift"
   }
 }
 }'
```

3. Accordez des autorisations.

Après avoir créé la base de données, vous pouvez accorder des autorisations aux utilisateurs de votre compte ou à Comptes AWS des organisations externes. Vous ne pourrez pas accorder d'autorisations d'écriture (insertion, suppression) et de métadonnées (modification, suppression, création) sur la base de données fédérée mappée à un partage de données Amazon Redshift.

Pour plus d'informations sur l'octroi d'autorisations, consultezGestion des autorisations relatives à Lake Formation.

> **Note**
>
> En tant qu'administrateur de data lake, vous ne pouvez consulter que les tables des bases de données fédérées. Pour effectuer toute autre action, vous devez vous accorder davantage d'autorisations sur ces tables.

Console

1. Sur l'écran Accorder des autorisations, sélectionnez les utilisateurs auxquels vous souhaitez accorder des autorisations.
2. Choisissez Grant (Accorder).

AWS CLI

Utilisez les exemples suivants pour accorder des autorisations de base de données et de table à l'aide de AWS CLI :

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
   "Principal": {
           "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/non-
admin"
   },
   "Resource": {
          "Database": {
                "CatalogId": "111122223333",
                 "Name": "tahoedb"
           }
    },
    "Permissions": [
             "DESCRIBE"
     ],
    "PermissionsWithGrantOption": [

     ]
```

```
 }
```

```
aws lakeformation grant-permissions --input-cli-json file://input.json

{
                   "Principal": {
                         "DataLakePrincipalIdentifier":
 "arn:aws:iam::111122223333:user/non-admin"
                   },
                  "Resource": {
                         "Table": {
                              "CatalogId": "111122223333",
                              "DatabaseName": "tahoedb",
                              "Name": "public.customer"
                        }
                  },
                 "Permissions": [
                        "SELECT"
                  ],
                 "PermissionsWithGrantOption": [
                         "SELECT"
                   ]
 }
```

Interrogation de bases de données fédérées

Après avoir accordé les autorisations, les utilisateurs peuvent se connecter et commencer à interroger la base de données fédérée à l'aide d'Amazon Redshift. Les utilisateurs peuvent désormais utiliser le nom de la base de données locale pour faire référence au partage de données Amazon Redshift dans les requêtes SQL. Dans Amazon Redshift, la table client du schéma public partagé via le partage de données sera associée à une table correspondante créée comme `public.customer` dans le catalogue de données.

1. Avant d'interroger la base de données fédérée à l'aide d'Amazon Redshift, l'administrateur du cluster crée une base de données à partir de la base de données du catalogue de données à l'aide de la commande suivante :

```
CREATE DATABASE sharedcustomerdb FROM ARN
 'arn:aws:glue:<region>:111122223333:database/tahoedb' WITH DATA CATALOG SCHEMA
 tahoedb
```

2. L'administrateur du cluster accorde des autorisations d'utilisation sur la base de données.

```
GRANT USAGE ON DATABASE sharedcustomerdb TO IAM:user;
```

3. Vous (l'utilisateur fédéré) pouvez désormais vous connecter aux outils SQL pour interroger la table.

```
Select * from sharedcustomerdb.public.customer limit 10;
```

Pour plus d'informations, consultez la section Requête AWS Glue Data Catalog dans le guide de gestion Amazon Redshift.

Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes

Grâce à la fédération des AWS Glue Data Catalog métadonnées (fédération du catalogue de données), vous pouvez connecter le catalogue de données à des métastores externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation Il n'est pas nécessaire de migrer les métadonnées du métastore externe vers le catalogue de données.

Le catalogue de données fournit un référentiel de métadonnées centralisé qui facilite la gestion et la découverte de données sur des systèmes disparates. Lorsque votre organisation gère les données du catalogue de données, vous pouvez les utiliser AWS Lake Formation pour contrôler l'accès à vos ensembles de données dans Amazon S3.

> **Note**
>
> Actuellement, nous prenons uniquement en charge la fédération de métastores Apache Hive (version 3 et supérieure).

Pour configurer la fédération de catalogues de données, nous fournissons une application AWS Serverless Application Model (AWS SAM) appelée GlueDataCatalogFederation- HiveMetastore dans le AWS Serverless Application Repository.

L'implémentation de référence est fournie GitHub sous forme de projet open source sur AWS Glue Data Catalog Federation - Hive Metastore.

L' AWS SAM application crée et déploie les ressources suivantes qui sont nécessaires pour connecter le catalogue de données au métastore Hive :

- Une AWS Lambda fonction : héberge l'implémentation du service de fédération qui communique entre le catalogue de données et le métastore Hive. AWS Glue invoque cette fonction Lambda pour récupérer des objets de métadonnées depuis le métastore Hive.
- Amazon API Gateway— Le point de connexion de votre métastore Hive qui agit comme un proxy pour acheminer toutes les invocations vers la fonction Lambda.
- Rôle IAM : rôle doté des autorisations nécessaires pour créer la connexion entre le catalogue de données et le métastore Hive.
- AWS Glue connection — Amazon API Gateway Type de AWS Glue connexion qui stocke le Amazon API Gateway point de terminaison et un rôle IAM pour l'invoquer.

Lorsque vous interrogez des tables, le AWS Glue service effectue un appel d'exécution vers le métastore Hive et récupère les métadonnées. La fonction Lambda agit comme un traducteur entre le métastore Hive et le catalogue de données.

Après avoir établi la connexion, afin de synchroniser les métadonnées du métastore Hive avec le catalogue de données, vous devez créer une base de données fédérée dans le catalogue de données à l'aide des détails de connexion du métastore Hive, et mapper cette base de données à la base de données Hive. Une base de données est qualifiée de base de données fédérée lorsqu'elle pointe vers une entité extérieure au catalogue de données.

Vous pouvez appliquer les autorisations de Lake Formation à l'aide du contrôle d'accès basé sur des balises et de la méthode des ressources nommées sur la base de données fédérée Comptes AWS AWS Organizations, et les partager entre plusieurs unités organisationnelles (UO). Vous pouvez également partager la base de données fédérée directement avec les principaux IAM depuis un autre compte.

Vous pouvez définir des autorisations détaillées au niveau des colonnes, des lignes et des cellules à l'aide des filtres de données Lake Formation sur les tables Hive externes. Vous pouvez utiliser

Amazon Athena, Amazon Redshift ou Amazon EMR pour interroger les tables Hive externes gérées par Lake Formation.

Pour plus d'informations sur le partage de données entre comptes et le filtrage des données, voir :

- Partage de données entre comptes dans Lake Formation
- Filtrage des données et sécurité au niveau des cellules dans Lake Formation

Étapes générales de fédération des métadonnées du catalogue de données

1. Vous créez des utilisateurs et des rôles IAM dotés des autorisations appropriées pour déployer l' AWS SAM application et créer des bases de données fédérées.
2. Vous enregistrez l'emplacement des données Amazon S3 auprès de Lake Formation en sélectionnant l'`Enable Data Catalog federation`option pour les ensembles de données qui utilisent un métastore Hive externe.
3. Vous configurez les paramètres de l' AWS SAM application (nom de AWS Glue connexion, URL du métastore Hive et paramètres de la fonction Lambda) et déployez l'application. AWS SAM
4. L' AWS SAM application déploie les ressources nécessaires pour connecter le métastore Hive externe au catalogue de données.
5. Pour appliquer les autorisations Lake Formation à la base de données et aux tables Hive, vous créez une base de données dans le catalogue de données à l'aide des détails de connexion au métastore Hive, et vous mappez cette base de données à la base de données Hive.
6. Accordez des autorisations sur les bases de données fédérées aux principaux de votre compte ou d'un autre compte.

> **Note**
>
> Vous pouvez connecter le catalogue de données à un mestastore Hive externe, créer des bases de données fédérées et exécuter des requêtes et des scripts ETL sur des bases de données et des tables Hive sans appliquer les autorisations de Lake Formation. Pour les données source dans Amazon S3 qui ne sont pas enregistrées auprès de Lake Formation, l'accès est déterminé par les politiques d'autorisation IAM pour Amazon S3 et AWS Glue les actions.

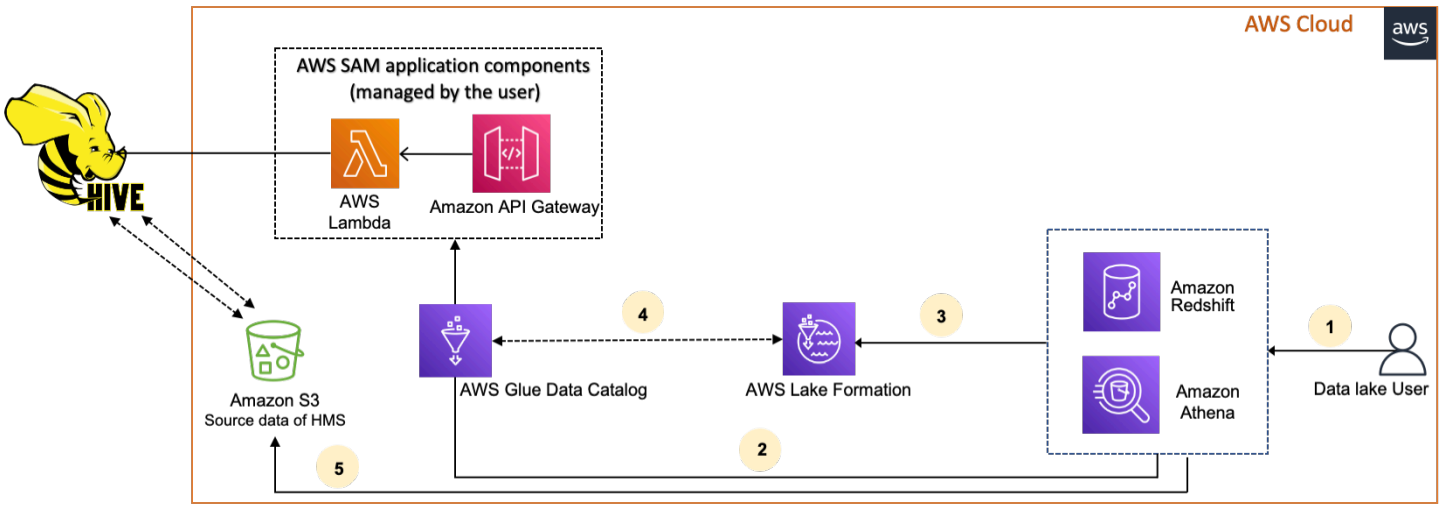
Pour connaître les limitations, veuillez consulter Considérations et limites relatives au partage des données du magasin de métadonnées Hive.

Rubriques

- Flux de travail
- Conditions préalables à la connexion du catalogue de données au métastore Hive
- Connexion du catalogue de données à un métastore Hive externe
- Ressources supplémentaires

Flux de travail

Le schéma suivant montre le flux de travail pour connecter le AWS Glue Data Catalog à un métastore Hive externe.



1. Un directeur soumet une requête à l'aide d'un service intégré tel qu'Athena ou Redshift Spectrum.
2. Le service intégré appelle le catalogue de données pour obtenir les métadonnées, qui à son tour appelle le point de terminaison du métastore Hive disponible derrière Amazon API Gateway et reçoit les réponses aux demandes de métadonnées.
3. Le service intégré envoie la demande à Lake Formation pour vérifier les informations de la table et les informations d'identification pour accéder à la table.
4. Lake Formation autorise la demande et fournit des informations d'identification temporaires à l'application intégrée, qui permet l'accès aux données.
5. À l'aide des informations d'identification temporaires reçues de Lake Formation, le service intégré lit les données d'Amazon S3 et partage les résultats avec le principal.

Conditions préalables à la connexion du catalogue de données au métastore Hive

Pour connecter le AWS Glue Data Catalog à un métastore Apache Hive externe et configurer les autorisations d'accès aux données, vous devez remplir les conditions suivantes :

> **Note**
>
> Nous recommandons qu'un administrateur de Lake Formation déploie l' AWS SAM application, et seul un utilisateur privilégié utilise la connexion au métastore Hive pour créer les bases de données fédérées correspondantes.

1. Création des rôles IAM.

   Pour déployer l' AWS SAM application

   - Créez un rôle doté des autorisations nécessaires pour déployer les ressources (fonction Lambda Amazon API Gateway, rôle IAM et AWS Glue connexion) requises pour créer une connexion au métastore Hive.

   Pour créer des bases de données fédérées

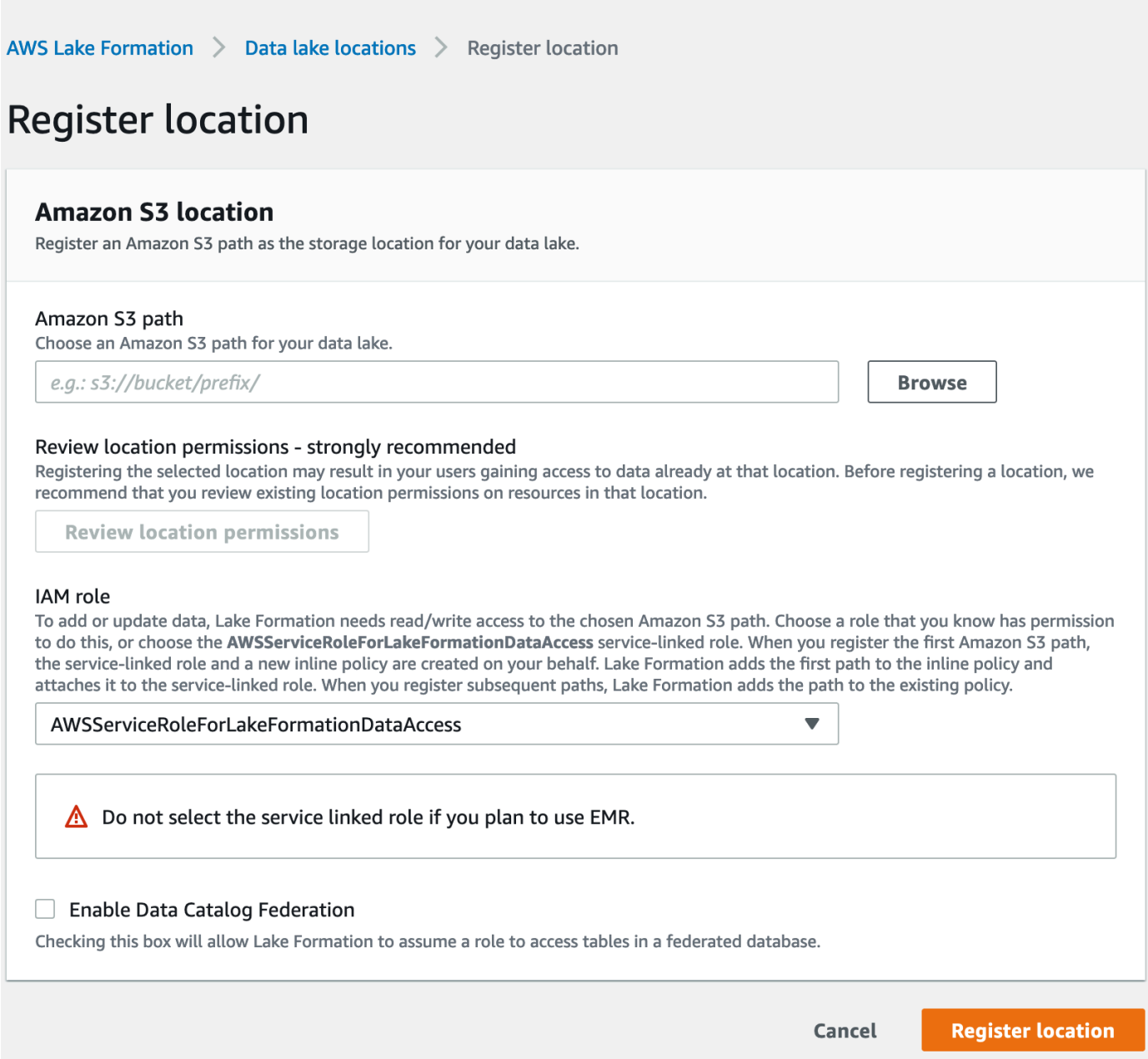
   Les autorisations suivantes sont requises pour les ressources :

   - `glue:CreateDatabase on resource arn:aws:glue:region:account-id:database/gluedatabasename`
   - `glue:PassConnection on resource arn:aws:glue:region:account-id:connection/hms_connection`
2. Enregistrez le site Amazon S3 auprès de Lake Formation.

   Pour utiliser Lake Formation afin de gérer et de sécuriser les données de votre lac de données, vous devez enregistrer l'emplacement Amazon S3 contenant les données des tables du métastore Hive auprès de Lake Formation. Lake Formation peut ainsi vendre des informations d'identification à des services AWS d'analyse tels qu'Athena, Redshift Spectrum et Amazon EMR.

Pour plus d'informations sur l'enregistrement d'un site Amazon S3, consultezAjouter un emplacement Amazon S3 à votre lac de données.

Lorsque vous enregistrez l'emplacement Amazon S3, cochez la case Enable Data Catalog Federation pour autoriser Lake Formation à assumer un rôle d'accès aux tables d'une base de données fédérée.



Pour plus d'informations sur l'enregistrement d'un emplacement de données auprès de Lake Formation, consultezConfigurer un emplacement Amazon S3 pour votre lac de données.

3. Utilisez la bonne version d'Amazon EMR.

   Pour utiliser Amazon EMR avec les bases de données de métastore Hive fédérées, vous devez disposer de la version 3.x ou supérieure de Hive et de la version 6.x ou supérieure d'Amazon EMR.

Connexion du catalogue de données à un métastore Hive externe

Pour le connecter AWS Glue Data Catalog à un métastore Hive, vous devez déployer une AWS SAM application appelée -. GlueDataCatalogFederation HiveMetastore Il crée les ressources nécessaires pour connecter le métastore Hive externe au catalogue de données. Vous pouvez accéder à l' AWS SAM application dans le AWS Serverless Application Repository.

L' AWS SAM application crée la connexion pour le métastore Hive derrière Amazon API Gateway à l'aide d'une fonction Lambda. L' AWS SAM application utilise un identifiant de ressource uniforme (URI) comme entrée par l'utilisateur et connecte le métastore Hive externe au catalogue de données. Lorsqu'un utilisateur exécute une requête sur des tables Hive, le catalogue de données appelle le point de terminaison API Gateway. Le point de terminaison invoque la fonction Lambda pour récupérer les métadonnées des tables Hive.

Pour connecter le catalogue de données au métastore Hive et configurer les autorisations

1. Déployez AWS SAM l'application.

   1. Connectez-vous au AWS Management Console et ouvrez le AWS Serverless Application Repository.
   2. Dans le volet de navigation, choisissez Applications.
   3. Choisissez Applications publiques.
   4. Sélectionnez l'option Show apps that create custom IAM roles or resource policies (Afficher les applications qui créent des rôles IAM ou des politiques de ressources personnalisés).
   5. Dans le champ de recherche, entrez le nom GlueDataCatalogFederation- HiveMetastore.
   6. Choisissez l'HiveMetastoreapplication GlueDataCatalogFederation-.
   7. Sous Paramètres de l'application, entrez les paramètres minimaux requis suivants pour votre fonction Lambda :
      - Nom de l'application : nom de votre AWS SAM application.
      - GlueConnectionName- Un nom pour la connexion.

- HiveMetastoreURI - L'URI de votre hôte de métastore Hive.
- LambdaMemory- La quantité de mémoire Lambda en Mo comprise entre 128 et 10240. La valeur par défaut est 1024.
- LambdaTimeout- Durée maximale d'invocation Lambda en secondes. La valeur par défaut est 30.
- VPC et SecurityGroupIds VPC : informations relatives au VPC sur SubnetIds lequel se trouve le métastore Hive.

8. Sélectionnez I acknowledge that this app creates custom IAM roles and resource policies (Je reconnais que cette application crée des politiques de ressources et rôles IAM personnalisés). Pour de plus amples informations, veuillez cliquer sur le lien Info.
9. En bas à droite de la section Application settings (Paramètres de l'application), choisissez Deploy (Déployer). Lorsque le déploiement est terminé, la fonction Lambda apparaît dans la section Resources (Ressources) dans la console Lambda.

L'application est déployée sur Lambda. Son nom est précédé de serverlessrepo- pour indiquer que l'application a été déployée à partir du. AWS Serverless Application Repository La sélection de l'application vous amène à la page Ressources où chacune des ressources de l'application qui ont été déployées est répertoriée. Les ressources incluent la fonction Lambda qui permet la communication entre le catalogue de données et le métastore Hive, la AWS Glue connexion et d'autres ressources nécessaires à la fédération de bases de données.

2. Créez une base de données fédérée dans le catalogue de données.

   Après avoir créé une connexion au métastore Hive, vous pouvez créer des bases de données fédérées dans le catalogue de données qui pointent vers les bases de données du métastore Hive externes. Vous devez créer une base de données correspondante dans le catalogue de données pour chaque base de données de métastore Hive que vous connectez au catalogue de données.

   Lake Formation console

   1. Sur la page Partage de données, choisissez l'onglet Bases de données partagées, puis sélectionnez Créer une base de données.
   2. Pour Nom de la connexion, choisissez le nom de votre connexion au métastore Hive dans le menu déroulant.

3. Entrez un nom de base de données unique et l'identifiant de source de fédération pour la base de données. Il s'agit du nom que vous utilisez dans vos instructions SQL lorsque vous interrogez des tables. Le nom peut comporter un maximum de 255 caractères et doit être unique au sein de votre compte.
4. Choisissez Créer une base de données.

AWS CLI

```
aws glue create-database \
'{
 "CatalogId": "<111122223333>",
  "database-input": {
    "Name":"<fed_glue_db>",
    "FederatedDatabase":{
        "Identifier":"<hive_db_on_emr>",
        "ConnectionName":"<hms_connection>"
     }
   }
 }'
```

3. Affichez les tables de la base de données fédérée.

Après avoir créé la base de données fédérée, vous pouvez consulter la liste des tables de votre métastore Hive à l'aide de la console Lake Formation ou du. AWS CLI

Lake Formation console

1. Sélectionnez le nom de la base de données dans l'onglet Bases de données partagées.
2. Sur la page Bases de données, choisissez Afficher les tables.

AWS CLI

Les exemples suivants montrent comment récupérer la définition de connexion, le nom de la base de données et certaines ou toutes les tables de la base de données. Remplacez l'ID du catalogue de données par l' Compte AWS ID valide que vous avez utilisé pour créer la base de données. Remplacez `hms_connection` par le nom de la connexion.

```
aws glue get-connection \
```

```
--name <hms_connection>  \
--catalog-id 111122223333
```

```
aws glue get-database \
--name <fed_glu_db> \
--catalog-id 111122223333
```

```
aws glue get-tables \
--database-name <fed_glue_db> \
--catalog-id 111122223333
```

```
aws glue get-table \
--database-name <fed_glue_db> \
--name <hive_table_name> \
--catalog-id 111122223333
```

4. Accordez des autorisations.

   Après avoir créé la base de données, vous pouvez accorder des autorisations à d'autres utilisateurs et rôles IAM dans votre compte ou à Comptes AWS des organisations externes. Vous ne pourrez pas accorder d'autorisations d'écriture (insertion, suppression) et de métadonnées (modification, suppression, création) sur les bases de données fédérées. Pour plus d'informations sur l'octroi d'autorisations, consultezGestion des autorisations relatives à Lake Formation.

5. Interrogez les bases de données fédérées.

   Après avoir accordé les autorisations, les utilisateurs peuvent se connecter et commencer à interroger la base de données fédérée à l'aide d'Athena et d'Amazon Redshift. Les utilisateurs peuvent désormais utiliser le nom de la base de données locale pour faire référence à la base de données Hive dans les requêtes SQL.

   Exemple de syntaxe Amazon Athena de requête

   `fed_glue_db`Remplacez-le par le nom de base de données locale que vous avez créé précédemment.

   ```
   Select * from fed_glue_db.customers limit 10;
   ```

Ressources supplémentaires

Le billet de blog suivant contient des instructions détaillées pour configurer les autorisations de Lake Formation sur une base de données et des tables de métastore Hive, et pour les interroger à l'aide d'Athena. Nous illustrons également un cas d'utilisation du partage entre comptes, dans lequel un responsable de Lake Formation inscrit sur le compte producteur A partage une base de données Hive fédérée et des tables en utilisant le tag LF sur le compte client B.

- Interrogez votre métastore Apache Hive avec des autorisations AWS Lake Formation

Sécurité dans AWS Lake Formation

La sécurité du cloud AWS est la priorité absolue. En tant que AWS client, vous bénéficiez d'un centre de données et d'une architecture réseau conçus pour répondre aux exigences des entreprises les plus sensibles en matière de sécurité.

La sécurité est une responsabilité partagée entre vous AWS et vous. Le modèle de responsabilité partagée décrit cette notion par les termes sécurité du cloud et sécurité dans le cloud :

- Sécurité du cloud : AWS est chargée de protéger l'infrastructure qui exécute les AWS services dans le AWS cloud. AWS vous fournit également des services que vous pouvez utiliser en toute sécurité. Des auditeurs tiers testent et vérifient régulièrement l'efficacité de notre sécurité dans le cadre des programmes de conformitéAWS. Pour en savoir plus sur les programmes de conformité qui s'appliquent à AWS Lake Formation, consultez la section AWS Services concernés par programme de conformité.
- Sécurité dans le cloud — Votre responsabilité est déterminée par le AWS service que vous utilisez. Vous êtes également responsable d'autres facteurs, y compris de la sensibilité de vos données, des exigences de votre entreprise, ainsi que de la législation et de la réglementation applicables.

Cette documentation vous aide à comprendre comment appliquer le modèle de responsabilité partagée lors de l'utilisation de Lake Formation. Les rubriques suivantes expliquent comment configurer Lake Formation pour répondre à vos objectifs de sécurité et de conformité. Vous apprendrez également à utiliser d'autres AWS services qui vous aident à surveiller et à sécuriser les ressources de vos Lake Formation.

Rubriques

- Protection des données dans le Lake Formation
- Sécurité de l'infrastructure dans AWS Lake Formation
- Prévention du problème de l'adjoint confus entre services
- Connexion aux événements de sécurité AWS Lake Formation

Protection des données dans le Lake Formation

Le modèle de responsabilité AWS partagée de s'applique à la protection des données dans AWS Lake Formation. Comme décrit dans ce modèle, AWS est chargé de protéger l'infrastructure

mondiale qui gère tous les AWS Cloud. La gestion du contrôle de votre contenu hébergé sur cette infrastructure relève de votre responsabilité. Vous êtes également responsable des tâches de configuration et de gestion de la sécurité des Services AWS que vous utilisez. Pour plus d'informations sur la confidentialité des données, consultez Questions fréquentes (FAQ) sur la confidentialité des données. Pour en savoir plus sur la protection des données en Europe, consultez le billet de blog Modèle de responsabilité partagée AWS et RGPD (Règlement général sur la protection des données) sur le Blog de sécuritéAWS .

À des fins de protection des données, nous vous recommandons de protéger les Compte AWS informations d'identification et de configurer les utilisateurs individuels avec AWS IAM Identity Center ou AWS Identity and Access Management (IAM). Ainsi, chaque utilisateur se voit attribuer uniquement les autorisations nécessaires pour exécuter ses tâches. Nous vous recommandons également de sécuriser vos données comme indiqué ci-dessous :

- Utilisez l'authentification multifactorielle (MFA) avec chaque compte.
- Utilisez le protocole SSL/TLS pour communiquer avec les ressources. AWS Nous exigeons TLS 1.2 et recommandons TLS 1.3.
- Configurez l'API et la journalisation de l'activité des utilisateurs avec AWS CloudTrail.
- Utilisez des solutions de AWS chiffrement, ainsi que tous les contrôles de sécurité par défaut qu'ils contiennent Services AWS.
- Utilisez des services de sécurité gérés avancés tels qu'Amazon Macie, qui contribuent à la découverte et à la sécurisation des données sensibles stockées dans Amazon S3.
- Si vous avez besoin de modules cryptographiques validés par la norme FIPS 140-2 pour accéder AWS via une interface de ligne de commande ou une API, utilisez un point de terminaison FIPS. Pour plus d'informations sur les points de terminaison FIPS (Federal Information Processing Standard) disponibles, consultez Federal Information Processing Standard (FIPS) 140-2 (Normes de traitement de l'information fédérale).

Nous vous recommandons fortement de ne jamais placer d'informations confidentielles ou sensibles, telles que les adresses e-mail de vos clients, dans des balises ou des champs de texte libre tels que le champ Name (Nom). Cela inclut lorsque vous travaillez avec Lake Formation ou une autre entreprise Services AWS à l'aide de la console, de l'API ou AWS des SDK. AWS CLI Toutes les données que vous entrez dans des balises ou des champs de texte de forme libre utilisés pour les noms peuvent être utilisées à des fins de facturation ou dans les journaux de diagnostic. Si vous fournissez une adresse URL à un serveur externe, nous vous recommandons fortement de ne pas

inclure d’informations d’identification dans l’adresse URL permettant de valider votre demande adressée à ce serveur.

Chiffrement au repos

AWS Lake Formation prend en charge le chiffrement des données dans les domaines suivants :

- Données de votre lac de données Amazon Simple Storage Service (Amazon S3).

  Lake Formation prend en charge le chiffrement des données avec AWS Key Management Service(AWS KMS). Les données sont généralement écrites dans le lac de données au moyen de tâches d'AWS Glueextraction, de transformation et de chargement (ETL). Pour plus d'informations sur le chiffrement des données écrites par les AWS Glue tâches, consultez la section Chiffrement des données écrites par les robots d'exploration, les tâches et les points de terminaison de développement dans le guide du développeur.AWS Glue
- Le AWS Glue Data Catalog, où Lake Formation stocke les tables de métadonnées décrivant les données du lac de données.

  Pour plus d'informations, consultez la section Chiffrer votre catalogue de données dans le guide du AWS Glue développeur.

Pour ajouter un emplacement Amazon S3 en tant que stockage dans votre lac de données, vous devez enregistrer l'emplacement auprès de AWS Lake Formation. Vous pouvez ensuite utiliser les autorisations de Lake Formation pour un contrôle d'accès précis aux AWS Glue Data Catalog objets pointant vers cet emplacement et aux données sous-jacentes de cet emplacement.

Lake Formation prend en charge l'enregistrement d'un emplacement Amazon S3 contenant des données chiffrées. Pour plus d'informations, voir Enregistrement d'un emplacement Amazon S3 chiffré.

Sécurité de l'infrastructure dans AWS Lake Formation

En tant que service géré, AWS Lake Formation il est protégé par les procédures de sécurité du réseau AWS mondial décrites dans le livre blanc Amazon Web Services : présentation des processus de sécurité.

Vous utilisez des appels d'API AWS publiés pour accéder à Lake Formation via le réseau. Les clients doivent supporter le protocole TLS (Sécurité de la couche transport) 1.0 ou une version ultérieure.

Nous recommandons TLS 1.2 ou version ultérieure. Les clients doivent aussi prendre en charge les suites de chiffrement PFS (Perfect Forward Secrecy) comme Ephemeral Diffie-Hellman (DHE) ou Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). La plupart des systèmes modernes tels que Java 7 et les versions ultérieures prennent en charge ces modes.

En outre, les demandes doivent être signées à l'aide d'un ID de clé d'accès et d'une clé d'accès secrète associée à un principal IAM. Vous pouvez également utiliser AWS Security Token Service (AWS STS) pour générer des informations d'identification de sécurité temporaires et signer les demandes.

Prévention du problème de l'adjoint confus entre services

Le problème de député confus est un problème de sécurité dans lequel une entité qui n'est pas autorisée à effectuer une action peut contraindre une entité plus privilégiée à le faire. En AWS, l'usurpation d'identité interservices peut entraîner la confusion des adjoints. L'usurpation d'identité entre services peut se produire lorsqu'un service (le service appelant) appelle un autre service (le service appelé). Le service appelant peut être manipulé et ses autorisations utilisées pour agir sur les ressources d'un autre client auxquelles on ne serait pas autorisé d'accéder autrement. Pour éviter cela, AWS fournit des outils qui vous aident à protéger vos données pour tous les services auprès des principaux fournisseurs de services qui ont obtenu l'accès aux ressources de votre compte.

Nous vous recommandons d'utiliser les clés de contexte de condition globale `aws:SourceArn` et `aws:SourceAccount` dans les politiques de ressources afin de limiter les autorisations à la ressource octroyées par AWS Lake Formation à un autre service. Si vous utilisez les deux clés de contexte de condition globale, la valeur `aws:SourceAccount` et le compte de la valeur `aws:SourceArn` doit utiliser le même ID de compte lorsqu'il est utilisé dans la même déclaration de stratégie.

Actuellement, Lake Formation ne prend `aws:SourceArn` en charge que les formats suivants :

```
arn:aws:lakeformation:aws-region:account-id:*
```

L'exemple suivant montre comment vous pouvez utiliser les clés contextuelles `aws:SourceArn` et les clés de contexte de condition `aws:SourceAccount` globale dans Lake Formation pour éviter le problème de confusion des adjoints.

```
{
```

```
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "ConfusedDeputyPreventionExamplePolicy",
      "Effect": "Allow",
      "Principal": {
        "Service": "lakeformation.amazonaws.com"
      },
      "Action": [
        "sts:AssumeRole"
      ],
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "account-id"
        },
        "ArnEquals": {
          "aws:SourceArn": "arn:aws:lakeformation:aws-region:account-id:*"
        }
      }
    }
  ]
}
```

Connexion aux événements de sécurité AWS Lake Formation

AWS Lake Formation est intégré à AWS CloudTrail un service qui fournit un enregistrement des actions entreprises par un utilisateur, un rôle ou un AWS service dans Lake Formation. CloudTrail capture tous les appels d'API pour Lake Formation sous forme d'événements. Les appels capturés incluent des appels provenant de la console Lake Formation AWS Command Line Interface, des appels et des appels de code vers les opérations de l'API Lake Formation.

Pour plus d'informations sur la journalisation des événements dans Lake Formation, consultezAppels d'API Logging AWS Lake Formation à l'aide de AWS CloudTrail.

> **Note**
>
> `GetTableObjectsUpdateTableObjects`, et `GetWorkUnitResults` sont des opérations de plan de données à volume élevé. Les appels à ces API ne sont actuellement pas enregistrés CloudTrail. Pour plus d'informations sur les opérations du plan de données dans CloudTrail, consultez la section Enregistrement des événements liés aux données pour les sentiers dans le guide de AWS CloudTrail l'utilisateur.

Les modifications de la Lake Formation destinées à soutenir d'autres CloudTrail événements seront documentées à l'adresseHistorique du document pour AWS Lake Formation.

Intégration de services tiers avec Lake Formation

L'intégration AWS Lake Formation permet aux services tiers d'accéder en toute sécurité aux données de leurs lacs de données basés sur Amazon S3. Vous pouvez utiliser Lake Formation comme moteur d'autorisation pour gérer ou appliquer les autorisations d'accès à votre lac de données grâce à AWS des services intégrés tels qu'Amazon Athena, Amazon EMR et Redshift Spectrum. Lake Formation propose deux options pour intégrer les services :

1. Paramètres d'intégration de l'application Lake Formation : Lake Formation peut vendre des informations d'identification temporaires limitées sous forme de jetons AWS STS aux sites Amazon S3 enregistrés en fonction des autorisations effectives, afin que les applications autorisées puissent accéder aux données pour le compte des utilisateurs.
2. Application centralisée : les opérations d'API d'interrogation de Lake Formation récupèrent les données d'Amazon S3 et filtrent les résultats en fonction des autorisations effectives. Le moteur ou l'application qui s'intègre au fonctionnement de l'API de requête peut dépendre de Lake Formation pour évaluer les autorisations de l'identité appelante et filtrer les données en toute sécurité en fonction de ces autorisations. Les moteurs de requêtes tiers ne voient et ne fonctionnent que sur les données filtrées.

Rubriques

- Utilisation de l'intégration de l'application Lake Formation

Utilisation de l'intégration de l'application Lake Formation

Lake Formation permet à des services tiers de s'intégrer à Lake Formation et d'obtenir un accès temporaire aux données Amazon S3 pour le compte de leurs utilisateurs en utilisant GetTemporaryGlueTableCredentialset en effectuant GetTemporaryGluePartitionCredentialsdes opérations. Cela permet aux services tiers d'utiliser la même fonctionnalité d'autorisation et de vente d'informations d'identification que les autres services d' AWS analyse. Cette section décrit comment utiliser ces opérations d'API pour intégrer un moteur de requête tiers àLake Formation.

Ces opérations d'API sont désactivées par défaut. Il existe deux options pour autoriser Lake Formation à intégrer des applications :

- Configurer les balises de session IAM qui sont validées chaque fois que les opérations de l'API d'intégration de l'application sont appelées

Pour plus d’informations, consultez Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications.

- Activez l'option qui permet aux moteurs externes d'accéder aux données des emplacements Amazon S3 avec un accès complet aux tables

  Cette option permet aux moteurs de requête et aux applications d'obtenir des informations d'identification sans balises de session IAM si l'utilisateur dispose d'un accès complet à la table. Il fournit des avantages en termes de performances aux moteurs de requêtes et aux applications, tout en simplifiant l'accès aux données. Amazon EMR sur Amazon EC2 est en mesure de tirer parti de ce paramètre.

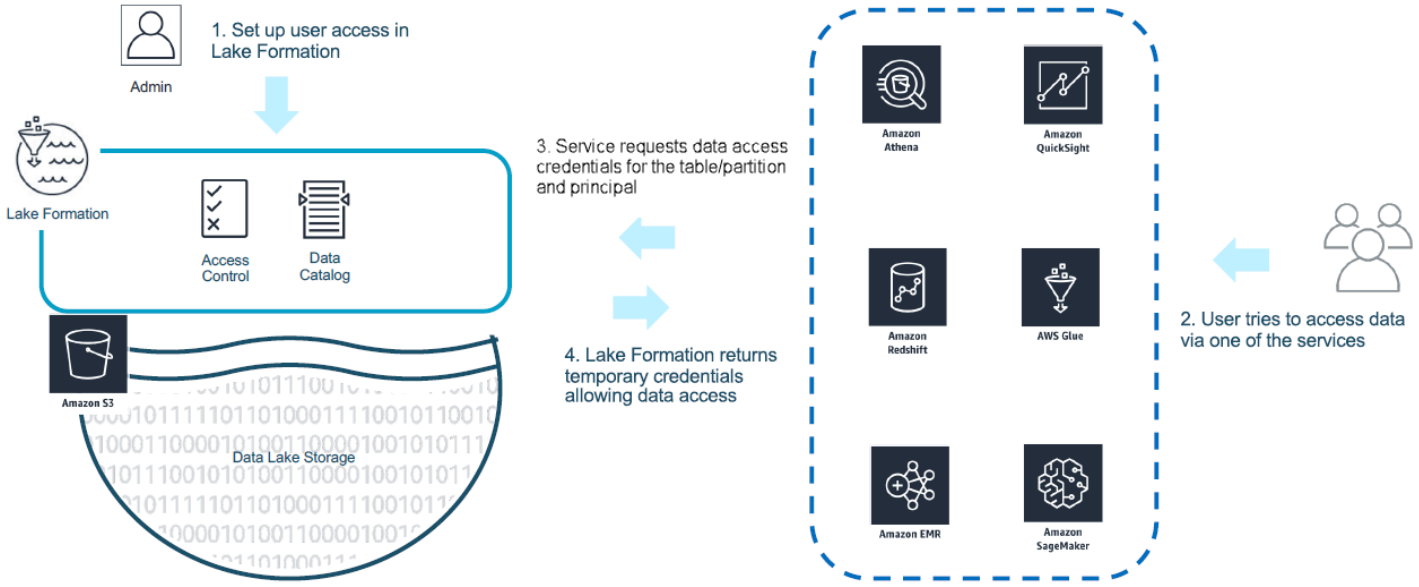
  Pour plus d’informations, consultez Intégration des applications pour un accès complet aux tables .

Rubriques

- Comment fonctionne l'intégration de l'application Lake Formation
- Rôles et responsabilités dans l'intégration de l'application Lake Formation
- Lake Formationflux de travail pour les opérations d'API d'intégration d'applications
- Enregistrement d'un moteur de requête tiers
- Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications
- Intégration des applications pour un accès complet aux tables

Comment fonctionne l'intégration de l'application Lake Formation

Cette section décrit comment utiliser les opérations de l'API d'intégration d'applications pour intégrer une application tierce (moteur de requête) àLake Formation.



1. L'Lake Formationadministrateur exécute les activités suivantes :
   - Enregistre un site Amazon S3 auprès de Lake Formation en fournissant un rôle IAM (utilisé pour les informations d'identification automatiques) doté des autorisations appropriées pour accéder aux données du site Amazon S3
   - Enregistre une application tierce pour pouvoir appeler les opérations de l'API de vente d'informations d'identification de Lake Formation. Consultez the section called “Enregistrement d'un moteur de requête tiers”.
   - Accorde aux utilisateurs des autorisations d'accès aux bases de données et aux tables

     Par exemple, si vous souhaitez publier un ensemble de données de sessions utilisateur comprenant des colonnes contenant des informations personnelles identifiables (PII), pour restreindre l'accès, vous devez attribuer à ces colonnes une balise LF-TBAC nommée « classification » avec la valeur « sensible ». Vous définissez ensuite une autorisation qui permet à un analyste commercial d'accéder aux données des sessions utilisateur, mais vous excluez les colonnes étiquetées avec classification = sensitive.
2. Un principal (utilisateur) soumet une requête à un service intégré.
3. L'application intégrée envoie la demande à Lake Formation pour demander les informations de la table et les informations d'identification pour accéder à la table.
4. Si le principal demandeur est autorisé à accéder à la table, Lake Formation renvoie les informations d'identification à l'application intégrée, qui autorise l'accès aux données.

> **Note**
>
> Lake Formation n'accède pas aux données sous-jacentes lors de la vente d'informations d'identification.

5. Le service intégré lit les données d'Amazon S3, filtre les colonnes en fonction des politiques reçues et renvoie les résultats au principal.

> **Important**
>
> Lake Formationles opérations de l'API de vente d'informations d'identification permettent une application distribuée avec un modèle de refus explicite en cas d'échec (fermeture en cas d'échec). Cela introduit un modèle de sécurité tripartite entre les clients, les services tiers et Lake Formation. Les services intégrés sont fiables pour appliquer correctement Lake Formation les autorisations (application distribuée).

Le service intégré est chargé de filtrer les données lues depuis Amazon S3 en fonction des politiques renvoyées Lake Formation avant que les données filtrées ne soient renvoyées à l'utilisateur. Les services intégrés suivent un modèle de fermeture automatique, ce qui signifie qu'ils doivent échouer à la requête s'ils ne sont pas en mesure d'appliquer les autorisations requisesLake Formation.

Rôles et responsabilités dans l'intégration de l'application Lake Formation

Rôle	Responsabilité
Le client	• Activez le paramètre d'intégration de l'application Lake Formation (voirthe section called “Enregistrement d'un moteur de requête tiers”). • Enregistre explicitement les tiers agréés auprès de Lake Formation (voirthe section called “Enregistrement d'un moteur de requête tiers”). • Teste et valide les solutions tierces avec les autorisations de Lake Formation. • Surveille et audite l'utilisation par des tiers des opérations de l'API de distribution automatique d'informations d'identification de Lake Formation.

Rôle	Responsabilité
Le tiers	• Documente publiquement les fonctionnalités prises en charge pour chaque révision logicielle et fournit des instructions pour les activer correctement. • Annonce avec précision les fonctionnalités prises en charge lors de l'appel aux opérations de l'API de vente d'informations d'identification de Lake Formation (conformément à la documentation). • Stocke et gère en toute sécurité les informations d'identification vendues afin d'éviter les fuites d'informations d'identification et l'augmentation des privilèges. • Applique les autorisations en fonction des fonctionnalités prises en charge et renvoie uniquement les données filtrées aux utilisateurs • Échoue la requête lorsqu'il est impossible d'appliquer correctement les autorisations requises
`AWS Lake Formation`	• Dérive et renvoie correctement les autorisations effectives pour un principal donné. • Valide les fonctionnalités prises en charge par des tiers sur la call-by-call base du fonctionnement de l'API. • Renvoie des informations d'identification IAM limitées uniquement lorsque les fonctionnalités annoncées du moteur correspondent à celles définies dans les ressources du catalogue, sinon renvoie une erreur.

Lake Formationflux de travail pour les opérations d'API d'intégration d'applications

Le flux de travail pour les opérations d'API d'intégration d'applications est le suivant :

1. Un utilisateur soumet une requête ou une demande de données à l'aide d'un moteur de requête tiers intégré. Le moteur de requête assume un rôle IAM qui représente l'utilisateur ou un groupe d'utilisateurs et récupère des informations d'identification fiables à utiliser lors de l'appel des opérations de l'API d'intégration de l'application.

2. Le moteur de requête appelle`GetUnfilteredTableMetadata`, et s'il s'agit d'une table partitionnée, le moteur de requête appelle `GetUnfilteredPartitionsMetadata` pour récupérer les métadonnées et les informations de politique à partir du catalogue de données.
3. Lake Formation autorise la demande. Si l'utilisateur ne dispose pas des autorisations appropriées sur la table, il AccessDeniedExceptionest renvoyé.
4. Dans le cadre de la demande, le moteur de requête envoie le filtrage qu'il prend en charge. Deux indicateurs peuvent être envoyés dans un tableau : COLUMN_PERMISSIONS et CELL_FILTER_PERMISSION. Si le moteur de requête ne prend en charge aucune de ces fonctionnalités et qu'une politique existe sur la table pour la fonctionnalité, un PermissionTypeMismatchExceptionest émis et la requête échoue. Cela permet d'éviter les fuites de données.
5. La réponse renvoyée contient les éléments suivants :
   - Le schéma complet de la table afin que les moteurs de requête puissent l'utiliser pour analyser les données stockées.
   - Liste des colonnes autorisées auxquelles l'utilisateur a accès. Si la liste des colonnes autorisées est vide, cela indique que l'utilisateur dispose d'`DESCRIBE`autorisations, mais pas d'`SELECT`autorisations, et la requête échoue.
   - Un drapeau`IsRegisteredWithLakeFormation`, qui indique si Lake Formation peut fournir des informations d'identification à ces données de ressources. Si le résultat est faux, les informations d'identification des clients doivent être utilisées pour accéder à Amazon S3.
   - Une liste indiquant, le `CellFilters` cas échéant, à appliquer aux lignes de données. Cette liste contient des colonnes et une expression permettant d'évaluer chaque ligne. Ce champ ne doit être renseigné que si CELL_FILTER_PERMISSION est envoyé dans le cadre de la demande et s'il existe un filtre de données dans la table pour l'utilisateur appelant.
6. Une fois les métadonnées récupérées, le moteur de requête appelle `GetTemporaryGlueTableCredentials` ou `GetTemporaryGluePartitionCredentials` pour obtenir des AWS informations d'identification afin de récupérer les données depuis l'emplacement Amazon S3.
7. Le moteur de requête lit les objets pertinents depuis Amazon S3, filtre les données en fonction des politiques reçues à l'étape 2 et renvoie les résultats à l'utilisateur.

Les opérations de l'API d'intégration d'applications Lake Formation contiennent du contenu supplémentaire pour configurer l'intégration avec des moteurs de requêtes tiers. Vous pouvez consulter les détails des opérations dans la section Opérations de l'API Credential vending.

Enregistrement d'un moteur de requête tiers

Avant qu'un moteur de requête tiers puisse utiliser les opérations de l'API d'intégration d'applications, vous devez explicitement autoriser le moteur de requêtes à appeler les opérations d'API en votre nom. Cela se fait en quelques étapes :

1. Vous devez spécifier les AWS comptes et les balises de session IAM qui nécessitent une autorisation pour appeler les opérations de l'API d'intégration des applications via la AWS Lake Formation console, l'API/SDK AWS CLI ou l'API/le.
2. Lorsque le moteur de requête tiers assume le rôle d'exécution dans votre compte, il doit associer une balise de session enregistrée auprès de Lake Formation représentant le moteur tiers. Lake Formationutilise cette balise pour le valider si la demande provient d'un moteur approuvé. Pour plus d'informations sur les balises de session, consultez la section Balises de session dans le guide de l'utilisateur IAM.
3. Lorsque vous configurez un rôle d'exécution de moteur de requête tiers, vous devez disposer des autorisations minimales suivantes dans la politique IAM :

```
{
  "Version": "2012-10-17",
  "Statement": {"Effect": "Allow",
    "Action": [
      "lakeformation:GetDataAccess",
      "glue:GetTable",
      "glue:GetTables",
      "glue:GetDatabase",
      "glue:GetDatabases",
      "glue:CreateDatabase",
      "glue:GetUserDefinedFunction",
      "glue:GetUserDefinedFunctions",
      "glue:GetPartition",
      "glue:GetPartitions"
    ],
    "Resource": "*"
  }
}
```

4. Configurez une politique d'approbation des rôles sur le rôle d'exécution du moteur de requête afin de contrôler avec précision la paire clé-valeur de balise de session qui peut être attachée à ce rôle. Dans l'exemple suivant, ce rôle est uniquement autorisé à associer une clé de balise de

session "LakeFormationAuthorizedCaller" et une valeur de balise "engine1" de session, et aucune autre paire clé-valeur de balise de session n'est autorisée.

```
{
    "Sid": "AllowPassSessionTags",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
    },
    "Action": "sts:TagSession",
    "Condition": {
    "StringLike": {
        "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"        }
    }
}
```

Lorsque LakeFormationAuthorizedCaller l'opération STS : AssumeRole API est appelée pour récupérer les informations d'identification à utiliser par le moteur de requête, le tag de session doit être inclus dans la AssumeRole demande. Les informations d'identification temporaires renvoyées peuvent être utilisées pour effectuer des demandes d'API d'intégration d'Lake Formationapplications.

Lake Formationles opérations de l'API d'intégration d'applications nécessitent que le principal appelant soit un rôle IAM. Le rôle IAM doit inclure une balise de session avec une valeur prédéterminée enregistrée avecLake Formation. Cette balise permet Lake Formation de vérifier que le rôle utilisé pour appeler les opérations de l'API d'intégration de l'application est autorisé à le faire.

Activation des autorisations permettant à un moteur de requête tiers d'appeler des opérations d'API d'intégration d'applications

Procédez comme suit pour autoriser un moteur de requête tiers à appeler les opérations de l'API d'intégration d'applications via la AWS Lake Formation console, l'API/SDK AWS CLI ou l'API/le.

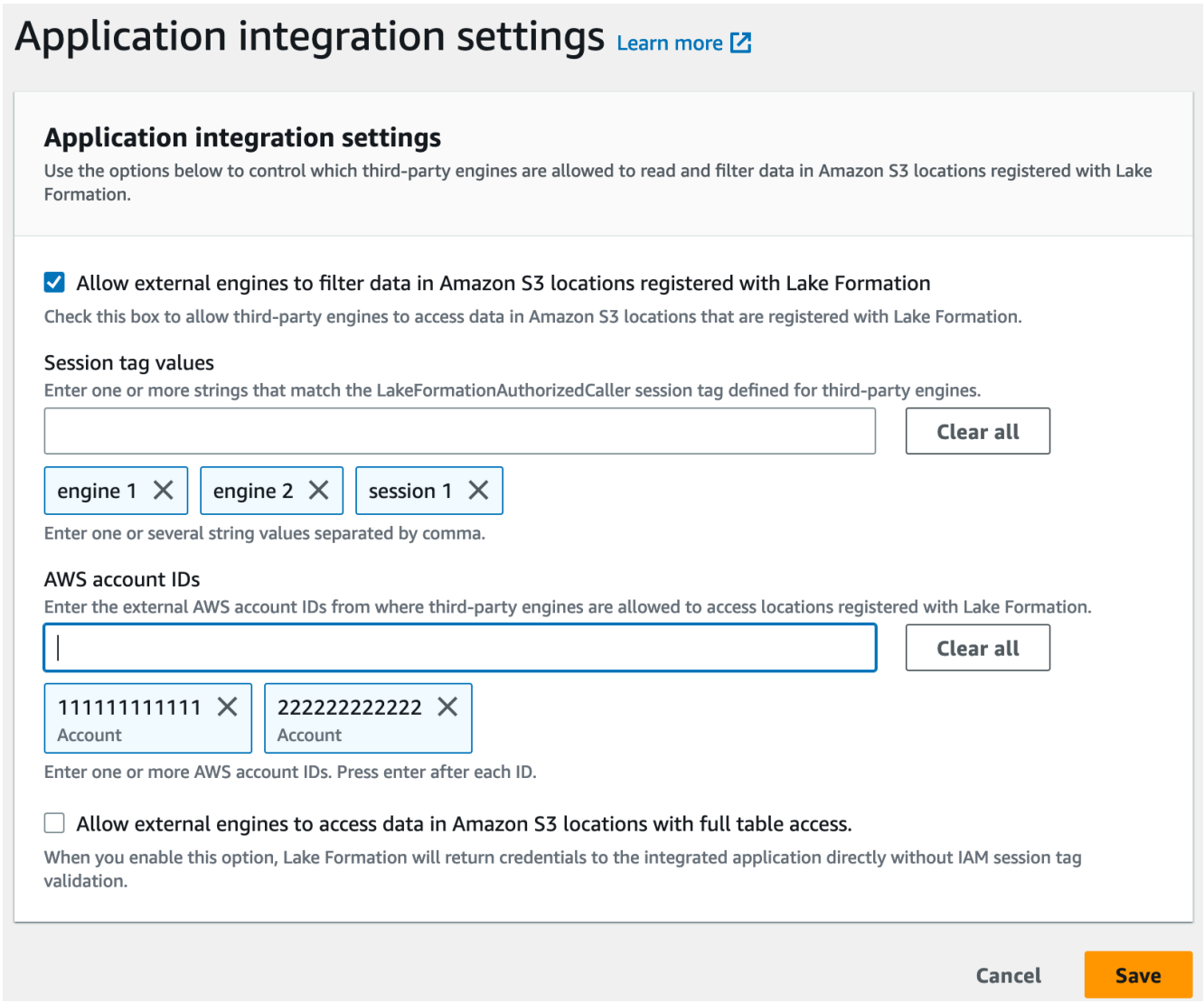
Console

Pour enregistrer votre compte pour le filtrage externe des données, procédez comme suit :

1. Connectez-vous à la AWS Management Console console Lake Formation et ouvrez-la à l'adresse https://console.aws.amazon.com/lakeformation/.

2. Dans le volet de navigation de gauche, développez Administration, puis choisissez Paramètre d'intégration des applications.
3. Sur la page des paramètres d'intégration des applications, choisissez l'option Autoriser les moteurs externes à filtrer les données dans les sites Amazon S3 enregistrés auprès deLake Formation.
4. Entrez les balises de session que vous avez créées pour le moteur tiers. Pour plus d'informations sur les balises de session, voir Transmission de balises de session dans AWS STS dans le Guide de AWS Identity and Access Management l'utilisateur.
5. Entrez les identifiants de compte des utilisateurs qui peuvent utiliser le moteur tiers pour accéder aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données des ressources du compte courant.

   Vous pouvez également utiliser le champ ID du AWS compte pour configurer l'accès entre comptes.



CLI

Utilisez la commande `put-data-lake-settings` CLI pour définir les paramètres suivants.

Trois champs sont à configurer lors de l'utilisation de cette AWS CLI commande :

- `allow-external-data-filtering` — (booléen) Indique qu'un moteur tiers peut accéder aux informations de métadonnées non filtrées et aux informations d'accès aux données des ressources du compte courant.
- `external-data-filtering-allow-list`— (tableau) Liste des identifiants de compte qui peuvent accéder aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données des ressources du compte courant lors de l'utilisation d'un moteur tiers.

- `authorized-sessions-tag-value-list`— (tableau) Liste des valeurs de balises de session autorisées (chaînes). Si un identifiant de rôle IAM a été associé à une paire clé-valeur autorisée, si le tag de session est inclus dans la liste, la session a accès aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données sur les ressources du compte configuré. La clé de balise de session autorisée est définie comme`*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess`- (booléen) S'il faut autoriser un moteur de requête tiers à obtenir des informations d'accès aux données sans balises de session lorsqu'un appelant dispose d'autorisations d'accès complètes aux données.

Par exemple :

```
aws lakeformation put-data-lake-settings --cli-input-json file://
datalakesettings.json

{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/lakeAdmin"
      }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "TrustedResourceOwners": [],
    "AllowExternalDataFiltering": true,
    "ExternalDataFilteringAllowList": [
        {"DataLakePrincipalIdentifier": "111111111111"}
        ],
    "AuthorizedSessionTagValueList": ["engine1"]
    }
    "AllowFullTableExternalDataAccess": false

}
```

API/SDK

Utilisez l'opération `PutDataLakeSetting` API pour définir les paramètres suivants.

Trois champs doivent être configurés lors de l'utilisation de cette opération d'API :

- `AllowExternalDataFiltering`— (Booléen) Indique si un moteur tiers peut accéder aux informations de métadonnées non filtrées et aux informations d'accès aux données des ressources du compte courant.
- `ExternalDataFilteringAllowList`— (tableau) Liste des identifiants de compte qui peuvent accéder aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données des ressources du compte courant à l'aide d'un moteur tiers.
- `AuthorizedSectionsTagValueList`— (tableau) Liste des valeurs de balises autorisées (chaînes). Si un identifiant de rôle IAM a été associé à une balise autorisée, la session a accès aux informations de métadonnées non filtrées et aux informations d'identification d'accès aux données sur les ressources du compte configuré. La clé de balise de session autorisée est définie comme`*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess`- (booléen) S'il faut autoriser un moteur de requête tiers à obtenir des informations d'accès aux données sans balises de session lorsqu'un appelant dispose d'autorisations d'accès complètes aux données.

Par exemple :

```
//Enable session tag on existing data lake settings
public void sessionTagSetUpForExternalFiltering(AWSLakeFormationClient
 lakeformation) {
    GetDataLakeSettingsResult getDataLakeSettingsResult =
 lfClient.getDataLakeSettings(new GetDataLakeSettingsRequest());
    DataLakeSettings dataLakeSettings =
 getDataLakeSettingsResult.getDataLakeSettings();

    //set account level flag to allow external filtering
    dataLakeSettings.setAllowExternalDataFiltering(true);

    //set account that are allowed to call credential vending or Glue
 GetFilteredMetadata API
    List<DataLakePrincipal> allowlist = new ArrayList<>();
    allowlist.add(new
 DataLakePrincipal().withDataLakePrincipalIdentifier("111111111111"));
    dataLakeSettings.setWhitelistedForExternalDataFiltering(allowlist);

    //set registered session tag values
    List<String> registeredTagValues = new ArrayList<>();
    registeredTagValues.add("engine1");
```

```
    dataLakeSettings.setAuthorizedSessionTagValueList(registeredTagValues);

    lakeformation.putDataLakeSettings(new
 PutDataLakeSettingsRequest().withDataLakeSettings(dataLakeSettings));
}
```

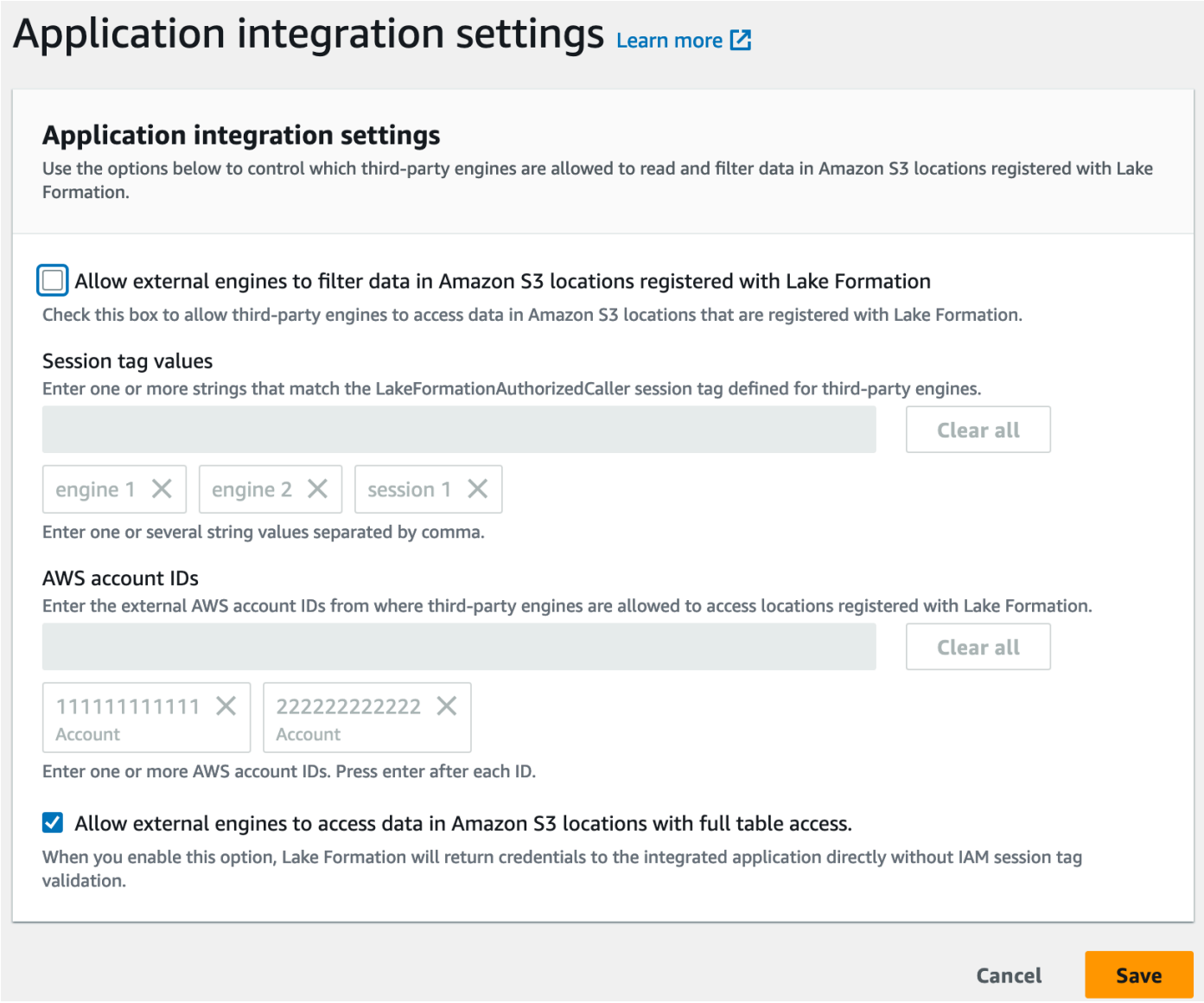
Intégration des applications pour un accès complet aux tables

Procédez comme suit pour permettre aux moteurs de requêtes tiers d'accéder aux données sans validation du tag de session IAM :

Console

1. Connectez-vous à la console Lake Formation à l'adresse https://console.aws.amazon.com/lakeformation/.
2. Dans le volet de navigation de gauche, développez Administration, puis sélectionnez Paramètres d'intégration des applications.
3. Sur la page des paramètres d'intégration des applications, choisissez l'option Autoriser les moteurs externes à accéder aux données dans les emplacements Amazon S3 avec accès complet aux tables.

   Lorsque vous activez cette option, Lake Formation renvoie les informations d'identification à l'application demandeuse directement sans validation des balises de session IAM.



AWS CLI

Utilisez la commande `put-data-lake-settings` CLI pour définir le `AllowFullTableExternalDataAccess` paramètre.

```
aws lakeformation put-data-lake-settings –cli-input-json file://put-data-lake-
settings.json –region ap-northeast-1
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111111111111:user/
lakeAdmin"
            }
```

```
        ],
        "AllowFullTableExternalDataAccess": true
    }
}
```

Collaboration avec d'autres AWS services

AWS des services tels qu'Amazon Athena AWS Glue, Amazon Redshift Spectrum et Amazon EMR peuvent être AWS Lake Formation utilisés pour accéder en toute sécurité aux données des sites Amazon S3 enregistrés auprès de Lake Formation. Avec Lake Formation, vous pouvez définir et gérer des autorisations de contrôle d'accès détaillées (FGAC) pour vos tables dans le. AWS Glue Data Catalog Chacun de ces AWS services est un interlocuteur fiable de Lake Formation, et Lake Formation fournit un accès aux données stockées dans Amazon S3 via des informations d'identification temporaires. Pour plus d'informations, consultez Comment fonctionne l'intégration de l'application Lake Formation.

Pour bénéficier de ces fonctionnalités, Lake Formation vous demande d'abord d'enregistrer l'emplacement Amazon S3 et d'attribuer les autorisations appropriées au principal IAM pour accéder à la table, à la base de données et à l'emplacement Amazon S3. Pour plus d'informations, veuillez consulter Gestion des autorisations relatives à Lake Formation.

Les tableaux suivants répertorient les types d'autorisations Lake Formation pris en charge par Amazon Athena, Amazon EMR et AWS Glue Amazon Redshift Spectrum pour accéder aux données à AWS Glue partir de tables standard et de tables transactionnelles (Apache Iceberg, Apache Hudi et Linux foundation Delta Lake) avec des données stockées dans Amazon S3 et des métadonnées de table dans le catalogue de données.

AWS services et types d'autorisations pris en charge pour les tables et les vues AWS Glue standard

AWS service	Autorisations au niveau du tableau	Autorisations au niveau des colonnes	Autorisations au niveau des lignes et des cellules
Athena SQL	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture
Athena Spark	Non pris en charge	Non pris en charge	Non pris en charge
Redshift Spectrum sur un cluster provisionné ou Amazon Redshift sans serveur	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture

AWS service	Autorisations au niveau du tableau	Autorisations au niveau des colonnes	Autorisations au niveau des lignes et des cellules
Apache Spark sur Amazon EMR (EC2)	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture
Apache Hive sur Amazon EMR (EC2)	Accès en lecture/écriture	Accès à la lecture	Non pris en charge
Apache Spark sur EMR sans serveur*	Accès en lecture/écriture	Accès à la lecture	Accès à la lecture
Apache Hive sur EMR sans serveur	Non pris en charge	Non pris en charge	Non pris en charge
Amazon EMR on EKS	Non pris en charge	Non pris en charge	Non pris en charge
Glue ETL	Accès en lecture/écriture	Non pris en charge	Non pris en charge

Considérations et restrictions

- Athena Spark ne prend pas en charge l'interrogation des tables du catalogue de données avec les autorisations Lake Formation.
- Les utilisateurs basés sur Athena SAML peuvent lire les sources de données sécurisées à l'aide des autorisations Lake Formation en activant la fédération basée sur SAML 2.0. Les utilisateurs de SAML peuvent insérer des données dans les tables Parquet.
- Apache Spark sur EMR Serverless ne prend pas en charge l'interrogation des vues du catalogue de données.
- Apache Hive sur EMR Serverless ne prend pas en charge l'interrogation de tables avec les autorisations Lake Formation.
- AWS Glue L'ETL nécessite un accès complet à l'intégralité de la table lors de la récupération des données depuis l'emplacement Amazon S3 sous-jacent. AWS Glue La tâche ETL échoue si vous appliquez des autorisations au niveau des colonnes sur une table.

AWS services et types d'autorisation pris en charge pour les formats de tables transactionnels

AWS service	Iceberg	Hudi	Delta Lake (natif)	Lac Delta (tableaux à liens symboliques)
Athena SQL	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture nécessitent un accès complet à la table.	Prend en charge les opérations de lecture et de création sur les tables avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.	Athena (version 3 du moteur) prend en charge la lecture des tables natives de Delta Lake avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.	Athena (version 3 du moteur) prend en charge la lecture des tables Delta Lake par liens symboliques avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.
Redshift Spectrum sur un cluster provisionné	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont	Non pris en charge	Permet de lire les tables de Delta Lake via un manifeste de liens symboliques avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les

AWS service	Iceberg	Hudi	Delta Lake (natif)	Lac Delta (tableaux à liens symboliques)
	pas prises en charge.	pas prises en charge.		opérations d'écriture ne sont pas prises en charge.
Apache Spark sur Amazon EMR (EC2)	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opération s d'écriture nécessitent un accès complet à la table.	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opération s d'écriture nécessitent un accès complet à la table.	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opérations d'écriture ne sont pas prises en charge.	Permet de lire des tableaux avec des autorisations au niveau des tables, des colonnes, des lignes et des cellules. Les opération s d'écriture nécessitent un accès complet à la table.
AWS Glue	Supporte la lecture/é criture sur les tables avec des autorisations au niveau des tables.	Supporte la lecture/é criture sur les tables avec des autorisations au niveau des tables.	Supporte la lecture/é criture sur les tables avec des autorisations au niveau des tables.	Supporte la lecture/é criture sur les tables avec des autorisations au niveau des tables.

Rubriques

- Utilisation AWS Lake Formation avec Amazon Athena
- Utilisation AWS Lake Formation avec Amazon Redshift Spectrum
- Utilisation AWS Lake Formation avec AWS Glue

- Utilisation AWS Lake Formation avec Amazon EMR
- Utilisation AWS Lake Formation avec Amazon QuickSight
- Utilisation AWS Lake Formation avec AWS CloudTrail Lake

Utilisation AWS Lake Formation avec Amazon Athena

Amazon Athena est un service de requête sans serveur qui vous aide à analyser les données structurées, semi-structurées et non structurées stockées dans Amazon S3. Vous pouvez utiliser Athena SQL pour interroger des données aux formats de données CSV, JSON, Parquet et Avro. Athena SQL prend également en charge les formats de table tels qu'Apache Hive, Apache Hudi et Apache Iceberg. Athena s'intègre au système AWS Glue Data Catalog pour stocker les métadonnées de vos ensembles de données dans Amazon S3. Athena peut utiliser Lake Formation pour définir et maintenir des politiques de contrôle d'accès sur ces ensembles de données.

Voici quelques cas d'utilisation courants dans lesquels vous pouvez utiliser Lake Formation avec Athena.

- Utilisez les autorisations de Lake Formation pour accéder aux ressources du catalogue de données (base de données et tables) depuis Athena. Vous pouvez utiliser la méthode de ressource nommée ou des balises LF pour définir des autorisations sur la base de données et les tables. Pour plus d'informations, consultez :
  - Octroi d'autorisations de base de données à l'aide de la méthode de ressource
  - Contrôle d'accès basé sur des balises Lake Formation

> **Note**
>
> Les autorisations de Lake Formation s'appliquent uniquement lorsque vous utilisez Athena SQL pour interroger les données sources d'Amazon S3 et les métadonnées du catalogue de données.
> Athena Spark ne prend pas en charge l'interrogation des tables du catalogue de données avec les autorisations Lake Formation. Les autorisations de Lake Formation prennent en charge les opérations de lecture et d'écriture sur les bases de données et les tables.

> **Note**
>
> Vous ne pouvez pas appliquer de filtres de données lorsque vous utilisez des balises LF pour gérer les autorisations sur les ressources du catalogue de données.

- Contrôlez les résultats des requêtes en Filtres de données dans Lake Formation sécurisant les tables de vos lacs de données Amazon S3 en accordant des autorisations au niveau des colonnes, des lignes et des cellules. Consultez les limites relatives à la projection de partitions dans le guide de l'utilisateur d'Amazon Athena.
- Appliquez un contrôle d'accès précis aux données mises à la disposition de l'utilisateur Athena basé sur SAML lors de l'exécution de requêtes fédérées.

  Les pilotes JDBC et ODBC Athena prennent en charge la configuration de l'accès fédéré à votre source de données à l'aide d'un fournisseur d'identité (IdP) basé sur SAML. Utilisez Amazon QuickSight intégré à Lake Formation avec votre rôle IAM existant ou des utilisateurs ou groupes SAML pour visualiser les résultats des requêtes Athena.

  > **Note**
  >
  > Les autorisations de Lake Formation pour les utilisateurs et les groupes SAML ne s'appliquent que lorsque vous soumettez des requêtes à Athena à l'aide du pilote JDBC ou ODBC.

  Pour plus d'informations, consultez la section Utilisation de Lake Formation et des pilotes JDBC et ODBC Athena pour un accès fédéré à Athena.

  > **Note**
  >
  > Actuellement, l'autorisation d'accès aux identités SAML dans Lake Formation n'est pas prise en charge dans les régions suivantes :
  >
  > - Moyen-Orient (Bahreïn) – me-south-1
  > - Asie-Pacifique (Hong Kong) – ap-east-1
  > - Afrique (Le Cap) – af-south-1
  > - Chine (Ningxia) – cn-northwest-1

- Asie-Pacifique (Osaka) – ap-northeast-3

- Partage de données entre comptes dans Lake FormationÀ utiliser pour interroger les tables d'un autre compte.

> **Note**
>
> Pour plus d'informations sur les limites liées à l'utilisation des autorisations de Lake Formation pour`Views`, voir Considérations et limites.

Support pour les formats de tables transactionnels

L'application des autorisations Lake Formation vous permet de sécuriser vos données transactionnelles dans vos lacs de données basés sur Amazon S3. Le tableau ci-dessous répertorie les formats de tables transactionnels pris en charge dans les autorisations Athena et Lake Formation. Lake Formation applique ces autorisations lorsque les utilisateurs d'Athena exécutent leurs requêtes.

Format de table	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Athena
Apache Hudi	Format utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Athena prend en charge les opérations de création et de lecture à l'aide des formats de table Apache Hudi sur les ensembles de données Amazon S3 pour les types de tables Copy on Write (CoW) et Merge On Read (MoR) Hudi. Athena ne prend pas en	Filtrage des données et sécurité au niveau des cellules dans Lake FormationÀ utiliser pour sécuriser la table Hudi à l'aide d'autorisations au niveau des tables, des colonnes, des lignes et des cellules.

Format de table	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Athena
	charge les opérations d'écritur e sur les tables Hudi. Utilisez Athena pour interroge r les ensembles de données Hudi.	
Apache Iceberg	Format de table ouvert qui gère de grandes collectio ns de fichiers sous forme de tables et prend en charge les opérations de lac de données analytiques modernes, telles que les requêtes d'insertion, de mise à jour, de suppression et de voyage dans le temps au niveau des enregistrements. Pour plus d'informations sur la prise en charge des tables Iceberg par Athena, consultez la section Utilisation des tables Iceberg.	Les autorisations au niveau des tables, des colonnes, des lignes et des cellules sont prises en charge. Actuellem ent, Lake Formation ne prend pas en charge la gestion des autorisations sur les opérations d'écriture telles que `VACUUMMERGE`, `UPDATE` et `OPTIMIZE` sur les tables dans Open Table Formats.

Format de table	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Athena
Linux Foundation Delta Lake	Delta Lake est un projet open source qui aide à implément er des architectures de lacs de données modernes généralem ent basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Athena prend en charge les tables Delta Lake créées à l'aide d'une définition de table manifeste basée sur des liens symboliques à AWS Glue Data Catalog partir d'une table Delta Lake. Pour plus d'informations, consultez les tables Crawl Delta Lake à l'aide de AWS Glue crawlers. Athena (version 3 du moteur) prend en charge la lecture des tables natives de Delta Lake. Pour plus d'informations, voir Présentation du support de table natif de Delta Lake avec des AWS Glue crawlers.	Les autorisations au niveau des tables, des colonnes, des lignes et des cellules sont prises en charge pour les tables de liens symboliques et les tables natives de Delta Lake.

Ressources supplémentaires

Articles de blog, vidéos et ateliers

- Interrogez un ensemble de données Apache Hudi dans un lac de données Amazon S3 avec Amazon Athena
- Créez un lac de données Apache Iceberg à l'aide d'Amazon Athena, Amazon EMR et AWS Glue
- Insérer, mettre à jour, supprimer sur Amazon S3 avec Athena et Apache Iceberg
- Atelier Lake Formation sur le contrôle d'accès basé sur des balises LF-Tag sur l'interrogation d'un lac de données.

Utilisation AWS Lake Formation avec Amazon Redshift Spectrum

Amazon Redshift Spectrum vous permet d'interroger et de récupérer des données dans les lacs de données Amazon S3 sans charger de données dans les nœuds du cluster Amazon Redshift.

Redshift Spectrum propose deux méthodes pour enregistrer un catalogue de AWS Glue données externe activé avec Lake Formation.

- Utilisation d'un rôle IAM attaché à un cluster autorisé à accéder au catalogue de données

  Pour créer un rôle IAM, suivez les étapes décrites dans la procédure ci-dessous.

  Pour créer un rôle IAM pour Amazon Redshift à l'aide d'un AWS Glue Data CatalogAWS Lake Formation
- Utilisation d'une identité IAM fédérée configurée pour gérer l'accès aux ressources externes AWS Glue Data Catalog

  Redshift Spectrum permet d'interroger les tables de Lake Formation à l'aide d'identités IAM fédérées. Les identités IAM peuvent être un utilisateur IAM ou un rôle IAM. Pour plus d'informations sur la fédération d'identité IAM dans Redshift Spectrum, consultez Utilisation d'une identité fédérée pour gérer l'accès d'Amazon Redshift aux ressources locales et aux tables externes Redshift Spectrum.

Grâce à l'intégration de Lake Formation à Redshift Spectrum, vous pouvez définir des autorisations de contrôle d'accès au niveau des lignes, des colonnes et des cellules sur les tables une fois vos données enregistrées auprès de Lake Formation.

Pour plus d'informations, voir Utilisation de Redshift Spectrum avec. AWS Lake Formation

Redshift Spectrum prend en charge les lectures ou les `SELECT` requêtes sur les tables de schéma externes gérées par Lake Formation.

Pour plus d'informations, consultez Création de schémas externes pour Redshift Spectrum.

Support pour les types de tables transactionnels

Ce tableau répertorie les formats de tables transactionnels pris en charge dans Redshift Spectrum et les autorisations Lake Formation applicables.

Formats de tableau pris en charge

Format de table	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Redshift Spectrum
Apache Hudi	Format utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Redshift Spectrum prend en charge les opérations d'écritur e d'insertion, de suppression et d'insertion à l'aide du format de table Apache Hudi Copy on Write (CoW) sur Amazon S3. Pour plus d'informations, voir Création de tables externes pour les données gérées dans Apache Hudi.	Filtrage des données et sécurité au niveau des cellules dans Lake FormationÀ utiliser pour sécuriser les tables Hudi à l'aide d'autorisations au niveau des tables, des colonnes, des lignes et des cellules.
Apache Iceberg	Format de table ouvert qui gère de grandes collectio ns de fichiers sous forme de tables et prend en charge les opérations de lac de données	Redshift Spectrum prend en charge les tables Apache Iceberg pour les requêtes.

Format de table	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Redshift Spectrum
	analytiques modernes, telles que les requêtes d'insertion, de mise à jour, de suppression et de voyage dans le temps au niveau des enregistrements. Pour plus d'informations, consultez la section Utilisation des tables Apache Iceberg avec Amazon Redshift.	
Linux Foundation Delta Lake	Delta Lake est un projet open source qui permet de mettre en œuvre des architectures de lacs de données modernes généralement basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Redshift Spectrum permet d'interroger les tables Delta Lake. Pour plus d'informations, voir Création de tables externes pour les données gérées dans Delta Lake.	Les autorisations au niveau des tables, des colonnes, des lignes et des cellules sont prises en charge.

Ressources supplémentaires

Articles de blog et ateliers

- Centralisez la gouvernance de votre lac de données AWS Lake Formation tout en mettant en place une architecture de données moderne avec Amazon Redshift Spectrum

- Utilisez Redshift Spectrum pour interroger les tables Apache HUDI Copy On Write (CoW) dans le lac de données Amazon S3

Utilisation AWS Lake Formation avec AWS Glue

Les ingénieurs des données et DevOps les professionnels utilisent AWS Glue Extract, Transform and Load (ETL) avec Apache Spark pour transformer leurs ensembles de données dans Amazon S3 et charger les données transformées dans des lacs de données et des entrepôts de données à des fins d'analyse, d'apprentissage automatique et de développement d'applications. Différentes équipes accédant au même ensemble de données dans Amazon S3, il est impératif d'accorder et de restreindre les autorisations en fonction de leurs rôles.

AWS Lake Formation est basé sur AWS Glue, et les services interagissent de la manière suivante :

- Lake Formation et AWS Glue partagent le même catalogue de données.
- Les fonctionnalités suivantes de la console Lake Formation invoquent la AWS Glue console :
  - Tâches — Pour plus d'informations, consultez la section Ajouter des tâches dans le Guide du AWS Glue développeur.
  - Crawlers — Pour plus d'informations, voir Catalogage de tables à l'aide d'un robot d'exploration dans le Guide du AWS Glue développeur.
- Les flux de travail générés lorsque vous utilisez un plan de Lake Formation sont des AWS Glue flux de travail. Vous pouvez consulter et gérer ces flux de travail à la fois dans la console Lake Formation et dans la AWS Glue console.
- Les transformations d'apprentissage automatique sont fournies avec Lake Formation et reposent sur des opérations AWS Glue d'API. Vous créez et gérez des transformations de machine learning sur la AWS Glue console. Pour plus d'informations, consultez Machine Learning Transforms dans le Guide du AWS Glue développeur.

Vous pouvez utiliser le contrôle d'accès détaillé de Lake Formation pour gérer les ressources de votre catalogue de données existantes et les emplacements de données Amazon S3.

> **Note**
>
> AWS Glue L'ETL nécessite un accès complet à l'intégralité de la table lors de la récupération des données depuis l'emplacement Amazon S3 sous-jacent. AWS Glue La tâche ETL échoue si vous appliquez des autorisations au niveau des colonnes sur une table. Toutefois, vous

pouvez créer une sécurité au niveau des colonnes et au niveau des lignes en définissant des filtres de données. Pour plus d'informations, voir Remarques et restrictions relatives au filtrage au niveau des colonnes Lake Formation évalue le filtre de données défini dans le tableau et extrait uniquement les données filtrées d'Amazon S3 requises pour la tâche AWS Glue ETL.

Support pour les types de tables transactionnels

L'application des autorisations Lake Formation vous permet de sécuriser vos données transactionnelles dans vos lacs de données basés sur Amazon S3. Le tableau ci-dessous répertorie les formats de tables transactionnels pris en charge dans les autorisations Lake Formation AWS Glue et les autorisations associées. Lake Formation applique ces autorisations pour les AWS Glue opérations.

Formats de tableau pris en charge

Format de table	Description et opérations autorisées	Permissions de Lake Formation prises en charge dans AWS Glue
Apache Hudi	Format de table ouvert utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Pour des exemples, voir Utilisation du framework Hudi dans AWS Glue.	Des autorisations au niveau des tables sont disponibles pour les tables Hudi. Pour plus d'informations, consultez Limites.
Apache Iceberg	Format de tableau ouvert qui gère de grandes collectio ns de fichiers sous forme de tableaux.	Des autorisations au niveau des tables sont disponibles pour les tables Iceberg. Pour plus d'informations, consultez Limites.

Format de table	Description et opérations autorisées	Permissions de Lake Formation prises en charge dans AWS Glue
	Pour des exemples, voir Utilisation du framework Iceberg dans AWS Glue.	
Linux Foundation Delta Lake	Delta Lake est un projet open source qui permet de mettre en œuvre des architectures de lacs de données modernes généralement basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Pour des exemples, voir Utilisation du framework Delta Lake dans AWS Glue.	Des autorisations au niveau des tables sont disponibles pour les tables Delta Lake. Pour plus d'informations, consultez Limites.

Ressources supplémentaires

Articles de blog et référentiels

- Utilisez le AWS Glue connecteur pour lire et écrire des tables Apache Iceberg avec des transactions ACID et voyager dans le temps
- Écrire dans des tables Apache Hudi à l'aide d'un connecteur AWS Glue personnalisé
- AWS référentiel du modèle Cloudformation et d'un exemple de code pyspark pour analyser les données de streaming à l'aide d' AWS Glue Apache Hudi et d'Amazon S3.

Utilisation AWS Lake Formation avec Amazon EMR

Amazon EMR est une plateforme de cluster AWS gérée flexible sur laquelle vous pouvez exécuter n'importe quel code personnalisé sur des frameworks de données compatibles tels que Hadoop Map-Reduce, Spark, Hive, Presto, etc. Organisations utilisent également Amazon EMR pour exécuter des

applications de traitement de données par lots et en flux sur un cluster hautement distribué. À l'aide d'Apache Spark sur Amazon EMR, vous pouvez exécuter vos transformations de données et votre code personnalisé sur une base de données et des tables dont les autorisations sont gérées par Lake Formation.

Il existe trois options pour déployer Amazon EMR :

- EMR sur EC2
- EMR sans serveur
- Amazon EMR on EKS

Pour plus d'informations, consultez Intégrer Amazon EMR à Lake Formation ou Utiliser EMR sans serveur pour un contrôle d'accès précis AWS Lake Formation

Support pour les formats de tables transactionnels

Les versions 6.15.0 et ultérieures d'Amazon EMR incluent la prise en charge des autorisations de contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules de Lake Formation sur les formats de table Apache Hudi, Apache Iceberg et Delta Lake lorsque vous lisez et écrivez des données avec Spark SQL.

Pour connaître les limites, consultez Considérations relatives à Amazon EMR with Lake Formation.

Formats de tableau pris en charge

Format de table	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Amazon EMR
Apache Hudi	Format de table ouvert utilisé pour simplifier le traitement incrémentiel des données et le développement de pipelines de données. Pour une liste des opération s prises en charge, consultez Apache Hudi et Lake Formation.	Amazon EMR prend en charge le contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules avec Apache Hudi.

Format de table	Description et opérations autorisées	Autorisations de Lake Formation prises en charge dans Amazon EMR
Apache Iceberg	Format de tableau ouvert qui gère de grandes collections de fichiers sous forme de tableaux. Pour une liste des opérations prises en charge, consultez Apache Iceberg et Lake Formation.	Amazon EMR prend en charge le contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules avec Apache Iceberg.
Linux Foundation Delta Lake	Delta Lake est un projet open source qui permet de mettre en œuvre des architectures de lacs de données modernes généralement basées sur Amazon S3 ou Hadoop Distributed File System (HDFS). Pour une liste des opérations prises en charge, voir Delta Lake and Lake Formation.	Amazon EMR prend en charge le contrôle d'accès au niveau des tables, des lignes, des colonnes et des cellules avec les tables Delta Lake.

Ressources supplémentaires

Guide de l'utilisateur, articles de blog et ateliers

- Intégration à Amazon EMR à l'aide de rôles d'exécution
- Démarrez rapidement avec Apache Hudi, Apache Iceberg et Delta Lake avec Amazon EMR sur EKS
- Utilisation de Delta Lake OSS avec EMR Serverless

Utilisation AWS Lake Formation avec Amazon QuickSight

Amazon QuickSight prend en charge l'exploration des ensembles de données gérés par les autorisations Lake Formation dans Amazon S3 à l'aide d'Athena.

Les utilisateurs des éditions Standard et Enterprise d'Amazon QuickSight s'intègrent à Lake Formation, mais de manière légèrement différente.

- Édition Enterprise : accordez des autorisations de contrôle d'accès détaillé (FGAC) à des QuickSight utilisateurs individuels, à des groupes et à des rôles IAM d'Amazon pour accéder aux bases de données et aux tables.
- Édition standard : accordez des autorisations aux rôles IAM pour accéder aux bases de données et aux tables.

> **Note**
>
> Par défaut, Amazon QuickSight utilise un rôle nomméaws-quicksight-service-role-v0. Vous pouvez également définir des rôles personnalisés avec les autorisations requises pour permettre QuickSight à Amazon d'accéder à Athena.

Pour plus d'informations, voir Autoriser les connexions via AWS Lake Formation

Ressources supplémentaires

Billets de blogs

- Activez des autorisations détaillées pour les auteurs Amazon QuickSight dans AWS Lake Formation
- Analysez vos données en toute sécurité avec AWS Lake Formation Amazon QuickSight

Utilisation AWS Lake Formation avec AWS CloudTrail Lake

AWS CloudTrail Lake prend en charge l'exploration des magasins de données d'événements Amazon Athena à l'aide d'autorisations détaillées dans. AWS Lake Formation

> **Note**
>
> CloudTrail Le lac ne peut être interrogé que par son intermédiaire. Amazon Athena

Pour enregistrer votre banque de données d'événements CloudTrail Lake auprès de Lake Formation, voir Fédérer une banque de données d'événements.

Appels d'API Logging AWS Lake Formation à l'aide de AWS CloudTrail

AWS Lake Formation est intégré à AWS CloudTrail un service qui fournit un enregistrement des actions entreprises par un utilisateur, un rôle ou un AWS service dans Lake Formation. CloudTrail capture tous les appels de l'API Lake Formation sous forme d'événements. Les appels capturés incluent les appels provenant de la console Lake Formation AWS Command Line Interface, les appels et les appels de code vers les actions de l'API Lake Formation. Si vous créez un suivi, vous pouvez activer la diffusion continue des CloudTrail événements vers un compartiment Amazon S3, y compris les événements relatifs à Lake Formation. Si vous ne configurez pas de suivi, vous pouvez toujours consulter les événements les plus récents dans la CloudTrail console dans Historique des événements. À l'aide des informations collectées par CloudTrail, vous pouvez déterminer la demande qui a été faite à Lake Formation, l'adresse IP à partir de laquelle la demande a été faite, qui a fait la demande, quand elle a été faite et des détails supplémentaires.

Pour en savoir plus CloudTrail, consultez le guide de AWS CloudTrail l'utilisateur.

Informations sur la formation des lacs en CloudTrail

CloudTrail est activé par défaut lorsque vous créez un nouveau AWS compte. Lorsqu'une activité se produit dans Lake Formation, cette activité est enregistrée en tant qu' CloudTrail événement avec d'autres événements de AWS service dans l'historique des événements. Un événement représente une demande émise par une source et comprend des informations sur l'action demandée, la date et l'heure de l'action, et les paramètres de la demande. En outre, chaque événement ou entrée de journal contient des informations sur l'auteur de la demande. Les informations relatives à l’identité permettent de déterminer les éléments suivants :

- Si la demande a été faite avec les informations d'identification de l'utilisateur root ou AWS Identity and Access Management (IAM).
- Si la demande a été effectuée avec les informations d’identification de sécurité temporaires d’un rôle ou d’un utilisateur fédéré.
- Si la demande a été faite par un autre AWS service.

Pour plus d'informations, consultez l'élément CloudTrail UserIdentity.

Vous pouvez consulter, rechercher et télécharger les événements récents pour votre AWS compte. Pour plus d'informations, consultez la section Affichage des événements à l'aide de l'historique des CloudTrail événements.

Pour un enregistrement continu des événements enregistrés sur votre AWS compte, y compris les événements liés à Lake Formation, créez un parcours. Un suivi permet CloudTrail de fournir des fichiers journaux à un compartiment Amazon S3. Par défaut, lorsque vous créez un journal de suivi dans la console, il s'applique à toutes les régions AWS . Le journal enregistre les événements de toutes les régions de la AWS partition et transmet les fichiers journaux au compartiment Amazon S3 que vous spécifiez. En outre, vous pouvez configurer d'autres AWS services, par exemple pour analyser plus en détail les données d'événements collectées dans les CloudTrail journaux et agir en conséquence. Amazon Athena CloudTrail peut également envoyer des fichiers journaux à Amazon CloudWatch Logs and CloudWatch Events.

Pour plus d’informations, consultez les ressources suivantes :

- Présentation de la création d’un journal de suivi
- CloudTrail services et intégrations pris en charge
- Configuration des notifications Amazon SNS pour CloudTrail
- Réception de fichiers CloudTrail journaux de plusieurs régions et réception de fichiers CloudTrail journaux de plusieurs comptes

Comprendre les événements liés à la formation des lacs

Toutes les actions de l'API Lake Formation sont enregistrées CloudTrail et documentées dans le Guide du AWS Lake Formation développeur. Par exemple, les appels aux `PutDataLakeSettings` `GrantPermissions`, et `RevokePermissions` les actions génèrent des entrées dans les fichiers CloudTrail journaux.

L'exemple suivant montre un CloudTrail événement associé à l'`GrantPermissions`action. L'entrée inclut l'utilisateur qui a accordé l'autorisation (`datalake_admin`), le principal auquel l'autorisation a été accordée (`datalake_user1`) et l'autorisation accordée (`CREATE_TABLE`). L'entrée indique également que l'autorisation a échoué car la base de données cible n'a pas été spécifiée dans l'`resource`argument.

```
{
    "eventVersion": "1.08",
    "userIdentity": {
```

```
        "type": "IAMUser",
        "principalId": "AIDAZKE67KM3P775X74U2",
        "arn": "arn:aws:iam::111122223333:user/datalake_admin",
        "accountId": "111122223333",
        "accessKeyId": "...",
        "userName": "datalake_admin"
    },
    "eventTime": "2021-02-06T00:43:21Z",
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GrantPermissions",
    "awsRegion": "us-east-1",
    "sourceIPAddress": "72.21.198.65",
    "userAgent": "aws-cli/1.19.0 Python/3.6.12
 Linux/4.9.230-0.1.ac.223.84.332.metal1.x86_64 botocore/1.20.0",
    "errorCode": "InvalidInputException",
    "errorMessage": "Resource must have one of the have either the catalog, table or
 database field populated.",
    "requestParameters": {
        "principal": {
            "dataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
        },
        "resource": {},
        "permissions": [
            "CREATE_TABLE"
        ]
    },
    "responseElements": null,
    "requestID": "b85e863f-e75d-4fc0-9ff0-97f943f706e7",
    "eventID": "8d2ccef0-55f3-42d3-9ede-3a6faedaa5c1",
    "readOnly": false,
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "eventCategory": "Management",
    "recipientAccountId": "111122223333"
}
```


L'exemple suivant montre une entrée de CloudTrail journal pour l'`GetDataAccess`action. Les directeurs n'appellent pas directement cette API. `GetDataAccess`Il est plutôt enregistré chaque fois qu'un AWS service principal ou intégré demande des informations d'identification temporaires pour accéder aux données d'un lac de données enregistré auprès de Lake Formation.


```
{
```

```
    "eventVersion": "1.05",
    "userIdentity": {
        "type": "AWSAccount",
        "principalId": "AROAQGFTBBBGOBWV2EMZA:GlueJobRunnerSession",
        "accountId": "111122223333"
    },
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GetDataAccess",
...
...
    "additionalEventData": {
        "requesterService": "GLUE_JOB",
        "lakeFormationPrincipal": "arn:aws:iam::111122223333:role/ETL-Glue-Role",
        "lakeFormationRoleSessionName": "AWSLF-00-GL-111122223333-G13T0Rmng2"
    },
...
}
```


consultez aussi

- Journalisation entre comptes CloudTrail

Meilleures pratiques, considérations et limites en matière de formation des lacs

Utilisez cette section pour trouver rapidement les meilleures pratiques, les considérations et les limites qu'elle contient AWS Lake Formation.

Consultez la section Quotas de service pour connaître le nombre maximal de ressources de service ou d'opérations pour votre Compte AWS.

Rubriques

- Meilleures pratiques et considérations relatives au partage de données entre comptes
- Limites d'accès aux données entre régions
- Considérations et limites relatives aux affichages du catalogue de données
- Limites du filtrage des données
- Considérations et limites relatives au mode d'accès hybride
- Considérations et limites relatives au partage des données du magasin de métadonnées Hive
- Limites du partage de données Amazon Redshift
- Limites de l'intégration d'IAM Identity Center
- Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation
- Formats pris en charge et limites pour le compactage géré des données

Meilleures pratiques et considérations relatives au partage de données entre comptes

Les fonctionnalités multi-comptes de Lake Formation permettent aux utilisateurs de partager en toute sécurité des lacs de données distribués entre plusieurs AWS organisations ou directement avec les responsables IAM d'un autre compte Comptes AWS, offrant ainsi un accès détaillé aux métadonnées du catalogue de données et aux données sous-jacentes.

Tenez compte des meilleures pratiques suivantes lorsque vous utilisez le partage de données entre comptes de Lake Formation :

- Il n'y a pas de limite au nombre d'autorisations de Lake Formation que vous pouvez accorder aux directeurs d'école pour votre propre AWS compte. Cependant, Lake Formation utilise AWS Resource Access Manager (AWS RAM) la capacité pour les subventions entre comptes que votre compte peut accorder avec la méthode de ressource nommée. Pour optimiser la AWS RAM capacité, suivez les meilleures pratiques suivantes pour la méthode de ressource nommée :
  - Utilisez le nouveau mode de subvention entre comptes (version 3 et supérieure, sous Paramètres des versions entre comptes) pour partager une ressource avec un externe Compte AWS. Pour plus d'informations, consultez Mise à jour des paramètres de version de partage de données entre comptes.
  - AWS Répartissez les comptes en organisations et accordez des autorisations à des organisations ou à des unités organisationnelles. Une subvention accordée à une organisation ou à une unité organisationnelle est considérée comme une subvention.

    L'octroi à des organisations ou à des unités organisationnelles élimine également le besoin d'accepter une AWS Resource Access Manager (AWS RAM) invitation à partager des ressources pour la subvention. Pour plus d'informations, consultez Accès aux tables et aux bases de données partagées du catalogue de données et affichage de celles-ci.
  - Au lieu d'accorder des autorisations sur de nombreuses tables individuelles d'une base de données, utilisez le caractère générique spécial Toutes les tables pour accorder des autorisations sur toutes les tables de la base de données. L'octroi d'une subvention sur toutes les tables est considéré comme une subvention unique. Pour plus d'informations, consultez Octroi et révocation d'autorisations sur les ressources du catalogue de données.

  > **Note**
  >
  > Pour plus d'informations sur la demande d'une limite plus élevée pour le nombre de partages de ressources dans AWS RAM, voir les quotas de AWS service dans le Références générales AWS.

- Vous devez créer un lien de ressource vers une base de données partagée pour que cette base de données apparaisse dans les éditeurs de requêtes Amazon Redshift Spectrum Amazon Athena et Amazon Redshift. De même, pour pouvoir interroger des tables partagées à l'aide d'Athena et Redshift Spectrum, vous devez créer des liens de ressources vers les tables. Les liens vers les ressources apparaissent ensuite dans la liste des tables des éditeurs de requêtes.

  Au lieu de créer des liens de ressources pour de nombreuses tables individuelles à des fins d'interrogation, vous pouvez utiliser le caractère générique Toutes les tables pour accorder des

autorisations sur toutes les tables d'une base de données. Ensuite, lorsque vous créez un lien de ressource pour cette base de données et que vous sélectionnez ce lien de ressource de base de données dans l'éditeur de requêtes, vous aurez accès à toutes les tables de cette base de données pour votre requête. Pour plus d'informations, consultez Création de liens vers des ressources.

- Lorsque vous partagez des ressources directement avec les principaux d'un autre compte, le principal IAM du compte destinataire n'est peut-être pas autorisé à créer des liens vers des ressources pour interroger les tables partagées à l'aide d'Athena et d'Amazon Redshift Spectrum. Au lieu de créer un lien de ressource pour chaque table partagée, l'administrateur du lac de données peut créer une base de données fictive et accorder des `CREATE_TABLE` autorisations au `ALLIAMPrincipal` groupe. Tous les principaux IAM du compte destinataire peuvent ensuite créer des liens de ressources dans la base de données d'espaces réservés et commencer à interroger les tables partagées.

  Consultez l'exemple de commande CLI pour accorder des autorisations `ALLIAMPrincipals` d'entréeOctroi d'autorisations de base de données à l'aide de la méthode de ressource.
- Athena et Redshift Spectrum prennent en charge le contrôle d'accès au niveau des colonnes, mais uniquement pour l'inclusion, et non pour l'exclusion. Le contrôle d'accès au niveau des colonnes n'est pas pris en charge dans les tâches AWS Glue ETL.
- Lorsqu'une ressource est partagée avec votre AWS compte, vous pouvez accorder des autorisations sur cette ressource uniquement aux utilisateurs de votre compte. Vous ne pouvez pas accorder d'autorisations sur la ressource à d'autres AWS comptes, à des organisations (pas même à votre propre organisation) ou au `IAMAllowedPrincipals` groupe.
- Vous ne pouvez pas accorder `DROP` ou attribuer une `Super` base de données à un compte externe.
- Révoquez les autorisations entre comptes avant de supprimer une base de données ou une table. Dans le cas contraire, vous devez supprimer les partages de ressources orphelins dans AWS Resource Access Manager.

Consultez aussi

- Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation

- `CREATE_TABLE`dans le Référence des autorisations de Lake Formation pour plus de règles et de limitations d'accès entre comptes.

Limites d'accès aux données entre régions

Lake Formation permet d'interroger les tables du catalogue de données. Régions AWS Vous pouvez accéder aux données d'une région depuis d'autres régions à l'aide Amazon Athena d'Amazon EMR et d' AWS Glue ETL en créant des liens de ressources dans d'autres régions pointant vers les bases de données et les tables sources. Grâce à l'accès aux tables entre régions, vous pouvez accéder aux données entre les régions sans copier les données sous-jacentes ou les métadonnées dans le catalogue de données.

Les restrictions suivantes s'appliquent à l'accès aux tables entre régions.

- Lake Formation ne permet pas d'interroger les tables du catalogue de données d'une autre région à l'aide d'Amazon Redshift Spectrum.
- Dans la console Lake Formation, les vues de base de données et de table n'affichent pas les noms des bases de données/tables de la région source.
- Pour afficher la liste des tables d'une base de données partagée d'une autre région, vous devez d'abord créer un lien de ressource vers la base de données partagée, puis sélectionner le lien de ressource et choisir Afficher les tables.
- La fonctionnalité d'accès aux tables entre régions ne fonctionne pas lorsque vous créez des liens de ressources Régions AWS pointant vers des bases de données partagées et des tables créées dans Opt in Regions.

  Pour plus d'informations, consultez la section Activer les régions sur la page Services pris en charge Régions AWS et services.
- Lake Formation ne prend pas en charge les appels de liens de ressources interrégionaux effectués par les utilisateurs de SAML.

Considérations et limites relatives aux affichages du catalogue de données

Dans AWS Glue Data Catalog, une vue est une table virtuelle dont le contenu est défini par une requête qui fait référence à une ou plusieurs tables. Vous pouvez créer une vue qui référence

jusqu'à 10 tables à l'aide des éditeurs SQL pour Amazon Athena ou Amazon Redshift. Les tables de référence sous-jacentes d'une vue peuvent appartenir à la même base de données ou à différentes bases de données au sein de la même base de données Compte AWS.

Les considérations et limites suivantes s'appliquent aux vues du catalogue de données.

- Amazon Redshift crée toujours des vues avec des colonnes varchar à partir de tables contenant des chaînes. Vous devez convertir les colonnes de chaîne en varchar avec une longueur explicite lorsque vous ajoutez des dialectes provenant d'autres moteurs.
- L'octroi d'autorisations de lac de données au `All views` sein d'une base de données permettra au bénéficiaire de disposer d'autorisations sur toutes les tables et vues de la base de données.
- Vous ne pouvez pas créer de vues :
  - Cela fait référence à d'autres points de vue.
  - Lorsque la référence à une table est un lien de ressource.
  - Lorsque les tables de référence disposent des autorisations `IAM_ALLOWED_GROUP` principales.
  - Lorsque la table de référence se trouve dans un autre compte.
  - À partir de métastores Hive externes.

Limites du filtrage des données

Lorsque vous accordez des autorisations à Lake Formation sur une table du catalogue de données, vous pouvez inclure des spécifications de filtrage des données afin de restreindre l'accès à certaines données dans les résultats des requêtes et les moteurs intégrés à Lake Formation. Lake Formation utilise le filtrage des données pour garantir la sécurité au niveau des colonnes, au niveau des lignes et au niveau des cellules. Vous pouvez définir et appliquer des filtres de données sur des colonnes imbriquées si vos données sources contiennent des structures imbriquées.

Tenez compte des remarques et restrictions suivantes concernant le filtrage au niveau des lignes et au niveau des cellules.

- La sécurité au niveau des cellules n'est pas prise en charge sur les colonnes, les vues et les liens de ressources imbriqués.
- Toutes les expressions prises en charge sur les colonnes de niveau supérieur sont également prises en charge sur les colonnes imbriquées. Cependant, les champs imbriqués sous les colonnes de partition ne doivent PAS être référencés lors de la définition d'expressions imbriquées au niveau des lignes.

- La sécurité au niveau des cellules est disponible dans toutes les régions lorsque vous utilisez le moteur Athena version 3 ou Amazon Redshift Spectrum. Pour les autres services, la sécurité au niveau des cellules n'est disponible que dans les régions mentionnées sur le. Régions prises en charge
- Les instructions `SELECT INTO` ne sont pas prises en charge.
- Les types `array` de `map` données et ne sont pas pris en charge dans les expressions de filtre de ligne. Le type `struct` de données est pris en charge.
- Il n'y a pas de limite au nombre de filtres de données pouvant être définis sur une table, mais il existe une limite de 100 `SELECT` autorisations de filtrage de données pour un seul principal sur une table.
- Le nombre maximum de filtres de données pouvant être inclus dans une subvention sur une table est de 10.
- Pour appliquer un filtre de données avec une expression de filtre de ligne, vous devez avoir `SELECT` l'option grant sur toutes les colonnes du tableau. Cette restriction ne s'applique pas aux administrateurs des comptes externes lorsque la subvention a été accordée au compte externe.
- Si un directeur est membre d'un groupe et que le principal et le groupe reçoivent des autorisations sur un sous-ensemble de lignes, les autorisations de ligne effectives du principal sont l'union des autorisations du principal et des autorisations du groupe.
- Les noms de colonnes suivants sont restreints dans un tableau pour le filtrage au niveau des lignes et au niveau des cellules :
  - ctid
  - oid
  - xmin
  - cmin
  - xmax
  - cmax
  - tabloïd
  - insérez un identifiant
  - supprimer exid
  - importoïde
  - identifiant unique du chat rouge

- Si vous appliquez l'expression de filtre toutes les lignes sur un tableau en même temps que d'autres expressions de filtre contenant des prédicats, l'expression de toutes les lignes prévaudra sur toutes les autres expressions de filtre.
- Lorsque des autorisations sur un sous-ensemble de lignes sont accordées à un AWS compte externe et que l'administrateur du lac de données du compte externe accorde ces autorisations au principal de ce compte, le prédicat de filtre effectif du principal est l'intersection du prédicat du compte et de tout prédicat directement accordé au principal.

  Par exemple, si le compte dispose d'autorisations de ligne avec le prédicat `dept='hr'` et que le principal a été autorisé séparément pour`country='us'`, le principal n'a accès qu'aux lignes avec `dept='hr'` et`country='us'`.

Pour plus d'informations sur le filtrage au niveau des cellules, consultez. Filtrage des données et sécurité au niveau des cellules dans Lake Formation

Considérations et limites relatives au mode d'accès hybride

Le mode d'accès hybride offre la flexibilité d'activer de manière sélective les autorisations de Lake Formation pour les bases de données et les tables de votre AWS Glue Data Catalog.
Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants.

Les considérations et limitations suivantes s'appliquent au mode d'accès hybride.

Limites

- Mettre à jour l'enregistrement d'une position Amazon S3 : vous ne pouvez pas modifier les paramètres d'une position enregistrée auprès de Lake Formation à l'aide d'un rôle lié à un service.
- Option d'activation lors de l'utilisation des balises LF — Lorsque vous pouvez accorder des autorisations à Lake Formation à l'aide de balises LF, vous pouvez activer des principes pour appliquer les autorisations de Lake Formation étape par étape en choisissant des bases de données et des tables auxquelles des balises LF sont attachées.
- Principaux d'adhésion — Actuellement, seul un rôle d'administrateur de lac de données peut attribuer des principes aux ressources.
- Activer toutes les tables d'une base de données — Dans le cas des autorisations entre comptes, lorsque vous accordez des autorisations et que vous activez toutes les tables d'une base de

données, vous devez également opter pour la base de données pour que les autorisations fonctionnent.

Considérations

- Mise à jour de l'emplacement Amazon S3 enregistré auprès de Lake Formation en mode d'accès hybride — Nous déconseillons de convertir un emplacement de données Amazon S3 déjà enregistré auprès de Lake Formation en mode d'accès hybride, bien que cela soit possible.
- Comportements de l'API lorsqu'un emplacement de données est enregistré en mode d'accès hybride
  - CreateTable — L'emplacement est considéré comme enregistré auprès de Lake Formation, quels que soient le drapeau du mode d'accès hybride et le statut d'inscription. L'utilisateur a donc besoin de l'autorisation de localisation des données pour créer une table.
  - CreatePartition/BatchCreatePartitions/UpdatePartitions (lorsque l'emplacement de la partition est mis à jour pour pointer vers l'emplacement enregistré auprès de l'hybride) — L'emplacement Amazon S3 est considéré comme enregistré auprès de Lake Formation, quels que soient l'indicateur du mode d'accès hybride et le statut d'option d'inscription. L'utilisateur a donc besoin de l'autorisation de localisation des données pour créer ou mettre à jour une base de données.
  - CreateDatabase/UpdateDatabase (lorsque l'emplacement de la base de données est mis à jour pour pointer vers l'emplacement enregistré en mode d'accès hybride) — L'emplacement est considéré comme enregistré auprès de Lake Formation, quels que soient le drapeau du mode d'accès hybride et le statut d'option d'inscription. L'utilisateur a donc besoin de l'autorisation de localisation des données pour créer ou mettre à jour une base de données.
  - UpdateTable (lorsqu'un emplacement de table est mis à jour pour pointer vers l'emplacement enregistré en mode d'accès hybride) — L'emplacement est considéré comme enregistré auprès de Lake Formation, quels que soient le drapeau du mode d'accès hybride et le statut d'option d'inscription. L'utilisateur a donc besoin d'une autorisation de localisation des données pour mettre à jour le tableau. Si l'emplacement de la table n'est pas mis à jour ou s'il pointe vers un emplacement qui n'est pas enregistré auprès de Lake Formation, l'utilisateur n'a pas besoin d'autorisation de localisation des données pour mettre à jour la table.

Considérations et limites relatives au partage des données du magasin de métadonnées Hive

Grâce à la fédération des AWS Glue Data Catalog métadonnées (fédération du catalogue de données), vous pouvez connecter le catalogue de données à des métastores externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation

Les considérations et limitations suivantes s'appliquent aux bases de données fédérées créées à partir de bases de données Hive :

Considérations

- AWS SAM support des applications : vous êtes responsable de la disponibilité des ressources de l'application AWS SAM déployée (Amazon API Gateway et de la fonction Lambda). Assurez-vous que la connexion entre le métastore AWS Glue Data Catalog et le métastore Hive fonctionne lorsque les utilisateurs exécutent des requêtes.
- Exigence de version du métastore Hive — Vous ne pouvez créer des bases de données fédérées qu'à l'aide d'Apache Hive version 3 ou supérieure.
- Exigence de base de données mappée — Chaque base de données Hive doit être mappée à une nouvelle base de données dans Lake Formation.
- Support de fédération au niveau de la base de données : vous pouvez vous connecter au métastore Hive uniquement au niveau de la base de données.
- Autorisations sur les bases de données fédérées : les autorisations appliquées à une base de données fédérée ou aux tables d'une base de données fédérée sont conservées même lorsqu'une table source ou une base de données est supprimée. Lorsque la base de données ou la table source est recréée, il n'est pas nécessaire de réoctroyer les autorisations. Lorsqu'une table fédérée dotée d'autorisations Lake Formation est supprimée à la source, les autorisations Lake Formation sont toujours visibles et vous pouvez les révoquer si nécessaire.

  Si un utilisateur supprime une base de données fédérée, toutes les autorisations correspondantes sont perdues. La recréation de la même base de données portant le même nom ne permet pas de récupérer les autorisations de Lake Formation. Les utilisateurs devront à nouveau configurer de nouvelles autorisations.
- Autorisations de AllowedPrincipal groupe IAM sur les bases de données fédérées — Sur la base de cela`DataLakeSettings`, Lake Formation peut attribuer des autorisations à toutes

les bases de données et tables à un groupe virtuel nommé. `IAMAllowedPrincipal` Il `IAMAllowedPrincipal` fait référence à tous les principaux IAM qui ont accès aux ressources du catalogue de données par le biais des politiques principales IAM et AWS Glue des politiques de ressources. Si ces autorisations existent sur une base de données ou une table, tous les principaux ont accès à la base de données ou à la table.

Cependant, Lake Formation n'autorise pas `IAMAllowedPrincipal` les autorisations sur les tables des bases de données fédérées. Lorsque vous créez des bases de données fédérées, assurez-vous de transmettre le `CreateTableDefaultPermissions` paramètre sous forme de liste vide.

Pour plus d'informations, consultez Modification des paramètres par défaut de votre lac de données.

- Joindre des tables dans des requêtes : vous pouvez joindre des tables de métastore Hive à des tables natives de Data Catalog pour exécuter des requêtes.

Limites

- Limitation de synchronisation des métadonnées entre le métastore AWS Glue Data Catalog et le métastore Hive — Après avoir établi la connexion au métastore Hive, vous devez créer une base de données fédérée pour synchroniser les métadonnées du métastore Hive avec le. AWS Glue Data Catalog Les tables de la base de données fédérée sont synchronisées au moment de l'exécution lorsque les utilisateurs exécutent des requêtes.
- Limitation relative à la création de nouvelles tables dans une base de données fédérée : vous ne pourrez pas créer de nouvelles tables dans des bases de données fédérées.
- Limitation des autorisations relatives aux données : le support pour les autorisations sur les vues tabulaires du métastore Hive n'est pas disponible.

Limites du partage de données Amazon Redshift

AWS Lake Formation vous permet de gérer en toute sécurité les données d'un partage de données d'Amazon Redshift. Amazon Redshift est un service d'entrepôt de données entièrement géré de plusieurs pétaoctets dans le cloud. AWS Grâce à la fonctionnalité de partage de données, Amazon Redshift vous permet de partager des données entre différentes entités. Comptes AWS Pour plus d'informations sur le partage de données Amazon Redshift, consultez Présentation du partage de données dans Amazon Redshift.

Les remarques et restrictions suivantes s'appliquent aux bases de données fédérées créées à partir de partages de données Amazon Redshift :

- Exigence de base de données mappée — Chaque partage de données Amazon Redshift doit être mappé vers une nouvelle base de données dans Lake Formation. Cela est nécessaire pour conserver des noms de table uniques lorsque la représentation des objets de partage de données est aplatie dans la base de données du catalogue de données.
- Limitation relative à la création de nouvelles tables dans une base de données fédérée : vous ne pourrez pas créer de nouvelles tables dans des bases de données fédérées.
- Autorisations sur les bases de données fédérées : les autorisations appliquées à une base de données fédérée ou à des tables d'une base de données fédérée sont conservées même lorsqu'une table source ou une base de données est supprimée. Lorsque la base de données ou la table source est recréée, il n'est pas nécessaire de réoctroyer les autorisations. Lorsqu'une table fédérée avec des autorisations Lake Formation est supprimée à la source, les autorisations Lake Formation seront toujours visibles et vous pouvez les révoquer si nécessaire.

  Si un utilisateur supprime une base de données fédérée, toutes les autorisations correspondantes sont perdues. La recréation de la même base de données portant le même nom ne permet pas de récupérer les autorisations de Lake Formation. Les utilisateurs devront à nouveau configurer de nouvelles autorisations.
- Autorisations de AllowedPrincipal groupe IAM sur les bases de données fédérées — Sur la base de cela`DataLakeSettings`, Lake Formation peut attribuer des autorisations à toutes les bases de données et tables à un groupe virtuel nommé. `IAMAllowedPrincipal` Il `IAMAllowedPrincipal` fait référence à tous les principaux IAM qui ont accès aux ressources du catalogue de données par le biais des politiques principales IAM et AWS Glue des politiques de ressources. Si ces autorisations existent sur une base de données ou une table, tous les principaux ont accès à la base de données ou à la table.

  Cependant, Lake Formation n'autorise pas `IAMAllowedPrincipal` les autorisations sur les tables des bases de données fédérées. Lorsque vous créez des bases de données fédérées, assurez-vous de transmettre le `CreateTableDefaultPermissions` paramètre sous forme de liste vide.

  Pour plus d’informations, consultez Modification des paramètres par défaut de votre lac de données.
- Filtrage des données : dans Lake Formation, vous pouvez accorder des autorisations sur une table d'une base de données fédérée avec un filtrage au niveau des colonnes et au niveau des lignes.

Toutefois, vous ne pouvez pas combiner le filtrage au niveau des colonnes et au niveau des lignes pour restreindre l'accès au niveau de la granularité au niveau des cellules aux tables des bases de données fédérées.

- Identifiant distinguant majuscules et minuscules : les objets de partage de données Amazon Redshift gérés par Lake Formation prennent en charge les noms de tables et de colonnes uniquement en minuscules. N'activez pas l'identifiant majuscules/minuscules pour les bases de données, les tables et les colonnes dans les partages de données Amazon Redshift, s'ils doivent être partagés et gérés à l'aide de Lake Formation.

Pour plus d'informations sur les limites liées à l'utilisation de partages de données dans Amazon Redshift, consultez la section Limitations relatives au partage de données dans le manuel Amazon Redshift Database Developer Guide.

Limites de l'intégration d'IAM Identity Center

Vous pouvez ainsi vous connecter à des fournisseurs d'identité (IdPs) et gérer de manière centralisée l'accès des utilisateurs et des groupes à travers les services AWS d'analyse. AWS IAM Identity Center Vous pouvez AWS Lake Formation le configurer en tant qu'application activée dans IAM Identity Center, et les administrateurs de data lake peuvent accorder des autorisations détaillées aux utilisateurs et aux groupes autorisés sur les ressources. AWS Glue Data Catalog

Les limites suivantes s'appliquent à l'intégration de Lake Formation à IAM Identity Center :

- Vous ne pouvez pas affecter des utilisateurs et des groupes IAM Identity Center en tant qu'administrateurs de data lake ou administrateurs en lecture seule dans Lake Formation.
- Les utilisateurs et les groupes IAM Identity Center peuvent interroger les ressources chiffrées du catalogue de données si vous utilisez un rôle IAM qui AWS Glue peut assumer en votre nom le chiffrement et le déchiffrement du catalogue de données. AWS les clés gérées ne prennent pas en charge la propagation d'identités fiables.
- Les utilisateurs et les groupes IAM Identity Center peuvent uniquement invoquer les opérations d'API répertoriées dans la `AWSIAMIdentityCenterAllowListForIdentityContext` politique fournie par IAM Identity Center.
- Lake Formation permet aux rôles IAM issus de comptes externes d'agir en tant que rôles de support au nom des utilisateurs et des groupes d'IAM Identity Center pour accéder aux ressources du catalogue de données, mais les autorisations ne peuvent être accordées que sur les ressources du catalogue de données du compte propriétaire. Si vous essayez d'accorder des autorisations aux

utilisateurs et aux groupes d'IAM Identity Center sur les ressources du catalogue de données d'un compte externe, Lake Formation génère le message d'erreur suivant : « Les autorisations entre comptes ne sont pas prises en charge pour le principal ».

Meilleures pratiques et considérations relatives au contrôle d'accès basé sur les balises Lake Formation

Vous pouvez créer, gérer et attribuer des balises LF pour contrôler l'accès aux bases de données, aux tables et aux colonnes du catalogue de données.

Tenez compte des meilleures pratiques suivantes lorsque vous utilisez le contrôle d'accès basé sur des balises Lake Formation :

- Toutes les balises LF doivent être prédéfinies avant de pouvoir être attribuées aux ressources du catalogue de données ou accordées aux principaux.

  L'administrateur du lac de données peut déléguer les tâches de gestion des balises en créant des balises LF avec les autorisations IAM requises. Les ingénieurs de données et les analystes décident des caractéristiques et des relations des balises LF. Les créateurs de balises LF créent et maintiennent ensuite les balises LF dans Lake Formation.
- Vous pouvez attribuer plusieurs balises LF aux ressources du catalogue de données. Une seule valeur pour une clé donnée peut être affectée à une ressource donnée.

  Par exemple, vous pouvez attribuer`module=Orders`,`region=West`,`division=Consumer`, etc. à une base de données, une table ou une colonne. Vous ne pouvez pas attribuer`module=Orders,Customers`.
- Vous ne pouvez pas attribuer de balises LF aux ressources lorsque vous créez la ressource. Vous ne pouvez ajouter des balises LF qu'aux ressources existantes.
- Vous pouvez accorder des expressions de balises LF, et pas simplement des balises LF uniques, à un principal.

  Une expression LF-Tag ressemble à ce qui suit (en pseudo-code).

  ```
  module=sales AND division=(consumer OR commercial)
  ```

  Un principal auquel cette expression LF-Tag est attribuée ne peut accéder qu'aux ressources du catalogue de données (bases de données, tables et colonnes) qui lui sont

attribuées *module=sales* ou*division=consumer*. division=commercial Si vous souhaitez que le directeur puisse accéder à des ressources qui incluent module=sales ou division=commercial non les deux dans la même subvention. Accordez deux subventions, une pour module=sales et une pourdivision=commercial.

L'expression LF-Tag la plus simple consiste en une seule balise LF, telle que. module=sales

- Un principal autorisé à accéder à une balise LF comportant plusieurs valeurs peut accéder aux ressources du catalogue de données avec l'une ou l'autre de ces valeurs. Par exemple, si un utilisateur se voit attribuer une balise LF avec key= module et values=orders,customers, il a accès aux ressources attribuées soit. module=orders module=customers
- Vous devez être Grant with LF-Tag expressions autorisé à accorder des autorisations de données sur les ressources du catalogue de données à l'aide de la méthode LF-TBAC. L'administrateur du lac de données et le créateur du tag LF reçoivent implicitement cette autorisation. Un principal Grant with LFTag expressions autorisé peut accorder des autorisations de données sur les ressources en utilisant :
  - la méthode de ressource nommée
  - la méthode LF-TBAC, mais en utilisant uniquement la même expression LF-Tag

    Supposons, par exemple, que l'administrateur du lac de données accorde l'autorisation suivante (en pseudo-code).

    ```
    GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH
     GRANT OPTION
    ```

    Dans ce cas, user1 vous pouvez attribuer des tables SELECT à d'autres principaux en utilisant la méthode LF-TBAC, mais uniquement avec l'expression LF-Tag complète. module=customers, region=west,south
- Si un principal obtient des autorisations sur une ressource à la fois avec la méthode LF-TBAC et la méthode de ressource nommée, les autorisations que le principal possède sur la ressource sont l'union des autorisations accordées par les deux méthodes.
- Lake Formation prend en charge l'attribution DESCRIBE et ASSOCIATE l'attribution de balises LF entre les comptes, ainsi que l'octroi d'autorisations sur les ressources du catalogue de données entre les comptes à l'aide de la méthode LF-TBAC. Dans les deux cas, le principal est un identifiant de AWS compte.

> **Note**
>
> Lake Formation soutient les subventions intercomptes aux organisations et aux unités organisationnelles en utilisant la méthode LF-TBAC. Pour utiliser cette fonctionnalité, vous devez mettre à jour les paramètres de version entre comptes vers la version 3.

Pour plus d'informations, consultez Partage de données entre comptes dans Lake Formation.

- Les ressources du catalogue de données créées dans un compte ne peuvent être étiquetées qu'à l'aide de balises LF créées dans le même compte. Les balises LF créées dans un compte ne peuvent pas être associées à des ressources partagées depuis un autre compte.
- L'utilisation du contrôle d'accès basé sur les balises de Lake Formation (LF-TBAC) pour accorder un accès entre comptes aux ressources du catalogue de données nécessite des ajouts à la politique de ressources du catalogue de données de votre compte. AWS Pour plus d'informations, consultez Prérequis.
- Les touches LF-Tag et les valeurs LF-Tag ne peuvent pas dépasser 50 caractères.
- Le nombre maximum de balises LF pouvant être attribuées à une ressource de catalogue de données est de 50.
- Les limites suivantes sont des limites souples :
  - Le nombre maximum de balises LF pouvant être créées est de 1 000.
  - Le nombre maximum de valeurs pouvant être définies pour une balise LF est de 1 000.
- Les balises, les clés et les valeurs sont converties en minuscules lors de leur stockage.
- Une seule valeur pour une balise LF peut être affectée à une ressource particulière.
- Si plusieurs balises LF sont accordées à un principal avec une seule autorisation, le principal ne peut accéder qu'aux ressources du catalogue de données contenant toutes les balises LF.
- AWS GlueLes tâches ETL nécessitent un accès complet aux tables. Les tâches échoueront si le rôle AWS Glue ETL n'a pas accès à toutes les colonnes d'une table. Il est possible d'appliquer des balises LF au niveau d'une colonne, mais cela peut entraîner la perte de l'accès complet aux tables pour les rôles AWS Glue ETL et l'échec des tâches. L'utilisation de filtres de données pour le filtrage des colonnes et/ou des lignes n'est pas affectée par cette limitation.
- Si l'évaluation d'une expression LF-Tag ne donne accès qu'à un sous-ensemble de colonnes de table, mais que l'autorisation Lake Formation accordée en cas de correspondance est l'une des autorisations nécessitant un accès complet aux colonnes, à savoir`Alter`, ou `Drop`

`InsertDelete`, alors aucune de ces autorisations n'est accordée. Au lieu de cela, seul `Describe` est accordé. Si l'autorisation accordée est `All` (`Super`), alors uniquement `Select` et `Describe` sont accordées.

- Les caractères génériques ne sont pas utilisés avec les balises LF. Pour attribuer une balise LF à toutes les colonnes d'une table, vous attribuez la balise LF à la table, et toutes les colonnes de la table héritent de la balise LF. Pour attribuer une balise LF à toutes les tables d'une base de données, vous attribuez la balise LF à la base de données, et toutes les tables de la base de données héritent de cette balise LF.

Formats pris en charge et limites pour le compactage géré des données

Pour améliorer les performances de lecture des services AWS d'analyse tels qu'Amazon Athena, Amazon EMR et les tâches AWS Glue ETL, AWS Glue Data Catalog propose un compactage géré (un processus qui compacte de petits objets Amazon S3 en objets plus grands) pour les tables Iceberg dans Data Catalog.

Le compactage des données prend en charge une variété de types de données et de formats de compression pour la lecture et l'écriture de données, y compris la lecture de données à partir de tables chiffrées.

Le compactage des données prend en charge :

- Types de fichiers : Parquet
- Types de données : booléen, entier, long, flottant, double, chaîne, décimal, date, heure, horodatage, chaîne, UUID, binaire
- Compression : zstd, gzip, snappy, non compressé
- Chiffrement : le compactage des données prend uniquement en charge le chiffrement Amazon S3 (SSE-S3) et le chiffrement KMS côté serveur (SSE-KMS).
- Compactage par regroupement
- Évolution du schéma
- Tableaux avec taille de fichier cible (écriture). target-file-size-bytes propriété en configuration iceberg) dans la plage incluse de 128 Mo à 512 Mo.
- Régions
  - Asie-Pacifique (Tokyo)

  - Asie-Pacifique (Séoul)
  - Asie-Pacifique (Mumbai)
  - Asie-Pacifique (Singapour)
  - Europe (Irlande)
  - Europe (Francfort)
  - USA Est (Virginie du Nord)
  - USA Est (Ohio)
  - USA Ouest (Californie du Nord)
  - Amérique du Sud (São Paulo)
- Vous pouvez exécuter le compactage depuis le compte où réside le catalogue de données lorsque le compartiment Amazon S3 qui stocke les données sous-jacentes se trouve dans un autre compte. Pour ce faire, le rôle de compactage nécessite l'accès au compartiment Amazon S3.

Le compactage des données ne prend pas en charge actuellement :

- Types de fichiers : Avro, ORC
- Types de données : fixe
- Compression : brotli, lz4
- Compactage des fichiers pendant que la spécification de partition évolue.
- Tri régulier ou tri par ordre Z
- Fusionner ou supprimer des fichiers : le processus de compactage ignore les fichiers de données auxquels des fichiers de suppression sont associés.
- Compactage sur des tables entre comptes : vous ne pouvez pas exécuter le compactage sur des tables entre comptes.
- Compaction sur des tables entre régions : vous ne pouvez pas exécuter de compactage sur des tables entre régions.
- Activation du compactage sur des liens de ressources
- Points de terminaison d'un VPC pour les compartiments Amazon S3

Résolution des problèmes liés à la formation du Lake

Si vous rencontrez des problèmes lorsque vous travaillez avec AWS Lake Formation, consultez les rubriques de cette section.

Rubriques

- Résolution de problème généraux
- Résolution des problèmes d'accès entre comptes
- Résolution des problèmes liés aux plans et aux flux de travail
- Problèmes connus pour AWS Lake Formation
- Message d'erreur mis à jour

Résolution de problème généraux

Utilisez les informations fournies ici pour vous aider à diagnostiquer et à résoudre divers problèmes liés à la Lake Formation.

Erreur : Autorisations insuffisantes pour Lake Formation sur <Amazon S3 location>

Une tentative a été faite pour créer ou modifier une ressource de catalogue de données sans autorisation de localisation des données sur l'emplacement Amazon S3 indiqué par la ressource.

Si une base de données ou une table de catalogue de données pointe vers un emplacement Amazon S3, lorsque vous accordez les autorisations `CREATE_TABLE` à Lake Formation`ALTER`, vous devez également accorder l'`DATA_LOCATION_ACCESS`autorisation sur cet emplacement. Si vous accordez ces autorisations à des comptes externes ou à des organisations, vous devez inclure l'option d'octroi.

Une fois ces autorisations accordées à un compte externe, l'administrateur du lac de données de ce compte doit les accorder aux principaux (utilisateurs ou rôles) du compte. Lorsque vous `DATA_LOCATION_ACCESS` accordez l'autorisation reçue d'un autre compte, vous devez spécifier l'ID de catalogue (ID de AWS compte) du compte propriétaire. Le compte propriétaire est le compte qui a enregistré l'emplacement.

Pour plus d’informations, consultez Contrôle d'accès aux données sous-jacent et Octroi d'autorisations de localisation des données.

Erreur : « Autorisations de clé de chiffrement insuffisantes pour l'API Glue »

Une tentative a été faite pour accorder des autorisations à Lake Formation sans autorisations AWS Identity and Access Management (IAM) sur la clé de AWS KMS chiffrement d'un catalogue de données chiffré.

Ma requête Amazon Athena ou celle d'Amazon Redshift qui utilise des manifestes échoue

Lake Formation ne prend pas en charge les requêtes utilisant des manifestes.

Erreur : « Autorisations de formation lacustres insuffisantes : création d'une balise requise dans le catalogue »

L'utilisateur/le rôle doit être un administrateur de data lake.

Erreur lors de la suppression d'administrateurs de lacs de données non valides

Vous devez supprimer simultanément tous les administrateurs de lacs de données non valides (rôles IAM supprimés définis comme administrateurs de lacs de données). Si vous essayez de supprimer séparément les administrateurs de data lake non valides, Lake Formation génère une erreur principale non valide.

Résolution des problèmes d'accès entre comptes

Utilisez les informations fournies ici pour vous aider à diagnostiquer et à résoudre les problèmes d'accès entre comptes.

Rubriques

- J'ai accordé l'autorisation d'utiliser plusieurs comptes Lake Formation, mais le destinataire ne peut pas voir la ressource
- Les utilisateurs principaux du compte destinataire peuvent voir la ressource du catalogue de données, mais ne peuvent pas accéder aux données sous-jacentes
- Erreur : « L'association a échoué car l'appelant n'était pas autorisé » lors de l'acceptation d'une invitation de partage de AWS RAM ressources

- Erreur : « Non autorisé à accorder des autorisations pour la ressource »
- Erreur : « Accès refusé pour récupérer les informations de AWS l'organisation »
- Erreur : « Organisation <organization-ID>introuvable »
- Erreur : « Permissions de formation lacustres insuffisantes : combinaison illégale »
- ConcurrentModificationException sur les demandes d'accord/de révocation adressées à des comptes externes
- Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes

J'ai accordé l'autorisation d'utiliser plusieurs comptes Lake Formation, mais le destinataire ne peut pas voir la ressource

- L'utilisateur du compte destinataire est-il un administrateur de data lake ? Seuls les administrateurs de data lake peuvent voir la ressource au moment du partage.
- Partagez-vous avec un compte externe à votre organisation en utilisant la méthode des ressources nommées ? Dans ce cas, l'administrateur du lac de données du compte destinataire doit accepter une invitation à partager des ressources dans AWS Resource Access Manager (AWS RAM).

  Pour plus d'informations, consultez the section called "Accepter une invitation à partager des AWS RAM ressources".
- Utilisez-vous des politiques de ressources au niveau du compte (catalogue de données) dans ? AWS Glue Dans l'affirmative, si vous utilisez la méthode des ressources nommées, vous devez inclure une déclaration spéciale dans la politique qui autorise le partage AWS RAM des politiques en votre nom.

  Pour plus d'informations, consultez the section called "Gestion des autorisations entre comptes à l'aide des deux AWS Glue et de Lake Formation".
- Disposez-vous des autorisations AWS Identity and Access Management (IAM) requises pour accorder un accès entre comptes ?

  Pour plus d'informations, consultez the section called "Prérequis".
- La ressource pour laquelle vous avez accordé des autorisations ne doit pas avoir d'autorisation Lake Formation accordée au `IAMAllowedPrincipals` groupe.
- Y a-t-il une `deny` déclaration sur la ressource dans la politique au niveau du compte ?

Les utilisateurs principaux du compte destinataire peuvent voir la ressource du catalogue de données, mais ne peuvent pas accéder aux données sous-jacentes

Les principaux associés au compte du destinataire doivent disposer des autorisations AWS Identity and Access Management (IAM) requises. Pour plus de détails, consultez Accès aux données sous-jacentes d'une table partagée.

Erreur : « L'association a échoué car l'appelant n'était pas autorisé » lors de l'acceptation d'une invitation de partage de AWS RAM ressources

Après avoir accordé l'accès à une ressource à un autre compte, lorsque le compte destinataire tente d'accepter l'invitation de partage de ressources, l'action échoue.

```
$ aws ram get-resource-share-associations --association-type PRINCIPAL --resource-
share-arns arn:aws:ram:aws-region:444444444444:resource-share/e1d1f4ba-xxxx-xxxx-xxxx-
xxxxxxxx5d8d
{
    "resourceShareAssociations": [
        {
            "resourceShareArn": "arn:aws:ram:aws-region:444444444444:resource-share/
e1d1f4ba-xxxx-xxxx-xxxx-xxxxxxxx5d8d
",
            "resourceShareName": "LakeFormation-MMCC0XQBH3Y",
            "associatedEntity": "5815803XXXXX",
            "associationType": "PRINCIPAL",
            "status": "FAILED",
            "statusMessage": "Association failed because the caller was not
 authorized.",
            "creationTime": "2021-07-12T02:20:10.267000+00:00",
            "lastUpdatedTime": "2021-07-12T02:20:51.830000+00:00",
            "external": true
        }
    ]
}
```

L'erreur se produit parce que le `glue:PutResourcePolicy` est invoqué AWS Glue lorsque le compte destinataire accepte l'invitation de partage de ressources. Pour résoudre le problème, autorisez l'`glue:PutResourcePolicy`action par le rôle assumé utilisé par le compte producteur/concédant.

Erreur : « Non autorisé à accorder des autorisations pour la ressource »

Une tentative a été faite pour accorder des autorisations entre comptes sur une base de données ou une table appartenant à un autre compte. Lorsqu'une base de données ou une table est partagée avec votre compte, en tant qu'administrateur de data lake, vous ne pouvez accorder des autorisations à ce sujet qu'aux utilisateurs de votre compte.

Erreur : « Accès refusé pour récupérer les informations de AWS l'organisation »

Votre compte est un compte de gestion des AWS Organisations et vous ne disposez pas des autorisations requises pour récupérer les informations relatives à l'organisation, telles que les unités organisationnelles du compte.

Pour plus d'informations, consultez Required permissions for cross-account grants.

Erreur : « Organisation <organization-ID>introuvable »

Une tentative a été faite pour partager une ressource avec une organisation, mais le partage avec les organisations n'est pas activé. Activez le partage des ressources avec les organisations.

Pour plus d'informations, consultez la section Activer le partage avec AWS les organisations dans le guide de AWS RAM l'utilisateur.

Erreur : « Permissions de formation lacustres insuffisantes : combinaison illégale »

Un utilisateur a partagé une ressource de catalogue de données alors que les autorisations de Lake Formation étaient accordées au `IAMAllowedPrincipals` groupe pour cette ressource. L'utilisateur doit révoquer toutes les autorisations de Lake Formation `IAMAllowedPrincipals` avant de partager la ressource.

ConcurrentModificationException sur les demandes d'accord/de révocation adressées à des comptes externes

Lorsque les utilisateurs font plusieurs demandes simultanées d'octroi et/ou de révocation d'autorisations pour un principal sur les politiques LF-Tag, Lake Formation les lance. ConcurrentModificationException Les utilisateurs doivent détecter l'exception et réessayer

la demande d'accord/de révocation qui a échoué. L'utilisation de versions par lots des opérations`GrantPermissions`/`RevokePermissions`API BatchRevokePermissionspermet de résoudre ce problème dans une certaine mesure en réduisant le nombre de demandes d'accord/de révocation simultanées. BatchGrantPermissions

Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes

Lorsque vous utilisez Amazon EMR pour accéder aux données partagées avec vous depuis un autre compte, certaines bibliothèques Spark tentent d'appeler l'opération d'API `Glue:GetUserDefinedFunctions`. Les versions 1 et 2 des autorisations AWS RAM gérées ne prenant pas en charge cette action, le message d'erreur suivant s'affiche :

```
"ERROR: User: arn:aws:sts::012345678901:assumed-role/my-
spark-role/i-06ab8c2b59299508a is not authorized to perform:
glue:GetUserDefinedFunctions on resource: arn:exampleCatalogResource
because no resource-based policy allows the glue:GetUserDefinedFunctions
action"
```

Pour résoudre cette erreur, l'administrateur du lac de données qui a créé le partage de ressources doit mettre à jour les autorisations AWS RAM gérées associées au partage de ressources. La version 3 des autorisations gérées AWS RAM permet aux nœuds principaux d'effectuer l'action `glue:GetUserDefinedFunctions`.

Si vous créez un nouveau partage de ressources, Lake Formation applique la dernière version de l'autorisation AWS RAM gérée par défaut, et aucune action n'est requise de votre part. Pour activer l'accès aux données entre comptes pour les partages de ressources existants, vous devez mettre à jour les autorisations AWS RAM gérées vers la version 3.

Vous pouvez consulter les AWS RAM autorisations attribuées aux ressources partagées avec vous dans AWS RAM. Les autorisations suivantes sont incluses dans la version 3 :

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase
```

```
AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Pour mettre à jour la version des autorisations AWS RAM gérées des partages de ressources existants

Vous (administrateur du lac de données) pouvez soit mettre à jour les autorisations AWS RAM gérées vers une version plus récente en suivant les instructions du guide de AWS RAM l'utilisateur, soit révoquer toutes les autorisations existantes pour le type de ressource et les réaccorder. Si vous révoquez les autorisations, le partage AWS RAM de AWS RAM ressources associé au type de ressource est supprimé. Lorsque vous réaccordez des autorisations, AWS RAM de nouveaux partages de ressources sont créés en y joignant la dernière version des autorisations AWS RAM gérées.

Résolution des problèmes liés aux plans et aux flux de travail

Utilisez les informations fournies ici pour vous aider à diagnostiquer et à résoudre les problèmes liés au plan et au flux de travail.

Rubriques

- <role-ARN>Mon plan a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à exécuter : iam : PassRole on resource : »
- <role-ARN>Mon flux de travail a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à effectuer : iam : PassRole on resource : »
- Un crawler de mon flux de travail a échoué avec le message « La ressource n'existe pas ou le demandeur n'est pas autorisé à accéder aux autorisations demandées »
- Un crawler de mon flux de travail a échoué avec « Une erreur s'est produite (AccessDeniedException) lors de l'appel de l' CreateTable opération... »

<role-ARN>Mon plan a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à exécuter : iam : PassRole on resource : »

Un utilisateur qui ne dispose pas des autorisations suffisantes pour transmettre le rôle choisi a tenté de créer un plan.

Mettez à jour la politique IAM de l'utilisateur pour pouvoir transmettre le rôle, ou demandez-lui de choisir un autre rôle avec les autorisations de mot de passe requises.

Pour plus d'informations, consultez the section called "Référence des personnalités de Lake Formation et des autorisations IAM".

<role-ARN>Mon flux de travail a échoué avec « L'utilisateur : <user-ARN>n'est pas autorisé à effectuer : iam : PassRole on resource : »

Le rôle que vous avez spécifié pour le flux de travail ne comportait pas de politique intégrée permettant au rôle de se transmettre de lui-même.

Pour plus d'informations, consultez the section called "(Facultatif) Créez un rôle IAM pour les flux de travail".

Un crawler de mon flux de travail a échoué avec le message « La ressource n'existe pas ou le demandeur n'est pas autorisé à accéder aux autorisations demandées »

L'une des causes possibles est que le rôle transmis ne disposait pas des autorisations suffisantes pour créer une table dans la base de données cible. Accordez au rôle l'`CREATE_TABLE`autorisation d'accéder à la base de données.

Un crawler de mon flux de travail a échoué avec « Une erreur s'est produite (AccessDeniedException) lors de l'appel de l' CreateTable opération... »

L'une des causes possibles est que le rôle de flux de travail ne disposait pas d'autorisations de localisation des données sur l'emplacement de stockage cible. Accordez des autorisations de localisation des données au rôle.

Pour plus d'informations, consultez the section called "`DATA_LOCATION_ACCESS`".

Problèmes connus pour AWS Lake Formation

Passez en revue ces problèmes connus pour AWS Lake Formation.

Rubriques

- Limitation du filtrage des métadonnées des tables

- Problème lié au changement de nom d'une colonne exclue
- Problème lié à la suppression de colonnes dans les tableaux CSV
- Les partitions de table doivent être ajoutées sous un chemin commun
- Problème lié à la création d'une base de données pendant la création du flux de travail
- Problème lié à la suppression puis à la recréation d'un utilisateur
- GetTableset SearchTables les API ne mettent pas à jour la valeur du IsRegisteredWithLakeFormation paramètre
- Les opérations de l'API Data Catalog ne mettent pas à jour la valeur du IsRegisteredWithLakeFormation paramètre
- Les opérations de Lake Formation ne prennent pas en charge AWS Glue le registre des schémas

Limitation du filtrage des métadonnées des tables

AWS Lake Formation les autorisations au niveau des colonnes peuvent être utilisées pour restreindre l'accès à des colonnes spécifiques d'un tableau. Lorsqu'un utilisateur extrait des métadonnées relatives à la table à l'aide de la console ou d'une API similaire`glue:GetTable`, la liste des colonnes de l'objet de table contient uniquement les champs auxquels il a accès. Il est important de comprendre les limites de ce filtrage des métadonnées.

Bien que Lake Formation mette à disposition des métadonnées relatives aux autorisations des colonnes pour les services intégrés, le filtrage des colonnes dans les réponses aux requêtes relève de la responsabilité du service intégré. Les clients de Lake Formation qui prennent en charge le filtrage au niveau des colonnes, notamment Amazon Athena, Amazon Redshift Spectrum et Amazon EMR, filtrent les données en fonction des autorisations de colonne enregistrées auprès de Lake Formation. Les utilisateurs ne pourront pas lire les données auxquelles ils ne devraient pas avoir accès. Actuellement, l'AWS GlueETL ne prend pas en charge le filtrage des colonnes.

> **Note**
>
> Les clusters EMR ne sont pas entièrement gérés par. AWS Il est donc de la responsabilité des administrateurs EMR de sécuriser correctement les clusters afin d'éviter tout accès non autorisé aux données.

Certaines applications ou certains formats peuvent stocker des métadonnées supplémentaires, notamment des noms et des types de colonnes, dans la `Parameters` carte sous forme de propriétés

de table. Ces propriétés sont renvoyées telles quelles et sont accessibles à tous les utilisateurs `SELECT` autorisés sur n'importe quelle colonne.

Par exemple, l'Avro SerDe stocke une représentation JSON du schéma de table dans une propriété de table nommée`avro.schema.literal`, qui est accessible à tous les utilisateurs ayant accès à la table. Nous vous recommandons d'éviter de stocker des informations sensibles dans les propriétés des tables et de savoir que les utilisateurs peuvent apprendre le schéma complet des tables au format Avro. Cette limitation est spécifique aux métadonnées relatives à une table.

AWS Lake Formation supprime toute propriété de table en commençant par « `spark.sql.sources.schema` lors de la réponse à une demande `glue:GetTable` ou à une demande similaire » si l'appelant n'a pas d'`SELECT`autorisations sur toutes les colonnes de la table. Cela empêche les utilisateurs d'accéder à des métadonnées supplémentaires concernant les tables créées avec Apache Spark. Lorsqu'elles sont exécutées sur Amazon EMR, les applications Apache Spark peuvent toujours lire ces tables, mais certaines optimisations risquent de ne pas être appliquées et les noms de colonnes distinguant majuscules et minuscules ne sont pas pris en charge. Si l'utilisateur a accès à toutes les colonnes de la table, Lake Formation renvoie la table non modifiée avec toutes les propriétés de la table.

Problème lié au changement de nom d'une colonne exclue

Si vous utilisez des autorisations au niveau des colonnes pour exclure une colonne puis renommer la colonne, celle-ci n'est plus exclue des requêtes, telles que. `SELECT *`

Problème lié à la suppression de colonnes dans les tableaux CSV

Si vous créez une table de catalogue de données au format CSV, puis que vous supprimez une colonne du schéma, les requêtes peuvent renvoyer des données erronées et les autorisations au niveau des colonnes risquent de ne pas être respectées.

Solution : créez plutôt une nouvelle table.

Les partitions de table doivent être ajoutées sous un chemin commun

Lake Formation s'attend à ce que toutes les partitions d'une table suivent un chemin commun défini dans le champ d'emplacement de la table. Lorsque vous utilisez le robot pour ajouter des partitions à un catalogue, cela fonctionne parfaitement. Mais si vous ajoutez des partitions manuellement et que ces partitions ne se trouvent pas à l'emplacement défini dans la table parent, l'accès aux données ne fonctionne pas.

Problème lié à la création d'une base de données pendant la création du flux de travail

Lorsque vous créez un flux de travail à partir d'un plan à l'aide de la console Lake Formation, vous pouvez créer la base de données cible si elle n'existe pas. Dans ce cas, l'utilisateur connecté obtient l'`CREATE_TABLE`autorisation d'accéder à la base de données créée. Cependant, le robot généré par le flux de travail assume le rôle du flux de travail lorsqu'il tente de créer une table. Cela échoue car le rôle n'a pas l'`CREATE_TABLE`autorisation d'accéder à la base de données.

Solution : Si vous créez la base de données via la console lors de la configuration du flux de travail, avant d'exécuter le flux de travail, vous devez accorder au rôle associé au flux de travail l'`CREATE_TABLE`autorisation sur la base de données que vous venez de créer.

Problème lié à la suppression puis à la recréation d'un utilisateur

Le scénario suivant se traduit par des autorisations erronées de Lake Formation renvoyées par `lakeformation:ListPermissions` :

1. Créez un utilisateur et accordez les autorisations de Lake Formation.
2. Supprimez l'utilisateur.
3. Recréez l'utilisateur portant le même nom.

`ListPermissions`renvoie deux entrées, une pour l'ancien utilisateur et une pour le nouvel utilisateur. Si vous essayez de révoquer les autorisations accordées à l'ancien utilisateur, les autorisations sont révoquées pour le nouvel utilisateur.

`GetTables`et **`SearchTables`** les API ne mettent pas à jour la valeur du **`IsRegisteredWithLakeFormation`** paramètre

Il existe une limite connue selon laquelle les opérations de l'API du catalogue de données, telles que la mise à jour `GetTables` et `SearchTables` la non-mise à jour de la valeur de`IsRegisteredWithLakeFormation parameter`, renvoient la valeur par défaut, qui est fausse. Il est recommandé d'utiliser l'`GetTable`API pour afficher la valeur correcte pour`IsRegisteredWithLakeFormation parameter`.

Les opérations de l'API Data Catalog ne mettent pas à jour la valeur du **`IsRegisteredWithLakeFormation`** paramètre

Il existe une limite connue selon laquelle les opérations de l'API du catalogue de données, telles que `GetTables` la mise à jour ou non de la valeur du `IsRegisteredWithLakeFormation` paramètre, renvoient la valeur par défaut, qui est fausse. `SearchTables` Il est recommandé d'utiliser l'`GetTable`API pour afficher la valeur correcte du `IsRegisteredWithLakeFormation` paramètre.

Les opérations de Lake Formation ne prennent pas en charge AWS Glue le registre des schémas

Les opérations de Lake Formation ne prennent pas en charge AWS Glue les tables contenant un `SchemaReference` dans le `StorageDescriptor` à utiliser dans le registre des schémas.

Message d'erreur mis à jour

AWS Lake Formation a mis à jour les exceptions spécifiques aux ressources par rapport au message `EntityNotFound` d'erreur général pour les opérations d'API suivantes afin de répondre aux objectifs de sécurité et de conformité.

- RevokePermissions
- GrantPermissions
- GetResourceBalises LF
- GetTable
- GetDatabase

AWS Lake Formation API

> **Note**
>
> La référence d'API mise à jour pour le AWS Lake Formation service est désormais disponible.

Table des matières

- API d'autorisations
  - Opérations
  - Les types de données
- API de paramètres du lac de données
  - Opérations
  - Les types de données
- API d'intégration d'IAM Identity Center
  - Opérations
  - Les types de données
- API en mode d'accès hybride
  - Opérations
  - Les types de données
- API de vente d'informations d'identification
  - Opérations
  - Les types de données
- API de balisage
  - Opérations
  - Les types de données
- API de filtrage de données
  - Opérations
  - Types de données
- Types de données courants

- ErrorDetail structure
- Modèles de chaîne

API d'autorisations

La section Permissions API décrit les opérations et les types de données requis pour accorder et révoquer des autorisations dans AWS Lake Formation. Consultez le guide de référence de l'API Lake Formation pour toutes les opérations AWS Lake Formation d'API et tous les types de données.

Opérations

- GrantPermissions
- RevokePermissions
- BatchGrantPermissions
- BatchRevokePermissions
- GetEffectivePermissionsForPath
- ListPermissions

Les types de données

- Ressource
- DatabaseResource
- TableResource
- TableWithColumnsResource
- DataCellsFilterResourcee
- DataLocationResource
- DataLakePrincipal
- PrincipalPermissions
- PrincipalResourcePermissions
- DetailsMap
- ColumnWildcard
- BatchPermissionsRequestEntry

- BatchPermissionsFailureEntry

API de paramètres du lac de données

Cette section contient les paramètres du lac de données, les opérations de l'API et les types de données permettant de gérer les administrateurs du lac de données.

Opérations

- GetDataLakeSettings
- PutDataLakeSettings

Les types de données

- DataLakeSettings

API d'intégration d'IAM Identity Center

Cette section contient les opérations de création et de gestion de l'intégration de Lake Formation avec IAM Identity Center.

Opérations

- CreateLakeFormationIdentityCenterConfiguration
- DeleteLakeFormationIdentityCenterConfiguration
- DescribeLakeFormationIdentityCenterConfiguration
- UpdateLakeFormationIdentityCenterConfiguration

Les types de données

- ExternalFilteringConfiguration

API en mode d'accès hybride

La section API du mode d'accès hybride décrit les opérations et les types de données requis pour configurer le mode d'accès hybride dans AWS Lake Formation. Consultez le guide de référence de l'API Lake Formation pour toutes les opérations AWS Lake Formation d'API et tous les types de données.

Opérations

- CreateLakeFormationOptIn
- DeleteLakeFormationOptIn
- ListLakeFormationOptIns

Les types de données

- Ressource
- DatabaseResource
- TableResource
- Informations sur les ressources
- LakeFormationOptInsInfo
- DataLocationResource

API de vente d'informations d'identification

La section API Credential Vending décrit les opérations et les types de données liés à l'utilisation du AWS Lake Formation service pour vendre des informations d'identification et pour enregistrer et gérer une ressource de lac de données.

Opérations

- RegisterResource
- DeregisterResource
- ListResources
- GetUnfilteredTableMetadata

- GetUnfilteredPartitionsMetadata
- GetTemporaryGluePartitionCredentials
- GetTemporaryGlueTableCredentials
- UpdateResource

Les types de données

- FilterCondition
- RowFilter
- ResourceInfo

API de balisage

La section API de balisage décrit les opérations et les types de données liés à une stratégie d'autorisation qui définit un modèle d'autorisation sur les attributs ou les balises de paires clé-valeur.

Opérations

- Ajouter LF TagsToResource
- Supprimer LF TagsFromResource
- GetResourceBalises LF
- Liste des balises LF
- Créer une balise LF
- Obtenir le tag LF
- Mettre à jour le tag LF
- Supprimer le tag ELF
- SearchTablesByBalises LF
- SearchDatabasesByBalises LF

Les types de données

- LF TagKeyResource

- LF TagPolicyResource
- TaggedTable
- TaggedDatabase
- Balise LF
- LF TagPair
- LF TagError
- Colonne LFTAG

API de filtrage de données

Les API de filtre de données décrivent comment gérer les filtres de cellules de données dans AWS Lake Formation.

Opérations

- CreateDataCellsFilter
- DeleteDataCellsFilter
- ListDataCellsFilter
- GetDataCellsFilter
- UpdateDataCellsFilter

Types de données

- DataCellsFilter
- RowFilter

Types de données courants

Les types de données courants décrivent des types de données courants variés dans AWS Lake Formation.

ErrorDetail structure

Contient des informations détaillées sur une erreur.

Champs

- `ErrorCode` – Chaîne UTF-8, d'une longueur comprise entre 1 et 255 octets, correspondant au Single-line string pattern.

  Code associé à cette erreur.
- `ErrorMessage` – Chaîne de description, d'une longueur maximale de 2 048 octets, correspondant au URI address multi-line string pattern.

  Message décrivant l'erreur.

Modèles de chaîne

L'API utilise les expressions régulières suivantes pour définir le contenu valide pour différents paramètres et membres de chaîne :

- Modèle de chaîne à ligne unique – `"[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\t]*"`
- Modèle de chaîne à plusieurs lignes d'adresse URI – `"[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*"`
- Schéma de chaîne personnalisé #3 — « `^\w+\.\w+\.\w+$` »
- Schéma de chaîne personnalisé #4 — « `^\w+\.\w+$` »
- Schéma de chaîne personnalisé #5 — « `arn:aws:iam::[0-9]*:role/.*` »
- Schéma de chaîne personnalisé #6 — « `arn:aws:iam::[0-9]*:user/.*` »
- Schéma de chaîne personnalisé #7 — « `arn:aws:iam::[0-9]*:group/.*` »
- Modèle de chaîne personnalisée #8 – `"arn:aws:iam::[0-9]*:saml-provider/.*"`
- Modèle de chaîne personnalisée #9 – `"^([\p{L}\p{Z}\p{N}_.:\/=+\-@%]*)$"`
- Modèle de chaîne personnalisée #10 – `"^([\p{L}\p{Z}\p{N}_.:\*\/=+\-@%]*)$"`
- Modèle de chaîne personnalisée #11 – `"[\p{L}\p{N}\p{P}]*"`

Régions prises en charge

Cette section contient des informations sur la prise en charge Régions AWS et les fonctionnalités de Lake Formation.

Disponibilité générale

Pour les services Régions AWS pris en charge par AWS Lake Formation, consultez la liste des AWS services disponibles par région.

Pour une liste des points de terminaison du service Lake Formation pour chaque région et des quotas du service Lake Formation, voir AWS Lake Formation Points de terminaison et quotas.

AWS GovCloud (US)

Pour un aperçu des différences entre AWS GovCloud (US) la région et la norme Régions AWS, voir En quoi AWS Lake Formation diffèrent pour AWS GovCloud (US).

Optimisation des transactions et du stockage

Les tables gouvernées, la prise en charge des transactions et les fonctionnalités d'optimisation du stockage pour Lake Formation sont disponibles dans les versions suivantes : Régions AWS

Nom de la région	Paramètre de région	Point de terminaison
US East (Virginie du Nord)	us-east-1	lakeformation.us-east-1.amazonaws.com lakeformation-fips.us-east-1.amazonaws.com
USA Est (Ohio)	us-east-2	lakeformation.us-east-2.amazonaws.com lakeformation-fips.us-east-2.amazonaws.com

Nom de la région	Paramètre de région	Point de terminaison
USA Ouest (Oregon)	us-west-2	lakeformation.us-west-2.amazonaws.com lakeformation-fips.us-west-2.amazonaws.com
Asie-Pacifique (Mumbai)	ap-south-1	lakeformation.ap-south-1.amazonaws.com
Asie-Pacifique (Séoul)	ap-northeast-2	lakeformation.ap-northeast-2.amazonaws.com
Asie-Pacifique (Singapour)	ap-southeast-1	lakeformation.ap-southeast-1.amazonaws.com
Asie-Pacifique (Sydney)	ap-southeast-2	lakeformation.ap-southeast-2.amazonaws.com
Asie-Pacifique (Tokyo)	ap-northeast-1	lakeformation.ap-northeast-1.amazonaws.com
Europe (Francfort)	eu-central-1	lakeformation.eu-central-1.amazonaws.com
Europe (Irlande)	eu-west-1	lakeformation.eu-west-1.amazonaws.com
Europe (Londres)	eu-west-2	lakeformation.eu-west-2.amazonaws.com
Europe (Stockholm)	eu-north-1	lakeformation.eu-north-1.amazonaws.com

Nom de la région	Paramètre de région	Point de terminaison
Canada (Centre)	ca-central-1	lakeformation.ca-central-1.amazonaws.com
Amérique du Sud (São Paulo)	sa-east-1	lakeformation.sa-east-1.amazonaws.com

Historique du document pour AWS Lake Formation

Le tableau suivant décrit les modifications importantes apportées à la documentation de AWS Lake Formation.

Modification	Description	Date
Modification de politique mise à jour	A documenté la modification (ajout d'identifiants de déclaration et suppression des autorisations redondantes) apportée aux AWSLakeFormationDataAdmin politiques AWSLakeFormationCrossAccountManager et.	14 mars 2024
Mise à jour de la configuration de Lake Formation	Les étapes de la AWS Lake Formation section Configuration ont été mises à jour.	7 février 2024
Modification de politique mise à jour	De nouvelles autorisations ont été ajoutées à la politique en ligne du rôle lié au service. Pour plus d'informations, consultez la section Utilisation de rôles liés à un service pour Lake Formation.	7 février 2024
Modification de politique mise à jour	A documenté le changement apporté à la LakeFormationDataAccessServiceRolePolicypolitique.	2 février 2024
Limites de la Formation lacustre consolidée	Création d'une section unifiée pour les limites et les considérations relatives à la Formation des Lacs. Pour plus	15 décembre 2023

	d'informations, consultez la section Limitations relatives aux Lake Formation.	
Ajout de documentation pour le compactage des icebergs	Pour améliorer les performances de lecture des services AWS d'analyse tels qu'Athena et Amazon EMR, ainsi que pour les tâches AWS Glue ETL, AWS Glue Data Catalog fournit un compactage géré (un processus qui compacte de petits objets Amazon S3 en objets plus grands) pour les tables Iceberg du catalogue de données. Pour plus d'informations, consultez Optimisation des tables Iceberg.	25 novembre 2023
Documentation ajoutée pour l'intégration d'IAM Identity Center	Les intégrations d'IAM Identity Center permettent aux utilisateurs et aux groupes d'accéder aux ressources du catalogue de données en appliquant les autorisations de Lake Formation. Pour plus d'informations, consultez la section Intégration d'IAM Identity Center.	25 novembre 2023

Documentation ajoutée pour les vues du catalogue de données	Vous pouvez créer des vues AWS Glue Data Catalog faisant référence à un maximum de 10 tables à l'aide des éditeurs SQL pour Amazon Athena ou Amazon Redshift. Pour plus d'informations, consultez la section Création de vues.	25 novembre 2023
Mise à jour du changement de politique	A documenté le changement apporté à la AWSLakeFormationCrossAccountManagerpolitique.	25 octobre 2023
Documentation ajoutée pour le mode d'accès hybride	Le mode d'accès hybride offre la flexibilité d'activer de manière sélective les autorisations de Lake Formation pour les bases de données et les tables de votre AWS Glue Data Catalog. Avec le mode d'accès hybride, vous disposez désormais d'un chemin incrémentiel qui vous permet de définir les autorisations de Lake Formation pour un ensemble spécifique d'utilisateurs sans interrompre les politiques d'autorisation des autres utilisateurs ou charges de travail existants. Pour plus d'informations, consultez la section Mode d'accès hybride.	26 septembre 2023

Ajout de documentation pour la création de tables Apache Iceberg	Vous pouvez désormais créer des tables Apache Iceberg qui utilisent le format de données Apache Parquet AWS Glue Data Catalog avec des données résidant dans Amazon S3. Pour plus d'informations, consultez la section Création de tables Iceberg.	16 août 2023
Documentation ajoutée pour l'accès aux données entre régions	Lake Formation permet d'interroger les tables du catalogue de données dans toutes les AWS régions. Vous pouvez accéder aux données d'une région depuis d'autres régions à l'aide d'Athena, Amazon EMR, et exécuter l' AWS Glue ETL en créant des liens de ressources dans d'autres régions pointant vers les bases de données et les tables sources. Vous pouvez connecter le catalogue de données à des métastore s externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation Pour plus d'informations, consultez la section Accès aux tables entre les régions.	30 juin 2023

Contenu réorganisé	Les chapitres du guide ont été réorganisés pour correspondre au parcours des utilisateurs de Lake Formation.	15 mai 2023
Ajout de documentation pour la fédération HMS	Vous pouvez connecter le catalogue de données à des métastores externes qui stockent les métadonnées de vos données Amazon S3 et gérer en toute sécurité les autorisations d'accès aux données à l'aide de. AWS Lake Formation Pour plus d'informations, consultez la section Gestion des autorisations sur les ensembles de données qui utilisent des métastores externes.	15 avril 2023
Ajout de documentation pour le partage de données Amazon Redshift	Vous pouvez désormais gérer en toute sécurité les données d'un partage de données depuis Amazon Redshift à l'aide des autorisations de Lake Formation. Lake Formation prend en charge l'accès sous licence à vos données via AWS Data Exchange. Pour plus d'informations, voir Partage de données dans AWS Lake Formation.	30 novembre 2022

Support pour le partage de données entre comptes directement avec les donneurs d'ordre	Ajout d'informations sur le partage de données directement avec les responsables IAM dans un autre compte. Pour plus d'informations, voir Partage de données entre comptes dans AWS Lake Formation.	10 novembre 2022
Support pour le partage de données AWS RAM activé à l'aide de TBAC	Ajout d'informations sur la méthode LF-TBAC d'octroi d'autorisations de catalogue de données utilisée AWS Resource Access Manager pour les autorisations entre comptes.	10 novembre 2022
Ajout d'une section sur le travail avec d'autres services	Informations supplémentaires sur la manière dont AWS des services tels qu'Athena, AWS Glue Redshift Spectrum et Amazon EMR peuvent utiliser Lake Formation pour accéder en toute sécurité aux données des sites Amazon S3 enregistrés auprès de Lake Formation. Pour plus d'informations, consultez Collaboration avec d'autres AWS services.	10 novembre 2022

???	Ajout d'informations sur la résolution d'une erreur lors de l'utilisation d'Amazon EMR pour accéder aux données entre comptes. Pour plus d'informations, consultez Erreur lors de l'utilisation d'Amazon EMR pour accéder aux données partagées via plusieurs comptes.	7 novembre 2022
Mises à jour du partage des ressources entre comptes	Ajout d'une description du fonctionnement des partages de ressources entre comptes dans Lake Formation. A documenté le changement apporté à la AWSLakeFormationCrossAccountManagerpolitique.	6 mai 2022
Nouveaux tutoriels	Ajout de nouveaux didacticiels pour la création de tables gouvernées, la sécurisation des lacs de données et le partage de lacs de données. Pour plus de détails, consultez la section Démarrage.	20 avril 2022

Page d'accueil de New Lake Formation	Mise à jour de la page d'accueil de Lake Formation pour inclure des liens vers des didacticiels fournissant des step-by-step instructions sur la façon de créer un lac de données, d'ingérer des données, de partager et de sécuriser des lacs de données à l'aide de Lake Formation.	20 avril 2022
Support pour la vente d'informations d'identification	Ajout d'informations sur la vente d'informations d'identif ication, qui permettent à Lake Formation de permettre à des services tiers de s'intégrer à Lake Formation en utilisant les opérations de l'API de vente d'informations d'identification. Pour plus d'informations, consultez Comment fonctionn e le distributeur automatique d'informations d'identification à Lake Formation.	28 février 2022

Support pour les tables gouvernées et le filtrage avancé des données	Ajout d'informations sur les tables gouvernées, qui prennent en charge les transactions ACID, le compactage automatique des données et les requêtes de voyage dans le temps. Ajout d'informations sur la création de filtres de données pour prendre en charge la sécurité au niveau des colonnes, la sécurité au niveau des lignes et la sécurité au niveau des cellules. Pour plus d'informations, consultez les sections Tables gouvernées dans la Lake Formation et Filtrage des données et Sécurité au niveau des cellules dans la Lake Formation.	30 novembre 2021
Support pour les points de terminaison d'interface VPC	Ajout d'informations sur la création d'un point de terminaison d'interface de cloud privé virtuel (VPC) pour Lake Formation, afin que les communications entre votre VPC et Lake Formation soient effectuées de manière complète et sécurisée au sein du réseau. AWS Pour plus d'informations, consultez la section Utilisation de Lake Formation avec des points de terminaison VPC.	11 octobre 2021

Support des politiques de point de terminaison d'un VPC	Ajout d'informations sur la prise en charge des politiques relatives aux points de terminaison Virtual Private Cloud (VPC) dans Lake Formation. Pour plus d'informations, consultez la section Utilisation de Lake Formation avec des points de terminaison VPC.	11 octobre 2021
Support pour le contrôle d'accès basé sur des balises	Le contrôle d'accès basé sur les balises Lake Formation fournit une nouvelle méthode plus évolutive pour gérer l'accès aux ressources du catalogue de données et aux données sous-jacentes à l'aide de balises LF. Pour plus d'informations, consultez la section Contrôle d'accès basé sur des balises Lake Formation.	7 mai 2021
Nouvelle exigence d'inscription pour le filtrage des données sur Amazon EMR.	Ajout d'informations concernant l'obligation d'opter pour autoriser Amazon EMR à filtrer les données gérées par Lake Formation. Pour plus d'informations, consultez Autoriser le filtrage des données sur Amazon EMR.	9 octobre 2020

Support pour l'octroi d'autorisations complètes entre comptes sur les bases de données du catalogue de données	Ajout d'informations sur l'octroi d'autorisations complètes à Lake Formation sur les bases de données du catalogue de données pour tous AWS les comptes, notamment `CREATE_TABLE` . Pour plus d'informations, voir Partage de bases de données de catalogues de données.	1er octobre 2020
Support pour l'authentification Amazon Athena des utilisateurs via SAML.	Ajout d'informations sur la prise en charge des utilisateurs d'Athena qui se connectent via le pilote JDBC ou ODBC et s'authentifient via des fournisseurs d'identité SAML tels qu'Okta et Microsoft Active Directory Federation Service (AD FS). Pour plus d'informations, voir Intégrations de AWS services avec Lake Formation.	30 septembre 2020
Support pour l'accès entre comptes avec un catalogue de données crypté	Ajout d'informations sur l'octroi d'autorisations entre comptes lorsque le catalogue de données est crypté. Pour plus d'informations, consultez la section Conditions préalables relatives à l'accès entre comptes.	30 juillet 2020

Support pour l'accès entre comptes au lac de données	Ajout d'informations sur l'octroi d' AWS Lake Formation autorisations sur les bases de données et les tables du catalogue de données à AWS des comptes et organisat ions externes, et sur l'accès aux objets du catalogue de données partagés à partir de comptes externes. Pour plus d'informations, consultez la section Accès entre comptes.	7 juillet 2020
Intégration avec Amazon QuickSight	Ajout d'informations sur la façon d'accorder des autorisations Lake Formation aux utilisateurs QuickSight d'Amazon Enterprise Edition afin qu'ils puissent accéder aux ensembles de données résidant dans des sites Amazon S3 enregistrés. Pour plus d'informations, consultez la section Octroi d'autorisations au catalogue de données.	29 juin 2020
Mises à jour des chapitres relatifs à la configuration et à la mise en route	Réorganisation et amélioration des chapitres Configuration et Démarrage. Les autorisations recommandées AWS Identity and Access Management (IAM) pour l'administrateur du lac de données ont été mises à jour.	27 février 2020

Support pour AWS Key Management Service	Ajout d'informations sur la façon dont le support de Lake Formation pour AWS Key Management Service (AWS KMS) simplifie la configuration de services intégrés pour lire et écrire des données chiffrées sur des sites Amazon Simple Storage Service (Amazon S3) enregistrés. Ajout d'informations sur la procédure d'enregistrement des sites Amazon S3 chiffrés avec AWS KMS keys. Pour plus d'informations, consultez the section called "Ajouter un emplacement Amazon S3 à votre lac de données".	27 février 2020
Mises à jour des plans et des politiques IAM des administrateurs de lacs de données	Paramètres d'entrée clarifiés pour les plans de base de données incrémentiels. Mise à jour des politiques IAM requises pour un administrateur de lac de données.	20 décembre 2019
Réécriture des chapitres sur la sécurité et révisions des chapitres de mise à niveau	Amélioration des chapitres relatifs à la sécurité et à la mise à niveau.	29 octobre 2019
La super autorisation remplace toutes les autorisations	Mise à jour des chapitres sur la sécurité et la mise à niveau pour refléter le remplacement de l'autorisation `All` par`Super`.	10 octobre 2019

Ajouts, corrections et clarifications	A apporté des ajouts, des corrections et des clarifications en fonction des commentaires. Révision du chapitre sur la sécurité. Mise à jour des chapitres sur la sécurité et la mise à niveau pour refléter le remplacement du groupe `Everyone` par`IAMAllowedPrincipals` .	11 septembre 2019
Nouveau guide	Il s'agit de la première version du Guide du développeur AWS Lake Formation .	8 août 2019

AWS Glossaire

Pour la AWS terminologie la plus récente, consultez le AWS glossaire dans la Glossaire AWS référence.

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.